Publicaciones Electrónicas
Sociedad Matemática Mexicana

Álgebra Lineal y Geometría I

Néstor Thome Coppo

www.smm.org.mx

Serie: Textos. Vol. 27 (2023)

ISBN: 978-607-8008-19-3

Álgebra Lineal y Geometría I

Néstor Thome Coppo

Catedrático de Universidad
Departamento de Matemática Aplicada
Instituto Universitario de Matemática Multidisciplinar
Universitat Politècnica de València, España
njthome@mat.upv.es

Publicaciones Electrónicas
Sociedad Matemática Mexicana

Índice general

Prólogo

El término *Matemáticas* fue acuñado por Pitágoras en el siglo IV antes de Cristo para indicar "lo que se conoce", "lo que se aprende".

El término *Álgebra* proviene de un libro persa titulado *Al-jebr w'al-muqâbalah* que significa: trasposición (coloquialmente, el pasaje de términos de un miembro a otro de una ecuación) y eliminación (la cancelación de términos iguales en ambos miembros). Fue escrito en Bagdad en el siglo IX (alrededor del año 825) por el matemático Mohammed ibn Mûsâ al-Khwârizmî, quien mostró en sus trabajos la primera fórmula general para la resolución de ecuaciones de primer y segundo grado; operó formalmente con símbolos permitiendo la abstracción de tipos de números concretos. El Álgebra fue considerada, hasta finales del siglo XIX, la ciencia que trataba la resolución de ecuaciones. En la segunda mitad del XIX, con el trabajo de Évariste Galois, nació el *Álgebra Abstracta*. Si bien mantenía el espíritu de la resolución de ecuaciones, el enfoque cambió principalmente al estudio de las *estructuras algebraicas* (como grupos, anillos, cuerpos, etc.) que se definen a partir de un pequeño número de *axiomas*[1].

Por su parte, la *Geometría* (*Sintética*[2]) tuvo su origen con los *Elementos*, que es una colección de trece libros escrita en torno al año 300 a.C. por el matemático griego Euclides, quien desarrolló su magnífica labor en

[1] Los *axiomas* son proposiciones evidentes que se aceptan en una teoría sin requerir demostración previa y de los que se deducen todos los resultados de dicha teoría.

[2] Llamada así en contraposición a la posterior Geometría Analítica que llevó a la dualidad análisis-síntesis.

Alejandría (Egipto). La obra cubre tanto geometría elemental (geometría plana y geometría de los cuerpos sólidos) como aritmética (razones, proporciones y teoría de números). En el siglo XVII, a partir del trabajo de René Descartes (Discurso del método, 1637) y Pierre de Fermat, se produjo una revolución científica al poder conectar el Álgebra con la Geometría (mediante la introducción de los sistemas de coordenadas), lo que originó lo que el propio Descartes llamó *Geometría Analítica*, que hoy se conoce como *Geometría Cartesiana*[3]. Informalmente, en el plano, la idea básica fue asociar una ecuación con una *curva*; consistía por tanto, en que dos variables se pudieran relacionar a través de una *función*, que se representaría gráficamente como una curva, donde el concepto de *variable* permitía tratar cantidades que no eran fijas. De este modo, sería entonces posible resolver el problema geométrico de la intersección de curvas mediante el problema algebraico de la resolución de sistemas de ecuaciones.

Las primeras incursiones que se conocen del *Álgebra Lineal* aparecen en el papiro Rhind (actualmente algunos fragmentos se encuentran en el Museo Británico y otros en el Museo de Brooklyn) que data del año 1650 a.C., fue escrito por un sacerdote egipcio, y donde se consideran ecuaciones de primer grado.

Las ecuaciones lineales (o de primer grado con n incógnitas) son relaciones de la forma

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b,$$

donde $a_1, a_2, \ldots, a_n$ son los coeficientes de la ecuación, b es el término independiente y $x_1, x_2, \ldots, x_n$ son las incógnitas. Un conjunto finito de ecuaciones de este tipo determina un sistema de ecuaciones lineales que es una de las herramientas básicas del Álgebra Lineal. Para su estudio se consideran: el método de eliminación de Gauss, el rango de una matriz, determinantes, independencia lineal, etc.

[3]En la actualidad, el nombre *Geometría Analítica* comprende no sólo la Geometría Cartesiana sino también la Geometría Diferencial (iniciada con Gauss) y la más reciente Geometría Algebraica.

Un gran cambio se produce en el Álgebra Lineal a partir de una observación realizada por Jean le Rond d'Alembert, matemático francés de la época de la Revolución Francesa. El hecho es que el conjunto solución de un sistema de ecuaciones lineales compatible forma una *variedad lineal*, que se estudiará en profundidad más adelante, en el tema de *espacios afines*. Otra observación realizada por d'Alembert (a la vez que Euler y Lagrange) fue que cuando dicho sistema es homogéneo, sus soluciones pueden expresarse como combinación lineal de algunas soluciones particulares. Se estaba a las puertas del descubrimiento del concepto de *vector* y *espacio vectorial* o *espacio lineal*, que forman los cimientos del Álgebra Lineal. Los primeros[4] se producen con William Hamilton[5] y Arthur Cayley[6] y la noción de espacio vectorial la introduce Hermann Günther Grassmann[7], quien suele ser considerado el padre del Álgebra Lineal. Por su parte, Cayley es considerado el fundador de la *Teoría de Matrices*. Curiosamente, el primero en usar el término *matriz* fue el matemático inglés Sylvester en 1850, mientras que, los *determinantes* datan del siglo II a.C. y fueron utilizados en sus orígenes por matemáticos chinos. Carl Friedrich Gauss[8] utilizó por primera vez el término *determinante* en sus Disquisitiones Arithmeticae publicadas en 1801.

De este modo, se han dado algunas pinceladas sobre las áreas de las Matemáticas de las cuales, en este libro, apenas se estudiarán algunos de sus conceptos elementales.

En la actualidad, el Álgebra Lineal se considera parte esencial de los requisitos formativos necesarios no sólo para futuros matemáticos, sino también para estudiantes de física, ingenierías en general, informática, economía, así como también es necesario en algunas ramas de la sociología, biología, etc. Esto acredita su importancia y es razonable predecir que poseerá innumerables

[4] A partir de los cuaterniones, precursores de los vectores n-dimensionales como se utilizan hoy en día.

[5] Matemático, físico y astrónomo irlandés.

[6] Matemático británico.

[7] Polímata alemán.

[8] Matemático alemán conocido como el Príncipe de las Matemáticas.

aplicaciones, como efectivamente ocurre. El libro está destinado a estudiantes del doble grado en Matemáticas con diferentes Ingenierías de la Universitat Politècnica de València.

En este libro se presentarán métodos matriciales básicos para resolver diferentes problemas del Álgebra Lineal, donde algunas de sus interpretaciones geométricas se estudian a partir de la Geometría Cartesiana. Se ha intentado buscar el equilibrio entre la sencillez en la presentación de los resultados y el rigor en los enunciados y sus demostraciones, ambos son aspectos fundamentales en la formación en matemáticas de un estudiante. Los resultados teóricos se interpretan mediante un amplio número de ejemplos resueltos. El libro contiene una serie de ejercicios propuestos al finalizar cada capítulo, de modo que el lector pueda familiarizarse con los temas de forma gradual.

Por una cuestión de completitud, se han incluido también algunos resultados (como los Anexos A, B, C, D y E) en los que no es necesario profundizar en una primera lectura del libro.

La intención, al presentar una gran cantidad y variedad de ejemplos resueltos, es la de ilustrar los conceptos que se desarrollan mediante una comprensión más rápida y sencilla de los mismos. Por otra parte, con el propósito de incentivar el espíritu crítico, el libro incluye una amplia cantidad de comentarios, complementos y observaciones que permitan al lector reflexionar sobre los conceptos fundamentales y discusiones constructivas acerca de la motivación de los mismos.

El temario (principal) del presente libro ha sido estructurado en un total de 11 capítulos.

En el capítulo 1 se presentan preliminares, conceptos básicos necesarios para comprender la forma en que se desarrolla, se piensa, y se debe proceder en Matemáticas. Una breve introducción a los métodos de demostración, a la teoría de conjuntos, al concepto de función y nociones elementales de estructuras algebraicas.

En el capítulo 2 se hace un repaso de cuestiones conocidas de matrices, se establecen las definiciones y notaciones a utilizar, siendo importante poner el enfoque, para una buena formación del alumno, en las técnicas de demostración utilizadas para probar las propiedades. Es probable que, además de la definición de matriz, los únicos hechos novedosos sean la forma de operar con matrices particionadas en bloques y la generalización, a matrices de tamaños arbitrarios, de algunos resultados conocidos.

En el capítulo 3 se sistematiza el método de eliminación gaussiana, se prueba con detalle que dicho método funciona y, a través de dicho método, se realiza la clasificación de los sistemas lineales.

El capítulo 4 está dedicado al estudio de la forma escalonada reducida de una matriz, garantizando su existencia y unicidad, lo que permitirá establecer el concepto de rango de una matriz.

En el capítulo 5 se aborda la definición, propiedades y algunas caracterizaciones del concepto de matriz inversa.

El capítulo 6 comienza con el estudio de matrices elementales y su utilización para el cálculo de la inversa (método de Gauss-Jordan). Tras establecer caracterizaciones de la equivalencia de matrices, equivalencia por filas y equivalencia por columnas, se prueba que determinan relaciones de equivalencia y se encuentra la forma normal de Hermite (como formas canónicas de clases de equivalencia en la equivalencia de matrices) y el rango (como un invariante[9] de esta relación).

En el capítulo 7 se aborda el concepto de determinante, introducido de forma recursiva. Se establecen y se demuestran sus propiedades por el método de inducción. Se analiza su relación con la existencia de la matriz inversa y se estudia la fórmula de Cauchy-Binet para el determinante de un producto de matrices. Como aplicaciones se demuestran la regla de Cramer, la fórmula que permite calcular la inversa de una matriz y se estudia la forma de utilizarlo para determinar el rango de una matriz. También se presentan aplicaciones a tipos concretos de matrices (por bloques, de Vandermonde, de compañía).

En el capítulo 8 se aborda uno de los temas centrales del Álgebra Lineal: los espacios vectoriales. Una vez definidos, se estudian los subespacios,

---

[9]En realidad, como un conjunto completo de invariantes.

las combinaciones lineales, subespacios generados y operaciones con subespacios. Una segunda parte del tema aborda los sistemas de generadores, la dependencia/independencia lineal, y el concepto de base que permite definir la dimensión de un espacio vectorial.

Con la intención de destacar la verdadera importancia que la introducción de coordenadas le confiere al tema de espacios vectoriales, se ha preferido dedicar el capítulo 9 completo a este tema, donde se caracterizan los espacios vectoriales isomorfos a partir de la dimensión, y se presenta este concepto como un invariante[10] de una relación de equivalencia adecuada. Se aplica la teoría desarrollada previamente para determinar la estructura del conjunto solución de un sistema de ecuaciones lineales homogéneo, encontrando las ecuaciones paramétricas y cartesianas de un subespacio.

Otro capítulo importante es el capítulo 10, donde se estudian los espacios euclídeos. Se introducen los espacios vectoriales con producto escalar y sobre ellos el concepto de norma y distancia inducidas y ángulo entre vectores. Se desarrolla con detalle el apartado de ortogonalidad y el método de Gram-Schmidt para el cálculo de bases ortogonales, con lo que se prueba la existencia de bases ortogonales en espacios euclídeos de dimensión finita. Por último, se caracterizan los espacios euclídeos de dimensión finita y se prueba que los productos escalares están determinados por las matrices de Gram.

El capítulo 11 está dedicado al estudio de los cuatro subespacios fundamentales asociados a una matriz. Se estudian sus dimensiones, se prueba el Teorema fundamental del rango y su relación con la equivalencia de matrices. Se muestra cómo los espacios euclídeos $\mathbb{R}^n$ y $\mathbb{R}^m$ pueden ser representados a partir de los cuatro subespacios fundamentales asociados a una matriz $A \in \mathbb{R}^{m\times n}$ en el llamado Teorema fundamental del Álgebra Lineal. El libro finaliza con tres resúmenes importantes. Uno se debe a la equivalencia de matrices donde se presentan todos los resultados obtenidos, un resultado que resume más de 20 caracterizaciones, estudiadas en capítulos previos, de la existencia de la inversa de una matriz, hecho clave en toda la Teoría de

[10] En realidad, como un conjunto completo de invariantes.

Matrices y en el Álgebra Lineal en general.

A continuación se presenta el Anexo A, donde se introduce el concepto de sumatorio y sus propiedades. El Anexo B trata sobre operaciones elementales en el cuerpo de los números complejos. Para el estudiante interesando en profundizar, en un Anexo C, se establece una definición rigurosa de polinomio y su relación con la función polinomial. El Apéndice D está dedicado a la demostración del Teorema de intercambio de Steinitz. Finalmente, en el Anexo E, se prueba la infinitud de la dimensión del espacio vectorial $\mathbb{R}$ sobre el cuerpo de escalares $\mathbb{Q}$ utilizando el Teorema fundamental de la Artimética.

Como cuestiones de estilo, se indica que los ejemplos comienzan a resolverse a partir del símbolo $\lhd$ y terminan con el símbolo $\rhd$. Las demostraciones de las proposiciones, teoremas, corolarios, etc. terminan con el símbolo $\square$.

Las fotografías que aparecen en este libro son de dominio público y la mayoría de ellas han sido tomadas de Wikipedia.

Me gustaría manifestar un especial agradecimiento a mis compañeros y amigos Laura Rueda, Marina Lattanzi y José Mas Marí por las enriquecedoras discusiones que hemos mantenido sobre temas relativos al libro y a todos los alumnos que, con sus comentarios, han permitido mejorar las versiones previas a esta.

Se agradecerá cualquier sugerencia o comentario que sea enviada al correo electrónico `njthome@mat.upv.es`.

El autor.

No es porque las cosas sean difíciles por lo que no nos atevemos;
sino que por no atrevernos se hacen difíciles. *Séneca.*

Tabla de símbolos

$\mathbb{N}$	conjunto de los números naturales: $\{1, 2, 3, \dots\}$
$\mathbb{Z}$	conjunto de los números enteros
$\mathbb{Q}$	conjunto de los números racionales
$\mathbb{R}$	conjunto de los números reales
$\mathbb{R}^+$	conjunto de los números reales positivos
$\mathbb{C}$	conjunto de los números complejos
$i\,\mathbb{R}$	conjunto de los números imaginarios puros
$\mathbb{I}_k = \{1, 2, \dots, k\}$	intervalo natural inicial
$\emptyset$	conjunto vacío
$:$ o bien $/$	tal que
$\in$	pertenece a
$\sim$	no
$\exists$	existe
$\forall$	para todo
$\wedge$	y
$\vee$	o (incluyente)
$\Rightarrow$	implica
$\Leftrightarrow$	si y sólo si
$\oplus$	suma directa
$\perp$	ortogonal
$\oplus^{\perp}$	suma directa ortogonal

Capítulo 1

Preliminares

Índice

1.1. Introducción

Con la intención de poner por escrito las ideas y los razonamientos que se realizan en el ámbito de las Matemáticas surge la necesidad de disponer de un lenguaje preciso. Para ello, en esta breve introducción, se darán unas nociones básicas de lógica proposicional. Estas pinceladas permitirán establecer el lenguaje básico que será necesario para abordar las definiciones, los enunciados de lemas, teoremas, etc. y poder comprender las demostraciones realizadas mediante razonamientos lógicos que se presentarán a lo largo de este libro.

1.2. Lógica proposicional

La **Lógica** surgió como una rama de la **Filosofía** que, como ciencia formal, estudia los principios y las estructuras del pensamiento. A pesar de haber vestigios en China e India, el filósofo, lógico y científico de la Antigua Grecia conocido como Aristóteles (384 a.C.-322) es considerado el fundador de la Lógica pues fue el primero en formalizar el sistema lógico.

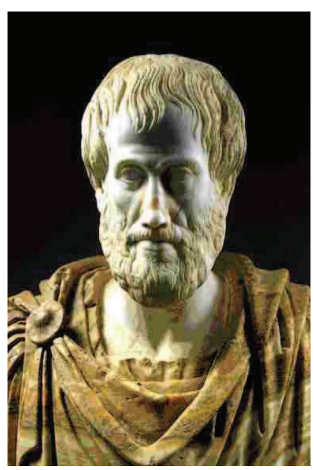


Figura 1.1: Busto de Aristóteles.

En la majestuosa obra renacentista "La Escuela de Atenas" del pintor Rafael Sanzio (entre 1510 y 1512) localizada en los Museos Vaticanos (véase la Figura 1.2), pueden apreciarse algunos personajes más de la antigua Grecia con una trascendencia indiscutible en Matemáticas como Zenón de Elea, Pitágoras de Samos, Hipatia de Alejandría, Sócrates, Platón, Aristóteles, Ptolomeo, Euclides o Arquímedes de Siracusa, entre otros.

Figura 1.2: La Escuela de Atenas.

Retornando a la lógica, su utilidad en **Matemáticas** consisite en considerarla como un sistema que permite verificar si un razonamiento es correcto o no.

Para abordar los métodos que permiten realizar la demostración de la verdad o falsedad de un enunciado, es necesario introducir algunos conceptos tales como proposiciones, conectivos lógicos, cuantificadores, etc.

1.2.1. Proposiciones

Una **proposición** es una oración de la cual se puede dar un **valor de verdad**, es decir, se puede indicar si es verdadera (V) o falsa (F).

Una proposición **simple** está formada por una oración simple.

Ejemplo 1.1. *Se consideran las siguientes oraciones:*

p: "Aristóteles fue un filósofo griego",
q: "4 es impar",
r: "¿Cuál es tu nombre?".

Indicar si son o no proposiciones simples y, en caso de serlo, establecer su valor de verdad.

$\triangleleft$ Las oraciones p y q son proposiciones y la oración r no lo es. El valor de verdad de la proposición p es verdadero (V) y el de la proposición q es falso (F). Esto se indica mediante la notación

$$v(p) = \mathrm{V},$$

mientras que

$$v(q) = \mathrm{F}.$$

$\triangleright$

Es posible operar con proposiciones simples. Para unir dos o más proposiciones simples, lo que formará una proposición **compuesta**, se necesitan **conectivos lógicos**. Los conectivos lógicos que se utilizarán son: "no", "y", "o", "implica" y "si y sólo si". En la Tabla 1.1 se indica la notación para cada uno de ellos.

La herramienta que permite determinar el valor de verdad de una proposición compuesta se conoce como **tabla de verdad**, que debe incluir todas las posibles variaciones de los valores de verdad de cada proposición simple.

A continuación se define cada uno de los conectivos.

Nombre	Conectivo lógico	Notación	Operación
Negación	no	$\sim$	$\sim p$
Conjunción	y	$\wedge$	$p \wedge q$
Disyunción	o	$\vee$	$p \vee q$
Implicación	si $\cdots$ entonces $\cdots$	$\Rightarrow$	$p \Rightarrow q$
Equivalencia	si y sólo si	$\Leftrightarrow$	$p \Leftrightarrow q$

Tabla 1.1: Conectivos básicos.

Negación: Es un conectivo que cambia el valor de verdad de una proposición. Consiste en anteponer "no" a la proposición.

Ejemplo 1.2. *La negación de la proposición p del Ejemplo 1.1 es*

$\sim p$: *Aristóteles no fue un filósofo griego.*

Conjunción: Vincula dos proposiciones mediante el conectivo lógico "y".

Ejemplo 1.3. *La conjunción de las proposiciones p y q del Ejemplo 1.1 es*

$p \wedge q$: *Aristóteles fue un filósofo griego y 4 es impar.*

Disyunción: Vincula dos proposiciones mediante el conectivo lógico "o".

Ejemplo 1.4. *La disyunción de las proposiciones p y q del Ejemplo 1.1 es*

$p \vee q$: *Aristóteles fue un filósofo griego o 4 es impar.*

Cabe destacar que el conectivo "o" empleado en Matemáticas es incluyente. Es decir, $p \vee q$ significa que se cumple p, se cumple q o bien se cumplen ambas.

Ejemplo 1.5. *Representar mediante operaciones lógicas la proposición compuesta*

$$s\text{: “Aristóteles fue un filósofo griego o 4 no es impar”.}$$

$\lhd$ La proposición compuesta s está formada a partir de las proposiciones simples p y q del Ejemplo 1.1. Se debe escribir: $p \vee \sim q$. $\rhd$

Implicación: Vincula dos proposiciones mediante el conectivo lógico

“si $\cdots$ entonces $\cdots$”,

y se representa por medio de la expresión $p \Rightarrow q$, que se lee “p implica q” o “si p entonces q”, donde p se llama **antecedente** y q **consecuente**.

Otras formas de leer p implica q son:

(a) q si p,

(b) p sólo si q,

(c) p es condición suficiente para q, o bien

(d) q es condición necesaria para p.

Si bien las tablas de verdad se definen de manera arbitraria, se han construido a partir del sentido común. El siguiente ejemplo dará sentido a la tabla de verdad correspondiente a la implicación (véase la Tabla 1.2 de la página 29). Se considera la proposición compuesta:

“Si el estudiante aprueba Álgebra entonces su padre le hará un regalo”,

formada a partir de las proposiciones simples

s: el estudiante aprueba Álgebra,
t: su padre le hará un regalo.

Simbólicamente, se puede escribir como $s \Rightarrow t$. Una vez acabado el curso, se presentan cuatro situaciones posibles:

- el estudiante ha aprobado Álgebra (V) y su padre le ha hecho un regalo (V),
- el estudiante ha aprobado Álgebra (V) y su padre no le ha hecho un regalo (F),
- el estudiante no ha aprobado Álgebra (F) y su padre le ha hecho un regalo (V),
- el estudiante no ha aprobado Álgebra (F) y su padre no le ha hecho un regalo (F).

Ahora se trata de asignar un valor de verdad a cada uno de los cuatro casos. La proposición compuesta $s \Rightarrow t$ puede verse como un compromiso condicionado por s.

En el primer caso, en que ambos cumplen, se asignará a la implicación el valor V; no hay dudas. En el segundo caso, como el estudiante ha cumplido el compromiso pero su padre no, se asignará el valor F a la implicación; tampoco hay dudas. Como en los dos últimos casos el estudiante no ha aprobado Álgebra, su padre se ve liberado de hacerle un regalo (puede hacerlo: V, o no: F) y, en ambos casos, ante la duda, se asignará el valor V a la implicación.

En conclusión, de la tabla de verdad correspondiente a la implicación, es posible interpretar que un antecedente verdadero no puede implicar un consecuente falso (pues la única F que aparece en la columna de $p \Rightarrow q$ de la Tabla 1.2 se produce en la segunda fila, es decir, el valor de verdad asignado a $p \Rightarrow q$ es F únicamente si p es V y q es F).

Sin embargo, la observación más importante de dicha tabla de verdad es que: si $p \Rightarrow q$ es V, la validez de p “implica” la validez de q, y de ahí su nombre.

El hecho anterior permite dar la siguiente *regla de inferencia*:

Modus ponens[1]:

$$\text{Si } p \Rightarrow q \text{ es V y } p \text{ es V entonces } q \text{ es V.}$$

En la columna de $[(p \Rightarrow q) \wedge p] \Rightarrow q$ de la Tabla 1.2 se aprecia que sus valores de verdad son todos V, independientemente de los valores que toman las proposiciones que la conforman[2] (esto se llama **tautología**).

Otra *regla de inferencia* que resultará de interés es la siguiente.

Modus tollens[3]:

$$\text{Si } p \Rightarrow q \text{ es V y } q \text{ es F entonces } p \text{ es F.}$$

Otra observación relativa al condicional que se aprecia en la Tabla 1.2 es que si $p \Rightarrow q$ es V y q es V entonces nada puede asegurarse acerca de p, puede ser V o F.

Una última observación sobre el condicional "si p entonces q" es que presenta la misma tabla de verdad que la disyunción $\sim p \vee q$, y pueden tomarse como fórmulas **lógicamente equivalentes**, lo que a veces se denota mediante (véase el Ejercicio 1d de la página 94):

$$(p \Rightarrow q) \quad \equiv \quad (\sim p \vee q).$$

Equivalencia: Se llama **bicondicional**, **equivalencia** o **doble implicación** a la operación lógica que vincula dos proposiciones mediante el conectivo lógico

"si y sólo si",

[1] En latín, *modus ponendo ponens* significa: "el modo que, al afirmar, afirma".

[2] En Lógica se suele distinguir entre el condicional (en sentido material) "si p entonces q" y la implicación "$p \Rightarrow q$". Cuando el condicional es lógicamente verdadero, puede decirse que antecedente implica consecuente (según la regla del modus ponens), que es la forma en que se utiliza en Matemáticas.

[3] En latín, *modus tollendo tollens* significa: "el modo que, al negar, niega".

y se representa por medio de la expresión $p \Leftrightarrow q$.

Se ha indicado que otra forma de leer p sólo si q es $p \Rightarrow q$. Si a esta expresión se añade p si q, que se ha indicado que corresponde a $q \Rightarrow p$, se tiene p si y sólo si q. Por ello, una formulación lógicamente equivalente a $p \Leftrightarrow q$ es $(p \Rightarrow q) \wedge (q \Rightarrow p)$, que se comprueba mediante la igualdad de sus tablas de verdad.

Las tablas de verdad de los conectivos anteriores se muestran en la Tabla 1.2. Las siguientes observaciones, representadas en color rojo en la tabla, se destacan como ejemplos representativos:

- $v(p \wedge q) =$ V se tiene únicamente cuando los valores de verdad de ambas proposiciones son verdaderos.

- $v(p \vee q) =$ F se tiene únicamente cuando los valores de verdad de ambas proposiciones son falsos.

- $v(p \Rightarrow q) =$ F se tiene únicamente cuando el antecedente es verdadero y el consecuente es falso.

- $v(p \Leftrightarrow q) =$ V se tiene únicamente cuando los valores de verdad de ambas proposiciones son iguales.

p	q	$\sim p$	$p \wedge q$	$p \vee q$	$p \Rightarrow q$	$p \Leftrightarrow q$	$[(p \Rightarrow q) \wedge p] \Rightarrow q$
V	V	F	V	V	V	V	V
V	F	F	F	V	F	F	V
F	V	V	F	V	V	F	V
F	F	V	F	F	V	V	V

Tabla 1.2: Tabla de verdad de los conectivos básicos y modus ponens.

1.2.2. Métodos de demostración

Las proposiciones en Matemáticas se pueden expresar siempre en la forma del condicional $p \Rightarrow q$, que se llama condicional **directo**, donde el antecedente pasa a llamarse **hipótesis** y el consecuente pasa a llamarse **tesis**. Es decir,

$$\underset{\textbf{hipótesis}}{p} \quad \Rightarrow \quad \underset{\textbf{tesis}}{q}$$

En realidad, algunas proposiciones pueden expresarse en la forma $p \Leftrightarrow q$, pero este caso se reduce al equivalente $(p \Rightarrow q) \wedge (q \Rightarrow p)$, por lo que sólo se analizará cómo demostrar un condicional directo $p \Rightarrow q$. Asociados a este condicional se tienen:

(a) **recíproco**: $q \Rightarrow p$,

(b) **contrarrecíproco**: $\sim q \Rightarrow \sim p$.

También se nombra **contrario** a $\sim p \Rightarrow \sim q$, pero en general carece de utilidad (por ser el contrarrecíproco de $q \Rightarrow p$).

A partir de su tabla de verdad, es posible probar que la proposición directa y su implicación contrarrecíproca son lógicamente equivalentes, es decir:

$$(p \Rightarrow q) \Leftrightarrow (\sim q \Rightarrow \sim p).$$

Esto significa que si la implicación directa es V entonces su contrarrecíproca también lo es y que si la contrarrecíproca es V su directa también lo es. Este hecho justifica el siguiente ejemplo.

Ejemplo 1.6. *Si $a \in \mathbb{Z}$, es lógicamente equivalente demostrar "a impar $\Rightarrow a^2$ impar" que demostrar "a^2 par $\Rightarrow a$ par".*

Sin embargo, si la implicación directa es V, no es posible afirmar la validez de su contraria ni de su recíproca.

Otro hecho a tener en cuenta es la **negación de una implicación**:

$$\sim (p \Rightarrow q) \Leftrightarrow (p \wedge \sim q),$$

que a continuación servirá para establecer el método del absurdo y la refutación (véanse también el Ejercicio 1 (apartado (1f)) y el Ejercicio 2).

En resumen, para demostrar que la proposición compuesta

$$p \Rightarrow q \text{ es verdadera,}$$

se dispone de los siguientes **métodos de demostración**:

- **Método directo**: Suponer que $v(p) = \mathrm{V}$ y establecer (mediante razonamientos válidos dentro de la teoría) que $v(q) = \mathrm{V}$.

- **Métodos indirectos**: Hay dos.

  - **Contrarrecíproco**: Suponer que $v(q) = \mathrm{F}$ y establecer (mediante razonamientos válidos dentro de la teoría) que $v(p) = \mathrm{F}$.

  - **Contradicción** o **reducción al absurdo**: Suponer que $v(p \Rightarrow q) = \mathrm{F}$ (es decir, $v(p) = \mathrm{V}$, $v(q) = \mathrm{F}$ y, por tanto, $v(p \wedge \sim q) = \mathrm{V}$) y establecer (mediante razonamientos válidos dentro de la teoría) que $v(p \wedge \sim q) = \mathrm{F}$, lo que lleva a una contradicción.

El método dado a continuación permite mostrar que el valor de verdad de una proposición es F.

Refutación: Para comprobar que el valor de verdad de una proposición $p \Rightarrow q$ es F se debe proporcionar un **contraejemplo**, es decir un ejemplo que muestre que $v(\sim (p \Rightarrow q)) = v(p \wedge \sim q) = \mathrm{V}$. Esto último se da si $v(p) = \mathrm{V}$ y $v(q) = \mathrm{F}$.

El próximo método permite mostrar que el valor de verdad de una proposición es V.

Demostración vacua o **vacía**: Para comprobar que el valor de verdad de una proposición $p \Rightarrow q$ es V alcanza con mostrar que siempre se cumple que $v(p) = \mathrm{F}$.

A continuación se indica un método que permite demostrar la unicidad de cierto elemento.

Unicidad: Se supondrá que hay dos elementos que satisfacen la condición que se esté tratando, y se llegará a demostrar que ambos son iguales. Por tanto, sólo podrá haber un elemento que satisfaga dicha condición[4].

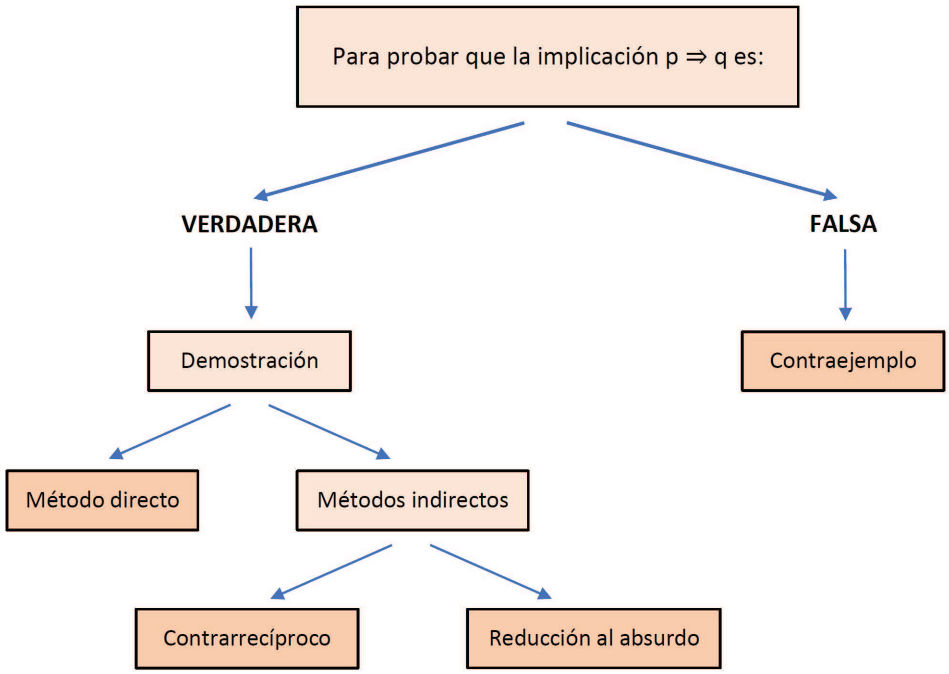


Figura 1.3: Métodos de demostración.

Al final de la próxima subsección se ilustran los métodos de demostración con una serie de ejemplos.

1.2.3. Funciones proposicionales

Una **función proposicional** o **predicado** en una variable x es toda oración en la que x aparece como sujeto, la cual se convierte en proposición

[4]Nótese que no se está asegurando que dicho elemento exista. Sólo se prueba que, si existe, es único.

cuando se especifica x. Puede ocurrir que, en lugar de una sola, haya más de una variable: x, y, z, etc.

Ejemplo 1.7. *Suponiendo que los hermanos Marta y Pedro son profesores y que Juan es médico, indicar el valor de verdad de las siguientes funciones proposicionales:*

$$P(x)\text{: “}x \text{ es un médico”},$$
$$Q(y)\text{: “}y \text{ es un número par”},$$
$$R(x,z)\text{: “}x \text{ y } z \text{ son hermanos”},$$

al sustituir $x = \text{Marta}$ *en la primera,* $y = 4$ *en la segunda y* $x = \text{Marta}$*,* $z = \text{Pedro}$ *en la tercera.*

$\triangleleft$ Las oraciones: $P(x)$, $Q(y)$ y $R(x,z)$ son funciones proposicionales, las dos primeras en una variable (x e y, respectivamente) y la tercera en dos variables (x y z).

Se tiene que: $P(\text{Marta})$ es falso, $Q(4)$ es verdadero y $R(\text{Marta}, \text{Pedro})$ es verdadero. $\triangleright$

Para hacer referencia a que hay algún valor de x para el cual una función proposicional $P(x)$ es verdadera se utiliza el símbolo $(\exists x)$, que se lee “existe x” y se llama **cuantificador existencial**. Luego,

$$\text{“Existe un } x \text{ tal que } P(x) \text{ es verdadero” se simboliza por:}$$
$$(\exists x) : P(x).$$

Análogamente, para hacer referencia a que $P(x)$ es verdadera para todos los valores posibles que pueda tomar la variable x se utiliza el símbolo $(\forall x)$, que se lee “para todo x” y se llama **cuantificador universal**. Luego,

$$\text{“Para todo } x \text{ se verifica que } P(x) \text{ es verdadero” se simboliza por:}$$
$$(\forall x) : P(x).$$

Aparecen con frecuencia expresiones dadas a partir de su negación. Se tiene que:

$$(\sim ((\exists x) : P(x))) \Leftrightarrow ((\forall x) : \sim P(x)),$$
$$(\sim ((\forall x) : P(x))) \Leftrightarrow ((\exists x) : \sim P(x)).$$

Ejemplo 1.8. *Escribir simbólicamente la función proposicional*

$P(x)$: "Todas las personas merecen una segunda oportunidad".

$\lhd$ La función proposicional $P(x)$ se puede simbolizar como

$$(\forall x) : ((x \text{ es una persona}) \Rightarrow (x \text{ merece una segunda oportunidad})).$$

$\rhd$

Ejemplo 1.9. *Escribir simbólicamente la función proposicional*

$R(x)$: "Todo número entero es par si y sólo si su cuadrado es par"

y su negación.

$\lhd$ La función proposicional $R(x)$ se puede simbolizar como

$$(\forall x \in \mathbb{Z}) : ((x \text{ es par}) \Leftrightarrow (x^2 \text{ es par})),$$

y llamando $P(x) : (x \text{ es par})$ y $Q(x) : (x^2 \text{ es par})$ queda

$$(\forall x \in \mathbb{Z}) : (P(x) \Leftrightarrow Q(x)).$$

La negación de la función proposicional anterior es:

$$\sim ((\forall x \in \mathbb{Z}) : (P(x) \Leftrightarrow Q(x))) \Leftrightarrow ((\exists x \in \mathbb{Z}) : \sim (P(x) \Leftrightarrow Q(x))).$$

Prescindiendo del cuantificador $(\exists x)$ y aplicando una ley de De Morgan (véase Ejercicio (1c)), el último predicado se simplifica como sigue:

$$\begin{aligned}
\sim ((P(x) \Leftrightarrow Q(x))) &\Leftrightarrow \sim ((P(x) \Rightarrow Q(x)) \wedge (Q(x) \Rightarrow P(x))) \\
&\Leftrightarrow (\sim (P(x) \Rightarrow Q(x))) \vee (\sim (Q(x) \Rightarrow P(x))) \\
&\Leftrightarrow (P(x) \wedge \sim Q(x)) \vee (Q(x) \wedge \sim P(x)),
\end{aligned}$$

donde en el último paso se ha usado la negación de la implicación dos veces. De este modo, la función proposicional $\sim R(x)$ se lee:

$\sim R(x)$: "Existe un número entero que es par y su cuadrado impar o bien el número es impar y su cuadrado par".

$\triangleright$

Ejemplo 1.10. *Expresar en símbolos la función proposicional*

$Q(x)$: *"Existe un número entero par que es primo"*

y su negación.

$\triangleleft$ La función proposicional se puede simbolizar como

$$(\exists x \in \mathbb{Z}) : ((x \text{ es par}) \wedge (x \text{ es primo})).$$

y, aplicando una ley de De Morgan, su negación como

$$(\forall x \in \mathbb{Z}) : ((x \text{ es impar}) \vee (x \text{ no es primo})).$$

$\triangleright$

Los métodos de demostración indicados anteriormente para tratar proposiciones se extienden al caso de funciones proposicionales con el único cuidado de tratar adecuadamente los cuantificadores.

A continuación se presentan ejemplos de los diferentes métodos de demostración.

Ejemplo 1.11 (Demostración directa)**.** *Sea $a \in \mathbb{Z}$. Probar que si a es un número par entonces a^2 es un número par.*

$\triangleleft$ En símbolos, si se consideran las proposiciones $P(a)$: "a es un número par" y $Q(a)$: "a^2 es un número par" se deberá probar que $(\forall a \in \mathbb{Z}) : P(a) \Rightarrow Q(a)$.

Prueba: Se supone que a es un número (entero) par. Puesto que todo número par es un múltiplo de 2, existe un número entero k tal que $a = 2k$. Entonces $a^2 = (2k)^2 = 4k^2 = 2(2k^2)$. Puesto que $s := 2k^2$ es un número entero, se ha probado que a^2 es un número par dado que $a^2 = 2s$, con s entero. $\triangleright$

Ejemplo 1.12 (Demostración por el contrarrecíproco). *Sea $a \in \mathbb{Z}$. Probar que si a es impar entonces $a+2$ es impar.*

$\triangleleft$ En símbolos, si se consideran las proposiciones $P(a)$: "a es impar" y $Q(a)$: "$a+2$ es impar" se deberá probar que $(\forall a \in \mathbb{Z}) : P(a) \Rightarrow Q(a)$.

Prueba: Para demostrarlo por el contrarrecíproco se debe probar que $\sim Q(a) \Rightarrow \sim P(a)$, es decir, se debe suponer que $a+2$ es par y se debe probar que a es par. En efecto, si se supone que $a+2$ es par entonces $a+2$ debe ser un múltiplo de 2, con lo cual existe un número entero k tal que $a+2 = 2k$. Por lo tanto, $a = 2k-2 = 2(k-1)$. Puesto que $s := k-1$ es un número entero, se ha llegado a que a es un número par dado que $a = 2s$, con s entero. $\triangleright$

Ejemplo 1.13 (Demostración por el absurdo). *Sea $a \in \mathbb{Z}$. Probar que si a es par entonces $a+3$ es impar.*

$\triangleleft$ En símbolos, si se consideran las proposiciones $P(a)$: "a es par" y $Q(a)$: "$a+3$ es impar" se deberá probar que $(\forall a \in \mathbb{Z}) : P(a) \Rightarrow Q(a)$.

Prueba: Para demostrarlo por el absurdo se debe suponer que la hipótesis "a es par" es V, la tesis "$a+3$ es impar" es F y llegar a una contradicción. En efecto, por ser $a+3$ par, debe existir un número entero k tal que $a+3 = 2k$. Además, por ser a un entero par, debe existir un número entero r de modo que $a = 2r$. Sustituyendo la última expresión de a en la primera se tiene $2k = a+3 = 2r+3$ y despejando se obtiene: $3 = 2k-2r = 2(k-r)$. La última expresión indica que 3 es un número par (pues $k-r \in \mathbb{Z}$), lo cual es un absurdo. Este absurdo proviene de suponer que $a+3$ no es impar. Luego, se ha probado que $a+3$ es impar. $\triangleright$

Ejemplo 1.14 (Demostración directa distinguiendo casos)**.** *Probar que si $n \in \mathbb{N}$ entonces $1 + (-1)^n(2n - 1)$ es un múltiplo entero de 4.*

$\lhd$ En símbolos, si se consideran las proposiciones $P(n)$: "$n \in \mathbb{N}$" y $Q(n)$: "$1 + (-1)^n(2n - 1)$ es un múltiplo entero de 4" se deberá probar que $(\forall n) : P(n) \Rightarrow Q(n)$.

Prueba: Puesto que la tesis contiene la expresión $(-1)^n$, para valores de n pares se obtiene un resultado y para valores de n impares, otro. Por ello, se debe dividir la demostración en dos casos y, en cada uno, se debe demostrar la tesis. Es preciso recordar que un número entero t es múltiplo de 4 si $t = 4p$ para algún entero p. En efecto,

- n par: En este caso, se tiene que $n = 2k$, para algún $k \in \mathbb{N}$, y es claro que $(-1)^n = 1$. Luego,

$$\begin{aligned} 1 + (-1)^n(2n - 1) &= 1 + 1[2(2k) - 1] \\ &= 2(2k) \\ &= 4k, \end{aligned}$$

  con $k \in \mathbb{N}$. De este modo, se ha probado que si n es par entonces $1 + (-1)^n(2n - 1)$ es un múltiplo (¡natural!) de 4.

- n impar: En este caso, se tiene que[5] $n = 2k - 1$, para algún $k \in \mathbb{N}$, y es claro que $(-1)^n = -1$. Luego,

$$\begin{aligned} 1 + (-1)^n(2n - 1) &= 1 + (-1)[2(2k - 1) - 1] \\ &= 1 - 4k + 2 + 1 \\ &= 4 - 4k = 4(1 - k), \end{aligned}$$

  con $1 - k \in \mathbb{Z}$. De este modo, se ha probado que si n es impar entonces $1 + (-1)^n(2n - 1)$ es un múltiplo (¡entero!) de 4.

[5]Nótese que no se puede poner $n = 2k + 1$ pues, en ese caso, para el primer natural $k = 1$, se tendría $n = 3$ y no $n = 1$, que es el primer número natural impar.

Debido a que, en ambos casos, se ha demostrado la tesis, se tiene que el resultado es válido para todo $n \in \mathbb{N}$, con lo que $(\forall n) : P(n) \Rightarrow Q(n)$. Observar que la función proposicional a demostrar se podría haber escrito simplemente como $(\forall n \in \mathbb{N}) : Q(n)$. $\triangleright$

Ejemplo 1.15 (Demostración por refutación)**.** *Decidir si la siguiente función proposicional es verdadera o no: "$n^2 - 14n + 45 \geq 0, \forall\, n \in \mathbb{N}$."*

$\triangleleft$ Para saber si la función proposicional es verdadera o no, se puede comenzar viendo qué valores de verdad se obtienen para unos cuantos valores de $n \in \mathbb{N}$. En caso de que sea verdadera para unos cuantos, se debería intentar una demostración general, válida para todo $n \in \mathbb{N}$. Si para algún valor de $n \in \mathbb{N}$ no fuese una función proposicional cierta, ese valor de n proporcionaría un contraejemplo.

En símbolos, si se considera la proposición $P(n)$: "$n^2 - 14n + 45 \geq 0$" se deberá probar que:

(a) $(\forall n \in \mathbb{N}) : P(n)$, o bien que

(b) $(\exists n_0 \in \mathbb{N}) : \sim P(n_0)$, o equivalentemente $(\exists n_0 \in \mathbb{N}) : n_0^2 - 14n_0 + 45 < 0$.

Prueba: La siguiente tabla presenta los resultados de calcular $n^2-14n+45$ para algunos valores de $n \in \mathbb{N}$:

n	1	2	3	4	5
$n^2 - 14n + 45$	32	21	12	5	0

Se observa que a medida que aumentan los valores de n, el valor de $n^2 - 14n + 45$ disminuye pero todos se mantienen estrictamente positivos, excepto para $n = 5$, que se obtiene $n^2 - 14n + 45 = 0$. Para $n = 6$, se tiene que $n^2 - 14n + 45 = -3 < 0$, lo que demuestra que la función proposicional es, en general, falsa. Por lo tanto, el valor de $n_0 = 6$ proporciona un contraejemplo[6]. $\triangleright$

[6] En realidad, cualquier número natural $n_0 \geq 6$ proporciona un contraejemplo, pero alcanza con encontrar uno para completar la prueba.

Ejemplo 1.16 (Demostración vacua). *Demostrar que si $n \in \mathbb{N}$ satisface que $n^2 < 0$ entonces $14n + 10 \geq 0$.*

$\lhd$ En símbolos, si se consideran las funciones proposicionales $P(n)$: "$n^2 < 0$" y $Q(n)$: "$14n + 10 \geq 0$", se deberá probar que: $(\forall n \in \mathbb{N}) : P(n) \Rightarrow Q(n)$.

Prueba: Como $P(n)$ es falsa para todo $n \in \mathbb{N}$, se cumple vacuamente que $(\forall n \in \mathbb{N}) : P(n) \Rightarrow Q(n)$. $\rhd$

En este último ejemplo es importante remarcar que no se ha probado que $Q(n)$ sea verdadera para todo $n \in \mathbb{N}$, sino que se ha asignado el valor de verdad V a la implicación $(\forall n \in \mathbb{N}) : P(n) \Rightarrow Q(n)$. Dado que $n^2 < 0$ nunca se cumple en $\mathbb{N}$, bajo esta hipótesis, en ningún caso podrá extraerse conclusión alguna acerca de la desigualdad $14n + 10 \geq 0$ puesto que la regla de inferencia modus ponens no es aplicable.

En la página 88 se puede encontrar un ejemplo de la *demostración de la unicidad*: En un grupo, el elemento neutro de la operación binaria es único.

A menos que sea necesario, los paréntesis en las expresiones $(\forall n)$ o $(\exists n)$ no se indicarán de forma explícita, escribiendo simplemente $\forall n$ ó $\exists n$.

1.3. Conjuntos y relaciones

Con el objetivo de describir la noción del infinito, el matemático alemán Georg Cantor (1845-1918) desarrolló entre 1879 y 1884 la teoría de conjuntos, que permite establecer la fundamentación de las matemáticas modernas. Para ello, consideró que: "Se entiende por conjunto la agrupación en un todo de objetos bien diferenciados de nuestra percepción o de nuestro pensamiento".

El lógico inglés Bertrand Russell (1872-1970) observó, en 1903, que la "definición" anterior conducía a paradojas: "Si la colección de todos los conjuntos que no son elementos de sí mismo es un conjunto entonces dicho conjunto verificaría la propiedad de ser elemento de sí mismo si y sólo si no es elemento de sí mismo". Este enunciado, un tanto enrevesado, se ilustra

Figura 1.4: Retrato de Georg Cantor hacia 1870.

más claramente mediante la conocida paradoja del barbero:

> "En un lejano poblado de un antiguo emirato había un barbero llamado As-Samet diestro en afeitar cabezas y barbas, maestro en escamondar pies y en poner sanguijuelas. Un día el emir se dio cuenta de la falta de barberos en el emirato, y ordenó que los barberos sólo afeitaran a aquellas personas que no pudieran hacerlo por sí mismas. Cierto día el emir llamó a As-Samet para que lo afeitara y él le contó sus angustias: 'En mi pueblo soy el único barbero. No puedo afeitar al barbero de mi pueblo, ¡que soy yo!, ya que si lo hago, entonces puedo afeitarme por mí mismo, por lo tanto ¡no debería afeitarme! Pero, si por el contrario no me afeito, entonces algún barbero debería afeitarme, ¡pero yo soy el único barbero de allí!' ".

Con la intención de preservar al máximo la teoría de Cantor y, a la vez, evitando las inconsistencias que dichas paradojas producían en dicha teoría, Ernst Zermelo (1871-1953) introdujo en 1908 la teoría axiomática de conjuntos, a partir de los llamados axiomas de Zermelo. La teoría de Zermelo fue completada por Adolf Fraenkel (1891-1965) en 1922. Sin embargo, dentro de la teoría de Zermelo-Fraenkel hay propiedades que no caracterizan a un conjunto, como pretendía Cantor en sus intuitivas ideas iniciales. Por ello, John von Neumann (1903-1957), Paul Bernays (1888-1977) y Kurt Gödel (1906-1978) resolvieron definitivamente el problema mediante la teoría axiomática de clases y conjuntos. Por indicarlo de forma somera, se llama *clase* a una colección o agrupación de objetos, *conjunto* a las clases que son objetos de otras clases, y *clase propia* a las clases incapaces de ser objetos de otra clase. De esta manera, la clase de todos los conjuntos *no es un conjunto* sino una clase propia.

Se trata de un tema delicado, profundo, y su desarrollo escapa del alcance de este libro. Sólo un comentario aproximado a este respecto. Una cuestión de fondo era saber si las matemáticas son *completas*, *consistentes* y *decidibles*. La intención de Gödel era demostrar formalmente, bajo ciertas condiciones, que todas las proposiciones verdaderas son demostrables. En 1931, pudo probar que no es así, lo que se conoce como *teorema de incompletitud de Gödel* (consistente implica incompleta, en toda teoría aritmética recursiva). Es decir, que una teoría de números necesita forzosamente contener proposiciones verdaderas pero indemostrables. Con esto, toda proposición no demostrada podía, en principio, ser indemostrable. Completando esta teoría, el matemático y lógico Alan Turing (1912-1954), que estudió problemas de decisión, probó, en 1936, que es imposible demostrar la solubilidad de un problema, a priori. Por ello, una proposición que permanece mucho tiempo sin demostrar puede ser porque es muy difícil (y que nadie lo haya conseguido aún) o simplemente porque es imposible demostrarla.

1.3.1. Conjuntos

A continuación se realizará una breve introducción a la teoría intuitiva de conjuntos, que será suficiente para desarrollar los objetivos de este libro.

Los términos **conjunto**, **elemento** y **pertenencia** son entes primitivos y, por tanto, se aceptarán sin definición. Una idea aproximada es:

> “Un conjunto consiste de una colección bien determinada de objetos que, por definición, son sus elementos”.

En general, los conjuntos se denotan por letras mayúsculas $A, B, C, \ldots$ y los elementos por letras minúsculas[7] $a, b, c, \ldots$.

La pertenencia de un elemento a un conjunto se indica con el símbolo $\in$. Es decir, el elemento a pertenece al conjunto A se indica mediante:

$$a \in A.$$

Para indicar que el elemento a no pertenece al conjunto A se escribe: $a \notin A$.

Definición 1.1. Se llama **conjunto universal** o de **referencia** al conjunto que incluye todos los elementos en discusión y se fija de antemano. En general, se denota por U.

Definición 1.2. Se llama **conjunto vacío** al conjunto que carece de elementos y se denota $\emptyset$. En símbolos,

$$\emptyset = \{x : x \neq x\}.$$

Definición 1.3. Se llama **conjunto unitario** a todo conjunto formado por un único elemento. En símbolos, si a es el único elemento del conjunto A entonces se escribe $A = \{a\}$.

Se recuerda que el símbolo $\Rightarrow$ se lee ‘implica’ y

$$p \Rightarrow q$$

significa que si p se cumple entonces también se cumple q.

[7] Salvo que se trate de un conjunto que sea, a su vez, un elemento de otro conjunto, en cuyo caso, ambos se denotarán con letras mayúsculas.

Además, el símbolo $\exists$ se lee 'existe'; el conectivo $\wedge$ se lee 'y' y el conectivo $\vee$ se lee 'o'. Por ejemplo, la expresión

$$(a \in A) \wedge (\exists b \in U : \, b \notin B)$$

significa que el elemento a pertenece al conjunto A y que existe un elemento b en el conjunto universal U que no pertence al conjunto B.

Definición 1.4. Sean A y B dos conjuntos. Se dice que A es un **subconjunto** de B o que A está **incluido** o **contenido** en B, y se escribe $A \subseteq B$, cuando todo elemento de A es un elemento de B. En símbolos,

$$A \subseteq B \qquad \Leftrightarrow \qquad (x \in A \Rightarrow x \in B).$$

Si, además, B contiene algún elemento que no está en A se dice que A es un **subconjunto propio** de B o que A está **incluido propiamente** en B y se denota por $A \subset B$. En símbolos,

$$A \subset B \qquad \Leftrightarrow \qquad ((x \in A \Rightarrow x \in B) \quad \wedge \quad (\exists y \in B : y \notin A)).$$

Por otra parte, el símbolo $\Leftrightarrow$ se lee 'si y sólo si' y

$$p \Leftrightarrow q$$

significa que si p se cumple entonces se cumple q y también que si q se cumple entonces se cumple p. En este caso se dice que p y q son **equivalentes**.

Definición 1.5. Se dice que los **conjuntos** A y B son **iguales**, y se denota $A = B$, si ambos contienen exactamente los mismos elementos. En símbolos,

$$A = B \qquad \Leftrightarrow \qquad ((x \in A \Rightarrow x \in B) \quad \wedge \quad (x \in B \Rightarrow x \in A)).$$

A continuación se definen algunas operaciones que se pueden realizar con uno o dos conjuntos: unión, intersección, complementario de un conjunto y conjunto de partes de un conjunto.

En todos los casos se considera que el conjunto universal o de referencia es U.

Definición 1.6. Sean A y B dos conjuntos. Se llama **unión** del conjunto A y el conjunto B, y se denota $A \cup B$, al conjunto formado por todos los elementos de A o de B. En símbolos,

$$A \cup B = \{x \in U : x \in A \vee x \in B\}.$$

La 'o' en la definición anterior es inclusiva y, por tanto, la unión puede contener elementos de ambos conjuntos.

Definición 1.7. Sean A y B dos conjuntos. Se llama **intersección** del conjunto A y el conjunto B, y se denota $A \cap B$, al conjunto formado por todos los elementos de A y de B. En símbolos,

$$A \cap B = \{x \in U : x \in A \wedge x \in B\}.$$

La 'y' en la definición anterior significa que la intersección debe contener únicamente elementos de ambos conjuntos.

Los **conjuntos** A y B se llaman **disjuntos** cuando la intersección de A y B es el conjunto vacío, es decir, si $A \cap B = \emptyset$.

Definición 1.8. Sean A y B dos conjuntos. Se llama **diferencia** de A y B, y se denota $A - B$ o también $A \backslash B$, al conjunto formado por todos los elementos que pertenecen a A y no pertenecen a B. En símbolos,

$$A - B = A \backslash B = \{x \in U : x \in A \wedge x \notin B\}.$$

Definición 1.9. Sea A un conjunto. Se llama **complementario** del conjunto A con respecto al conjunto U, y se denota A' o $U \backslash A$, al conjunto formado por todos los elementos de U que no están en A. En símbolos,

$$A' = U \backslash A = \{x \in U : x \notin A\}.$$

De la definición es inmediato que: Si $x \in U$ entonces $x \in A' \Leftrightarrow x \notin A$.

Figura 1.5: Vitral del comedor del Caius College (Cambridge) en homenaje a John Venn y su creación.

Las operaciones indicadas hasta ahora pueden representarse gráficamente mediante los conocidos como **diagramas de Venn**[8]. Cada círculo representa un conjunto y en él pueden haber elementos, que se representan mediante un pequeño punto o una cruz. Permiten realizar una representación diagramática

[8]Estos diagramas se llaman así en honor al matemático y filósofo británico John Venn, 1834-1923. Una conmemoración al creador de estos diagramas se aprecia en un ventanal del Colegio de Gonville y Caius mostrado en la Figura 1.5.

de los razonamientos mediante estos círculos superpuestos que expresan la relación entre los elementos considerados en cada conjunto.

Definición 1.10. Sea A un conjunto. Se llama **conjunto de partes** o **conjunto potencia** del conjunto A, y se denota $\mathcal{P}(A)$, al conjunto formado por todos los subconjuntos de A. En símbolos,

$$\mathcal{P}(A) = \{X \subseteq U : X \subseteq A\}.$$

De la definición se tiene que $\mathcal{P}(A)$ es un conjunto de conjuntos. Además, es inmediato que:

1. $\mathcal{P}(\emptyset) = \{\emptyset\}$.

2. $X \in \mathcal{P}(A) \Leftrightarrow X \subseteq A$.

3. Entre los elementos de $\mathcal{P}(A)$ siempre se encuentran $\emptyset$ y el propio A; es decir $\emptyset \in \mathcal{P}(A)$ y $A \in \mathcal{P}(A)$.

4. $a \in A \Leftrightarrow \{a\} \subseteq A \Leftrightarrow \{a\} \in \mathcal{P}(A)$.

Definición 1.11. Sean A y B dos conjuntos. Se llama **par ordenado**[a] de los elementos $a \in A$ y $b \in B$, y se denota (a, b), al conjunto

$$(a, b) = \{\{a\}, \{a, b\}\}.$$

[a]Esta definición se debe a C. Kuratowski [13].

Es fácil probar que se tiene la **igualdad** de dos pares ordenados (a, b) y (c, d) si y sólo si $a = c$ y $b = d$.

Definición 1.12. Dados dos conjuntos A y B, se llama **producto cartesiano** de A por B, y se denota $A \times B$, al conjunto de todos los pares ordenados con a en A y b en B. En símbolos,

$$A \times B = \{(a, b) : a \in A, b \in B\}.$$

Se observa que, por definición, $A \times B = \emptyset \Leftrightarrow A = \emptyset$ ó $B = \emptyset$.

Familias de conjuntos[9]

En esta sección se utiliza la noción de familia finita o infinita (numerable o no) de conjuntos de forma intuitiva. En la página 80 se establece la definición formal de conjunto finito, infinito numerable e infinito no numerable.

Por tanto, se presentan tres situaciones:

- La definición de unión e interección de dos conjuntos puede extenderse a un número finito de ellos.

  Sea $\{A_1, A_2, \dots, A_n\}$ una familia (finita) de conjuntos. Se define la unión $A_1 \cup A_2 \cup \cdots \cup A_n$ y la intersección $A_1 \cap A_2 \cap \cdots \cap A_n$ de los elementos de la familia por:

  $$\bigcup_{i=1}^{n} A_i = \bigcup_{i \in \{1,2,\dots,n\}} A_i = \{x \in U : x \in A_i, \text{ para algún } i = 1, 2, \dots, n\}$$

  y

  $$\bigcap_{i=1}^{n} A_i = \bigcap_{i \in \{1,2,\dots,n\}} A_i = \{x \in U : x \in A_i, \text{ para todo } i = 1, 2, \dots, n\},$$

- Ambas nociones pueden ampliarse a conjuntos infinitos (numerables).

  Por ejemplo, si el conjunto de índices es el conjunto de los números naturales $\mathbb{N}$, para la familia (no vacía) $\{A_1, A_2, \dots, A_n, \dots\} = \{A_i\}_{i=1}^{\infty}$

[9] La definición de familia de conjuntos se formaliza en la Definición 1.27 de la página 71.

(infinito numerable) se tiene que la unión $A_1 \cup A_2 \cup \cdots \cup A_n \cup \ldots$ se define como

$$\bigcup_{i=1}^{\infty} A_i = \bigcup_{i\in\mathbb{N}} A_i = \{x \in U : x \in A_i, \text{ para algún } i = 1, 2, \ldots, n, \ldots\}$$

y la intersección $A_1 \cap A_2 \cap \cdots \cap A_n \cap \ldots$ se define como

$$\bigcap_{i=1}^{\infty} A_i = \bigcap_{i\in\mathbb{N}} A_i = \{x \in U : x \in A_i, \text{ para todo } i = 1, 2, \ldots, n, \ldots\}.$$

- Más aún, las nociones pueden extenderse al caso de conjuntos infinitos del tipo $I = [0, 1]$ (infinito no numerable), el intervalo de los números reales comprendidos entre 0 y 1. En este caso, para la familia (no vacía) $\{A_i\}_{i\in I}$ se tiene

$$\bigcup_{i\in I} A_i = \{x \in U : x \in A_i, \text{ para algún } i \in I\}$$

y

$$\bigcap_{i\in I} A_i = \{x \in U : x \in A_i, \text{ para todo } i \in I\}.$$

1.3.2. Relaciones de equivalencia

En esta sección, primero se establecerá la definición de relación (binaria) que permite relacionar los elementos de dos conjuntos y, posteriormente, un tipo especial de relaciones binarias, correspondientes a las relaciones de equivalencia, que permiten determinar particiones en un conjunto.

Definición 1.13. Dados dos conjuntos A y B, se llama **relación** o **relación binaria** entre los elementos de A y de B, y se denota $\mathcal{R}$, a todo subconjunto del producto cartesiano $A \times B$. En símbolos,

$$\mathcal{R} \subseteq A \times B.$$

Si $A = B$, se llama simplemente **relación en** A.

Si $(a, b) \in \mathcal{R}$ se escribe $a\mathcal{R}b$ y se lee: a está en relación $\mathcal{R}$ con b.

Uno de los conceptos importantes del Álgebra, y de las Matemáticas en general, es el de *relación de equivalencia* sobre un conjunto. Este concepto permite formar diferentes grupos, cada uno de los cuales posee todos los elementos del conjunto dado que tengan en común una característica (predeterminada). Dicha característica se debe indicar en la definición de la relación binaria (de equivalencia) propuesta.

Definición 1.14. Sea A un conjunto no vacío y $\mathcal{R}$ una relación en A. Se dice que $\mathcal{R}$ es una **relación de equivalencia** si cumple las propiedades reflexiva, simétrica y transitiva. En símbolos, la relación $\mathcal{R} \subseteq A \times A$ es de equivalencia si:

(E1) Reflexiva: para todo $a \in A$ se cumple $a\mathcal{R}a$,

(E2) Simétrica: si $a, b \in A$ son tales que $a\mathcal{R}b$ entonces $b\mathcal{R}a$,

(E3) Transitiva: si $a, b, c \in A$ son tales que $a\mathcal{R}b$ y $b\mathcal{R}c$ entonces $a\mathcal{R}c$.

Si $\mathcal{R}$ es una relación de equivalencia definida en un conjunto A, la notación $a\mathcal{R}b$ se lee: a es equivalente a b por $\mathcal{R}$.

Definición 1.15. Sea $A \neq \emptyset$ y $\mathcal{R}$ una relación de equivalencia sobre A. Se llama **clase de equivalencia** del elemento $a \in A$, y se denota[a] por C_a, al conjunto de todos los elementos de A equivalentes a a, es decir,

$$C_a = \{x \in A : x\mathcal{R}a\}.$$

[a] A veces también se denota $\bar{a}$ o también $[a]$.

El siguiente resultado aclara cuándo dos clases de equivalencia coinciden.

Lema 1.1. Sea $A \neq \emptyset$ y $\mathcal{R}$ una relación de equivalencia sobre A. Los elementos $a, b \in A$ satisfacen $a\mathcal{R}b$ si y sólo si $C_a = C_b$.

Demostración. ($\Longrightarrow$) Se supone que $a\mathcal{R}b$. Si $c \in C_a$ entonces $c\mathcal{R}a$. Por la propiedad transitiva se tiene que $c\mathcal{R}b$, con lo cual $c \in C_b$. Esto prueba que $C_a \subseteq C_b$. Análogamente, si $c \in C_b$ entonces $c\mathcal{R}b$. Por la propiedad simétrica se tiene que $b\mathcal{R}a$. Ahora, por la propiedad transitiva se llega a $c\mathcal{R}a$, con lo cual $c \in C_a$. Esto prueba que $C_b \subseteq C_a$. De ambas inclusiones se obtiene $C_a = C_b$.

($\Longleftarrow$) Se supone que $a, b \in A$ son tales que $C_a = C_b$. Por la propiedad reflexiva, se sabe que $x\mathcal{R}x$, para todo $x \in A$; en particular, para $x = a$. Luego, $a \in C_a$. Como $C_a = C_b$, se llega a $a \in C_b$, de donde se deduce que $a\mathcal{R}b$. $\square$

Se observa que son necesarias las tres propiedades (reflexiva, simétrica y transitiva) para establecer la demostración anterior.

Además, el lema establece que se puede cambiar la relación binaria $a\mathcal{R}b$ por la relación de igualdad $C_a = C_b$, donde en cada clase se han identificado todos los elementos equivalentes entre sí.

Definición 1.16. Sea $A \neq \emptyset$ y $\mathcal{R}$ una relación de equivalencia sobre A. Se llama **conjunto cociente** por la relación de equivalencia $\mathcal{R}$ definida en A al conjunto, denotado $A/\mathcal{R}$, formado por todas las clases de equivalencia por $\mathcal{R}$ en A (distintas entre sí). En símbolos,

$$A/\mathcal{R} = \{C_a : a \in A\}.$$

Recordando que el conjunto de partes de un conjunto dado A es el conjunto de todos los subconjuntos de A (es decir, $\mathcal{P}(A) = \{X : X \subseteq A\}$), es claro que $A/\mathcal{R} \subseteq \mathcal{P}(A)$ y $C_a \in \mathcal{P}(A)$ para todo $a \in A$.

Definición 1.17. Sea $A \neq \emptyset$. Se llama **partición** de A a toda familia $\{A_i\}_{i \in I}$ de subconjuntos no vacíos de A, disjuntos dos a dos, cuya unión sea A. En símbolos, si se satisface que:

(P1) Para cada $i \in I$ se cumple que $A_i \subseteq A$ con $A_i \neq \emptyset$,

(P2) Para $i, j \in I$ se tiene que $i \neq j \Longrightarrow A_i \cap A_j = \emptyset$,

(P3) $\cup_{i \in I} A_i = A$.

Observación 1.1. Si $\{A_i\}_{i \in I}$ es una partición de A, la condición (P3) fuerza a que sea $I \neq \emptyset$ puesto que $A \neq \emptyset$. Por lo tanto, para cada $i_0 \in I$, existe $A_{i_0} \in \{A_i\}_{i \in I}$ y, por (P1) debe ser $A_{i_0} \neq \emptyset$. Además, por el apartado (P2), la familia $\{A_i\}_{i \in I}$ está formada por conjuntos distintos dos a dos puesto que deben ser disjuntos dos a dos.

Teorema 1.1. Sea $A \neq \emptyset$ y $\mathcal{R}$ una relación de equivalencia sobre A. El conjunto de las clases de equivalencia de $\mathcal{R}$ forma una partición de A.

Demostración. Si $\mathcal{R}$ es una relación de equivalencia sobre el conjunto no vacío A, se debe probar que $A/\mathcal{R}$ es una partición de A. En efecto,

(P1) Al ser $A \neq \emptyset$, existe $a \in A$. Como $\mathcal{R}$ es reflexiva, $a\mathcal{R}a$. Luego, $a \in C_a$. Así, $C_a \neq \emptyset$, para todo $a \in A$ y es claro que $C_a \subseteq A$, para todo $a \in A$.

(P2) Sean $a, b \in A$ tales que $C_a \cap C_b \neq \emptyset$. Entonces existe $c \in C_a \cap C_b$, de donde se tiene que $c \in C_a$ y $c \in C_b$. Por la definición de clase de equivalencia, se tiene que $c\mathcal{R}a$ y $c\mathcal{R}b$. Por la propiedad simétrica se obtiene $a\mathcal{R}c$, y por la propiedad transitiva se concluye que $a\mathcal{R}b$. Aplicando el Lema 1.1, $C_a = C_b$. Por lo tanto, por el contrarrecíproco, se ha probado que si $C_a \neq C_b$ entonces $C_a \cap C_b = \emptyset$.

(P3) Si $x \in A$, por (P1) se tiene que $x \in C_x \subseteq \cup_{a \in A} C_a$. Entonces $A \subseteq \cup_{a \in A} C_a$. Para probar la otra inclusión, como para cada $a \in A$, $C_a \subseteq A$ se tiene que $\cup_{a \in A} C_a \subseteq A$. De ambas inclusiones de deduce que $\cup_{a \in A} C_a = A$.

Por lo tanto, $A/\mathcal{R}$ es una partición de A. $\square$

Es decir, toda relación de equivalencia sobre A induce una partición sobre dicho conjunto.

El Teorema 1.1 conjuntamente con el siguiente resultado prueban que el concepto de partición es equivalente al de relación de equivalencia en el sentido que existe una biyección[10] entre el conjunto de las particiones de un conjunto no vacío A y el conjunto de las relaciones de equivalencia definidas sobre A.

Teorema 1.2. Sean $A \neq \emptyset$ y $\{A_i\}_{i \in I}$ una partición de A. Si se define la relación binaria $\mathcal{R}$ en A del siguiente modo: Para $a, b \in A$,

$$a\mathcal{R}b \iff \text{existe } i \in I \text{ tal que } a \in A_i \text{ y } b \in A_i,$$

entonces $\mathcal{R}$ es una relación de equivalencia sobre A y sus clases de equivalencia son los subconjuntos A_i, con $i \in I$, de la partición dada.

Demostración. Para cada $a \in A = \cup_{i \in I} A_i$ se tiene que existe $i \in I$ tal que $a \in A_i$. Entonces $a\mathcal{R}a$, y así, $\mathcal{R}$ es reflexiva.

Si $a\mathcal{R}b$ entonces existe $i \in I$ tal que $a \in A_i$ y $b \in A_i$, de donde claramente se deduce que $b\mathcal{R}a$, y por lo tanto $\mathcal{R}$ es simétrica.

Si $a\mathcal{R}b$ y $b\mathcal{R}c$ entonces existe $i \in I$ tal que $a \in A_i$ y $b \in A_i$ y también existe $j \in I$ tal que $b \in A_j$ y $c \in A_j$. Luego $b \in A_i \cap A_j$, de donde se deduce que $A_i \cap A_j \neq \emptyset$. Por la definición de partición, $A_i = A_j$. Entonces existe $i \in I$ tal que $a \in A_i$ y $c \in A_i$, es decir, $a\mathcal{R}c$. Luego, $\mathcal{R}$ es transitiva.

Por lo tanto, $\mathcal{R}$ es una relación de equivalencia sobre A.

[10] Véase la Subsección 1.4.1 para la definición de biyección.

Sea $a \in A$. Como $A = \cup_{i\in I} A_i$, existe $i_0 \in I$ tal que $a \in A_{i_0}$. Se probará que $C_a = A_{i_0}$. En efecto,

($\subseteq$) si $x \in C_a$ entonces $x\mathcal{R}a$, es decir, existe $i \in I$ tal que $x \in A_i$ y $a \in A_i$. Luego, $a \in A_{i_0} \cap A_i$. Por la definición de partición debe ser $A_i = A_{i_0}$ y, por lo tanto, $i = i_0$. De este modo, $x \in A_{i_0}$, con lo que $C_a \subseteq A_{i_0}$.

($\supseteq$) si $x \in A_{i_0}$, como $a \in A_{i_0}$ se tiene que $x\mathcal{R}a$ y, por lo tanto, $x \in C_a$. De este modo, $A_{i_0} \subseteq C_a$. $\square$

Es posible identificar cada clase de equivalencia a partir de un sólo elemento de la misma, lo que se define a continuación.

Definición 1.18. Sean A un conjunto no vacío y la familia $\mathcal{P} = \{A_i\}_{i\in I}$ una partición de A. Se llama **conjunto completo de representantes** de la partición $\mathcal{P}$ a todo subconjunto R de A que satisface[a]:

(CR1) para cada $A_i \in \mathcal{P}$, existe $b \in R$ tal que $b \in A_i$,

(CR2) para cualquier par $b_1, b_2 \in R$, si existe $i_0 \in I$ tal que $b_1 \in A_{i_0}$ y $b_2 \in A_{i_0}$ entonces $b_1 = b_2$.

[a] De forma alternativa, se puede definir **conjunto completo de representantes** de una partición $\mathcal{P} = \{A_i\}_{i\in I}$ a todo subconjunto R de A tal que, para cada $i \in I$, la intersección $R \cap A_i$ contiene exactamente un elemento.

La condición (CR1) indica que R contiene *al menos* un representante b de cada conjunto A_i. La condición (CR2) indica que dos elementos distintos de R deben corresponder a dos elementos (que son conjuntos) distintos de $\mathcal{P}$, es decir que R contiene *a lo sumo* un representante de cada conjunto A_i. Por lo tanto, R debe contener un único **representante** de cada conjunto A_i de la partición $\mathcal{P}$.

Un ejemplo importante de la Geometría

Tras considerar que se acepta la existencia de un conjunto llamado **plano ordinario**[11], que se denota π, cuyos elementos se llaman **puntos**, y teniendo en cuenta los Postulados de la Geometría (axiomática) de Euclides, es conocido que para cada par de puntos A y B distintos del plano, existe exactamente una recta que los contiene. Además, ambos puntos determinan los segmentos $\overline{AB}$ y $\overline{BA}$ para los cuales se cumple que $\overline{AB} = \overline{BA}$. Si sobre dichos segmentos se considera una orientación, el segmento $\overline{AB}$ se puede recorrer de dos formas: desde A hasta B y, en este caso, se dice que tiene origen en A y extremo en B; o bien desde B hasta A y, en este caso, se dice que dicho segmento tiene origen en B y extremo en A.

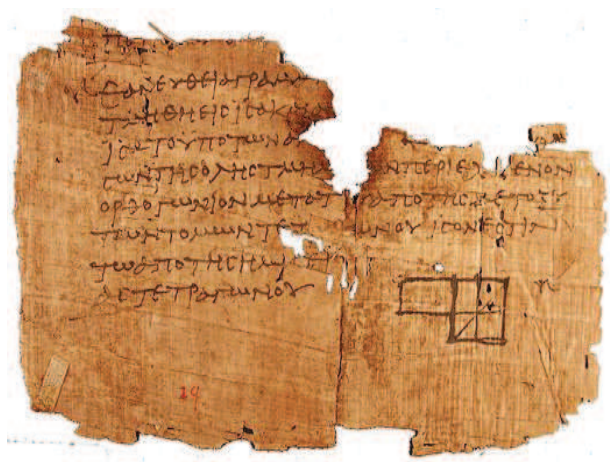


Figura 1.6: Fragmento de papiro de los Elementos de Euclides.

Al segmento orientado con **origen** en A y **extremo** en B se lo llama **vector fijo**, y se lo denota por $\overrightarrow{AB}$. Cuando los puntos A y B coinciden, el **vector fijo** se llama **nulo**.

Un vector fijo $\overrightarrow{AB}$ posee los tres elementos siguientes: su **módulo** $\|\overrightarrow{AB}\|$ (que se define como su longitud), su **dirección** (determinada por la recta

[11] Puede pensarse este plano, por ejemplo, como abstracción de una pizarra infinita.

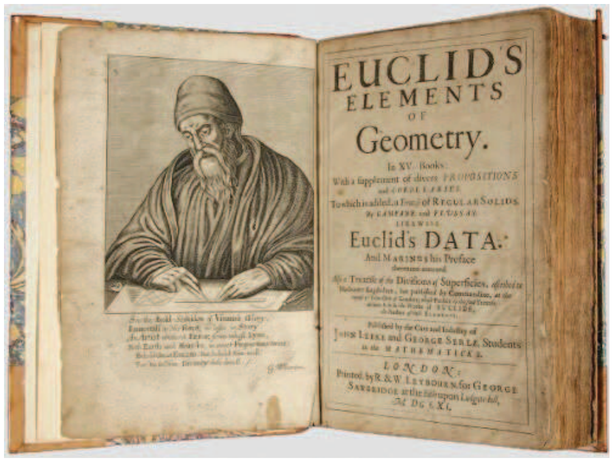


Figura 1.7: Elementos de Euclides, Universidad de Glasgow, Londres, 1661.

que pasa por los puntos A y B o por una de sus paralelas), y su **sentido** (determinado a partir de su origen y de su extremo). De un vector fijo nulo sólo es posible decir que tiene módulo 0.

Sobre el conjunto de los **vectores fijos** del plano ordinario se define la siguiente relación binaria: dos vectores fijos no nulos $\overrightarrow{AB}$ y $\overrightarrow{CD}$ se llaman **equipolentes**[12], y se denota $\overrightarrow{AB}\ \mathcal{R}\ \overrightarrow{CD}$, si tienen el mismo módulo, la misma dirección y el mismo sentido[13]. Por definición, se considera que todos los vectores fijos nulos son equipolentes entre sí.

Se prueba que esta relación (binaria) es una relación de equivalencia sobre el conjunto de vectores fijos de π. La clase de equivalencia de un vector fijo no nulo $\overrightarrow{AB}$ es

$$C_{\overrightarrow{AB}} = \{\overrightarrow{CD} : \overrightarrow{CD} \text{ tiene mismo módulo, dirección y sentido que } \overrightarrow{AB}\}.$$

[12]Del latín, misma capacidad, misma potencia.

[13]En la Geometría Afín se establece esta definición, más formalmente, diciendo que: $\overrightarrow{AB}\ \mathcal{R}\ \overrightarrow{CD}$ si (a) $ABDC$ determina un paralelogramo o bien si (b) existe otro vector fijo $\overrightarrow{EF}$ tal que $ABFE$ determina un paralelogramo y $CDFE$ también determina un paralelogramo. La opción (a) corresponde a vectores que no están alineados y la opción (b) a dos vectores alineados. Esta definición sólo utiliza el concepto de paralelismo y no el de longitud de un segmento.

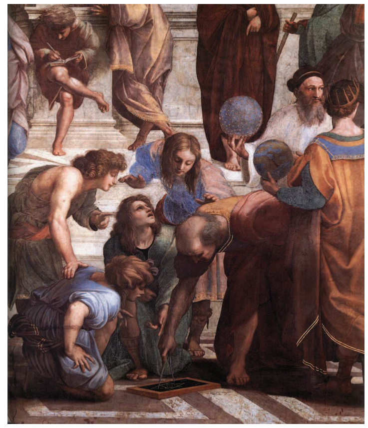


Figura 1.8: Ampliación de un fragmento de “La Escuela de Atenas” de Rafael que incluye a Euclides (dibujando con un compás).

Debido a su importancia, cada una de estas clases recibe el nombre de **vector libre**. En concreto, si se toma un vector fijo $\overrightarrow{AB}$ particular entonces $C_{\overrightarrow{AB}}$ es un vector libre y uno de sus representantes es $\overrightarrow{AB}$. Se observa que, en este caso, cada clase de equivalencia $C_{\overrightarrow{AB}}$ contiene infinitos vectores. A su vez, para cada vector fijo $\overrightarrow{AB}$, hay una clase a la que pertenece, con lo que en general hay infinitas clases de equivalencia; los vectores fijos nulos dan lugar al **vector libre nulo**.

Recordando que el **conjunto cociente** de π por la relación $\mathcal{R}$ es el conjunto formado por todas las clases de equivalencia (distintas), se puede escribir

$$\pi/\mathcal{R} = \{C_{\overrightarrow{AB}} : \overrightarrow{AB} \text{ vector fijo de } \pi\},$$

considerando sólo las clases $C_{\overrightarrow{AB}}$ distintas entre sí. Por tanto, los vectores libres son los elementos de $\pi/\mathcal{R}$. Además, el propio plano π se puede escribir como la unión (disjunta) de todas las clases de equivalencias por la relación $\mathcal{R}$, es decir,

$$\pi = \bigcup_{\overrightarrow{AB} \text{ vector fijo de } \pi} C_{\overrightarrow{AB}}.$$

A diferencia de los vectores fijos de π (que tienen origen y extremo fijos en el plano), cada representante de un vector libre de $\pi/\mathcal{R}$ se caracteriza por tener el mismo módulo, dirección y sentido[14], pero cada uno de ellos puede tener su origen "libremente" en cualquier punto del plano π, siempre que mantengan estos tres elementos invariantes. Esta caracterización de los representantes de cada vector libre permite deducir la siguiente:

Propiedad: Fijado un punto O del plano π, para cada vector fijo $\overrightarrow{AB}$ del plano π, existe un único punto P del plano π tal que $\overrightarrow{AB}$ es equipolente a $\overrightarrow{OP}$.

Los vectores fijos que comienzan en O se suelen llamar **vectores de posición**. Por lo tanto, para un vector fijo determinado, el vector libre que lo contiene, contiene también un único representante con origen en O. El conjunto formado por todos estos representantes con origen en O (es decir, por todos los vectores de posición) proporcionan un **conjunto completo de representantes** para los vectores libres que determinan la partición de π en sus clases de equivalencia.

Como la relación de equipolencia se define con el objetivo que, posteriormente, permita identificar todos los vectores fijos que comparten esa propiedad, a veces se suele hablar de **igualdad** de vectores en lugar de **equipolencia**.

Observación 1.2. La *relación de igualdad* sobre los elementos de un conjunto no vacío es una relación de equivalencia. Sin embargo, no toda relación de equivalencia corresponde a una igualdad. Por lo tanto, el concepto de relación de equivalencia es una generalización del concepto de igualdad.

[14] Exceptuando el vector nulo en estos últimos dos casos.

Ejemplo 1.17. *Sea A el conjunto de todas las rectas del plano ordinario. Se define la relación binaria: Si $r, s \in A$ entonces "$r\mathcal{R}s \Leftrightarrow r$ es paralela a s". Analizar si $\mathcal{R}$ es una relación de equivalencia. En caso afirmativo, hallar las clases de equivalencia, el conjunto cociente y un conjunto completo de representantes.*

$\lhd$ Se cumple que:

- $r\mathcal{R}r$, puesto que toda recta es paralela a sí misma.
- Si $r\mathcal{R}s$ entonces $s\mathcal{R}r$, puesto que si una recta es paralela a otra, la segunda es paralela a la primera.
- Si $r\mathcal{R}s$ y $s\mathcal{R}t$, puesto si una recta es paralela a otra y la segunda es paralela a una tercera, claramente la primera es paralela a la tercera.

Entonces $\mathcal{R}$ es una relación de equivalencia sobre A. La clase de equivalencia de una recta $r \in A$ viene dada por

$$C_r = \{s \in A : r \text{ es paralela a } s\}.$$

La relación de equivalencia produce una partición en el conjunto A formada por las clases de equivalencia. El conjunto cociente $A/\mathcal{R}$ es el conjunto de todas las clases de equivalencia (distintas):

$$A/\mathcal{R} = \{C_r : r \text{ es una recta de } A\}.$$

Fijado un punto O del plano, un conjunto completo de representantes de la partición está formado por

$$\{r \in A : r \text{ pasa por } O\}.$$

Todas las rectas de una clase C_r tienen una propiedad en común: el *paralelismo* entre rectas del plano ordinario. Esto permite hablar de la *dirección* de la recta, en realidad la dirección de una cualquiera de las rectas de dicha clase.

Por tanto, todas las rectas de la misma clase (que son paralelas) tienen igual dirección. En consecuencia, es posible distinguir una clase de equivalencia de otra a partir de la dirección de una recta cualquiera que la representante. $\triangleright$

Ejemplo 1.18 (Congruencia módulo m)**.** *Sea* $m \in \mathbb{N}$*. Para* $x, y \in \mathbb{Z}$ *se define la relación*

$$x \equiv y (\text{mód}\, m) \quad \Leftrightarrow \quad x \text{ e } y \text{ dan el mismo resto al dividirlos por } m.$$

Se dice que x **es congruente con** y **módulo** m*. Demostrar que la congruencia módulo* m *es una relación de equivalencia, hallar las clases de equivalencia, el conjunto cociente y un conjunto completo de representantes de la partición asociada.*

$\triangleleft$ En primer lugar se probarán las propiedades reflexiva, simétrica y transitiva:

- Reflexiva: $x \equiv x \pmod{m}$ pues x y x dan el mismo resto al dividirlos por m,
- Simétrica: Si $x \equiv y \pmod{m} \Rightarrow y \equiv x \pmod{m}$ pues si x e y dan el mismo resto al dividirlos por m entonces y y x dan el mismo resto al dividirlos por m,
- Transitiva: Si $x \equiv y \pmod{m}$ e $y \equiv z \pmod{m} \Rightarrow x \equiv z \pmod{m}$ pues si x e y dan el mismo resto al dividirlos por m e y y z también entonces x y z dan el mismo resto al dividirlos por m.

De este modo, la relación binaria $\equiv \pmod{m}$ es una relación de equivalencia sobre $\mathbb{Z}$.

Sea $x \in \mathbb{Z}$ fijo. Entonces $x \equiv y \pmod{m} \Leftrightarrow$ existen $c_x, c_y, r \in \mathbb{Z}$ unívocamente determinados tales que $x = c_x m + r$ e $y = c_y m + r$ con $0 \leq r < m$. Restando miembro a miembro, $x - y = (c_x - c_y)m$, con $c_x - c_y \in \mathbb{Z}$, es decir, $x - y$ es un múltiplo de m. Recíprocamente, si $x - y$ es múltiplo de m

entonces $x - y = km$ para un cierto $k \in \mathbb{Z}$ (en este caso, el resto de dividir $x-y$ por m es cero). Sean c_x, c_y los cocientes y r_x, r_y los restos, unívocamente determinados, de dividir x e y por m, respectivamente; es decir,

$$x = c_x m + r_x \qquad \text{e} \qquad y = c_y m + r_y$$

con $0 \leq r_x < m$ y $0 \leq r_y < m$. Luego, $x - y = (c_x - c_y)m + (r_x - r_y)$ con[15] $0 \leq |r_x - r_y| < m$. Puesto que el resto de la división es único y no negativo, comparando las dos expresiones de $x - y$ (o bien de $y - x$), debe ser $r_x - r_y = 0$, con lo que $r_x = r_y$ y así, $x \equiv y \pmod{m}$. En definitiva, se ha probado que

$$x \equiv y (\text{mód}\, m) \qquad \Leftrightarrow \qquad x - y = km, \text{ para cierto } k \in \mathbb{Z}.$$

Luego, la clase de equivalencia de x es

$$\begin{aligned} C_x &= \{y \in \mathbb{Z} : x \equiv y(\text{mód}\, m)\} \\ &= \{y \in \mathbb{Z} : y \equiv x(\text{mód}\, m)\} \\ &= \{y \in \mathbb{Z} : y - x = km, \text{ para cierto } k \in \mathbb{Z}\} \\ &= \{x + km : k \in \mathbb{Z}\}. \end{aligned}$$

Por ejemplo, si $m = 4$ entonces todas las clases de equivalencia son

$$\begin{aligned} C_0 &= \overline{0} = \{\dots, -8, -4, 0, 4, 8, \dots\}, \\ C_1 &= \overline{1} = \{\dots, -7, -3, 1, 5, 9, \dots\}, \\ C_2 &= \overline{2} = \{\dots, -6, -2, 2, 6, 10, \dots\}, \\ C_3 &= \overline{3} = \{\dots, -5, -1, 3, 7, 11, \dots\}. \end{aligned}$$

Por lo tanto, en este caso particular, el conjunto cociente es $\mathbb{Z}/\!\equiv\; = \{\overline{0}, \overline{1}, \overline{2}, \overline{3}\}$, que suele denotarse por $\mathbb{Z}_4$ y un conjunto completo de representantes es $R = \{0, 1, 2, 3\}$. En general,

$$\mathbb{Z}/\!\equiv\; = \{\overline{0}, \overline{1}, \overline{2}, \dots, \overline{m-1}\} = \mathbb{Z}_m,$$

[15] Sumando $0 \leq r_x < m$ y $-m < -r_y \leq 0$, se tiene $-m < r_x - r_y < m$. Por propiedades del valor absoluto, $0 \leq |r_x - r_y| < m$.

y un conjunto completo de representantes es $R = \{0, 1, 2, \ldots, m-1\}$. $\triangleright$

Es fácil probar las siguientes **propiedades**. Si $x \equiv y \pmod{m}$ y $z \in \mathbb{Z}$ entonces

(C1) compatibilidad de $\equiv \pmod{m}$ con la adición: $x + z \equiv y + z \pmod{m}$,

(C2) compatibilidad de $\equiv \pmod{m}$ con la multiplicación: $xz \equiv yz \pmod{m}$.

Para terminar esta sección, se comenta que un concepto importante tanto en el Álgebra como en la Geometría es el concepto de **invariante** asociado a una relación de equivalencia (y conjunto completo de invariantes), que se definirá con precisión en la Definición 6.5 de la página 259, momento en que se estudiará el primer ejemplo de invariante en este libro.

1.4. Aplicación o función

La idea de *función* como dependencia entre dos cantidades variables fue establecida por René Descartes (1596-1650), Isaac Newton (1642-1727) y Gottfried Leibniz (1646-1716). En particular, Leibniz acuñó los términos “función”, “variable”, “constante” y “parámetro”. La notación $f(x)$ fue utilizada por primera vez por Alexis Claude Clairaut (1713-1765) y por Leonhard Euler (1707-1783) en un artículo publicado en Memorias de la Academia Imperial de Ciencias de San Petersburgo en 1736. Euler fue el principal matemático del siglo XVIII y uno de los más grandes y prolíficos de todos los tiempos.

El concepto de función es clave en toda la Matemática. Aquí, se formalizará su definición, la operación de composición, se proporcionará su clasificación y se establecerá una caracterización para la existencia de la función inversa.

Figura 1.9: Leonhard Euler, por J.E. Handmann (hacia 1756) (Deutsches Museum, Múnich).

Definición 1.19. Sean A y B dos conjuntos no vacíos. Una relación f entre A y B (es decir, $f \subseteq A \times B$) se llama **aplicación** o **función**[a] de A en B si:

(F1) Existencia: $\forall a \in A$, $\exists b \in B$: $(a, b) \in f$,

(F2) Unicidad: $\forall a \in A, \forall b_1, b_2 \in B$: $(a, b_1) \in f \ \wedge \ (a, b_2) \in f \ \Rightarrow \ b_1 = b_2$.

[a] El término aplicación suele utilizarse de forma más general, mientras que cuando A es un subconjunto de $\mathbb{R}^n$ y B uno de de $\mathbb{R}^m$, para algunos números naturales m y n, se suele hablar de función. Sin embargo, pueden considerarse como sinónimos.

La definición indica que todo elemento de A tiene un correspondiente en B (por (F1)) y que este correspondiente es único (por (F2)).

Los conjuntos A y B se llaman **dominio** (y se denota $Dm(f)$) y **codominio** o **conjunto de llegada** de f, respectivamente.

Si $(a,b) \in f$ se dice que b es la **imagen** o **transformado** de a por f, y se denota $f(a) = b$.

La notación $f : A \to B$ se lee: f es una aplicación de A en B. Para indicar que, además, a cada $a \in A$ se le hace corresponder su imagen $b \in B$, se escribe

$$\begin{array}{cccc} f & : & A & \to & B \\ & & a & \mapsto & f(a) = b \end{array}$$

Dada una aplicación $f : A \to B$, se llama **imagen** de A por f al conjunto

$$f(A) = Im(f) = \{b \in B : \exists a \in A, f(a) = b\} = \{f(a) : a \in A\}.$$

Definición 1.20. Se dice que las **aplicaciones** $f : A \to B$ y $g : A \to B$ son **iguales**, y se denota $f = g$, si se cumple que:

$$f(a) = g(a), \quad \forall a \in A.$$

Se observa que para que dos aplicaciones sean iguales es condición necesaria que los dominios (y codominios) de ambas coincidan, y la definición requiere que ambas aplicaciones coincidan sobre todos los elementos del dominio A.

Una operación importante que se puede realizar entre algunos pares de aplicaciones es su composición.

Definición 1.21. Se llama **composición** de las aplicaciones $f : A \to B$ y $g : B \to C$ a la aplicación, denotada por $g \circ f$, tal que $g \circ f : A \to C$ y está definida mediante

$$(g \circ f)(a) := g(f(a)), \quad \forall a \in A.$$

Al realizar la composición de las aplicaciones f y g se obtiene la llamada aplicación **compuesta** de las aplicaciones f y g.

Se observa que para que tenga sentido esta operación entre aplicaciones, es necesario que el dominio de g (que es el conjunto B) coincida en el codominio de f. En realidad, no hace falta que ambos conjuntos coincidan, alcanza con que la imagen de f esté incluida en el dominio de g.

La composición de aplicaciones cumple la propiedad asociativa.

Lema 1.2. Las aplicaciones $f : A \to B$, $g : B \to C$ y $h : C \to D$ satisfacen la propiedad asociativa:

$$(h \circ g) \circ f = h \circ (g \circ f).$$

Demostración. Primero se observa que $h \circ g : B \to D$ y $g \circ f : A \to C$. Luego,

$$(h \circ g) \circ f : A \to D \quad \text{y} \quad h \circ (g \circ f) : A \to D,$$

es decir, ambas aplicaciones compuestas presentan el mismo dominio y el mismo codominio. Ahora se debe probar que ambas aplicaciones son iguales sobre todos los elementos del dominio. En efecto, sea $a \in A$. Entonces

$$[(h \circ g) \circ f](a) = (h \circ g)(f(a)) = h[g(f(a)].$$

Por otro lado,

$$[h \circ (g \circ f)](a) = h[(g \circ f)(a)] = h[g(f(a))].$$

Se ha comprobado que ambas aplicaciones son iguales para todo $a \in A$ y, por tanto, coinciden. $\square$

Ejemplo 1.19. *Sea A un conjunto no vacío. Se llama aplicación* **identidad** *en A a la aplicación $id_A : A \to A$ definida mediante*

$$id_A(a) = a, \quad \forall a \in A.$$

Se observa que la aplicación identidad sobre un conjunto deja invariantes todos los elementos de dicho conjunto.

Ejemplo 1.20. *Sean A y B dos conjuntos no vacíos. Si $f : A \to B$ es una aplicación entonces se cumple que, para todo $a \in A$, se tiene:*

$$(f \circ id_A)(a) = f(id_A(a)) = f(a) \quad \text{y} \quad (id_B \circ f)(a) = id_B(f(a)) = f(a),$$

con lo cual

$$f \circ id_A = f \quad \text{y} \quad id_B \circ f = f. \tag{1.1}$$

En este ejemplo se observa que la aplicación id_A actúa a derecha de f y la aplicación id_B lo hace a izquierda de f. Además, cuando $A = B$, se dice que la aplicación identidad id_A es el **elemento neutro** del conjunto de todas las aplicaciones de A en A pues cumple $f \circ id_A = id_A \circ f = f$, para toda aplicación $f : A \to A$.

1.4.1. Tipos de aplicaciones

A la hora de comparar ciertos aspectos de dos conjuntos, son necesarios algunos tipos especiales de aplicaciones que se definen a continuación.

Sean A y B dos conjuntos no vacíos.

Definición 1.22. Una aplicación $f : A \to B$ se llama **inyectiva** o **1-1** si cumple que:

$$\forall a_1 \in A, \forall a_2 \in A : a_1 \neq a_2 \quad \Rightarrow \quad f(a_1) \neq f(a_2),$$

o equivalentemente,

$$\forall a_1 \in A, \forall a_2 \in A : f(a_1) = f(a_2) \quad \Rightarrow \quad a_1 = a_2.$$

Es decir, una aplicación es inyectiva si elementos distintos del dominio se transforman en elementos distintos de su imagen.

Definición 1.23. Una aplicación $f : A \to B$ se llama **sobreyectiva**, **epiyectiva** o **exhaustiva** si cumple que:

$$\forall b \in B, \exists a \in A : f(a) = b.$$

Es fácil comprobar que f es sobreyectiva si y sólo si $Im(f) = B$.

Definición 1.24. Una aplicación $f : A \to B$ se llama **biyectiva** si es inyectiva y sobreyectiva.

Si entre dos conjuntos A y B es posible definir una aplicación biyectiva o **biyección**, se dice que entre ellos se ha establecido una **correspondencia biunívoca**.

Definición 1.25. Se dice que una aplicación $f : A \to B$ admite aplicación **inversa** si existe una aplicación $g : B \to A$ cumple que

$$g \circ f = id_A \qquad \text{y} \qquad f \circ g = id_B.$$

En caso contrario, se dice que f no admite aplicación inversa.

El concepto de aplicación inversa puede caracterizarse a partir del de aplicación biyectiva.

Teorema 1.3. Sea $f : A \to B$ una aplicación. Entonces f admite aplicación inversa si y sólo si f es biyectiva.

Demostración. Se considera una aplicación $f : A \to B$. Se deben demostrar las dos implicaciones que proporcionarán la validez del si y sólo si que establece el resultado.

($\Rightarrow$) Se supone que f admite aplicación inversa. Es decir, existe una aplicación $g : B \to A$ tal que

$$g \circ f = id_A \qquad \text{y} \qquad f \circ g = id_B.$$

Se debe probar que f es biyectiva, es decir, f debe ser inyectiva y sobreyectiva.

- f es inyectiva. En efecto, sean $a_1, a_2 \in A$ tales que $f(a_1) = f(a_2)$. Calculando la imagen de estos elementos (iguales) de B por la aplicación g, se tiene que $g(f(a_1)) = g(f(a_2))$. Por definición de composición, $(g \circ f)(a_1) = (g \circ f)(a_2)$. Teniendo en cuenta que $g \circ f = id_A$, se tiene que $id_A(a_1) = id_A(a_2)$. Se llega a que $a_1 = a_2$, lo que prueba que f es inyectiva.

- f es sobreyectiva. En efecto, sea $b \in B$. Utilizando que $f \circ g = id_B$, se tiene que $(f \circ g)(b) = id_B(b)$, es decir, $f(g(b)) = b$. Como $g : B \to A$, tomando $a := g(b)$, se tiene que $a \in A$ y $f(a) = f(g(b)) = b$. Esto prueba que f es sobreyectiva.

Por lo tanto, f es biyectiva.

($\Leftarrow$) Se supone que la aplicación $f : A \to B$ es biyectiva. Por definición, se tiene que f es inyectiva y sobreyectiva. Se debe probar que f admite aplicación inversa. Para ello, se construirá una aplicación $g : B \to A$ que sea (una) inversa de f.

El primer paso en la construcción de $g : B \to A$ es que cumpla las condiciones de ser aplicación (existencia y unicidad) requeridas en la Definición 1.19 y, el segundo, que cumpla las dos condiciones de la definición de aplicación inversa de f, a saber,

$$g \circ f = id_A \qquad \text{y} \qquad f \circ g = id_B.$$

En efecto, como f es sobreyectiva, para $b \in B$, existe (al menos un) $a \in A$ tal que $f(a) = b$. Además, dicho a es único, pues si existiesen $a_1, a_2 \in A$ tales que $f(a_1) = b$ y $f(a_2) = b$ entonces $f(a_1) = f(a_2)$ y, al ser f inyectiva, se deduce que $a_1 = a_2$. En definitiva, para cada $b \in B$ existe un único $a \in A$ tal que $f(a) = b$.

Este hecho permite establecer una aplicación $g : B \to A$ definida, para cada $b \in B$, mediante:

$$g(b) := a, \text{ donde } a \in A \text{ el único elemento tal que } f(a) = b. \tag{1.2}$$

Para que la aplicación g definida de esta manera sea (una) inversa de f queda por probar que

$$g \circ f = id_A \qquad \text{y} \qquad f \circ g = id_B.$$

En efecto, utilizando la definición de composición y (1.2) se tiene que

$$(g \circ f)(a) = g(f(a)) = g(b) = a = id_A(a)$$

y

$$(f \circ g)(b) = f(g(b)) = f(a) = b = id_B(b).$$

Observar que ambas identidades se verifican para cada par de elementos a y b relacionados mediante (1.2). Como estos elementos b recorren todo el conjunto B, se ha probado que $g \circ f = id_A$ y $f \circ g = id_B$. □

En el resultado anterior se ha demostrado que si una aplicación $f : A \to B$ es biyectiva entonces admite al menos una inversa.

En el siguiente teorema se probará que, en tal caso, dicha inversa es única y, además, es biyectiva.

Teorema 1.4. Sea $f : A \to B$ una aplicación. Si f admite inversa entonces dicha inversa es única y biyectiva.

Demostración. Se supone que f admite (al menos una) aplicación inversa.

Un método que permite demostrar unicidad es suponer que existen dos aplicaciones que cumplen esas condiciones y probar que son iguales. Es decir, suponer que existen dos aplicaciones

$$g_1 : B \to A \qquad \text{y} \qquad g_2 : B \to A$$

tales que

$$g_1 \circ f = id_A \qquad \text{y} \qquad f \circ g_1 = id_B$$

y

$$g_2 \circ f = id_A \qquad \text{y} \qquad f \circ g_2 = id_B.$$

Ahora, utilizando (1.1) y el hecho que la composición de aplicaciones es asociativa, se obtiene

$$g_1 = g_1 \circ id_B = g_1 \circ (f \circ g_2) = (g_1 \circ f) \circ g_2 = id_A \circ g_2 = g_2,$$

con lo que se ha probado que $g_1 = g_2$ y, por lo tanto, existe una única aplicación inversa de f.

Por otro lado, como f admite inversa, existe una aplicación $g : B \to A$ tal que

$$g \circ f = id_A \qquad \text{y} \qquad f \circ g = id_B. \tag{1.3}$$

Esta situación tiene otra lectura: para la aplicación $g : B \to A$, existe una aplicación $f : A \to B$ tal que se cumple (1.3). Es decir, g admite inversa. Del Teorema 1.3 se deduce que g es biyectiva. □

La aplicación $g : B \to A$, que es la única inversa de una aplicación biyectiva $f : A \to B$, se denota f^{-1} y cumple

$$f^{-1} \circ f = id_A \qquad \text{y} \qquad f \circ f^{-1} = id_B.$$

1.5. Números naturales. Principio de inducción

A partir de los dos elementos distinguidos (y distintos) 0 y 1 de los números reales[16] $\mathbb{R}$ y de la siguiente propiedad:

Propiedad (Consistencia de la relación de orden con la adición de $\mathbb{R}$): cualesquiera que sean los números reales a, b y c, se tiene que

$$a < b \qquad \Rightarrow \qquad a + c < b + c,$$

es posible asegurar que

$$0 < 1 \qquad \Rightarrow \qquad 0 + 1 < 1 + 1.$$

Como $0 + 1 = 1$ (pues 0 es el elemento neutro de la adición en $\mathbb{R}$) y $1 + 1$ se indica con la notación 2, se tiene que $1 < 2$. Aplicando de nuevo la propiedad, ahora con $a = 1$, $b = 2$ y $c = 1$, se tiene que

$$1 < 2 \qquad \Rightarrow \qquad 1 + 1 < 2 + 1,$$

y como $2 + 1$ se representa con 3, se tiene que $2 < 3$. Es decir, comenzando por 1, si se continúa de este modo, mediante el proceso de sumar 1, se obtiene de manera intuitiva el conjunto de los **números naturales** a partir de los sucesores del número 1.

Se observa que, para todo $a \in \mathbb{R}$, siempre se cumple que $a + 1 \neq a$ (pues $1 \neq 0$) y, por tanto, los sucesores que se construyen mediante este proceso son siempre diferentes de los anteriores.

Con la intención de formalizar estos hechos intuitivos se presenta el concepto de conjunto inductivo.

[16] Este conjunto se introduce y se estudia en detalle en Análisis Matemático.

Definición 1.26. Un **subconjunto** $K \subseteq \mathbb{R}$ se llama **inductivo** si se cumple:

$(I.1)$ $1 \in K$,

$(I.2)$ Si $k \in K$ entonces $k + 1 \in K$.

Ejemplo 1.21. *Analizar si los siguientes subconjuntos son subconjuntos inductivos de* $\mathbb{R}$*:*

$$K_1 = \mathbb{R}, \quad K_2 = \left\{x \in \mathbb{R} : \frac{1}{2} < x\right\}, \quad K_3 = \{0, 1\} \cup \left\{x \in \mathbb{R} : \frac{3}{2} < x\right\}$$

$$K_4 = \emptyset, \quad K_5 = \{1, 2, 3\}, \quad K_6 = \{x \in \mathbb{R} : 2 \leq x < 6\}.$$

$\triangleleft$ Los conjuntos K_1, K_2 y K_3 son subconjuntos inductivos de $\mathbb{R}$. Sin embargo, K_4, K_5 y K_6 no lo son (pues $1 \notin K_4$, $3 \in K_5$ pero $4 \notin K_5$ y $1 \notin K_6$ y, además, $5 \in K_6$ pero $6 \notin K_6$). $\triangleright$

Se observa que si K es un subconjunto inductivo de $\mathbb{R}$ entonces:

$$1 \in K, \quad 2 := 1 + 1 \in K, \quad 3 := 2 + 1 \in K, \quad 4 := 3 + 1 \in K, \quad \ldots$$

Como estos son todos los elementos que debe tener el conjunto de los números naturales (y sólo estos), parece interesante buscar, de alguna forma, el menor de los conjuntos K en estas condiciones, para rescatar la propiedad que todos ellos comparten.

Antes de presentar este hecho en el Lema 1.3, utilizando el concepto de función, es necesario precisar el de familia de conjuntos.

Definición 1.27. Sean A e I dos conjuntos. Una **familia de elementos** del conjunto A indexada por el **conjunto de índices** I es una aplicación $f : I \to A$ definida por $f(i) = a_i \in A$ para cada $i \in I$. Se la suele representar por $\{a_i\}_{i \in I}$, donde se omite mencionar la aplicación f.

En particular, cuando los elementos del conjunto A a su vez son conjuntos, se habla de **familia de conjuntos**.

Nótese que, en una familia indexada, dos elementos pueden estar repetidos, es decir, puede ocurrir que $a_i = a_j$ con $i, j \in I$ donde $i \neq j$.

Si los elementos de A son vectores, se habla de **sistema de vectores**[17].

Al hablar de familia está implícito el hecho de que puede ser vacía.

Ejemplo 1.22. *Se considera el conjunto*

$$A = \{C_a = \{x \in \mathbb{R} : a \leq x < a + 1\} : a \in \mathbb{R}\}$$

y el conjunto $I = \{1, 2, 3, 4, \dots\}$. *Escribir la familia que describe la función* $f(i) = C_i$ *para* $i \in I$.

$\lhd$ Se trata de la familia de conjuntos $\{C_i\}_{i\in I} = \{C_1, C_2, C_3, \dots\} = \{\{x \in \mathbb{R} : 1 \leq x < 2\}, \{x \in \mathbb{R} : 2 \leq x < 3\}, \{x \in \mathbb{R} : 3 \leq x < 4\}, \dots\}$. $\rhd$

Ejemplo 1.23. *Se consideran los conjuntos* $A = I = \{1, 2, 3, 4, \dots\}$. *Escribir la familia que describe la función* $f(i) = \left[\frac{i+1}{2}\right]$ *para* $i \in I$ *siendo* $[a]$ *la parte entera*[a] *del número real* a.

[a]Si $a \in \mathbb{R}$, la **parte entera** de a, denotada por [a], se define como el mayor número entero menor o igual que a.

$\lhd$ Sea $i \in I$. Por definición de la función parte entera se tiene que

- Si i es impar entonces $i+1$ es par. Luego, $\frac{i+1}{2} \in I$ con lo que $\left[\frac{i+1}{2}\right] = \frac{i+1}{2}$.
- Si i es par entonces $i + 1$ es impar. Luego, $\frac{i+1}{2} \notin I$ pero $\frac{i}{2} \in I$ con lo que $\left[\frac{i+1}{2}\right] = \left[\frac{i}{2} + \frac{1}{2}\right] = \frac{i}{2}$.

[17]Esto será útil en el tema de espacios vectoriales. Por ejemplo, un sistema de s vectores de un espacio vectorial V es un subconjunto de vectores de V, no necesariamente distintos, de la forma $\{u_1, u_2, \dots, u_s\}$.

Por ejemplo, $f(1) = 1$, $f(2) = 1$, $f(3) = 2$, $f(4) = 2$, $f(5) = 3$, $f(6) = 3$, $f(7) = 4$, $f(8) = 4, \ldots$. Se trata entonces de la familia

$$\{f_i\}_{i\in I} = \{1, 1, 2, 2, 3, 3, 4, 4, \ldots\}$$

que posee elementos repetidos. $\triangleright$

Lema 1.3. La intersección arbitraria (no vacía) de conjuntos inductivos de $\mathbb{R}$ es un conjunto inductivo de $\mathbb{R}$.

Demostración. Sea $\{K_i\}_{i\in I}$ una familia arbitraria de conjuntos inductivos de $\mathbb{R}$. Observar que, al tratarse de una familia arbitraria, el conjunto de índices I puede ser finito (no vacío) o infinito. Se debe probar que el conjunto $K := \cap_{i\in I} K_i$ es inductivo. En efecto,

(I.1) $1 \in K$, pues, como para cada $i \in I$, se tiene que K_i un conjunto inductivo, es posible afirmar que $1 \in K_i$ para todo $i \in I$. Por definición de intersección, $1 \in \cap_{i\in I} K_i =: K$.

(I.2) Si $k \in K$ entonces $k + 1 \in K$. Para demostrar este hecho, se considera $k \in K$. Por definición de intersección de una familia, $k \in K_i$ para todo $i \in I$. Como, para cada $i \in I$, el conjunto K_i es inductivo, por definición (de conjunto inductivo) se tiene que $k + 1 \in K_i$. Luego, $k + 1 \in \cap_{i\in I} K_i =: K$.

De (I.1) e (I.2) se concluye que K es un conjunto inductivo de $\mathbb{R}$. $\square$

Definición 1.28. Se llama conjunto de **números naturales**, y se denota $\mathbb{N}$, a la intersección de todos los subconjuntos inductivos de $\mathbb{R}$.

Esta definición del conjunto de números naturales requiere suponer conocido el conjunto de los números reales.

Otra forma de construir el conjunto de números naturales es a partir de un conjunto de axiomas propuestos por Giuseppe Peano (1858-1932).

Figura 1.10: Giuseppe Peano.

Observación 1.3. Se realizan las siguientes observaciones sobre conjuntos inductivos:

(a) La intersección de todos los subconjuntos inductivos de $\mathbb{R}$ es un conjunto no vacío con $1 \in \mathbb{N}$ (es decir, pues 1 pertenece a todos los subconjuntos inductivos de $\mathbb{R}$, según se ha probado en el Lema 1.3).

(b) El conjunto $\mathbb{N}$, tal como se ha presentado en la Definición 1.28, es un subconjunto inductivo de $\mathbb{R}$ (por el Lema 1.3).

(c) Si K es un subconjunto inductivo de $\mathbb{R}$ entonces $\mathbb{N} \subseteq K$. Es decir, $\mathbb{N}$ es el menor (en el sentido de la inclusión) de los subconjuntos inductivos de $\mathbb{R}$.

El siguiente resultado es una propiedad fundamental de los números naturales, permite realizar demostraciones de propiedades válidas para todo número natural. Su importancia es de tal calibre que peano lo incluyó como uno de los axiomas al construir los númeors naturales de forma axiomática.

Teorema 1.5 (Principio de inducción)**.** Sea $P(n)$ una función proposicional con n recorriendo el conjunto de los números naturales $\mathbb{N}$.
Si se cumple que:

(Paso base) $P(1)$ es verdadera,

(Paso inductivo) Para cada $k \in \mathbb{N}$, si $P(k)$ es verdadera entonces $P(k+1)$ es verdadera,

entonces $P(n)$ es verdadera para todo $n \in \mathbb{N}$.

Demostración. Primero se probará que el conjunto

$$K := \{n \in \mathbb{N} : \ P(n) \text{ es verdadera}\}$$

es un subconjunto inductivo de $\mathbb{R}$. En efecto,

$(I.1)$ $1 \in K$ pues $1 \in \mathbb{N}$ y $P(1)$ es verdadera (por la hipótesis llamada paso base).

$(I.2)$ Se debe probar que: Si $k \in K$ entonces $k+1 \in K$. Para ello, sea $k \in K$. Por definición de K, se sabe que $k \in \mathbb{N}$ y $P(k)$ es verdadera. Al ser $\mathbb{N}$ un subconjunto inductivo se tiene que $k+1 \in \mathbb{N}$. Ahora se está en las condiciones de la hipótesis del paso inductivo, su aplicación permite obtener que $P(k+1)$ es verdadera. Esto significa que $k+1 \in K$.

Ahora, como K es un conjunto inductivo de $\mathbb{R}$ y $\mathbb{N}$ es el menor de los conjuntos inductivos de $\mathbb{R}$, se tiene que $\mathbb{N} \subseteq K$.

Por otro lado, por la propia construcción, $K \subseteq \mathbb{N}$. Se ha probado entonces que $K = \mathbb{N}$, lo que significa que $P(n)$ es verdadera para todo $n \in \mathbb{N}$. □

En el paso inductivo, la suposición que se realiza: "Para cada $k \in \mathbb{N}$, $P(k)$ es verdadera" (con la cual se debe demostrar que $P(k+1)$ es verdadera) se llama **hipótesis de inducción**.

Ejemplo 1.24. *Demostrar, por el principio de inducción, que*

$$1 + 2 + 3 + \cdots + n = \frac{n(n+1)}{2}, \qquad \forall n \in \mathbb{N}. \tag{1.4}$$

De forma más compacta, la suma[a] $1 + 2 + 3 + \cdots + n$ *se podría escribir como* $\sum_{i=1}^{n} i$*. La expresión (1.4) quedaría* $\sum_{i=1}^{n} i = \frac{n(n+1)}{2}, \forall n \in \mathbb{N}$.

[a] El símbolo $\sum_{i=1}^{n}$ se lee "sumatorio desde $i = 1$ hasta n" y se define como $\sum_{i=1}^{n} a_i = a_1 + a_2 + \cdots + a_n$. Para sus propiedades, véase el Anexo A.

$\lhd$ Se quiere probar que la función proposicional

$$P(n) : 1 + 2 + 3 + \cdots + n = \frac{n(n+1)}{2}$$

es válida $\forall n \in \mathbb{N}$. Al tratarse de una propiedad sobre los números naturales, es adecuado utilizar el principio de inducción para su demostración.

Paso base: Para $n = 1$ la igualdad es válida, pues sustituyendo $n = 1$ en ambos miembros se obtiene la igualdad evidente $1 = \frac{1 \cdot 2}{2}$. Es decir, $P(1)$ es verdadera.

Paso inductivo: Para cada $k \in \mathbb{N}$, $k > 1$, se supone por hipótesis de inducción que $P(k)$ es verdadera, es decir, es sabido que:

$$1 + 2 + 3 + \cdots + k = \frac{k(k+1)}{2}.$$

Se debe probar que $P(k+1)$ es verdadera, es decir,

$$1 + 2 + 3 + \cdots + k + (k+1) = \frac{(k+1)[(k+1)+1]}{2}.$$

Para ello, aplicando la propiedad asociativa y la hipótesis de inducción, se tiene:

$$\begin{aligned}1+2+3+\cdots+k+(k+1) &= [1+2+3+\cdots+k]+(k+1)\\ &= \frac{k(k+1)}{2}+(k+1)\\ &= \frac{k(k+1)+2(k+1)}{2}\\ &= \frac{(k+1)[k+2]}{2}\\ &= \frac{(k+1)[(k+1)+1]}{2},\end{aligned}$$

lo que prueba que $P(k+1)$ es verdadera. Aplicando ahora el principio de inducción se puede afirmar que $P(n)$ es verdadera para todo $n \in \mathbb{N}$. $\triangleright$

En el principio de inducción, la elección del número 1 para comprobar la validez del paso base "$P(1)$ es V", se realiza por ser 1 el menor número natural. En realidad, el principio puede enunciarse como en el siguiente resultado, en el que puede elegirse cualquier número natural para comprobar la validez del paso base.

Teorema 1.6 (Principio de inducción (comenzando en $n_0 \in \mathbb{N}$))**.** Sea $n_0 \in \mathbb{N}$ y sea $P(n)$ una función proposicional con n recorriendo el conjunto de los números naturales $\mathbb{N}$ con $n \geq n_0$.
Si se cumple que:

(Paso base) $P(n_0)$ es verdadera,

(Paso inductivo) Para cada $k \in \mathbb{N}$ con $k \geq n_0$, si $P(k)$ es verdadera entonces $P(k+1)$ es verdadera,

entonces $P(n)$ es verdadera para todo $n \in \mathbb{N}$ tal que $n \geq n_0$.

Demostración. Si $n_0 = 1$, el resultado coincide con el del Teorema 1.5. Sea $n_0 > 1$, es decir, $n_0 \geq 2$. Primero se probará que el conjunto

$$K := \{1, 2, \ldots, n_0 - 1\} \cup \{n \in \mathbb{N} : \ P(n) \text{ es verdadera}\}$$

es un subconjunto inductivo de $\mathbb{R}$. En efecto,

$(I.1)$ $1 \in K$ por la definición de K.

$(I.2)$ Si $k \in K$ entonces $k+1 \in K$. Por la definición de K, es claro que esta condición es cierta $k = 1, 2, \ldots, n_0 - 2$. Sea ahora $k = n_0 - 1$. Entonces $k+1 = n_0 \in K$ pues $P(n_0)$ es verdadera por la hipótesis del paso base. Por último, sea $k \geq n_0$, es decir, $k \in K - \{1, 2, \ldots, n_0 - 1\}$. Se debe probar que, también en estos casos, $k \in K \Rightarrow k+1 \in K$. En efecto, de $k \in K$ se tiene que $k \in \mathbb{N}$ y $P(k)$ es verdadera y por el paso inductivo se tiene que $P(k+1)$ es verdadera puesto que $k+1 \in \mathbb{N}$ (por ser $\mathbb{N}$ inductivo); esto significa que $k+1 \in K$.

Ahora, como K es un conjunto inductivo de $\mathbb{R}$ y $\mathbb{N}$ es el menor de los conjuntos inductivos de $\mathbb{R}$, se tiene que $\mathbb{N} \subseteq K$.

Por otro lado, por la propia construcción, $K \subseteq \mathbb{N}$. Se prueba así que $K = \mathbb{N}$. Por lo tanto, $K - \{1, 2, \ldots, n_0 - 1\} = \mathbb{N} - \{1, 2, \ldots, n_0 - 1\}$, lo que significa que $P(n)$ es verdadera para todo $n \in \mathbb{N}$ tal que $n \geq n_0$. $\square$

Ejemplo 1.25. *Demostrar, por el principio de inducción*[a]*, que*

$$n! > 3^{n-2}, \qquad \forall n \in \mathbb{N} \text{ tal que } n \geq 2.$$

[a] Se recuerda que el **factorial de** $n \in \mathbb{N}$ se define como $n! = n \cdot (n-1) \cdot (n-2) \ldots 3 \cdot 2 \cdot 1$.

$\lhd$ Se quiere probar que la función proposicional

$$P(n) : n! > 3^{n-2}$$

es válida $\forall n \in \mathbb{N}$, $n \geq 2$.

Como $n = 2$ es el primer valor a partir del cual es válida la desigualdad (en $\mathbb{N}$), es apropiado aplicar la versión del principio de inducción del Teorema 1.6 (con $n_0 = 2$).

<u>Paso base</u>: Para $n = 2$ la desigualdad es válida, pues se tiene la desigualdad trivial $2! = 2 \cdot 1 = 2 > 1 = 3^0 = 3^{2-2}$. Luego, $P(2)$ es verdadera.

Paso inductivo: Para cada $k \in \mathbb{N}$, $k > 2$, se supone por hipótesis de inducción que $P(k)$ es verdadera, es decir, es conocido que:

$$k! > 3^{k-2}.$$

Se debe probar que $P(k+1)$ es verdadera, es decir,

$$(k+1)! > 3^{(k+1)-2}. \tag{1.5}$$

Para ello, se comienza con reescribiendo la definición de factorial y luego se aplica la hipótesis de inducción:

$$\begin{aligned} (k+1)! &= (k+1)k! \\ &> (k+1)3^{k-2}. \end{aligned}$$

Para demostrar la tesis buscada, que es (1.5), a partir de la expresión $(k+1)3^{k-2}$, se debe encontrar una relación de desigualdad con $3^{(k+1)-2}$. Como $3^{(k+1)-2} = 3^{1+(k-2)} = 3 \cdot 3^{k-2}$, hace falta comparar $k+1$ con 3. Al ser $k > 2$ (y, por tanto, $k+1 > 3$) se tiene que

$$\begin{aligned} (k+1)! &> (k+1)3^{k-2} \\ &> 3 \cdot 3^{k-2} \\ &= 3^{1+(k-2)} \\ &= 3^{(k+1)-2}, \end{aligned}$$

lo que prueba que $P(k+1)$ es verdadera. Aplicando ahora el principio de inducción del Teorema 1.6 se puede afirmar que $P(n)$ es verdadera para todo $n \in \mathbb{N}$ tal que $n \geq 2$. ▷

El concepto de aplicación biyectiva permite definir con precisión las nociones de conjuntos finito e infinito y, junto a la definición del conjunto de los números naturales, permite establecer "grados de infinitud" mediante los conceptos de conjuntos infinito numerable e infinito no numerable[18].

[18] La noción de aplicación biyectiva fue introducida por Georg Cantor (1845-1918) para resolver el problema de "lo infinitamente grande".

Definición 1.29. Un **conjunto** A se llama

- **finito** si es el conjunto vacío o bien si existe una aplicación biunívoca entre A y el subconjunto $I_n := \{1, 2, \ldots, n\}$ de $\mathbb{N}$, para algún $n \in \mathbb{N}$.
- **infinito**, si no es finito.
- **infinito numerable** si existe una aplicación biyectiva entre A y el conjunto $\mathbb{N}$ de los números naturales.
- **infinito no numerable** si no es finito ni infinito numerable.

De forma alternativa, (se puede demostrar que) un **conjunto** A es **finito** si es vacío o si no puede ponerse en correspondencia biunívoca con ninguno de sus subconjuntos propios[19]. Al escribir $\{a_1, a_2, \ldots, a_n\}$, y a menos que se diga lo contrario, se supone que todos los elementos de dicho conjunto (finito) son diferentes entre sí, es decir,

$$a_i \neq a_j, \qquad \forall i, j \in I_n : \quad i \neq j.$$

Se deduce que un **conjunto** A es **infinito** si puede ponerse en correspondencia biunívoca con alguno de sus subconjuntos propios.

Para un análisis detallado de los conceptos de conjunto finito e infinito se puede consultar la referencia [1].

Ejemplo 1.26. *Si a, b, c y d son números reales diferentes dos a dos, decir si los siguientes conjuntos son finitos, infinitos numerables o infinitos no numerables:*

$$A = \{a, b, c, d\}, \qquad P = \{m \in \mathbb{N} : m \text{ es par}\}, \qquad U = [0, 1].$$

$\lhd$ Es claro que, por ser a, b, c y d son números reales diferentes dos a dos, es posible establecer una biyección entre A e I_4. Por ejemplo, la aplicación

[19] Es decir, no existe una función biyectiva entre ellos.

que envía $a \mapsto 1$, $b \mapsto 2$, $c \mapsto 3$ y $d \mapsto 4$ define una tal biyección. Luego, A es finito (tiene 4 elementos).

Definiendo la aplicación $f : \mathbb{N} \longrightarrow P$ dada por $f(n) = 2n$ es inmediato probar que f es biyectiva. Además, su inversa es $f^{-1}(n) = \frac{n}{2}$. Luego, P es infinito numerable.

Para el tercer conjunto, no es posible definir una biyección entre U y $\mathbb{N}$, con lo que U es infinito no numerable. Para ver que es así, se supone por reducción al absurdo, que sí es posible definir una aplicación biyectiva de $\mathbb{N}$ sobre $[0, 1]$.

Los números de dicho intervalo son de la forma

$$0, a_1 a_2 a_3 \ldots,$$

donde $a_1, a_2, a_3, \cdots$ son los infinitos decimales (entre 0 y 9) de dicha representación (que pueden ser 0 a partir de uno en adelante, como en el caso de $0,500000\ldots$; pueden repetirse, en el caso de expresiones decimales periódicas, como en $0,\widehat{89} = 0,89898989\ldots$; o pueden ser todos diferentes en el caso de expresiones no periódicas, es decir, si es un número irracional). Los números como $\frac{1}{100}$, que admiten dos posibles representaciones: $0,00\widehat{9} = 0,0099999\ldots$ y $0,0100000\ldots$, pueden incluso considerarse mediante ambas representaciones.

Se tiene que dicha biyección es de la forma

$$\begin{array}{ccc}
1 & \mapsto & 0, a_{11}a_{12}a_{13}a_{14}\ldots \\
2 & \mapsto & 0, a_{21}a_{22}a_{23}a_{24}\ldots \\
3 & \mapsto & 0, a_{31}a_{32}a_{33}a_{34}\ldots \\
4 & \mapsto & 0, a_{41}a_{42}a_{43}a_{44}\ldots \\
\vdots & \vdots & \vdots
\end{array}$$

Observar cómo se han distribuido los subíndices en cada número (comparten el primer subíndice y el segundo se va incrementando indefinidamente de uno en uno). Es importante tener en cuenta que, en dicha lista, los números que aparecen a la derecha son todos los números distintos del intervalo $[0, 1]$, ya que la biyección ha permitido enumerarlos una vez a todos, salvo los que tienen duplicidad en las representaciones.

Ahora se debe poner la atención en "la diagonal" (cuyos elementos se han marcado en rojo):

$$\begin{array}{ccl} 1 & \mapsto & \alpha_1 = 0, a_{11}a_{12}a_{13}a_{14}\dots \\ 2 & \mapsto & \alpha_2 = 0, a_{21}a_{22}a_{23}a_{24}\dots \\ 3 & \mapsto & \alpha_3 = 0, a_{31}a_{32}a_{33}a_{34}\dots \\ 4 & \mapsto & \alpha_4 = 0, a_{41}a_{42}a_{43}a_{44}\dots \\ \vdots & \vdots & \qquad\quad \vdots \end{array}$$

Se construye un número α con representación decimal:

$$\alpha = 0, b_1b_2b_3b_4\dots,$$

donde b_i se elige distinto de 0, de 9 y de a_{ii} para todo $i = 1, 2, 3, \dots$. Por ser b_i distinto de 0 y de 9, el número α no es uno de los que admite dos representaciones decimales. Por ser $b_i \neq a_{ii}$, $\forall i$, se tiene que $\alpha \neq \alpha_1$ porque difieren en la primera cifra decimal; $\alpha \neq \alpha_2$ porque difieren en la segunda cifra decimal; $\alpha \neq \alpha_3$ porque difieren en la tercera cifra decimal, etc. En general, se tiene que $\alpha \in [0, 1]$ y no está en la lista anterior, lo que produce una contradicción. Esto prueba que el conjunto $[0, 1]$ es infinito no numerable.

El procedimiento utilizado se conoce como el **argumento de la diagonal de G. Cantor** proporcionado en 1891.

$\triangleright$

Teniendo en mente que un conjunto A (no vacío) es finito si existe una aplicación biyectiva con I_n, puede decirse que A tiene n elementos. Observar que, si se mantuviese esta idea, se estaría diciendo según el ejemplo anterior que

¡P tiene "la misma cantidad de elementos que" $\mathbb{N}$!

En el caso de conjuntos infinitos, esta última afirmación carece de sentido.

Moraleja: al tratar con infinitos, la intuición puede fallar. De forma natural, se tendería a pensar que la cantidad de números naturales es el doble que la de números pares, pero sólo cobra sentido hablar de *cantidad de elementos* cuando se hace referencia a conjuntos finitos.

En la Sección 1.6 se recordarán las definiciones de grupo, anillo y cuerpo; conceptos todos ellos que aparecerán a lo largo del libro.

1.6. Estructuras algebraicas

Si bien fue Ferdinand Georg Frobenius (1849-1917) quien definió el concepto de **grupo** (abstracto) mediante un sistema de axiomas a finales del siglo XIX, este concepto tiene su origen en el estudio de las ecuaciones algebraicas, en la teoría de números y en la geometría.

Las bases de esta rama del Álgebra fueron establecidas por Euler, Gauss, Lagrange, Abel y Galois, entre otros.

Fue el propio Évariste Galois (1811-1832), un matemático revolucionario y de los más originales de todos los tiempos, quien acuñó el término grupo y estableció la relación entre la teoría de grupos y la teoría de cuerpos, permitiendo el estudio de la simetría y de las ecuaciones algebraicas conocida hoy en día como la Teoría de Galois.

Sin duda, otra de las grandes científicas que estudió el Álgebra Abstracta (y la Física teórica) fue Emmy Noether (1882-1935); matemática alemana, de ascendencia judía, especialista en la teoría de invariantes, teoría de anillos, teoría de cuerpos, etc.

En Física, el teorema de Noether explica la importante conexión entre la simetría y las leyes de conservación. En Matemáticas, cimentó el Álgebra Abstracta con sus aportaciones a la teoría de anillos (introduciendo la teoría de ideales, un tipo especial de subanillos), entre otras cosas.

A continuación se definen las estructuras algebraicas de grupo, anillo y cuerpo.

Figura 1.11: Retrato de Évariste Galois.

Definición 1.30. Sea A un conjunto no vacío. Se llama **ley de composición interna** en A, o bien **operación binaria** en A, a toda aplicación

$$* : A \times A \to A.$$

La operación binaria se denotará mediante $*(a_1, a_2) = a_1 * a_2$, para todo $a_1, a_2 \in A$.

La operación $*(a_1, a_2)$ suele representarse $a_1 \cdot a_2$ (notación multiplicativa) o bien $a_1 + a_2$ (notación aditiva) cuando se refiere a las operaciones habituales.

Cuando se tiene una ley de composición interna $*$ en A, se suele decir que A es **cerrado** para la operación $*$ (pues al operar en A con $*$ se obtiene un elemento de A).

Figura 1.12: Retrato de Emmy Noether.

Ejemplo 1.27. *¿Es la adición de números enteros*

$$* : \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z},$$

definida mediante $*(a, b) = a + b$, *con* $a, b \in \mathbb{Z}$, *una ley de composición interna u operación binaria?*

$\lhd$ Sí, lo es, puesto que al sumar dos números enteros se obtiene otro número entero. $\rhd$

Al abordar la estructura algebraica de un espacio vectorial (en el Capítulo 8) será necesario operar con elementos de dos conjuntos diferentes, de modo que el resultado de dicha operción sea un elemento de uno de ellos. Para ello es necesaria la siguiente definición.

Definición 1.31. Sean A y B dos conjuntos no vacíos. Se llama **ley de composición externa**, o bien **operación externa**, definida en B

con operadores en A, a toda aplicación

$$*: A \times B \to B,$$

donde $A \times B$ es el producto cartesiano de A por B. La ley se denotará mediante $*(a, b) = b'$, para todo $a \in A, b \in B$.

Ejemplo 1.28. *Se considera el conjunto denotado por $B = \mathbb{R}[x]$ de todos los polinomios[a] a coeficientes reales. Analizar si las expresiones*

(a) $*(r, p(x)) = r \cdot p(x)$, con $r \in \mathbb{R}$ y $p \in B$, es decir, el producto del número real r por el polinomio $p(x)$,

(b) $*(z, p(x)) = z \cdot p(x)$, con $z \in \mathbb{C}$ y $p \in B$, es decir, el producto del número complejo z por el polinomio $p(x)$,

son leyes de composición externa.

[a] Las definiciones detalladas de **polinomio** y de **función polinomial** se pueden encontrar en el Apéndice C.

$\lhd$ La función definida por:

(a) $*(r, p(x)) = r \cdot p(x)$, con $r \in \mathbb{R}$ y $p \in B$, cumple que

$$*: \mathbb{R} \times B \to B,$$

con lo que $*$ es una ley de composición externa definida en $\mathbb{R}[x]$ con operadores en $\mathbb{R}$.

(b) $*(z, p(x)) = z \cdot p(x)$, con $z \in \mathbb{C}$ y $p \in B$, cumple que

$$*: \mathbb{C} \times B \to C,$$

donde $C = \mathbb{C}[x]$ es el conjunto de todos los polinomios a coeficientes complejos, con lo que no es una ley de composición externa definida en

$\mathbb{R}[x]$ con operadores en $\mathbb{R}$. Un contraejemplo se construye con $z = i$ y $p(x) = x$, donde se tiene que $z \cdot p(x) = ix \notin \mathbb{R}[x]$.

Sin embargo, la ley $*(z, p(x)) = z \cdot p(x)$, con $z \in \mathbb{C}$ y $p \in C$, cumple que

$$* : \mathbb{C} \times C \to C.$$

con lo que sí es una ley de composición externa definida en $\mathbb{C}[x]$ con operadores en $\mathbb{C}$. ▷

Definición 1.32. Sea G un conjunto no vacío y $*$ una operación binaria en G. El par $(G, *)$ se llama **grupo** si se cumplen las siguientes condiciones:

G1) Asociatividad: $(g_1 * g_2) * g_3 = g_1 * (g_2 * g_3)$, para todo $g_1, g_2, g_3 \in G$,

G2) Existencia de elemento neutro: $\exists e \in G$: $g * e = e * g = g$, para todo $g \in G$,

G3) Existencia de elementos inversos: Para todo $g \in G$, existe $g' \in G$: $g * g' = g' * g = e$.

Si además se cumple:

G4) Conmutatividad: $g_1 * g_2 = g_2 * g_1$, para todo $g_1, g_2 \in G$,

el par $(G, *)$ se llama **grupo conmutativo** o **abeliano**[a].

[a] El adjetivo abeliano es en reconocimineto del matemático noruego Niels Henrik Abel (1802-1829) quien estudió este tipo de estructuras algebraicas.

Ejemplo 1.29. *Los conjuntos numéricos*

$$(\mathbb{Z}, +), \qquad (\mathbb{Q}, +), \qquad (\mathbb{R}, +) \qquad y \qquad (\mathbb{C}, +)$$

de los números enteros, racionales, reales y complejos son grupos abelianos aditivos (infinitos).

Ejemplo 1.30. *Si* $\mathbb{K} \in \{\mathbb{Q}, \mathbb{R}, \mathbb{C}\}$ *entonces los conjuntos*

$$(\mathbb{K}^* := \mathbb{K} - \{0\}, \cdot)$$

son grupos abelianos multiplicativos (infinitos).

Ejemplo 1.31. *Los conjuntos*

$$(\mathbb{N}, +) \qquad y \qquad (\mathbb{Q}, \cdot)$$

no son grupos.

De forma más específica, en un grupo suelen definirse inversos a izquierda y a derecha. Si se cumple que: "Para un elemento $g \in G$, existe $g' \in G$: $g' * g = e$", el elemento g' se llama **inverso a izquierda** de g. Y si se cumple que: "Para un elemento $g \in G$, existe $g' \in G$: $g * g' = e$", el elemento g' se llama **inverso a derecha** de g.

En notación aditiva[20], el elemento neutro se nota con 0 y el elemento inverso de g se llama **simétrico u opuesto** y se nota por $-g$.

Lema 1.4. Sea $(G, *)$ un grupo. El elemento neutro e de G para la operación binaria $*$ es único.

Demostración. Si e y e' fuesen dos elementos neutros para la operación binaria $*$ de un grupo G entonces se cumplirían las condiciones

$$g * e = e * g = e \qquad \text{y} \qquad g * e' = e' * g = e',$$

para todo $g \in G$. Se debe probar que $e = e'$. Para ello, de las primeras igualdades se tiene que $e = e * e'$ y de las segundas igualdades se tiene que $e' = e * e'$. Luego, $e = e'$, como se quería demostrar. $\square$

[20]Véanse los Ejercicios 33 y 34.

Definición 1.33. Un **anillo** es una terna $(A, +, \cdot)$ donde A es un conjunto no vacío, $+$ y $\cdot$ son dos operaciones binarias que cumplen las siguientes condiciones:

A1) $(A, +)$ es un grupo abeliano,

A2) Asociatividad con respecto a $\cdot$: $(a_1 \cdot a_2) \cdot a_3 = a_1 \cdot (a_2 \cdot a_3)$, para todo $a_1, a_2, a_3 \in A$,

A3) Distributividad de $\cdot$ con respecto a $+$ a izquierda y a derecha:

$$a_1 \cdot (a_2 + a_3) = a_1 \cdot a_2 + a_1 \cdot a_3, \text{ para todo } a_1, a_2, a_3 \in A,$$

y

$$(a_1 + a_2) \cdot a_3 = a_1 \cdot a_3 + a_2 \cdot a_3, \text{ para todo } a_1, a_2, a_3 \in A.$$

Si, además, se cumple

A4) Existencia de elemento neutro con respecto a $\cdot$: existe $1 \in A$ tal que $a \cdot 1 = 1 \cdot a = a$ para todo $a \in A$,

la terna $(A, +, \cdot)$ se llama **anillo con unidad**. Y si también se satisface

A5) Conmutatividad con respecto a $\cdot$: $a_1 \cdot a_2 = a_2 \cdot a_1$ para todo $a_1, a_2 \in A$,

entonces $(A, +, \cdot)$ se llama **anillo conmutativo con unidad**.

Ejemplo 1.32. *Los conjuntos numéricos* $(\mathbb{Z}, +, \cdot)$, $(\mathbb{Q}, +, \cdot)$, $(\mathbb{R}, +, \cdot)$ *y* $(\mathbb{C}, +, \cdot)$ *de los números enteros, racionales, reales y complejos con sus operaciones habituales son anillos (infinitos) conmutativos con unidad. Sin embargo,* $(\mathbb{N}, +, \cdot)$ *no es un anillo.*

Ejemplo 1.33. *El conjunto* $(\mathbb{K}[x], +, \cdot)$ *de las funciones polinomiales en la variable* x *coeficientes en* $\mathbb{K}$*, con* $\mathbb{K} \in \{\mathbb{Q}, \mathbb{R}, \mathbb{C}\}$*, es un anillo conmutativo con unidad.*

Definición 1.34. Un anillo A con unidad 1 ($\neq 0$) se dice que tiene **característica** m si existe un entero positivo m tal que la unidad 1 sumada m veces da como resultado 0, el neutro de la adición. En símbolos, si

$$\underbrace{1 + 1 + \cdots + 1}_{m \text{ veces}} = 0.$$

Si tal m no existe, se dice que el anillo tiene **característica** 0.

Ejemplo 1.34. *Los anillos* $\mathbb{Z}$*,* $\mathbb{Q}$*,* $\mathbb{R}$ *y* $\mathbb{C}$ *tienen característica 0.*

Un caso (algo patológico) que surge con frecuencia es cuando $m = 2$. En este caso, la condición $1 + 1 \neq 0$ indica que se trata de un anillo de característica distinta de 2.

Ejemplo 1.35. *En el Ejercicio 36 de la página 99 se cumple que* $b+b = a$*, con lo que se trata de un anillo de característica 2.*

Definición 1.35. Sea $(A, +, \cdot)$ un anillo. Un elemento $a \in A$ se llama **divisor de cero a izquierda** si cumple que:

(a) $a \neq 0$,

(b) $\exists b \in A$: $b \neq 0$ tal que $a \cdot b = 0$.

De forma análoga se define **divisor de cero a derecha**.

Definición 1.36. Sea $(A, +, \cdot)$ un anillo. Un elemento $a \in A$ se llama **divisor de cero** si a es un divisor de cero a izquierda y a derecha.

Definición 1.37. Sea $(A, +, \cdot)$ un anillo y $a \in A$. Se dice que en A se cumple la

- **ley de cancelación a izquierda** por a si

$$\forall b, c \in A : \quad a \cdot b = a \cdot c \quad \Rightarrow \quad b = c.$$

- **ley de cancelación a derecha** por a si

$$\forall b, c \in A : \quad b \cdot a = c \cdot a \quad \Rightarrow \quad b = c.$$

Proposición 1.1. Sea $(A, +, \cdot)$ un anillo y $a \in A$ con $a \neq 0$. Entonces se cumple la ley de cancelación a izquierda (a derecha) por a si y sólo si a no es un divisor de cero a izquierda (a derecha).

Demostración. Sólo se hará la demostración para un divisor de cero a izquierda pues para uno a derecha es similar.

Sea $a \neq 0$.

Si a no es divisor de cero a izquierda entonces para cualquier $b \in A$ se tiene que $a \cdot b = 0$ implica $b = 0$. Luego, si $a \cdot x = a \cdot y$ entonces $a \cdot (x - y) = a \cdot x - a \cdot y = 0$, por lo que $x - y = 0$, es decir $x = y$.

Recíprocamente, si se cumple la ley de cancelación a izquierda por a, como $a \cdot b = 0 = a \cdot 0$ (véase Ejercicio 35) implica $b = 0$ entonces a no es divisor de cero a izquierda. $\square$

Definición 1.38. Un anillo $(\mathbb{K}, +, \cdot)$ conmutativo con unidad $1 \neq 0$ que, además, satisface la condición:

C) <u>Existencia de elementos inversos con respecto a $\cdot$</u>: para cada $a \in$

$\mathbb{K} - \{0\}$, existe $a' \in \mathbb{K}$ tal que $a \cdot a' = a' \cdot a = 1$,

se llama **cuerpo**.

El inverso de $a \in \mathbb{K} - \{0\}$, que es único, suele denotarse mediante a^{-1}.

Ejemplo 1.36. *Los conjuntos numéricos $\mathbb{Q}$, $\mathbb{R}$ y $\mathbb{C}$, con las operaciones habituales, son cuerpos (infinitos). Sin embargo, $\mathbb{Z}$ no lo es.*

Ejemplo 1.37. *El conjunto $(\mathbb{K}[x], +, \cdot)$ de las funciones polinomiales en la variable x con coeficientes en $\mathbb{K}$, con $\mathbb{K} \in \{\mathbb{Q}, \mathbb{R}, \mathbb{C}\}$, no es un cuerpo.*

$\lhd$ En el anillo $(\mathbb{K}[x], +, \cdot)$, no todo elemento no nulo es invertible, puesto que los únicos polinomios cuyo inverso es otro polinomio son los polinomios constantes no nulos. $\rhd$

Ejemplo 1.38. *La terna $(A, +, \cdot)$ del Ejercicio 36 de la página 99 es un cuerpo finito (con 2 elementos).*

Proposición 1.2. Si $(\mathbb{K}, +, \cdot)$ es un cuerpo entonces $\mathbb{K}$ no tiene divisores de cero.

Demostración. Sea $a \in \mathbb{K} - \{0\}$ y sea $b \in \mathbb{K}$ tal que $a \cdot b = 0$. Como $\mathbb{K}$ es un cuerpo, existe $a^{-1} \in \mathbb{K}$. Luego, se tiene que $a^{-1} \cdot (a \cdot b) = a^{-1} \cdot 0$. Por la definición de neutro, de inverso, la propiedad asociativa y el Ejercicio 35, se obtiene

$$b = 1 \cdot b = (a^{-1} \cdot a) \cdot b = a^{-1} \cdot (a \cdot b) = a^{-1} \cdot 0 = 0.$$

Un razonamiento semejante es válido para los divisores de cero a derecha. $\square$

Corolario 1.1. Si $(\mathbb{K}, +, \cdot)$ es un cuerpo y $a \in \mathbb{K}$ con $a \neq 0$ entonces en $\mathbb{K}$ se cumple la ley de cancelación a derecha e izquierda por a.

Demostración. Se deduce inmediatamente de la Proposición 1.1 y de la Proposición 1.2. $\square$

Ejemplo 1.39. *Si en el conjunto (cociente) $\mathbb{Z}_m$ del Ejemplo 1.18 de la página 59 se definen, para $\overline{x}, \overline{y} \in \mathbb{Z}_m$, la adición y la multiplicación como*

$$\overline{x} + \overline{y} := \overline{x + y} \qquad y \qquad \overline{x} \cdot \overline{y} := \overline{xy},$$

probar que $(\mathbb{Z}_m, +, \cdot)$ es un anillo conmutativo con unidad.

$\triangleleft$ Al estar definidas las nuevas operaciones sobre un conjunto cociente, es necesario asegurar que al elegir representantes de $\overline{x}$ y de $\overline{y}$ para operar, dichas operaciones no dependen de esos representantes. En efecto, se tiene que ambas operaciones están bien definidas por la compatibilidad de la relación con la adición (C1) y la multiplicación (C2) indicadas en la página 61. Luego, las clases no dependen de los representantes elegidos para realizar las operaciones.

Por otra parte, $\overline{x}+\overline{y} = \overline{x + y} = \overline{y + x} = \overline{y}+\overline{x}$, para todo $\overline{x}, \overline{y} \in \mathbb{Z}_m$. Luego, $+$ cumple la propiedad conmutativa en $\mathbb{Z}_m$. Se observa que la demostración se ha basado en utilizar la propiedad conmutativa en $\mathbb{Z}$. Las demás propiedades se prueban de forma similar (y se plantean como ejercicios). Luego, $(\mathbb{Z}_m, +, \cdot)$ es un anillo conmutativo con unidad $\overline{1}$. $\triangleright$

En cursos superiores de Estructuras Algebraicas se prueba que:

$$(\mathbb{Z}_p, +, \cdot) \text{ es un cuerpo} \qquad \Leftrightarrow \qquad p \text{ es primo positivo.}$$

De este modo, por ejemplo, $\mathbb{Z}_2$, $\mathbb{Z}_3$ y $\mathbb{Z}_5$ son cuerpos pero $\mathbb{Z}_4$ y $\mathbb{Z}_6$ no lo son.

A lo largo de libro se utilizarán principalmente los cuerpos $\mathbb{K} = \mathbb{Q}$, $\mathbb{K} = \mathbb{R}$, $\mathbb{K} = \mathbb{C}$ y, en menor medida, $\mathbb{Z}_p$ (con p primo positivo) y sus elementos se llamarán **escalares**.

1.7. EJERCICIOS

(1) Demostrar que las siguientes equivalencias son ciertas comprobando que son **tautologías**, es decir, sus tablas de verdad proporcionan siempre valores de verdad V, independientemente de los valores de verdad de las proposiciones componentes.

(a) $\sim(\sim p) \Leftrightarrow p$ (doble negación),

(b) $\sim(p \vee q) \Leftrightarrow \sim p \wedge \sim q$ (ley de De Morgan),

(c) $\sim(p \wedge q) \Leftrightarrow \sim p \vee \sim q$ (ley de De Morgan),

(d) $(p \Rightarrow q) \Leftrightarrow (\sim p \vee q)$ (la implicación en términos de $\sim$ y $\vee$),

(e) $(p \Rightarrow q) \Leftrightarrow (\sim q \Rightarrow \sim p)$ (directo y contrarrecíproco son equivalentes),

(f) $\sim(p \Rightarrow q) \Leftrightarrow (p \wedge \sim q)$ (negación de una implicación),

(g) $(p \wedge (p \Rightarrow q)) \Rightarrow q$ (modus ponens),

(h) $(\sim q \wedge (p \Rightarrow q)) \Rightarrow \sim p$ (modus tollens),

(i) $((p \Rightarrow q) \wedge (q \Rightarrow r)) \Rightarrow (p \Rightarrow r)$ (silogismo hipotético).

(2) Demostrar que la siguiente proposición es una **contradicción**, es decir, su tabla de verdad proporciona siempre valores de verdad F, independientemente de los valores de verdad de las proposiciones componentes.

(a) $p \wedge \sim p$.

(b) $p \Leftrightarrow \sim p$.

(3) Escribir en símbolos las funciones proposicionales:

(a) "La diferencia de los cuadrados de dos números naturales consecutivos es un número (entero) impar",

(b) "El producto de un número racional no nulo por uno irracional es un número irracional", y demostrarlas. (Ayuda: Para la segunda suponer conocido que el producto de dos números racionales es un número racional).

(4) Se consideran los conjuntos

$$A = \{x \in \mathbb{Z} : x^2 = 1\}, \quad B = \{x \in \mathbb{Z} : x^3 - x = 0\} \quad \text{y} \quad C = \{-1, 0, 1\}.$$

(I) Probar que $A \subseteq B$. ¿Es cierto que $A \subset B$? Justificar.

(II) ¿Es cierto que $C \subset A$? ¿Y que $B = C$? Justificar.

(5) Mostrar que $A = B$ si y sólo si $A \subseteq B$ y $B \subseteq A$.

(6) Mostrar que $A \subset B$ si y sólo si $A \subseteq B$ y $A \neq B$.

(7) Demostrar que los siguientes conjuntos son iguales:

$$A = \{x \in \mathbb{N} : x \text{ es par}\} \qquad \text{y} \qquad B = \{x \in \mathbb{N} : x^2 \text{ es par}\}.$$

¿Produce algún problema el caso $x^2 = 2$, que es par, para que se cumpla $B \subseteq A$? Justificar.

(8) Demostrar que:

(I) El conjunto vacío está incluido en cualquier conjunto A, es decir, $\emptyset \subseteq A$, para todo conjunto A.

(II) Todo conjunto A está incluido en el conjunto universal U, es decir, $A \subseteq U$, para todo conjunto A.

(III) El conjunto vacío es único. (Ayuda: Supóngase que, además de $\emptyset$, hay otro conjunto vacío y probar que deben ser iguales).

(IV) Si $A \subseteq \emptyset$ entonces $A = \emptyset$.

(9) Probar que si U denota el conjunto universal, se cumple que:

$$A \cap A' = \emptyset \qquad \text{y} \qquad A \cup A' = U.$$

(10) Dados los conjuntos A y B, demostrar que se cumple que:

$$A \subseteq B \quad \Leftrightarrow \quad A \cap B = A.$$

Enunciar y demostrar una propiedad similar cambiando la intersección por la unión.

(11) Demostrar las leyes de De Morgan: El complemento de la unión de dos conjuntos es igual a la intersección de sus complementos. Enunciar y demostrar la propiedad dual para el complemento de la intersección.

(12) Dados los conjuntos A y B, demostrar que: $A \subseteq B$ implica $B' \subseteq A'$.

(13) Dados los conjuntos A y B, demostrar que: $A - B = A \cap B'$.

(14) Sean A y B dos conjuntos disjuntos tales que $A \cup B = U$. Probar que $B = A'$.

(15) Sean A y B dos conjuntos tales que $A \subseteq B$. Demostrar que

$$B = A \cup (B - A) \qquad \text{con} \qquad A \cap (B - A) = \emptyset.$$

(16) Sean A y B dos conjuntos tales que $a, c \in A$ y $b, d \in B$. Demostrar que dos pares ordenados (a, b) y (c, d) son iguales si y sólo si $a = c$ y $b = d$. Concluir que si $A = B$ entonces $(a, b) \neq (b, a)$ si y sólo si $a \neq b$.

(17) En el conjunto $A = \{1, 2, 3, 4, 5\}$ se definen las relaciones binarias dadas por:

$$R_1 = \{(1,1), (2,2), (3,3), (4,4), (5,5)\},$$

$$R_2 = R_1 \cup \{(1,3), (3,1), (1,4), (4,1), (2,5), (5,2), (3,4), (4,3)\}$$

y

$$R_3 = A \times A.$$

Comprobar que, en todos los casos, se trata de una relación de equivalencia. Determinar las clases de equivalencia, el conjunto cociente y la partición que se produce en A en cada caso.

(18) Sea A un conjunto no vacío y R una relación de equivalencia en A. Demostrar que si $a, b \in A$ entonces aRb si y sólo si a y b pertenecen a una misma clase de equivalencia.

(19) Escribir las familias $\bigcup_{i \in I} A_i$ siendo $\{A_i\}_{i \in I}$ donde $A_i = \{i + 2\}$, con $i \in I$, son conjuntos unitarios (es decir, formados por un solo elemento). Considerar los siguientes casos:

(I) $I = \{1, 2\}$,

(II) $I = \{1, 2, \ldots, n\}$,

(III) $I = [0, 1]$.

(20) Analizar si las relaciones R_1, R_2 y R_3 definidas en cada uno de los siguientes apartados son o no una relación de equivalencia. En caso afirmativo, encontrar las clases de equivalencia y el conjunto cociente en cada caso.

(a) Para $x, y \in \mathbb{R}$, sea $x \ R_1 \ y \Leftrightarrow 4x^2 - y^2 = 0$,

(b) Para $x, y \in \mathbb{R}$, sea $x \ R_2 \ y \Leftrightarrow x^2 - y^2 = 0$,

(c) Para $x, y \in \mathbb{Z}$, sea $x \ R_3 \ y \Leftrightarrow y - x$ es un número par.

(21) Probar que la relación de equipolencia definida entre vectores fijos del plano ordinario es una relación de equivalencia.

(22) Proporcionar un contraejemplo de dos funciones f y g que permitan constatar que, en general, $g \circ f \neq f \circ g$.

(23) Sea $f : \mathbb{R} \to \mathbb{R}$ la función definida mediante la ley $f(x) = 2x - 3$. Demostrar que f es biyectiva y calcular su inversa f^{-1}. Comprobar que $f \circ f^{-1} = id_{\mathbb{R}}$ y que $f^{-1} \circ f = id_{\mathbb{R}}$.

(24) Analizar la existencia de la función inversa f^{-1} de la función $f : \mathbb{R} \to \mathbb{R}$ definida mediante la ley $f(x) = x^2$. En caso de existir su inversa calcularla, y si no existe, restringir el dominio y el codominio de f a subconjuntos adecuados de $\mathbb{R}$ en los que exista su inversa y calcularla.

La **restricción** de la función $f : Dm(f) \to Im(f)$ a un subconjunto $S \subset Dm(f)$ suele denotarse por $f|_S$ y es la función $f|_S : S \to Im(f)$ definida por $f|_S(x) = f(x)$ para todo $x \in S$.

(25) Demostrar que:

(a) La composición de dos funciones inyectivas es inyectiva.

(b) La composición de dos funciones sobreyectivas es sobreyectiva.

(c) La composición de dos funciones biyectivas es biyectiva.

(26) Comprobar que los conjuntos K_1, K_2 y K_3 del Ejemplo 1.21 son inductivos en $\mathbb{R}$.

(27) Probar que la intersección de la familia $\{K_n\}_{n\in\mathbb{N}}$ es un conjunto inductivo de $\mathbb{R}$ siendo $K_n = \{1\} \cup \{x \in \mathbb{R} : \frac{3}{2} - \frac{1}{n+1} < x\}$. ¿Es $\cap_{n\in\mathbb{N}} K_n = \mathbb{N}$? Justificar.

(28) Demostrar, por inducción, que:

(a) $\frac{1}{1\cdot 2} + \frac{1}{2\cdot 3} + \cdots + \frac{1}{n\cdot(n+1)} = \frac{n}{n+1}$, $\forall n \in \mathbb{N}$.

(b) $3 + 5 + 7 + \cdots + (2n+1) = n(n+2)$, $\quad \forall n \in \mathbb{N}$.

(c) $3 + 7 + 11 + \cdots + (4n-1) = n(2n+1)$, $\quad \forall n \in \mathbb{N}$.

(d) $1^2 + 2^2 + 3^2 + \cdots + n^2 = \frac{n(n+1)(2n+1)}{6}$, $\quad \forall n \in \mathbb{N}$.

(e) $1^3 + 2^3 + 3^3 + \cdots + n^3 = \frac{n^2(n+1)^2}{4}$, $\quad \forall n \in \mathbb{N}$.

(f) $1 + x + x^2 + \cdots + x^{n-1} + x^n = \frac{x^{n+1}-1}{x-1}$, $\forall n \in \mathbb{N}$, siendo $x \in \mathbb{R}$ un número positivo fijo, $x \neq 1$.

(29) Probar que la función $h : \{0\} \cup \mathbb{N} \to \mathbb{Z}$ definida por

$$h(n) = \begin{cases} -\frac{n}{2}, & \text{si } n \text{ es par} \\ \frac{n+1}{2}, & \text{si } n \text{ es impar} \end{cases} \qquad \text{para} \quad n \in \{0\} \cup \mathbb{N}$$

es biyectiva. Hallar h^{-1}. ¿Alcanza con que h sea biyectiva para concluir que $\mathbb{Z}$ es un conjunto infinito numerable? Justificar.

(30) Probar que $(\mathbb{Z}, +)$, $(\mathbb{Q}^* = \mathbb{Q} - \{0\}, \cdot)$, $(\mathbb{R}, +)$ y $(\mathbb{C}, +)$ son grupos abelianos (infinitos).

(31) Mostrar que los conjuntos numéricos $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ y $\mathbb{C}$ con las operaciones habituales son anillos conmutativos con unidad (infinitos).

(32) Demostrar que el conjunto $2 \cdot \mathbb{Z} = \{2k : k \in \mathbb{Z}\}$ con las operaciones habituales de $\mathbb{Z}$ es un anillo conmutativo sin unidad (infinito).

(33) Demostrar que el elemento neutro en un grupo es único.

(34) Demostrar que, para cada elemento g de un grupo G, sólo existe un elemento inverso de g en G.

(35) Si $(A, +, \cdot)$ es un anillo y 0_A es el neutro de la adición probar que $0_A \cdot a = 0_A$, $\forall a \in A$.

(36) Se considera el conjunto $A = \{a, b\}$ con $a \neq b$ y las operaciones binarias definidas mediante las tablas

$+$	a	b
a	a	b
b	b	a

$\cdot$	a	b
a	a	a
b	a	b

Demostrar que $(A, +, \cdot)$ es un anillo conmutativo con unidad (finito). ¿Es un cuerpo? Justificar la respuesta. Al estudiar Estructuras Algebraicas en profundidad, se prueba que este conjunto se comporta como $\mathbb{Z}_2$.

(37) Se considera el conjunto $K = \{a, b, c\}$ con $a \neq b \neq c \neq a$ y las operaciones binarias definidas mediante las tablas

$+$	a	b	c
a	a	b	c
b	b	c	a
c	c	a	b

$\cdot$	a	b	c
a	a	a	a
b	a	b	c
c	a	c	b

Demostrar que $(K, +, \cdot)$ es un cuerpo. Al estudiar Estructuras Algebraicas con mayor detalle, se prueba que este conjunto se comporta como $\mathbb{Z}_3$.

(38) Sea A un conjunto no vacío. Se considera el conjunto

$$\mathcal{B}(A) = \{f / f : A \to A \text{ es una aplicación}\}$$

de todas las aplicaciones de A en sí mismo. Probar que $(\mathcal{B}(A), \circ)$ es un grupo no abeliano, siendo $\circ$ la composición de aplicaciones.

Parte I

Análisis Matricial

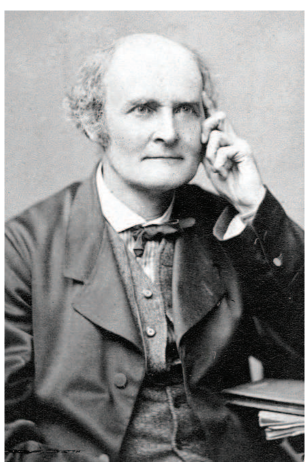
Capítulo 2

Matrices

Índice

2.1. Introducción

Si bien algunos conceptos en los que figuran matrices (y sistemas de ecuaciones lineales) aparecieron ya en algunos escritos babilonios sobre el 300 a.C. y en otros escritos chinos sobre el 100 a.C. estudiando cantidades de granos producidos en sus cosechas, fue el francés Augustin Louis Cauchy (1789-1857) quien en 1826 habló por primera vez de "cuadro" refiriéndose a una matriz. También, James Joseph Sylvester (1814-1897), en publicaciones de 1851 (en las que estudiaba la intersección de dos cónicas o cuádricas), habló explícitamente de matriz (como la madre de los menores de un determinante) y comenzó a utilizar notación matricial. En 1846, Sylvester inició una estrecha colaboración con Arthur Cayley (1821-1895) quien se sintió motivado por la idea de matriz. Más adelante, sería Cayley quien demostrara uno de los más importantes teoremas del Álgebra Lineal[1].



(a) Arthur Cayley

(b) James Joseph Sylvester

Figura 2.1: Matemáticos británicos que hicieron grandes aportaciones a la teoría de matrices.

Las matrices no sólo servirán como cuadros de números donde almace-

[1] Teorema de Cayley-Hamilton: Toda matriz cuadrada satisface su ecuación característica.

nar ciertos datos; sino que estudiando su estructura y desarrollando métodos específicos, se podrá extraer mucha información que permitirá resolver problemas que, en principio, eran impensables e inesperado que así fuese.

Las matrices son una herramienta básica del Álgebra Lineal que permiten representar y manipular de manera sencilla sistemas de ecuaciones lineales, vectores en espacios vectoriales de dimensión n, aplicaciones lineales, isometrías, formas bilineales, cónicas, cuádricas, etc.

Si bien el desarrollo básico de matrices se realiza generalmente sobre los números reales, a finales del siglo XIX y principios del XX se estudiaron numerosas propiedades de matrices con elementos sobre cuerpos abstractos[2].

Aprovechando las capacidades actuales de los ordenadores, algunos paquetes informáticos permiten trabajar con matrices de manera natural y resolver problemas de gran envergadura.

2.2. Definición

La mayoría de los textos básicos de Álgebra Lineal[3] definen el concepto de matriz de la siguiente manera:

Sean $m, n \in \mathbb{N}$. Se llama **matriz** de tamaño $m \times n$ a coeficientes en un cuerpo $\mathbb{K}$ a una tabla o cuadro rectangular A formado por $m \cdot n$ escalares de $\mathbb{K}$ dispuestos en m filas y n columnas de la siguiente forma

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}, \tag{2.1}$$

[2]Al respecto, el físico matemático Peter Guthrie Tait (1831-1901), miembro de la Sociedad Real de Edimburgo, dijo: “Cayley está forjando las armas para las futuras generaciones de físicos”.

[3]Sobre todo, los dedicados más bien a los estudiantes de ingeniería.

donde el elemento $a_{ij} \in \mathbb{K}$ se encuentra en el lugar (i, j) de la matriz A que corresponde a la intersección entre la fila i-ésima y columna j-ésima para $i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, n$.

Sin embargo, las expresiones "tabla" o "cuadro rectangular", "filas" y "columnas" no son definidas de manera precisa.

Con la intención de formalizar la definición de matriz, se recuerda que el conjunto de todas las aplicaciones $f : X \to Y$ suele denotarse mediante Y^X, es decir,

$$Y^X = \{f : f \text{ es una aplicación de } X \text{ en } Y\}.$$

Se utilizará la notación $\mathbb{I}_k = \{1, 2, \ldots, k\}$ para indicar el intervalo de los primeros k números naturales.

Definición 2.1. Se llama **matriz** de tamaño $m \times n$ a coeficientes en un cuerpo[a] $\mathbb{K}$ a cualquier aplicación $A : \mathbb{I}_m \times \mathbb{I}_n \to \mathbb{K}$ que a cada par ordenado $(i, j) \in \mathbb{I}_m \times \mathbb{I}_n$ le asigna un elemento $a_{ij} \in \mathbb{K}$, es decir $A(i, j) = a_{ij}$, llamado el elemento del lugar (i, j) de A.

[a]En realidad, es posible definirlas sobre un anillo.

Puesto que una matriz A quedará bien definida una vez conocido el conjunto de todas las imágenes de $(i, j) \in \mathbb{I}_m \times \mathbb{I}_n$ por la aplicación A:

$$\begin{matrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{matrix}$$

se las suele escribir en un cuadro de $m \cdot n$ escalares de $\mathbb{K}$ dispuestos en m **filas** y n **columnas**. En cada fila se escriben ordenadamente las imágenes de todos los pares ordenados que tienen la misma primera componente, y en cada columna se representan ordenadamente las imágenes de todos los pares ordenados que comparten la segunda componente. El **elemento** o **entrada** de la matriz que figura en la fila i-ésima y en la columna j-ésima se denota a_{ij} y se escribe con la notación indicada en (2.1).

A la matriz A de tamaño $m \times n$ a coeficientes en $\mathbb{K}$ se la suele denotar mediante $A = [a_{ij}]_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}}$, por $A = [a_{ij}]_{m \times n}$ o simplemente $A = [a_{ij}]$. Si A es una matriz de tamaño $n \times n$, se la suele denotar mediante $A = [a_{ij}]_{i,j=1}^{n}$.

En notación de aplicaciones, la matriz A será un elemento de $\mathbb{K}^{\mathbb{I}_m \times \mathbb{I}_n}$ o abreviadamente, de $\mathbb{K}^{m \times n}$. De ahora en más, al conjunto de todas las matrices de tamaño $m \times n$ a coeficientes en $\mathbb{K}$ se lo denotará por $\mathbb{K}^{m \times n}$; es decir,

$$\mathbb{K}^{m \times n} = \{A = [a_{ij}]_{m \times n} : a_{ij} \in \mathbb{K}, i = 1, 2, \ldots, m, j = 1, 2, \ldots, n\}.$$

Ejemplo 2.1. *Indicar el tamaño de la siguiente matriz y el conjunto al que pertenece*

$$A = \begin{bmatrix} 2 & 0 & 5 & -1 \\ -4 & -1 & 0 & -3 \\ 1 & 3 & 7 & 2 \end{bmatrix}.$$

◁ La matriz A tiene 3 filas y 4 columnas, con lo que es de tamaño 3×4. Su elemento a_{13} es 5. Puede considerarse que $A \in \mathbb{Z}^{3\times 4}$, $A \in \mathbb{Q}^{3\times 4}$, $A \in \mathbb{R}^{3\times 4}$ o incluso que $A \in \mathbb{C}^{3\times 4}$. ▷

Ejemplo 2.2. *La matriz* $A = \begin{bmatrix} 2+i & 0 & -i \\ -4i & -1 & 0 \\ 1 & 3 & 7-2i \end{bmatrix}$ *pertenece a* $\mathbb{C}^{3\times 3}$.

Definición 2.2. Sean $A = [a_{ij}] \in \mathbb{K}^{m\times n}$ y $B = [b_{ij}] \in \mathbb{K}^{r\times s}$. Se dice que las **matrices** A y B son **iguales** si tienen el mismo tamaño y coinciden los elementos de los lugares correspondientes, en símbolos, si $m = r$, $n = s$ y $a_{ij} = b_{ij}$ para todo $i = 1, 2, \ldots, m$ y para todo $j = 1, 2, \ldots, n$.

2.3. Tipos especiales de matrices

A continuación se definen algunos tipos especiales de matrices.

Sea $A = [a_{ij}] \in \mathbb{K}^{m\times n}$ una **matriz** dada. Se dice que:

- A es **cuadrada** si $m = n$ y A es **rectangular** si $m \neq n$.
- si $m = n$, los elementos $a_{11}, a_{22}, \dots, a_{nn}$ forman la **diagonal principal** de A.
- si $m = n$, los elementos a_{ij} para los que $i+j = n+1$ (para $i = 1, 2, \dots, n$, $j = 1, 2, \dots, n$) forman la **diagonal secundaria** de A.
- A es **diagonal** si es cuadrada y se anulan todos los elementos de fuera de su diagonal principal, en símbolos, si $m = n$ y $a_{ij} = 0$ para todo $i \in \{1, 2, \dots, n\}$ y para todo $j \in \{1, 2, \dots, n\}$ tales que $i \neq j$.
- si $m = n$, A es **triangular superior** si son cero todos los elementos por debajo de su diagonal principal, en símbolos, si $a_{ij} = 0$ para todo $i \in \{1, 2, \dots, n\}$ y para todo $j \in \{1, 2, \dots, n\}$ tales que $i > j$.
- si $m = n$, A es **triangular inferior** si son cero todos los elementos por encima de su diagonal principal, en símbolos, si $a_{ij} = 0$ para todo $i \in \{1, 2, \dots, n\}$ y para todo $j \in \{1, 2, \dots, n\}$ tales que $i < j$.
- A es **escalar** si es una matriz diagonal con todos los elementos de su diagonal iguales entre sí, en símbolos, si $m = n$, $a_{ij} = 0$ para todo $i \in \{1, 2, \dots, n\}$ y para todo $j \in \{1, 2, \dots, n\}$ tales que $i \neq j$ y existe $k \in \mathbb{K}$ tal que $a_{ii} = k$ para todo $i = 1, 2, \dots, n$.
- A es una **matriz fila** si sólo tiene una fila, es decir, si $m = 1$.
- A es una **matriz columna** si sólo tiene una columna, es decir, si $n = 1$.
- A es la **matriz nula** si todos sus elementos son cero, es decir, si $a_{ij} = 0$ para todo $i \in \{1, 2, \dots, m\}$ y para todo $j \in \{1, 2, \dots, n\}$, se denota $O_{m\times n}$ o simplemente O.

- A es la **matriz identidad** si es una matriz escalar con $k = 1$, en símbolos, si $m = n$ y si $a_{ij} = 0$ para todo $i, j \in \{1, 2, \ldots, n\}$ y $a_{ii} = 1$ para todo $i \in \{1, 2, \ldots, n\}$. Utilizando la notación de la delta de Kronecker[4] definida por

$$\delta_{ij} = \begin{cases} 1, & \text{si } i = j \\ 0, & \text{si } i \neq j \end{cases},$$

la matriz identidad es $A = [\delta_{ij}] \in \mathbb{K}^{n \times n}$. Se denota I_n o simplemente I.

Ejemplo 2.3. *La matriz nula de tamaño 3×2 es* $O = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$.

Ejemplo 2.4. $A = \begin{bmatrix} 2 & 0 & 5 & -1 \end{bmatrix}$ *es una matriz fila y* $B = \begin{bmatrix} 2 \\ 0 \\ -1 \end{bmatrix}$ *es una matriz columna, A y B son rectangulares y no son iguales.*

Ejemplo 2.5. *La matriz identidad de tamaño 3×3 es cuadrada y es*

$$I_3 = I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

[4]Leopold Kronecker (1823-1891), matemático alemán.

Ejemplo 2.6. *La matriz*

$$C = \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}$$

es diagonal para todo valor de los escalares $a, b \in \mathbb{R}$, y además C es escalar si y sólo si $a = b$.

Ejemplo 2.7. *Las matrices*

$$D = \begin{bmatrix} 1 & 6 & -9 \\ 0 & 3 & 0 \\ 0 & 0 & -11 \end{bmatrix} \qquad y \qquad E = \begin{bmatrix} 1 & 0 & 0 \\ 6 & 3 & 0 \\ 1 & 0 & -1 \end{bmatrix}$$

son triangular superior y triangular inferior, respectivamente. La diagonal secundaria de la matriz E está formada por los números $e_{13} = 0, e_{22} = 3$ y $e_{31} = 1$ (sus subíndices suman $i + j = 4$). Las matrices D y E son cuadradas.

2.4. Álgebra de matrices

En el conjunto de todas las matrices se pueden definir algunas operaciones que permitirán realizar numerosas aplicaciones en áreas tales como Teoría de códigos, Teoría de grafos, Economía, Química, Programación lineal, y un amplio etcétera que inlcuye, por supuesto, importantes aplicaciones en Ingeniería como el estudio del Procesamineto de la señal, el Cálculo de estructuras en Ingeniería Estructural (Arquitectura e Ingeniería Civil), Informática, Administración de Empresas, Geodesia, etc.

Para ello, el requisito fundamental es tener en cuenta los tamaños de las matrices involucradas. En esta sección se comienza con algunas operaciones definidas sobre matrices como la adición, la multiplicación de un escalar por una matriz, la multiplicación de matrices y la potencia de matrices cuadradas.

2.4.1. Adición de matrices

Definición 2.3. Se llama **adición** de matrices a la operación binaria

$$\begin{array}{rcl} + \ : \ \mathbb{K}^{m\times n} \times \mathbb{K}^{m\times n} & \to & \mathbb{K}^{m\times n} \\ (A,B) & \mapsto & +(A,B) = A + B, \end{array}$$

donde $A + B$, llamada **suma**[a] de las matrices $A = [a_{ij}] \in \mathbb{K}^{m\times n}$ y $B = [b_{ij}] \in \mathbb{K}^{m\times n}$, se define por

$$A + B := [a_{ij} + b_{ij}] = \begin{bmatrix} a_{11}+b_{11} & a_{12}+b_{12} & \cdots & a_{1n}+b_{1n} \\ a_{21}+b_{21} & a_{22}+b_{22} & \cdots & a_{2n}+b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}+b_{m1} & a_{m2}+b_{m2} & \cdots & a_{mn}+b_{mn} \end{bmatrix}.$$

[a]Si bien, la operación binaria se llama adición y el resultado de dicha operación se llama suma, por abuso de lenguaje, se suele llamar suma en ambos casos.

Observar que, para poder definir la adición, ambas matrices deben tener el mismo tamaño, y la suma se obtiene sumando (en $\mathbb{K}$) elemento a elemento los correspondientes a las mismas posiciones de A y de B.

Observación 2.1. Nótese que $(A, B) \mapsto A+B$ no es una operación binaria en el conjuntos de todas las matrices (de todos los tamaños), pero sí lo es en el conjunto $\mathbb{K}^{m\times n}$ para dos cantidades $m, n \in \mathbb{N}$ fijas.

Ejemplo 2.8. *Calcular la suma de las matrices*

$$A = \begin{bmatrix} 2 & 0 & 5 & -1 \\ -4 & 1 & 0 & 3 \\ 1 & -3 & 7 & -2 \end{bmatrix} \in \mathbb{Q}^{3\times 4} \quad y \quad B = \begin{bmatrix} 0 & 5 & 3 & -4 \\ -1 & 2 & 4 & 5 \\ 1 & -2 & 1 & 2 \end{bmatrix} \in \mathbb{Q}^{3\times 4}.$$

◁ La suma es

$$A + B = \begin{bmatrix} 2 & 5 & 8 & -5 \\ -5 & 3 & 4 & 8 \\ 2 & -5 & 8 & 0 \end{bmatrix} \in \mathbb{Q}^{3\times 4}.$$

▷

Esta operación de matrices verifica las siguientes propiedades.

Proposición 2.1. La adición de matrices verifica las siguientes propiedades:

Asociativa: $(A+B)+C = A+(B+C)$, $\forall A, B, C \in \mathbb{K}^{m\times n}$.

Elemento neutro: $A + O_{m\times n} = O_{m\times n} + A = A, \forall A \in \mathbb{K}^{m\times n}$.

Elementos opuestos: $\forall A \in \mathbb{K}^{m\times n}, \exists B \in \mathbb{K}^{m\times n} : A+B = B+A = O_{m\times n}$.

Conmutativa: $A+B = B+A$, $\forall A, B \in \mathbb{K}^{m\times n}$.

Demostración. Sean $A = [a_{ij}]$, $B = [b_{ij}]$ y $C = [c_{ij}]$ tres matrices en $\mathbb{K}^{m\times n}$. Las demostraciones se basan en utilizar la definición de adición, y luego aplicar las correspondientes propiedades en el cuerpo $\mathbb{K}$.

Para probar la propiedad asociativa se aplica dos veces la definición de adición obteniendo

$$(A+B)+C = ([a_{ij}] + [b_{ij}]) + [c_{ij}] = [a_{ij} + b_{ij}] + [c_{ij}] = [(a_{ij} + b_{ij}) + c_{ij}].$$

De manera similar,

$$A + (B+C) = [a_{ij}] + ([b_{ij}] + [c_{ij}]) = [a_{ij}] + [b_{ij} + c_{ij}] = [a_{ij} + (b_{ij} + c_{ij})].$$

Puesto que la propiedad asociativa se cumple en $\mathbb{K}$, se tiene que $(a_{ij} + b_{ij}) + c_{ij} = a_{ij} + (b_{ij} + c_{ij})$, lo que prueba la igualdad $(A+B)+C = A+(B+C)$.

Observar que habría sido más directo escribir

$$\begin{aligned}
(A+B)+C &= ([a_{ij}]+[b_{ij}])+[c_{ij}] \\
&= [a_{ij}+b_{ij}]+[c_{ij}] \\
&= [(a_{ij}+b_{ij})+c_{ij}] \\
&= [a_{ij}+(b_{ij}+c_{ij})] \\
&= [a_{ij}]+[b_{ij}+c_{ij}] \\
&= [a_{ij}]+([b_{ij}]+[c_{ij}]) \\
&= A+(B+C).
\end{aligned}$$

La matriz $O_{m\times n}$ es elemento neutro para la adición de matrices, pues para $A \in \mathbb{K}^{m\times n}$,

$$A + O_{m\times n} = [a_{ij}] + [0] = [a_{ij} + 0] = [a_{ij}] = A,$$

y, de manera similar, se prueba que $O_{m\times n} + A = A$.

La demostración de la propiedad relativa a elementos opuestos y la propiedad conmutativa se proponen como ejercicios. □

Observación 2.2. En las propiedades relativas al elemento neutro y a los elementos opuestos (a veces también llamados **simétricos**) se ha probado su existencia, pero no se ha dicho nada acerca de su unicidad.

- En el Ejercicio 2 (página 144) se probará que el elemento neutro es único (y por tanto no habrá ambigüedad a la hora de nombrar el neutro para la adición con la misma notación utilizada para la matriz nula, pues ambos son iguales).
- Para una matriz dada A, en el Ejercicio 3 (página 144) se probará la existencia de una única matriz B, opuesta a ella. Esto permitirá denotar sin ambigüedad $B = -A$, para indicar que $A+(-A) = O_{m\times n}$.

Observación 2.3. Sean $A, B \in \mathbb{K}^{m\times n}$. A partir de la notación $-A$ para la matriz opuesta de la matriz A, se puede definir la **sustracción** A menos B como

$$A - B := A + (-B).$$

La expresión $A+(-B)$ se llama la **resta** de A y B; y $A-B$ se lee A menos B.

Corolario 2.1. El conjunto $\mathbb{K}^{m\times n}$ de todas las matrices de tamaño $m \times n$ a coeficientes en $\mathbb{K}$ forma un grupo abeliano con la operación binaria de adición de matrices.

Demostración. Es un hecho inmediato que se deduce de la definición de grupo (Definición 1.32) y de la Proposición 2.1. □

2.4.2. Multiplicación de un escalar por una matriz

Definición 2.4. Se llama **multiplicación** de un escalar por una matriz a la operación externa

$$\begin{array}{rccc} \cdot & : & \mathbb{K} \times \mathbb{K}^{m\times n} & \to & \mathbb{K}^{m\times n} \\ & & (\lambda, A) & \mapsto & \cdot(\lambda, A) = \lambda A, \end{array}$$

donde λA, llamado **producto**[a] del escalar $\lambda \in \mathbb{K}$ por la matriz $A = [a_{ij}] \in \mathbb{K}^{m\times n}$, se define por

$$\lambda A = [\lambda a_{ij}] = \begin{bmatrix} \lambda a_{11} & \lambda a_{12} & \cdots & \lambda a_{1n} \\ \lambda a_{21} & \lambda a_{22} & \cdots & \lambda a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda a_{m1} & \lambda a_{m2} & \cdots & \lambda a_{mn} \end{bmatrix}.$$

[a] Si bien, la operación externa se llama multiplicación de un escalar por una matriz y el resultado de dicha operación se llama producto de λ por A, por abuso de lenguaje, se suele llamar producto en ambos casos.

Observación 2.4. Nótese que $(\lambda, A) \mapsto \lambda A$ es una operación externa, con operadores en $\mathbb{K}$, sobre cualquier conjunto de matrices, por ejemplo en el conjunto $\mathbb{K}^{m\times n}$, siendo $m, n \in \mathbb{N}$ dos cantidades fijas, en varios conjuntos del tipo anterior, o incluso, sobre el conjunto de todas las matrices de todos los tamaños.

Ejemplo 2.9. *Calcular la multiplicación de $\lambda = 4$ por la matriz*

$$A = \begin{bmatrix} 2 & 0 & 5 & -1 \\ -4 & 1 & 0 & 3 \\ 1 & -3 & 7 & -2 \end{bmatrix}.$$

◁ El producto es

$$4A = \begin{bmatrix} 8 & 0 & 20 & -4 \\ -16 & 4 & 0 & 12 \\ 4 & -12 & 28 & -8 \end{bmatrix}.$$

▷

Esta operación de matrices verifica las siguientes propiedades.

Proposición 2.2. La multiplicación de un escalar por una matriz verifica las siguientes propiedades:

<u>Distributiva respecto de la suma de escalares</u>: $(\lambda + \mu)A = \lambda A + \mu A$, $\forall \lambda, \mu \in \mathbb{K}, \forall A \in \mathbb{K}^{m\times n}$.

<u>Distributiva respecto de la suma de matrices</u>: $\lambda(A + B) = \lambda A + \lambda B, \forall \lambda \in \mathbb{K}, \forall A, B \in \mathbb{K}^{m\times n}$.

<u>Pseudoasociativa</u>: $(\lambda\mu)A = \lambda(\mu A)$, $\forall \lambda, \mu \in \mathbb{K}, \forall A \in \mathbb{K}^{m\times n}$.

<u>Multiplicación del neutro de $\mathbb{K}$ por una matriz</u>: $1_{\mathbb{K}} A = A$, $\forall A \in \mathbb{K}^{m\times n}$.

Demostración. Las demostraciones se basan en utilizar la definición de multiplicación de un escalar por una matriz, y luego aplicar las correspondientes propiedades en el cuerpo $\mathbb{K}$.

Sean $A = [a_{ij}] \in \mathbb{K}^{m\times n}$ y $\lambda, \mu \in \mathbb{K}$. Entonces

$$\begin{aligned}(\lambda+\mu)A &= (\lambda+\mu)[a_{ij}]\\ &= [(\lambda+\mu)a_{ij}]\\ &= [\lambda a_{ij} + \mu a_{ij}]\\ &= [\lambda a_{ij}] + [\mu a_{ij}]\\ &= \lambda[a_{ij}] + \mu[a_{ij}]\\ &= \lambda A + \mu A.\end{aligned}$$

La restantes demostraciones se proponen como ejercicios. □

2.4.3. Producto de matrices

En 1855, Cayley consideró las aplicaciones (lineales) $f, g : \mathbb{R}^2 \to \mathbb{R}^2$ definidas por

$$f(x,y) = (ax+by, cx+dy) \qquad \text{y} \qquad g(x,y) = (\alpha x + \beta y, \gamma x + \delta y),$$

donde $a, b, c, d, \alpha, \beta, \gamma, \delta \in \mathbb{R}$ están fijos, y al realizar la composición obtuvo

$$\begin{aligned}(g\circ f)(x,y) &= g(f(x,y))\\ &= g((ax+by, cx+dy))\\ &= (\alpha(ax+by)+\beta(cx+dy), \gamma(ax+by)+\delta(cx+dy))\\ &= ((\alpha a+\beta c)x + (\alpha b+\beta d)y, (\gamma a+\delta c)x + (\gamma b+\delta d)y).\end{aligned}$$

Si se realizan las siguientes identificaciones de funciones con matrices[5]

$$g \leftrightarrow \begin{bmatrix}\alpha & \beta\\ \gamma & \delta\end{bmatrix} \qquad \text{y} \qquad f \leftrightarrow \begin{bmatrix} a & b\\ c & d\end{bmatrix}$$

[5] En 1801, Gauss ya había realizado esta composición de aplicaciones lineales pero no la identificación con matrices.

se observa que a la composición le corresponde la matriz

$$g \circ f \leftrightarrow \begin{bmatrix} \alpha a + \beta c & \alpha b + \beta d \\ \gamma a + \delta c & \gamma b + \delta d \end{bmatrix},$$

que es lo que permitirá introducir la operación de producto de matrices en el caso general.

Si ahora se considera una aplicación (lineal, como antes) $g : \mathbb{R}^n \to \mathbb{R}^m$, para que la composición esté bien definida, debe ocurrir que $f : \mathbb{R}^p \to \mathbb{R}^n$ (también lineal), de modo que $g \circ f : \mathbb{R}^p \to \mathbb{R}^m$. Realizando el mismo razonamiento que antes, ahora en general, este hecho sugiere que se defina el producto de dos matrices del siguiente modo.

Definición 2.5. Sean las matrices $A = [a_{ij}] \in \mathbb{K}^{m\times n}$ y $B = [b_{ij}] \in \mathbb{K}^{n\times p}$. Se llama **producto** de A por B a la matriz de tamaño $m \times p$, denotada por AB, definida como la matriz $C = [c_{ij}]$ donde, para cada $i = 1, 2, \dots, m$ y para cada $j = 1, 2, \dots, p$, el elemento c_{ij} es

$$c_{ij} := \sum_{k=1}^{n} a_{ik}b_{kj} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj}.$$

En símbolos,

$$AB = \begin{bmatrix} a_{11}b_{11} + \cdots + a_{1n}b_{n1} & \cdots & a_{11}b_{1p} + \cdots + a_{1n}b_{np} \\ a_{21}b_{11} + \cdots + a_{2n}b_{n1} & \cdots & a_{21}b_{1p} + \cdots + a_{2n}b_{np} \\ \vdots & \ddots & \vdots \\ a_{m1}b_{11} + \cdots + a_{mn}b_{n1} & \cdots & a_{m1}b_{1p} + \cdots + a_{mn}b_{np} \end{bmatrix}.$$

Observar que, para poder definir el producto de dos matrices, la cantidad de columnas de la primera matriz debe coincidir con la cantidad de filas de la segunda.

Observación 2.5. Es claro, de la definición anterior, que la operación $(A, B) \mapsto AB$ no es una operación binaria (sobre ninguno de los espacios considerados, ni en $\mathbb{K}^{m\times n}$, ni en $\mathbb{K}^{n\times p}$, ni sobre $\mathbb{K}^{m\times p}$). Para conseguir una operación binaria, se debe operar con matrices cuadradas del mismo tamaño.

Definición 2.6. Se llama **multiplicación** de matrices a la operación binaria

$$\begin{array}{rccc} \cdot \ : & \mathbb{K}^{n\times n} \times \mathbb{K}^{n\times n} & \to & \mathbb{K}^{n\times n} \\ & (A, B) & \mapsto & \cdot(A, B) = AB, \end{array}$$

donde AB es el **producto**[a] de A por B de la Definición 2.5 (restringido al caso $m = n = p$).

[a] Si bien, la operación binaria se llama multiplicación y el resultado de dicha operación se llama producto, por abuso de lenguaje, se suele llamar producto en ambos casos.

Ejemplo 2.10. *Analizar si es posible realizar las multiplicaciones AB y BA siendo $A \in \mathbb{R}^{3\times 1}$ y $B \in \mathbb{R}^{1\times 2}$.*

$\lhd$ Una matriz $A \in \mathbb{R}^{3\times 1}$ se puede multiplicar por una matriz $B \in \mathbb{R}^{1\times 2}$, puesto que la cantidad de columnas de A coincide con la cantidad de filas de B, con lo que $AB \in \mathbb{R}^{3\times 2}$.

Sin embargo, no es posible calcular BA, puesto que B tiene 2 columnas y no coincide con la cantidad de filas de A, que es 3. $\rhd$

Ejemplo 2.11. *Realizar el producto de*

$$A = \begin{bmatrix} 2 & 0 & 5 & -1 \\ -4 & 1 & 0 & 3 \\ 1 & -3 & 7 & -2 \end{bmatrix} \in \mathbb{R}^{3\times 4} \quad \textit{por} \quad B = \begin{bmatrix} 0 & 5 \\ -1 & 2 \\ 1 & -2 \\ 0 & -1 \end{bmatrix} \in \mathbb{R}^{4\times 2}.$$

$\lhd$ El producto de las matrices AB es

$$AB = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \\ c_{31} & c_{32} \end{bmatrix} = \begin{bmatrix} 5 & 1 \\ -1 & -21 \\ 10 & -13 \end{bmatrix}.$$

donde

$$\begin{aligned} c_{11} &= 2\cdot 0 + 0\cdot(-1) + 5\cdot 1 + (-1)\cdot 0 = 5, \\ c_{12} &= 2\cdot 5 + 0\cdot 2 + 5\cdot(-2) + (-1)\cdot(-1) = 1, \\ c_{21} &= (-4)\cdot 0 + 1\cdot(-1) + 0\cdot 1 + 3\cdot 0 = -1, \\ c_{22} &= (-4)\cdot 5 + 1\cdot 2 + 0\cdot(-2) + 3\cdot(-1) = -21, \\ c_{31} &= 1\cdot 0 + (-3)\cdot(-1) + 7\cdot 1 + (-2)\cdot 0 = 10, \\ c_{32} &= 1\cdot 5 + (-3)\cdot 2 + 7\cdot(-2) + (-2)\cdot(-1) = -13. \end{aligned}$$

$\rhd$

El producto de matrices verifica las siguientes propiedades.

Proposición 2.3. Si todas las matrices involucradas tienen tamaños adecuados entonces el producto de matrices verifica las siguientes propiedades:

Asociativa: $(AB)C = A(BC)$.

Elemento neutro a izquierda y derecha: $I_m A = A = A I_n$, $\forall A \in \mathbb{K}^{m\times n}$.

Distributiva con respecto a la suma a izquierda: $A(B+C) = AB + AC$.

Distributiva con respecto a la suma a derecha: $(A+B)C = AC+BC$.

Pseudoasociativa: $\lambda(AB) = (\lambda A)B$.

Elemento absorbente: $O_{m\times m}A = O_{m\times n} = AO_{n\times n}$, $\forall A \in \mathbb{K}^{m\times n}$.

Demostración. Sean $A = [a_{ij}] \in \mathbb{K}^{m\times n}$, $B = [b_{ij}] \in \mathbb{K}^{n\times p}$ y $C = [c_{ij}] \in \mathbb{K}^{p\times q}$. Para las demostraciones se utilizará la definición de producto, y luego se aplicarán las correspondientes propiedades en $\mathbb{K}$.

Para probar la propiedad asociativa se debe tener en cuenta que $AB \in \mathbb{K}^{m\times p}$, $BC \in \mathbb{K}^{n\times q}$ y se aplica la definición de producto cuatro veces para obtener

$$\begin{aligned}
(AB)C &= ([a_{ij}][b_{ij}])[c_{ij}] \\
&= \left[\sum_{k=1}^{n} a_{ik}b_{kj}\right][c_{ij}] \\
&= \left[\sum_{\ell=1}^{p}\left(\sum_{k=1}^{n} a_{ik}b_{k\ell}\right)c_{\ell j}\right] \\
&= \left[\sum_{k=1}^{n} a_{ik}\left(\sum_{\ell=1}^{p} b_{k\ell}c_{\ell j}\right)\right] \\
&= [a_{ij}]\left[\sum_{\ell=1}^{p} b_{i\ell}c_{\ell j}\right] \\
&= [a_{ij}]\left([b_{ij}][c_{ij}]\right) \\
&= A(BC).
\end{aligned}$$

Es importante observar que los sumatorios anteriores se han podido intercambiar en virtud de las propiedades distributiva y asociativa que se cumplen en el cuerpo $\mathbb{K}$.

Para probar que I_m es un elemento neutro a izquierda de A se considera la notación con la delta de Kronecker $I = [\delta_{ij}]$. En efecto,

$$\begin{aligned} I_m A &= [\delta_{ij}][a_{ij}] \\ &= \left[\sum_{k=1}^{m} \delta_{ik} a_{kj}\right] \\ &= [\delta_{i1}a_{1j} + \delta_{i2}a_{2j} + \cdots + \delta_{ii}a_{ij} + \cdots + \delta_{im}a_{mj}] \\ &= [a_{ij}] \\ &= A. \end{aligned}$$

Del mismo modo se prueba que $AI_n = A$.

La demostración de la restantes propiedades se proponen como ejercicios.

□

Observación 2.6. En general, el producto de matrices no cumple la propiedad conmutativa. El siguiente contraejemplo lo constata:

$$\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix},$$

y

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

Observación 2.7. Obsérvese, también, que (a diferencia de lo que ocurre en los cuerpos $\mathbb{R}$ y $\mathbb{C}$, por ejemplo) puede ocurrir que el producto de matrices sea la matriz nula mientras que ninguna de ellas lo sea. Es decir, en $\mathbb{K}^{m\times n}$ hay divisores de cero (con $m > 1$ ó $n > 1$). Por la Proposición 1.2, $\mathbb{K}^{n\times n}$ no puede ser un cuerpo con dichas operaciones.

Corolario 2.2. El conjunto $\mathbb{K}^{n\times n}$ con la adición y la multiplicación de matrices es un anillo con unidad y es no conmutativo si $n > 1$.

Demostración. Es inmediato del Corolario 2.1 y de la Proposición 2.3. □

2.4.4. Potenciación de matrices cuadradas

Definición 2.7. Se llama **potenciación** de matrices a la operación externa

$$\begin{array}{rcccc} \hat{} & : & (\{0\} \cup \mathbb{N}) \times \mathbb{K}^{n\times n} & \to & \mathbb{K}^{n\times n} \\ & & (k, A) & \mapsto & \hat{}(k, A) = A^k, \end{array}$$

donde A^k, llamada **potencia** k-ésima de A, y denotada por A^k, es la matriz de tamaño $n \times n$ definida por

$$\begin{array}{rcl} A^0 & = & I_n, \\ A^k & = & A^{k-1}A, \quad k = 1, 2, 3, \ldots \end{array}$$

Por convención, $O_n^0 = I_n$.

La definición recursiva anterior expresa que para calcular la potencia k-ésima de A se debe realizar el producto de A por sí misma k veces.

Ejemplo 2.12. *Calcular* A^2 *y* A^3 *siendo*

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

◁ El cuadrado de la matriz A es

$$A^2 = AA = \begin{bmatrix} 4 & 0 \\ 9 & 1 \end{bmatrix}$$

y su potencia cúbica es

$$A^3 = A^2 A = \begin{bmatrix} 8 & 0 \\ 21 & 1 \end{bmatrix}.$$

▷

Esta operación de matrices verifica las siguientes propiedades.

Proposición 2.4. Sea $A \in \mathbb{K}^{n\times n}$ y sean $k, r \in \{0\} \cup \mathbb{N}$. Entonces

Primera potencia: $A^1 = A$.

Producto de potencias: $A^k A^r = A^{k+r}$.

Potencia de otra potencia: $(A^k)^r = A^{kr}$.

Demostración. Es claro que, por definición, $A^1 = A^0 A = I_n A = A$.

Para demostrar la segunda propiedad se comienza por los casos triviales $k = 0$ y $r = 0$. En efecto,

- $k = 0, r \geq 0$: $A^{k+r} = A^{0+r} = A^r = I_n A^r = A^0 A^r = A^k A^r$.
- $k \geq 0, r = 0$: $A^{k+r} = A^{k+0} = A^k = A^k I_n = A^k A^0 = A^k A^r$.

Falta demostrar la propiedad para los casos $k, r \in \mathbb{N}$. Sea ahora $k \in \mathbb{N}$ fijo pero arbitrario. Se demostrará por inducción sobre r. Para ello, se considera la función proposicional

$$P(r) : A^k A^r = A^{k+r}, \text{ con } r \in \mathbb{N}.$$

- Caso base: $P(1)$ es verdadera. En efecto, utilizando la propiedad correspondiente a la potencia 1 y la definición de potencia se tiene que $A^k A^1 = A^k A = A^{k+1}$.
- Paso inductivo: Sea $s > 1$, se supone que $P(s)$ es verdadera y se debe probar que $P(s+1)$ es verdadera. En efecto, por hipótesis de inducción, se supone que $A^k A^s = A^{k+s}$. Se debe probar que $A^k A^{s+1} = A^{k+(s+1)}$. En efecto, $A^k A^{s+1} = A^k(A^s A) = (A^k A^s)A = A^{k+s} A = A^{(k+s)+1}$, donde se ha aplicado la definición de potencia y la propiedad asociativa.

De este modo, por el principio de inducción, $P(r)$ es verdadera para todo $r \in \mathbb{N}$. Como, además, k es fijo pero arbitrario, también es válida la propiedad para todo $k \in \mathbb{N}$.

La demostración restante se propone como ejercicio. □

2.4.5. Trasposición de matrices

Definición 2.8. Sea $A = [a_{ij}] \in \mathbb{K}^{m\times n}$. Se llama matriz **traspuesta** de A a la matriz de tamaño $n \times m$, denotada por A^t, definida por $A^t = [a^t_{ij}]$ con $a^t_{ij} = a_{ji}$. En términos de los elementos a_{ij} de A se tiene

$$A^t = [a^t_{ij}] = [a_{ji}] = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{bmatrix}.$$

Para encontrar la traspuesta de una matriz basta escribir la primera fila de A como primera columna de A^t, la segunda fila de A como segunda columna de A^t y así siguiendo hasta la última fila de A. Al proceso de calcular la matriz traspuesta de una matriz dada se lo denomina **trasposición**.

Ejemplo 2.13. *Calcular la traspuesta de la matriz*

$$A = \begin{bmatrix} 2 & 0 & 5 \\ -4 & 1 & 0 \\ -11 & -3 & 7 \\ 3 & 1 & -2 \end{bmatrix}.$$

◁ La traspuesta de A es

$$A^t = \begin{bmatrix} 2 & -4 & -11 & 3 \\ 0 & 1 & -3 & 1 \\ 5 & 0 & 7 & -2 \end{bmatrix}.$$

▷

Al calcular la matriz traspuesta, tras aplicar las operaciones definidas anteriormente, se cumplen las siguientes propiedades.

Proposición 2.5. La trasposición de matrices satisface las siguientes propiedades:

Trasposición de la adición: $(A+B)^t = A^t + B^t$, $\forall A, B \in \mathbb{K}^{m\times n}$.

Trasposición del producto de un escalar por una matriz:

$$(\lambda A)^t = \lambda A^t, \quad \forall \lambda \in \mathbb{K}, \forall A \in \mathbb{K}^{m\times n}.$$

Trasposición del producto:

$$(AB)^t = B^t A^t, \quad \forall A \in \mathbb{K}^{m\times n}, B \in \mathbb{K}^{n\times p}.$$

Trasposición de una trasposición: $(A^t)^t = A$, $\forall A \in \mathbb{K}^{m\times n}$.

Demostración. Las demostraciones se basan en la utilización de la definición de la traspuesta de una matriz, y luego de la aplicación de las correspondientes propiedades en el cuerpo $\mathbb{K}$.

Sean $A = [a_{ij}]$ y $B = [b_{ij}]$ dos matrices en $\mathbb{K}^{m\times n}$. Entonces, denotando $s_{ij} := a_{ij} + b_{ij}$, se tiene

$$\begin{aligned}
(A+B)^t &= ([a_{ij}] + [b_{ij}])^t \\
&= [a_{ij} + b_{ij}]^t \\
&= [s_{ij}]^t \\
&= [s^t_{ij}] \\
&= [s_{ji}] \\
&= [a_{ji} + b_{ji}] \\
&= [a_{ji}] + [b_{ji}] \\
&= [a^t_{ij}] + [b^t_{ij}] \\
&= A^t + B^t.
\end{aligned}$$

Sean ahora $A \in \mathbb{K}^{m\times n}$ y $B \in \mathbb{K}^{n\times p}$. Entonces, se tiene que

$$\begin{aligned}
(AB)^t &= ([a_{ij}][b_{ij}])^t \\
&= \left[\sum_{k=1}^{n} a_{ik}b_{kj}\right]^t \\
&= \left[\left(\sum_{k=1}^{n} a_{ik}b_{kj}\right)^t\right] \\
&= \left[\sum_{k=1}^{n} a_{jk}b_{ki}\right] \\
&= \left[\sum_{k=1}^{n} b_{ki}a_{jk}\right] \\
&= \left[\sum_{k=1}^{n} b^t_{ik}a^t_{kj}\right] \\
&= [b^t_{ij}][a^t_{ij}] \\
&= B^tA^t.
\end{aligned}$$

La restantes demostraciones se proponen como ejercicios. □

2.5. Matrices simétricas y antisimétricas. Traza de una matriz

En esta sección se presentan los conceptos de matriz simétrica, matriz antisimétrica y traza de una matriz.

2.5.1. Matrices simétricas y antisimétricas

Definición 2.9. Sea $A = [a_{ij}] \in \mathbb{K}^{n\times n}$. La matriz A se llama **simétrica** si coincide con su traspuesta. En símbolos, si $A^t = A$, o bien si $a_{ji} = a_{ij}$ para todo $i, j = 1, 2, \ldots, n$.

Observar que una condición necesaria para que una matriz sea simétrica es que sea cuadrada.

Ejemplo 2.14. *La matriz*

$$A = \begin{bmatrix} 2 & 0 & -1 \\ 0 & 1 & 3 \\ -1 & 3 & 7 \end{bmatrix} \text{ es simétrica}$$

mientras que

$$B = \begin{bmatrix} 2 & 10 & -1 \\ 0 & 1 & -3 \\ -1 & -3 & 7 \end{bmatrix} \text{ no es simétrica.}$$

$\triangleleft$ Es fácil ver que $A^t = A$ y $B^t \neq B$ (observando en B los elementos b_{12} y b_{21}). $\triangleright$

Definición 2.10. Sea $A = [a_{ij}] \in \mathbb{K}^{n\times n}$. La matriz A se llama **antisimétrica** si coincide con la opuesta de su traspuesta. En símbolos, si $A^t = -A$, o bien si $a_{ji} = -a_{ij}$ para todo $i, j = 1, 2, \ldots, n$.

Observar que una condición necesaria para que una matriz sea antisimétrica es que sea cuadrada.

Observación 2.8. Si una matriz $A \in \mathbb{K}^{n\times n}$ es antisimétrica y en $\mathbb{K}$ se cumple que[a] $1 + 1 \neq 0$ entonces los elementos de la diagonal deben valer cero. En efecto, los elementos de la diagonal se obtienen haciendo $i = j$ en los subíndices de los elementos de A. Luego, $a_{ii} = -a_{ii}$ implica $a_{ii} + a_{ii} = 0$, de donde $(1+1)a_{ii} = 0$. Por el Corolario 1.1, $\mathbb{K}$ no tiene divisores de cero. Al ser $1 + 1 \neq 0$, se obtiene $a_{ii} = 0$, y ese hecho es válido para todo $i = 1, 2, \ldots, n$.

[a] Es decir, $\mathbb{K}$ es un cuerpo de característica distinta de 2.

Ejemplo 2.15. *La matriz*

$$A = \begin{bmatrix} 0 & 4 & -1 \\ -4 & 0 & -3 \\ 1 & 3 & 0 \end{bmatrix} \text{ es antisimétrica}$$

mientras que

$$B = \begin{bmatrix} 2 & 10 & -1 \\ -10 & 0 & -3 \\ 1 & 3 & 0 \end{bmatrix} \text{ no es antisimétrica.}$$

$\lhd$ Es fácil ver que $A^t = -A$ y $B^t \neq -B$ (observando en B el elemento b_{11}). $\rhd$

Las matrices simétricas y antisimétricas cumplen las siguientes propiedades.

Proposición 2.6. Sea $A \in \mathbb{K}^{m\times n}$. Entonces:

(a) Si $m = n$ se tiene que $A + A^t$ es simétrica.

(b) Si $m = n$ se tiene que $A - A^t$ es antisimétrica.

(c) $AA^t \in \mathbb{K}^{m\times m}$ y es simétrica.

(d) $A^tA \in \mathbb{K}^{n\times n}$ y es simétrica.

Demostración. Sólo se probará la primera de las propiedades. En efecto, utilizando las propiedades de la trasposición se tiene

$$(A + A^t)^t = A^t + (A^t)^t = A^t + A = A + A^t.$$

La restantes demostraciones se proponen como ejercicios. $\square$

2.5.2. Traza de una matriz cuadrada

Definición 2.11. Sea $A = [a_{ij}] \in \mathbb{K}^{n\times n}$. Se llama **traza** de la matriz A, y se denota $\operatorname{tr}(A)$, al escalar

$$\operatorname{tr}(A) = a_{11} + a_{22} + \cdots + a_{nn} = \sum_{r=1}^{n} a_{rr}.$$

Ejemplo 2.16. *Calcular la traza de*

- *la matriz identidad* I_n.
- *la matriz nula* $O_{n\times n}$.
- *la matriz* $A = \begin{bmatrix} -1 & 4 \\ -4 & 4 \end{bmatrix} \in \mathbb{R}^{2\times 2}$.

$\lhd$ Se tiene que: $\operatorname{tr}(I_n) = n$, $\operatorname{tr}(O_{n\times n}) = 0$ y $\operatorname{tr}(A) = -1 + 4 = 3$. $\rhd$

La traza de una matriz cumple las siguientes propiedades.

Proposición 2.7. Sea $A, B \in \mathbb{K}^{n\times n}$ y $\lambda \in \mathbb{K}$. Entonces

<u>Traza de la suma</u>: $\operatorname{tr}(A + B) = \operatorname{tr}(A) + \operatorname{tr}(B)$.

<u>Traza del producto de un escalar por una matriz</u>: $\operatorname{tr}(\lambda A) = \lambda\operatorname{tr}(A)$.

<u>Traza de un producto</u>: $\operatorname{tr}(AB) = \operatorname{tr}(BA)$.

Demostración. Sólo se probará la tercera de las propiedades. En efecto, llamando $C := AB = [c_{ij}]$ y $D := BA = [d_{ij}]$ se tiene $\operatorname{tr}(AB) = \operatorname{tr}(C) =$

$\sum_{r=1}^{n} c_{rr}$. Luego, utilizando la definición de producto,

$$\begin{aligned}
\operatorname{tr}(AB) &= \sum_{r=1}^{n}\left(\sum_{k=1}^{n} a_{rk}b_{kr}\right)\\
&= \sum_{k=1}^{n}\left(\sum_{r=1}^{n} b_{kr}a_{rk}\right)\\
&= \sum_{k=1}^{n} d_{kk}\\
&= \operatorname{tr}(BA).
\end{aligned}$$

Es importante observar que los sumatorios anteriores se han podido intercambiar en virtud de la propiedad asociativa y conmutativa de la suma válidas en el cuerpo $\mathbb{K}$.

La restantes demostraciones se proponen como ejercicios. □

2.6. Partición de matrices en bloques

En Informática, la computación en paralelo, por ejemplo, requiere de la resolución de problemas de grandes dimensiones por lo que es conveniente contar con técnicas que permitan tratar un problema matricial a partir de problemas de tamaños menores y poder distrubuir las tareas entre varios procesadores. En Análisis Numérico suelen aparecer matrices vacías, es decir, matrices de gran tamaño con muchos ceros (muchas entradas nulas); y si es posible determinar estructuras donde estos bloques de ceros estén agrupados, se podrán encontrar estrategias para resolver más eficientemente los problemas que las involucran. Para ello, la partición de una matriz en bloques de tamaños adecuados proporciona una herramienta útil que permite explotar posibles propiedades de la estructura que presenta la matriz original. Estas son sólo algunas aplicaciones de la partición de matrices en bloques.

En aspectos de Ingeniería como comunicaciones, electrónica, cálculo de vigas, etc. es adecuado recurrir a la partición de una matriz en bloques para simplificar algunos problemas. En Geodesia se utiliza para calcular con precisión las coordenadas de puntos en la superficie de la Tierra, así como en Administración de Empresas en la aplicación de métodos estadísticos que requieren del cálculo de matrices de covarianza y el estudio de las matrices de correlación (por bloques).

Desde un punto de vista algebraico, presenta grandes ventajas, por ejemplo, a la hora de trabajar con la matriz aumentada en un sistema de ecuaciones lineales (véase Capítulo 4.2), al buscar formas canónicas (véase Capítulo 6), etc. También, se utilizan en numerosas áreas muy activas de investigación en Teoría de Matrices actuales.

En este apartado se mostrará cómo operar con matrices particionadas en bloques.

La idea intuitiva es dividir una matriz en *bloques* de modo que estos queden bien determinados mediante las dos reglas siguientes:

- todos los bloques que conforman una nueva fila deben tener la misma cantidad de filas,
- todos los bloques que conforman una nueva columna deben tener la misma cantidad de columnas.

Se busca, por ejemplo, algo del tipo

$$A = \left[\begin{array}{cc|ccc|c} 1 & 2 & 5 & 6 & 7 & 11 \\ 3 & 4 & 8 & 9 & 10 & 12 \\ \hline 13 & 14 & 19 & 20 & 21 & 28 \\ 15 & 16 & 22 & 23 & 24 & 29 \\ 17 & 18 & 25 & 26 & 27 & 30 \end{array}\right] = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \end{bmatrix}, \qquad (2.3)$$

donde los bloques en que se ha particionado la matriz A son:

$$A_{11} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \in \mathbb{R}^{2\times 2}, A_{12} = \begin{bmatrix} 5 & 6 & 7 \\ 8 & 9 & 10 \end{bmatrix} \in \mathbb{R}^{2\times 3}, A_{13} = \begin{bmatrix} 11 \\ 12 \end{bmatrix} \in \mathbb{R}^{2\times 1},$$

$$A_{21} = \begin{bmatrix} 13 & 14 \\ 15 & 16 \\ 17 & 18 \end{bmatrix} \in \mathbb{R}^{3\times 2}, A_{22} = \begin{bmatrix} 19 & 20 & 21 \\ 22 & 23 & 24 \\ 25 & 26 & 27 \end{bmatrix} \in \mathbb{R}^{3\times 3}, A_{23} = \begin{bmatrix} 28 \\ 29 \\ 30 \end{bmatrix} \in \mathbb{R}^{3\times 1}.$$

Definición 2.12. Sea $A \in \mathbb{K}^{m\times n}$. Una **submatriz** de A es una matriz de menor tamaño que el de A que se obtiene seleccionando los elementos que se encuentran en las intersecciones de un conjunto determinado de filas y de columnas de A. Cuando las filas seleccionadas son consecutivas y las columnas también, la submatriz de A se denomina **bloque**.

Por ejemplo, en la matriz A anterior, la siguiente es una submatriz:

$$A([2:4],[2,5]) = \begin{bmatrix} 4 & 10 \\ 14 & 21 \\ 16 & 24 \end{bmatrix},$$

donde el significado de las notaciones es:

- $[r:s]$ (para $r, s \in \mathbb{N}$ con $r < s$) indica que corresponde a las filas (por encontrarse en la primera posición de $A(\cdot,\cdot)$) consecutivas $r, r+1, r+2, \ldots, s$,
- $[r_1, r_2, \ldots, r_t]$ indica que corresponde a las columnas (por encontrarse en la segunda posición de $A(\cdot,\cdot)$), no necesariamente consecutivas $r_1, r_2, \ldots, r_t$ donde los números naturales $r_1, r_2, \ldots, r_t$ están ordenados por $r_1 < r_2 < \ldots < r_t$.

Otra submatriz de la misma matriz A es

$$A([1:2],[1:3]) = \begin{bmatrix} 1 & 2 & 5 \\ 3 & 4 & 8 \end{bmatrix},$$

que constituye un bloque por haber seleccionado filas consecutivas y columnas consecutivas.

Siguiendo esta notación, los bloques en que se ha particionado la matriz A anterior se escriben (a partir de filas consecutivas y columnas consecutivas) como

$$A_{11} = A([1:2],[1:2]), \qquad A_{12} = A([1:2],[3:5]), \qquad A_{13} = A([1:2],[6]),$$

$A_{21} = A([3:5],[1:2]), \qquad A_{22} = A([3:5],[3:5]), \qquad A_{23} = A([3:5],[6]).$

Definición 2.13. Se dice que una matriz $A \in \mathbb{K}^{m\times n}$ presenta una descomposición[a] en una **partición del tipo** $(f_1 + f_2 + \cdots + f_r) \times (c_1 + c_2 + \cdots + c_s)$ si el elemento del lugar (i, j) es el bloque A_{ij} con $i = 1, 2, \ldots, r$ y $j = 1, 2, \ldots, s$, es decir

$$A = \begin{bmatrix} A_{11} & A_{12} & \ldots & A_{1s} \\ A_{21} & A_{22} & \ldots & A_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ A_{r1} & A_{r2} & \ldots & A_{rs} \end{bmatrix} \in \mathbb{K}^{(f_1+f_2+\cdots+f_r)\times(c_1+c_2+\cdots+c_s)},$$

y se cumple que:

- $m = f_1 + f_2 + \cdots + f_r$, con $f_i \in \mathbb{N}$ para todo $i = 1, 2, \ldots, r$, $n = c_1 + c_2 + \cdots + c_s$, con $c_j \in \mathbb{N}$ para todo $j = 1, 2, \ldots, s$, con las cantidades f_i y c_j en ese orden,
- el bloque $A_{ij} \in \mathbb{K}^{f_i \times c_j}$ contiene los elementos de:
  - las f_i filas consecutivas que aparecen en A después de las $f_1 + \cdots + f_{i-1}$ filas anteriores si $i > 1$ o bien, si $i = 1$, sólo de las f_1 filas consecutivas (es decir, en matrices del tipo A_{1j}),
  - las c_j columnas consecutivas que aparecen en A después de las $c_1 + \cdots + c_{j-1}$ columnas anteriores si $j > 1$ o bien, si $j = 1$, sólo de las c_1 columnas consecutivas (es decir, en matrices del tipo A_{i1}).

Cuando los tamaños de los bloques están claros del contexto se dice simplemente que la **matriz** A está **particionada** o que está **dividida en bloques**.

[a] O bien, está dividida mediante una partición del tipo $(f_1 + f_2 + \cdots + f_r) \times (c_1 + c_2 + \cdots + c_s)$.

Para indicar que la matriz $A \in \mathbb{R}^{5\times 6}$ del ejemplo anterior está particionada se escribe

$$A = \left[\begin{array}{cc|ccc|c} 1 & 2 & 5 & 6 & 7 & 11 \\ 3 & 4 & 8 & 9 & 10 & 12 \\ \hline 13 & 14 & 19 & 20 & 21 & 28 \\ 15 & 16 & 22 & 23 & 24 & 29 \\ 17 & 18 & 25 & 26 & 27 & 30 \end{array}\right] = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \end{bmatrix} \in \mathbb{R}^{(2+3)\times(2+3+1)},$$

donde $f_1 = 2$, $f_2 = 3$, $c_1 = 2$, $c_2 = 3$ y $c_3 = 1$. El bloque A_{21} se encuentra en el lugar $(2, 1)$ de la matriz A expresada según sus bloques A_{ij} en (2.3), tiene tamaño $f_2 \times c_1 = 3 \times 2$, y contiene los elementos de las 3 filas consecutivas que figuran en A después de las $f_1 = 2$ primeras filas y en las $c_1 = 2$ primeras columnas consecutivas de A.

2.6.1. Suma de matrices particionadas

Sean $A, B \in \mathbb{K}^{m\times n}$ dos matrices, ambas descompuestas según una partición del tipo $(f_1 + f_2 + \cdots + f_r) \times (c_1 + c_2 + \cdots + c_s)$. La suma de las matrices A y B puede realizarse por bloques de la siguiente forma:

$$A + B = [A_{ij} + B_{ij}] = \begin{bmatrix} A_{11} + B_{11} & A_{12} + B_{12} & \dots & A_{1s} + B_{1s} \\ A_{21} + B_{21} & A_{22} + B_{22} & \dots & A_{2s} + B_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ A_{r1} + B_{r1} & A_{r2} + B_{r2} & \dots & A_{rs} + B_{rs} \end{bmatrix}.$$

2.6.2. Producto de un escalar por una matriz particionada

Sea $A \in \mathbb{K}^{m\times n}$ una matriz descompuesta según una partición del tipo $(f_1 + f_2 + \cdots + f_r) \times (c_1 + c_2 + \cdots + c_s)$ y sea $\lambda \in \mathbb{K}$. El producto del escalar

λ por la matriz A puede realizarse por bloques de la siguiente forma:

$$\lambda A = [\lambda A_{ij}] = \begin{bmatrix} \lambda A_{11} & \lambda A_{12} & \dots & \lambda A_{1s} \\ \lambda A_{21} & \lambda A_{22} & \dots & \lambda A_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda A_{r1} & \lambda A_{r2} & \dots & \lambda A_{rs} \end{bmatrix}.$$

2.6.3. Producto de matrices particionadas

Sean $A \in \mathbb{K}^{m\times n}$ y $B \in \mathbb{K}^{n\times p}$ dos matrices descompuestas según las particiones del tipo $m = f_1+f_2+\cdots+f_r$, $n = c_1+c_2+\cdots+c_s$ y $p = q_1+q_2+\cdots+q_\ell$. Teniendo en cuenta la propiedad asociativa de la suma de matrices, el producto de A por B puede realizarse por bloques de la siguiente forma:

$$\begin{aligned} AB &= \left[\sum_{k=1}^{s} A_{ik}B_{kj}\right] \qquad (2.4) \\ &= \begin{bmatrix} A_{11}B_{11}+\cdots+A_{1s}B_{s1} & \cdots & A_{11}B_{1\ell}+\cdots+A_{1s}B_{s\ell} \\ A_{21}B_{11}+\cdots+A_{2s}B_{s1} & \cdots & A_{21}B_{1\ell}+\cdots+A_{2s}B_{s\ell} \\ \vdots & \ddots & \vdots \\ A_{r1}B_{11}+\cdots+A_{rs}B_{s1} & \cdots & A_{r1}B_{1\ell}+\cdots+A_{rs}B_{s\ell} \end{bmatrix}. \end{aligned}$$

Ejemplo 2.17. *Calcular AB para las siguientes matrices particionadas*

$$A = \left[\begin{array}{rr|rrr} 2 & 0 & 3 & 3 & 1 \\ 0 & 2 & 3 & 3 & 1 \\ \hline -1 & 6 & 1 & 0 & 4 \end{array}\right] = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = \begin{bmatrix} 2I_2 & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

y

$$B = \left[\begin{array}{rr|rr} 1 & 0 & 1 & 4 \\ 0 & 1 & 6 & 3 \\ \hline 0 & 0 & 1 & 1 \\ 0 & 0 & 2 & 2 \\ 0 & 0 & 4 & 5 \end{array}\right] = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} = \begin{bmatrix} I_2 & B_{12} \\ O_{3\times 2} & B_{22} \end{bmatrix}.$$

◁ Las siguientes operaciones por bloques se pueden realizar de manera sencilla:

$$A_{11}B_{11} + A_{12}B_{21} = \underbrace{\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}}_{2I_2} \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}}_{I_2} + \begin{bmatrix} 3 & 3 & 1 \\ 3 & 3 & 1 \end{bmatrix} \underbrace{\begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}}_{O_{3\times 2}} = 2I_2,$$

$$A_{21}B_{11} + A_{22}B_{21} = \begin{bmatrix} -1 & 6 \end{bmatrix} \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}}_{I_2} + \begin{bmatrix} 1 & 0 & 4 \end{bmatrix} \underbrace{\begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}}_{O_{3\times 2}} = \begin{bmatrix} -1 & 6 \end{bmatrix}.$$

Calculando, además,

$$A_{11}B_{12} + A_{12}B_{22} = \underbrace{\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}}_{2I_2} \begin{bmatrix} 1 & 4 \\ 6 & 3 \end{bmatrix} + \begin{bmatrix} 3 & 3 & 1 \\ 3 & 3 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 2 & 2 \\ 4 & 5 \end{bmatrix} = \begin{bmatrix} 15 & 22 \\ 25 & 20 \end{bmatrix}$$

y

$$A_{21}B_{12} + A_{22}B_{22} = \begin{bmatrix} -1 & 6 \end{bmatrix} \begin{bmatrix} 1 & 4 \\ 6 & 3 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 4 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 2 & 2 \\ 4 & 5 \end{bmatrix} = \begin{bmatrix} 52 & 35 \end{bmatrix}$$

se tiene que su producto es

$$\begin{aligned} AB &= \begin{bmatrix} A_{11}B_{11} + A_{12}B_{21} & A_{11}B_{12} + A_{12}B_{22} \\ A_{21}B_{11} + A_{22}B_{21} & A_{21}B_{12} + A_{22}B_{22} \end{bmatrix} \\ &= \begin{bmatrix} 2I_2 & 2B_{12} + A_{12}B_{22} \\ A_{21} & A_{21}B_{12} + A_{22}B_{22} \end{bmatrix} \\ &= \begin{bmatrix} 2 & 0 & 15 & 22 \\ 0 & 2 & 25 & 20 \\ -1 & 6 & 52 & 35 \end{bmatrix}, \end{aligned}$$

que, evidentemente, coincide con el resultado obtenido según la definición de producto. ▷

Una aplicación muy importante: A continuación se dan dos casos particulares extremadamente útiles del producto de matrices por bloques. En temas posteriores, simplificarán sobremanera los cálculos y es importante habituarse lo antes posible a pensar/reescribir los productos de estas formas que involucran matrices filas y/o columnas.

El primer resultado técnico permite escribir una combinación lineal[6] de matrices columnas de $\mathbb{K}^{m\times 1}$ como un producto de una matriz por una matriz columna. Además, en el Ejercicio (32) de la página 148 se pide probar que el producto de una matriz fila por una matriz es combinación lineal de las filas de dicha matriz.

Proposición 2.8 (Producto especial)**.** Sea $A \in \mathbb{K}^{m\times n}$ una matriz particionada según sus columnas y sea $x \in \mathbb{K}^{n\times 1}$ de modo que

$$A = \left[\begin{array}{cccc} a_1 & a_2 & \dots & a_n \end{array}\right] \qquad \text{y} \qquad x = \left[\begin{array}{c} x_1 \\ x_2 \\ \vdots \\ x_n \end{array}\right].$$

Entonces

$$Ax = x_1 a_1 + x_2 a_2 + \cdots + x_n a_n,$$

es decir, el producto de una matriz A por una matriz columna x se escribe como combinación lineal de las columnas de A donde los escalares de dicha combinación lineal son las componentes $x_1, x_2, \dots, x_n$ de la matriz columna x.

Demostración. Se propone como ejercicio. □

[6]El concepto de combinación lineal se estudia en detalle en el capítulo dedicado a Espacios Vectoriales. En este caso, significa que se realizarán productos de escalares por matrices columnas (o filas) y se sumarán.

El segundo resultado de este tipo muestra cómo realizar el producto de dos matrices considerando a una de ellas particionada según sus columnas o sus filas.

Observación 2.9. **Procesamiento en paralelo**: En la actualidad los ordenadores utilizan esta propiedad para realizar el cálculo efectivo del producto de dos matrices, lo cual depende de la forma en que el computador almacene los datos en su memoria. Para ello, los algoritmos de alto rendimiento aplican la propiedad (a) de la Proposición 2.9 mediante la cual particionan la matriz B por columnas, se distribuyen la matriz A y cada una de las columnas b_i (o grupos de ellas) a diferentes procesadores, cada procesador en paralelo realiza un cálculo más pequeño, y se devuelve la información al procesador encargado de generar el resultado.

Proposición 2.9 (Dos productos especiales)**.** Sean $A \in \mathbb{K}^{m\times n}$ y $B \in \mathbb{K}^{n\times p}$. Se cumple que:

(a) Si $B \in \mathbb{K}^{f_1\times(c_1+c_2+\cdots+c_p)}$ con $f_1 = n$, $c_j = 1$, $j = 1, \ldots, p$ entonces

$$A\begin{bmatrix} b_1 & b_2 & \ldots & b_p \end{bmatrix} = \begin{bmatrix} Ab_1 & Ab_2 & \ldots & Ab_p \end{bmatrix}.$$

(b) Si $A \in \mathbb{K}^{(f_1+f_2+\cdots+f_m)\times c_1}$ con $f_i = 1$, $i = 1, \ldots, m$, $c_1 = n$ entonces

$$\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix} B = \begin{bmatrix} a_1B \\ a_2B \\ \vdots \\ a_mB \end{bmatrix}.$$

Demostración. Se probará (a) pues (b) es similar y se propone como ejercicio.

(a) Del enunciado se observa que la primera columna de B se denota por b_1, la segunda por b_2, y así hasta la última columna denotada por b_p.

Si bien no es necesario, para fijar ideas se analizará en primer lugar el caso correspondiente a la primera columna de AB y, luego se abordará el caso general. En efecto, cada elemento c_{i1} del producto AB, para $i = 1, 2, \ldots, m$, se obtiene de multiplicar la fila i-ésima de A por la primera columna de B. Con esto, la primera columna de $AB = C = [c_{ij}]$ queda

$$\begin{aligned}
\begin{bmatrix} c_{11} \\ c_{21} \\ \vdots \\ c_{m1} \end{bmatrix} &= \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} + \cdots + a_{1n}b_{n1} \\ a_{21}b_{11} + a_{22}b_{21} + \cdots + a_{2n}b_{n1} \\ \vdots \\ a_{m1}b_{11} + a_{m2}b_{21} + \cdots + a_{mn}b_{n1} \end{bmatrix} \\
&= \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix} \begin{bmatrix} b_{11} \\ b_{21} \\ \vdots \\ b_{n1} \end{bmatrix} \\
&= Ab_1.
\end{aligned}$$

Un razonamiento similar es válido para las columnas restantes. Para cualquier $j = 1, 2, \ldots, p$, la columna j-ésima se puede expresar como

$$\begin{aligned}
\begin{bmatrix} c_{1j} \\ c_{2j} \\ \vdots \\ c_{mj} \end{bmatrix} &= \begin{bmatrix} a_{11}b_{1j} + a_{12}b_{2j} + \cdots + a_{1n}b_{nj} \\ a_{21}b_{1j} + a_{22}b_{2j} + \cdots + a_{2n}b_{nj} \\ \vdots \\ a_{m1}b_{1j} + a_{m2}b_{2j} + \cdots + a_{mn}b_{nj} \end{bmatrix} \\
&= \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix} \begin{bmatrix} b_{1j} \\ b_{2j} \\ \vdots \\ b_{nj} \end{bmatrix} \\
&= Ab_j.
\end{aligned}$$

En definitiva, reuniendo toda la información se tiene

$$AB = \begin{bmatrix} Ab_1 & Ab_2 & \ldots & Ab_p \end{bmatrix}.$$

$\square$

Ejemplo 2.18. *Calcular AB utilizando los productos especiales de las dos proposiciones anteriores para las siguientes matrices particionadas según sus columnas:*

$$A = \begin{bmatrix} 2 & 3 \\ 6 & 1 \end{bmatrix} \qquad y \qquad B = \begin{bmatrix} 5 & 4 \\ 0 & 7 \end{bmatrix}.$$

$\lhd$ Se requiere calcular AB. Se hará primero en general y luego se particularizará. Para ello, se consideran las particiones siguientes:

$$A = \begin{bmatrix} a_1 & a_2 \end{bmatrix} \in \mathbb{R}^{2\times(1+1)}, \quad B = \begin{bmatrix} b_1 & b_2 \end{bmatrix} \in \mathbb{R}^{2\times(1+1)},$$

$$b_1 = \begin{bmatrix} b_{11} \\ b_{21} \end{bmatrix} \in \mathbb{R}^{(1+1)\times 1} \quad \text{y} \quad b_2 = \begin{bmatrix} b_{12} \\ b_{22} \end{bmatrix} \in \mathbb{R}^{(1+1)\times 1}.$$

Luego,

$$\begin{aligned} AB &= A\begin{bmatrix} b_1 & b_2 \end{bmatrix} \\ &= \begin{bmatrix} Ab_1 & Ab_2 \end{bmatrix} \\ &= \begin{bmatrix} \begin{bmatrix} a_1 & a_2 \end{bmatrix}\begin{bmatrix} b_{11} \\ b_{21} \end{bmatrix} & \begin{bmatrix} a_1 & a_2 \end{bmatrix}\begin{bmatrix} b_{12} \\ b_{22} \end{bmatrix} \end{bmatrix} \\ &= \begin{bmatrix} b_{11}a_1 + b_{21}a_2 & b_{12}a_1 + b_{22}a_2 \end{bmatrix}, \end{aligned}$$

donde cada columna del producto AB se expresa como combinación lineal de las columnas de A y los escalares de dichas combinaciones lineales se toman de las columnas respectivas de B. En particular,

$$AB = \begin{bmatrix} 5\begin{bmatrix} 2 \\ 6 \end{bmatrix} + 0\begin{bmatrix} 3 \\ 1 \end{bmatrix} & 4\begin{bmatrix} 2 \\ 6 \end{bmatrix} + 7\begin{bmatrix} 3 \\ 1 \end{bmatrix} \end{bmatrix} = \begin{bmatrix} 10 & 29 \\ 30 & 31 \end{bmatrix}.$$

$\rhd$

Observación 2.10. Si en el apartado (a) de la Proposición 2.9 una columna de B es nula, el producto AB también contiene esa misma columna nula. Del mismo modo, si en el apartado (b), una fila de A es nula, esa misma fila nula se replica en el producto AB.

Si bien se han señalado dos tipos de productos especiales, es recomendable manejar con soltura los diferentes productos por bloques que se presentan en los últimos ejercicios de este tema.

2.6.4. Trasposición de matrices particionadas

Se considera el siguiente ejemplo que motivará el procedimiento general. Se trata de calcular la traspuesta de la matriz particionada

$$A = \left[\begin{array}{r|rr|r} 2 & 0 & 1 & 3 \\ 4 & 5 & 6 & -1 \\ \hline -3 & 5 & 7 & -2 \\ 6 & 7 & 2 & -6 \\ -4 & 8 & 6 & 0 \end{array}\right]. \tag{2.5}$$

Según la Definición 2.8, la traspuesta de la matriz

$$A = \begin{bmatrix} 2 & 0 & 1 & 3 \\ 4 & 5 & 6 & -1 \\ -3 & 5 & 7 & -2 \\ 6 & 7 & 2 & -6 \\ -4 & 8 & 6 & 0 \end{bmatrix} \quad \text{es} \quad A^t = \begin{bmatrix} 2 & 4 & -3 & 6 & -4 \\ 0 & 5 & 5 & 7 & 8 \\ 1 & 6 & 7 & 2 & 6 \\ 3 & -1 & -2 & -6 & 0 \end{bmatrix}.$$

Dividiendo en bloques A^t de acuerdo a los bloques señalados en (2.5) para A queda

$$A^t = \left[\begin{array}{lr|rrr} 2 & 4 & -3 & 6 & -4 \\ \hline 0 & 5 & 5 & 7 & 8 \\ 1 & 6 & 7 & 2 & 6 \\ \hline 3 & -1 & -2 & -6 & 0 \end{array}\right].$$

Además de intercambiar los bloques de cada fila de A a la correspondiente columna de A^t se observa que, a su vez, es necesario trasponer cada uno de los bloques. Por ejemplo, el bloque $A_{12} = \begin{bmatrix} 0 & 1 \\ 5 & 6 \end{bmatrix}$ cambia de la posición $(1, 2)$ de A a la posición $(2, 1)$ de A^t como $A^t_{21} = (A_{12})^t = \begin{bmatrix} 0 & 5 \\ 1 & 6 \end{bmatrix}$, donde A^t_{21} es la notación del elemento de la posición $(2, 1)$ de A^t y $(A_{12})^t$ es la traspuesta de la matriz A_{12}.

En general, si se tiene la matriz particionada

$$A = \begin{bmatrix} A_{11} & A_{12} & \dots & A_{1s} \\ A_{21} & A_{22} & \dots & A_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ A_{r1} & A_{r2} & \dots & A_{rs} \end{bmatrix} \in \mathbb{K}^{(f_1+f_2+\cdots+f_r)\times(c_1+c_2+\cdots+c_s)},$$

la traspuesta de A puede calcularse por bloques mediante

$$A^t = \begin{bmatrix} A^t_{11} & A^t_{21} & \dots & A^t_{r1} \\ A^t_{12} & A^t_{22} & \dots & A^t_{r2} \\ \vdots & \vdots & \ddots & \vdots \\ A^t_{1s} & A^t_{2s} & \dots & A^t_{rs} \end{bmatrix} \in \mathbb{K}^{(c_1+c_2+\cdots+c_s)\times(f_1+f_2+\cdots+f_r)}.$$

2.7. EJERCICIOS

(1) Probar la propiedad conmutativa de la adición de matrices de la Proposición 2.1.

(2) Demostrar que el elemento neutro para la adición de matrices es único.

(3) Para cada matriz $A \in \mathbb{K}^{m\times n}$, probar que existe una única matriz $B \in \mathbb{K}^{m\times n}$ tal que $A + B = B + A = O_{m\times n}$.

(4) Probar las propiedades relativas a la multiplicación de un escalar por una matriz de la Proposición 2.2.

(5) Probar las propiedades relativas a la multiplicación de matrices de la Proposición 2.3.

(6) Dadas las matrices

$$A = \begin{bmatrix} 1 & 3 \\ -2 & 4 \end{bmatrix} \qquad \text{y} \qquad B = \begin{bmatrix} 3 & 6 \\ 9 & 0 \end{bmatrix},$$

hallar una matriz triangular inferior C con la misma diagonal principal que $-A$ y una matriz triangular superior D de modo que $A+D = B-C$. ¿Son únicas las matrices C y D?

(7) Demostrar que si $A, B \in \mathbb{R}^{n\times n}$ entonces

$$(A+B)^2 = A^2 + 2AB + B^2 \qquad \Leftrightarrow \qquad AB = BA.$$

Encontrar dos matrices A y B para las cuales $(A+B)^2 \neq A^2+2AB+B^2$.

(8) Encontrar todas las matrices a coeficientes reales que conmutan con

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

y expresar el resultado utilizando potencias de la matriz A.

(9) Probar la propiedad relativa a la potencia de otra potencia de matrices de la Proposición 2.4.

(10) Encontrar el resultado de la potencia n-ésima A^n de la matriz

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

y demostrar que es correcto por inducción sobre $n \in \mathbb{N}$. Idem para

$$B = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}.$$

(11) Probar las propiedades relativas a la trasposición de matrices de la Proposición 2.5.

(12) Demostrar las propiedades relativas a la simetría de la Proposición 2.6.

(13) Sea $\mathbb{K}$ un cuerpo de característica distinta de 2 ¿Existe en $\mathbb{K}^{n\times n}$ alguna matriz que sea simétrica y antisimétrica a la vez? Justificar la respuesta. En caso afirmativo, encontrar todas las matrices en esas condiciones. ¿Qué ocurre si $\mathbb{K}$ es un cuerpo de característica 2? En este caso encontrar una de dichas matrices para $\mathbb{K} = \mathbb{Z}_2$ y $n = 2$.

(14) Probar que toda matriz en $\mathbb{K}^{n\times n}$ se puede escribir como suma de una matriz simétrica y otra antisimétrica. Supóngase que $\mathbb{K}$ es un cuerpo de característica distinta de 2.

(15) Demostrar que las matrices escalares a coeficientes en $\mathbb{K}$ de tamaño $n \times n$ conmutan con todas las matrices de $\mathbb{K}^{n\times n}$ y que son las únicas en estas condiciones.

(16) El producto de matrices, en general, es no conmutativo. Dar un ejemplo que permita afirmar que $AB \neq BA$ para dos matrices $A, B \in \mathbb{R}^{n\times n}$ para $n \geq 2$ arbitrario.

(17) Dar un ejemplo que permita afirmar que $AB = AC$ no implica $B = C$ para matrices para las que los productos están bien definidos.

(18) Sean $A, B \in \mathbb{R}^{n\times n}$. Indicar si la siguiente afirmación es verdadera o falsa y justificar adecuadamente la respuesta:

(a) $AB = O_n \quad \Rightarrow \quad BA = O_n$.

(b) Si A y B son simétricas entonces $A - B$ es simétrica.

(19) Se consideran dos matrices particionadas $A \in \mathbb{K}^{(a_1+a_2)\times(a_3+a_4)}$ y $B \in \mathbb{K}^{(a_3+a_4)\times(b_1+b_2)}$. Comprobar que la multiplicación de A por B puede realizarse por bloques como se indica en (2.4).

(20) Calcular la matriz A^6 para

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

(a) Evaluando directamente.

(b) Dividiendo la matriz A en cuatro bloques de tamaño 2×2 cada uno.

¿Cuál de las dos formas es más sencilla?

(21) Sean $A, C \in \mathbb{K}^{r\times r}$ y $B, D \in \mathbb{K}^{s\times s}$. Demostrar que

$$\begin{bmatrix} A & O \\ O & B \end{bmatrix} \begin{bmatrix} C & O \\ O & D \end{bmatrix} = \begin{bmatrix} AC & O \\ O & BD \end{bmatrix}.$$

Es decir, que el producto de matrices diagonales por bloques es una matriz diagonal por bloques y se calcula multiplicando los bloques respectivos.

(22) Demostrar las propiedades restantes en la Proposición 2.7.

(23) Demostrar que la traza de una matriz $A \in \mathbb{K}^{n\times n}$ coincide con la de su traspuesta.

(24) Sean A y B matrices reales de tamaños $m \times n$ y $n \times m$, respectivamente, tales que $AB = I_m$ y $BA = I_n$. Demostrar que $m = n$.

(25) Demostrar que $\operatorname{tr}(AB) = \operatorname{tr}(BA)$ para cualquier par de matrices $A \in \mathbb{K}^{m\times n}$ y $B \in \mathbb{K}^{n\times m}$. Comparar con la Proposición 2.7.

(26) Demostrar que $\operatorname{tr}(ABC) = \operatorname{tr}(CAB) = \operatorname{tr}(BCA)$ donde A, B, C son matrices a coeficientes en $\mathbb{K}$ de tamaños apropiados (que se deben indicar). Comprobar que, en general, $\operatorname{tr}(ABC) \neq \operatorname{tr}(BAC)$.

(27) Demostrar que el producto de matrices triangulares superiores (respectivamente, inferiores) de $\mathbb{K}^{n\times n}$ es una matriz triangular superior (respectivamente, inferior). ¿Qué elementos aparecen en la diagonal del producto?

(28) Sea A la matriz

$$A = \begin{bmatrix} \cos(2\pi/k) & \operatorname{sen}(2\pi/k) \\ -\operatorname{sen}(2\pi/k) & \cos(2\pi/k) \end{bmatrix},$$

donde k es un entero positivo. Encontrar A^m para cada $m \in \mathbb{N}$. ¿Es posible encontrar una notación para la matriz A que sea más representativa atendiendo a su definición? ¿Cuánto da A^k?

(29) Sea $A \in \mathbb{R}^{n\times n}$ una matriz antisimétrica y $m \in \mathbb{N}$. Indicar si las siguientes afirmaciones son verdaderas o falsas y justificar adecuadamente:

(a) A^m es antisimétrica si m es impar.

(b) A^m es simétrica si m es par.

Enunciar y demostrar un resultado similar sobre la propiedad de simetría de las potencias naturales de una matriz simétrica $A \in \mathbb{R}^{n\times n}$.

(30) Hallar, si existen, todas las matrices $A \in \mathbb{R}^{3\times 3}$ tales que $AA^t = B$, donde

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

Resolverlo de dos formas distintas.

En lo que sigue, se propone una serie de ejercicios que permitirán realizar **productos especiales** *a partir de diferentes particiones de las matrices.*

(31) Demostrar la Proposición 2.8.

(32) En la Proposición 2.8 se ha mostrado que el producto de una matriz por una matriz columna es combinación lineal de las columnas de dicha matriz. Enunciar detalladamente y probar una propiadad dual: El producto de una matriz fila por una matriz es combinación lineal de las filas de dicha matriz.

(33) Sean $D \in \mathbb{K}^{n\times n}$ una matriz diagonal y $A \in \mathbb{K}^{m\times n}$ expresada mediante una partición del tipo $1 \times (c_1 + c_2 + \cdots + c_n)$ con $c_1 = c_2 = \cdots = c_n = 1$. Probar que si las columnas de A son $a_1, a_2, \ldots, a_n \in \mathbb{K}^{m\times 1}$ entonces el producto AD es

$$\begin{bmatrix} a_1 & a_2 & \ldots & a_n \end{bmatrix} \begin{bmatrix} d_1 & 0 & \ldots & 0 \\ 0 & d_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & d_n \end{bmatrix} = \begin{bmatrix} d_1a_1 & d_2a_2 & \ldots & d_na_n \end{bmatrix}.$$

Observar que el producto AD es una matriz del mismo tipo que el de A. Enunciar y probar un resultado similar para el producto DB para $B \in \mathbb{K}^{n\times p}$ particionada por filas.

(34) Sean $A = \begin{bmatrix} a^{(1)} \\ a^{(2)} \\ \vdots \\ a^{(m)} \end{bmatrix} \in \mathbb{K}^{(f_1+f_2+\cdots+f_m)\times n}$ con $f_1 = f_2 = \cdots = f_m = 1$ y

$B = \begin{bmatrix} b_1 & b_2 & \ldots & b_p \end{bmatrix} \in \mathbb{K}^{n\times(c_1+c_2+\cdots+c_p)}$ con $c_1 = c_2 = \cdots = c_p = 1$. Probar que el producto AB puede escribirse como una matriz de tipo $(f_1 + f_2 + \cdots + f_m) \times (c_1 + c_2 + \cdots + c_p)$, siendo $a^{(i)}b_j$ su elemento de la posición (i, j) para $i = 1, 2, \ldots, m$ y $j = 1, 2, \ldots, p$.

(35) Demostrar el apartado (b) de la Proposición 2.9.

(36) Sean $A = \begin{bmatrix} a_1 & a_2 & \ldots & a_n \end{bmatrix} \in \mathbb{K}^{m\times(c_1+c_2+\cdots+c_n)}$ con $c_1 = c_2 = \cdots = c_n = 1$ y $B = \begin{bmatrix} b^{(1)} \\ b^{(2)} \\ \vdots \\ b^{(n)} \end{bmatrix} \in \mathbb{K}^{(f_1+f_2+\cdots+f_n)\times p}$ con $f_1 = f_2 = \cdots = f_n = 1$. Probar que el producto AB puede escribirse como

$$AB = a_1b^{(1)} + a_2b^{(2)} + \cdots + a_nb^{(n)}.$$

(37) Sean $A \in \mathbb{K}^{m\times n}$, $B \in \mathbb{K}^{n\times p}$ y $C \in \mathbb{K}^{n\times q}$. Probar que

$$A \begin{bmatrix} B & C \end{bmatrix} = \begin{bmatrix} AB & AC \end{bmatrix}.$$

Indicando previamente cuáles deben ser ahora los tamaños de las matrices A, B y C, probar que

$$\begin{bmatrix} A \\ B \end{bmatrix} C = \begin{bmatrix} AC \\ BC \end{bmatrix}.$$

Para esta segunda prueba recurrir a la de trasposición de matrices por bloques.

Capítulo 3

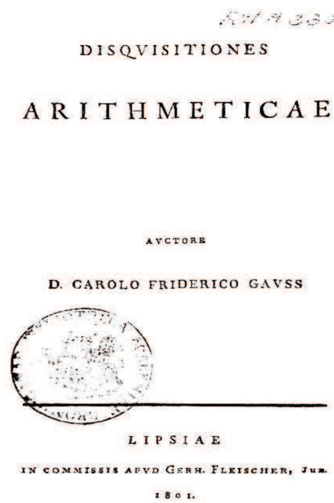
Sistemas de ecuaciones lineales

Índice

3.1. Introducción

En la obra china *Nueve capítulos sobre el Arte Matemático*, originaria de la Dinastía Zhou y utilizada durante los siglos II y I a.C, se incluyeron sistemáticamente todos los conocimientos matemáticos de la época. En su capítulo ocho aparecen algunos problemas donde se resuelven sistemas de ecuaciones lineales. Es conocido que esta obra fue consultada por el matemático, físico y astrónomo alemán Carl Friederich Gauss (1777-1855) al estudiar la órbita del asteroide Palas (el asteriode Ceres). El retrato de Gauss se puede apreciar en la Figura 3.1. A partir de estas observaciones, tomadas

(a) Carl Friedrich Gauss



DISQVISITIONES

ARITHMETICAE

AVCTORE

D. CAROLO FRIDERICO GAVSS

LIPSIAE

IN COMMISSIS APVD GERH. FLEISCHER, JUN.

1801.

(b) Disquisitiones Arithmeticae

Figura 3.1: Carl Friedrich Gauss y sus Disquisitiones Arithmeticae.

entre los años 1803 y 1809, Gauss obtuvo un sistema de seis ecuaciones lineales con seis incógnitas que resolvió mediante un método sistemático, conocido en la actualidad como *método de eliminación de Gauss*.

En este capítulo se presenta el método de eliminación de Gauss. Se prueba que dicho método funciona correctamente y que permite realizar una clasificación de los sistemas de ecuaciones lineales analizando los posibles tipos de soluciones.

Definición 3.1. Sea n un número natural. Una ecuación del tipo

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b$$

se llama **ecuación lineal**. Los escalares $a_1, a_2, \ldots, a_n$ son elementos conocidos de un cuerpo $\mathbb{K}$ y se llaman **coeficientes** y el escalar $b \in \mathbb{K}$, también conocido, se llama **término independiente**. Resolver la ecuación lineal significa encontrar todos los valores de los símbolos $x_1, x_2, \ldots, x_n$, llamados **incógnitas**, que satisfagan la ecuación. Más precisamente, una **solución** de la ecuación lineal es una n-tupla $(k_1, k_2, \ldots, k_n)$ con $k_i \in \mathbb{K}$ para $i \in \{1, 2, \ldots, n\}$, de modo que al sustituir x_i por k_i para cada $i \in \{1, 2, \ldots, n\}$, la ecuación se transforme en una igualdad, es decir, cumpla

$$a_1k_1 + a_2k_2 + \cdots + a_nk_n = b.$$

Ejemplo 3.1. *Las ecuaciones*

$$x_1 - 6x_2 = 0 \qquad y \qquad 2x_1 - 3x_2 + x_3 = \frac{1}{3}$$

son ecuaciones lineales con 2 y 3 incógnitas, respectivamente. La dupla $(6, 1)$ *es una solución de la primera y la tripla* $(\frac{1}{6}, 0, 0)$ *una solución de la segunda, independientemente de si se considera* $\mathbb{K}$ *igual a* $\mathbb{Q}, \mathbb{R}$ *ó* $\mathbb{C}$*. Otras soluciones de la primera ecuación son del tipo* $(6k, k)$*, para cualquier* $k \in \mathbb{K}$ *y otras soluciones de la segunda son* $(k_1, k_2, \frac{1}{3} - 2k_1 + 3k_2)$*, para cualquier par* $k_1, k_2 \in \mathbb{K}$*.*

Ejemplo 3.2. *Las ecuaciones*

$$e^x - 6x = 0, \qquad 2x_1^2 + 3x_2 - \cos(x_3) = \frac{1}{3}, \qquad \operatorname{sen}(x) = x$$

y

$$ax^2 + bx + c = 0, \qquad a, b, c \in \mathbb{K},\ a \neq 0$$

son ecuaciones no lineales (debido a la exponencial, al cuadrado (dos veces), al coseno y al seno que involucran), tres de ellas en 1 incógnita y la segunda en 3 incógnitas.

Para tratar ecuaciones no lineales como las del Ejemplo 3.2 serán necesarios métodos que escapan al alcance de este libro y son temas que suelen abordarse en Análisis Numérico.

3.2. Sistemas de ecuaciones lineales

Definición 3.2. Sean m y n dos números naturales. Al conjunto de m ecuaciones lineales con n incógnitas

$$S : \begin{cases} E_1 : & a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n & = & b_1 \\ E_2 : & a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n & = & b_2 \\ \vdots & & & \vdots \\ E_m : & a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n & = & b_m \end{cases} \tag{3.1}$$

se lo llama **sistema de m ecuaciones lineales con n incógnitas**. Con $E_1, E_2, \ldots, E_m$ se indican cada una de las ecuaciones del sistema. Los escalares a_{ij}, para números naturales i, j tales que $1 \leq i \leq m$, $1 \leq j \leq n$, son elementos conocidos de un cuerpo $\mathbb{K}$ y se llaman **coeficientes** del sistema y los escalares $b_i \in \mathbb{K}$, para números naturales i tales que $1 \leq i \leq m$, también conocidos, se llaman **términos independientes** del sistema. Resolver el sistema de ecuaciones lineales significa encontrar todos los valores de los símbolos $x_1, x_2, \ldots, x_n$, llamados **incógnitas**, que satisfagan las m ecuaciones lineales simultáneamente. Más precisamente, una **solución** del sistema de ecuaciones lineales es una n-tupla $(k_1, k_2, \ldots, k_n)$ con $k_i \in \mathbb{K}$, que sea solución de cada una de las m ecuaciones E_i, $i \in \{1, 2, \ldots, m\}$. Se llama **solución general** o **conjunto solución** del sistema S al conjunto de todas las soluciones del mismo. Si $b_1 = b_2 = \cdots = b_m = 0$, el sistema se llama **homogéneo** y si para algún $i_0 \in \{1, 2, \ldots, m\}$ se tiene $b_{i_0} \neq 0$, el sistema se llama **no homogéneo**.

Ejemplo 3.3. *Si $\mathbb{K} = \mathbb{R}$, las soluciones de los sistemas de ecuaciones lineales*

$$S_1 : \left\{ \begin{array}{rcrcr} 2x_1 & + & x_2 & = & 0 \\ x_1 & + & x_2 & = & 1 \end{array} \right. \qquad y \qquad S_2 : \left\{ \begin{array}{rcrcr} x_1 & + & 3x_2 & = & -2 \\ -2x_1 & - & 6x_2 & = & 4 \end{array} \right.$$

son iguales a $(-1, 2)$ y $(-2 - 3k, k)$ para $k \in \mathbb{R}$, respectivamente.

Ejemplo 3.4. *El sistema de ecuaciones lineales*

$$\left\{ \begin{array}{rcrcr} x_1 & - & \sqrt{2}x_2 & = & 0 \\ 2x_1 & - & 3x_2 & = & \frac{1}{3} \end{array} \right.$$

no tiene solución en $\mathbb{Q}$, aunque sí tiene solución en $\mathbb{R}$ y en $\mathbb{C}$. La única solución es $\left(-\frac{4}{3} - \sqrt{2}, -1 - \frac{2}{3}\sqrt{2}\right)$.

Ejemplo 3.5. *El conjunto solución del sistema de ecuaciones lineales*

$$\left\{ \begin{array}{rcrcr} x_1 & - & ix_2 & = & i \\ ix_1 & + & x_2 & = & -1 \end{array} \right.$$

en $\mathbb{C}$ es

$$\{(i(1 + k), k) : \, k \in \mathbb{C}\}.$$

La única solución tanto en $\mathbb{Q}$ como en $\mathbb{R}$ es $(0, -1)$.

Definición 3.3. Dos sistemas de ecuaciones lineales S_1 y S_2 con las mismas incógnitas $x_1, x_2, \ldots, x_n$ se llaman **sistemas equivalentes** si ambos tienen el mismo conjunto solución (eventualmente vacío), es decir, si ambos son conjuntos vacíos o, en caso contrario, si todas las soluciones de S_1 son soluciones de S_2 y todas las soluciones de S_2 lo son de S_1.

Ejemplo 3.6. *La solución del sistema de ecuaciones lineales*

$$S_3 : \begin{cases} 4x_1 & + & x_2 & = & -2 \\ x_1 & - & 3x_2 & = & -7 \end{cases}$$

es $(-1, 2)$. *Por tanto, los sistemas* S_1 *del Ejemplo 3.3 y* S_3 *son equivalentes, mientras que* S_2 *y* S_3 *no lo son. Debido a que el sistema*

$$S_4 : \begin{cases} x_1 & - & 3x_2 & = & 1 \\ x_1 & - & 3x_2 & = & -3 \end{cases}$$

no tiene solución, se tiene que S_3 *y* S_4 *no son sistemas equivalentes.*

3.3. Método de eliminación de Gauss

En esta sección se establece el método de eliminación de Gauss y se justifica que, efectivamente, funciona.

3.3.1. Operaciones elementales

Se consideran los siguientes tipos de operaciones entre ecuaciones que se llamarán **operaciones elementales**:

Tipo I (Intercambio): Intercambiar la ecuación i-ésima con la j-ésima (se denota $E_i \leftrightarrow E_j$).

Tipo II (Escalado): Multiplicar la ecuación E_i por un escalar no nulo $k \in \mathbb{K}$ (se denota $kE_i \to E_i$).

Tipo III (Eliminación): Sumar a la ecuación E_j la ecuación E_i ($i \neq j$) previamente multiplicada por un escalar $k \in \mathbb{K}$ (se denota $E_j + kE_i \to E_j$).

Observación 3.1. Las operaciones elementales son elegidas de este modo con la finalidad que, tras ser aplicadas a un sistema de ecuaciones lineales, preserven la linealidad del sistema, o sea, proporcionen ecuaciones del mismo tipo que las originales (es decir, ecuaciones lineales).

3.3.2. Justificación del método de eliminación de Gauss

A continuación se prueba que las operaciones elementales no alteran la equivalencia entre el sistema dado y el obtenido mediante su aplicación. Está claro entonces que las operaciones elementales se podrán utilizar para transformar el sistema en uno más sencillo de resolver.

Teorema 3.1 (Teorema de equivalencia de sistemas lineales por operaciones elementales)**.** Si a un sistema de ecuaciones lineales con coeficientes en un cuerpo $\mathbb{K}$ se le aplican operaciones elementales de tipo I, II o III entonces se obtiene un sistema de ecuaciones lineales equivalente.

Demostración. Se probará que la aplicación de operaciones elementales no modifica el conjunto solución del sistema. Es decir, se debe demostrar la invariancia del conjunto solución de un sistema de ecuaciones lineales, tras la aplicación de cada una de las operaciones elementales.

En primer lugar se supone que el conjunto solución de un sistema de ecuaciones lineales S es no vacío y se analiza cada operación elemental por separado.

- Tipo I: Es claro que si en el sistema de ecuaciones lineales S se intercambian dos ecuaciones entre sí, se obtiene otro sistema de ecuaciones lineales con las mismas soluciones.

Para los dos tipos restantes de operaciones elementales se considera el sistema de ecuaciones lineales

$$S : \begin{cases} E_1 : & e_1(x) & = & 0 \\ \vdots & & \vdots & \\ E_i : & e_i(x) & = & 0 \\ \vdots & & \vdots & \\ E_j : & e_j(x) & = & 0 \\ \vdots & & \vdots & \\ E_m : & e_m(x) & = & 0 \end{cases}, \tag{3.2}$$

donde, para una n-tupla $x = (x_1, x_2, \ldots, x_n)$, se ha definido

$$e_t(x) := a_{t1}x_1 + a_{t2}x_2 + \cdots + a_{tn}x_n - b_t, \quad \text{para cada } t \in \{1, 2, \ldots, m\},$$

con lo que se cumple que, si $k_i \in \mathbb{K}$, la n-tupla $k_0 := (k_1, k_2, \ldots, k_n)$ es solución del sistema lineal (3.2) si y sólo si $e_i(k_0) = 0$ para todo $i \in \{1, 2, \ldots, m\}$.

- Tipo II: Si al sistema (3.2) se le efectúa la operación elemental de tipo II dada por $kE_i \to E_i$, con $k \in \mathbb{K} - \{0\}$, se obtiene el sistema:

$$S' : \begin{cases} E'_1 : & e_1(x) & = & 0 \\ \vdots & & \vdots & \\ E'_i : & ke_i(x) & = & 0 \\ \vdots & & \vdots & \\ E'_j : & e_j(x) & = & 0 \\ \vdots & & \vdots & \\ E'_m : & e_m(x) & = & 0 \end{cases}, \tag{3.3}$$

  donde coinciden todas las ecuaciones excepto la i-ésima.

Es claro que si $k_0 := (k_1, k_2, \ldots, k_n)$ es solución del sistema (3.2), se tiene que $e_\ell(k_0) = 0$, para toda $\ell = 1, \ldots, i, \ldots, j, \ldots, m$. Entonces, también lo es del sistema (3.3) (pues $ke_i(k_0) = 0$ y las restantes ecuaciones son las mismas). Recíprocamente, si $k_0 := (k_1, k_2, \ldots, k_n)$ es solución del sistema (3.3), también lo es del sistema (3.2), puesto que al multiplicar la ecuación $ke_i(x) = 0$ por $k^{-1} \in \mathbb{K}$ (que existe por ser $k \neq 0$ y $\mathbb{K}$ un cuerpo) se obtiene la ecuación $e_i(x) = 0$. Luego, $ke_i(k_0) = 0$ implica $e_i(k_0) = 0$ (y las restantes ecuaciones son las mismas).

- Tipo III: Si al sistema (3.2) se le efectúa la operación elemental de tipo III dada por $E_j + kE_i \to E_j$, con $k \in \mathbb{K}$, se obtiene el sistema:

$$S' : \left\{ \begin{array}{lccc} E_1' : & e_1(x) & = & 0 \\ \vdots & & \vdots & \\ E_i' : & e_i(x) & = & 0 \\ \vdots & & \vdots & \\ E_j' : & e_j(x) + ke_i(x) & = & 0 \\ \vdots & & \vdots & \\ E_m' : & e_m(x) & = & 0 \end{array} \right. . \tag{3.4}$$

Es claro que si $k_0 := (k_1, k_2, \ldots, k_n)$ es solución del sistema (3.2) entonces $e_i(k_0) = 0$ y $e_j(k_0) = 0$, y por tanto $e_j(k_0) + ke_i(k_0) = 0$, con lo que también $k_0 := (k_1, k_2, \ldots, k_n)$ es solución del sistema (3.4). Recíprocamente, si $k_0 := (k_1, k_2, \ldots, k_n)$ es solución del sistema (3.4), se tiene que $e_i(k_0) = 0$ y $e_j(k_0) + ke_i(k_0) = 0$, de donde se deduce que

$$e_j(k_0) = (e_j(k_0) + ke_i(k_0)) - ke_i(k_0) = 0 - k0 = 0,$$

es decir $k_0 := (k_1, k_2, \ldots, k_n)$ es solución del sistema (3.2), pues las restantes ecuaciones coinciden en ambos sistemas.

Para los tres tipos de operaciones elementales se ha probado que las ecuaciones de un sistema dado y de un sistema obtenido tras aplicar cualesquiera de ellas proporciona las mismas soluciones, por tanto, cuando el conjunto solución del sistema de partida es no vacío, ambos sistemas son equivalentes.

En segundo lugar, se debe analizar el caso en que el conjunto solución del sistema de partida es vacío. De nuevo, se analiza cada tipo de operación por separado.

- Tipo I: Es claro que si en un sistema de ecuaciones lineales se intercambian dos ecuaciones entre sí, si uno de ellos no tiene solución, el otro tampoco, pues de tenerla el transformado, el original también la tendría.

- Tipo II: Si el sistema de partida no tiene solución y se multiplica una ecuación por $k \neq 0$, el sistema obtenido no puede tener solución, pues en caso de que la tuviese, multiplicando esa misma ecuación por k^{-1} se volvería al sistema de partida, el cual tendría solución, en contra de lo supuesto.

- Tipo III: Si el sistema de partida no tiene solución y a una ecuación se suma otra previamente multiplicada por un escalar, el nuevo sistema no puede tener solución. De tenerla, de la ecuación modificada se podría restar la misma ecuación utilizada para modificarla, previamente multiplicada por el mismo escalar; esto llevaría a que el conjunto solución del sistema de partida fuese no vacío, lo cual sería una contradicción.

$\square$

Corolario 3.1 (Las operaciones elementales son reversibles)**.** Cada operación elemental aplicada a un sistema de ecuaciones lineales es reversible. Esto es, si se pasa de un sistema S a uno S' aplicando una operación elemental de tipo I, tipo II o tipo III entonces se puede pasar también de S' a S mediante una operación elemental del mismo tipo obteniendo un sistema equivalente.

Demostración. Es una consecuencia de la demostración del Teorema 3.1. Más específicamente, si se pasa de S a S':

- intercambiando la ecuación i-ésima con la j-ésima,
- multiplicando la ecuación i-ésima por $k \neq 0$,
- sumando a la ecuación j-ésima la ecuación i-ésima previamente multiplicada por k,

se puede pasar de S' a S:

- intercambiando la ecuación j-ésima con la i-ésima,
- multiplicando la ecuación i-ésima por k^{-1},
- sumando a la ecuación j-ésima la ecuación i-ésima previamente multiplicada por $-k$,

respectivamente. Esto completa la prueba. □

3.3.3. Descripción del método

El **método de eliminación Gauss** consiste en la aplicación de operaciones elementales, convenientemente elegidas, a un sistema de ecuaciones lineales.

Se comienza por la primera incógnita, escogida de las ecuaciones donde posea un coeficiente no nulo e intercambiada con la primera ecuación si fuese necesario. Este coeficiente no nulo se llama **pivote**.

Una vez fijada esa incógnita (es decir, elegido el pivote), se eliminan los coeficientes en las ecuaciones que acompañan a la misma incógnita, por debajo de ella. A partir de aquí quedará fijada esta primera ecuación.

Ahora se procede de la misma forma con la segunda incógnita. Es decir, desde la segunda ecuación hasta la última, se elige un pivote no nulo correspondiente a la segunda incógnita (intercambiando ecuaciones si fuese necesario) y se eliminan los coeficientes que acompañan a la segunda incógnita, por debajo de ella.

El proceso continúa de manera similar, es decir, en cada ecuación obtenida con coeficientes no todos nulos, la estrategia consiste en elegir un pivote, y utilizarlo para eliminar todos los coeficientes que están por debajo de él (y, por tanto, correspondientes a la misma incógnita).

Si en algún paso se obtiene una ecuación con todos los coeficientes nulos, se la debe intercambiar con una no nula para poder continuar (manteniendo todas las nulas en la parte inferior del sistema). Si en algún paso se obtiene una ecuación del tipo $0x_1 + 0x_2 + \cdots + 0x_n = b$, con $b \neq 0$, se debe detener el proceso puesto que el sistema no tiene solución.

De este modo, el Teorema 3.1 establece que los sistemas intermedios obtenidos serán *todos* equivalentes al original y más sencillos de resolver.

Se repite el proceso un número finito de pasos y, si el sistema tiene solución, se conseguirá un sistema lineal cuyas soluciones se obtengan fácilmente mediante **sustitución regresiva**. Esto es, de la última ecuación con no todos los coeficientes nulos se despeja la incógnita que se encuentra a la izquierda de todas y se la sustituye en la ecuación anterior. Así se prosigue hasta sustituir en la primera ecuación las incógnitas encontradas en los pasos anteriores. De este modo se obtendrá una única solución o bien más de una solución.

En los siguientes ejemplos se muestra el procedimiento a seguir.

Ejemplo 3.7. *Mediante el método de eliminación de Gauss resolver el sistema de ecuaciones lineales (considerando sus coeficientes en $\mathbb{R}$)*

$$\left\{\begin{array}{rcrcrcrcr} 2x_1 & - & 3x_2 & - & x_3 & + & 3x_4 & = & 5 \\ x_1 & - & 3x_2 & + & x_3 & & & = & 1 \\ -x_1 & - & x_2 & + & x_3 & - & 2x_4 & = & 1 \\ & & & & 3x_3 & + & x_4 & = & -5 \end{array}\right. .$$

$\lhd$ En primer lugar se observa que el primer coeficiente no nulo de la primera ecuación es $2 \neq 1$. Conviene intercambiar las ecuaciones 1 y 2:

$$E_1 \leftrightarrow E_2 \quad \left\{ \begin{array}{rcrcrcrcr} x_1 & - & 3x_2 & + & x_3 & & & = & 1 \\ 2x_1 & - & 3x_2 & - & x_3 & + & 3x_4 & = & 5 \\ -x_1 & - & x_2 & + & x_3 & - & 2x_4 & = & 1 \\ & & & & 3x_3 & + & x_4 & = & -5 \end{array} \right. .$$

Ahora, el coeficiente 1 que acompaña a x_1 de la ecuación E_1 permite eliminar los coeficientes correspondienes a la incógnita x_1 que se encuentran por debajo de él,

$$\begin{array}{c} E_2 + (-2)E_1 \to E_2 \\ E_3 + E_1 \to E_3 \end{array} \quad \left\{ \begin{array}{rcrcrcrcr} x_1 & - & 3x_2 & + & x_3 & & & = & 1 \\ & & 3x_2 & - & 3x_3 & + & 3x_4 & = & 3 \\ & - & 4x_2 & + & 2x_3 & - & 2x_4 & = & 2 \\ & & & & 3x_3 & + & x_4 & = & -5 \end{array} \right. .$$

La primera ecuación quedará fija de ahora en adelante. Como en el sistema lineal obtenido todos los términos de la segunda ecuación se encuentran multiplicados por 3, se aplica una operación elemental de tipo II:

$$\frac{1}{3}E_2 \to E_2 \quad \left\{ \begin{array}{rcrcrcrcr} x_1 & - & 3x_2 & + & x_3 & & & = & 1 \\ & & x_2 & - & x_3 & + & x_4 & = & 1 \\ & - & 4x_2 & + & 2x_3 & - & 2x_4 & = & 2 \\ & & & & 3x_3 & + & x_4 & = & -5 \end{array} \right. .$$

De ahora en más, la segunda ecuación también queda fija. Se elimina el término $-4x_2$ de la tercera ecuación:

$$E_3 + 4E_2 \to E_3 \quad \left\{ \begin{array}{rcrcrcrcr} x_1 & - & 3x_2 & + & x_3 & & & = & 1 \\ & & x_2 & - & x_3 & + & x_4 & = & 1 \\ & & & - & 2x_3 & + & 2x_4 & = & 6 \\ & & & & 3x_3 & + & x_4 & = & -5 \end{array} \right. .$$

Se simplifica la tercera ecuación a partir de una operación elemental de tipo

II:

$$-\frac{1}{2}E_3 \to E_3 \quad \left\{ \begin{array}{rcrcrcrcr} x_1 & - & 3x_2 & + & x_3 & & & = & 1 \\ & & x_2 & - & x_3 & + & x_4 & = & 1 \\ & & & & x_3 & - & x_4 & = & -3 \\ & & & & 3x_3 & + & x_4 & = & -5 \end{array} \right. .$$

Finalmente, se elimina el término $3x_3$ de la cuarta ecuación:

$$E_4 + (-3)E_3 \to E_4 \quad \left\{ \begin{array}{rcrcrcrcr} x_1 & - & 3x_2 & + & x_3 & & & = & 1 \\ & & x_2 & - & x_3 & + & x_4 & = & 1 \\ & & & & x_3 & - & x_4 & = & -3 \\ & & & & & & 4x_4 & = & 4 \end{array} \right. .$$

Ahora la solución se obtiene mediante sustitución regresiva. Es decir, se despeja x_4 de la última ecuación: $x_4 = 1$. Sustituyendo en la anterior y despejando x_3 se encuentra que $x_3 = -2$. Sustituyendo ambos valores en la segunda ecuación y despejando x_2 se tiene $x_2 = -2$. Finalmente, de la primera ecuación se llega a $x_1 = -3$.

La solución general del sistema es el conjunto $\{(-3, -2, -2, 1)\}$ formado por un solo elemento. $\triangleright$

Es necesario remarcar que, al pasar de un sistema al siguiente, las ecuaciones deberían ser renombradas (por ejemplo, con $E'_1, E'_2, \ldots, E'_m$) puesto que algunas de ellas cambian. Por sencillez, se realiza un abuso en la notación y, se las denota igual que a las anteriores con $E_1, E_2, \ldots, E_m$.

Ejemplo 3.8. *Mediante el método de eliminación de Gauss resolver, considerando sus coeficientes en* $\mathbb{R}$*, el sistema de ecuaciones lineales*

$$\left\{ \begin{array}{rcrcrcrcr} 2x_1 & - & 3x_2 & - & x_3 & + & 3x_4 & = & 5 \\ x_1 & - & 3x_2 & + & x_3 & & & = & 1 \\ -x_1 & - & x_2 & + & x_3 & - & 2x_4 & = & 1 \end{array} \right. .$$

$\triangleleft$ Se trata del mismo sistema que en el ejemplo anterior en el que se ha quitado la última ecuación. Por lo tanto, siguiendo los mismos pasos se obtiene el sistema equivalente

$$\left\{\begin{array}{rcrcrcrcr} x_1 & - & 3x_2 & + & x_3 & & & = & 1 \\ & & x_2 & - & x_3 & + & x_4 & = & 1 \\ & & & & x_3 & - & x_4 & = & -3 \end{array}\right. .$$

Se observa que, en este caso, hay más incógnitas que ecuaciones. Asignando el valor $x_4 = k$, para $k \in \mathbb{R}$ arbitrario, y procediendo por sustitución regresiva como en el ejemplo anterior, se tiene que el conjunto solución es $\{(-2-k, -2, -3+k, k) : k \in \mathbb{R}\}$, que contiene infinitos elementos. La solución general se puede expresar en notación matricial como

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} -2-k \\ -2 \\ -3+k \\ k \end{bmatrix} = \begin{bmatrix} -2 \\ -2 \\ -3 \\ 0 \end{bmatrix} + k \begin{bmatrix} -1 \\ 0 \\ 1 \\ 1 \end{bmatrix}, \quad \forall k \in \mathbb{R}.$$

$\triangleright$

Ejemplo 3.9. *Mediante el método de eliminación de Gauss, resolver el sistema de ecuaciones lineales con coeficientes en $\mathbb{R}$:*

$$\left\{\begin{array}{rcrcrcrcr} 2x_1 & - & 3x_2 & - & x_3 & + & 3x_4 & = & 5 \\ x_1 & - & 3x_2 & + & x_3 & & & = & 1 \\ -x_1 & - & x_2 & + & x_3 & - & 2x_4 & = & 1 \\ & & & & x_3 & - & x_4 & = & 1 \end{array}\right. .$$

$\triangleleft$ Se trata del sistema del Ejemplo 3.7, donde se han mantenido las mismas tres primeras ecuaciones y la última ha sido modificada. Por lo tanto, siguiendo los mismos pasos para relizar las eliminaciones en las tres primeras ecuaciones se obtiene el sistema equivalente

$$\left\{\begin{array}{rcrcrcrcr} x_1 & - & 3x_2 & + & x_3 & & & = & 1 \\ & & x_2 & - & x_3 & + & x_4 & = & 1 \\ & & & & x_3 & - & x_4 & = & -3 \\ & & & & x_3 & - & x_4 & = & 1 \end{array}\right.$$

y realizando la operación elemental $E_4 + (-1)E_3 \to E_4$ se llega a

$$\left\{\begin{array}{rcrcrcrcr} x_1 & - & 3x_2 & + & x_3 & & & = & 1 \\ & & x_2 & - & x_3 & + & x_4 & = & 1 \\ & & & & x_3 & - & x_4 & = & -3 \\ & & & & & & 0\,x_4 & = & 4 \end{array}\right. . \tag{3.5}$$

Se observa ahora que la última ecuación no se cumple para ningún valor de $x_4 \in \mathbb{R}$. En este caso, el conjunto solución es vacío. ▷

3.4. Clasificación de los sistemas lineales

En general, dado un sistema de ecuaciones lineales con m ecuaciones y n incógnitas del tipo (3.1), tras la aplicación del método de eliminación de Gauss, una vez realizadas un número finito de operaciones elementales, se obtiene:

- un **sistema triangular**[1] del tipo

$$S' : \left\{\begin{array}{lrcl} E_1' : & g_{11}x_1 + g_{12}x_2 + \cdots + g_{1r}x_r + \cdots + g_{1n}x_n & = & c_1 \\ E_2' : & g_{22}x_2 + \cdots + g_{2r}x_r + \cdots + g_{2n}x_n & = & c_2 \\ \vdots & & \vdots & \\ E_r' : & g_{rr}x_r + \cdots + g_{rn}x_n & = & c_r \end{array}\right. , \tag{3.6}$$

  con g_{11}, g_{22}, $\ldots$, g_{rr} elementos no nulos de $\mathbb{K}$ $(r \leq m)$ y donde las incógnitas pueden aparecer eventualmente reordenadas[2]. Ahora, pueden ocurrir dos situaciones:

[1] Por analogía con la definición de matriz triangular superior.

[2] Esta reordenación se debe realizar si alguna de las incógnitas de la “diagonal” del sistema se anulase. Al reordenarlas, el sistema mantendrá sus coeficientes diagonales no nulos. Si no se reordenasen, tendría aspecto de **sistema escalonado**, que se definirá con detalle en el Capítulo 4.

- Si $r = n$ entonces la última ecuación de S' es $g_{nn}x_n = c_n$, que tiene solución única en $\mathbb{K}$ (pues $g_{nn} \neq 0$). Sustituyendo el valor hallado de x_n en la ecuación anterior E'_{r-1}, se obtiene un único valor para x_{n-1} y, continuando de este modo (por sustitución regresiva), se obtienen valores únicos para $x_{n-2}, \ldots, x_1$. En este caso, la solución es única y el sistema se llama **compatible determinado**.
- Si $r < n$ entonces asignando valores (fijos pero arbitrarios) a las incógnitas $x_{r+1}, \ldots, x_n$ en la ecuación E'_r, se obtiene un único valor para x_r. De la ecuación E'_{r-1} anterior, se obtiene un único valor para x_{r-1}, y se continúa así (por sustitución regresiva) hasta obtener un único valor para x_1. De este modo, debido a la arbitrariedad en los valores asignados, la solución no es única y, en este caso, el sistema se llama **compatible indeterminado**. Observar que hay exactamente $n - r$ incógnitas que se pueden asignar libremente con valores arbitrarios; a $x_{r+1}, \ldots, x_n$ se las suele llamar **incógnitas libres** y a $x_1, \ldots, x_r$ **incógnitas principales**.

- que alguna de sus ecuaciones es del tipo:

$$0x_1 + 0x_2 + \cdots + 0x_n = b.$$

En este caso, de nuevo, hay dos opciones:

  - Si $b \neq 0$ entonces su conjunto solución es vacío, dado que ninguna asignación realizada a las incógnitas $x_1, x_2, \ldots, x_n$ puede satisfacer la condición $0 = b$, con $b \neq 0$. En este caso, el sistema se denomina **incompatible**.
  - Si $b = 0$ entonces cualquier asignación dada a las incógnitas satisface esta ecuación (será una identidad, es decir, se cumple cualesquiera sean los valores asignados a $x_1, x_2, \ldots, x_n$), y dicha ecuación se puede **eliminar del sistema** puesto que no impondrá restricción alguna sobre el resto de ecuaciones[3].

[3] En el sistema (3.6) las ecuaciones $E'_{r+1}, \ldots, E'_m$ que faltan son de este tipo y se han

En la práctica, la expresión "discutir un sistema" suele referirse a que se lo clasifique en alguno de los tres casos anteriores y que se proporcione su conjunto solución.

En el siguiente esquema se resumen las diferentes situaciones a las que da lugar el análisis anterior, clasificando así los sistemas de ecuaciones lineales de acuerdo a la naturaleza de su conjunto solución.

$$\begin{array}{c}\underline{\text{Sistema de}}\\ \underline{\text{ecuaciones}}\\ \underline{\text{lineales}}\end{array} \begin{cases} \underline{\text{Compatible}}: \begin{cases} \underline{\text{Determinado}}: \text{Tiene solución única} \\ \\ \underline{\text{Indeterminado}}: \text{Tiene más de una solución} \end{cases} \\ \\ \underline{\text{Incompatible}}: \text{No tiene solución} \end{cases}$$

Cuando el sistema es compatible indeterminado no es posible afirmar que, en general, tenga infinitas soluciones puesto que depende del cuerpo (finito o infinito) en el que se esté resolviendo el sistema.

eliminado por este mismo motivo.

3.5. EJERCICIOS

(1) Plantear y resolver (en $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ y $\mathbb{Z}_2$) tres sistemas de 2 ecuaciones lineales con 2 incógnitas, uno de cada tipo: compatible determinado, compatible indeterminado e incompatible.

(2) Resolver y clasificar los siguientes sistemas de ecuaciones lineales:

$$\left\{\begin{array}{rcrcrcrcr} 2x & + & 3y & + & 3z & + & 3t & = & 3 \\ -2x & + & 11y & + & 11z & + & 10t & = & 11 \\ -4x & + & 8y & + & 8z & + & 7t & = & 7 \end{array}\right.$$

y

$$\left\{\begin{array}{rcrcrcr} 2x & + & 4y & + & 2z & = & 4 \\ -x & + & 4y & + & 8z & = & 4 \\ -4x & + & 8y & + & 7z & = & 8 \end{array}\right. .$$

(3) Resolver y clasificar el siguiente sistema de ecuaciones lineales

$$\left\{\begin{array}{rcrcrcrcr} 2x_1 & + & 3x_2 & + & 3x_3 & + & 3x_4 & = & 3 \\ 2x_1 & + & 4x_2 & + & 4x_3 & + & \frac{7}{2}x_4 & = & \frac{7}{2} \\ 2x_1 & + & 3x_2 & + & 3x_3 & + & 4x_4 & = & 4 \end{array}\right. .$$

Observar que la incógnita x_3 no aparece en la última ecuación del sistema equivalente transformado. Reordenar las incógnitas en el sistema original y resolver de nuevo, de modo que el sistema obtenido tenga aspecto triangular (es decir, que los elementos g_{ii} de la expresión (3.6) sean todos no nulos).

(4) Encontrar los valores de a y b reales tales que los sistemas de ecuaciones lineales

$$\left\{\begin{array}{rcrcr} 2x & - & 3y & = & 8 \\ 5x & - & 2y & = & 9 \end{array}\right. \quad \text{y} \quad \left\{\begin{array}{rcrcr} ax & - & y & = & 6 \\ x & - & 2b\,y & = & 9 \end{array}\right.$$

sean equivalentes.

(5) Encontrar un sistema de ecuaciones equivalente a cada uno de los siguientes sistemas:

$$\left\{\begin{array}{rcrcrcl} x_1 & + & x_2 & + & x_3 & = & 2 \\ x_1 & + & 2x_2 & + & 3x_3 & = & 2 \\ x_1 & + & 3x_2 & + & 6x_3 & = & 1 \end{array}\right. ,$$

$$\left\{\begin{array}{rcrcl} x_1 & + & 2x_2 & = & -1 \\ -2x_1 & - & 4x_2 & = & 2 \end{array}\right. \qquad \text{y} \qquad \left\{\begin{array}{rcrcl} x_1 & - & 3x_2 & = & 0 \\ -2x_1 & + & 6x_2 & = & 1 \end{array}\right. .$$

En caso de sistemas compatibles, encontrar su conjunto solución. Si es compatible indeterminado, expresar su solución general en forma matricial.

(6) Dadas las matrices

$$A = \begin{bmatrix} 2 & -1 \\ 5 & -3 \end{bmatrix} \qquad \text{y} \qquad B = \begin{bmatrix} 1 & 5 \\ -2 & 3 \end{bmatrix},$$

resolver la ecuación matricial $AX = B$ planteando y resolviendo un sistema de ecuaciones lineales.

(7) Determinar los valores de a, b y c reales para que los puntos $(1, 4)$, $(-1, 6)$ y $(2, 9)$ pertenezcan a la parábola de ecuación $y = ax^2 + bx + c$. Pintar la gráfica de dicha parábola realizando previamente el completamiento de cuadrados de modo que quede del tipo $y = m(x - h)^2 + k$.

(8) Encontrar, si existen, todos los valores de θ, α, β y $\gamma \in \mathbb{R}$ que sean soluciones de los siguientes sistemas de ecuaciones:

(a)

$$\left\{\begin{array}{rcl} \operatorname{sen}\theta - 2\cos\theta & = & 2 \\ 2\operatorname{sen}\theta - 2\cos\theta & = & 2 \end{array}\right.$$

(b)

$$\left\{\begin{array}{rcrcl} 2\operatorname{sen}\theta & + & \sqrt{2}\,\tan\theta & = & 4 \\ 4\operatorname{sen}\theta & - & 3\sqrt{2}\tan\theta & = & 8 \end{array}\right.$$

(c)

$$\left\{\begin{array}{rcrcrcl} 2\operatorname{sen}\alpha & - & \cos\beta & + & 3\tan\gamma & = & 3 \\ 4\operatorname{sen}\alpha & + & 2\cos\beta & - & 2\tan\gamma & = & 2 \\ 6\operatorname{sen}\alpha & - & 3\cos\beta & + & \tan\gamma & = & 9 \end{array}\right. .$$

Indicar si los sistemas anteriores son lineales o no lineales.

(9) Sean a, b, c y d números reales fijos. Demostrar que los valores de λ para los cuales el sistema

$$\left\{\begin{array}{rcl} (a-\lambda)x + by & = & 0 \\ cx + (d-\lambda)y & = & 0 \end{array}\right.$$

tiene soluciones no triviales deben satisfacer la ecuación cuadrática (en la incógnita λ)

$$\lambda^2 - (a+d)\lambda + (ad - bc) = 0.$$

¿Son números reales los valores de λ obtenidos? ¿Y si $b = c$? Justificar.

(10) Aplicando el método de eliminación de Gauss, encontrar los valores de $b \in \mathbb{R}$ para que el sistema

$$\begin{bmatrix} 1 & b & -2 \\ -1 & b-2 & 2 \\ 2 & 2 & b-4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ -2 \\ 3 \end{bmatrix}$$

(a) tenga una única solución.

(b) tenga infinitas soluciones.

(c) no tenga solución.

Capítulo 4

Matrices y sistemas de ecuaciones lineales. Rango

Índice

4.1. Introducción

En este capítulo se estudia la relación entre las matrices y los sistemas de ecuaciones lineales, llegando al concepto de rango de una matriz.

El concepto de rango se atribuye a Ferdinand Georg Frobenius (1849-1917), quien en 1879 lo introdujo en otro campo (el de las formas bilineales, para indicar que el rango de una matriz es el mayor orden que pueden alcanzar los menores no nulos de dicha matriz). Sin embargo, este concepto había sido utilizado previamente por James J. Sylvester (1814-1897) en 1851.

Antes de abordar el concepto de rango se establece el método llamado de Gauss-Jordan. Se llama así en consideración del trabajo realizado por Wilhelm Jordan (1842-1899) en problemas de triangulación en Geodesia en 1888, donde lo utilizaba únicamente para sistemas simétricos. En este método se mejora el método de eliminación de Gauss: ahora la matriz de coeficientes reducida se buscará diagonal en lugar de triangular (si es posible, la identidad). Íntimanente relacionado con estos conceptos está el de forma escalonada reducida por filas (forma normal de Hermite por filas), que se introducirá en la Definición 4.3 y se estudiará en detalle en el Capítulo 6.

Un resultado clave será el Teorema de Rouché-Frobenius cuya autoría varía en función del país en que se lo presente. Permite decidir la compatibilidad o no de un sistema de ecuaciones lineales. Si bien lo enunció el matemático francés Eugène Rouché (1832-1910) en 1875, el matemático Georges Fontené (1848-1923) reclamó la autoría de la demostración y, en 1905, el matemático alemán Kronecker reconoció que ambos lo habían establecido previamente. En España e Italia, se lo conoce como Teorema de Rouché-Frobenius, en Francia, se le atribuye a Rouché-Fontené mientras que, en Rusia, como Teorema de Kronecker-Capelli en honor a Alfredo Capelli (1855-1910).

(a) Wilhelm Jordan

(b) Ferdinand Georg Frobenius

Figura 4.1: Matemáticos que realizaron grandes aportaciones a la Teoría de Matrices.

4.2. Matriz asociada a un sistema lineal

Dado que las operaciones elementales sobre las ecuaciones consisten en reordenarlas, multiplicarlas por escalares no nulos y/o sumarlas a otra previamente multiplicada por un escalar, para aplicar el método de eliminación de Gauss es suficiente considerar los coeficientes involucrados, sin necesidad de escribir las incógnitas de manera explícita, lo que permite realizar un trabajo más esquemático.

Representación matricial de un sistema lineal

Utilizando notación y operaciones matriciales, el sistema de m ecuaciones lineales con n incógnitas

$$S : \left\{ \begin{array}{lrcl} E_1 : & a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n & = & b_1 \\ E_2 : & a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n & = & b_2 \\ \vdots & & \vdots & \\ E_m : & a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n & = & b_m \end{array} \right. \tag{4.1}$$

puede escribirse de forma equivalente mediante la igualdad de matrices

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

Utilizando la notación

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}, \quad x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad \text{y} \quad b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix},$$

el sistema de ecuaciones lineales (4.1) se representa mediante su expresión matricial como

$$Ax = b,$$

que es claramente más compacta.

Por lo tanto, asociados al sistema de ecuaciones lineales (4.1), se tiene la matriz A, que se conoce como **matriz de coeficientes**, la matriz columna b, llamada **vector**[1] **de términos independientes**, y el conocido como **vector de incógnitas**, que es la matriz columna denotada por x.

Biyección entre sistemas lineales y sus matrices ampliadas

Dado que las operaciones elementales del método de eliminación de Gauss deben realizarse sobre cada una de las ecuaciones de (4.1), es interesante considerar solamente la matriz que contenga todos los coeficientes necesarios en las operaciones a realizar, incluyendo tanto los de la matriz de coeficientes como los términos independientes.

Es decir, dado un sistema S de m ecuaciones lineales con n incógnitas con coeficientes en $\mathbb{K}$, es posible definir una matriz asociada al mismo, que

[1] En el tema dedicado a Espacios Vectoriales se justificará el nombre de *vector*.

se llama **matriz ampliada** o **aumentada**, formada por los bloques determinados por la matriz de coeficientes $A \in \mathbb{K}^{m\times n}$ y el vector de términos independientes $b \in \mathbb{K}^{m\times 1}$ como sigue:

$$\left[\begin{array}{cc} A & b \end{array}\right] = \left[\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{array}\right] \in \mathbb{K}^{m\times(n+1)}. \tag{4.2}$$

Recíprocamente, dada una matriz en bloques $\left[\begin{array}{cc} A & b \end{array}\right] \in \mathbb{K}^{m\times(n+1)}$, es posible asociarle un sistema de m ecuaciones lineales con n incógnitas, que será el sistema S dado en (4.1), de forma que su matriz ampliada sea $\left[\begin{array}{cc} A & b \end{array}\right]$.

De esta manera, se establece una biyección entre el conjunto de sistemas lineales a coeficientes en $\mathbb{K}$ de m ecuaciones con n incógnitas y el conjunto de las matrices en bloques de $\mathbb{K}^{m\times(n+1)}$.

Este hecho permitirá trasladar información de los sistemas de ecuaciones lineales a las matrices y recíprocamente. Esto significa que los resultados se podrán expresar en términos de sistemas lineales o en términos de matrices.

Ejemplo 4.1. *Indicar la matriz ampliada asociada al sistema de ecuaciones lineales*

$$S : \left\{\begin{array}{lrcrcrcrcrcr} E_1 : & x_1 & + & 3x_2 & & & + & 2x_4 & + & 16x_5 & = & 0 \\ E_2 : & -2x_1 & - & 6x_2 & + & x_3 & & & - & 9x_5 & = & -6 \\ E_3 : & & & & & 2x_3 & + & x_4 & + & 11x_5 & = & -5 \\ E_4 : & x_1 & + & 3x_2 & - & x_3 & + & x_4 & + & 8x_5 & = & 3 \end{array}\right.$$

$\lhd$ La matriz ampliada asociada al sistema de ecuaciones lineales dado es

$$\left[\begin{array}{cc} A & b \end{array}\right] = \left[\begin{array}{rrrrr|r} 1 & 3 & 0 & 2 & 16 & 0 \\ -2 & -6 & 1 & 0 & -9 & -6 \\ 0 & 0 & 2 & 1 & 11 & -5 \\ 1 & 3 & -1 & 1 & 8 & 3 \end{array}\right].$$

$\rhd$

4.3. Método de Gauss-Jordan

El **método de Gauss-Jordan**, al igual que el método de eliminación de Gauss, también utiliza un número finito de operaciones elementales de tipo I (intercambio), tipo II (escalado) y/o tipo III (eliminación). Tras aplicar el método de eliminación de Gauss, una vez llegado al sistema "triangular" (o "escalonado") equivalente, en lugar de resolver el sistema mediante sustitución regresiva, se continúa simplificándolo de modo de conseguir un nuevo sistema equivalente en el que las incógnitas se puedan reconocer/determinar de forma evidente (con tan sólo observar/despejar en el sistema encontrado). Para ello, se transforman todos los pivotes (se hallan en la diagonal si el sistema ha quedado triangular) en 1 y se utilizan para eliminar los restantes elementos no nulos de la respectiva columna en la que se encuentra dicho pivote, obteniendo así, en dicha columna, todos ceros salvo un 1 que corresponde al pivote.

Se realizarán operaciones elementales sobre las filas de la matriz ampliada, asociada al sistema lineal dado que, en lenguaje matricial, son:

Tipo I (Intercambio): Intercambiar la fila i-ésima con la j-ésima (se denota $F_i \leftrightarrow F_j$).

Tipo II (Escalado): Multiplicar la fila F_i por un escalar no nulo $k \in \mathbb{K}$ (se denota $kF_i \to F_i$, $k \in \mathbb{K} - \{0\}$).

Tipo III (Eliminación): Sumar a la fila F_j la fila F_i ($i \neq j$) previamente multiplicada por un escalar $k \in \mathbb{K}$ (se denota $F_j + kF_i \to F_j$).

Se mostrará la aplicación del método con un ejemplo y, posteriormente, se formalizará.

Ejemplo 4.2. *Resolver el sistema del Ejemplo 4.1 por el método de Gauss-Jordan con coeficientes en* $\mathbb{R}$.

◁ Realizando operaciones elementales se tiene:

$$\left[\begin{array}{cc} A & b \end{array}\right] = \left[\begin{array}{rrrrr|r} 1 & 3 & 0 & 2 & 16 & 0 \\ -2 & -6 & 1 & 0 & -9 & -6 \\ 0 & 0 & 2 & 1 & 11 & -5 \\ 1 & 3 & -1 & 1 & 8 & 3 \end{array}\right] \rightarrow \left[\begin{array}{rrrrr|r} 1 & 3 & 0 & 2 & 16 & 0 \\ 0 & 0 & 1 & 4 & 23 & -6 \\ 0 & 0 & 2 & 1 & 11 & -5 \\ 0 & 0 & -1 & -1 & -8 & 3 \end{array}\right]$$

$$\rightarrow \left[\begin{array}{rrrrr|r} 1 & 3 & 0 & 2 & 16 & 0 \\ 0 & 0 & 1 & 4 & 23 & -6 \\ 0 & 0 & 0 & -7 & -35 & 7 \\ 0 & 0 & 0 & 3 & 15 & -3 \end{array}\right] \rightarrow \left[\begin{array}{rrrrr|r} 1 & 3 & 0 & 2 & 16 & 0 \\ 0 & 0 & 1 & 4 & 23 & -6 \\ 0 & 0 & 0 & 1 & 5 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

$$\rightarrow \left[\begin{array}{rrrrr|r} 1 & 3 & 0 & 0 & 6 & 2 \\ 0 & 0 & 1 & 0 & 3 & -2 \\ 0 & 0 & 0 & 1 & 5 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

Los tres primeros pasos (flechas) corresponden al método de eliminación de Gauss. Las operaciones que se han realizado son: $F_2 + 2F_1 \rightarrow F_2$, $F_4 + (-1)F_1 \rightarrow F_4$; luego, $F_3 + (-2)F_2 \rightarrow F_3$, $F_4 + F_2 \rightarrow F_4$ y, finalmente, $(-\frac{1}{7})F_3 \rightarrow F_3$, y por último, $F_4 + (-3)F_3 \rightarrow F_4$.

La novedad se encuentra en la obtención de la última matriz ampliada. Se utiliza el 1 de la posición $(3, 4)$ para eliminar el 4 de la posición $(2, 4)$ y el 2 de la posición $(1, 4)$. Las correspondientes operaciones elementales por filas son: $F_2 + (-4)F_3 \rightarrow F_2$ y $F_1 + (-2)F_3 \rightarrow F_1$. Por último, se han redondeado los primeros elementos no nulos de cada fila comenzando por la izquierda.

Ahora, el sistema equivalente asociado es

$$S' : \left\{ \begin{array}{lccccccccc} E_1' : & x_1 & + & 3x_2 & & & & + & 6x_5 & = & 2 \\ E_2' : & & & & & x_3 & & + & 3x_5 & = & -2 \\ E_3' : & & & & & & x_4 & + & 5x_5 & = & -1 \end{array} \right. .$$

Este sistema es compatible indeterminado y, asignando valores arbitrarios a las variables libres x_5 y x_2, su conjunto solución está dado por

$$\{(2 - 3r - 6k, r, -2 - 3k, -1 - 5k, k) : r, k \in \mathbb{R}\},$$

que tiene infinitos elementos. ▷

Observación 4.1. Como se verá más adelante, el método de eliminación de Gauss-Jordan se utilizará para conseguir cuatro objetivos diferentes:

(GJ1) Obtención de una matriz escalonada reducida (por filas) de una matriz dada.

(GJ2) Resolución de sistemas de ecuaciones lineales.

(GJ3) Cálculo del rango de una matriz.

(GJ4) Cálculo de la matriz inversa.

Es importante resaltar que en el método de Gauss-Jordan se han realizado únicamente operaciones sobre las filas (y no sobre las columnas) de la matriz de partida.

Se ha observado en la Sección 3.4 que si se llega a un sistema compatible (eventualmente reordenando las incógnitas) por el método de eliminación de Gauss, se consigue un sistema triangular. Sin embargo, la última matriz ampliada del ejemplo anterior, en que las incógnitas (determinadas por las correspondientes columnas) no han sido reordenadas, presenta una estructura ligeramente diferente a la de una triangular superior: una forma "escalonada". Se observa que:

- la fila nula aparece en la última posición de las filas,
- el primer elemento no nulo de cada fila no nula es un 1,
- el primer elemento no nulo de una fila (si no es nula y es posterior a la primera) se encuentra situado a la derecha del (primer) 1 de la fila anterior.

Estos unos, que en el Ejemplo 4.2 han sido encerrados en un círculo, juegan un papel importante pues son el único elemento no nulo de la columna a la que pertenecen. Por lo tanto, es inmediato reconocer el valor de la incógnita asociada (que es una incógnita principal), ya sea su valor numérico determinado, o bien en función de las incógnitas libres.

El concepto de forma escalonada, mencionado de forma intuitiva, se puede refinar obteniendo la llamada matriz escalonada reducida (por filas) de una matriz. Si bien en estos momentos su utilidad está enfocada a la resolución de sistemas lineales, será estudiada ampliamente en el Capítulo 6.

4.4. Matriz escalonada reducida por filas

Definición 4.1. Sea $R \in \mathbb{K}^{m \times n}$. Se dice que R es una matriz **escalonada reducida por filas** si se cumplen las siguientes condiciones:

(a) Toda fila nula de R debe aparecer debajo de todas las filas no nulas.

(b) El primer elemento no nulo de cada fila no nula es 1, y se llama **uno principal** y además, las columnas de R que contienen a esos 1 tienen los restantes elementos iguales a 0, y se llaman **columnas básicas**[a].

(c) Si las filas no nulas de R son $1, 2, \dots, r$ y si el primer elemento no nulo de la fila i-ésima aparece en la columna j_i-ésima entonces

$$j_1 < j_2 < \cdots < j_r.$$

[a]En el Capítulo 11, dedicado a los subespacios fundamentales de una matriz, se justificará el adjetivo “básicas”.

El sistema de ecuaciones lineales asociado a una matriz escalonada reducida por filas se denomina **sistema de ecuaciones escalonado reducido por filas**.

Observación 4.2. El apartado (a) de la Definición 4.1 indica que si aparecen filas nulas, deben estar agrupadas entre las últimas. El apartado (b) establece que si una fila es no nula, su primer elemento comenzando por la izquierda debe ser un 1 y la columna que contiene ese 1 tiene el siguiente aspecto $\begin{bmatrix} 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \end{bmatrix}^t$, es decir, tiene ceros por arriba y por debajo de dicho 1, lo que justifica el adjetivo de *reducida*. El apartado (c) establece los escalones de la *matriz escalonada*.

Ejemplo 4.3. *Las matriz identidad I_n y la matriz nula $O_{m\times n}$ son matrices escalonadas reducidas por filas.*

Ejemplo 4.4. *Las matrices reales A y B son escalonadas reducidas por filas:*

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

y las matrices C y D no lo son

$$C = \begin{bmatrix} 0 & 2 & 3 & 1 & 4 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}, \qquad D = \begin{bmatrix} 1 & 2 & 3 & 0 \\ 0 & 1 & 2 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

$\triangleleft$ La matriz C no es escalonada reducida por filas puesto que el primer elemento no nulo de la primera fila (no nula) es 2 (distinto de 1); además el primer elemento no nulo de la segunda fila es 1, pero en esa misma columna aparece otro elemento no nulo (otro 1). La matriz D no lo es pues tiene su segunda fila no nula, la primera entrada de dicha fila es un 1 en la columna 2, pero no tiene ceros en los restantes elementos de esa columna (tiene un 2 en la posición de arriba). ▷

4.4.1. Matrices equivalentes por filas

Se recuerda que al pasar de un sistema a otro mediante el método de eliminación de Gauss se obtiene un sistema equivalente. Un concepto similar se establecerá para matrices.

Definición 4.2. Sean $A, B \in \mathbb{K}^{m\times n}$. Se dice que A **es equivalente por filas** a B, y se denota mediante $A \sim_f B$, si B se obtiene de aplicar a A un número finito de operaciones elementales por filas.

Ejemplo 4.5. *En el Ejemplo 4.2 todas las matrices obtenidas a partir de la matriz aumentada inicial son equivalentes por filas a dicha matriz.*

El siguiente resultado entra dentro de la categoría de los teoremas constructivos. Es decir, no sólo se probará la existencia de una matriz escalonada reducidas por filas asociada a una matriz arbitraria, sino que la demostración da un método explícito (un algoritmo) que indica cómo obtener dicha matriz.

Teorema 4.1. Toda matriz $A \in \mathbb{K}^{m\times n}$ es equivalente por filas a una matriz escalonada reducida por filas.

Demostración. Si $A = O_{m\times n}$, dicha matriz es escalonada reducida por filas y el teorema se verifica.

Sea $A = [a_{ij}] \neq O_{m\times n}$. Suponer que la primera columna no nula de A es la j_1-ésima. Intercambiando la primera fila de A con alguna otra, si fuese necesario, se tiene que $a_{1j_1} \neq 0$. Por sencillez en la notación se llamará A a todas las matrices intermedias obtenidas mediante las transformaciones.

Utilizando el pivote a_{1j_1}, mediante operaciones elementales por filas, se convierte en 1 el elemento a_{1j_1} y se eliminan los restantes elementos de la columna j_1-ésima. Se obtiene una matriz de la forma

$$A = \begin{bmatrix} O & 1 & x \\ O & O & B \end{bmatrix},$$

donde x y B son matrices de tamaños adecuados y el bloque O de la posición $(1,1)$ de A es de tamaño $1 \times (j_1 - 1)$, el cual no aparece si $j_1 = 1$.

Si $B = O$ entonces la matriz hallada es escalonada reducida por filas. Si $B \neq O$, se selecciona la primera columna no nula de $\begin{bmatrix} O & O & B \end{bmatrix}$, y se supone que dicha columna es la j_2-ésima. Es claro que $j_1 < j_2$.

Se elige $a_{2j_2} \neq 0$, intercambiando la segunda fila con alguna fila inferior, si fuese necesario. Luego, se convierte en 1 el elemento a_{2j_2} y se eliminan los restantes elementos de la columna j_2-ésima de A (tanto los que se encuentran por debajo como por encima del 1).

Es importante observar que estas operaciones no alteran las primeras $j_2 - 1$ columnas de la matriz A (pues $a_{2j} = 0$ para todo $j < j_2$).

De este modo, se obtiene una matriz de la forma

$$A = \begin{bmatrix} O & 1 & y & 0 & z \\ O & 0 & O & 1 & u \\ O & O & O & O & C \end{bmatrix}.$$

Procediendo de forma similar, es posible reducir A a una matriz escalonada reducida por filas en un número finito de pasos. $\square$

4.4.2. Existencia y unicidad de la forma escalonada reducida por filas

Los pasos a realizar para obtener "una" matriz escalonada reducida por filas no son únicos, pues dependen de las operaciones elementales elegidas y del orden en que se apliquen las mismas. Sin embargo, la matriz equivalente por filas obtenida, que es escalonada reducida por filas, sí que es única, como se probará en el Teorema 4.2. Se podrá hablar, entonces, de *la* **forma escalonada reducida por filas** de una matriz.

Previamente a ello, se establecerán dos resultados auxiliares que serán necesarios en su demostración.

Es importante destacar que un sistema homogéneo del tipo $Ax = 0$ es siempre compatible, pues admite al menos la **solución trivial** $x = 0$.

Lema 4.1. Sean $A, B \in \mathbb{K}^{m\times n}$. Si A es equivalente por filas a B (es decir, $A \sim_f B$) entonces los sistemas lineales (homogéneos) $Ax = 0$ y $Bx = 0$ son equivalentes.

Demostración. Por hipótesis, $A \sim_f B$. Es evidente que

$$\begin{bmatrix} A & 0 \end{bmatrix} \sim_f \begin{bmatrix} B & 0 \end{bmatrix} \tag{4.3}$$

como matrices de $\mathbb{K}^{m\times(n+1)}$. Por la biyección existente entre matrices ampliadas y sistemas lineales, las matrices de (4.3) son las matrices ampliadas asociadas a los sistemas lineales homogéneos $Ax = 0$ y $Bx = 0$.

Si S_A y S_B denotan el conjunto solución de los sistemas $Ax = 0$ y $Bx = 0$, respectivamente, se debe probar que $S_A = S_B$. Por ser homogéneos, es claro que ambos conjuntos solución son no vacíos, es decir, se cumple que $S_A \neq \emptyset \neq S_B$.

($S_A \subseteq S_B$): Aplicando el Teorema 3.1, las soluciones de $Ax = 0$ también son soluciones de $Bx = 0$, pues B se obtiene mediante un número finito de operaciones elementales por filas a partir de A.

($S_B \subseteq S_A$): En el Corolario 3.1 se ha probado que las operaciones elementales por filas que permiten llevar un sistema en otro son reversibles. Este hecho implica que, por (4.3), las soluciones de $Bx = 0$ también son soluciones de $Ax = 0$.

Se ha demostrado entonces que $Ax = 0$ y $Bx = 0$ tienen el mismo conjunto solución. $\square$

Lema 4.2. Sean $A \in \mathbb{K}^{m\times n}$ y $b \in \mathbb{K}^{m\times 1}$. Si ciertas operaciones elementales por filas llevan la matriz por bloques $\begin{bmatrix} A & b \end{bmatrix}$ a una matriz escalonada reducida por filas entonces las mismas operaciones elementales también llevan A a una matriz escalonada reducida por filas.

Demostración. Es inmediato, pues basta aplicar a A las mismas operaciones elementales que se aplican a $\begin{bmatrix} A & b \end{bmatrix}$. □

Teorema 4.2 (Existencia y unicidad de la matriz escalonada reducida por filas)**.** Toda matriz $A \in \mathbb{K}^{m\times n}$ es equivalente por filas a una única matriz escalonada reducida por filas.

Demostración. De la definición, se deduce inmediatamente que la matriz escalonada reducida por filas equivalente por filas a la matriz nula es (únicamente) ella misma. Por lo tanto, se deberá probar el resultado para matrices no nulas. Esto es, es necesario demostrar que toda matriz no nula es equivalente por filas a una matriz escalonada reducida por filas (esto probará su existencia) y, además, que admite sólo una matriz escalonada reducida por filas equivalente por filas a ella (esto probará su unicidad). En efecto,

- EXISTENCIA: Se ha demostrado en el Teorema 4.1.

- UNICIDAD: La demostración se realizará por inducción sobre el número de columnas n (para lo que se supone que el número de filas m es fijo, aunque puede ser arbitrario).

  Si $n = 1$ entonces

$$A = \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix} \in \mathbb{K}^{m\times 1}$$

y, por ser $A \neq O$, existe $i_0 \in \{1, 2, \ldots, m\}$ tal que $a_{i_0 1} \neq 0$. Realizando operaciones elementales por filas sobre A, se puede convertir el elemento $a_{i_0 1}$ en 1, llevarlo a la posición $(1, 1)$ y usarlo como pivote para eliminar el resto de elementos no nulos (si los hay) de la única columna que determina la matriz A. Luego, la matriz escalonada reducida por filas equivalente por filas a A es

$$R = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

y, claramente, en este caso, es la única posible.

Ahora, sea $n > 1$. Se supondrá que el resultado es cierto para matrices con m filas y $n-1$ columnas y se probará que es válido para matrices con m filas y n columnas, todas con coeficientes en $\mathbb{K}$.

Para ello, se considera una matriz $A \in \mathbb{K}^{m \times n}$ particionada en dos bloques, separando la última columna como sigue:

$$A = \left[\begin{array}{cccc|c} a_{11} & a_{12} & \ldots & a_{1n-1} & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n-1} & a_{2n} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn-1} & a_{mn} \end{array}\right] = \begin{bmatrix} A' & a_n \end{bmatrix}, \tag{4.4}$$

siendo $A' \in \mathbb{K}^{m \times (n-1)}$ y $a_n \in \mathbb{K}^{m \times 1}$.

Por la existencia, hay al menos una matriz, escalonada reducida por filas, equivalente por filas a A. Si se supone que B y C son dos matrices escalonadas reducidas por filas, equivalentes por filas a A, para demostrar la unicidad requerida, se deberá probar que $B = C$.

Al ser

$$B = \begin{bmatrix} B' & b_n \end{bmatrix} \qquad \text{y} \qquad C = \begin{bmatrix} C' & c_n \end{bmatrix} \tag{4.5}$$

dos matrices escalonadas reducidas por filas, equivalentes por filas a A, por el Lema 4.2, los bloques $B' \in \mathbb{K}^{m \times (n-1)}$ y $C' \in \mathbb{K}^{m \times (n-1)}$ deben ser matrices escalonadas reducidas por filas, equivalentes por filas a A'. Puesto que A' es una matriz que posee m filas y $n-1$ columnas, la hipótesis de inducción asegura que $B' = C'$.

Supóngase, por reducción al absurdo, que $B \neq C$. Puesto que $B' = C'$, las matrices B y C deben diferir únicamente en la última columna[2], es decir $b_n \neq c_n$. Por lo tanto, existirá $j_0 \in \{1, 2, \ldots, m\}$ tal que $b_{j_0 n} \neq c_{j_0 n}$ siendo $b_n = [b_{jn}]$ y $c_n = [c_{jn}]$.

Si $u \in \mathbb{K}^{n\times 1}$ satisface el sistema homogéneo[3] $Au = 0$ entonces $Bu = 0$ y $Cu = 0$ (por tratarse de sistemas equivalentes, como se deduce del Lema 4.1). Restando, se tiene $(B - C)u = 0$, y en términos de sus bloques,

$$\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \left[\begin{array}{c|c} B' - C' & b_n - c_n \end{array}\right] u = \left[\begin{array}{c|c} O_{m,n-1} & b_n - c_n \end{array}\right] \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}.$$

Al multiplicar la j_0-ésima fila de la última matriz en bloques por u se tiene que $(b_{j_0 n} - c_{j_0 n})u_n = 0$, con lo que $u_n = 0$. Por lo tanto, cualquier solución de $Bu = 0$ y de $Cu = 0$ debe ser tal que $u_n = 0$, es decir, dicha solución será del tipo

$$u = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_{n-1} \\ 0 \end{bmatrix} \in \mathbb{K}^{n\times 1}. \tag{4.6}$$

Esto obliga a las n-ésimas columnas de B y de C a contener un uno principal. En efecto, si (por reducción al absurdo) esto no se diese, los unos principales estarían todos en $B'(= C')$, y en consecuencia, las filas

[2]Observar que b_n y c_n pueden no ser escalonadas reducidas.

[3]Recordar que los sistemas homogéneos son siempre compatibles.

m-ésimas tanto de la matriz (completa) B como de la matriz (completa) C sólo tendrían ceros o bien la fila m-ésima de $B'(= C')$ contendría un uno principal, por lo que los sistemas homogéneos $\begin{bmatrix} B' & b_n \end{bmatrix} u = 0$ y $\begin{bmatrix} C' & c_n \end{bmatrix} u = 0$ admitirían soluciones diferentes de (4.6) ya que u_n podría elegirse de forma arbitraria por ser la última una columna libre (no básica), lo que llevaría a una contradicción debido a que debe ser $u_n = 0$.

Como en las n-ésimas columnas de B y de C hay un uno principal, los restantes elementos de dichas columnas deben ser ceros (por ser B y C matrices escalonadas reducidas por filas).

Al ser $B' = C'$, la fila en la cual debe aparecer este pivote 1 debe ser la misma en b_n que en c_n (que coincide con la primera fila no nula de la matriz escalonada reducida $B'(= C')$, equivalente por filas a A'). Luego, $b_n = c_n$, lo que contradice $b_n \neq c_n$. Se ha probado entonces que $B = C$.

Por el principio de inducción queda probado el teorema para todo $n \in \mathbb{N}$.

Como m se ha considerado fijo, pero es arbitrario, se tiene probada la unicidad para toda $A \in \mathbb{K}^{m\times n}$, siendo m y n números naturales arbitrarios.

$\square$

Este teorema de existencia y unicidad permite establecer la siguiente definición.

Definición 4.3. Sea $A \in \mathbb{K}^{m\times n}$. Se llama **forma normal de Hermite por filas** o **forma escalonada reducida por filas** de A, y se denota R_A, a la única matriz escalonada reducida por filas que se obtiene de A aplicando operaciones elementales por filas.

Ejemplo 4.6. *Calcular la forma escalonada reducida por filas de la matriz*

$$A = \begin{bmatrix} -4 & 4 & -4 & 20 \\ 1 & 2 & -2 & 4 \\ 4 & 2 & -2 & -2 \end{bmatrix}.$$

◁ Realizando operaciones elementales por filas se tiene

$$A \sim_f \begin{bmatrix} 1 & 2 & -2 & 4 \\ -4 & 4 & -4 & 20 \\ 4 & 2 & -2 & -2 \end{bmatrix} \sim_f \begin{bmatrix} 1 & 2 & -2 & 4 \\ 1 & -1 & 1 & -5 \\ 2 & 1 & -1 & -1 \end{bmatrix} \sim_f$$

$$\sim_f \begin{bmatrix} 1 & 2 & -2 & 4 \\ 0 & -3 & 3 & -9 \\ 0 & -3 & 3 & -9 \end{bmatrix} \sim_f \begin{bmatrix} 1 & 2 & -2 & 4 \\ 0 & 1 & -1 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix} \sim_f \begin{bmatrix} 1 & 0 & 0 & -2 \\ 0 & 1 & -1 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

La última matriz es la forma escalonada reducida por filas R_A de A con 2 unos principales. ▷

El método de Gauss-Jordan se puede enunciar en el lenguaje de las matrices escalonadas reducidas por filas de la siguiente forma.

Teorema 4.3. Sea $Ax = b$ un sistema de ecuaciones lineales con $A \in \mathbb{K}^{m\times n}$, $b \in \mathbb{K}^{n\times 1}$ y sea $\mathcal{A} := \begin{bmatrix} A & b \end{bmatrix}$ su matriz ampliada. Con la notación

$$R_{\mathcal{A}} := \begin{bmatrix} C & d \end{bmatrix}, \qquad \text{donde } d \text{ es una matriz columna,}$$

se indica la partición en bloques de la forma escalonada reducida por filas de $\mathcal{A}$. Entonces el sistema de ecuaciones lineales

$$Cx = d$$

es equivalente al sistema $Ax = b$.

Demostración. Utilizando la biyección entre sistemas lineales y matrices se construye la matriz ampliada

$$\mathcal{A} := \begin{bmatrix} A & b \end{bmatrix}$$

asociada al sistema de ecuaciones lineales $Ax = b$ dado. Ahora se calcula la forma escalonada reducida por filas de $\mathcal{A}$ y se denota, como en el enunciado, por

$$R_{\mathcal{A}} := \begin{bmatrix} C & d \end{bmatrix},$$

donde se ha particionado en bloques la matriz hallada $R_{\mathcal{A}}$, separando la última columna.

De nuevo, utilizando la biyección entre sistemas lineales y matrices se tiene que el sistema asociado a

$$R_{\mathcal{A}} := \begin{bmatrix} C & d \end{bmatrix}$$

es $Cx = d$. Como este último sistema se obtiene de aplicar operaciones elementales al sistema $Ax = b$, el Teorema 3.1 asegura que ambos sistemas son equivalentes. □

Ejemplo 4.7. *Resolver el sistema de ecuaciones lineales $Ax = b$ para*

$$A = \begin{bmatrix} -4 & 4 & -4 \\ 1 & 2 & -2 \\ 4 & 2 & -2 \end{bmatrix} \qquad y \qquad b = \begin{bmatrix} 20 \\ 4 \\ -2 \end{bmatrix}$$

utilizando el sistema equivalente obtenido al realizar la forma escalonada reducida.

◁ El Teorema 4.3 asegura que el sistema de ecuaciones lineales $Ax = b$ es equivalente, usando el resultado del Ejemplo 4.6, al sistema lineal $Cx = d$ donde

$$C = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix} \qquad \text{y} \qquad d = \begin{bmatrix} -2 \\ 3 \\ 0 \end{bmatrix}.$$

Por tanto, el sistema $Ax = b$ es compatible indeterminado y el conjunto solución está formado por elementos de la forma $\begin{bmatrix} -2 \\ 3+k \\ k \end{bmatrix}$, $k \in \mathbb{R}$. ▷

4.4.3. Matriz escalonada reducida por columnas

El trabajo realizado por filas tiene su contrapartida operando por columnas. Sólo se establecerá la definición pues el resto de la teoría es similar.

Las operaciones elementales se deben considerar aplicadas a las columnas (en lugar de aplicadas a las filas) de la matriz que se esté analizando.

Definición 4.4. Sea $R \in \mathbb{K}^{m\times n}$. Se dice que R es **escalonada reducida por columnas** si se cumplen las siguientes condiciones:

(a) Toda columna nula de R debe aparecer después de todas las columnas no nulas.

(b) El primer elemento no nulo de cada columna no nula, llamado **uno principal**, es 1 y, las filas de R que contienen a esos 1 tienen los restantes elementos iguales a 0, y se llaman **filas básicas**.

(c) Si las columnas no nulas de R son $1, 2, \ldots, r$ y si el primer elemento no nulo de la columna j-ésima aparece en la fila i_j-ésima entonces $i_1 < i_2 < \ldots < i_r$.

Ejemplo 4.8. *Las matriz identidad I_n y la matriz nula $O_{m\times n}$ son escalonadas reducidas por columnas.*

Ejemplo 4.9. *Las matrices reales A y B son escalonadas reducidas por columnas:*

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

y las matrices C y D no lo son

$$C = \begin{bmatrix} 0 & 2 & 3 & 1 & 4 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}, \qquad D = \begin{bmatrix} 1 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 2 & -1 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

$\triangleleft$ La primera columna de C es nula y hay columnas no nulas a posteriori. El primer elemento no nulo de la primera columna no nula de la matriz D es 1 pero en la fila que lo contiene figura un $3 \neq 0$. $\triangleright$

4.5. Rango de una matriz

Definición 4.5. Sea $A \in \mathbb{K}^{m\times n}$. Se llama **rango**[a] de A, y se denota $\operatorname{rg}(A)$, a la cantidad de filas no nulas de su forma escalonada reducida por filas R_A.

[a] Si bien se ha definido el rango de A como el número de *filas* no nulas de su forma escalonada reducida por filas R_A, en realidad, este número se debería llamar **rango fila** de A. De manera similar, se podría definir el **rango columna** de A como el número de *columnas* no nulas de su forma escalonada reducida por columnas. Más adelante se volverá sobre este punto.

Lema 4.3 (Cálculo del rango). Sea $A \in \mathbb{K}^{m\times n}$ una matriz con forma escalonada reducida por filas R_A. Entonces

$$\begin{aligned}\mathrm{rg}(A) &= \text{cantidad de unos principales en } R_A\\ &= \text{cantidad de columnas básicas en } R_A.\end{aligned}$$

Demostración. La prueba es inmediata por la existencia y unicidad de la forma escalonada reducida por filas R_A de A (Teorema 4.2) y de la propia Definición 4.1. Se recuerda que los unos principales se cuentan de las filas de R_A y las columnas básicas son las columnas que contienen esos unos principales, por tanto, coinciden en cantidad. □

Ejemplo 4.10. *El rango de la matriz del Ejemplo 4.6 es 2 puesto que R_A tiene 2 filas no nulas o bien pues tiene 2 unos principales o, también, por tener 2 columnas básicas.*

Ejemplo 4.11. *Del Ejemplo 4.3 se tiene que* $\mathrm{rg}(I_n) = n$ *y* $\mathrm{rg}(O_{m\times n}) = 0$.

Definición 4.6. Dada una matriz $A \in \mathbb{K}^{m\times n}$, se suele llamar **matriz escalonada por filas** de A a toda matriz $E \in \mathbb{K}^{m\times n}$ que cumpla sólo las condiciones (a) y (c) de la Definición 4.1.

Observación 4.3. Es evidente que, mientras que la forma escalonada reducida por filas de A es única, esta matriz E escalonada por filas de A de la Definición 4.6 no tiene por qué serlo pues, por ejemplo, no requiere pivotes iguales a uno. Sin embargo, a la hora de hallar el rango de una matriz es suficiente calcular una forma escalonada (sin necesidad de llegar hasta la forma escalonada reducida por filas), puesto que la cantidad de filas no nulas será la misma en ambas formas.

La siguiente es la primera propiedad básica del rango de una matriz.

Proposición 4.1. Sea $A \in \mathbb{K}^{m\times n}$. Entonces $\mathrm{rg}(A) \leq \text{mín}\{m, n\}$.

Demostración. Por definición, es evidente que el valor del rango de A no puede superar el número de filas de R_A (que coincide con el de A), es decir, $\mathrm{rg}(A) \leq m$. Por otro lado, teniendo en cuenta el Lema 4.3, $\mathrm{rg}(A)$ es la cantidad de unos principales en R_A, que están situados en columnas concretas (las básicas) de R_A, y por tanto no puede superar n, que es el número de columnas de A. Ambas situaciones se expresan simultáneamente mediante $\mathrm{rg}(A) \leq \text{mín}\{m, n\}$. □

Ejemplo 4.12. *¿Cuáles son los posibles valores del rango de una matriz no nula $A \in \mathbb{R}^{3\times 2}$? Justificar.*

◁ Se ha visto que el rango de la matriz nula es 0. Además, es la única matriz que tiene rango 0. Entonces, puesto que $A \neq O$, se tiene que $\mathrm{rg}(A) \geq 1$. En este caso, $m = 3$ y $n = 2$, y el mínimo entre ellos es 2. Luego, $\mathrm{rg}(A) \leq 2$. Las posibilidades son: $\mathrm{rg}(A) = 1$ ó $\mathrm{rg}(A) = 2$. ▷

4.5.1. Clasificación y compatibilidad de sistemas lineales

En esta sección se estudian los sistemas de ecuaciones lineales desde dos puntos de vista:

(1) Se realiza su clasificación utilizando la forma normal de Hermite (por filas) asociada a la matriz ampliada.

(2) Se caracteriza su compatibilidad comparando el rango de la matriz de coeficientes con el de su matriz ampliada.

(1) <u>Mediante la forma normal de Hermite (por filas) asociada a la matriz ampliada</u>:

Dado un sistema de m ecuaciones lineales con n incógnitas del tipo (3.1) (página 154), se considera su matriz ampliada $\begin{bmatrix} A & b \end{bmatrix}$, donde la matriz A es la matriz de coeficientes, la matriz columna b contiene los términos independientes y se denota $r := \mathrm{rg}(A)$.

Se busca la forma escalonada reducida por filas $\begin{bmatrix} R & c \end{bmatrix}$ de la matriz ampliada. En la Sección 3.4 se ha realizado la clasificación a partir del método de eliminación de Gauss; ahora se clasificará nuevamente, en función de su forma normal de Hermite por filas:

- Si al realizar los cálculos, alguna de las matrices intermedias o bien la propia matriz $\begin{bmatrix} R & c \end{bmatrix}$ tiene una fila con todos los elementos nulos excepto el último, es decir del tipo

$$\left[\begin{array}{cccc|c} 0 & 0 & \dots & 0 & \alpha \end{array}\right], \text{ con } \alpha \neq 0,$$

  entonces la ecuación correspondiente en el sistema asociado será del tipo:

$$0x_1 + 0x_2 + \cdots + 0x_n = \alpha,$$

  que evidentemente no tiene solución. Se trata de un **sistema incompatible** (abreviadamente, SI).

- Si en la matriz $\begin{bmatrix} R & c \end{bmatrix}$ no aparecen filas del tipo mencionado, se trata de un **sistema compatible** (abreviadamente, SC).

  Ahora se escribe el sistema lineal $Rx = c$ y las incógnitas correspondientes a las columnas básicas, digamos $x_{j_1}, x_{j_2}, \dots, x_{j_r}$, figuran cada una en una sola ecuación con coeficiente 1.

  - Si $r = n$ entonces el sistema tiene solución única en $\mathbb{K}$. En este caso, el sistema es **compatible determinado** (abreviadamente, SCD).
  - Si $r < n$ entonces se asignan valores arbitrarios a las $n-r$ incógnitas restantes (es decir, a las que no corresponden a columnas básicas) y se calculan en función de ellos los valores de las incógnitas

básicas $x_{j_1}, x_{j_2}, \ldots, x_{j_r}$. Se obtienen así todas las soluciones del sistema, que en este caso tiene más de una solución. En este caso, se trata de un **sistema compatible indeterminado**[4] (abreviadamente, SCI).

(2) Caracterización de su compatibilidad: la proporciona el conocido como Teorema de Rouché-Frobenius.

El rango de una matriz, en combinación con el de la matriz ampliada asociada a un sistema de ecuaciones lineales, es una herramienta clave para establecer la compatibilidad de dicho sistema.

La importancia del Teorema de Rouché-Frobenius radica en que permite clasificar un sistema de ecuaciones lineales sobre un cuerpo arbitrario $\mathbb{K}$ a partir de los rangos de la matriz de coeficientes del sistema y de la matriz ampliada, sin necesidad de resolverlo.

Teorema 4.4 (Teorema de Rouché-Frobenius)**.** Se considera el sistema lineal $Ax = b$ con $A \in \mathbb{K}^{m\times n}$ y $b \in \mathbb{K}^{m\times 1}$. Entonces

(a) $Ax = b$ es un SC $\quad\Leftrightarrow\quad \operatorname{rg}(A) = \operatorname{rg}([\ A \quad b\ ])$.

- (I) $Ax = b$ es un SCD $\quad\Leftrightarrow\quad \operatorname{rg}(A) = \operatorname{rg}([\ A \quad b\ ]) = n$.
- (II) $Ax = b$ es un SCI $\quad\Leftrightarrow\quad \operatorname{rg}(A) = \operatorname{rg}([\ A \quad b\ ]) < n$.

(b) $Ax = b$ es un SI $\quad\Leftrightarrow\quad \operatorname{rg}(A) < \operatorname{rg}([\ A \quad b\ ])$.

Si $Ax = b$ es un SCI, todas sus soluciones se pueden expresar en términos de $n - \operatorname{rg}(A)$ parámetros.

Demostración. Sea $[\ R \quad c\ ]$ la forma escalonada reducida por filas de la matriz ampliada $[\ A \quad b\ ]$ (que existe y es única por el Teorema 4.2). Aplicando

[4]Como se ha visto previamente, el adjetivo *indeterminado* se debe a que el sistema puede tener un número finito o infinito de soluciones, en función del cuerpo $\mathbb{K}$ considerado.

el Lema 4.2, se tiene que la forma escalonada reducida por filas de la matriz A es R.

(a) Por lo demostrado en el Apartado (1) anterior, el sistema $Ax = b$ es compatible si y sólo si al calcular la forma escalonada reducida de su matriz ampliada no aparecen filas del tipo $\begin{bmatrix} 0 & 0 & \dots & 0 & \alpha \end{bmatrix}$, con $\alpha \neq 0$. Esta última condición es equivalente a decir que R y $[\ R \quad c\ ]$ tienen el mismo número de filas no nulas y, por definición, a $\operatorname{rg}(A) = \operatorname{rg}([\ A \quad b\ ])$.

   (I) En este caso, la condición $\operatorname{rg}(A) = n$ es necesaria y suficiente para que todas las incógnitas del sistema $Ax = b$ sean básicas (es decir, asociadas a columnas básicas). Por el Apartado (1), esto equivale a decir que el sistema es compatible determinado.

   (II) Se deduce del contrarrecíproco de (a)-(I) y recordando que, en general, $\operatorname{rg}(A) \leq n$.

(b) Sigue de forma inmediata tras la negación del apartado (a) y teniendo en cuenta que, en general, $\operatorname{rg}(A) \leq \operatorname{rg}([\ A \quad b\ ])$.

Cuando el sistema $Ax = b$ es compatible indeterminado, se despejan las incógnitas básicas en función de las libres. De este modo, la solución general se puede expresar en términos de esos $n - \operatorname{rg}(A)$ parámetros. $\square$

4.5.2. Soluciones de sistemas lineales homogéneos

Un caso especialmente importante es el de los sistemas de ecuaciones lineales homogéneos.

Comos se ha observado con anterioridad, el sistema $Ax = 0$ es siempre compatible, pues admite al menos la solución trivial $x = 0$.

Los resultados siguientes proporcionan condiciones para determinar si la solución trivial es su única solución.

Proposición 4.2 (Soluciones no triviales de un sistema homogéneo)**.**
Sea $A \in \mathbb{K}^{m\times n}$. Si $n > m$ entonces el sistema lineal homogéneo

$$Ax = 0$$

admite al menos una solución no trivial (es decir, diferente de la solución trivial $x = 0$).

Demostración. Como siempre se cumple que $\mathrm{rg}(A) \leq m$, por hipótesis, se tiene que $\mathrm{rg}(A) \leq m < n$. Luego, es claro que

$$\mathrm{rg}\left(\left[\begin{array}{cc} A & 0 \end{array}\right]\right) = \mathrm{rg}(A) < n.$$

Entonces, el Teorema de Rouché-Frobenius asegura que el sistema $Ax = 0$ es compatible indeterminado con $n - \mathrm{rg}(A) > 0$ parámetros. Asignando a una incógnita libre (no básica) el valor 1 ($0 \neq 1 \in \mathbb{K}$) se consigue una solución no trivial de $Ax = 0$. □

La siguiente es una consecuencia inmediata del Teorema de Rouché-Frobenius.

Corolario 4.1 (Soluciones de un sistema homogéneo)**.** Sea $A \in \mathbb{K}^{m\times n}$. El sistema lineal

$$Ax = 0$$

admite:

(a) únicamente la solución trivial $x = 0$ si y sólo si $\mathrm{rg}(A) = n$.

(b) al menos una solución no trivial si y sólo si $\mathrm{rg}(A) < n$.

Demostración. Aplicar el Teorema de Rouché-Frobenius al caso $b = 0$. □

La Tabla 4.1 muestra la clasificación de los sistemas lineales utilizando el Teorema de Rouché-Frobenius.

Clasificación		**No homogéneo** ($Ax = b$)	**Homogéneo** ($Ax = 0$)
SC	SCD	$\text{rg}(A) = \text{rg}(\begin{bmatrix} A & b \end{bmatrix}) = n$	$\text{rg}(A) = n$ ($x = 0$, $n \leq m$)
	SCI	$\text{rg}(A) = \text{rg}(\begin{bmatrix} A & b \end{bmatrix}) < n$	$\text{rg}(A) < n$
SI		$\text{rg}(A) < \text{rg}(\begin{bmatrix} A & b \end{bmatrix})$	Imposible

Tabla 4.1: Clasificación de sistemas lineales

Algoritmo: Resolución de un sistema lineal $Ax = b$ con n incógnitas:

- Aplicar el Algoritmo de Gauss a la matriz ampliada $\begin{bmatrix} A & b \end{bmatrix}$ hasta obtener una forma escalonada.
- Si $\text{rg}\left(\begin{bmatrix} A & b \end{bmatrix}\right) \neq \text{rg}(A)$ entonces el sistema es incompatible. En caso contrario,
    - Si $\text{rg}(A) = n$ entonces el sistema es determinado.
    - Si $\text{rg}(A) < n$ entonces el sistema es indeterminado.

  Para resolverlo (es decir, encontrar todas las soluciones):

    - Completar el algoritmo de Gauss-Jordan hasta obtener la forma escalonada reducida por filas de la matriz ampliada.
    - En cada ecuación del sistema lineal reducido, despejar la incógnita básica.
    - Sustituir las incógnitas no básicas por parámetros.

4.6. EJERCICIOS

(1) Sea $\mathcal{S}$ el conjunto formado por los sistemas de m ecuaciones lineales con n incógnitas a coeficientes en $\mathbb{K}$. Demostrar que existe una biyección entre $\mathcal{S}$ y $\mathbb{K}^{m\times(n+1)}$. Particularizar esta biyección al caso de sistemas de ecuaciones lineales homogéneos.

(2) Utilizando notación matricial, expresar el sistema de ecuaciones lineales (4.1) como

$$x_1c_1 + x_2c_2 + \cdots + x_nc_n = b,$$

donde, para cada $j = 1, 2, \ldots, n$,

$$c_j := \begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{bmatrix} \qquad \text{y} \qquad b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix},$$

y, de manera equivalente, mediante

$$\begin{bmatrix} c_1 & c_2 & \cdots & c_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = b.$$

(3) Aplicar el método de Gauss-Jordan para resolver el sistema de ecuaciones lineales

$$\left\{\begin{array}{rcrcrcrcrcl} x_1 & + & 5x_2 & - & 3x_3 & + & 4x_4 & + & 6x_5 & = & b_1 \\ 3x_1 & + & 6x_2 & - & x_3 & + & 2x_4 & + & 8x_5 & = & b_2 \\ x_1 & - & x_2 & + & 2x_3 & + & 3x_4 & - & x_5 & = & b_3 \end{array}\right.$$

para los casos:

(a) $b_1 = 2, b_2 = 2, b_3 = 4.$

(b) $b_1 = b_2 = b_3 = 0.$

Para el primer caso, si se trata de un sistema compatible indeterminado, indicar el número de parámetros necesarios para escribir el conjunto solución, su solución general en notación matricial y contrastar con la información obtenida al respecto en el Teorema de Rouché-Frobenius. Contrastar el segundo caso con la Proposición 4.2 y con el Corolario 4.1.

(4) Encontrar la forma escalonada reducida por filas de la matriz

$$\begin{bmatrix} -1 & -1 & -1 & 0 & 1 \\ 1 & 2 & 1 & 4 & 4 \\ 1 & 1 & 1 & 2 & 1 \end{bmatrix}.$$

Identificar los unos principales, sus columnas básicas y hallar su rango.

(5) De todas las soluciones del siguiente sistema de ecuaciones lineales homogéneo

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

determinar aquellas que también satisfacen la relación no lineal $z = y^2(1-x)$. Indicar el significado geométrico de esta situación.

(6) Calcular el rango de las siguientes matrices en el cuerpo de los números reales:

$$A = \begin{bmatrix} 1 & 2 & 5 \\ 2 & 4 & 8 \end{bmatrix}, \qquad B = \begin{bmatrix} 1 & 3 & 2 \\ 2 & 6 & 4 \\ -1 & -3 & -2 \\ 0 & 0 & 0 \end{bmatrix}$$

y

$$C = \begin{bmatrix} 15 & 1 & 17 & -30 & 51 \\ 23 & 2 & 26 & -46 & 78 \\ 17 & 3 & 21 & -34 & 63 \\ 45 & 4 & 50 & -90 & 150 \end{bmatrix}.$$

(7) Se considera la matriz

$$A = \begin{bmatrix} 1 & 2 & -1 \\ -2 & -3 & -1 \\ 3 & 7 & -6 \end{bmatrix}.$$

(a) ¿Existe $b \in \mathbb{R}^3$ tal que el sistema $Ax = b$ sea compatible determinado? Justificar.

(b) Utilizar el método de Gauss-Jordan para resolver simultáneamente los dos sistemas $Ax = b_1$ y $Ax = b_2$ siendo

$$b_1 = \begin{bmatrix} 1 \\ -2 \\ -1 \end{bmatrix} \qquad \text{y} \qquad b_2 = \begin{bmatrix} -1 \\ 5 \\ 0 \end{bmatrix}.$$

(c) Hallar la forma escalonada reducida y el rango de A.

(8) Sea $A \in \mathbb{K}^{m \times n}$. Demostrar que $\mathrm{rg}(kA) = \mathrm{rg}(A)$, para todo $k \in \mathbb{K}$, $k \neq 0$.

(9) Demostrar que un sistema de m ecuaciones lineales a coeficientes en un cuerpo $\mathbb{K}$ con n incógnitas que cumple $m < n$ sólo puede ser incompatible o compatible indeterminado.

(10) Discutir la compatibilidad del sistema en función de los parámetros $a, b \in \mathbb{R}$

$$\begin{cases} x + by + z = 0 \\ x + y + z = 1 \\ 2x + (b+1)y + az = 1 \end{cases}$$

y, si es posible, encontrar el conjunto solución.

(11) Comprobar que cada uno de los sistemas siguientes es compatible indeterminado

$$\begin{cases} x = 1 \\ 3y - 2z = 1 \end{cases} \qquad \text{y} \qquad \begin{cases} 3x - (b+1)z = 2 \\ 3y - 2z = 3a - 2 \end{cases}$$

Luego, buscar las condiciones que deben satisfacer $a, b \in \mathbb{R}$ para que ambos sistemas tengan una solución en común. Determinar dicha solución.

(12) Resolver los siguientes sistemas de ecuaciones lineales con coeficientes en el cuerpo $\mathbb{K} = \mathbb{Z}_2$

$$\left\{\begin{array}{ccccccc} x_1 & + & x_2 & + & x_3 & = & 1 \\ x_1 & & & + & x_3 & = & 1 \\ x_1 & + & x_2 & & & = & 0 \end{array}\right. , \qquad \left\{\begin{array}{ccccccc} x_1 & + & x_2 & & & = & 0 \\ x_1 & & & + & x_3 & = & 0 \\ & & x_2 & + & x_3 & = & 0 \end{array}\right.$$

y

$$\left\{\begin{array}{ccccccc} x_1 & + & x_2 & - & x_3 & = & 0 \\ x_1 & - & x_2 & - & x_3 & = & 0 \end{array}\right.$$

¿Cuántas soluciones tienen estos sistemas de ecuaciones? Resolver el segundo sistema considerando que sus coeficientes pertenecen a $\mathbb{R}$ y comparar con lo obtenido. Hallar la forma escalonada reducida por filas de la matriz ampliada de cada sistema y su rango (en $\mathbb{K} = \mathbb{Z}_2$).

Capítulo 5

Matrices invertibles

Índice

5.1. Introducción

Como el propio Cayley dijo, la idea de matriz (y muchas de sus propiedades) ya estaba clara antes de haber sido introducido propiamente dicho concepto. Habían sido desarrolladas a medida que se estudiaban los determinantes, aunque sin aislar un concepto del otro. Sin embargo, históricamente, el orden en que surgieron fue el inverso.

Una de las importantes propiedades de una matriz cuadrada es la de ser (o no) invertible. Arthur Cayley introdujo por primera vez el concepto de matriz inversa, en el mismo artículo en el que consideró el producto de matrices, titulado "A Memoir on the Theory of Matrices", Philosophical Transactions of the Royal Society of London, vol. CXLVIII, 17–37, 1858 y presentó la fórmula que permite calcular la inversa a partir del cálculo del determinante para una matriz de tamaño 3×3.

En consideración a la claridad en la notación que ofrecía el concepto de matriz de Cayley, el matemático francés Pierre-Simon Laplace (1749-1827) dijo:

> "Tal es la ventaja de un lenguaje bien construido que su notación simplificada frecuentemente llega a ser la fuente de profundas teorías".

Muchos problemas reales se encuentran modelizados mediante una ecuación del tipo $Ax = y$, que permite obtener el valor de la salida y a partir de una entrada x, donde la salida se obtiene premultiplicando por A la entrada x. Si se conoce que A es invertible, el proceso será reversible, y será posible encontrar la entrada x a partir de la salida y y de la matriz inversa A^{-1} mediante el cálculo $x = A^{-1}y$.

En este capítulo se caracteriza cuándo una matriz (cuadrada) es invertible desde diferentes puntos de vista. En uno de los enfoques se necesita el concepto de matrices elementales. Estas matrices permiten traducir las operaciones elementales (estudiadas en capítulos anteriores) a productos de matrices. También se deduce el método de Gauss-Jordan para el cálculo de la inversa de una matriz dada, y se estudia la inversa de matrices particionadas mediante la fórmula de Banachiewicz-Schur.

5.2. Definición

Definición 5.1. Sea $A \in \mathbb{K}^{n\times n}$. Se dice que A es una matriz **invertible**, **regular** o **no singular**[a] si existe una matriz $B \in \mathbb{K}^{n\times n}$ tal que

$$AB = BA = I_n.$$

En este caso, la matriz B se llama **inversa** de A.

[a] En el Capítulo 7 se retomará el concepto de regular o no singular.

Del mismo modo que se hizo al definir elemento inverso a derecha e izquierda en un grupo, se puede hacer en matrices. Más aún, teniendo en cuenta que dos matrices, en general, no conmutan.

Definición 5.2. Sea $A \in \mathbb{K}^{n\times n}$. Se dice que[a]:

- A es **invertible a izquierda** si existe una matriz $B \in \mathbb{K}^{n\times n}$ tal que $BA = I_n$. En este caso, la matriz B se llama **inversa a izquierda** de A.
- A es **invertible a derecha** si existe una matriz $C \in \mathbb{K}^{n\times n}$ tal que $AC = I_n$. En este caso, la matriz C se llama **inversa a derecha** de A.

[a] Estas definiciones pueden hacerse, de manera más general, para matrices rectangulares.

Al considerar matrices inversas a derecha e izquierda se habla, en general, de **inversas laterales**.

Ejemplo 5.1. *Dadas las matrices*

$$A_1 = \begin{bmatrix} 1 & 4 \\ 1 & 5 \end{bmatrix} \in \mathbb{R}^{2\times 2} \qquad y \qquad A_2 = \begin{bmatrix} 1 & 1 \\ -2 & -2 \end{bmatrix} \in \mathbb{R}^{2\times 2},$$

comprobar que la matriz

$$B_1 = \begin{bmatrix} 5 & -4 \\ -1 & 1 \end{bmatrix} \in \mathbb{R}^{2\times 2}$$

es (una) inversa de A_1; mientras que, la matriz A_2 no es invertible.

$\triangleleft$ Es fácil comprobar que se cumple que $A_1B_1 = B_1A_1 = I_2$, luego A_1 es invertible.

Ahora, suponiendo que

$$B_2 = \begin{bmatrix} x & y \\ z & t \end{bmatrix}, \qquad \text{con } \; x, y, z, t \in \mathbb{R},$$

se tiene que, tras multiplicar las matrices A_2 por B_2 e igualar dicho producto a la matriz I_2, el sistema de ecuaciones lineales al que se llega es incompatible. Esto garantiza que la matriz A_2 no es invertible. Se propone completar todos los detalles como ejercicio. $\triangleright$

Del ejemplo anterior se aprecia que hay matrices que pueden carecer de inversa. En el siguiente resultado se probará que puede existir, a lo sumo, una matriz que satisfaga las condiciones de la Definición 5.1. Es decir, no siempre existe la inversa de una matriz cuadrada, pero cuando existe, debe ser única.

Proposición 5.1. Sea $A \in \mathbb{K}^{n\times n}$. La inversa de la matriz A, si existe, es única.

Demostración. Sean $B_1, B_2 \in \mathbb{K}^{n\times n}$ dos matrices tales que

$$AB_1 = B_1A = I_n \qquad \text{y} \qquad AB_2 = B_2A = I_n.$$

Entonces

$$B_1 = B_1 I_n = B_1(AB_2) = (B_1A)B_2 = I_nB_2 = B_2.$$

Por tanto, si existe un elemento $B \in \mathbb{K}^{n\times n}$ tal que $AB = BA = I_n$, entonces es único. $\square$

Ahora es posible denotar por A^{-1} a *la* (única) inversa de una matriz invertible $A \in \mathbb{K}^{n\times n}$.

Observación 5.1. Otra observación importante a realizar sobre matrices invertibles es que únicamente se pueden considerar sobre el anillo de matrices cuadradas $\mathbb{K}^{n\times n}$ (véase el Ejercicio 24 de la página 147, que es válido, no sólo en $\mathbb{R}^{n\times n}$, sino también sobre un cuerpo arbitrario $\mathbb{K}$).

La Proposición 5.1 permite deducir la siguiente consecuencia.

Corolario 5.1. Sea $A \in \mathbb{K}^{n\times n}$. Si A admite una inversa a izquierda $B_1 \in \mathbb{K}^{n\times n}$ y una inversa a derecha $B_2 \in \mathbb{K}^{n\times n}$ entonces $B_1 = B_2$; en consecuencia, A es invertible.

Demostración. Repasando con detalle la demostración de la Proposición 5.1, se observa que sólo se han usado las igualdades $B_1A = I_n$ y $AB_2 = I_n$, que son precisamente las hipótesis de este corolario. $\square$

Si bien en el corolario anterior se pide como hipótesis la existencia de una inversa a izquierda y de una inversa a derecha de una matriz $A \in \mathbb{K}^{n\times n}$ dada, se verá en el Teorema de caracterización de matriz invertible que con asegurar la existencia de una sola de las inversas laterales, se conseguirá regularidad en la matriz cuadrada A.

5.3. Propiedades

La inversa de una matriz cumple las siguientes propiedades.

Proposición 5.2. Sean $A, B \in \mathbb{K}^{n\times n}$. Se cumple que:

Inversa de una inversa: Si A es invertible entonces A^{-1} también lo es y, además, $(A^{-1})^{-1} = A$.

Inversa de un producto: Si A y B son invertibles entonces AB también lo es y, además, $(AB)^{-1} = B^{-1}A^{-1}$. Esta propiedad es conocida como **ley del orden inverso**.

Inversa de una traspuesta: Si A es invertible entonces A^t también lo es y, además, $(A^t)^{-1} = (A^{-1})^t$.

Demostración. En todos los casos, para probar que la inversa de una matriz existe se debe encontrar cierta matriz que multiplicada a izquierda y a derecha por la matriz de partida den la matriz identidad[1].

- Para probar que la inversa de A^{-1} es A se debe comprobar que el producto de ambas (en ambos sentidos) es la matriz I_n. En este caso, de la propia definición de A^{-1}, es evidente que $AA^{-1} = A^{-1}A = I_n$.

- Si A y B son invertibles, aplicando la propiedad asociativa, es posible realizar el producto:

  $$(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AI_nA^{-1} = (AI_n)A^{-1} = AA^{-1} = I_n.$$

  Análogamente, se prueba que $(B^{-1}A^{-1})(AB) = I_n$.

- Si A es invertible entonces, aplicando propiedades de la traspuesta, se tiene que

  $$A^t(A^{-1})^t = (A^{-1}A)^t = I_n^t = I_n.$$

  Análogamente, se prueba que $(A^{-1})^tA^t = I_n$.

$\square$

[1]Observar que con un único cálculo se prueban ambas cuestiones simultáneamente: que la matriz de partida es invertible y cuál es su inversa.

El conjunto de matrices invertibles de $\mathbb{K}^{n\times n}$, con el producto de matrices, forma un grupo.

Corolario 5.2. Para cada $n \in \mathbb{N}$, el conjunto

$$GL_n(\mathbb{K}) := \{A \in \mathbb{K}^{n\times n} : A \text{ es invertible}\},$$

considerando como operación binaria la multiplicación de matrices, es un grupo.

Demostración. En la Proposición 5.2 se ha probado que el producto de dos matrices de $GL_n(\mathbb{K})$ es invertible, con lo que el conjunto es cerrado para la operación multiplicación. Es decir, si $A, B \in GL_n(\mathbb{K})$ entonces $AB \in GL_n(\mathbb{K})$.

Dado que la asociatividad es válida en todo $\mathbb{K}^{n\times n}$, en particular lo es para las matrices de $GL_n(\mathbb{K})$.

Es fácil ver que la matriz I_n es invertible, y es conocido que es el elemento neutro para la multiplicación en $\mathbb{K}^{n\times n}$, en particular lo es para las matrices de $GL_n(\mathbb{K})$. Es decir, existe $I_n \in GL_n(\mathbb{K})$ tal que $AI_n = I_nA = A$, $\forall A \in GL_n(\mathbb{K})$.

En la Proposición 5.2 se ha probado que la inversa de una matriz de $GL_n(\mathbb{K})$ es invertible, de donde se tiene que: si $A \in GL_n(\mathbb{K})$ entonces $A^{-1} \in GL_n(\mathbb{K})$. $\square$

El grupo multiplicativo $GL_n(\mathbb{K})$ se llama **grupo lineal general** de orden n en $\mathbb{K}$.

Los resultados siguientes aseguran que si se conoce la existencia de una inversa lateral de una matriz invertible A, se puede asegurar que dicha inversa lateral coincide con la inversa de A.

Proposición 5.3. Sea $A \in \mathbb{K}^{n\times n}$ una matriz invertible. Si existe una matriz $B \in \mathbb{K}^{n\times n}$ tal que $BA = I_n$ entonces $B = A^{-1}$.

Demostración. Se supone que A es invertible. Por lo tanto, existe una (única) matriz $A^{-1} \in \mathbb{K}^{n\times n}$ tal que $AA^{-1} = A^{-1}A = I_n$. Si se cumple que $BA = I_n$ entonces

$$B = BI_n = B(AA^{-1}) = (BA)A^{-1} = I_nA^{-1} = A^{-1},$$

como se quería probar. □

La proposición anterior garantiza que si una matriz es invertible, para hallar A^{-1}, basta encontrar "una" inversa a derecha de A (pues será la única). El siguiente resultado es análogo y se demuestra de manera similar. La demostración se propone como ejercicio.

Proposición 5.4. Sea $A \in \mathbb{K}^{n\times n}$ una matriz invertible. Si existe una matriz $C \in \mathbb{K}^{n\times n}$ tal que $AC = I_n$ entonces $C = A^{-1}$.

En los dos resultados anteriores, la hipótesis sobre la regularidad de la matriz A se puede quitar, como se verá en el Teorema 5.1. En la Subsección 5.5 se probará, por ejemplo, que la existencia de una inversa lateral de una matriz cuadrada garantiza la regularidad de la matriz. Antes de ello, se introducirá la versión matricial de las operaciones elementales.

5.4. Matrices elementales

Definición 5.3. Sea $E \in \mathbb{K}^{n\times n}$. Se dice que E es una **matriz elemental** (por filas[a]) si se obtiene de la matriz identidad I_n por medio de la aplicación de una única operación elemental por filas.

[a]Las (mismas) matrices elementales se pueden definir utilizando operaciones elementales sobre las columnas de la matriz identidad.

Hay tres tipos de matrices elementales por filas, uno por cada tipo de operación elemental:

Matriz elemental de tipo I o de permutación: Se denota E_{ij} a la matriz que se obtiene intercambiando la fila i-ésima con la fila j-ésima de I_n.

Matriz elemental de tipo II o de escalado: Se denota $E_i(k)$ a la matriz que se obtiene multiplicando los elementos de la fila i-ésima de I_n por $k \in \mathbb{K}$, con $k \neq 0$.

Matriz elemental de tipo III o de eliminación: Se denota $E_{ij}(k)$ a la matriz que se obtiene sumando a la fila i-ésima la fila j-ésima de I_n multiplicada previamente por $k \in \mathbb{K}$.

Ejemplo 5.2. *Identificar las matrices siguientes como matrices elementales:*

$$E_{12} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \; E_2(-4) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -4 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \; E_{32}(6) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 6 & 1 \end{bmatrix}.$$

$\lhd$ La matriz E_{12} es una matriz elemental de tipo I (intercambio de primera y segunda filas de I_3).

La matriz $E_2(-4)$ es una matriz elemental de tipo II (la segunda fila de I_3 está multiplicada por -4).

Finalmente, la matriz $E_{32}(6)$ es una matriz elemental de tipo III (a la tercera fila de I_3 se le suma la segunda previamente multiplicada por 6). $\rhd$

Inversas de matrices elementales:

Proposición 5.5 (Las matrices elementales son invertibles)**.** Las matrices elementales de $\mathbb{K}^{n\times n}$ son invertibles y cada inversa es una matriz elemental del mismo tipo. Más precisamente,

$$[E_{ij}]^{-1} = E_{ji}, \qquad [E_i(k)]^{-1} = E_i\left(\frac{1}{k}\right), \qquad [E_{ij}(k)]^{-1} = E_{ij}(-k).$$

Demostración. Observar que, realmente, estas propiedades cobran sentido si $n \geq 2$.

Se supondrá que $1 \leq i < j \leq n$, que los bloques de ceros representan matrices de tamaños adecuados y en los casos extremos $i = 1$ y $j = n$, los bloques respectivos I_{i-1} e I_{n-j} pueden estar ausentes.

Para probar que $[E_{ij}]^{-1} = E_{ji}$, se multiplican matrices diagonales por bloques como sigue:

$$
\begin{aligned}
& E_{ij}E_{ji} = E_{ji}E_{ij} = \\
&= \left[\begin{array}{c|ccc|c} I_{i-1} & 0 & \dots & 0 & 0 \\ \hline 0 & 0 & \dots & 1 & 0 \\ \vdots & \vdots & I_{j-i-1} & \vdots & \vdots \\ 0 & 1 & \dots & 0 & 0 \\ \hline 0 & 0 & \dots & 0 & I_{n-j} \end{array}\right] \left[\begin{array}{c|ccc|c} I_{i-1} & 0 & \dots & 0 & 0 \\ \hline 0 & 0 & \dots & 1 & 0 \\ \vdots & \vdots & I_{j-i-1} & \vdots & \vdots \\ 0 & 1 & \dots & 0 & 0 \\ \hline 0 & 0 & \dots & 0 & I_{n-j} \end{array}\right] \\
&= \left[\begin{array}{c|ccc|c} I_{i-1} & 0 & \dots & 0 & 0 \\ \hline 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & I_{j-i-1} & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \\ \hline 0 & 0 & \dots & 0 & I_{n-j} \end{array}\right] \\
&= I_n.
\end{aligned}
$$

Observar que el bloque central (quitando 2 elementos que están escritos explícitamente) tiene

$$j - i - 1 = n - (n - j) - (i - 1) - 2$$

elementos, que se calculan como la cantidad de elementos totales (n) menos la cantidad de elementos del bloque inferior ($n - j$) menos la cantidad de elementos del bloque superior $i - 1$ menos los 2 elementos citados explícitamente.

Del mismo modo se comprueba que $[E_i(k)]^{-1} = E_i\left(\frac{1}{k}\right)$, siempre que $k \neq 0$:

$$\begin{aligned}
E_i(k)E_i\left(\frac{1}{k}\right) &= \left[\begin{array}{c|c|c} I_{i-1} & O & O \\ \hline O & k & O \\ \hline O & O & I_{n-i} \end{array}\right] \left[\begin{array}{c|c|c} I_{i-1} & O & O \\ \hline O & \frac{1}{k} & O \\ \hline 0 & O & I_{n-i} \end{array}\right] \\
&= \left[\begin{array}{c|c|c} I_{i-1} & O & O \\ \hline O & 1 & O \\ \hline O & O & I_{n-i} \end{array}\right] \\
&= I_n,
\end{aligned}$$

y de forma similar se prueba que $E_i\left(\frac{1}{k}\right)E_i(k) = I_n$.

Recordando que el producto de matrices triangulares superiores es triangular superior con el producto de sus elementos diagonales en la diagonal del resultado, el producto $E_{ij}(k)E_{ij}(-k)$ se calcula como

$$\begin{aligned}
&E_{ij}(k)E_{ij}(-k) = \\
&= \left[\begin{array}{c|cccc|c} I_{i-1} & 0 & \dots & & 0 & 0 \\ \hline 0 & 1 & \dots & & k & 0 \\ \vdots & \vdots & & I_{j-i-1} & \vdots & \vdots \\ 0 & 0 & \dots & & 1 & 0 \\ \hline 0 & 0 & \dots & & 0 & I_{n-j} \end{array}\right] \left[\begin{array}{c|cccc|c} I_{i-1} & 0 & \dots & & 0 & 0 \\ \hline 0 & 1 & \dots & & -k & 0 \\ \vdots & \vdots & & I_{j-i-1} & \vdots & \vdots \\ 0 & 0 & \dots & & 1 & 0 \\ \hline 0 & 0 & \dots & & 0 & I_{n-j} \end{array}\right] \\
&= \left[\begin{array}{c|cccc|c} I_{i-1} & 0 & \dots & & 0 & 0 \\ \hline 0 & 1 & \dots & & 0 & 0 \\ \vdots & \vdots & & I_{j-i-1} & \vdots & \vdots \\ 0 & 0 & \dots & & 1 & 0 \\ \hline 0 & 0 & \dots & & 0 & I_{n-j} \end{array}\right] \\
&= I_n,
\end{aligned}$$

y de forma similar se prueba que $E_{ij}(-k)E_{ij}(k) = I_n$. $\square$

Como observación, compárese la demostración de la Proposición 5.5 con la del Teorema 3.1 y su corolario.

Traspuesta de matrices elementales:

El siguiente resultado será necesario para demostrar una importante propiedad de determinantes en el Capítulo 7.

Proposición 5.6 (Las traspuestas de matrices elementales son elementales)**.** La matriz traspuesta de cada matriz elemental de $\mathbb{K}^{n\times n}$ es elemental del mismo tipo. Más precisamente,

$$[E_{ij}]^t = E_{ij}, \qquad [E_i(k)]^t = E_i\,(k)\,, \qquad [E_{ij}(k)]^t = E_{ji}(k).$$

Demostración. Como se ha realizado anteriormente, para probar estas propiedades, se utiliza el cálculo de matrices traspuestas por bloques y se supone que $1 \leq i < j \leq n$ y que los bloques de ceros representan matrices de tamaños adecuados:

$$[E_{ij}]^t = \left[\begin{array}{c|cccc|c} I_{i-1} & 0 & \dots & & 0 & 0 \\ \hline 0 & 0 & \dots & & 1 & 0 \\ \vdots & \vdots & & I_{j-i-1} & \vdots & \vdots \\ 0 & 1 & \dots & & 0 & 0 \\ \hline 0 & 0 & \dots & & 0 & I_{n-j} \end{array}\right]^t = \left[\begin{array}{c|cccc|c} I_{i-1} & 0 & \dots & & 0 & 0 \\ \hline 0 & 0 & \dots & & 1 & 0 \\ \vdots & \vdots & & I_{j-i-1} & \vdots & \vdots \\ 0 & 1 & \dots & & 0 & 0 \\ \hline 0 & 0 & \dots & & 0 & I_{n-j} \end{array}\right] = E_{ij}.$$

Del mismo modo se comprueba que $[E_i(k)]^t = E_i\,(k)$, siempre que $k \neq 0$:

$$[E_i(k)]^t = \left[\begin{array}{c|c|c} I_{i-1} & O & O \\ \hline O & k & O \\ \hline O & O & I_{n-i} \end{array}\right]^t = \left[\begin{array}{c|c|c} I_{i-1} & O & O \\ \hline O & k & O \\ \hline O & O & I_{n-i} \end{array}\right] = E_i(k).$$

Por último, se debe comprobar que $[E_{ij}(k)]^t = E_{ji}(k)$:

$$
\begin{aligned}
&[E_{ij}(k)]^t = \\
&= \left[\begin{array}{c|ccc|c}
I_{i-1} & 0 & \dots & 0 & 0 \\ \hline
0 & 1 & \dots & k & 0 \\
\vdots & \vdots & I_{j-i-1} & \vdots & \vdots \\
0 & 0 & \dots & 1 & 0 \\ \hline
0 & 0 & \dots & 0 & I_{n-j}
\end{array}\right]^t \\
&= \left[\begin{array}{c|ccc|c}
I_{i-1} & 0 & \dots & 0 & 0 \\ \hline
0 & 1 & \dots & 0 & 0 \\
\vdots & \vdots & I_{j-i-1} & \vdots & \vdots \\
0 & k & \dots & 1 & 0 \\ \hline
0 & 0 & \dots & 0 & I_{n-j}
\end{array}\right] \\
&= E_{ji}(k).
\end{aligned}
$$

□

En el teorema anterior, se observa que en el último caso la matriz elemental es del mismo tipo aunque hay un intercambio entre las filas y las columnas $[E_{ij}(k)]^t = E_{ji}(k)$.

Efecto de multiplicar una matriz cualquiera por una matriz elemental:

Ahora se analizará el efecto de multiplicar una matriz cualquiera $A \in \mathbb{K}^{m\times n}$ por una matriz elemental por la izquierda. Para ello son necesarias las siguientes <u>notaciones</u>:

- Para hacer referencia a una cualquiera de las tres operaciones elementales se escribirá simplemente e.
- $A \longrightarrow_{f,e} B$ indicará que $A \sim_f B$ tras aplicar únicamente la operación elemental e por filas.

- $E \in \mathbb{K}^{m\times m}$ (con E mayúscula) es la matriz elemental (de tipo I, II ó III) que resulta de aplicar la operación elemental e (con e minúscula) por filas a I_m, es decir, $E = e(I_m)$.

Proposición 5.7 (Operaciones elementales por filas versus matrices elementales)**.** Sean $A, B \in \mathbb{K}^{m\times n}$. Entonces

$$A \longrightarrow_{f,e} B \qquad \text{es equivalente a} \qquad EA = B. \tag{5.1}$$

Demostración. Al premultiplicar A por una matriz de permutación $E_{ij} \in \mathbb{K}^{m\times m}$ se obtiene

$$\begin{aligned} E_{ij}A &= \left[\begin{array}{c|ccc|c} I_{i-1} & 0 & \dots & 0 & 0 \\ \hline 0 & 0 & \dots & 1 & 0 \\ \vdots & \vdots & I_{j-i-1} & \vdots & \vdots \\ 0 & 1 & \dots & 0 & 0 \\ \hline 0 & 0 & \dots & 0 & I_{m-j} \end{array}\right] \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{i1} & a_{i2} & \dots & a_{in} \\ \vdots & \vdots & \vdots & \vdots \\ a_{j1} & a_{j2} & \dots & a_{jn} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \\ &= \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{j1} & a_{j2} & \dots & a_{jn} \\ \vdots & \vdots & \vdots & \vdots \\ a_{i1} & a_{i2} & \dots & a_{in} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}, \end{aligned}$$

que equivale a la operación elemental de intercambiar las filas i-ésima y j-ésima de A entre sí, y donde se ha supuesto que $i < j$.

Al considerar una matriz elemental de tipo escalado se tiene:

$E_i(k)A =$

$$= \left[\begin{array}{c|c|c} I_{i-1} & O & O \\ \hline O & k & O \\ \hline O & O & I_{m-i} \end{array}\right] \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{i1} & a_{i2} & \dots & a_{in} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ ka_{i1} & ka_{i2} & \dots & ka_{in} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix},$$

que equivale a la operación elemental que multiplica la fila i-ésima de la matriz A por el escalar $k \neq 0$.

Para una matriz elemental de eliminación $(i < j)$ se tiene:

$$E_{ij}(k)A =$$

$$= \left[\begin{array}{c|cccc|c} I_{i-1} & 0 & \dots & 0 & 0 \\ \hline 0 & 1 & \dots & k & 0 \\ \vdots & \vdots & I_{j-i-1} & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \\ \hline 0 & 0 & \dots & 0 & I_{m-j} \end{array}\right] \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{i1} & a_{i2} & \dots & a_{in} \\ \vdots & \vdots & \vdots & \vdots \\ a_{j1} & a_{j2} & \dots & a_{jn} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}$$

$$= \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{i1} + ka_{j1} & a_{i2} + ka_{j2} & \dots & a_{in} + ka_{jn} \\ \vdots & \vdots & \vdots & \vdots \\ a_{j1} & a_{j2} & \dots & a_{jn} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix},$$

que equivale a la operación elemental que a la fila i-ésima de A le suma la fila j-ésima previamente multiplicada por k.

Se observa que todos los razonamientos son válidos en ambos sentidos, tanto para la condición necesaria como para la suficiente del enunciado. $\square$

Parafraseando la implicación ($\Rightarrow$) del resultado anterior, la misma indica que EA es la matriz que se obtiene aplicando a las filas de A la misma operación elemental e con la que se obtiene E a partir de la identidad I_m.

Observación 5.2. El resultado indica que (en lenguaje de operaciones elementales) realizar una operación elemental e sobre las filas de una matriz es lo mismo que (en lenguaje de producto de matrices) premultiplicar la matriz por la correspondiente matriz elemental E (que es la que corresponde a efectuar la misma operación elemental e sobre la identidad I). En símbolos, si $A \in \mathbb{K}^{m\times n}$ entonces $e(A) = e(I_m)A$.

Es posible establecer un resultado similar para las columnas. Ahora, $A \longrightarrow_{c,e} B$ indica que $A \sim_c B$ tras aplicar únicamente la operación elemental columna e.

Proposición 5.8 (Operaciones elementales por columnas versus matrices elementales)**.** Sean $A, B \in \mathbb{K}^{m\times n}$. Entonces

$$A \longrightarrow_{c,e} B \qquad \text{es equivalente a} \qquad AE = B. \tag{5.2}$$

Demostración. La prueba del resultado sobre las columnas es análoga a la realizada para filas y se propone como ejercicio. □

El resultado indica que AE es la matriz que se obtiene aplicando a las columnas de A la misma operación elemental sobre las columnas con la que se obtiene E a partir de la identidad I_n.

Ejemplo 5.3. *Analizar el efecto de premultiplicar y posmultiplicar cada matriz elemental del Ejemplo 5.2 por una matriz arbitraria $A \in \mathbb{R}^{3\times 3}$.*

◁ El efecto de premultiplicar la primera matriz elemental del Ejemplo 5.2

por una matriz arbitraria $A \in \mathbb{R}^{3\times 3}$ es

$$E_{12}A = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} a_{21} & a_{22} & a_{23} \\ a_{11} & a_{12} & a_{13} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}.$$

Al posmultiplicar se obtiene

$$AE_{12} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} a_{12} & a_{11} & a_{13} \\ a_{22} & a_{21} & a_{23} \\ a_{32} & a_{31} & a_{33} \end{bmatrix}.$$

El efecto de premultiplicar la segunda matriz elemental del Ejemplo 5.2 a una matriz arbitraria $A \in \mathbb{R}^{3\times 3}$ es

$$E_2(-4)A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -4 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ -4a_{21} & -4a_{22} & -4a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}.$$

Al posmultiplicar se obtiene

$$AE_2(-4) = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & -4 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} a_{11} & -4a_{12} & a_{13} \\ a_{21} & -4a_{22} & a_{23} \\ a_{31} & -4a_{32} & a_{33} \end{bmatrix}.$$

El efecto de premultiplicar por la tercera matriz elemental del Ejemplo 5.2 a una matriz arbitraria $A \in \mathbb{R}^{3\times 3}$ es

$$\begin{aligned} E_{32}(6)A &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 6 & 1 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \\ &= \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31}+6a_{21} & a_{32}+6a_{22} & a_{33}+6a_{23} \end{bmatrix}. \end{aligned}$$

Al posmultiplicar se obtiene

$$
\begin{aligned}
AE_{32}(6) &= \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 6 \\ 0 & 0 & 1 \end{bmatrix} \\
&= \begin{bmatrix} a_{11} & a_{12} & a_{13} + 6a_{12} \\ a_{21} & a_{22} & a_{23} + 6a_{22} \\ a_{31} & a_{32} & a_{33} + 6a_{32} \end{bmatrix}.
\end{aligned}
$$

$\triangleright$

5.5. Caracterizaciones de matriz invertible

Hasta ahora se ha tratado la unicidad de la inversa de una matriz; se ha probado que si A es invertible entonces su inversa es única.

Al tratarse de un tema central, es importante establecer relaciones entre la no singularidad de una matriz (cuadrada) y diferentes aspectos del Álgebra Lineal.

En el siguiente resultado se aborda, desde diferentes puntos de vista, cuándo una matriz es invertible, es decir, mientras que antes se estudió su unicidad, ahora es el momento de analizar la existencia. Por ahora, para este objetivo se utilizarán:

(a) inversas laterales,

(b) sistemas de ecuaciones lineales: homogéneos y no homogéneos,

(c) la forma normal de Hermite por filas,

(d) el rango.

Más adelante, se encontrarán nuevas caracterizaciones para que una matriz sea invertible.

Teorema 5.1 (Caracterizaciones de matriz invertible)**.** Sea $A \in \mathbb{K}^{n\times n}$. Las siguientes condiciones son equivalentes:

(a) A es invertible.

(b) A admite una inversa a izquierda.

(c) A admite una inversa a derecha.

(d) El sistema homogéneo $Ax = 0$ tiene una única solución (la solución trivial $x = 0$).

(e) Todo sistema $Ax = b$ tiene una única solución, cualquiera que sea $b \in \mathbb{K}^{n\times 1}$.

(f) La forma escalonada reducida por filas R_A de A es la matriz identidad I_n.

(g) $A \sim_f I_n$.

(h) $\mathrm{rg}(A) = n$.

(i) $A = E_1E_2 \dots E_s$, para ciertas matrices elementales por filas $E_1, E_2, \dots, E_s \in \mathbb{K}^{n\times n}$.

Demostración. Primero se probará la cadena de implicaciones:

$$(a) \Rightarrow (b) \Rightarrow (d) \Rightarrow (f) \Rightarrow (g) \Rightarrow (i) \Rightarrow (a).$$

(a) $\Rightarrow$ (b) Es evidente que si A es invertible, en particular, A admite inversa a izquierda.

(b) $\Rightarrow$ (d) Se supone que A admite inversa a izquierda, es decir, existe $B \in \mathbb{K}^{n\times n}$ tal que $BA = I_n$. Se debe probar que si x satisface $Ax = 0$ entonces $x = 0$. En efecto, si $Ax = 0$ entonces multiplicando a izquierda ambos miembros de esta igualdad por B se tiene

$$x = I_nx = (BA)x = B(Ax) = B0 = 0,$$

por tanto, $x = 0$.

(d) $\Rightarrow$ (f) Por el Teorema 4.2, $A \sim_f R_A$ y, por el Lema 4.1, los sistemas lineales $Ax = 0$ y $R_A x = 0$ son equivalentes. Como, por hipótesis, el sistema $Ax = 0$ sólo tiene la solución trivial, el sistema $R_A x = 0$ debe tener únicamente la solución trivial. Este hecho, junto a que R_A es una matriz escalonada reducida por filas, obliga a que todas las columnas de R_A sean básicas (pues si R_A tuviese alguna fila nula, el sistema admitiría alguna solución no trivial[2]). Luego, $R_A = I_n$.

(f) $\Rightarrow$ (g) Se supone que $R_A = I_n$. Por el Teorema 4.2, es conocido que A es equivalente por filas a la (única) matriz escalonada reducida por filas R_A. Luego, $A \sim_f R_A = I_n$.

(g) $\Rightarrow$ (i) Si $A \sim_f I_n$ entonces existen s (un número finito de) operaciones elementales por filas que transforman A en I_n, es decir

$$A \longrightarrow_{f,e_1} B_1 \longrightarrow_{f,e_2} B_2 \longrightarrow \cdots \longrightarrow_{f,e_{s-1}} B_{s-1} \longrightarrow_{f,e_s} I_n.$$

Aplicando la Proposición 5.7 a cada operación elemental, existen s matrices elementales por filas $F_1, F_2, \ldots, F_s \in \mathbb{K}^{n\times n}$ tales que

$$F_s \ldots F_2 F_1 A = I_n.$$

Como las matrices elementales son invertibles, premultiplicando por sus inversas $F_s^{-1}, \ldots, F_2^{-1}, F_1^{-1}$ (en ese orden), es posible escribir

$$A = F_1^{-1} F_2^{-1} \ldots F_s^{-1} I_n = F_1^{-1} F_2^{-1} \ldots F_s^{-1}.$$

Ahora, como las inversas de las matrices elementales son también matrices elementales (véase Proposición 5.5), basta tomar $E_i = F_i^{-1}$ para $i = 1, 2, \ldots, s$ y se obtiene el resultado.

(i) $\Rightarrow$ (a) Si A es producto de matrices elementales por filas, el resultado es inmediato del hecho que las matrices elementales son invertibles y utilizando que el producto de matrices invertibles es invertible.

Ahora se probará la cadena de implicaciones:

[2]Quitando las filas nulas, se obtendría un sistema homogéneo con más incógnitas que ecuaciones y la Proposición 4.2 aseguraría la existencia de soluciones no triviales.

$$(c) \Rightarrow (a) \Rightarrow (e) \Rightarrow (c).$$

(c) $\Rightarrow$ (a) Si A admite una inversa a derecha, existe una matriz $C \in \mathbb{K}^{n\times n}$ tal que $AC = I_n$. Esto indica que la matriz C admite como inversa a izquierda a la matriz A. Puesto que se ha probado que (b) $\Leftrightarrow$ (a), se concluye que C es invertible. Por la Proposición 5.3, $A = C^{-1}$. Multiplicando a izquierda $A = C^{-1}$ por C se tiene $CA = CC^{-1} = I_n$. Luego, $AC = CA = I_n$. Por tanto, A es invertible.

(a) $\Rightarrow$ (e) Se supone que A es invertible y $b \in \mathbb{K}^{n\times 1}$ está fijo (pero es arbitrario). Entonces el sistema $Ax = b$ tiene solución pues

$$A(A^{-1}b) = (AA^{-1})b = I_n b = b,$$

es decir, $A^{-1}b$ es una solución de $Ax = b$.

Además, dicha solución es única pues si x_1 y x_2 son soluciones de $Ax = b$ entonces $Ax_1 = b = Ax_2$. Multiplicando a izquierda $Ax_1 = Ax_2$ por A^{-1} se obtiene

$$x_1 = I_n x_1 = (A^{-1}A)x_1 = A^{-1}(Ax_1) = A^{-1}(Ax_2) = (A^{-1}A)x_2 = I_n x_2 = x_2,$$

con lo que la solución es única.

(e) $\Rightarrow$ (c) Se consideran las matrices columna de $\mathbb{K}^{n\times 1}$:

$$e_1 := \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \qquad e_2 := \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \qquad \ldots, \qquad e_n := \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix},$$

que dan lugar a los sistemas de ecuaciones lineales

$$Ax_i = e_i, \qquad \text{para } i = 1, 2, \ldots, n.$$

Por hipótesis, cada uno de estos sistemas tiene solución única, por lo tanto, las respectivas soluciones y_i cumplen que $Ay_i = e_i$, $i = 1, 2, \ldots, n$. Luego, la matriz

$$B := \begin{bmatrix} y_1 & y_2 & \ldots & y_n \end{bmatrix},$$

satisface que

$$\begin{aligned} AB &= A\begin{bmatrix} y_1 & y_2 & \dots & y_n \end{bmatrix} \\ &= \begin{bmatrix} Ay_1 & Ay_2 & \dots & Ay_n \end{bmatrix} \\ &= \begin{bmatrix} e_1 & e_2 & \dots & e_n \end{bmatrix} \\ &= I_n. \end{aligned}$$

Se ha probado que $AB = I_n$, con lo que A admite inversa a derecha, lo que demuestra (c).

Por último falta demostrar que:

$$\text{(f)} \Leftrightarrow \text{(h)}.$$

En efecto,

(f) $\Rightarrow$ (h) Si $R_A = I_n$, las n columnas de la forma escalonada reducida por filas de A son básicas. Por el Lema 4.3, es evidente que $\operatorname{rg}(A) = n$.

(h) $\Rightarrow$ (f) Si $\operatorname{rg}(A) = n$, por el Lema 4.3, las n columnas de la forma escalonada reducida por filas de A son básicas. De la definición de forma escalonada reducida por filas, se sigue que $R_A = I_n$. □

Observación 5.3. De acuerdo al Teorema 5.1 es importante remarcar que, para $A \in \mathbb{K}^{n\times n}$:

- Alcanza con la regularidad a izquierda (o derecha) de A para garantizar que A es invertible.
- Los apartados (a), (b), (d), (f), (g), (h) e (i) están relacionados con la regularidad a izquierda de la matriz, mientras que los apartados (a), (c) y (e) están vinculados a la regularidad a derecha de A.

Ejemplo 5.4. *Analizar si la matriz*

$$A = \begin{bmatrix} \frac{2}{3} & \frac{1}{3} \\ -4 & -2 \end{bmatrix} \in \mathbb{R}^{2\times 2}$$

es invertible.

$\lhd$ Se calcula la forma escalonada reducida por filas de A y se utiliza el apartado (f) del Teorema 5.1. En efecto, efectuando operaciones elementales por filas

$$A = \begin{bmatrix} \frac{2}{3} & \frac{1}{3} \\ -4 & -2 \end{bmatrix} \sim_f \begin{bmatrix} 1 & \frac{1}{2} \\ -4 & -2 \end{bmatrix} \sim_f \begin{bmatrix} 1 & \frac{1}{2} \\ 0 & 0 \end{bmatrix} = R_A \neq I_2.$$

Por tanto, A no es invertible. Se puede concluir, además, que existe un sistema lineal $Ax = b$ (es decir, para alguna matriz columna $b \in \mathbb{R}^{2\times 1}$) que tendrá más de una solución. También pude concluirse que el sistema homogéneo $Ax = 0$ admitirá alguna solución no trivial. $\rhd$

Interpretación geométrica: La condición (i) del Teorema 5.1 tiene una interpretación geométrica interesante para el caso en que A sea una matriz invertible en $\mathbb{R}^{2\times 2}$. En efecto, las matrices elementales por filas en $\mathbb{R}^{2\times 2}$ son del tipo:

$$E_{12} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad E_1(k) = \begin{bmatrix} k & 0 \\ 0 & 1 \end{bmatrix}, \quad E_2(k) = \begin{bmatrix} 1 & 0 \\ 0 & k \end{bmatrix}, \quad k \neq 0,$$

$$E_{21}(c) = \begin{bmatrix} 1 & 0 \\ c & 1 \end{bmatrix}, \quad E_{12}(c) = \begin{bmatrix} 1 & c \\ 0 & 1 \end{bmatrix}.$$

El efecto que produce multiplicar cada una de estas matrices por un vector[3] $\begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^{2\times 1}$ (es decir, del plano OXY) es:

[3]En el capítulo de espacios vectoriales se estudiará en profundidad el concepto de vector. Por ahora, se puede leer este párrafo con la idea intuitiva que se tiene de vector de Geometría elemental del plano.

- $E_{12}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} y \\ x \end{bmatrix}$, que es el simétrico respecto de la bisectriz del primer cuadrante.

- Si $k > 0$, $E_1(k)\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} kx \\ y \end{bmatrix}$, que es una compresión $(0 < k < 1)$ o elongación $(k > 1)$ en un factor k paralela al eje OX.

  Por otro lado, $E_2(k)\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x \\ ky \end{bmatrix}$, que es una compresión $(0 < k < 1)$ o elongación $(k > 1)$ en un factor k paralela al eje OY.

- Si $k < 0$, $E_1(k)\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} kx \\ y \end{bmatrix}$, que es una compresión $(-1 < k < 0)$ o elongación $(k < -1)$ en un factor $-k$ paralela al eje OX seguida de una reflexión con respecto al eje OY.

  Por otro lado, $E_2(k)\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x \\ ky \end{bmatrix}$, que es una compresión $(-1 < k < 0)$ o elongación $(k < -1)$ en un factor $-k$ paralela al eje OY seguida de una reflexión con respecto al eje OX.

- $E_{21}(c)\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x \\ cx + y \end{bmatrix}$, que es una transvección paralela al eje OY.

  Por otro lado, $E_{12}(c)\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x + cy \\ y \end{bmatrix}$, que es una transvección paralela al eje OX.

En definitiva, si $A \in \mathbb{R}^{2\times 2}$ es una matriz invertible (y, por tanto, producto de matrices elementales por filas) y $\begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^{2\times 1}$ es un vector arbitrario entonces el producto $A\begin{bmatrix} x \\ y \end{bmatrix}$ puede ser representado como la aplicación sucesiva de compresiones/elongaciones unidimensionales, reflexiones y/o transvecciones.

5.6. Matriz inversa: método de Gauss-Jordan

En el Teorema 5.1 se probó que una matriz $A \in \mathbb{K}^{n\times n}$ es invertible si y sólo si su forma escalonada reducida por filas R_A es I_n o, equivalentemente $A \sim_f I_n$. También se probó que si A es invertible, existen matrices elementales por filas (que son invertibles) $E_1, E_2, \ldots, E_s \in \mathbb{K}^{n\times n}$ tales que

$$E_s \ldots E_2 E_1 A = I_n.$$

Multiplicando ambos miembros a derecha por A^{-1} se tiene

$$A^{-1} = E_s \ldots E_2 E_1 I_n.$$

Este hecho permite establecer el método general.

Método de Gauss-Jordan para el cálculo de la matriz inversa: Si se multiplica a izquierda por matrices elementales por filas $E_1, E_2, \ldots, E_s$ a una matriz invertible A hasta llegar a I_n, las mismas matrices elementales $E_1, E_2, \ldots, E_s$ premultiplicadas por I_n en el mismo orden, proporcionan la matriz inversa A^{-1}.

De forma resumida, se premultiplica $\left[\begin{array}{c|c} A & I_n \end{array}\right]$ por matrices elementales por filas $E_1, E_2, \ldots, E_s$ hasta conseguir $\left[\begin{array}{c|c} I_n & A^{-1} \end{array}\right]$. En símbolos,

$$\begin{aligned} E_s \ldots E_2 E_1 \left[\begin{array}{c|c} A & I_n \end{array}\right] &= \left[\begin{array}{c|c} E_s \ldots E_2 E_1 A & E_s \ldots E_2 E_1 I_n \end{array}\right] \\ &= \left[\begin{array}{c|c} I_n & A^{-1} \end{array}\right]. \end{aligned}$$

Ejemplo 5.5. *Calcular la inversa de la matriz*

$$A = \begin{bmatrix} 1 & 2 & 0 \\ 2 & 5 & -1 \\ 3 & 5 & 2 \end{bmatrix}$$

utilizando el método de Gauss-Jordan.

◁ Aplicando el método de Gauss-Jordan se tiene

$$\left[\begin{array}{c|c} A & I_3 \end{array}\right] = \left[\begin{array}{ccc|ccc} 1 & 2 & 0 & 1 & 0 & 0 \\ 2 & 5 & -1 & 0 & 1 & 0 \\ 3 & 5 & 2 & 0 & 0 & 1 \end{array}\right] \sim_f \left[\begin{array}{ccc|ccc} 1 & 2 & 0 & 1 & 0 & 0 \\ 0 & 1 & -1 & -2 & 1 & 0 \\ 0 & -1 & 2 & -3 & 0 & 1 \end{array}\right] \sim_f$$

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 2 & 5 & -2 & 0 \\ 0 & 1 & -1 & -2 & 1 & 0 \\ 0 & 0 & 1 & -5 & 1 & 1 \end{array}\right] \sim_f \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 15 & -4 & -2 \\ 0 & 1 & 0 & -7 & 2 & 1 \\ 0 & 0 & 1 & -5 & 1 & 1 \end{array}\right] = \left[\begin{array}{c|c} I_3 & A^{-1} \end{array}\right].$$

Luego,

$$A^{-1} = \begin{bmatrix} 15 & -4 & -2 \\ -7 & 2 & 1 \\ -5 & 1 & 1 \end{bmatrix}.$$

▷

La demostración de (g) ⇒ (i) del Teorema 5.1 proporciona un método de cómo proceder para expresar una matriz invertible como producto de matrices elementales por filas.

Ejemplo 5.6. *Escribir, si es posible, la matriz*

$$A = \begin{bmatrix} 1 & 2 & 0 \\ 2 & 5 & -1 \\ 3 & 5 & 2 \end{bmatrix}$$

y su inversa como producto de matrices elementales por filas.

◁ Del ejemplo anterior se sabe que A es una matriz invertible. Las operaciones elementales por fila del Ejemplo 5.5 se escriben matricialmente como

$$E_{13}(-2)E_{23}(1)E_{32}(1)E_{12}(-2)E_{31}(-3)E_{21}(-2)A = I_3.$$

Multiplicando a derecha por A^{-1} ambos miembros de la igualdad anterior se tiene

$$A^{-1} = E_{13}(-2)E_{23}(1)E_{32}(1)E_{12}(-2)E_{31}(-3)E_{21}(-2),$$

con lo que se ha obtenido A^{-1} escrita como producto de matrices elementales.

Al ser todas las matrices E_{ij} invertibles y sus inversas matrices invertibles del mismo tipo, A se escribe como producto de matrices elementales:

$$\begin{aligned} A &= [E_{21}(-2)]^{-1}[E_{31}(-3)]^{-1}[E_{12}(-2)]^{-1}[E_{32}(1)]^{-1}[E_{23}(1)]^{-1}[E_{13}(-2)]^{-1} \\ &= E_{21}(2)E_{31}(3)E_{12}(2)E_{32}(-1)E_{23}(-1)E_{13}(2). \end{aligned}$$

▷

Observación 5.4. Sea $A \in \mathbb{K}^{n\times n}$. Si al realizar operaciones elementales sobre $\left[\begin{array}{c|c} A & I_n \end{array}\right]$ no se llega a la forma escalonada reducida por filas $R_A = I_n$ en el primer bloque, entonces A no es invertible, pues se ha probado que, de ser A invertible, se debería cumplir que $A \sim_f R_A = I_n$.

Ejemplo 5.7. *Calcular, si existe, la inversa de la matriz*

$$A = \begin{bmatrix} 1 & 2 & 0 \\ 2 & 5 & -1 \\ 4 & 9 & -1 \end{bmatrix}$$

utilizando el método de Gauss-Jordan.

◁ Aplicando el método de Gauss-Jordan se tiene

$$\left[\begin{array}{c|c} A & I_3 \end{array}\right] = \left[\begin{array}{rrr|rrr} 1 & 2 & 0 & 1 & 0 & 0 \\ 2 & 5 & -1 & 0 & 1 & 0 \\ 4 & 9 & -1 & 0 & 0 & 1 \end{array}\right] \sim_f \left[\begin{array}{rrr|rrr} 1 & 2 & 0 & 1 & 0 & 0 \\ 0 & 1 & -1 & -2 & 1 & 0 \\ 0 & 1 & -1 & -4 & 0 & 1 \end{array}\right] \sim_f$$

$$\sim_f \left[\begin{array}{rrr|rrr} 1 & 2 & 0 & 1 & 0 & 0 \\ 0 & 1 & -1 & -2 & 1 & 0 \\ 0 & 0 & 0 & -2 & -1 & 1 \end{array}\right].$$

Al haber obtenido una fila de ceros en el primer bloque, se concluye que la matriz A no es invertible. ▷

Resumiendo: Si al aplicar operaciones elementales por filas se obtiene

$$\begin{bmatrix} A & I_n \end{bmatrix} \sim_f \cdots \sim_f \begin{bmatrix} R_A & P \end{bmatrix},$$

hay dos opciones:

- Si $R_A = I_n$ entonces A es invertible y $A^{-1} = P$.
- Si $R_A \neq I_n$ entonces A no es invertible.

5.7. Inversa de matrices particionadas

Se considera la matriz particionada $M \in \mathbb{K}^{(r+s)\times(r+s)}$ siguiente:

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix}. \tag{5.3}$$

Si $A \in \mathbb{K}^{r\times r}$ es invertible, se llama **complemento de Schur** de A en M a la matriz $S = D - CA^{-1}B$.

Proposición 5.9. Sea $M \in \mathbb{K}^{(r+s)\times(r+s)}$ definida como en (5.3). Si $A \in \mathbb{K}^{r\times r}$ y el complemento de Schur S de A en M son invertibles entonces M es invertible y, además,

$$M^{-1} = \begin{bmatrix} A^{-1} + A^{-1}BS^{-1}CA^{-1} & -A^{-1}BS^{-1} \\ -S^{-1}CA^{-1} & S^{-1} \end{bmatrix}.$$

Esta fórmula para la inversa de una matriz por bloques se conoce como **fórmula de Banachiewicz-Schur**.

Demostración. Por la equivalencia entre (a) y (b) del Teorema 5.1, basta con comprobar que $MM^{-1} = I_{r+s}$. En efecto, teniendo en cuenta que $SS^{-1} = I_s$

se tiene que

$$
\begin{aligned}
MM^{-1} &= \\
&= \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} A^{-1} + A^{-1}BS^{-1}CA^{-1} & -A^{-1}BS^{-1} \\ -S^{-1}CA^{-1} & S^{-1} \end{bmatrix} \\
&= \begin{bmatrix} A(A^{-1} + A^{-1}BS^{-1}CA^{-1}) - BS^{-1}CA^{-1} & -AA^{-1}BS^{-1} + BS^{-1} \\ C(A^{-1} + A^{-1}BS^{-1}CA^{-1}) - DS^{-1}CA^{-1} & -CA^{-1}BS^{-1} + DS^{-1} \end{bmatrix} \\
&= \begin{bmatrix} I_r + BS^{-1}CA^{-1} - BS^{-1}CA^{-1} & -BS^{-1} + BS^{-1} \\ CA^{-1} + CA^{-1}BS^{-1}CA^{-1} - DS^{-1}CA^{-1} & (D - CA^{-1}B)S^{-1} \end{bmatrix} \\
&= \begin{bmatrix} I_r & O_{r\times s} \\ CA^{-1} - (D - CA^{-1}B)S^{-1}CA^{-1} & I_s \end{bmatrix} \\
&= \begin{bmatrix} I_r & O_{r\times s} \\ CA^{-1} - CA^{-1} & I_s \end{bmatrix} \\
&= \begin{bmatrix} I_r & O_{r\times s} \\ O_{s\times r} & I_s \end{bmatrix} \\
&= I_{r+s}.
\end{aligned}
$$

Luego,

$$
M^{-1} = \begin{bmatrix} A^{-1} + A^{-1}BS^{-1}CA^{-1} & -A^{-1}BS^{-1} \\ -S^{-1}CA^{-1} & S^{-1} \end{bmatrix}.
$$

□

Proposición 5.10. Sea $M \in \mathbb{K}^{(r+s)\times(r+s)}$ definida como en (5.3) con $C = O$. Si $A \in \mathbb{K}^{r\times r}$ y $D \in \mathbb{K}^{s\times s}$ son invertibles entonces M es invertible y además

$$
M^{-1} = \begin{bmatrix} A^{-1} & -A^{-1}BD^{-1} \\ O & D^{-1} \end{bmatrix}.
$$

Demostración. Basta particularizar $C = O$ en la Proposición 5.9. □

5.8. EJERCICIOS

(1) Comparar la demostración de la Proposición 5.1 con el Ejercicio 34 de la página 99.

(2) Enunciar y demostrar la **ley del orden inverso generalizada** a un número finito de matrices $A_1, A_2, \ldots, A_s \in \mathbb{K}^{n\times n}$ con $s \geq 2$. Probar que si todas son invertibles entonces su producto lo es y encontrar la fórmula para la inversa de dicho producto.

(3) De la ley del orden inverso deducir la siguiente propiedad: Si $\lambda \in \mathbb{K} - \{0\}$ y $A \in \mathbb{K}^{n\times n}$ es una matriz invertible entonces λA es una matriz invertible y $(\lambda A)^{-1} = \lambda^{-1}A^{-1}$.

(4) Si $A \in \mathbb{K}^{n\times n}$ es invertible y $b \in \mathbb{K}^{n\times 1}$ entonces el sistema de ecuaciones lineales $Ax = b$ tiene una única solución (Observar que se debe probar la existencia y unicidad de la solución).

(5) Sea $A \in \mathbb{K}^{n\times n}$ una matriz invertible. Demostrar que los sistemas matriciales $AX = B$, para cualquier matriz $B \in \mathbb{K}^{n\times m}$, tienen solución única. Encontrar una expresión para dicha solución. ¿Cómo se relaciona este resultado con el del ejercicio anterior?

(6) Si la matriz $A \in \mathbb{R}^{n\times n}$ satisface

$$A^6 - 2A^4 - 9A^2 - 14I_n = O,$$

probar que A es una matriz invertible y expresar su inversa en términos de potencias de A. ¿Cuál es el término en el primer miembro de la igualdad que permite garantizar la existencia de la inversa de la matriz A? ¿Por qué?

(7) Las matrices A y B siguientes corresponden a las matrices ampliadas, en

forma escalonada (por filas), de un sistema de ecuaciones lineales

$$A = \begin{bmatrix} \square & * & * \\ 0 & \square & * \\ 0 & 0 & \square \end{bmatrix} \qquad \text{y} \qquad B = \begin{bmatrix} \square & * & * & * & * & * \\ 0 & 0 & \square & * & * & * \\ 0 & 0 & 0 & \square & * & * \end{bmatrix},$$

donde la notación $\square$ representa un número real diferente de cero arbitrario y con el símbolo $*$ se indica un valor real arbitrario (incluyendo el cero).

(a) Determinar, para cada matriz, su forma escalonada reducida por filas.

(b) En cada caso determinar si el sistema es compatible o no. En caso de serlo, determinar si su solución es única.

(c) Responder las preguntas anteriores si se cambia la matriz B por otra en la que se quita la última columna. Justificar.

(d) ¿Y si se quitan la segunda y la quinta columnas de B? Justificar.

(8) Sean $A \in \mathbb{K}^{r\times r}$ y $B \in \mathbb{K}^{s\times s}$ dos matrices invertibles. Demostrar que

$$\begin{bmatrix} A & O \\ O & B \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} & O \\ O & B^{-1} \end{bmatrix}.$$

Es decir, la inversa de una matriz diagonal por bloques (invertibles) se puede calcular a partir de la inversa de cada bloque.

(9) Se consideran los siguientes sistemas de ecuaciones lineales homogéneos:

$$\begin{cases} x_1 + 5x_2 - 3x_3 = 0 \\ 3x_1 + 6x_2 - x_3 = 0 \\ x_1 - x_2 + 2x_3 = 0 \end{cases} \qquad \begin{cases} x_1 + \frac{4}{5}x_2 = 0 \\ \frac{1}{2}x_1 + \frac{2}{5}x_2 = 0 \end{cases}$$

Se pide:

(a) Resolver cada sistema utlizando operaciones elementales por filas.

(b) ¿Es posible deducir del apartado anterior si la matriz de coeficientes es invertible o no? Justificar adecuadamente.

(c) Aplicar ahora el método de Gauss-Jordan, para decidir si existen las inversas de las matrices de coeficientes de los sistemas anteriores y, en caso afirmativo, calcularlas.

(10) Sea $A \in \mathbb{R}^{n\times n}$ una matriz invertible y $x, y \in \mathbb{R}^{n\times 1}$. Demostrar que la matriz $A + xy^t$ es invertible si y sólo si $1 + y^t A^{-1} x \neq 0$. En este caso,

$$(A + xy^t)^{-1} = A^{-1} - \frac{1}{1 + y^t A^{-1} x} A^{-1} x y^t A^{-1}.$$

Esta expresión se conoce como la **fórmula de Sherman-Morrison**. (Como ayuda observar los siguientes hechos. Cuando $x = 0$, la equivalencia se cumple vacuamente; es decir, el caso interesante es $x \neq 0$. Si A y $A + xy^t$ son invertibles entonces el producto $(A + xy^t)A^{-1}$ también lo es. Además, el producto de matrices $y^t A^{-1} x$ es un número real, en realidad, una matriz de tamaño 1×1.)

(11) Demostrar que toda matriz triangular superior (respectivamente, inferior) $A \in \mathbb{K}^{n\times n}$ con elementos no nulos en la diagonal es invertible, su inversa es triangular superior (respectivamente, inferior) y en la diagonal de la inversa aparecen los inversos multiplicativos de los elementos de la diagonal de A en las respectivas posiciones. (Ayuda: Proceder por inducción y utilizar la Proposición 5.10).

(12) Sea la matriz

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 1/3 & -2 & 0 \\ 5 & -4 & 3 \end{bmatrix}.$$

(a) ¿Tiene la matriz A alguna estructura especial?

(b) Usando la estructura indicada en el apartado (a), hallar A^{-1}.

(c) Comparar con el cálculo de la inversa mediante la fórmula conocida

$$A^{-1} = \frac{1}{\det(A)}\mathrm{Adj}(A^t).$$

(13) Se consideran las matrices

$$M = \begin{bmatrix} I_r & O_{r\times s} \\ C & D \end{bmatrix} \qquad \text{y} \qquad Q = \begin{bmatrix} Q_1 & O \\ Q_2 & Q_3 \end{bmatrix},$$

siendo $D \in \mathbb{R}^{s\times s}$ una matriz invertible. Encontrar los tamaños adecuados de los bloques Q_i, $i = 1, 2, 3$ y calcular el producto $M^{-1}Q$ justificando previamente que M es invertible.

(14) Visualizar en qué se transforma el cuadrado de lado 1 cuyos vértices se hallan en el origen del plano y sobre los ejes coordenados, utilizando las diferentes matrices elementales por filas y considerando los valores de $k = \pm 2$ y $c = \pm 2$ en la interpretación geométrica de la página 227.

(15) Escribir, si es posible, la matriz

$$A = \begin{bmatrix} 1 & 4 \\ 1 & 5 \end{bmatrix}$$

y su inversa como producto de matrices elementales.

(16) Comprobar que existen inversas a izquierda de la matriz

$$A = \begin{bmatrix} 1 & 2 \\ 1 & 1 \\ 3 & 0 \end{bmatrix} \in \mathbb{R}^{3\times 2}$$

y calcularlas. ¿Es posible garantizar que la existencia de inversas a izquierda implican la existencia de la inversa de A? Justificar, analizando si se contradicen las equivalencias (a) y (b) del Teorema de caracterizaciones de matriz invertible.

Capítulo 6

Equivalencia de matrices

Índice

6.1. Introducción

El matemático francés Charles Hermite (1822-1901) estudió la reducción de matrices utilizando operaciones elementales por filas, por lo que la forma escalonada reducida suele llamarse también forma normal de Hermite en su honor. En realidad la forma que encontró difiere de la presentada en este libro, pero se suele atribuir a Hermite la forma escalonada reducida presentada en este capítulo en consideración a su trabajo en este tema.

Figura 6.1: Charles Hermite.

Uno de los principales hechos que se utilizará en este capítulo es la forma de escribir en notación matricial las operaciones elementales efectuadas sobre una matriz; lo que se ha realizado a través de las llamadas matrices elementales.

El principal objetivo de este capítulo es encontrar nuevas caracterizaciones de la equivalencia por filas y por columnas, se introducirá la equivalencia (bilateral) de matrices, donde estará permitido realizar tanto operaciones elementales por filas como por columnas para reducir una matriz a la forma "más sencilla" posible. Una caracterización de la equivalencia de matrices viene dada a partir del rango y este ente matemático se destacará como un invariante[1] de la teoría de equivalencia de matrices. En definitiva, se analiza la forma canónica a la que se puede llevar cualquier matriz, llamada forma normal de Hermite.

6.2. Equivalencia de matrices por filas y por columnas

Hasta ahora se ha utilizado la equivalencia de matrices por filas (o por columnas) básicamente con el objetivo de hallar la forma escalonada reducida por filas (o por columnas). En lo que queda del capítulo, se estudiará la equivalencia de matrices como relación de equivalencia estableciendo la relación con un invariante fundamental en la Teoría de Matrices y, en el Álgebra Lineal en general, como es el rango de una matriz.

6.2.1. Equivalencia de matrices por filas

En el Capítulo 4, se introdujo en concepto de matrices equivalentes por filas. Se recuerda que:

Definición 6.1. Sean $A, B \in \mathbb{K}^{m\times n}$. Se dice que A **es equivalente por filas** a B, y se denota $A \sim_f B$, si B se obtiene de aplicar un número finito de operaciones elementales por filas a A.

Es natural la pregunta: ¿Es la relación binaria $\sim_f$ una relación de equivalencia sobre $\mathbb{K}^{m\times n}$? La respuesta es afirmativa y se prueba en el siguiente resultado.

Proposición 6.1. La relación binaria $\sim_f$ es una relación de equivalencia sobre el conjunto $\mathbb{K}^{m\times n}$.

[1] En realidad, determinará un conjunto completo de invariantes.

Demostración. En efecto, sean $A, B, C \in \mathbb{K}^{m\times n}$. Aplicando el Teorema 3.1 (y su Corolario 3.1) es fácil ver que:

- Reflexiva: $A \sim_f A$. No hace falta aplicar operación elemental alguna.
- Simétrica: $A \sim_f B \Rightarrow B \sim_f A$. Si se aplican un número finito de operaciones elementales por filas para ir de A a B, al ser dichas operaciones reversibles, estas operaciones inversas se pueden aplicar (en orden inverso al anterior) para ir de B a A. Por lo tanto, $B \sim_f A$.
- Transitiva: $A \sim_f B \ \wedge \ B \sim_f C \Rightarrow A \sim_f C$. Si se aplican un número finito de operaciones elementales por filas para ir de A a B, y otro número finito para ir de B a C, aplicando todas sucesivamente (en el mismo orden), se puede ir de A a C. Luego, $A \sim_f C$.

□

Para estudiar la partición que produce dicha relación de equivalencia sobre $\mathbb{K}^{m\times n}$ (véase Teorema 1.1), a continuación se presenta una caracterización de la equivalencia de matrices por filas.

Teorema 6.1 (Caracterización de matrices equivalentes por filas)**.** Sean $A, B \in \mathbb{K}^{m\times n}$. Entonces las siguientes condiciones son equivalentes:

(a) $A \sim_f B$.

(b) Existen matrices elementales por filas $E_1, E_2, \ldots, E_s \in \mathbb{K}^{m\times m}$ tales que $E_s \ldots E_2 E_1 A = B$.

(c) Existe una matriz invertible $Q \in \mathbb{K}^{m\times m}$ tal que $QA = B$.

(d) $R_A = R_B$ (es decir, A y B tienen la misma forma escalonada reducida por filas).

Demostración. (a) $\Leftrightarrow$ (b) Se deduce de aplicar la Proposición 5.7, de forma reiterada, un número finito de veces.

(b) $\Rightarrow$ (c) Si se cumple (b), tomando $Q := E_s \dots E_2E_1$, el resultado sigue recordando dos hechos: que las matrices elementales son invertibles (Proposición 5.5) y que el producto de matrices invertibles del mismo tamaño es invertible (Ejercicio 2 de página 234).

(c) $\Rightarrow$ (b) Sea $Q \in \mathbb{K}^{m\times m}$ una matriz invertible tal que $QA = B$. Como Q es invertible, por el Teorema 5.1, Q es producto de matrices elementales por filas, con lo que (b) sigue de forma directa.

(a) $\Rightarrow$ (d) Por el Teorema 4.2, $A \sim_f R_A$ y $B \sim_f R_B$, siendo R_A y R_B las respectivas formas escalonadas reducidas por filas (que son únicas).

Si $A \sim_f B$, al ser $\sim_f$ una relación de equivalencia, debe ser $R_A \sim_f R_B$ (puesto que $R_A \sim_f A$, $A \sim_f B$ y $B \sim_f R_B$). Como R_A y R_B son matrices escalonadas reducidas por filas y además son equivalentes por filas, de nuevo el Teorema 4.2 asegura que $R_A = R_B$.

(d) $\Rightarrow$ (a) Si $R_A = R_B$, al ser $A \sim_f R_A$ y $B \sim_f R_B$ (Teorema 4.2) y, además, $\sim_f$ una relación de equivalencia (Proposición 6.1), se tiene que $A \sim_f R_A = R_B \sim_f B$, por tanto, $A \sim_f B$. $\square$

Por lo tanto, la clase de equivalencia por la relación $\sim_f$ de una matriz fija $A \in \mathbb{K}^{m\times n}$, denotada C_A, es:

$$C_A = \{QA : Q \in \mathbb{K}^{m\times m} \text{ es invertible}\}.$$

Ejemplo 6.1. *Las cinco matrices del Ejemplo 4.6, desde A hasta su forma escalonada reducida por filas (con ella misma incluida), son matrices equivalentes por filas a A. Luego, todas ellas pertenecen a su clase de equivalencia por la relación $\sim_f$ y cada una es de la forma QA para alguna matriz invertible $Q \in \mathbb{R}^{3\times 3}$.*

Ejemplo 6.2. *Encontrar una matriz invertible Q tal que $QA = R_A$ donde*

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

◁ Es fácil probar que

$$\text{Si} \qquad Q\begin{bmatrix} A & I_4 \end{bmatrix} = \begin{bmatrix} R_A & Z \end{bmatrix} \qquad \Rightarrow \qquad QA = RA \quad \text{y} \quad Q = Z$$

y este cálculo proporciona un método para hallar la matriz invertible Q tal que $QA = R_A$.

Efectuando las mismas operaciones elementales por filas sobre la matriz A e I_4 a la vez se tiene

$$\begin{bmatrix} A & I_4 \end{bmatrix} = \left[\begin{array}{cc|cccc} 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 2 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right] \sim_f \left[\begin{array}{cc|cccc} 1 & 0 & 1 & -1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right].$$

Luego,

$$Q = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & -2 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad \text{y} \qquad R_A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

cumplen que $QA = R_A$. ▷

6.2.2. Equivalencia de matrices por columnas

De forma análoga a la equivalencia por filas, se define el concepto de equivalencia de matrices por columnas.

Definición 6.2. Sean $A, B \in \mathbb{K}^{m\times n}$. Se dice que A **es equivalente por columna** a B, y se denota $A \sim_c B$, si B se obtiene de aplicar un número finito de operaciones elementales por columnas a A.

Es posible establecer un resultado similar al de la Proposición 6.1.

Proposición 6.2. La relación binaria $\sim_c$ es una relación de equivalencia sobre el conjunto $\mathbb{K}^{m\times n}$.

Demostración. La demostración es similar a la de la Proposición 6.1 y se propone como ejercicio. □

A continuación se presenta un resultado dual al Teorema 6.1 que caracteriza las matrices equivalentes por columnas.

Teorema 6.2 (Caracterización de matrices equivalentes por columnas)**.** Sean $A, B \in \mathbb{K}^{m\times n}$. Entonces las siguientes condiciones son equivalentes:

(a) $A \sim_c B$.

(b) Existen matrices elementales por columnas $F_1, F_2, \ldots, F_\ell \in \mathbb{K}^{n\times n}$ tales que $AF_1F_2 \ldots F_\ell = B$.

(c) Existe una matriz invertible $P \in \mathbb{K}^{n\times n}$ tal que $AP = B$.

(d) $R_{A^t} = R_{B^t}$ (es decir, A^t y B^t tienen la misma forma escalonada reducida por filas).

Demostración. Es fácil comprobar (véase Ejercicio (4) de este capítulo) que

$$A \sim_c B \qquad \Leftrightarrow \qquad A^t \sim_f B^t.$$

Ahora se aplica el Teorema 6.1 y se toman traspuestas para obtener (b) y (c). Los detalles se proponen como ejercicio. □

Observación 6.1. Para obtener una mayor simetría (como el Teorema 6.1), el apartado (d) del Teorema 6.2 se puede sustituir por la condición:

> (d') A y B tienen la misma forma escalonada reducida por columnas (es decir, ${}_AR = {}_BR$).

Sin embargo, el enunciado (d) suele resultar más sencillo de comprobar en la práctica.

Para demostrar la equivalencia entre (d) y (d') basta probar la equivalencia entre (a) y (d'). En efecto, dado que *se puede probar* que toda matriz es equivalente por columnas a una única matriz escalonada reducida por columnas, si $A \sim_c B$ entonces ${}_AR \sim_c A \sim_c B \sim_c {}_BR$, con lo que ${}_AR = {}_BR$. Recíprocamente, si ${}_AR = {}_BR$ entonces $A \sim_c {}_AR = {}_BR \sim_c B$. Al ser $\sim_c$ una relación de equivalencia, la transitividad implica que $A \sim_c B$.

Por lo tanto, la clase de equivalencia por la relación $\sim_c$, de una matriz fija $A \in \mathbb{K}^{m\times n}$, denotada ${}_AC$, es:

$${}_AC = \{AP : P \in \mathbb{K}^{n\times n} \text{ es invertible}\}.$$

Ejemplo 6.3. *Encontrar una matriz invertible P tal que $AP = {}_AR$ donde*

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

$\triangleleft$ Efectuando operaciones elementales por columnas sobre la matriz A se tiene

$$\left[\begin{array}{c} A \\ \hline I_3 \end{array}\right] = \left[\begin{array}{ccc} 1 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \\ \hline 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right] \sim_c \left[\begin{array}{ccc} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \\ \hline 1 & -1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right] \sim_c \left[\begin{array}{ccc} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ \hline 1 & -1 & -1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{array}\right] = \left[\begin{array}{c} {}_AR \\ \hline P \end{array}\right].$$

Luego, la matriz intermedia entre A y su forma escalonada reducida por columnas ${}_AR = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ es equivalente por columnas a A y a ${}_AR$; y todas ellas pertenecen a ${}_AC$. Además (véase el Ejercicio 3 de este capítulo), se tiene que

$$P := \begin{bmatrix} 1 & -1 & -1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \qquad \text{es tal que} \qquad AP = {}_AR.$$

$\triangleright$

Observación 6.2. Notar que en el ejemplo anterior se han hecho operaciones elementales por columnas partiendo de la matriz por bloques $\begin{bmatrix} A \\ I_3 \end{bmatrix}$. Volviendo al Ejemplo 5.5, se observa que allí un procedimiento similar se ha aplicado por filas a $\begin{bmatrix} A & I_3 \end{bmatrix}$; en aquel caso se ha obtenido una matriz $Q := \begin{bmatrix} 15 & -4 & -2 \\ -7 & 2 & 1 \\ -5 & 1 & 1 \end{bmatrix}$ tal que $QA = R_A = I_3$.

6.3. Equivalencia de matrices

Si se permite realizar operaciones elementales por filas y por columnas sobre una matriz A se consigue una relación de equivalencia cuya forma escalonada reducida es especialmente sencilla.

Definición 6.3. Sean $A, B \in \mathbb{K}^{m\times n}$. Se dice que A **es equivalente**[a] a B, y se denota $A \sim B$, si B se obtiene de aplicar un número finito de operaciones elementales por filas y por columnas a A.

[a]Cuando no se especifica nada más que equivalente, significa que lo es por filas y por columnas utilizando ambos tipos de operaciones a la vez.

De los resultados probados para filas y para columnas se tiene de manera inmediata el siguiente resultado.

Proposición 6.3. La relación binaria $\sim$ es una relación de equivalencia sobre el conjunto $\mathbb{K}^{m\times n}$.

Demostración. Se propone como ejercicio. □

Proposición 6.4 (Equivalencia por filas o columnas implica equivalencia)**.** Sean $A, B \in \mathbb{K}^{m\times n}$. Si $A \sim_f B$ ó $A \sim_c B$ entonces $A \sim B$.

Demostración. Se propone como ejercicio. □

Observación 6.3. Una pregunta natural es si la implicación recíproca de la proposición anterior es cierta. La respuesta es negativa; es decir, si se cumple $A \sim B$, no se puede deducir ni $A \sim_f B$ ni $A \sim_c B$.

A continuación de presenta una caracterización de la equivalencia de matrices.

Teorema 6.3 (Caracterización de matrices equivalentes)**.** Sean $A, B \in \mathbb{K}^{m\times n}$. Entonces las siguientes condiciones son equivalentes:

(a) $A \sim B$.

(b) Existen matrices invertibles $Q \in \mathbb{K}^{m\times m}$ y $P \in \mathbb{K}^{n\times n}$ tales que

$$QAP = B.$$

Demostración. (a) $\Rightarrow$ (b) Aplicando las Proposiciones 5.7 y 5.8 un número finito de veces cada una, es posible asegurar que existen matrices elementales por filas $E_1, E_2, \dots, E_s \in \mathbb{K}^{m\times m}$ y por columnas $F_1, F_2, \dots, F_\ell \in \mathbb{K}^{n\times n}$ tales que $E_s \dots E_2E_1AF_1F_2 \dots F_\ell = B$. Llamando $Q := E_s \dots E_2E_1$ y $P := F_1F_2 \dots F_\ell$ se tiene que ambas son invertibles (por ser producto de invertibles) y $QAP = B$.

(b) $\Rightarrow$ (a) Si $QAP = B$ para ciertas matrices invertibles $Q \in \mathbb{K}^{m\times m}$ y $P \in \mathbb{K}^{n\times n}$, puesto que Q y P^t son producto de matrices elementales por filas (Teorema 5.1), P será producto de matrices elementales por columnas y, de la Definición 6.3, se tiene que $A \sim B$. □

Por lo tanto, la clase de equivalencia por la relación $\sim$, de una matriz fija $A \in \mathbb{K}^{m\times n}$, es el conjunto:

$$\{QAP : Q \in \mathbb{K}^{m\times m} \ \wedge \ P \in \mathbb{K}^{n\times n} \text{ son invertibles}\}.$$

Ahora se trata de buscar la forma más sencilla que puede tomar QAP variando las matrices invertibles $Q \in \mathbb{K}^{m\times m}$ y $P \in \mathbb{K}^{n\times n}$.

Combinando los procedimientos realizados anteriormente en los Ejemplos 6.2 y 6.3 se pueden encontrar matrices invertibles P y Q para que una matriz A sea equivalente a otra matriz B. El siguiente ejemplo lo ilustra.

Ejemplo 6.4. *Hallar matrices invertibles Q y P tales que la matriz QAP presente el "aspecto más simplificado posible" (con unos en la diagonal y ceros en el resto) donde*

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

◁ Efectuando las mismas operaciones elementales por filas sobre la matriz $\left[\begin{array}{c|c} A & I_4 \end{array}\right]$ y por columnas sobre la matriz $\left[\begin{array}{c} A \\ \hline I_3 \end{array}\right]$ se tiene[2]

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

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

[2]Los lugares vacíos se deberían rellenar con ceros para tener una matriz 7×7. Como al operar seguirán siendo ceros, al no ponerlos queda más claro el resultado que se busca.

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

De acuerdo al Ejercicio (8) de la página 262, se tiene que $QAP = R$, siendo

$$Q = \begin{bmatrix} 3 & -2 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}, \ P = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}, \ R = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

$\triangleright$

En el ejemplo anterior se observa que la forma obtenida para la matriz R está determinada por una identidad (I_2 en este caso) y bloques de matrices nulas a su derecha y en filas inferiores, es decir

$$A \sim \left[\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \hline 0 & 0 & 0 \\ 0 & 0 & 0 \end{array}\right].$$

Si se compara con el Ejemplo 5.5, donde se han realizado únicamente operaciones elementales por filas, se observa que también se llegó a la matriz identidad (sin el añadido de bloques nulos en aquel caso) y se puede ver que $QA = I_3$, que también se puede escribir como $QAI_3 = I_3$. Luego, $A \sim I_3$. En ambos casos, son las formas equivalentes más sencillas a las que se pueden llevar las matrices A consideradas. Estas observaciones pueden generalizarse, como se hará en el Teorema 6.4.

Antes, se proporciona el siguiente resultado que es de gran utilidad. Establece que la multiplicación por matrices invertibles (a derecha y/o a izquierda) no cambia el rango de una matriz.

Lema 6.1 (Invariancia del rango al multiplicar por matrices invertibles)**.** Sea $A \in \mathbb{K}^{m\times n}$ y sean $Q \in \mathbb{K}^{m\times m}$ y $P \in \mathbb{K}^{n\times n}$ dos matrices invertibles. Entonces se cumple la invariancia del rango para

matrices equivalentes por filas: $\operatorname{rg}(QA) = \operatorname{rg}(A)$.

matrices equivalentes por columnas: $\operatorname{rg}(AP) = \operatorname{rg}(A)$.

matrices equivalentes: $\operatorname{rg}(QAP) = \operatorname{rg}(A)$.

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

Para probar la invariancia del rango para matrices equivalentes por filas, se utilizan las equivalencias entre (a) y (c) del Teorema 6.1 para obtener que $A \sim_f QA$ y las equivalencias (a) y (d) del mismo teorema aseguran que $R_A = R_{QA}$. Por definición de rango, $\operatorname{rg}(A) = \operatorname{rg}(QA)$.

Ahora se prueba la invariancia del rango para matrices equivalentes por columnas. En efecto, es claro que el resultado se cumple si $A = O$. Sea ahora $A \neq O$, luego $\operatorname{rg}(A) = r > 0$. Si se lleva A a su forma escalonada reducida por filas mediante las matrices elementales $E_1, E_2, \ldots, E_s$ se tiene que $E_s \ldots E_2E_1A = R_A$, donde las primeras r filas de R_A son no nulas y las últimas $m - r$ son nulas. Luego,

$$AP \sim_f E_s \ldots E_2E_1(AP) = (E_s \ldots E_2E_1A)P = R_AP.$$

La Observación 2.10 asegura que en el producto R_AP también aparecerán las $m - r$ últimas filas nulas, con lo que $\operatorname{rg}(R_AP)$ no puede exceder a r. Así, $s := \operatorname{rg}(AP) \leq r = \operatorname{rg}(A)$. El mismo razonamiento permite garantizar que, como $\operatorname{rg}(AP) = s$ y P es invertible, entonces

$$r = \operatorname{rg}(A) = \operatorname{rg}((AP)P^{-1}) \leq \operatorname{rg}(AP) = s.$$

Se sigue que $s \leq r$ y $r \leq s$, es decir, $r = s$. Por lo tanto, $\operatorname{rg}(AP) = \operatorname{rg}(A)$.

Finalmente, la invariancia del rango para matrices equivalentes se deduce de los casos anteriores como sigue:

$$\operatorname{rg}(QAP) = \operatorname{rg}(Q(AP)) = \operatorname{rg}(AP) = \operatorname{rg}(A),$$

usando primero que Q es invertible y luego que P lo es. $\square$

6.3.1. Forma escalonada reducida

Es conocido que la matriz nula $O_{m\times n}$ tiene rango 0 y es la única en esta situación (es decir, si $\mathrm{rg}(A) = 0$ entonces $A = O$). En definitiva,

$$\mathrm{rg}(A) = 0 \qquad \Leftrightarrow \qquad A = O.$$

Teorema 6.4. Sea $A \in \mathbb{K}^{m\times n}$ una matriz no nula. Las siguientes condiciones son equivalentes:

(a) $\mathrm{rg}(A) = r$.

(b) $A \sim \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$.

Demostración. (a) $\Rightarrow$ (b) Si $\mathrm{rg}(A) = r$ entonces la forma escalonada reducida por filas R_A de A tiene r filas no nulas, y también r columnas básicas (con pivote igual a 1 y los restantes elementos de la columna que los contiene iguales a cero). Como $A \sim_f R_A$, se tiene que $QA = R_A$ para alguna matriz invertible $Q \in \mathbb{K}^{m\times m}$. Multiplicando R_A a derecha por matrices elementales de permutación, es posible reunir todas las columnas básicas en las primeras r columnas, es decir, existe una matriz invertible $P_1 \in \mathbb{K}^{n\times n}$ que cumple

$$QAP_1 = R_A P_1 = \begin{bmatrix} I_r & K \\ 0 & 0 \end{bmatrix},$$

donde P_1 es el producto de las matrices de permutación necesarias y la matriz $K \in \mathbb{K}^{r\times(n-r)}$ contiene los restantes elementos de las primeras r filas de las columnas no básicas de R_A. Si se considera la matriz invertible

$$P_2 := \begin{bmatrix} I_r & -K \\ 0 & I_{n-r} \end{bmatrix}$$

entonces

$$QAP_1P_2 = \begin{bmatrix} I_r & K \\ 0 & 0 \end{bmatrix} \begin{bmatrix} I_r & -K \\ 0 & I_{n-r} \end{bmatrix} = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}.$$

Puesto que $P := P_1P_2$ y Q son matrices invertibles y satisfacen

$$QAP = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix},$$

del Teorema 6.3 se tiene que $A \sim \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$.

(b) $\Rightarrow$ (a) Si $A \sim \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$ entonces $QAP = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$, para ciertas matrices invertibles $Q \in \mathbb{K}^{m\times m}$ y $P \in \mathbb{K}^{n\times n}$. Por el Lema 6.1,

$$\operatorname{rg}(A) = \operatorname{rg}(QAP) = \operatorname{rg}\left(\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}\right) = r,$$

por tratarse esta última de un matriz escalonada reducida por filas.

□

Definición 6.4. Sea $A \in \mathbb{K}^{m\times n}$ con $\operatorname{rg}(A) = r > 0$. La matriz

$$FNH_r := \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix},$$

tal que $A \sim \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$ se llama **forma escalonada reducida** de A, **forma normal de Hermite** de A o **forma normal de rango** de A. Por convención, la **forma normal de Hermite** de la matriz nula es ella misma.

Ejemplo 6.5. *En el Ejemplo 6.4 se ha calculado la forma normal de Hermite* $FNH_2 = \left[\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \hline 0 & 0 & 0 \\ 0 & 0 & 0 \end{array}\right]$ *de la matriz A y las matrices Q y P que la llevan a dicha forma, además* $\text{rg}(A) = 2$.

El siguiente resultado proporciona un criterio práctico para decidir si dos matrices son equivalentes por filas, por columnas o equivalentes (bilateralmente).

Teorema 6.5 (Caracterización de cada tipo de equivalencia)**.** Sean $A, B \in \mathbb{K}^{m\times n}$ matrices no nulas. Entonces

(a) $A \sim_f B \;\Leftrightarrow\; R_A = R_B$.

(b) $A \sim_c B \;\Leftrightarrow\; R_{A^t} = R_{B^t}$.

(c) $A \sim B \;\Leftrightarrow\; \text{rg}(A) = \text{rg}(B)$.

Demostración. Los apartados (a) y (b) se han probado en los Teoremas 6.1 y 6.2, respectivamente.

(c) Sean $\text{rg}(A) = r$ y $\text{rg}(B) = s$. El Teorema 6.4 asegura que

$$A \sim \begin{bmatrix} I_r & O \\ O & O \end{bmatrix} \qquad \text{y} \qquad B \sim \begin{bmatrix} I_s & O \\ O & O \end{bmatrix}. \tag{6.1}$$

($\Rightarrow$) Si $A \sim B$ entonces, por (6.1) y dado que $\sim$ es relación de equivalencia,

$$\begin{bmatrix} I_r & O \\ O & O \end{bmatrix} \sim \begin{bmatrix} I_s & O \\ O & O \end{bmatrix}.$$

Por el Teorema 6.3, existen matrices invertibles $Q \in \mathbb{K}^{m\times m}$ y $P \in \mathbb{K}^{n\times n}$ tales que

$$Q \begin{bmatrix} I_r & O \\ O & O \end{bmatrix} P = \begin{bmatrix} I_s & O \\ O & O \end{bmatrix}. \tag{6.2}$$

Por la invariancia del rango por matrices invertibles, la igualdad de (6.2) lleva a $\mathrm{rg}(A) = \mathrm{rg}(B)$ pues

$$r = \mathrm{rg}\left(\begin{bmatrix} I_r & O \\ O & O \end{bmatrix}\right) = \mathrm{rg}\left(Q \begin{bmatrix} I_r & O \\ O & O \end{bmatrix} P\right) = \mathrm{rg}\left(\begin{bmatrix} I_s & O \\ O & O \end{bmatrix}\right) = s.$$

($\Leftarrow$) Si $r = s$, de (6.1) y teniendo en cuenta que $\sim$ es una relación de equivalencia, es claro que $A \sim \begin{bmatrix} I_r & O \\ O & O \end{bmatrix} \sim B$. □

En definitiva, la clase de equivalencia por la relación $\sim$, de una matriz fija $A \in \mathbb{K}^{m\times n}$ de rango r, es:

$$\begin{aligned} C_{FNH_r} &:= \{QAP \in \mathbb{K}^{m\times n} : Q \in \mathbb{K}^{m\times m} \wedge P \in \mathbb{K}^{n\times n} \text{ son invertibles}\} \\ &= \{B \in \mathbb{K}^{m\times n} :\ \mathrm{rg}(B) = r\}. \end{aligned}$$

Dado que el **conjunto cociente** $\mathbb{K}^{m\times n}/\!\sim$ es el conjunto formado por todas las clases de equivalencia (distintas) por la relación $\sim$, se tiene que

$$\mathbb{K}^{m\times n}/\!\sim\, = \{C_{FNH_r} : r \in \{0, 1, 2, \ldots, \text{mín}\{m, n\}\}\}.$$

Ahora bien, el propio conjunto $\mathbb{K}^{m\times n}$ se puede escribir como la unión (disjunta) de las clases de equivalencia distintas como

$$\begin{aligned} \mathbb{K}^{m\times n} &= \cup_{r=0}^{\text{mín}\{m,n\}} C_{FNH_r} \\ &= \cup_{r=0}^{\text{mín}\{m,n\}} C_{\begin{bmatrix} I_r & O \\ O & O_{(m-r)\times(n-r)} \end{bmatrix}}. \end{aligned}$$

Como $C_{FNH_r} \neq C_{FNH_s} \Leftrightarrow C_{FNH_r} \cap C_{FNH_s} = \emptyset \Leftrightarrow r \neq s$, hay $1 + \text{mín}\{m, n\}$ clases de equivalencia distintas por la relación $\sim$. Por último, un **conjunto completo de representantes** viene dado por

$$\left\{ \begin{bmatrix} I_r & O \\ O & O \end{bmatrix} \in \mathbb{K}^{m\times n} : r = 0, 1, 2, \ldots, \text{mín}\{m, n\} \right\},$$

donde $r = 0$ corresponde a la matriz nula[3].

6.3.2. Más propiedades del rango

Hay dos propiedades sobre el rango que cerrarán el capítulo y que son importantes a la hora de trabajar con matrices. Una está relacionada con el rango de la traspuesta de una matriz y la otra con el rango de un producto de matrices.

Proposición 6.5 (Rango de la matriz traspuesta de una matriz)**.** Sea $A \in \mathbb{K}^{m\times n}$. Entonces

$$\mathrm{rg}(A^t) = \mathrm{rg}(A).$$

Demostración. El resultado es evidente para $A = O$: $\mathrm{rg}(A^t) = \mathrm{rg}(A) = 0$.

Sea $A \neq O$ una matriz tal que $\mathrm{rg}(A) = r > 0$. El Teorema 6.4 asegura que $A \sim \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$. Por definición, existen matrices invertibles $Q \in \mathbb{K}^{m\times m}$ y $P \in \mathbb{K}^{n\times n}$ tales que

$$QAP = \begin{bmatrix} I_r & 0_{r\times(n-r)} \\ 0_{(m-r)\times r} & 0_{(m-r)\times(n-r)} \end{bmatrix}.$$

Luego,

$$A = Q^{-1} \begin{bmatrix} I_r & 0_{r\times(n-r)} \\ 0_{(m-r)\times r} & 0_{(m-r)\times(n-r)} \end{bmatrix} P^{-1},$$

y, por tanto,

$$A^t = (P^{-1})^t \begin{bmatrix} I_r & 0_{r\times(m-r)} \\ 0_{(n-r)\times r} & 0_{(n-r)\times(m-r)} \end{bmatrix} (Q^{-1})^t.$$

Como $(P^{-1})^t$ y $(Q^{-1})^t$ son matrices invertibles, se tiene que

$$A^t \sim \begin{bmatrix} I_r & 0_{r\times(m-r)} \\ 0_{(n-r)\times r} & 0_{(n-r)\times(m-r)} \end{bmatrix}.$$

[3]En ese caso, la matriz identidad I_r estaría ausente.

Por el Teorema 6.4, $\mathrm{rg}(A^t) = r$. Esto prueba que $\mathrm{rg}(A^t) = \mathrm{rg}(A)$. □

Proposición 6.6 (Cotas del rango de un producto de matrices)**.** Sean $A \in \mathbb{K}^{m\times n}$ y $B \in \mathbb{K}^{n\times p}$. Entonces

$$\mathrm{rg}(AB) \leq \text{mín}\{\mathrm{rg}(A), \mathrm{rg}(B)\}.$$

Demostración. El resultado es claro si $A = O$:

$$\mathrm{rg}(AB) = 0 \leq 0 = \text{mín}\{\mathrm{rg}(A), \mathrm{rg}(B)\}.$$

Se considera entonces $A \neq O$ tal que $\mathrm{rg}(A) = r > 0$. Como $A \sim_f R_A$, existe una matriz invertible $Q \in \mathbb{K}^{m\times m}$ tal que

$$QA = R_A.$$

Luego, $QAB = R_AB$, con lo que $AB \sim_f R_AB$. Entonces ambas matrices tendrán la misma forma escalonada reducida por filas, lo que garantiza que

$$\mathrm{rg}(AB) = \mathrm{rg}(R_AB).$$

Como R_A tiene sus últimas $m - r$ filas nulas, la Observación 2.10 asegura que estas filas nulas se replican en R_AB, con lo que $\mathrm{rg}(R_AB) \leq r = \mathrm{rg}(A)$. Así, se ha probado que

$$\mathrm{rg}(AB) \leq \mathrm{rg}(A), \tag{6.3}$$

es decir, el rango de un producto no excede el rango del primero de los factores.

Ahora, aplicando la Proposición 6.5 dos veces y el resultado que se acaba de probar se tiene:

$$\mathrm{rg}(AB) = \mathrm{rg}[(AB)^t] = \mathrm{rg}(B^tA^t) \leq \mathrm{rg}(B^t) = \mathrm{rg}(B), \tag{6.4}$$

con lo que el rango de un producto no excede el del segundo factor. Reuniendo los resultados de (6.3) y (6.4) se tiene que $\mathrm{rg}(AB) \leq \text{mín}\{\mathrm{rg}(A), \mathrm{rg}(B)\}$. □

Resumen: La siguiente definición formaliza el concepto de invariante mencionado en la introducción.

Definición 6.5. Sea $\sim$ una relación de equivalencia definida sobre un conjunto S no vacío. Una aplicación f definida sobre S (en un cierto conjunto, generalmente numérico) se llama **invariante** para la relación de equivalencia $\sim$ si se cumple que

$$a \sim b \qquad \Rightarrow \qquad f(a) = f(b),$$

es decir, f toma el mismo valor sobre todos los elementos equivalentes. Una colección de invariantes $f_1, f_2, \ldots, f_s$ se llama un **conjunto completo de invariantes** para la relación de equivalencia $\sim$ si se cumple que

$$a \sim b \qquad \Leftrightarrow \qquad f_1(a) = f_1(b),\ f_2(a) = f_2(b),\ \ldots,\ f_s(a) = f_s(b).$$

De la definición se desprende que para una relación de equivalencia $\sim$ sobre S:

- Si f es un invariante para $\sim$ y $f(a) \neq f(b)$ entonces $a \nsim b$, es decir, permite detectar elementos que no son equivalentes.
- Un conjunto completo de invariantes permite caracterizar la equivalencia de dos elementos a y b de S con tan solo comprobar la igualdad de las aplicaciones f_i, $i = 1, 2, \ldots, s$, sobre ellos.

A modo de resumen se presenta la Tabla 6.1, donde se ha denotado por:

- $FERF$: la forma escalonada reducida por filas de una matriz.
- $FERC$: la forma escalonada reducida por columnas de una matriz.

En la fila de equivalencia, la palabra **rango** está resaltada con negrita por tratarse (no sólo de un invariante sino) de un conjunto completo de invariantes (Teorema 6.5 (c)). Mientras tanto, en la equivalencia por filas y por

Relación binaria	Tamaño	Invariante	Formas canónicas
Equivalencia por filas	$m \times n$	rango	$FERF$
Equivalencia por columnas	$m \times n$	rango	$FERC$
Equivalencia	$m \times n$	**rango**	FNH_r

Tabla 6.1: Tipos de equivalencias de matrices.

columnas no lo está por ser sólo un invariante ($A \sim_f B$ o bien $A \sim_c B$ implican $A \sim B$ y aplicando el Teorema 6.5 (c) se llega, en ambos casos, a $\mathrm{rg}(A) = \mathrm{rg}(B)$) pero no un conjunto completo de invariantes.

Debe quedar claro que mientras que, a todas las matrices de $\mathbb{K}^{m\times n}$ que tengan el mismo rango les corresponde una única FNH_r (y, por tanto, hay un número finito de clases de equivalencia), en el caso de equivalencia por filas y por columnas el número de clases de equivalencia no es necesariamente finito. Es decir, mientras el conjunto cociente por la relación $\sim$ tiene un número finito de clases, el conjunto cociente por $\sim_f$ y por $\sim_c$ podría contener infintas (dependiendo del cuerpo $\mathbb{K}$). Por ejemplo, en $\mathbb{R}^{3\times 3}$,

$$\begin{bmatrix} 1 & 0 & a \\ 0 & 1 & b \\ 0 & 0 & 0 \end{bmatrix}$$

proporciona FERF diferentes al variar los valores de a y b en $\mathbb{R}$.

Además, es necesario aclarar el nuevo concepto introducido: la expresión forma canónica. Según el Dicionario "The definitive glossary of Higher Mathematical Jargon", Math Vault[4], una **forma canónica** es una representación que se ha visto como la más simple de un objeto, la cual permite identificar al objeto de forma única.

[4]https://mathvault.ca/math-glossary/canonical.

6.4. EJERCICIOS

(1) Escribir todas las posibles formas de las matrices elementales por filas en $\mathbb{K}^{2\times 2}$. Repetir para las matrices elementales por columnas y establecer una relación entre unas y otras.

(2) Demostrar que la matriz identidad I_2 se puede transformar en la matriz elemental E_{12} (de tipo I) utilizando únicamente 4 operaciones elementales por filas de tipo II y de tipo III. Generalizando este hecho a matrices de tamaño arbitrario, se deduce que la operación elemental de tipo I no es independiente de las demás (y, por tanto, ¡es redundante!).

(3) Demostrar que si $A \in \mathbb{K}^{m\times n}$ y existe una matriz $P \in \mathbb{K}^{n\times n}$ tal que

$$\begin{bmatrix} A \\ I_n \end{bmatrix} P = \begin{bmatrix} W \\ C \end{bmatrix}$$

entonces $AC = W$. Enunciar y demostrar un resultado similar operando por filas. Comparar con las matrices P y ${}_AR$ obtenidas en el Ejemplo 6.3.

(4) Demostrar que si $A, B \in \mathbb{K}^{m\times n}$ entonces $A \sim_c B \Leftrightarrow A^t \sim_f B^t$.

(5) Proporcionar un ejemplo de dos matrices que cumplan $A \sim B$ y, sin embargo, $A \nsim_f B$.

(6) Hallar la forma escalonada reducida por filas R_A de las siguientes matrices A y una matriz Q que permita expresar $QA = R_A$:

(a) $A = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 0 & 1 & 2 & 0 \end{bmatrix}$,

(b) $A = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 1 & 2 \\ 0 & 1 & 1 \end{bmatrix}$,

(c) $A = \begin{bmatrix} 1 & -1 & 0 & 1 & 2 & 4 \\ 1 & 1 & 2 & 0 & 3 & 1 \\ 2 & 6 & 4 & -3 & 6 & 1 \end{bmatrix}$.

(7) Se consideran las matrices

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

(a) Escribir la matriz A como producto de una matriz invertible (que debe ser calculada utilizando matrices elementales) por su forma escalonada reducida por filas.

(b) Hallar la forma escalonada reducida de A.

(c) Indicar si $A \sim B$ y si $A \sim_f B$. Justificar.

(8) Sean $A, R \in \mathbb{K}^{m\times n}$, $Q, U \in \mathbb{K}^{m\times m}$ y $P, V \in \mathbb{K}^{n\times n}$. Si

$$\begin{bmatrix} Q & O \\ O & I_n \end{bmatrix} \begin{bmatrix} A & I_m \\ I_n & O \end{bmatrix} \begin{bmatrix} P & O \\ O & I_m \end{bmatrix} = \begin{bmatrix} R & U \\ V & O \end{bmatrix},$$

probar que $QAP = R$, $U = Q$ y $V = P$. Deducir de este resultado el procedimiento utilizado en el Ejemplo 6.4.

(9) Demostrar que si $A \in \mathbb{K}^{m\times n}$ y $B \in \mathbb{K}^{p\times q}$ entonces

$$\operatorname{rg}\left(\begin{bmatrix} A & O \\ O & B \end{bmatrix}\right) = \operatorname{rg}(A) + \operatorname{rg}(B).$$

(10) Dar un ejemplo de dos matrices reales A y B que cumplan $\operatorname{rg}(AB) < \text{mín}\{\operatorname{rg}(A), \operatorname{rg}(B)\}$ y otro donde se cumpla la igualdad de ambas cantidades.

(11) Hallar la forma escalonada por filas (triangulada, sin reducirla) de una matriz $A = [a_{ij}] \in \mathbb{R}^{3\times 3}$ suponiendo que $a_{11} \neq 0$ y $\delta := a_{11}a_{22} - a_{21}a_{12} \neq 0$. ¿Qué expresión aparece en la posición $(3, 3)$ de la matriz reducida?

(12) Sean $A, R \in \mathbb{K}^{m\times n}$. Si R_A es la forma escalonada reducida por filas de A y A es una matriz escalonada reducida por filas, probar que $A = R_A$.

(13) Sea $A \in \mathbb{K}^{m\times n}$. En este ejercicio se estudian las inversas laterales (a izquierda y a derecha) de A.

(a) Si existe una inversa a izquierda de A, probar que $\operatorname{rg}(A) = n$.

(b) Suponiendo que $\operatorname{rg}(A) = n$, justificar la existencia de una matriz invertible $Q \in \mathbb{K}^{m\times m}$ tal que

$$QA = \begin{bmatrix} I_n \\ O_{(m-n)\times n} \end{bmatrix}.$$

Ahora calcular el producto $Q\begin{bmatrix} A & I_m \end{bmatrix}$ y, particionando la matriz Q mediante

$$Q = \begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix} \in \mathbb{K}^{(n+(m-n))\times m},$$

probar que la matriz $Q_1 \in \mathbb{K}^{n\times m}$ es tal que

$$Q_1 A = I_n.$$

Concluir que este procedimiento permite hallar una inversa a izquierda de A (si existe).

(c) Deducir que existe una inversa a izquierda de A si y sólo si $\operatorname{rg}(A) = n$.

(d) Si la matriz Q del apartado (b) viene dada por

$$Q = \begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix} \in \mathbb{K}^{(n+(m-n))\times m},$$

probar que las matrices $Q_1 + XQ_2$ son inversas a izquierda cualquiera que sea $X \in \mathbb{K}^{n\times m}$.

(e) En uno de los apartados anteriores, el tamaño de una de sus matrices es erróneo. ¿De cuál se trata?

(f) Si $X_0 \in \mathbb{K}^{n\times m}$ satisface $X_0 A = I_n$, probar que

$$\{X \in \mathbb{K}^{n\times m} : XA = I_n\} = X_0 + \{Z \in \mathbb{K}^{n\times m} : ZA = O_n\},$$

es decir, que toda inversa a izquierda de A se determina como suma de una inversa a izquierda particular de A (solución del sistema no homogéneo $XA = I_n$) más todas las matrices Z solución del sistema homogéneo $ZA = O_n$.

(g) Establecer resultados duales a los anteriores para las inversas a derecha.

(h) Justificar si las siguientes matrices admiten inversas a izquierda y, en caso afirmativo, hallar una de ellas. Repetir el ejercicio para las inversas a derecha:

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

Capítulo 7

Determinantes

Índice

7.1. Introducción

Los determinantes fueron estudiados en profundidad a partir de mediados del siglo XVIII. En 1815, Augustin Louis Cauchy realizó la disposición de los elementos en una tabla de números y la notación de los dobles subíndices para la notación moderna de los determinantes. Estos fueron introducidos simplemente como una notación de esta herramienta considerablemente útil en diversas áreas de las Matemáticas.

Por su parte, Pierre-Simon, Marquês de Laplace (1749-1827) utilizaba implícitamente determinantes para discutir las soluciones de sistemas de ecuaciones lineales. Relativo a esta notación, Laplace llegó a decir:

> *"Tanta es la ventaja de un lenguaje bien construido que su notación simplificada a menudo se convierte en fuente de teorías profundas".*

Uno de los resultados más importantes en esta teoría establece que el determinante de un producto de dos matrices cuadradas del mismo tamaño es el producto de sus determinantes. En 1773, Joseph-Louis Lagrange (1736-1813) ya había demostrado este resultado para el caso de matrices 3×3. Y, si bien, fue enunciado en general por Jacques Philippe Marie Binet (1786-1856) en 1812, su demostración no era correcta. El resultado general fue demostrado por Cauchy en 1841.

Algunas propiedades importantes de los determinantes se deben a Heinrich F. Scherk (1798-1855). En su Disertación Matemática (Mathematische Abhandlungen) de 1825 probó que el determinante es una aplicación multilineal en las filas y la invariancia al aplicar operaciones elementales (Proposiciones 7.1, 7.2 y 7.3); también calculó el determinante de una matriz triangular.

En 1750, Gabriel Cramer (1704-1752) introdujo la regla que lleva su nombre para la resolución de sistemas de ecuaciones lineales (cuadrados compatibles determinados).

Una importante aplicación de los determinantes se debe a Carl Gustav Jakob Jacobi (1804-1851) que, si bien no hizo incursiones en la Geometría

(a) Augustin Louis Cauchy

(b) Pierre Simon Laplace

Figura 7.1: Augustin Louis Cauchy y Pierre Simon Laplace.

Analítica donde tienen grandes aplicaciones, los utilizó en 1827 en cuestiones de Cálculo donde el determinante de una matriz cuadrada cuyos elementos son ciertas derivadas se conoce como *Jacobiano*.

7.1.1. Determinantes y sistemas de ecuaciones lineales

Si bien se ha situado su estudio sistemático a mediados del siglo XVIII, en 1683 el matemático japonés Kōwa Seki ya utilizaba la idea de determinantes para resolver sistemas de ecuaciones lineales. Si, por ejemplo, se resuelve por el método de eliminación de Gauss el sistema

$$S : \left\{ \begin{array}{rcl} ax + by & = & c \\ dx + ey & = & f \end{array} \right. ,$$

suponiendo $a \neq 0$, se tiene

$$\left\{ \begin{array}{ccccccc} ax & + & b & y & = & c \\ & & (e - b\frac{d}{a}) & y & = & f - c\frac{d}{a} \end{array} \right. ,$$

y si $e - b\frac{d}{a} \neq 0$ entonces $y = \frac{af-dc}{ae-db}$. Sustituyendo en la primera ecuación, $x = \frac{ce-fb}{ae-db}$. Disponiendo los elementos a, b, d, e en una matriz

$$A = \begin{bmatrix} a & b \\ d & e \end{bmatrix},$$

es posible asociar a la matriz A el número: $\det(A) := ae - db$, que se llamará **determinante**[1] de A, pues permite *determinar* las soluciones del sistema lineal mediante:

$$x = \frac{\det\left(\begin{bmatrix} c & b \\ f & e \end{bmatrix}\right)}{\det(A)}, \qquad y = \frac{\det\left(\begin{bmatrix} a & c \\ d & f \end{bmatrix}\right)}{\det(A)}.$$

Si se repite el procedimiento para un sistema lineal de 3 ecuaciones con 3 incógnitas x, y y z, se obtendrá que el sistema tiene una única solución si y sólo si un cierto número es diferente de 0. A ese número se le pondrá el nombre de determinante de una matriz 3×3. Surge así, de manera natural, el concepto de determinante.

En este capítulo, para cada $n \in \mathbb{N}$, se dará una fórmula que permita decidir si un sistema de ecuaciones lineales cuadrado es compatible determinado o no y servirá también para decidir si una matriz $A \in \mathbb{K}^{n\times n}$ es invertible o no. Como es lógico, dicha fórmula se complica según aumenta n, el tamaño de la matriz A.

7.1.2. Determinantes y Geometría

A continuación se mencionará una aplicación del concepto de vector[2], de la que también surgirá de manera natural el concepto de determinante, esta vez desde un punto de vista geométrico. Utilizando vectores, se pondrá de manifiesto una relación entre la noción geométrica de **área de un paralelogramo** y el concepto algebraico de **determinante**, que están estrechamente

[1] Fue Gauss quien lo llamó *determinante* pues determina si un sistema de ecuaciones lineales (cuadrado) tiene o no solución.

[2] Recordar el apartado llamado “un ejemplo importante” mencionado en la página 54.

relacionados en $\mathbb{R}^2$. Este hecho permitirá motivar el interés por las propiedades de los determinantes, que se podrán interpretar geométricamente de manera sencilla en $\mathbb{R}^2$ y $\mathbb{R}^3$ y, así, dichas propiedades resultarán intuitivas y sus generalizaciones a $\mathbb{R}^n$ serán inmediatas.

Si se dispone de un sistema de ejes cartesianos rectangulares en el plano euclídeo[3] $\mathbb{R}^2$, mediante el concepto de coordenadas de Geometría Analítica, se consideran dos vectores columna (no nulos) tales que ninguno sea múltiplo del otro:

$$u = \begin{bmatrix} a \\ b \end{bmatrix} \qquad \text{y} \qquad v = \begin{bmatrix} c \\ d \end{bmatrix}.$$

Se recuerda que:

- módulo de u: $\|u\| = \sqrt{a^2 + b^2}$.
- producto escalar de u por v: $\langle u, v\rangle = ac + bd = \|u\|\|v\|\cos(\alpha)$, siendo α el ángulo formado por u y v.

Los vectores u y v determinan un paralelogramo $\mathcal{P}_{u,v}$ en $\mathbb{R}^2$ cuya área se denotará por $\acute{A}rea(\mathcal{P}_{u,v})$. Una gráfica que ilustra esta situación se aprecia en la Figura 7.2. Para calcularla, se considera la longitud de u como (la

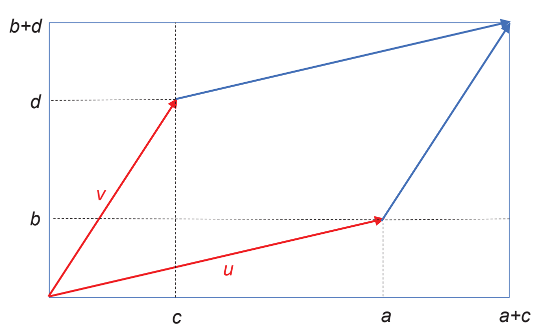


Figura 7.2: Área del paralelogramo $\mathcal{P}_{u,v}$.

medida de) la base del paralelogramo. Además, si se denota por h la altura

[3]El plano euclídeo es el conocido de Secundaria.

correspondiente a esa base, dicha altura será la longitud del segmento con origen en el punto extremo de v, que tiene pie en la recta que determina u y cuya dirección es la determinada por la perpendicular a la recta que determina u y pasa por el punto extremo de v. En esta situación, se cumple que: $\operatorname{sen}(\alpha) = \frac{h}{\|v\|}$, siendo α el ángulo formado por u y v. Luego,

$$\acute{A}rea(\mathcal{P}_{u,v}) = \|u\|h = \|u\|\|v\||\operatorname{sen}(\alpha)|,$$

donde el valor absoluto permite cambiar el signo en caso de ser negativo.

Ahora, recurriendo a la identidad trigonométrica $\operatorname{sen}^2(\alpha) + \cos^2(\alpha) = 1$, se tiene que

$$|\operatorname{sen}(\alpha)| = \sqrt{1-\cos^2(\alpha)} = \sqrt{1 - \frac{\langle u,v\rangle^2}{\|u\|^2\|v\|^2}} = \frac{\sqrt{\|u\|^2\|v\|^2 - \langle u,v\rangle^2}}{\|u\|\|v\|}.$$

Luego,

$$\begin{aligned}
\acute{A}rea(\mathcal{P}_{u,v}) &= \|u\|\|v\||\operatorname{sen}(\alpha)| \\
&= \|u\|\|v\|\frac{\sqrt{\|u\|^2\|v\|^2 - \langle u,v\rangle^2}}{\|u\|\|v\|} \\
&= \sqrt{\|u\|^2\|v\|^2 - \langle u,v\rangle^2} \\
&= \sqrt{(a^2+b^2)(c^2+d^2) - (ac+bd)^2} \\
&= \sqrt{a^2c^2 + a^2d^2 + b^2c^2 + b^2d^2 - a^2c^2 - 2acbd - b^2d^2} \\
&= \sqrt{a^2d^2 - 2adbc + b^2c^2} \\
&= \sqrt{(ad-bc)^2} \\
&= |ad-bc|.
\end{aligned}$$

Si se considera la matriz por bloques que determinan los vectores u y v (por columnas):

$$A = \begin{bmatrix} a & c \\ b & d \end{bmatrix}$$

y se utiliza la notación

$$\det(A) = \det(u,v) := ad - bc,$$

que se llamará **determinante** de A (para una matriz de tamaño 2×2), se tiene que

$$\acute{A}rea(\mathcal{P}_{u,v}) = |\det(A)|,$$

es decir, el área del paralelogramo que determinan u y v coincide con el valor absoluto del determinante de la matriz que forman.

Si se escribe $\acute{A}rea(\mathcal{P}_{u,v}) = \pm\det(A)$, se suele hablar de **área con signo** o **área orientada**[4].

Ahora se dispondrá de tres vectores en $\mathbb{R}^3$ no nulos y no coplanarios,

$$u = \begin{bmatrix} a \\ b \\ c \end{bmatrix}, \qquad v = \begin{bmatrix} d \\ e \\ f \end{bmatrix} \qquad \text{y} \qquad w = \begin{bmatrix} g \\ h \\ i \end{bmatrix},$$

y $\mathcal{P}_{u,v,w}$ será el paralelepípedo que ellos determinan. Utilizando la notación

$$\det(A) = \det(u, v, w) := aei + dhc + bfg - gec - hfa - dbi,$$

que se llamará **determinante** de A (para una matriz de tamaño 3×3), y el concepto de producto mixto, una idea similar a la anterior lleva a:

$$Volumen(\mathcal{P}_{u,v,w}) = |\det(u, v, w)| = |\det(A)|,$$

donde A es la matriz por bloques (según sus columnas)

$$A = \begin{bmatrix} u & v & w \end{bmatrix} = \begin{bmatrix} a & d & g \\ b & e & h \\ c & f & i \end{bmatrix}.$$

Ahora el **volumen con signo** u **orientado** es $Volumen(\mathcal{P}_{u,v,w}) = \pm\det(A)$. Esta interpretación apareció por primera vez en un trabajo de 1773 del matemático (oriundo de Italia) Joseph-Louis Lagrange (1736-1813).

Valen resultados similares disponiendo los vectores A por filas en lugar de por columnas.

[4]Este hecho indica el orden u orientación en que se han tomado los vectores que determinan sus lados.

Figura 7.3: Joseph-Louis Lagrange.

7.2. Definición

Hay diferentes formas de introducir la definición de determinante de una matriz cuadrada[5]. Se elige la que se considera más sencilla (por recurrencia) evitando así introducir, previamente, nuevos conceptos que alargarían demasiado la introducción al tema. Con la intención de evitar que la definición posea un aspecto poco intuitivo, se han establecido las aplicaciones previas, que le confieren naturalidad en los casos de matrices de tamaños pequeños. Se trata ahora de generalizar esas ideas.

El objetivo consiste en asignar a una matriz cuadrada $A \in \mathbb{K}^{n\times n}$ un elemento de $\mathbb{K}$. La aplicación

$$\det : \mathbb{K}^{n\times n} \to \mathbb{K}$$

que permita hacerlo se llamará **determinante**.

[5]Tres enfoques en que puede introducirse el concepto de determinantes son: (a) como una forma multilineal alternada (se requiere conocer aplicaciones lineales), (b) a partir del concepto de permutación (y de su clasificación en tipo par e impar) que se debe a Leibniz, (c) utilizando complementos algebraicos y recurrencia.

Definición 7.1. Para toda matriz cuadrada $A = [a_{11}] \in \mathbb{K}^{1\times 1}$ se llama **determinante de orden** 1 al elemento a_{11} de $\mathbb{K}$; en símbolos,

$$\det(A) = |A| = a_{11}.$$

Sea $n > 1$. Suponiendo conocido el valor del determinante de toda matriz de tamaño $(n-1)\times(n-1)$ a coeficientes en $\mathbb{K}$, para cada matriz

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \in \mathbb{K}^{n\times n}$$

se llama **determinante de orden** n de A al elemento de $\mathbb{K}$, denotado por $\det(A) = |A|$, dado por:

$$\det(A) = |A| = a_{11}A_{11} + a_{21}A_{21} + \cdots + a_{n1}A_{n1}, \tag{7.1}$$

donde el elemento A_{ij} se llama el **complemento algebraico** o **cofactor** de la posición[a] (i,j) y viene dado por

$$A_{ij} = (-1)^{i+j}\det(M_{ij}),$$

siendo M_{ij} la submatriz que se obtiene de A al suprimir la fila i-ésima y la columna j-ésima, y se llama **menor complementario** de la posición (i,j).

[a] Es importante observar que para definir el complemento algebraico A_{ij} lo relevante no es el valor que tome el propio elemento a_{ij} en sí, sino la posición (i,j) que ocupa, que serán la fila y la columna a eliminar en el cálculo de M_{ij}.

La fórmula (7.1) expresa que el determinante de A se define como la suma de los elementos de la primera columna de A multiplicados por los complementos algebraicos de las respectivas posiciones de dichos elementos:

$$\det(A) = \sum_{k=1}^{n} a_{k1} A_{k1} = \sum_{k=1}^{n} (-1)^{k+1} a_{k1} \det(M_{k1}).$$

Se la conoce con el nombre de **desarrollo de Laplace** del determinante de A por los elementos de su primera columna. Se observa que el patrón de signos de la expresión $(-1)^{i+j}$ en los cofactores depende únicamente de la posición (i, j) que ocupa cada elemento y presenta el siguiente aspecto:

$$\text{signo}([(-1)^{i+j}]) = \begin{bmatrix} + & - & + & - & \cdots \\ - & + & - & + & \cdots \\ + & - & + & - & \cdots \\ - & + & - & + & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix},$$

donde la *función signo* aplicada a una matriz a coeficientes reales no nulos, proporciona signo $+$ si el valor es positivo y $-$ si es negativo.

A continuación se escriben de manera explícita los determinantes de matrices de tamaño 2×2 y 3×3:

- Si $A \in \mathbb{K}^{2\times 2}$ entonces

$$\begin{aligned} \det(A) &= \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \\ &= a_{11}A_{11} + a_{21}A_{21} \\ &= a_{11}(-1)^{1+1}|M_{11}| + a_{21}(-1)^{2+1}|M_{21}| \\ &= a_{11}|M_{11}| - a_{21}|M_{21}| \\ &= a_{11}|a_{22}| - a_{21}|a_{12}| \\ &= a_{11}a_{22} - a_{21}a_{12}. \end{aligned}$$

- Si $A \in \mathbb{K}^{3\times 3}$ entonces

$$\begin{aligned}
\det(A) &= \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} \\
&= a_{11}A_{11} + a_{21}A_{21} + a_{31}A_{31} \\
&= a_{11}(-1)^{1+1}\left|M_{11}\right| + a_{21}(-1)^{2+1}\left|M_{21}\right| + a_{31}(-1)^{3+1}\left|M_{31}\right| \\
&= a_{11}\left|M_{11}\right| - a_{21}\left|M_{21}\right| + a_{31}\left|M_{31}\right| \\
&= a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{21}\begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} + a_{31}\begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix} \\
&= a_{11}[a_{22}a_{33} - a_{32}a_{23}] - a_{21}[a_{12}a_{33} - a_{32}a_{13}] + \\
&\quad + a_{31}[a_{12}a_{23} - a_{22}a_{13}] \\
&= a_{11}a_{22}a_{33} - a_{11}a_{32}a_{23} - a_{21}a_{12}a_{33} + a_{21}a_{32}a_{13} + \\
&\quad + a_{31}a_{12}a_{23} - a_{31}a_{22}a_{13}.
\end{aligned}$$

Regla de Sarrus: Es una regla mnemotécnica para recordar el valor del determinante de una matriz de tamaño 3×3 que consiste en[6] repetir las dos primeras filas de la matriz debajo de la tercera fila y luego hacer productos de tres elementos por diagonales con el signo adecuado y sumarlos todos. Se comienza por la diagonal principal y las dos que están por debajo de ella (se les antepone signo positivo) y, luego, se sigue por la diagonal secundaria y las dos por debajo de ella (se les antepone signo negativo)[7]:

$$\begin{vmatrix} \color{red}{a_{11}} & a_{12} & a_{13} \\ \color{brown}{a_{21}} & \color{red}{a_{22}} & a_{23} \\ \color{blue}{a_{31}} & \color{brown}{a_{32}} & \color{red}{a_{33}} \\ a_{11} & \color{blue}{a_{12}} & \color{brown}{a_{13}} \\ a_{21} & a_{22} & \color{blue}{a_{23}} \end{vmatrix} \text{(signo positivo)}, \quad \begin{vmatrix} a_{11} & a_{12} & \color{red}{a_{13}} \\ a_{21} & \color{red}{a_{22}} & \color{brown}{a_{23}} \\ \color{red}{a_{31}} & \color{brown}{a_{32}} & \color{blue}{a_{33}} \\ \color{brown}{a_{11}} & \color{blue}{a_{12}} & a_{13} \\ \color{blue}{a_{21}} & a_{22} & a_{23} \end{vmatrix} \text{(signo negativo)}.$$

[6] Es probable que sea más conocida la que utiliza diagonales (principal y secundaria) y sus paralelas formando un triángulo con el elemento que queda en la esquina contraria.

[7] O bien, repetir las dos primeras columnas a continuación de la tercera, empezando por la diagonal principal y siguiendo por las diagonales paralelas hacia su derecha (positivos) y desde la secundaria siguiendo por diagonales paralelas hacia su izquierda (negativos).

Ejemplo 7.1. *El determinante de la matriz identidad I_n es 1, para todo $n \in \mathbb{N}$.*

$\lhd$ Se probará por inducción sobre el número natural n, que determina el tamaño de la matriz. En efecto, es claro que $\det(I_1) = 1$, por la Definición 7.1, por tanto la propiedad es válida para $n = 1$. Sea $n > 1$. Se supone, por hipótesis de inducción que $\det(I_{n-1}) = 1$. Luego, por la Definición 7.1 y de la hipótesis de inducción,

$$\begin{aligned} \det(I_n) &= 1(-1)^{1+1}\det(M_{11}) + 0A_{21} + \cdots + 0A_{n1} \\ &= \det(I_{n-1}) \\ &= 1. \end{aligned}$$

Por el principio de inducción, la propiedad se cumple para cualquier valor de $n \in \mathbb{N}$. Es decir, $\det(I_n) = 1$, $\forall n \in \mathbb{N}$. $\rhd$

Ejemplo 7.2. *El determinante de la matriz nula O_n es 0.*

$\lhd$ Se prueba por inducción sobre n. La prueba es inmediata y se propone como ejercicio. $\rhd$

Ejemplo 7.3. *Calcular el determinante de la matriz*

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

$\lhd$ Al ser nulos dos de los elementos de la primera columna de A ($a_{11} =$

$0 = a_{41}$), por definición se tiene

$$\begin{aligned}
\det(A) &= 1(-1)^{2+1}\begin{vmatrix} 0 & 4 & 1 \\ -2 & 1 & 3 \\ 0 & 1 & 2 \end{vmatrix} + 2(-1)^{3+1}\begin{vmatrix} 0 & 4 & 1 \\ -1 & 3 & 2 \\ 0 & 1 & 2 \end{vmatrix} \\
&= -(-2)(-1)^{2+1}\begin{vmatrix} 4 & 1 \\ 1 & 2 \end{vmatrix} + 2(-1)(-1)^{2+1}\begin{vmatrix} 4 & 1 \\ 1 & 2 \end{vmatrix} \\
&= -2(4 \cdot 2 - 1 \cdot 1) + 2(4 \cdot 2 - 1 \cdot 1) \\
&= 0.
\end{aligned}$$

$\triangleright$

El siguiente es un determinante importante de conocer para calcular en la práctica el de otras matrices, reduciéndolas previamente a triangulares, como se verá más adelante.

Ejemplo 7.4. *El determinante de una matriz triangular superior $T \in \mathbb{K}^{n\times n}$ coincide con el producto de los elementos de la diagonal principal. En símbolos,*

$$\det\begin{pmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1,n-1} & a_{1n} \\ 0 & a_{22} & a_{23} & \dots & a_{2,n-1} & a_{2n} \\ 0 & 0 & a_{33} & \dots & a_{3,n-1} & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & a_{n-1,n-1} & a_{n-1,n} \\ 0 & 0 & 0 & \dots & 0 & a_{nn} \end{pmatrix} = a_{11}a_{22}a_{33}\dots a_{n-1,n-1}a_{nn}.$$

$\triangleleft$ Se realizará por inducción sobre el número natural n, que determina el tamaño de la matriz.

Si $n = 1$, por definición,

$$\det\left(\begin{bmatrix} a_{11} \end{bmatrix}\right) = a_{11}.$$

Sea $n > 1$ y supóngase que el resultado es verdadero para toda matriz triangular superior de tamaño $(n-1) \times (n-1)$. Se debe probar que la propiedad vale para toda matriz triangular superior de tamaño $n \times n$. En efecto, como todos los elementos de la primera columna son cero excepto el primero, por definición se tiene que si

$$T = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \ldots & a_{1,n-1} & a_{1n} \\ 0 & a_{22} & a_{23} & \ldots & a_{2,n-1} & a_{2n} \\ 0 & 0 & a_{33} & \ldots & a_{3,n-1} & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & a_{n-1,n-1} & a_{n-1,n} \\ 0 & 0 & 0 & \ldots & 0 & a_{nn} \end{bmatrix}$$

entonces

$$\begin{aligned} \det(T) &= a_{11}(-1)^{1+1}\det(M_{11}) \\ &= a_{11}\det\begin{pmatrix} a_{22} & a_{23} & \ldots & a_{2,n-1} & a_{2n} \\ 0 & a_{33} & \ldots & a_{3,n-1} & a_{3n} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & a_{n-1,n-1} & a_{n-1,n} \\ 0 & 0 & \ldots & 0 & a_{nn} \end{pmatrix} \\ &= a_{11}(a_{22}a_{33}\ldots a_{n-1,n-1}a_{nn}) \\ &= a_{11}a_{22}a_{33}\ldots a_{n-1,n-1}a_{nn}. \end{aligned}$$

Como el determinante de M_{11} corresponde al de una matriz triangular superior de tamaño $(n-1)\times(n-1)$, se ha aplicado la hipótesis de inducción para obtenerlo.

Por el principio de inducción, el resultado es válido para todo $n \in \mathbb{N}$. $\triangleright$

7.3. Propiedades. Interpretaciones geométricas

En esta sección se estudian las propiedades de los determinantes.

El determinante es una aplicación multilineal en las filas

Se ha visto que el determinante de una matriz cuadrada A se define como una aplicación $\det : \mathbb{K}^{n\times n} \to \mathbb{K}$. Si se fijan todos los elementos de la matriz A excepto los de la fila i-ésima, el determinante será función de los elementos de dicha fila. Teniendo esta idea en mente, se tienen los dos resultados siguientes en los que una fila se sustituirá por la suma de otras dos, o bien una fila se sustituirá por la que resulta de multiplicarla por un escalar.

Proposición 7.1 (Aplicación aditiva en las filas)**.** Sean $A, B, C \in \mathbb{K}^{n\times n}$ tres matrices con todos sus elementos iguales excepto los de la i-ésima fila de A, que son la suma de los respectivos elementos de las filas i-ésimas de B y C. Entonces $\det(A) = \det(B) + \det(C)$. Utilizando elementos,

$$\begin{vmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ b_{i1}+c_{i1} & \dots & b_{in}+c_{in} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{vmatrix} = \begin{vmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ b_{i1} & \dots & b_{in} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{vmatrix} + \begin{vmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ c_{i1} & \dots & c_{in} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{vmatrix}.$$

Demostración. Se realizará por inducción sobre el número natural n, que determina el tamaño de la matriz.

Si $n = 1$, por definición de determinante, $\det\left(\begin{bmatrix} b_{11}+c_{11} \end{bmatrix}\right) = b_{11}+c_{11} = \det\left(\begin{bmatrix} b_{11} \end{bmatrix}\right) + \det\left(\begin{bmatrix} c_{11} \end{bmatrix}\right)$, con lo que la propiedad se cumple.

Sea $n > 1$ y supóngase que la propiedad se cumple para todas las matrices de tamaño $(n-1)\times(n-1)$. Se debe probar que la propiedad se cumple para matrices de tamaño $n \times n$. En efecto, se consideran tres matrices A, B, C de tamaño $n \times n$, con la misma notación del enunciado. Utilizando la notación $a_{ij} := b_{ij} + c_{ij}$ para la fila i-ésima $(j = 1, 2, \dots, n)$ de A, por definición de determinante, se tiene

$$\begin{aligned}
\det(A) &= a_{11}A_{11} + \cdots + a_{i1}A_{i1} + \cdots + a_{n1}A_{n1} \\
&= a_{11}A_{11} + \cdots + (b_{i1} + c_{i1})A_{i1} + \cdots + a_{n1}A_{n1} \\
&= a_{11}A_{11} + \cdots + b_{i1}A_{i1} + c_{i1}A_{i1} + \cdots + a_{n1}A_{n1}. \qquad (7.2)
\end{aligned}$$

Si A_{i1}, B_{i1} y C_{i1} son los complementos algebraicos correspondientes a la posición $(i,1)$ de las matrices A, B y C (i.e., en este caso, asociados a los elementos a_{i1}, b_{i1} y c_{i1}), respectivamente, entonces $A_{i1} = B_{i1} = C_{i1}$. En efecto,

$$A_{i1} = (-1)^{i+1}\det(M_{i1}^A) = (-1)^{i+1}\det(M_{i1}^B) = B_{i1},$$

donde M_{i1}^A y M_{i1}^B son los menores complementarios correspondientes a las posiciones $(i,1)$ de las matrices A y B, respectivamente. Se debe tener en cuenta que $M_{i1}^A = M_{i1}^B$ pues en las matrices A y B se suprime la fila i-ésima (además de la primera columna, que no afecta) que es la única en la que difieren. De la misma forma $A_{i1} = C_{i1}$.

Mientras tanto, los complementos algebraicos A_{k1}, B_{k1} y C_{k1} cumplen $A_{k1} = B_{k1} + C_{k1}$ para cada $k \neq i$ $(k = 1, 2, \ldots, n)$. En efecto, por hipótesis de inducción, $\det(M_{k1}^A) = \det(M_{k1}^B) + \det(M_{k1}^C)$ por tratarse de determinantes de matrices de tamaño $(n-1) \times (n-1)$ donde una fila de M_{k1}^A es suma de otras dos: las filas correspondientes de M_{k1}^B y de M_{k1}^C. Luego,

$$\begin{aligned}
A_{k1} &= (-1)^{k+1}\det(M_{k1}^A) \\
&= (-1)^{k+1}[\det(M_{k1}^B) + \det(M_{k1}^C)] \\
&= (-1)^{k+1}\det(M_{k1}^B) + (-1)^{k+1}\det(M_{k1}^C) \\
&= B_{k1} + C_{k1}.
\end{aligned}$$

Sustituyendo en (7.2) y aplicando dos veces la definición de determinante,

$$\begin{aligned}
\det(A) &= a_{11}A_{11} + \cdots + b_{i1}A_{i1} + c_{i1}A_{i1} + \cdots + a_{n1}A_{n1} \\
&= a_{11}(B_{11} + C_{11}) + \cdots + b_{i1}B_{i1} + c_{i1}C_{i1} + \cdots + a_{n1}(B_{n1} + C_{n1}) \\
&= (a_{11}B_{11} + \cdots + b_{i1}B_{i1} + \cdots + a_{n1}B_{n1}) + \\
&\quad +(a_{11}C_{11} + \cdots + c_{i1}C_{i1} + \cdots + a_{n1}C_{n1}) \\
&= \det(B) + \det(C).
\end{aligned}$$

Por el principio de inducción, la propiedad se cumple para cualquier valor de $n \in \mathbb{N}$. $\square$

Interpretación geométrica: Se consideran tres matrices $A, B, C \in \mathbb{R}^{2\times 2}$ que comparten la primera fila denotada por a. Se denotan por b la segunda fila de B y c la segunda fila de C. La segunda fila de A es $b + c$. Se ha visto que:

- $|\det(B)| = \acute{A}rea(\mathcal{P}_{a,b})$, el área del paralelogramo determinado por a y b;
- $|\det(C)| = \acute{A}rea(\mathcal{P}_{a,c})$, el área del paralelogramo determinado por a y c; y
- $|\det(A)| = \acute{A}rea(\mathcal{P}_{a,b+c})$, el área del paralelogramo determinado por a y $b + c$.

En una gráfica en el plano euclídeo $\mathbb{R}^2$ se puede observar, que el área $\acute{A}rea(\mathcal{P}_{a,b+c})$ coincide con la suma de $\acute{A}rea(\mathcal{P}_{a,b})$ y $\acute{A}rea(\mathcal{P}_{a,c})$. Para ello, construir un rectángulo de lados a y b no paralelos, trasladar el vector a sobre su lado paralelo sobre dicho rectángulo y colocar c con origen en el extremo de b. Ahora comparar áreas.

Proposición 7.2 (Aplicación homogénea en las filas)**.** Sean $A, B \in \mathbb{K}^{n\times n}$ dos matrices con todos sus elementos iguales excepto los de la i-ésima fila de B, que son el producto de $\alpha \in \mathbb{K}$ por los respectivos elementos de la i-ésima fila de A. Entonces $\det(B) = \alpha \det(A)$. Utilizando elementos,

$$\begin{vmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ \alpha a_{i1} & \dots & \alpha a_{in} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{vmatrix} = \alpha \begin{vmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{i1} & \dots & a_{in} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{vmatrix}.$$

Demostración. Se realizará por inducción sobre el número natural n, que determina el tamaño de la matriz.

Si $n = 1$, la propiedad se cumple puesto que, por definición de determinante, $\det\left(\begin{bmatrix} \alpha a_{11} \end{bmatrix}\right) = \alpha a_{11} = \alpha \det\left(\begin{bmatrix} a_{11} \end{bmatrix}\right)$.

Sea $n > 1$ y supóngase que la propiedad se cumple para todas las matrices de tamaño $(n-1)\times(n-1)$. Se debe probar que la propiedad se cumple para matrices de tamaño $n \times n$. En efecto, se consideran dos matrices A y B de tamaño $n \times n$, con la misma notación del enunciado. Utilizando la notación $b_{ij} := \alpha a_{ij}$ para la fila i-ésima ($j = 1, 2, \dots, n$) de B, por definición de determinante, se tiene

$$\begin{aligned} \det(B) &= a_{11}B_{11} + \cdots + b_{i1}B_{i1} + \cdots + a_{n1}B_{n1} \\ &= a_{11}B_{11} + \cdots + (\alpha a_{i1})B_{i1} + \cdots + a_{n1}B_{n1}. \end{aligned} \tag{7.3}$$

Los complementos algebraicos B_{i1} así como A_{i1} correspondientes a la posición $(i, 1)$ de las matrices B y A (i.e, en este caso, asociados a los elementos b_{i1} y a_{i1}), respectivamente, satisfacen que $B_{i1} = A_{i1}$. En efecto,

$$B_{i1} = (-1)^{i+1}\det(M_{i1}^B) = (-1)^{i+1}\det(M_{i1}^A) = A_{i1},$$

pues $M_{i1}^B = M_{i1}^A$ debido a que en las dos matrices se suprime la fila i-ésima (además de la primera columna, que no afecta) que es la única en la que ambas matrices difieren.

Mientras tanto, los complementos algebraicos B_{k1} y A_{k1} satisfacen que $B_{k1} = \alpha A_{k1}$ para cada $k \neq i$ $(k = 1, 2, \dots, n)$. En efecto,

$$B_{k1} = (-1)^{k+1} \det(M_{k1}^B) = (-1)^{k+1} \alpha \det(M_{k1}^A) = \alpha A_{k1},$$

puesto que, por hipótesis de inducción, $\det(M_{k1}^B) = \alpha \det(M_{k1}^A)$, por tratarse de determinantes de matrices de tamaño $(n-1)\times(n-1)$ donde una fila de M_{k1}^B es producto del escalar α por la fila correspondiente de M_{k1}^A. Sustituyendo en (7.3) y aplicando dos veces la definición de determinante,

$$\begin{aligned}
\det(B) &= a_{11}B_{11} + \cdots + \alpha a_{i1}B_{i1} + \cdots + a_{n1}B_{n1} \\
&= a_{11}\alpha A_{11} + \cdots + \alpha a_{i1}A_{i1} + \cdots + a_{n1}\alpha A_{n1} \\
&= \alpha[a_{11}A_{11} + \cdots + a_{i1}A_{i1} + \cdots + a_{n1}A_{n1}] \\
&= \alpha \det(A).
\end{aligned}$$

Por el principio de inducción, la propiedad se cumple para cualquier valor $n \in \mathbb{N}$. $\square$

Las dos propiedades anteriores, juntas, indican que la aplicación determinante es **lineal** en las filas. Al ser válidas para cualesquiera de las filas de las matrices (con las demás fijas), se habla de aplicación[8] **multilineal**.

[8]En realidad, para resaltar que det una aplicación con imagen sobre $\mathbb{K}$, se llama **forma** en lugar de aplicación.

Interpretación geométrica: Se consideran dos matrices $A, B \in \mathbb{R}^{2\times 2}$ que comparten la primera fila denotada por a. Se denota por b la segunda fila de A. La segunda fila de B es αb para algún $\alpha > 0$. Se ha visto que:

- $|\det(A)| = \acute{A}rea(\mathcal{P}_{a,b})$, que es el área del paralelogramo determinado por a y b; y
- $|\det(B)| = \acute{A}rea(\mathcal{P}_{a,\alpha b})$, que es el área del paralelogramo determinado por a y αb.

En una gráfica en el plano euclídeo $\mathbb{R}^2$ se puede observar que el área $\acute{A}rea(\mathcal{P}_{a,\alpha b})$ coincide con el producto de α por $\acute{A}rea(\mathcal{P}_{a,b})$. Para ello, considerando por ejemplo $\alpha = 3$, construir un rectángulo de lados a y b no paralelos, trasladar el vector a sobre su lado paralelo sobre dicho rectángulo y colocar b con origen en el extremo de b original y, de nuevo, trasladar a sobre su paralelo en el nuevo rectángulo y colocar otra vez b a continuación del vector repetido b por segunda vez. Ahora comparar áreas.

Corolario 7.1 (Determinante de una matriz con una fila nula)**.** Sean $A \in \mathbb{K}^{n\times n}$ una matriz con todos los elementos de una fila iguales a cero. Entonces $\det(A) = 0$. Utilizando elementos,

$$\begin{vmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ 0 & 0 & 0 \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{vmatrix} = 0.$$

Demostración. Basta tomar $\alpha = 0$ en la Proposición 7.2. $\square$

Interpretación geométrica: Es inmediata a partir de la interpretación geométrica anterior pues en este caso, un vector es nulo y, por tanto, se trata de un paralelogramo con área nula.

El determinante es una aplicación alternada en las filas

La propiedad alternada, que se mostrará a continuación, indica que un intercambio de filas implica un cambio de signo en el determinante.

Proposición 7.3 (Aplicación alternada en las filas)**.** Sean $A, B \in \mathbb{K}^{n\times n}$ con $n \geq 2$. Si la matriz B se obtiene al intercambiar dos filas de la matriz A entonces $\det(B) = -\det(A)$. Utilizando elementos, si $i < j$,

$$\begin{vmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{j1} & \dots & a_{jn} \\ \vdots & \ddots & \vdots \\ a_{i1} & \dots & a_{in} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{vmatrix} = - \begin{vmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{i1} & \dots & a_{in} \\ \vdots & \ddots & \vdots \\ a_{j1} & \dots & a_{jn} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{vmatrix}.$$

Demostración. Se probará por inducción sobre n. En efecto, si $n = 2$, por definición de determinante, se llega a $\det(B) = -\det(A)$ pues

$$\begin{vmatrix} a_{21} & a_{22} \\ a_{11} & a_{12} \end{vmatrix} = a_{21}a_{12} - a_{11}a_{22} = -(a_{11}a_{22} - a_{21}a_{12}) = - \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}.$$

Sea $n > 2$. La prueba se realizará en dos etapas. Para la primera supóngase que la propiedad se cumple para toda matriz de tamaño $(n-1) \times (n-1)$ con dos filas consecutivas intercambiadas. Se debe probar que la propiedad se cumple para matrices de tamaño $n \times n$ con dos filas consecutivas intercambiadas. En efecto, sea $A = [a_{ij}] \in \mathbb{K}^{n\times n}$ y sea $B \in \mathbb{K}^{n\times n}$ una matriz igual a A excepto que tiene intercambiadas dos filas consecutivas, la i-ésima con la $(i+1)$-ésima filas, es decir,

$$B = \begin{array}{c} \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{i+1,1} & \dots & a_{i+1,n} \\ a_{i1} & \dots & a_{in} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{bmatrix} \end{array} \begin{array}{l} \\ \\ \longleftarrow \text{ fila } \quad i - \text{ésima} \\ \longleftarrow \text{ fila } \ (i+1) - \text{ésima} \\ \\ \\ \end{array}$$

Se debe probar que $\det(B) = -\det(A)$. En efecto, por definición de determinante,

$$\det(B) = a_{11}B_{11} + \cdots + a_{i+1,1}B_{i1} + a_{i1}B_{i+1,1} + \cdots + a_{n1}B_{n1},$$

donde A_{ij} y B_{ij} son los complementos algebraicos correspondientes a la posición (i,j) de las matrices A y B, respectivamente (en este caso, con elementos $a_{i+1,1}$ y a_{i1} en B). Al calcular los menores complementarios correspondientes a las filas i-ésima e $(i+1)$-ésima y primera columna se tiene $M^B_{i1} = M^A_{i+1,1}$ y $M^B_{i+1,1} = M^A_{i1}$ y así $\det(M^B_{i1}) = \det(M^A_{i+1,1})$ y $\det(M^B_{i+1,1}) = \det(M^A_{i1})$.

Por otro lado, por hipótesis de inducción, $\det(M^B_{k1}) = -\det(M^A_{k1})$, para todo $k = 1, 2, \dots, i-1, i+2, \dots, n$ por tratarse de determinantes de matrices de tamaño $(n-1) \times (n-1)$ con dos filas consecutivas intercambiadas.

Luego, sustituyendo los valores encontrados y aplicando la definición de determinante,

$$\begin{aligned} \det(B) &= a_{11}(-1)^{1+1}|M^B_{11}| + \cdots + a_{i+1,1}(-1)^{i+1}|M^B_{i1}| + \\ &\quad + a_{i1}(-1)^{(i+1)+1}|M^B_{i+1,1}| + \cdots + a_{n1}(-1)^{n+1}|M^B_{n1}| \\ &= a_{11}(-1)^{1+1}(-|M^A_{11}|) + \cdots + a_{i+1,1}(-1)^{i+1}|M^A_{i+1,1}| + \\ &\quad + a_{i1}(-1)^{(i+1)+1}|M^A_{i1}| + \cdots + a_{n1}(-1)^{n+1}(-|M^A_{n1}|) \\ &= -[a_{11}(-1)^{1+1}|M^A_{11}| + \cdots + a_{i+1,1}(-1)^{(i+1)+1}|M^A_{i+1,1}| + \\ &\quad + a_{i1}(-1)^{i+1}|M^A_{i1}| + \cdots + a_{n1}(-1)^{n+1}|M^A_{n1}|] \\ &= -\det(A), \end{aligned}$$

donde se deben conmutar los términos i-ésimo e $(i+1)$-ésimo de la última suma. Por el principio de inducción, la propiedad (con la suposición de tener dos filas consecutivas intercambiadas) es válida para todo $n \in \mathbb{N}$ tal que $n \geq 2$.

Para la segunda etapa se supone que $A = [a_{ij}] \in \mathbb{K}^{n\times n}$ y $B \in \mathbb{K}^{n\times n}$ es una matriz igual que A excepto que tiene intercambiadas dos filas no consecutivas, la i-ésima con la j-ésima filas, con $i \neq j$. De nuevo, se debe probar que $\det(B) = -\det(A)$. En efecto, se aplica la etapa anterior un total de $2|i-j|-1$ veces[9] para intercambiar la fila i-ésima con la j-ésima hasta obtener

$$\det(B) = (-1)^{2|i-j|-1}\det(A) = -\det(A),$$

dado que $2|i-j|-1$ es un número impar. $\square$

Corolario 7.2 (Determinante de una matriz con dos filas iguales)**.** Sea $n \geq 2$. Si $A \in \mathbb{K}^{n\times n}$ es una matriz con dos filas iguales y $1+1 \neq 0$ entonces $\det(A) = 0$. Utilizando elementos,

$$\begin{vmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{i1} & \dots & a_{in} \\ \vdots & \ddots & \vdots \\ a_{i1} & \dots & a_{in} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{vmatrix} = 0.$$

Demostración. Aplicando la Proposición 7.3, intercambiando las filas iguales

[9]Para ver este paso más fácilmente, poner por ejemplo 7 filas y contar los cambios necesarios para intercambiar la fila $i=2$ con la fila $j=6$. Serán necesarios 4 cambios para llevar la fila 2 al lugar 6 y 3 cambios (es decir, uno menos) para llevar la fila 6 a la 2. En total, 7 cambios, que escritos en términos de $i=2$ y $j=6$ queda $7 = 2\cdot 4 - 1 = 2|2-6|-1$.

entre sí, se obtiene la misma matriz, con lo que

$$\det(A) = -\det(A),$$

de donde

$$(1+1)\det(A) = 0.$$

Al ser $1+1 \neq 0$ (i.e., $\mathbb{K}$ es un cuerpo de característica distinta de 2), se tiene que $\det(A) = 0$ pues $\mathbb{K}$ no tiene divisores de cero. $\square$

Observación 7.1. La hipótesis $1+1 \neq 0$ del Corolario 7.2 no es esencial, es decir, se puede remover y el resultado sigue siendo válido. En el Ejercicio (16) de la página 327 se indica cómo realizar una demostración alternativa sin el requerimiento de dicha hipótesis.

Propiedades relativas a las operaciones elementales en las filas

Corolario 7.3 (Invariancia al aplicar una operación elemental de tipo III en las filas)**.** Si B es una matriz obtenida tras cambiar sólo una fila en una matriz $A \in \mathbb{K}^{n\times n}$, con $n \geq 2$, de modo que se le sume otra fila previamente multiplicada por un escalar $k \in \mathbb{K}$ entonces $\det(B) = \det(A)$. Utilizando elementos, para $i, j \in \{1, 2, \ldots, n\}$ tales que $i \neq j$, se cumple

$$\begin{vmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & \vdots & \vdots \\ a_{i1} + ka_{j1} & \ldots & a_{in} + ka_{jn} \\ \vdots & \vdots & \vdots \\ a_{j1} & \ldots & a_{jn} \\ \vdots & \vdots & \vdots \\ a_{n1} & \ldots & a_{nn} \end{vmatrix} = \begin{vmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & \vdots & \vdots \\ a_{i1} & \ldots & a_{in} \\ \vdots & \vdots & \vdots \\ a_{j1} & \ldots & a_{jn} \\ \vdots & \vdots & \vdots \\ a_{n1} & \ldots & a_{nn} \end{vmatrix}.$$

Demostración. Es consecuencia inmediata de las Proposiciones 7.1, 7.2 y 7.3:

$$\begin{aligned}
\begin{vmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \vdots & \vdots \\ a_{i1}+ka_{j1} & \dots & a_{in}+ka_{jn} \\ \vdots & \vdots & \vdots \\ a_{j1} & \dots & a_{jn} \\ \vdots & \vdots & \vdots \\ a_{n1} & \dots & a_{nn} \end{vmatrix} &= \begin{vmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \vdots & \vdots \\ a_{i1} & \dots & a_{in} \\ \vdots & \vdots & \vdots \\ a_{j1} & \dots & a_{jn} \\ \vdots & \vdots & \vdots \\ a_{n1} & \dots & a_{nn} \end{vmatrix} + k \begin{vmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \vdots & \vdots \\ a_{j1} & \dots & a_{jn} \\ \vdots & \vdots & \vdots \\ a_{j1} & \dots & a_{jn} \\ \vdots & \vdots & \vdots \\ a_{n1} & \dots & a_{nn} \end{vmatrix} \\
&= \det(A) + k0 \\
&= \det(A).
\end{aligned}$$

□

Interpretación geométrica: Se consideran dos matrices $A, B \in \mathbb{R}^{2\times 2}$ que comparten la primera fila denotada por a. Se denota por b la segunda fila de A. La segunda fila de B es $b + \alpha a$ para algún $\alpha > 0$. Se ha visto que:

- $|\det(A)| = \acute{A}rea(\mathcal{P}_{a,b})$, que es el área del paralelogramo determinado por a y b; y

- $|\det(B)| = \acute{A}rea(\mathcal{P}_{a,b+\alpha a})$, que es el área del paralelogramo determinado por a y $b + \alpha a$.

En una gráfica en el plano euclídeo $\mathbb{R}^2$ se puede observar fácilmente que el área $\acute{A}rea(\mathcal{P}_{a,b+\alpha a})$ coincide con $\acute{A}rea(\mathcal{P}_{a,b})$ (las longitudes de las bases y las alturas de ambos paralelogramos coinciden).

Proposición 7.4 (Efecto al aplicar operaciones elementales en las filas)**.** Sea $A \in \mathbb{K}^{n\times n}$ con $n \geq 2$. Entonces

(OEI) $\det(E_{ij}A) = -\det(A)$, siendo E_{ij} una operación elemental de intercambio. En particular, $\det(E_{ij}) = -1$.

(OEII) $\det(E_i(k)A) = k\det(A)$, $\forall k \in \mathbb{K}$, $k \neq 0$. En particular, $\det(E_i(k)) = k$.

(OEIII) $\det(E_{ij}(k)A) = \det(A)$, $\forall k \in \mathbb{K}$. En particular, $\det(E_{ij}(k)) = 1$.

Además, si $E_1, E_2, \ldots, E_s \in \mathbb{K}^{n\times n}$ son matrices elementales por filas entonces

$$\det(E_s \ldots E_2E_1A) = \det(E_s)\ldots\det(E_2)\det(E_1)\det(A). \tag{7.4}$$

Por último, si $E \in \mathbb{K}^{n\times n}$ es una matriz elemental por filas entonces

$$\det(E^t) = \det(E). \tag{7.5}$$

Demostración. Para demostrar las propiedades (OEI), (OEII) y (OEIII), basta con expresar las operaciones elementales subyacentes en la Proposición 7.3, la Proposición 7.2 y el Corolario 7.3 en términos de matrices elementales, como se probó en la Proposición 5.7.

Los casos particulares se obtienen todos de particularizar la matriz A a la identidad I_n.

La fórmula (7.4) se prueba por inducción sobre la cantidad s de matrices elementales. En efecto, si $s = 1$, de los resultados anteriores se obtiene

$$\begin{aligned}
\det(E_{ij}A) &= -\det(A) &= \det(E_{ij})\det(A),\\
\det(E_i(k)A) &= k\det(A) &= \det(E_i(k))\det(A),\\
\det(E_{ij}(k)A) &= \det(A) &= \det(E_{ij}(k))\det(A),
\end{aligned}$$

con lo cual la propiedad está probada. Este caso se puede resumir de la siguiente forma: Si $E \in \mathbb{K}^{n\times n}$ es una matriz elemental por filas entonces

$$\det(EA) = \det(E)\det(A).$$

Sea $s > 1$ y supóngase que la propiedad es cierta para $s-1$ matrices elementales por filas. Se quiere probar que si $E_1, E_2, \ldots, E_s \in \mathbb{K}^{n\times n}$ son matrices elementales por filas entonces

$$\det(E_s \ldots E_2 E_1 A) = \det(E_s)\ldots\det(E_2)\det(E_1)\det(A).$$

En efecto, utilizando la propiedad asociativa del producto de matrices y el caso base ($s = 1$) se tiene

$$\begin{aligned}
\det(E_s E_{s-1} \ldots E_2 E_1 A) &= \det(E_s(E_{s-1} \ldots E_2 E_1 A)) \\
&= \det(E_s)\det(E_{s-1} \ldots E_2 E_1 A) \\
&= \det(E_s)[\det(E_{s-1})\ldots\det(E_2)\det(E_1)\det(A)] \\
&= \det(E_s)\det(E_{s-1})\ldots\det(E_2)\det(E_1)\det(A),
\end{aligned}$$

donde en el penúltimo paso se ha aplicado la hipótesis de inducción y la asociatividad en $\mathbb{K}$.

De la Proposición 5.6 se deducen inmediatamente que los valores de los determinantes de las traspuestas de las matrices elementales son $\det([E_{ij}]^t) = -1$, $\det([E_i(k)]^t) = k$ y $\det([E_{ij}(k)]^t) = 1$, que coinciden con los determinantes de las respectivas matrices sin trasponer. $\square$

Observación 7.2. Del resultado anterior se observa que el determinante de las matrices elementales es siempre un número no nulo y de la Proposición 5.5 se sabe que son todas invertibles.

Determinante y existencia de la inversa de una matriz

La observación anterior es un caso particular del siguiente resultado general[10].

[10] Una interpretación geométrica de este resultado se puede ver en la página 295 bajo el título: **Una aplicación de la fórmula de Cauchy-Binet**.

Teorema 7.1 (Existencia de la inversa de una matriz)**.** Sea $A \in \mathbb{K}^{n\times n}$. Entonces

$$A \text{ es invertible} \qquad \Leftrightarrow \qquad \det(A) \neq 0.$$

Demostración. ($\Rightarrow$) Si A es una matriz invertible, por el Teorema 5.1, existen matrices elementales por filas $E_1, E_2, \ldots, E_s \in \mathbb{K}^{n\times n}$ tales que $A = E_1E_2\cdots E_s$. Por la Proposición 7.4,

$$\det(A) = \det(E_1E_2\cdots E_s) = \det(E_1)\det(E_2)\cdots\det(E_s).$$

Por la Observación 7.2, $\det(E_i) \neq 0$, para todo $i = 1, 2, \ldots, s$. Al ser $\mathbb{K}$ un cuerpo, la Proposición 1.2 asegura que $\mathbb{K}$ no tiene divisores de cero. Por lo tanto, $\det(A) = \det(E_1)\det(E_2)\cdots\det(E_s) \neq 0$.

($\Leftarrow$) Se supone que $\det(A) \neq 0$. Por el Teorema 5.1, probar que A es invertible es equivalente a probar que $R_A = I_n$, siendo R_A la forma escalonada reducida por filas de A. Se supone, por reducción al absurdo, que $R_A \neq I_n$. Este hecho implica que R_A debe tener (al menos) una fila de ceros. El Corolario 7.1 garantiza que $\det(R_A) = 0$. Ahora, como $R_A = E_s\cdots E_2E_1A$ para ciertas matrices elementales por filas $E_1, E_2, \ldots, E_s \in \mathbb{K}^{n\times n}$, por la fórmula (7.4) de la Proposición 7.4, se tiene que

$$\det(R_A) = \det(E_s\cdots E_2E_1A) = \det(E_s)\cdots\det(E_2)\det(E_1)\det(A).$$

Al ser $\det(E_i) \neq 0$ para todo $i = 1, 2, \ldots, s$, es $\det(E_s)\cdots\det(E_2)\det(E_1) \neq 0$ (pues en $\mathbb{K}$ no hay divisores de cero). Por lo tanto,

$$0 = \det(R_A) = [\det(E_s)\cdots\det(E_2)\det(E_1)]\det(A) \quad \Rightarrow \quad \det(A) = 0,$$

que es una contradicción. Esta contradicción proviene de suponer que A no es invertible (o que $R_A \neq I_n$), con lo cual se ha probado que A es invertible. $\square$

En la definición de matriz invertible también se usaron los adjetivos regular y no singular. Se dice que una matriz $A \in \mathbb{K}^{n\times n}$ se llama **singular** o **no regular** si $\det(A) = 0$; en caso contrario se llama **no singular** o **regular**.

Con singular, se expresa que $\det(A)$ es un elemento que carece de inverso en el cuerpo $\mathbb{K}$. El contrarrecíproco del Teorema 7.1 queda expresado como

$$A \text{ es no invertible} \qquad \Leftrightarrow \qquad A \text{ es singular.}$$

Este hecho justifica el nombre de regulares o no singulares también atribuido a las matrices invertibles.

Ejemplo 7.5. *Analizar si las siguientes matrices son o no invertibles:*

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

$\lhd$ Los determinantes de estas matrices son $\det(A) = 0$ (tras observar que la tercera fila es -11 veces la primera más -3 veces la segunda y aplicar la aditividad y la Proposición 7.3) y $\det(B) = 2 \neq 0$ (por ser triangular superior). Luego, A carece de inversa (es singular) y B es invertible (regular o no singular). $\rhd$

Determinante y producto de matrices

En la fórmula (7.4) se probó que el determinante de un producto de dos matrices (cuadradas del mismo tamaño) es el producto de los determinantes de cada una, siempre que una de ellas (a la izquierda) sea una matriz elemental. Es decir, si se tienen $A, E \in \mathbb{K}^{n\times n}$ con E matriz elemental entonces

$$\det(EA) = \det(E)\det(A).$$

Esta propiedad es válida en general, sin ser E necesariamente una matriz elemental y se conoce como determinante de un producto de dos matrices o fórmula de Cauchy-Binet.

Teorema 7.2 (Fórmula de Cauchy-Binet)**.** Sean $A, B \in \mathbb{K}^{n\times n}$. Entonces

$$\det(AB) = \det(A)\det(B).$$

Demostración. A tenor de los resultados precedentes, se dividirá la prueba en dos casos:

- A es no singular: Como $\det(A) \neq 0$, por el Teorema 7.1, A es invertible. Aplicando el Teorema 5.1, se tiene que existen matrices elementales por filas $E_1, E_2, \ldots, E_s \in \mathbb{K}^{n\times n}$ tales que $A = E_1E_2\cdots E_s$. Por la fórmula (7.4),

$$\begin{aligned} \det(AB) &= \det(E_1E_2\cdots E_sB) \\ &= \det(E_1)\det(E_2)\cdots\det(E_s)\det(B) \\ &= \det(E_1E_2\cdots E_s)\det(B) \\ &= \det(A)\det(B). \end{aligned}$$

- A es singular: En este caso, $\det(A) = 0$, y por tanto $\det(A)\det(B) = 0$, independientemente del valor de $\det(B)$.

  Como A no es invertible, por el Teorema 5.1, $R_A \neq I_n$, donde R_A es la forma escalonada reducida por filas de A. Luego, existen matrices elementales por filas $E_1, E_2, \ldots, E_s \in \mathbb{K}^{n\times n}$ tales que $E_s\cdots E_2E_1A = R_A$, donde la forma escalonada reducida por filas R_A tiene al menos una fila de ceros. Luego, R_AB también tiene (al menos) una fila de ceros. Por el Corolario 7.1, $\det(R_AB) = 0$. Aplicando la fórmula (7.4),

$$0 = \det(E_s\cdots E_2E_1AB) = \det(E_s)\cdots\det(E_2)\det(E_1)\det(AB).$$

  Al ser $\det(E_i) \neq 0$, para todo $i = 1, 2, \ldots, s$, se llega a que $\det(AB) = 0$.

Por lo tanto, la igualdad requerida en la tesis se cumple en ambos casos. $\square$

Es interesante observar la sencillez de la fórmula de Cauchy-Binet a pesar de la complicación de la definición de determinante y de la cantidad de cálculos que involucra un producto de matrices.

Corolario 7.4 (Determinante de la inversa de una matriz)**.** Sea $A \in \mathbb{K}^{n\times n}$. Si A es invertible entonces

$$\det(A^{-1}) = \frac{1}{\det(A)}.$$

Demostración. La hipótesis garantiza la existencia de la matriz A^{-1}. Aplicando la fórmula de Cauchy-Binet a $A^{-1}A = I_n$ se tiene

$$\det(A^{-1})\det(A) = \det(A^{-1}A) = \det(I_n) = 1.$$

Del Teorema 7.1 se tiene que $\det(A) \neq 0$, lo que permite obtener la expresión $\det(A^{-1}) = \frac{1}{\det(A)}$. □

Observar que en el último paso de la demostación anterior no haría falta aplicar el Teorema 7.1. Alcanzaría con observar que como en $\mathbb{K}$ no hay divisores de cero, y el producto de los elementos $\det(A^{-1})$ y $\det(A)$ es no nulo, entonces ambos deben serlo (un elemento es el inverso multiplicativo del otro).

A continuación se muestra una nueva interpretación geométrica del determinante de una matriz.

Una aplicación de la fórmula de Cauchy-Binet

Sean u y v dos vectores de $\mathbb{R}^2$ no nulos y tal que ninguno sea múltiplo del otro. El efecto de multiplicar una matriz

$$A = \begin{bmatrix} a & c \\ b & d \end{bmatrix} \in \mathbb{R}^{2\times 2}$$

por cada uno de estos vectores es el de transformarlos en los vectores Au y Av, que también pertenecen a $\mathbb{R}^2$. Disponiendo estos últimos vectores por columnas en una matriz y utilizando la Proposición 2.9 se obtiene

$$\begin{bmatrix} Au & Av \end{bmatrix} = A \begin{bmatrix} u & v \end{bmatrix}.$$

Aplicando determinantes y la fórmula de Cauchy-Binet se puede escribir

$$\left|\det\left(\begin{bmatrix} Au & Av \end{bmatrix}\right)\right| = |\det(A)| \left|\det\left(\begin{bmatrix} u & v \end{bmatrix}\right)\right|, \tag{7.6}$$

donde se ha tomado, además, valor absoluto en ambos miembros y se ha utilizado la propiedad distributiva del valor absoluto respecto del producto de dos números reales.

Se recuerda que una interpretación geométrica del determinante permite establecer que

$$\left|\det\left(\begin{bmatrix} u & v \end{bmatrix}\right)\right| = \text{Área}(\mathcal{P}_{u,v}),$$

siendo $\mathcal{P}_{u,v}$ el paralelogramo determinado por los vectores u y v. Dicha área es no nula pues u y v son no nulos y ninguno es múltiplo del otro.

Algo similar ocurrirá con

$$\left|\det\left(\begin{bmatrix} Au & Av \end{bmatrix}\right)\right| = \text{Área}(\mathcal{P}_{Au,Av}),$$

siempre que los vectores Au y Av sean no nulos y ninguno de ellos múltiplo del otro.

Si $\det(A) \neq 0$, se obtiene que $\text{Área}(\mathcal{P}_{Au,Av}) > 0$ y la relación (7.6) indica que

$$\text{Área}(\mathcal{P}_{Au,Av}) = |\det(A)|\, \text{Área}(\mathcal{P}_{u,v}),$$

de donde $|\det(A)|$ es un *factor de proporcionalidad de áreas*, que indica por cuánto se debe multiplicar el área del paralelogramo original para obtener la del transformado, lo que ofrece una nueva interpretación geométrica del concepto de determinante. Estos mismos cálculos también permiten asegurar que si se transforma un paralelogramo con área estrictamente positiva mediante una matriz invertible, el paralelogramo obtenido también tendrá área estrictamente positiva[11].

[11] Este hecho se utiliza en Cálculo Integral de funciones de dos variables.

En el caso (degenerado) en que sea $\det(A) = 0$, se tiene que $\text{Área}(\mathcal{P}_{Au,Av}) = 0$, lo que produce que uno de los vectores Au o Av sea múltiplo del otro o bien que al menos uno de ellos sea nulo.

Determinante y traspuesta de una matriz

La fórmula (7.5) de la página 290 indica que el determinante de una matriz elemental coincide con el de su traspuesta. Esta propiedad es válida, no sólo para matrices elementales sino, en general.

Teorema 7.3 (Determinante de la matriz traspuesta de una matriz)**.**
Sea $A \in \mathbb{K}^{n\times n}$. Entonces

$$\det(A^t) = \det(A).$$

Demostración. De nuevo, se dividirá la prueba en dos casos:

- A es no singular: Como $\det(A) \neq 0$, por el Teorema 7.1, A es invertible. Por el Teorema 5.1, existen matrices elementales por filas $E_1, E_2, \ldots, E_s \in \mathbb{K}^{n\times n}$ tales que $A = E_1E_2 \ldots E_s$. Luego,

$$A^t = (E_1E_2 \ldots E_s)^t = E_s^t \ldots E_2^t E_1^t.$$

De la fórmula de Cauchy-Binet extendida por inducción a s matrices (véase Ejercicio 8 de la página 325), se tiene que

$$\begin{aligned}
\det(A^t) &= \det(E_s^t \ldots E_2^t E_1^t) \\
&= \det(E_s^t) \ldots \det(E_2^t) \det(E_1^t) \\
&= \det(E_s) \ldots \det(E_2) \det(E_1) \\
&= \det(E_1) \det(E_2) \ldots \det(E_s) \\
&= \det(E_1E_2 \ldots E_s) \\
&= \det(A).
\end{aligned}$$

- A es singular: En este caso, $\det(A) = 0$. Por el Teorema 7.1, A no es invertible. Luego, $\mathrm{rg}(A) < n$ de acuerdo al Teorema 5.1. Por la Proposición 6.5, $\mathrm{rg}(A^t) = \mathrm{rg}(A) < n$. Esto garantiza que A^t no es invertible y, por tanto, $\det(A^t) = 0$.

Así, la igualdad requerida en la tesis se cumple en ambos casos. □

Desarrollo por cualesquiera de las columnas o cualesquiera de las filas

Por último, se probará una propiedad que indica que la primera columna no es especial para realizar el desarrollo de Laplace por los elementos de ella, y dicho desarrollo se puede realizar utilizando una columna cualquiera.

Teorema 7.4 (Desarrollo por los elementos de cualquier columna)**.** Sea $A \in \mathbb{K}^{n\times n}$. Entonces para cualquier $s \in \{2, 3, \ldots, n\}$ se tiene que

$$\det(A) = \sum_{i=1}^{n} a_{is} A_{is}.$$

Demostración. Sea $A = [a_{ij}] \in \mathbb{K}^{n\times n}$ (y, por tanto, n está fijo) y sea

$$B := \begin{bmatrix} a_{1s} & a_{11} & \ldots & a_{1,s-1} & a_{1,s+1} & \ldots & a_{1n} \\ a_{2s} & a_{21} & \ldots & a_{2,s-1} & a_{2,+1} & \ldots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{ns} & a_{n1} & \ldots & a_{n,s-1} & a_{n,s+1} & \ldots & a_{nn} \end{bmatrix} \in \mathbb{K}^{n\times n}$$

la matriz obtenida a partir de A, moviendo su columna s-ésima a la primera posición. Al cambiar la columna s-ésima con la $(s-1)$-ésima se realiza un intercambio, si ahora se la cambia con la $(s-2)$-ésima se realizan 2 intercambios, y continuando así, intercambiando con cada columna consecutiva que le precede hasta llegar la primera posición, se debe efecuar un total de $s-1$ cambios. Dado que el determinante es una aplicación alternada en las filas (Proposición 7.3) y el determinante de una matriz coincide con el de su traspuesta (Teorema 7.3), se tiene que cada intercambio de columna obliga a un cambio en el signo del determinante, es decir,

$$\det(B) = \det(B^t) = (-1)^{s-1}\det(A^t) = (-1)^{s-1}\det(A).$$

Por definición[12], si se llama M_{ij}^B al menor complementario asociado a la posición (i,j) de la matriz B, se tiene

$$\begin{aligned}(-1)^{s-1}\det(A) &= \det(B)\\ &= a_{1s}\det(M_{11}^B) - a_{2s}\det(M_{21}^B) + \cdots + (-1)^{n+1}a_{ns}\det(M_{n1}^B).\end{aligned}$$

Ahora se debe escribir cada uno de estos menores M_{i1}^B, para $i = 1, 2, \ldots, n$, en términos de los menores M_{ij}^A de A. En efecto, para construir M_{i1}^B se debe eliminar la primera columna de B y su fila i-ésima. Pero es fácil observar que se obtiene la misma submatriz que al eliminar la columna s-ésima de A y su fila i-ésima[13]. Esto es, $M_{i1}^B = M_{is}^A$. Así,

$$(-1)^{s-1}\det(A) = a_{1s}\det(M_{1s}^A) - a_{2s}\det(M_{2s}^A) + \cdots + (-1)^{n+1}a_{ns}\det(M_{ns}^A).$$

Sólo queda por arreglar los signos de modo que aparezcan los determinantes de los menores M_{ij}^A con el signo adecuado en los complementos algebraicos A_{ij}. Para ello, se debe tener en cuenta que: $A_{1s} = (-1)^{1+s}\det(M_{1s}^A)$, $A_{2s} = (-1)^{2+s}\det(M_{2s}^A)$, …, $A_{ns} = (-1)^{n+s}\det(M_{ns}^A)$. Con ello,

$$\begin{aligned}&(-1)^{s-1}\det(A) =\\ &= (-1)^{1+s}a_{1s}A_{1s} - (-1)^{2+s}a_{2s}A_{2s} + \cdots + (-1)^{n+s}(-1)^{n+1}a_{ns}A_{ns}.\end{aligned}$$

Multiplicando ambos miembros de la igualdad anterior por la expresión $(-1)^{1-s}$, se cancelan todas las eses de las potencias de -1 quedando:

$$\begin{aligned}\det(A) &= (-1)^{1+1}a_{1s}A_{1s} - (-1)^{2+1}a_{2s}A_{2s} + \cdots + (-1)^{2n+1+1}a_{ns}A_{ns}\\ &= (-1)^2a_{1s}A_{1s} - (-1)^3a_{2s}A_{2s} + \cdots + (-1)^{2n+2}a_{ns}A_{ns}\\ &= a_{1s}A_{1s} + a_{2s}A_{2s} + \cdots + a_{ns}A_{ns},\end{aligned}$$

[12]Recordar que el determinante se define a partir del desarrollo de Laplace por los elementos de la primera columna de la matriz.

[13]Si, por ejemplo, en una matriz de tamaño 4×4 se lleva la tercera columna de A a la primera posición, al suprimir la primera columna de B quedan las mismas columnas que en A una vez suprimida su tercera columna, es decir quedan las columnas 1, 2 y 4 de A.

que es el desarrollo de $\det(A)$ por la s-ésima columna. Como el resultado ha sido demostrado para un valor fijo de n, teniendo en cuenta que este puede ser arbitrario, se ha probado el resultado para todo $n \in \mathbb{N}$. $\square$

Como se prueba a continuación, los resultados anteriores permiten establecer que es posible realizar el desarrollo de Laplace a lo largo de cualquier fila de la matriz dada.

Corolario 7.5 (Desarrollo por los elementos de cualquier fila)**.** Sea $A \in \mathbb{K}^{n\times n}$. Entonces para cualquier $s \in \{1, 2, 3, \dots, n\}$ se tiene que

$$\det(A) = \sum_{j=1}^{n} a_{sj} A_{sj}.$$

Demostración. Sean

$$A = [a_{ij}] \in \mathbb{K}^{n\times n} \qquad \text{y} \qquad A^t = [a^t_{ij}] = [a_{ji}] \in \mathbb{K}^{n\times n}.$$

Se denota A^t_{ks} al complemento algebraico (¡no representa traspuesta!) asociado a la posición (k, s) de la matriz A^t y M^t_{ks} su menor complementario (¡no representa traspuesta!).

Es claro que al eliminar[14] la fila k-ésima y la columna s-ésima de la matriz A^t se obtiene la matriz traspuesta de eliminar la fila s-ésima y la columna k-ésima de A, es decir $M^t_{ks} = (M_{sk})^t$, donde la última es la traspuesta de la matriz M_{sk}. Por el Teorema 7.3 se tiene que $\det(M^t_{ks}) = \det((M^A_{sk})^t) = \det(M^A_{sk})$.

Ahora, el Teorema 7.3 y el Teorema 7.4 aseguran que desarrollando por

[14] Para comprender este paso sería recomendable poner un ejemplo de una matriz de tamaño al menos 3×3.

la columna s-ésima de A^t se tiene

$$\begin{aligned}
\det(A) &= \det(A^t) \\
&= \sum_{j=1}^{n} a^t_{js} A^t_{js} \\
&= \sum_{j=1}^{n} a_{sj} (-1)^{j+s} \det(M^t_{js}) \\
&= \sum_{j=1}^{n} a_{sj} (-1)^{s+j} \det(M^A_{sj}) \\
&= \sum_{j=1}^{n} a_{sj} A_{sj}.
\end{aligned}$$

□

Regla de cálculo de determinantes

Utilizando las propiedades demostradas en la Proposición 7.4, el determinante cambia de signo al aplicar una operación de intercambio, queda multiplicado por el escalar k si se aplica una operación elemental de escalado, y se mantiene invariante si se aplica una operación de eliminación. Luego, es posible aplicar operaciones elementales sobre una matriz para hallar su determinante, de forma similar a como se hace en el método de eliminación de Gauss, con el objetivo de conseguir una matriz triangular cuyo determinante, se ha probado que, es el producto de los elementos de la diagonal. Este método es general y permite calcular determinantes de órdenes pequeños.

Ejemplo 7.6. *Calcular el determinante de la matriz*

$$A = \begin{bmatrix} 2 & 3 & 0 & 0 \\ 9 & 7 & 3 & 2 \\ -5 & 3 & 3 & 0 \\ 1 & 1 & 1 & 0 \end{bmatrix}.$$

$\triangleleft$ Aplicando operaciones elementales se tiene que:

$$\begin{vmatrix} 2 & 3 & 0 & 0 \\ 9 & 7 & 3 & 2 \\ -5 & 3 & 3 & 0 \\ 1 & 1 & 1 & 0 \end{vmatrix} = -\begin{vmatrix} 1 & 1 & 1 & 0 \\ 2 & 3 & 0 & 0 \\ 9 & 7 & 3 & 2 \\ -5 & 3 & 3 & 0 \end{vmatrix} = -\begin{vmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & -2 & 0 \\ 0 & -2 & -6 & 2 \\ 0 & 8 & 8 & 0 \end{vmatrix} =$$

$$= -(-2)8\begin{vmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & -2 & 0 \\ 0 & 1 & 3 & -1 \\ 0 & 1 & 1 & 0 \end{vmatrix} = 16\begin{vmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & -2 & 0 \\ 0 & 0 & 5 & -1 \\ 0 & 0 & 3 & 0 \end{vmatrix} = 48\begin{vmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & -2 & 0 \\ 0 & 0 & 5 & -1 \\ 0 & 0 & 1 & 0 \end{vmatrix}$$

$$= -48\begin{vmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & -2 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 5 & -1 \end{vmatrix} = -48\begin{vmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & -2 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{vmatrix} = (-48)(-1) = 48.$$

Otra estrategia para realizar el mismo cálculo es la siguiente. La línea (fila o columna) que más ceros tiene es la cuarta columna, por lo tanto, conviene desarrollar por los elementos de ella:

$$\det(A) = 2(-1)^{2+4}\begin{vmatrix} 2 & 3 & 0 \\ -5 & 3 & 3 \\ 1 & 1 & 1 \end{vmatrix}.$$

Como figura un 0 en la posición $(1, 3)$, se eliminará alguno de los elementos no nulos, por ejemplo, de la tercera columna. Si a la segunda fila se le suma la tercera multiplicada previamente por -3 el determinante no cambia, con lo cual queda

$$\det(A) = 2\begin{vmatrix} 2 & 3 & 0 \\ -8 & 0 & 0 \\ 1 & 1 & 1 \end{vmatrix},$$

que se puede desarrollar fácilmente por la tercera columna mediante

$$\det(A) = 2(-1)^{3+3}\begin{vmatrix} 2 & 3 \\ -8 & 0 \end{vmatrix} = 48.$$

$\triangleright$

7.4. Aplicaciones de los determinantes

Hasta ahora se ha visto como aplicación de los determinantes, el cálculo del área de un paralelogramo en $\mathbb{R}^2$ o el volumen de un paralelepípedo en $\mathbb{R}^3$.

En las próximas subsecciones se presentan tres aplicaciones más de los determinantes que permiten: (a) calcular la matriz inversa, (b) resolver sistemas de ecuaciones lineales, y (c) calcular el rango de una matriz.

7.4.1. Cálculo de la matriz inversa

Para $A = [a_{ij}] \in \mathbb{K}^{n\times n}$ se han definido los **complementos algebraicos** o **cofactores** asociados a cada posición (i, j) como

$$A_{ij} = (-1)^{i+j} \det(M_{ij}),$$

siendo el **menor complementario** correspondiente a la posición (i, j) la submatriz M_{ij} que se obtiene al eliminar la fila i-ésima y la columna j-ésima de A.

Se presenta una importante propiedad de los complementos algebraicos.

Lema 7.1 (Lema de las filas paralelas)**.** Sea $A = \begin{bmatrix} a_{ij} \end{bmatrix} \in \mathbb{K}^{n\times n}$ y sea A_{ij} el complemento algebraico del elemento de A correspondiente a la posición (i, j). Si $t \in \{1, 2, \ldots, n\}$ está fijo, se cumple que:

$$a_{k1}A_{t1} + a_{k2}A_{t2} + \cdots + a_{kn}A_{tn} = \begin{cases} \det(A), & \text{si } k = t, \\ 0, & \text{si } k \neq t, \end{cases}$$

para todo $k \in \{1, 2, \ldots, n\}$.

Demostración. Si $k = t$ entonces, por el Corolario 7.5,

$$a_{t1}A_{t1} + a_{t2}A_{t2} + \cdots + a_{tn}A_{tn} = \det(A),$$

que es el propio desarrollo del determinante de A por la fila t-ésima.

Si $k \neq t$ entonces se debe considerar la matriz B que tiene en la fila k-ésima los elementos $a_{k1}, a_{k2}, ..., a_{kn}$ y en la fila t-ésima esos mismos elementos, y en el resto B coincide con A, es decir,

$$B = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{k1} & \dots & a_{kn} \\ \vdots & \ddots & \vdots \\ a_{k1} & \dots & a_{kn} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{bmatrix} \begin{array}{l} \\ \\ \longleftarrow \quad \text{Fila} \quad k - \text{ésima} \\ \\ \longleftarrow \quad \text{Fila} \quad t - \text{ésima} \\ \\ \\ \end{array}$$

De ese modo, por un lado, $\det(B) = 0$ por tener dos filas iguales y, por otro, el desarrollo de Laplace por la fila t-ésima proporciona $\det(B) = a_{k1}A_{t1} + a_{k2}A_{t2} + \cdots + a_{kn}A_{tn}$ pues el elemento de A que acompaña al complemento algebraico A_{ts} es a_{ks}. □

Este lema se puede parafrasear diciendo que: la suma de los productos de los elementos de una línea (fila o columna) de A por los complementos algebraicos de sus respectivas posiciones es igual al determinante de A, mientras que, la suma de los productos de los elementos de una línea (fila o columna) de A por los complementos algebraicos de las respectivas posiciones de una línea paralela es 0.

Definición 7.2. Sea $A = [a_{ij}] \in \mathbb{K}^{n\times n}$. Se llama matriz **adjunta** de la matriz A, y se denota $\mathrm{Adj}(A)$, a la matriz formada por los complementos algebraicos de los elementos a_{ij} en sus respectivas posiciones. En símbolos,

$$\mathrm{Adj}(A) = [A_{ij}] = \begin{bmatrix} A_{11} & A_{12} & \dots & A_{1n} \\ A_{21} & A_{22} & \dots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1} & A_{n2} & \dots & A_{nn} \end{bmatrix}.$$

La matriz adjunta conjuntamente con el determinante permiten dar una fórmula para calcular la inversa de una matriz.

Teorema 7.5. Si $A \in \mathbb{K}^{n\times n}$ es invertible entonces

$$A^{-1} = \frac{1}{\det(A)}[\mathrm{Adj}(A)]^t.$$

Demostración. Es necesario calcular el producto

$$A\,[\mathrm{Adj}(A)]^t \quad = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \begin{bmatrix} A_{11} & A_{21} & \dots & A_{n1} \\ A_{12} & A_{22} & \dots & A_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ A_{1n} & A_{2n} & \dots & A_{nn} \end{bmatrix}.$$

Aplicando el Lema 7.1, se tiene que los elementos de la diagonal son

$$a_{j1}A_{j1} + a_{j2}A_{j2} + \cdots + a_{jn}A_{jn} = \det(A), \qquad \text{para } j = 1, 2, \dots, n,$$

y el elemento de la posición (j, k) (es decir, de fuera de la diagonal) es

$$a_{j1}A_{k1} + a_{j2}A_{k2} + \cdots + a_{jn}A_{kn} = 0, \qquad \text{para } j, k = 1, 2, \dots, n, \text{ con } j \neq k.$$

Luego,

$$A\,[\mathrm{Adj}(A)]^t = \begin{bmatrix} \det(A) & 0 & \dots & 0 \\ 0 & \det(A) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \det(A) \end{bmatrix} = \det(A) I_n.$$

Al ser A invertible, es regular con lo que $\det(A) \neq 0$ y así se obtiene que

$$A\left[\frac{1}{\det(A)}\,[\mathrm{Adj}(A)]^t\right] = I_n,$$

de donde se concluye que $A^{-1} = \frac{1}{\det(A)}\,[\mathrm{Adj}(A)]^t$. □

Ejemplo 7.7. *Hallar la inversa de la matriz*

$$A = \begin{bmatrix} 3 & 0 & 0 \\ 7 & 3 & 2 \\ 1 & 1 & 0 \end{bmatrix}.$$

$\lhd$ En este caso, realizando el desarrollo de Laplace por la primera fila de A se tiene que $\det(A) = -6 \neq 0$, con lo cual A es invertible.

Se debe calcular la matriz adjunta de A, a partir de los complementos algebraicos, donde, por ejemplo,

$$A_{11} = (-1)^{1+1} \det\left(\begin{bmatrix} 3 & 2 \\ 1 & 0 \end{bmatrix} \right) = -2.$$

Los restantes complementos algebraicos se calculan de forma similar.

Por el Teorema 7.5,

$$\begin{aligned} A^{-1} &= \frac{1}{\det(A)}[\mathrm{Adj}(A)]^t \\ &= \frac{1}{-6}\begin{bmatrix} -2 & 0 & 0 \\ 2 & 0 & -6 \\ -4 & -3 & 9 \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{3} & 0 & 0 \\ -\frac{1}{3} & 0 & 1 \\ \frac{2}{3} & \frac{1}{2} & -\frac{3}{2} \end{bmatrix}. \end{aligned}$$

$\rhd$

7.4.2. Regla de Cramer para resolver sistemas lineales

El método presentado a continuación utiliza determinantes para resolver sistemas de ecuaciones lineales con la misma cantidad de ecuaciones que de incógnitas, siempre que la matriz de coeficientes sea no singular.

Teorema 7.6 (Regla de Cramer). Sea $A \in \mathbb{K}^{n\times n}$ una matriz no singular particionada por columnas como $A = \begin{bmatrix} c_1 & c_2 & \dots & c_n \end{bmatrix}$ y sea $b \in \mathbb{K}^{n\times 1}$. Entonces el sistema lineal $Ax = b$ tiene solución única y denotando por

$$A_i := \begin{bmatrix} c_1 & \dots & c_{i-1} & b & c_{i+1} & \dots & c_n \end{bmatrix} \quad \text{y} \quad x := \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix},$$

donde la matriz A_i se obtiene cambiando la columna i-ésima de A por la de términos independientes b, se tiene

$$x_i = \frac{\det(A_i)}{\det(A)}, \qquad \text{para } i = 1, 2, \dots, n.$$

En consecuencia,

$$x = \frac{1}{\det(A)} \begin{bmatrix} \det(A_1) \\ \det(A_2) \\ \vdots \\ \det(A_n) \end{bmatrix}.$$

Demostración. Si A es una matriz no singular, $\det(A) \neq 0$, con lo cual A es invertible y la única solución del sistema $Ax = b$ es $x = A^{-1}b$. Luego, denotando

$$x := \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \qquad \text{y} \qquad b := \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

del Teorema 7.5 se tiene que

$$
\begin{aligned}
x &= A^{-1}b \\
&= \frac{1}{\det(A)}[\mathrm{Adj}(A)]^t b \\
&= \frac{1}{\det(A)}\begin{bmatrix} A_{11} & A_{21} & \dots & A_{n1} \\ A_{12} & A_{22} & \dots & A_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ A_{1n} & A_{2n} & \dots & A_{nn} \end{bmatrix}\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} \\
&= \frac{1}{\det(A)}\begin{bmatrix} A_{11}b_1 + A_{21}b_2 + \cdots + A_{n1}b_n \\ A_{12}b_1 + A_{22}b_2 + \cdots + A_{n2}b_n \\ \vdots \\ A_{1n}b_1 + A_{2n}b_2 + \cdots + A_{nn}b_n \end{bmatrix} \\
&= \frac{1}{\det(A)}\begin{bmatrix} \det(A_1) \\ \det(A_2) \\ \vdots \\ \det(A_n) \end{bmatrix},
\end{aligned}
$$

donde en el último paso se ha utilizado el Ejercicio 19 de la página 327. □

Ejemplo 7.8. *Resolver el sistema de ecuaciones lineales $Ax = b$ siendo*

$$
A = \begin{bmatrix} 2 & 3 & 0 & 0 \\ 9 & 7 & 3 & 2 \\ -5 & 3 & 3 & 0 \\ 1 & 1 & 1 & 0 \end{bmatrix} \qquad \text{y} \qquad b = \begin{bmatrix} 4 \\ 2 \\ 0 \\ -2 \end{bmatrix}.
$$

◁ Se ha visto en el Ejemplo 7.6 que $\det(A) = 48 \neq 0$. Esto asegura que A es no singular y puede aplicarse la regla de Cramer con lo que

$$
x_i = \frac{\det(A_i)}{\det(A)}, \qquad \text{para} \quad i = 1, 2, 3, 4.
$$

En efecto, calculando los determinantes indicados queda

$$x_1 = \frac{1}{48}\begin{vmatrix} 4 & 3 & 0 & 0 \\ 2 & 7 & 3 & 2 \\ 0 & 3 & 3 & 0 \\ -2 & 1 & 1 & 0 \end{vmatrix} = -\frac{3}{4}, \qquad x_2 = \frac{1}{48}\begin{vmatrix} 2 & 4 & 0 & 0 \\ 9 & 2 & 3 & 2 \\ -5 & 0 & 3 & 0 \\ 1 & -2 & 1 & 0 \end{vmatrix} = \frac{11}{6},$$

$$x_3 = \frac{1}{48}\begin{vmatrix} 2 & 3 & 4 & 0 \\ 9 & 7 & 2 & 2 \\ -5 & 3 & 0 & 0 \\ 1 & 1 & -2 & 0 \end{vmatrix} = -\frac{37}{12}, \qquad x_4 = \frac{1}{48}\begin{vmatrix} 2 & 3 & 0 & 4 \\ 9 & 7 & 3 & 2 \\ -5 & 3 & 3 & 0 \\ 1 & 1 & 1 & -2 \end{vmatrix} = \frac{31}{12}.$$

$\triangleright$

Al involucrar determinantes, la regla de Cramer se podría utilizar en la práctica únicamente para resolver sistemas de ecuaciones lineales de tamaño pequeño, debido a la cantidad de operaciones que involucra. Si bien se incluye como resultado teórico clásico, el método recomendable para resolver un sistema de ecuaciones lineales es el método de eliminación de Gauss.

7.4.3. Cálculo del rango de una matriz

Otra aplicación de los determinantes permite calcular el rango por un método diferente a los vistos anteriormente.

Se ha definido el rango de una matriz $A \in \mathbb{K}^{m\times n}$ como el número de filas no nulas de su forma escalonada reducida por filas R_A, lo que coincide con el número de columnas básicas de R_A.

Para el próximo objetivo será necesario trabajar con unos determinantes especiales que se definen a continuación.

Definición 7.3. Sea $A \in \mathbb{K}^{m\times n}$ y $k \in \mathbb{N}$ con $1 \leq k \leq \min\{m, n\}$. Se llama **menor de orden** k de A al determinante de cualquier submatriz (cuadrada) de tamaño $k \times k$ de A.

Definición 7.4. Sea $A \in \mathbb{K}^{m\times n}$ y $k \in \mathbb{N}$ con $1 \leq k \leq \text{mín}\{m,n\}$. Si M es una submatriz cuadrada de tamaño $k \times k$ de una matriz A, se dice que una **submatriz** M' de A de tamaño $(k+1)\times(k+1)$ **se obtiene orlando**[a] M si M es submatriz de M'. Además[b], se dice que el **menor** $\det(M')$ se obtiene **orlando al menor** $\det(M)$.

[a]Un sinónimo de la palabra *orlar* es *bordear*.
[b]Por abuso de lenguaje.

De la Definición 7.4, se sigue que debe ser $2 \leq k+1 \leq \text{mín}\{m,n\}$.

Puesto que es conocido que $A = O$ si y sólo si $\text{rg}(A) = 0$, para calcular el rango de una matriz, sólo es necesario considerar matrices no nulas. El próximo resultado se debe a F. G. Frobenius.

Proposición 7.5. Sea $A \in \mathbb{K}^{m\times n}$ una matriz no nula y $r \in \mathbb{N}$ con $1 \leq r \leq \text{mín}\{m,n\}$. Las siguientes condiciones son equivalentes:

(a) $\text{rg}(A) = r$,

(b) Existe un menor no nulo de orden r de A tal que todos los menores que se obtienen orlándolo son nulos.

Demostración. (a) $\Rightarrow$ (b) Si $\text{rg}(A) = r$ entonces la forma escalonada reducida por filas R_A de A satisface que $PA = R_A$ para alguna $P \in \mathbb{K}^{m\times m}$ invertible y donde R_A contiene r columnas básicas. Si R_A tiene sus columnas básicas en las posiciones $j_1 < j_2 < \ldots < j_r$ entonces eligiendo las correspondientes columnas $c_{j_1}, c_{j_2}, \ldots, c_{j_r}$ de A, que ocupan esas mismas posiciones, de $PA = R_A$ se tiene, en particular, que

$$P \begin{bmatrix} c_{j_1} & c_{j_2} & \ldots & c_{j_r} \end{bmatrix} = \begin{bmatrix} e_1 & e_2 & \ldots & e_r \end{bmatrix} = \begin{bmatrix} I_r \\ O \end{bmatrix},$$

siendo e_j matrices columna con un 1 en la fila j-ésima y ceros en las demás posiciones.

Hasta ahora se ha seleccionado la submatriz de A con r columnas

$$B := \left[\begin{array}{cccc} c_{j_1} & c_{j_2} & \dots & c_{j_r} \end{array}\right] \in \mathbb{K}^{m\times r}.$$

Como multiplicar por matrices invertibles no altera el rango, de $PB = \left[\begin{array}{c} I_r \\ O \end{array}\right]$ se tiene que $\mathrm{rg}(B) = \mathrm{rg}(PB) = r$.

Por la Proposición 6.5 se sabe que $\mathrm{rg}(B) = \mathrm{rg}(B^t)$. Se puede considerar entonces la forma escalonada reducida por filas R_{B^t} de B^t cumple que $P_1B^t = R_{B^t} \in \mathbb{K}^{r\times m}$, para cierta matriz invertible $P_1 \in \mathbb{K}^{r\times r}$. Al ser $\mathrm{rg}(B^t) = r$, de nuevo se seleccionan las r columnas básicas de R_{B^t}, lo que permite determinar r columnas $\alpha_1, \alpha_2, \dots, \alpha_r$ en B^t tales que $P_1 \left[\begin{array}{cccc} \alpha_1 & \alpha_2 & \dots & \alpha_r \end{array}\right] = I_r$. Puesto que las columnas de B^t son filas de B, tomando ahora

$$M := \left[\begin{array}{cccc} \alpha_1 & \alpha_2 & \dots & \alpha_r \end{array}\right]^t \in \mathbb{K}^{r\times r}$$

se observa que es claramente una submatriz invertible de A y $\det(M)$ es el menor de orden r no nulo buscado.

Falta ver que todos los menores que se obtienen orlando a $\det(M)$ son nulos. En efecto, sea $M' \in \mathbb{K}^{(r+1)\times(r+1)}$ una submatriz orlada a partir de M. Supońgase, por el absurdo, que M' tuviese determinante no nulo. Entonces M' sería invertible y su forma escalonada reducida por filas sería I_{r+1}, con $r+1$ unos principales. Entonces $r = \mathrm{rg}(A) \geq r+1$, que es una contradicción. Luego, todos los menores de orden $r+1$ deben ser nulos.

(b) $\Rightarrow$ (a) Supóngase que $M \in \mathbb{K}^{r\times r}$ es una submatriz cuadrada de A con determinante no nulo tal que todas las submatrices orladas a partir de M tengan determinante nulo. Se debe probar que $\mathrm{rg}(A) = r$. Por la hipótesis, la submatriz M es invertible, con lo que su forma escalonada reducida por filas es I_r y por tanto $\mathrm{rg}(M) = r$, y así, $\mathrm{rg}(A) \geq r$. Si, por el absurdo, fuese $\mathrm{rg}(A) > r$ entonces se tendría que poder encontrar una matriz $M' \in \mathbb{K}^{(r+1)\times(r+1)}$ orlada a partir de M de modo que sea invertible (para que su forma escalonada reducida por filas sea I_{r+1}), eso asegura la existencia de un menor no nulo de orden $r+1$, en contra de la hipótesis. Se ha probado entonces que $\mathrm{rg}(A) = r$. $\square$

Ejemplo 7.9. *Utilizar el procedimiento de la demostración anterior para hallar un menor no nulo de igual orden que el rango de la siguiente matriz*

$$A = \begin{bmatrix} 1 & 2 & 0 & 4 & 5 & 0 \\ 2 & 4 & 1 & 2 & 0 & 1 \\ 4 & 8 & 1 & 10 & 10 & 1 \\ 3 & 6 & 2 & 0 & 0 & 0 \end{bmatrix} \in \mathbb{R}^{4\times 6}.$$

◁ La forma escalonada reducida por filas de A es

$$R_A = \begin{bmatrix} 1 & 2 & 0 & 4 & 0 & 2 \\ 0 & 0 & 1 & -6 & 0 & -3 \\ 0 & 0 & 0 & 0 & 1 & -2/5 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

cuyas columnas básicas se encuentran en la primera, tercera y quinta posiciones y $\operatorname{rg}(A) = 3$. Se considera la submatriz B de A que contiene esas columnas y se calcula la forma de Hermite escalonada reducida por filas de B^t

$$B = \begin{bmatrix} 1 & 0 & 5 \\ 2 & 1 & 0 \\ 4 & 1 & 10 \\ 3 & 2 & 0 \end{bmatrix} \in \mathbb{R}^{4\times 3} \qquad R_{B^t} = \begin{bmatrix} 1 & 0 & 2 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \in \mathbb{R}^{3\times 4}.$$

Las columnas primera, segunda y cuarta de B^t determinan, trasponiendo, la submatriz M invertible de A buscada

$$M = \begin{bmatrix} 1 & 0 & 5 \\ 2 & 1 & 0 \\ 3 & 2 & 0 \end{bmatrix} \in \mathbb{R}^{3\times 3}.$$

▷

La Proposición anterior establece que el rango de una matriz es el mayor orden que pueden alcanzar los menores no nulos de A. Este hecho proporciona el siguiente método para hallar el rango de una matriz.

Método del orlado para calcular el rango de una matriz: Sea $A \in \mathbb{K}^{m\times n}$ una matriz no nula. Hay, por lo menos, una submatriz de A de tamaño 1×1 con determinante no nulo. Luego, $\mathrm{rg}(A) \geq 1$. Se busca una submatriz de tamaño 2×2 cuyo determinante sea distinto de 0. Si no hay, $\mathrm{rg}(A) = 1$. Caso contrario, se orla la matriz encontrada a partir de filas y columnas no utilizadas hasta hallar una submatriz de tamaño 3×3 con determinante no nulo. Si no la hay, $\mathrm{rg}(A) = 3$. Se repite este proceso orlando las matrices encontradas sucesivamente, con filas y columnas no utilizadas previamente, hasta encontrar una submatriz de tamaño $r \times r$ con determinante no nulo y tal que todas las orladas a partir de ella tengan determinante nulo. Se tiene que $\mathrm{rg}(A) = r$.

Ejemplo 7.10. *Calcular el rango de la matriz del ejemplo anterior utilizando el método del orlado.*

$\triangleleft$ Un menor de orden 1 no nulo lo proporciona $M_1 = \begin{bmatrix} 1 \end{bmatrix}$ considerando la primera fila y primera columna de A. Así, $\mathrm{rg}(A) \geq 1$. Orlándolo con la segunda columna y cada una de las filas restantes todos los menores que se obtienen son nulos. Orlando M_1 con la tercera columna y segunda fila se obtiene $M_2 = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix}$, con $\det(M_2) \neq 0$. Así, $\mathrm{rg}(A) \geq 2$. Orlando M_2 con la cuarta columna y las dos filas restantes se obtienen menores nulos. Orlando con la quinta columna y cuarta fila el menor es nulo mientras que la quinta columna con la quinta fila se tiene $M_3 = \begin{bmatrix} 1 & 0 & 5 \\ 2 & 1 & 0 \\ 3 & 2 & 0 \end{bmatrix}$, con $\det(M_3) \neq 0$. Así, $\mathrm{rg}(A) \geq 3$. Pero cualquier menor orlado a partir de este es nulo. Luego, $\mathrm{rg}(A) = 3$. $\triangleright$

7.4.4. Determinantes de matrices por bloques

En este apartado se proporcionan fórmulas que permiten calcular el determinante de una matriz triangular (superior o inferior) por bloques y el determinante de una matriz por bloques sabiendo que uno de los bloques diagonales es invertible.

Proposición 7.6. Sean $A \in \mathbb{K}^{n\times n}$, $B \in \mathbb{K}^{n\times m}$ y $C \in \mathbb{K}^{m\times m}$. Entonces

$$\det\left(\begin{bmatrix} A & B \\ O & C \end{bmatrix}\right) = \det(A)\det(C).$$

Demostración. Sea $m \in \mathbb{N}$ fijo (pero arbitrario). Se procederá por inducción sobre el número natural n, desarrollando por los elementos de la primera columna de la matriz en cuestión.

Si $n = 1$ entonces $A = \begin{bmatrix} a_{11} \end{bmatrix}$ y $B \in \mathbb{K}^{1\times m}$. Mediante el desarrollo de Laplace por la primera columna se tiene

$$\begin{aligned} \det\left(\begin{bmatrix} A & B \\ O & C \end{bmatrix}\right) &= \det\left(\begin{bmatrix} a_{11} & B \\ O & C \end{bmatrix}\right) \\ &= a_{11}A_{11}^S \\ &= a_{11}(-1)^{1+1}\det(C) \\ &= \det(A)\det(C), \end{aligned}$$

donde

$$S := \begin{bmatrix} a_{11} & B \\ O & C \end{bmatrix}.$$

Sea $n > 1$ y supóngase que la propiedad se cumple para todas las matrices de la forma dada siendo A de tamaño $(n-1)\times(n-1)$. Se probará que se cumple para las matrices de la forma dada siendo A de tamaño $n \times n$. En efecto, se considera la matriz particionada

$$S := \begin{bmatrix} A & B \\ O & C \end{bmatrix} \in \mathbb{K}^{(n+m)\times(n+m)},$$

es decir con $A \in \mathbb{K}^{n\times n}$. Desarrollando por la primera columna de S se tiene

$$\det(S) = a_{11}A_{11}^S + a_{21}A_{21}^S + \cdots + a_{n1}A_{n1}^S, \tag{7.8}$$

donde A_{i1}^S denota el complemento algebraico correspondiente a la fila i-ésima y primera columna de la matriz S. Obsérvese que en este desarrollo no se han indicado los ceros de las últimas $m-n$ filas (y primera columna) de S. Se sabe que $A_{i1}^S = (-1)^{i+1}\det(M_{i1}^S)$ donde M_{i1}^S se obtiene de eliminar la fila i-ésima y la primera columna de la matriz S. Se denota M_{i1}^A la matriz que se obtiene de eliminar la fila i-ésima y la primera columna de la matriz A. La forma de las matrices M_{i1}^S es

$$M_{i1}^S = \begin{bmatrix} M_{i1}^A & \widetilde{B}_{i1} \\ O & C \end{bmatrix} \in \mathbb{K}^{((n-1)+m)\times((n-1)+m)},$$

donde las matrices $\widetilde{B}_{i1}$ no se escriben explícitamente pues no intervienen en la demostración. Puesto que M_{i1}^S está en las condiciones de la hipótesis de inducción, se tiene que $\det(M_{i1}^S) = \det(M_{i1}^A)\det(C)$. Sustituyendo en (7.8) se tiene

$$\begin{aligned}
\det(S) &= a_{11}(-1)^{1+1}\det(M_{11}^S) + a_{21}(-1)^{2+1}\det(M_{21}^S) + \cdots + \\
&\quad +a_{n1}(-1)^{n+1}\det(M_{n1}^S) \\
&= a_{11}(-1)^{1+1}\det(M_{11}^A)\det(C) + a_{21}(-1)^{2+1}\det(M_{21}^A)\det(C) + \cdots + \\
&\quad +a_{n1}(-1)^{n+1}\det(M_{n1}^A)\det(C) \\
&= [a_{11}(-1)^{1+1}\det(M_{11}^A) + a_{21}(-1)^{2+1}\det(M_{21}^A) + \cdots + \\
&\quad +a_{n1}(-1)^{n+1}\det(M_{n1}^A)]\det(C) \\
&= \det(A)\det(C),
\end{aligned}$$

donde la última igualdad sigue de la definición de determinante de A.

Por lo tanto, la propiedad es cierta para las matrices de la forma dada siendo A de tamaño $n \times n$, $\forall n \in \mathbb{N}$. Al ser $m \in \mathbb{N}$ un número arbitrario, la propiedad queda establecida. □

La idea de la próxima demostración consiste en realizar operaciones elementales del mismo modo que se hace con matrices dadas por sus elementos, pero ahora utilizando bloques. Se anularán bloques (en lugar de elementos), de modo que los bloques que se obtengan en la nueva matriz sean los que requiere la tesis de la Proposición. Se debe tener en cuenta el orden en que aparecen las matrices puesto que, en general, no conmutan, y que los tamaños estén en concordancia para poder realizar las operaciones.

Proposición 7.7. Sean $A \in \mathbb{K}^{n\times n}$ y $D \in \mathbb{K}^{m\times m}$ con A una matriz invertible. Entonces

$$\det\left(\begin{bmatrix} A & B \\ C & D \end{bmatrix}\right) = \det(A)\det(D - CA^{-1}B).$$

Demostración. Utilizando que A es invertible se puede eliminar el bloque $(2,2)$ que contiene la submatriz C como sigue:

$$\begin{bmatrix} I_n & O \\ -CA^{-1} & I_m \end{bmatrix}\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} A & B \\ O & D - CA^{-1}B \end{bmatrix}.$$

Del Teorema 7.3,

$$\begin{aligned} \det\left(\begin{bmatrix} I_n & O \\ -CA^{-1} & I_m \end{bmatrix}\right) &= \det\left(\begin{bmatrix} I_n & O \\ -CA^{-1} & I_m \end{bmatrix}^t\right) \\ &= \det\left(\begin{bmatrix} I_n & (-CA^{-1})^t \\ O & I_m \end{bmatrix}\right) \\ &= 1, \end{aligned}$$

por tratarse de una matriz triangular superior con unos en la diagonal (Ejemplo 7.4). Aplicando la propiedad que establece que el determinante de un producto es el producto de los determinantes de cada uno de los factores (siempre que se trate de matrices cuadradas) se llega a

$$\begin{aligned}
\det\left(\begin{bmatrix} A & B \\ C & D \end{bmatrix}\right) &= \det\left(\begin{bmatrix} I_n & O \\ -CA^{-1} & I_m \end{bmatrix}\right)\det\left(\begin{bmatrix} A & B \\ C & D \end{bmatrix}\right) \\
&= \det\left(\begin{bmatrix} I_n & O \\ -CA^{-1} & I_m \end{bmatrix}\begin{bmatrix} A & B \\ C & D \end{bmatrix}\right) \\
&= \det\left(\begin{bmatrix} A & B \\ O & D - CA^{-1}B \end{bmatrix}\right) \\
&= \det(A)\det\left(D - CA^{-1}B\right),
\end{aligned}$$

donde en el último paso se ha utilizado la Proposición 7.6. □

Una conclusión interesante del resultado anterior es que si A y la expresión $S := D - CA^{-1}B$ son invertibles entonces $M = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$ es invertible. La expresión $S = D - CA^{-1}B$ se llama el **complemento de Schur** de A en M.

7.4.5. El determinante de Vandermonde

En la práctica suele aparecer la conocida como **matriz de Vandermonde**[15] correspondiente a los elementos $x_1, x_2, \ldots, x_n \in \mathbb{K} - \{0\}$, con $n \geq 2$, que se define por $V(x_1, x_2, \ldots, x_n) = \begin{bmatrix} v_{ij} \end{bmatrix} \in \mathbb{K}^{n\times n}$ donde $v_{ij} := x_i^{j-1}$ para $i, j \in \{1, 2, \ldots, n\}$. Muchas veces se requiere conocer su determinante.

Si $n = 2$, su determinante queda

$$\det(V(x_1, x_2)) = \det\left(\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \end{bmatrix}\right) = x_2 - x_1.$$

Si $n = 3$, para calcular su determinante se utilizan las operaciones ele-

[15] Alexandre-Théophile Vandermonde (1735-1796), matemático francés.

mentales por columna $C_3 + (-x_1)C_2 \to C_3$ y $C_2 + (-x_1)C_1 \to C_2$

$$\begin{aligned}
\det(V(x_1, x_2, x_3)) &= \det\left(\begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \end{bmatrix}\right) \\
&= \det\left(\begin{bmatrix} 1 & 0 & 0 \\ 1 & x_2 - x_1 & x_2^2 - x_1x_2 \\ 1 & x_3 - x_1 & x_3^2 - x_1x_3 \end{bmatrix}\right) \\
&= \det\left(\begin{bmatrix} 1 & 0 & 0 \\ 1 & x_2 - x_1 & (x_2 - x_1)x_2 \\ 1 & x_3 - x_1 & (x_3 - x_1)x_3 \end{bmatrix}\right) \\
&= \det\left(\begin{bmatrix} x_2 - x_1 & (x_2 - x_1)x_2 \\ x_3 - x_1 & (x_3 - x_1)x_3 \end{bmatrix}\right) \\
&= (x_2 - x_1)(x_3 - x_1)\det\left(\begin{bmatrix} 1 & x_2 \\ 1 & x_3 \end{bmatrix}\right) \\
&= (x_2 - x_1)(x_3 - x_1)(x_3 - x_2).
\end{aligned}$$

De estos casos particulares se puede inferir el caso general.

Se recuerda la siguiente notación utilizando el símbolo productorio:

$$\prod_{i=1}^{n} x_i = x_1 \cdot x_2 \cdot \cdots \cdot x_n.$$

Proposición 7.8 (Determinante de la matriz de Vandermonde)**.** Sean $x_1, x_2, \ldots, x_n \in \mathbb{K}$ y $n \in \mathbb{N}$ con $n \geq 2$. Entonces

$$\det\left(\begin{bmatrix} 1 & x_1 & x_1^2 & \ldots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \ldots & x_2^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \ldots & x_n^{n-1} \end{bmatrix}\right) = \prod_{1 \leq i < j \leq n} (x_j - x_i).$$

Demostración. Se demostrará por inducción sobre n.

Si $n = 2$,

$$\det\left(\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \end{bmatrix}\right) = x_2 - x_1,$$

que coincide con

$$\prod_{1 \leq i < j \leq 2} (x_j - x_i) = x_2 - x_1.$$

Sea $n > 2$ y supóngase que el resultado se cumple para toda matriz de Vandermonde de tamaño $(n-1) \times (n-1)$. Se probará para matrices de ese tipo de tamaño $n \times n$.

En efecto, si se considera $V(x_1, x_2, \ldots, x_n)$, se trata de eliminar todos los elementos de la primera fila excepto el de la posición $(1, 1)$ que es 1. Para ello, se aplican las operaciones elementales siguientes y en el orden indicado:

$$\begin{gathered} C_n + (-x_1)C_{n-1} \to C_n \\ C_{n-1} + (-x_1)C_{n-2} \to C_{n-1} \\ \vdots \\ C_2 + (-x_1)C_1 \to C_2 \end{gathered}$$

y, desarrollando luego por la primera fila, queda

$$\begin{aligned}
&\det(V(x_1,x_2,\dots,x_n)) = \\
&= \det\left(\begin{bmatrix} 1 & x_1 & x_1^2 & \dots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \dots & x_2^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \dots & x_n^{n-1} \end{bmatrix}\right) \\
&= \det\left(\begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 & x_2 - x_1 & x_2^2 - x_1x_2 & \dots & x_2^{n-1} - x_1x_2^{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n - x_1 & x_n^2 - x_1x_n & \dots & x_n^{n-1} - x_1x_n^{n-2} \end{bmatrix}\right) \\
&= \det\left(\begin{bmatrix} x_2 - x_1 & (x_2 - x_1)x_2 & \dots & (x_2 - x_1)x_2^{n-2} \\ x_3 - x_1 & (x_3 - x_1)x_3 & \dots & (x_3 - x_1)x_3^{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ x_n - x_1 & (x_n - x_1)x_n & \dots & (x_n - x_1)x_n^{n-2} \end{bmatrix}\right) \\
&= (x_2 - x_1)(x_3 - x_1)\dots(x_n - x_1)\det\left(\begin{bmatrix} 1 & x_2 & \dots & x_2^{n-2} \\ 1 & x_3 & \dots & x_3^{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & \dots & x_n^{n-2} \end{bmatrix}\right).
\end{aligned}$$

Al ser el último determinante el de la matriz $V(x_2,\dots,x_n) \in \mathbb{K}^{(n-1)\times(n-1)}$, por hipótesis de inducción, se tiene que

$$\det(V(x_2,\dots,x_n)) = \prod_{2\le i<j\le n} (x_j - x_i).$$

Finalmente,

$$\begin{aligned}
\det(V(x_1,x_2,\dots,x_n)) &= (x_2 - x_1)(x_3 - x_1)\dots(x_n - x_1)\prod_{2\le i<j\le n}(x_j - x_i) \\
&= \prod_{1\le i<j\le n}(x_j - x_i).
\end{aligned}$$

$\square$

Observación 7.3. Que los valores $x_i \in \mathbb{K}$ sean no nulos, para todo $i \in \{1, 2, \dots, n\}$, en realidad no es una restricción importante. En efecto, si existe algún $i_0 \in \{1, 2, \dots, n\}$ tal que $x_{i_0} = 0$, el determinante de Vandermonde $\det(V(x_1, x_2, \dots, x_n))$ se puede reducir fácilmente a uno de un tamaño menor desarrollando por la fila donde aparecen $n-1$ ceros. El nuevo determinante obtenido es también de Vandermonde.

7.4.6. Un determinante asociado a la matriz de compañía

A un polinomio mónico $p(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_{n-1}x^{n-1} + x^n \in \mathbb{K}[x]$ no constante dado, es posible asociarle la matriz

$$C(p) = \begin{bmatrix} 0 & 0 & 0 & 0 & \dots & 0 & 0 & -a_0 \\ 1 & 0 & 0 & 0 & \dots & 0 & 0 & -a_1 \\ 0 & 1 & 0 & 0 & \dots & 0 & 0 & -a_2 \\ 0 & 0 & 1 & 0 & \dots & 0 & 0 & -a_3 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 & -a_{n-3} \\ 0 & 0 & 0 & 0 & \dots & 1 & 0 & -a_{n-2} \\ 0 & 0 & 0 & 0 & \dots & 0 & 1 & -a_{n-1} \end{bmatrix},$$

llamada **matriz de compañía** de p. El siguiente resultado proporciona un determinante que será necesario conocer en Álgebra Lineal.

Proposición 7.9. Si $C(p)$ es la matriz de compañía asociada al polinomio (mónico) $p(x) = a_0 + a_1x + a_2x^2 + \cdots + a_{n-1}x^{n-1} + x^n \in \mathbb{K}[x]$ no constante entonces

$$\det(xI_n - C(p)) = \begin{vmatrix} x & 0 & 0 & 0 & \ldots & 0 & 0 & a_0 \\ -1 & x & 0 & 0 & \ldots & 0 & 0 & a_1 \\ 0 & -1 & x & 0 & \ldots & 0 & 0 & a_2 \\ 0 & 0 & -1 & x & \ldots & 0 & 0 & a_3 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & 0 & 0 & a_{n-3} \\ 0 & 0 & 0 & 0 & \ldots & -1 & x & a_{n-2} \\ 0 & 0 & 0 & 0 & \ldots & 0 & -1 & x + a_{n-1} \end{vmatrix} = p(x).$$

Demostración. Se probará por inducción sobre n.

Si $n = 1$, el polinomio mónico de grado 1 es $p(x) = a_0 + x$ y el determinante de la matriz $xI_1 - \begin{bmatrix} -a_0 \end{bmatrix} = \begin{bmatrix} x + a_0 \end{bmatrix}$ es claramente $x + a_0 = p(x)$.

Sea $n > 1$ y supóngase que el resultado es válido para toda matriz del tipo $xI_{n-1} - C(p)$ de tamaño $(n-1) \times (n-1)$. Se debe probar que es válido para una matriz de este tipo de tamaño $n \times n$. En efecto, realizando el desarrollo del determinante por la primera fila se observa que el determinante que resulta de eliminar la primera fila y última columna es el de una matriz triangular superior (y, por tanto, coincide con el producto de los elementos de su diagonal, que es -1 multiplicado por sí mismo $n-1$ veces). Si tras el desarrollo indicado se aplica la hipótesis de inducción entonces se tiene que

$$\det(xI_n - C(p)) = \begin{vmatrix} x & 0 & 0 & 0 & \dots & 0 & 0 & a_0 \\ -1 & x & 0 & 0 & \dots & 0 & 0 & a_1 \\ 0 & -1 & x & 0 & \dots & 0 & 0 & a_2 \\ 0 & 0 & -1 & x & \dots & 0 & 0 & a_3 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 & a_{n-3} \\ 0 & 0 & 0 & 0 & \dots & -1 & x & a_{n-2} \\ 0 & 0 & 0 & 0 & \dots & 0 & -1 & x + a_{n-1} \end{vmatrix}$$

$$\begin{aligned} &= x \det \left(\begin{bmatrix} x & 0 & 0 & \dots & 0 & 0 & a_1 \\ -1 & x & 0 & \dots & 0 & 0 & a_2 \\ 0 & -1 & x & \dots & 0 & 0 & a_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & 0 & a_{n-3} \\ 0 & 0 & 0 & \dots & -1 & x & a_{n-2} \\ 0 & 0 & 0 & \dots & 0 & -1 & x + a_{n-1} \end{bmatrix} \right) + (-1)^{1+n} a_0 (-1)^{n-1} \\ &= x(a_1 + a_2 x + a_3 x^2 + \cdots + a_{n-2} x^{n-3} + a_{n-1} x^{n-2} + x^{n-1}) + a_0 \\ &= a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \cdots + a_{n-2} x^{n-2} + a_{n-1} x^{n-1} + x^n \\ &= p(x). \end{aligned}$$

$\square$

7.5. EJERCICIOS

(1) Calcular el área del paralelogramo de la Figura 7.2 de la página 269 utilizando el área de triángulos y rectángulos.

(2) Sea $A \in \mathbb{K}^{2\times 2}$. Demostrar que el desarrollo de Laplace se puede realizar utilizando cualquier columna o cualquier fila de A.

(3) Sea $A \in \mathbb{K}^{3\times 3}$. Demostrar que la definición de determinante coincide con el desarrollo de Laplace utilizando:

(a) la segunda fila de A.

(b) la tercera columna de A.

(4) Para matrices reales invertibles A y B de tamaño 3×3 y matrices reales C_1 y C_2 de tamaños adecuados, se tiene la ecuación matricial siguiente

$$(3A^{-1}B^t)^{-1} + [\det(4(B^{-1})^t) \cdot X]^t = \begin{bmatrix} C_1 \\ C_2 \end{bmatrix}^t .$$

(a) Despejar la matriz X.

(b) Usando el apartado (a), encontrar una matriz X que satisfaga la ecuación matricial dada, para:

$$A = \begin{bmatrix} 6 & 12 & -3 \\ -24 & 18 & -6 \\ -12 & 30 & 6 \end{bmatrix}, \qquad B = \begin{bmatrix} 1 & -1 & 0 \\ 3 & 3 & 2 \\ 0 & -4 & -1 \end{bmatrix},$$

$$C_1 = \begin{bmatrix} 1 \\ -1 \\ 3 \end{bmatrix}^t \qquad \text{y} \qquad C_2 = \begin{bmatrix} 1 & 1 \\ -2 & 0 \\ 3 & -1 \end{bmatrix}^t .$$

(c) El apartado (b) puede resolverse reemplazando directamente las matrices A, B, C_1 y C_2 en la ecuación matricial y calculando todas las operaciones indicadas. ¿Qué ventajas tiene resolverlo a partir del primer apartado?

(5) Sean α y β dos números reales fijos y sea la matriz

$$A = \begin{bmatrix} x & \alpha & \beta \\ \beta & x & \alpha \\ \alpha & \beta & x \end{bmatrix}.$$

(a) Encontrar todos los valores de $x \in \mathbb{C}$ que anulen el determinante de A. (Ayuda: se obtendrá una raíz real y dos complejas).

(b) Aplicando propiedades de determinantes, comprobar que las raíces obtenidas en el apartado anterior verifican la condición requerida.

(6) Sea la matriz

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

Calcular $|\det(A)|$ y el área del paralelogramo que determinan los dos vectores que forman las columnas de la matriz A. ¿Qué se puede observar? Representar gráficamente. ¿Se puede generalizar a dos vectores cualesquiera de $\mathbb{R}^2$?

(7) Realizar dibujos en $\mathbb{R}^2$ que permitan visualizar todas las interpretaciones geométricas realizadas en las páginas 281, 284, 285 y 289.

(8) Demostrar las tres primeras propiedades de la Proposición 7.4 utilizando la Proposición 5.7 y propiedades de determinantes probadas previamente.

(9) Probar que la fórmula de Cauchy-Binet es válida para un número finito de matrices cuadradas con coeficientes en un cuerpo $\mathbb{K}$, todas del mismo tamaño.

(10) ¿Cuánto vale el determinante de una matriz compleja A de tamaño $n \times n$ que verifica $A^{p+1} = A$, donde $p \in \mathbb{N}$? Justificar la respuesta.

(11) (a) Dados dos puntos distintos (x_1, y_1) y (x_2, y_2) en el plano $\mathbb{R}^2$, demostrar que existe una única recta que pasa por ellos y dicha ecuación

está dada por

$$\begin{vmatrix} x & y & 1 \\ x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \end{vmatrix} = 0.$$

(b) Encontrar la ecuación de la recta que pasa por los puntos $(2,3)$ y $(4,1)$.

(12) Sean $a, b, c \in \mathbb{R}$. Demostrar que las raíces de la ecuación

$$\det \begin{bmatrix} a-x & b \\ b & c-x \end{bmatrix} = 0$$

son números reales. Comparar la dificultad en su resolución con el mismo ejercicio planteado en la página 171.

(13) Demostrar que si la suma de cada par de elementos correspondientes de dos filas del determinante de una matriz $A \in \mathbb{K}^{3\times 3}$ es un múltiplo (constante) de los correspondientes elementos de otra fila entonces el determinante de A es cero.

(14) Indicar si cada una de las siguientes afirmaciones es verdadera o falsa. Si es verdadera, demostrarla; si es falsa, dar un contraejemplo.

- $\det(A+B) = \det(A) + \det(B)$, $\forall A, B \in \mathbb{K}^{n\times n}$.
- $\det(\lambda A) = \lambda \det(A)$, $\forall A \in \mathbb{K}^{n\times n}$, $\forall \lambda \in \mathbb{K}$.
- $\det(\lambda A) = \lambda^n \det(A)$, $\forall A \in \mathbb{K}^{n\times n}$, $\forall \lambda \in \mathbb{K}$.
- $\det(A^m) = [\det(A)]^m$, $\forall A \in \mathbb{K}^{n\times n}$, $\forall m \in \mathbb{N}$.
- $\mathrm{Adj}(\mathrm{A}^{\mathrm{t}}) = [\mathrm{Adj}(A)]^t$, $\forall A \in \mathbb{K}^{n\times n}$.

(15) Sea $n \in \mathbb{N}$. Utilizar únicamente la definición para demostrar, por el método de inducción, que el determinante de una matriz $A \in \mathbb{K}^{n\times n}$ que tiene una fila completa de ceros es nulo.

(16) La propiedad del Corolario 7.2 requiere que $\operatorname{car}(\mathbb{K}) \neq 2$ pues se demuestra como consecuencia de la Proposición 7.3. Demostrar, por inducción, que la misma propiedad es válida sin suponer dicha restricción. (Ayuda: distinguir los casos de filas iguales consecutivas y no consecutivas).

(17) Sea $n \in \mathbb{N}$. Demostrar que el determinante de una matriz triangular inferior $A \in \mathbb{K}^{n\times n}$ es el producto de los elementos de su diagonal. Hacerlo de dos formas: la primera utilizando como definición de determinante el desarrollo de Laplace por la primera columna y el método de inducción (se puede utilizar la propiedad del determinante de una matriz con una fila nula) y la segunda utilizando la relación entre el determinante de una matriz y el de su traspuesta.

(18) Sea $A \in \mathbb{K}^{n\times n}$. Demostrar que $A[\operatorname{Adj}(A)]^t = \det(A)I_n$. Primero calcular el producto $P = [p_{ij}] := A[\operatorname{Adj}(A)]^t$. Luego, analizar los elementos p_{ij} separando los casos $i = j$ e $i \neq j$.

(19) Sea $A = \begin{bmatrix} a_{ij} \end{bmatrix} = \begin{bmatrix} a_1 & \dots & a_i & \dots & a_n \end{bmatrix} \in \mathbb{K}^{n\times n}$ una matriz particionada según sus columnas a_j $(j = 1, \dots, n)$ y sean, para cada $i = 1, \dots, n$

$$A_i = \begin{bmatrix} a_1 & \dots & b & \dots & a_n \end{bmatrix} \in \mathbb{K}^{n\times n} \qquad \text{y} \qquad b = \begin{bmatrix} b_1 \\ \vdots \\ b_i \\ \vdots \\ b_n \end{bmatrix},$$

donde la matriz A_i se obtiene cambiando la columna i-ésima de A por la columna b. Probar que, para cada $i = 1, \dots, n$, se cumple que:

(a) $\det(A_i) = b_1 A_{1i} + \cdots + b_i A_{ii} + \cdots + b_n A_{ni}$, donde A_{ij} es el complemento algebraico correspondiente a la posición (i, j) de A.

(b) si $x = \begin{bmatrix} x_1 \\ \vdots \\ x_i \\ \vdots \\ x_n \end{bmatrix}$ satisface $Ax = b$ entonces $\det(A_i) = x_i \det(A)$.

(Ayuda: Para el segundo apartado, sustituir cada b_i en $\det(A_i)$ por la relación que ofrece el sistema lineal. Luego, aplicar propiedades de determinantes).

(20) Demostrar la regla de Cramer utilizando la fórmula de la inversa de una matriz y el ejercicio anterior.

(21) Hallar todos los valores de $\alpha \in \mathbb{C}$ tales que

$$\det\left(A \begin{bmatrix} 4+4i & \alpha^2 \\ -\alpha & 1-i \end{bmatrix} A^{-1} \right) = 0,$$

para toda matriz invertible $A \in \mathbb{C}^{2\times 2}$. Comprobar que los valores obtenidos son soluciones de la ecuación indicada.

(22) Hallar condiciones necesarias y suficientes para que una matriz de Vandermonde sea invertible.

(23) Analizar la validez del siguiente razonamiento:

"Sean $A, B \in \mathbb{K}^{n\times n}$. Si existe $(I_n - AB)^{-1}$ entonces existe $(I_n - BA)^{-1}$."

En efecto, aplicando la propiedad distributiva se tiene que $B(I_n - AB) = B - BAB = (I_n - BA)B$. Puesto que $I_n - AB$ es invertible, multiplicando a derecha ambos miembros de $B(I_n - AB) = (I_n - BA)B$ por $(I_n - AB)^{-1}$ se llega a $B = (I_n - BA)B(I_n - AB)^{-1}$. Usando la fórmula de Cauchy-Binet queda $\det(B) = \det(I_n - BA)\det(B)\det((I_n - AB)^{-1})$. Luego, $\det(I_n - BA)\det((I_n - AB)^{-1}) = 1$. Como $\mathbb{K}$ es un cuerpo, no tiene dividores de cero y, además, $\det((I_n - AB)^{-1}) \neq 0$ por ser $(I_n - AB)^{-1}$ invertible, entonces se obtiene que $\det(I_n - BA) \neq 0$, con lo que $I_n - BA$ es invertible.

(24) Sean $A, B \in \mathbb{K}^{n\times n}$. Si $I_n - AB$ es invertible entonces probar que $I_n - BA$ es invertible y además

$$(I_n - BA)^{-1} = I_n + B(I_n - AB)^{-1}A.$$

¿Es válido el resultado si $A \in \mathbb{K}^{m\times n}$ y $B \in \mathbb{K}^{n\times m}$? Justificar.

(25) Demostrar que si $A, B \in \mathbb{K}^{n\times n}$ son tales que $I_n + AB$ es invertible entonces

$$\begin{bmatrix} I_n & B \\ -A & I_n \end{bmatrix}$$

es invertible.

Parte II

Álgebra Lineal

Capítulo 8

Espacios vectoriales

Índice

8.1. Introducción

El lingüista y matemático[1] polaco Hermann Günther Grassmann (1808-1887) (véase la Figura 8.1), quien como matemático fue autodidacta, fue el

Figura 8.1: Hermann G. Grassmann.

primero en desarrollar (en 1844) las ideas de espacios vectoriales, subespacios, bases, dimensión, etc.

Debido a la dificultad y poca claridad en la nomenclatura, sus trabajos, que no estaban escritos en el lenguaje actual, no tuvieron la acogida esperada. Desafortunadamente, pasaron desapercibidos hasta que en 1888 el matemático italiano Giuseppe Peano (1858-1932), tras investigar a partir de los trabajos de Grassmann, introdujo el concepto de espacio vectorial de forma axiomática.

La comunidad científica comenzó a trabajar en ellos a partir de estudios del matemático alemán Hermann Weyl (1885-1955) (véase la Figura 8.2) sobre 1920.

Por otra parte, William Rowan Hamilton (1805-1865) (véase la Figura 8.3), dedicó buena parte de su vida al desarrollo de la teoría de los números cuaterniones, lo que ofreció una base sólida al concepto de vector.

[1] En realidad fue un polímata pues además fue filósofo, físico y teólogo.

Figura 8.2: Hermann Weyl.

En su libro *History of Analitic Geometry*, Carl B. Boyer señala que Hamilton pretendía construir, sin utilizar coordenadas cartesianas, un cálculo de vectores en el espacio ordinario y para ello fijó la atención en operaciones con cuatro parámetros que transforma un vector en otro. De hecho la notación de los vectores $\mathbf{i}, \mathbf{j}$ y $\mathbf{k}$ que introdujo aún se utiliza hoy en día. Sin embargo, Grassmann fue menos restrictivo en su concepción e involucró un número infinito de dimensiones.

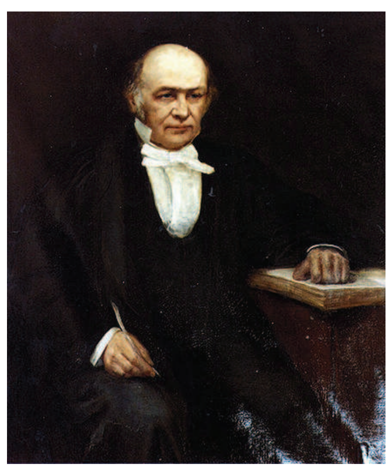


Figura 8.3: William Rowan Hamilton.

8.2. Modelos geométricos y algebraicos

Se comienza con modelos elementales que serán utilizados como instrumentos intuitivos para establecer la teoría general de espacios vectoriales abstractos.

En el plano de la Geometría Elemental se consideran los vectores representados como segmentos dirigidos u orientados, suponiendo como vectores iguales aquellos que se obtienen tras realizar desplazamientos paralelos de uno fijado. Se definen la suma de vectores (mediante la regla del paralelogramo) y el producto de un escalar por un vector (alargando o acortando su longitud y considerando el sentido adecuado) y se llama $\mathbb{VL}^2$ al conjunto de tales vectores (libres) del plano. Los vectores de este conjunto satisfacen ciertas propiedades en relación a las operaciones.

Es importante observar que $\mathbb{VL}^2$ no es lo mismo que el plano cartesiano

$$\mathbb{R}^2 = \mathbb{R} \times \mathbb{R} = \{(x, y) : x, y \in \mathbb{R}\}$$

puesto que $\mathbb{VL}^2$ está formado por vectores libres y $\mathbb{R}^2$ por pares ordenados de elementos de $\mathbb{R}$. Sin embargo, $\mathbb{R}^2$ puede utilizarse como un modelo abstracto (algebraico) del modelo geométrico $\mathbb{VL}^2$ identificando cada vector libre con sus coordenadas. Gracias a esta identificación, muchos resultados del Álgebra Lineal se podrán interpretar geométricamente en el plano y podrán ser utilizados en el desarrollo de la Geometría Cartesiana a través de métodos de la Geometría Analítica.

A continuación se detallan estas ideas.

Los modelos geométricos $\mathbb{VL}^2$ y $\mathbb{VL}^3$

Se recuerda que[2] en un conjunto $\mathcal{E}_2$ llamado **plano**, cuyos elementos se llaman **puntos**, se han considerado segmentos orientados.

El segmento orientado con **origen** en el punto A y **extremo** en el punto B, denotado por $\overrightarrow{AB}$, se llama **vector fijo** y se representa gráficamente en

[2]Véase el ejemplo de la página 54 donde $\mathcal{E}_2$ se denotó por π.

$\mathcal{E}_2$ por una flecha señalando cuáles son los puntos A y B. Tres elementos que posee un vector fijo $\overrightarrow{AB}$ son: su **módulo** $\|\overrightarrow{AB}\|$, su **dirección** y su **sentido**. El conjunto de vectores fijos del plano $\mathcal{E}_2$ se denota $\mathbb{VF}^2$.

Suma de vectores fijos: Si dos vectores fijos (no nulos) de $\mathcal{E}_2$ tienen el mismo origen A, como $\overrightarrow{AB}$ y $\overrightarrow{AC}$, y distinta dirección, se llama *suma* de estos vectores al vector fijo $\overrightarrow{AD}$ donde D es el punto de $\mathcal{E}_2$ tal que $ABDC$ es un paralelogramo, y se denota

$$\overrightarrow{AD} = \overrightarrow{AB} + \overrightarrow{AC}.$$

Cuando los vectores $\overrightarrow{AB}$ y $\overrightarrow{AC}$ tienen la misma dirección y mismo sentido, la *suma* se define como el vector $\overrightarrow{AD}$ cuyo módulo es $\|\overrightarrow{AD}\| = \|\overrightarrow{AB}\| + \|\overrightarrow{AC}\|$ y con la misma dirección y sentido que $\overrightarrow{AB}$ (ó $\overrightarrow{AC}$). Cuando los vectores $\overrightarrow{AB}$ y $\overrightarrow{AC}$ tienen la misma dirección y sentido contrario, la *suma* se define como el vector $\overrightarrow{AD}$ cuyo módulo $\|\overrightarrow{AD}\|$ es la resta de los módulos $\|\overrightarrow{AB}\|$ y $\|\overrightarrow{AC}\|$ (el de mayor menos el de menor módulo) y con la misma dirección que ambos y el mismo sentido que el de mayor módulo; y se define como el vector nulo si ambos tienen el mismo módulo. La *suma* del vector fijo $\overrightarrow{AB}$ con el vector nulo $\overrightarrow{AA}$ se define como $\overrightarrow{AB}$.

Multiplicación de un número real por un vector fijo: Dado un vector fijo $\overrightarrow{AB}$ (no nulo) en $\mathcal{E}_2$ y un número real λ, se llama *multiplicación* del número real por el vector fijo al vector fijo $\overrightarrow{AC}$ tal que los puntos A, B y C están en la misma recta; B se encuentra en el segmento $\overline{AC}$ cuando $\lambda \geq 1$, C se encuentra en el segmento $\overline{AB}$ cuando $0 < \lambda < 1$ y A en el segmento $\overline{CB}$ cuando $\lambda < 0$; y el módulo del vector $\overrightarrow{AC}$ es el producto del módulo del vector $\overrightarrow{AB}$ por el valor absoluto de λ (es decir, $\|\overrightarrow{AC}\| = |\lambda|\|\overrightarrow{AB}\|$). Se denota

$$\overrightarrow{AC} = \lambda\overrightarrow{AB}.$$

Si el vector fijo es $\overrightarrow{AA}$, es decir, el nulo, o bien si $\lambda = 0$ entonces la multiplicación se define como el vector fijo nulo $\overrightarrow{AA}$.

Nota: Debido a que, para que se pueda realizar la suma de vectores fijos, se requiere la condición de que ambos posean *el mismo origen*, no todo par de vectores fijos se podrá sumar. Será necesario pues, identificar todos los vectores fijos *con distinto origen* que tengan el mismo módulo, dirección y sentido, para que dicha operación se pueda realizar de manera efectiva. La técnica algebraica que permite realizar esta identificación es la consideración de una relación de equivalencia y la partición que produce en el conjunto junto al conjunto cociente que queda determinado.

Relación de equipolencia: Sobre el conjunto $\mathbb{VF}^2$ de los **vectores fijos** del plano se define la siguiente relación binaria: dos vectores fijos no nulos $\overrightarrow{AB}$ y $\overrightarrow{CD}$ se llaman **equipolentes**, y se denota $\overrightarrow{AB}\ \mathcal{R}\ \overrightarrow{CD}$, si tienen el mismo módulo, la misma dirección y el mismo sentido. Por definición, se considera que todos los vectores fijos nulos son **equipolentes** entre sí.

Esta relación binaria es una relación de equivalencia sobre el conjunto $\mathbb{VF}^2$ de los vectores fijos del plano $\mathcal{E}_2$.

La clase de equivalencia de un vector fijo (no nulo) $\overrightarrow{AB}$ es

$$C_{\overrightarrow{AB}} = \{\overrightarrow{CD} : \overrightarrow{CD} \text{ tiene mismo módulo, dirección y sentido que } \overrightarrow{AB}\}$$

y se llama **vector libre**, **vector geométrico** o simplemente **vector**. El **conjunto cociente** del conjunto de los **vectores fijos** $\mathbb{VF}^2$ del plano $\mathcal{E}_2$ por la relación $\mathcal{R}$ es

$$\mathbb{VL}^2 := \mathbb{VF}^2/\mathcal{R} = \{C_{\overrightarrow{AB}} :\ \overrightarrow{AB} \text{ es un vector fijo de } \mathcal{E}_2\}$$

y se denomina **conjunto de vectores libres** del plano $\mathcal{E}_2$. Por tanto, los vectores libres son los elementos de $\mathbb{VL}^2$ y cada uno es un conjunto de vectores fijos de $\mathcal{E}_2$.

Cada vector fijo $\overrightarrow{CD}$ perteneciente a la clase de equivalencia $C_{\overrightarrow{AB}}$ es un **representante** de dicha clase. Se llama módulo, dirección y sentido de un vector libre no nulo al de uno cualquiera de sus representantes. El vector libre formado por todos los vectores fijos nulos se llama vector libre nulo, se designa por $C_{\overrightarrow{AA}}$ (si A pertenece a $\mathcal{E}_2$), y carece de dirección y sentido.

Si se fija un punto O en el plano $\mathcal{E}_2$, para cada vector fijo $\overrightarrow{AB}$, existe un único punto P tal que $\overrightarrow{AB}$ es equipolente a $\overrightarrow{OP}$. Los vectores fijos que comienzan en O se suelen llamar **vectores de posición**[3]. En este caso, es el vector de posición de P con respecto a O.

Por lo tanto, para un vector fijo determinado, el vector libre que lo contiene, contiene también un único representante con origen en O. El conjunto $\mathbb{VL}^2_O$ formado por todos los vectores de posición asociados a cada vector fijo (es decir, por todos estos representantes con origen en O) ofrecen un **conjunto completo de representantes** para los vectores libres que determinan la partición del conjunto de los vectores fijos del plano $\mathcal{E}_2$ en sus clases de equivalencia por $\mathcal{R}$.

Puesto que el vector $\overrightarrow{OP}$, con origen en el punto O y extremo en el punto P, queda completamente determinado por el punto P (pues O está fijado), es posible identificar dicho vector $\overrightarrow{OP}$ con el propio punto P. Este hecho establece una biyección entre el conjunto $\mathbb{VL}^2_O$ (formado por los representantes de los vectores libres de $\mathcal{E}_2$ que comienzan en O) y $\mathcal{E}_2$.

Con vectores libres sí es posible operar. Para ello, se define la adición y la multiplicación de un número real por un vector libre. La razón es que dados dos vectores libres, es posible encontrar un representante de cada una de las clases de equivalencia que los definen, de modo que ambos representantes tengan el mismo origen O.

Fijado un punto O, se consideran las siguientes operaciones:

Adición de vectores libres: La operación

$$\begin{array}{rccc} + : & \mathbb{VL}^2 \times \mathbb{VL}^2 & \rightarrow & \mathbb{VL}^2 \\ & (\vec{u}, \vec{v}) & \longmapsto & \vec{u} + \vec{v} \end{array}$$

se define de la siguiente forma:

- se eligen dos puntos U y V en $\mathcal{E}_2$ tales que $\vec{u} = C_{\overrightarrow{OU}}$ y $\vec{v} = C_{\overrightarrow{OV}}$,
- y ahora se suman $\overrightarrow{OU}$ y $\overrightarrow{OV}$ como vectores fijos (mediante la regla del paralelogramo)

[3]El hecho de fijar un punto O es determinante para la estructura algebraica de Espacio Vectorial, y será lo que la distinga de la de Espacio Afín.

obteniendo un representante $\overrightarrow{OS} = \overrightarrow{OU} + \overrightarrow{OV}$ de la clase $\vec{u} + \vec{v} = C_{\overrightarrow{OS}}$.

Para que esté bien definida, uno de los hechos que debe cumplir la adición de dos vectores libres, es que su resultado sea otro vector libre (es decir, que esté en $\mathbb{VL}^2$), lo cual se cumple atendiendo a la definición previa (realizada mediante la regla del paralelogramo, y considerando la clase del vector obtenido como suma de vectores fijos con igual origen). Además, se puede ver que si se cambian los representantes para realizar la suma, se obtiene la misma clase resultado, este hecho indica que la operación es independiente de los representantes elegidos para realizar la operación. Con esto, es posible afirmar que la operación adición está **bien definida** y su resultado es la **suma** de ambos vectores libres.

Multiplicación de un número real por un vector libre: La operación

$$\begin{array}{rccc} \cdot : & \mathbb{R} \times \mathbb{VL}^2 & \to & \mathbb{VL}^2 \\ & (\lambda, \vec{u}) & \longmapsto & \lambda \cdot \vec{u} \end{array}$$

se define de la siguiente forma:

- se elige un punto U en $\mathcal{E}_2$ tal que $\vec{u} = C_{\overrightarrow{OU}}$,
- y ahora se multiplica λ por $\overrightarrow{OU}$ como vectores fijos

obteniendo un representante $\overrightarrow{OM} = \lambda\overrightarrow{OU}$ de la clase $\lambda \cdot \vec{u} = C_{\overrightarrow{OM}}$.

El **producto** $\lambda \cdot \vec{u}$ de un número real por un vector libre también está **bien definido**.

Si $\lambda = -1$, $\lambda \cdot \vec{u} = (-1)\vec{u}$ y se denota $-\vec{u}$ (y cumple que $\vec{u} + (-\vec{u}) = \vec{0} = C_{\overrightarrow{OO}}$).

En el **espacio ordinario** $\mathcal{E}_3$ es posible realizar un procedimiento completamente análogo al realizado en un plano $\mathcal{E}_2$ (o **plano ordinario**). Se obtendrá el conjunto cociente $\mathbb{VL}^3$ por una relación de equivalencia sobre el conjunto de vectores fijos $\mathbb{VF}^3$ de $\mathcal{E}_3$, similar a la establecida en $\mathbb{VF}^2$.

Así, denotando $\mathbb{VL}$ tanto a $\mathbb{VL}^2$ como a $\mathbb{VL}^3$, es posible disponer de dos ejemplos importantes de la Geometría Elemental que poseen las siguientes propiedades:

Propiedad asociativa: $\forall \vec{u}, \vec{v}, \vec{w} \in \mathbb{VL} : (\vec{u} + \vec{v}) + \vec{w} = \vec{u} + (\vec{v} + \vec{w})$.

Elemento neutro: $\exists \vec{0} \in \mathbb{VL}$, $\forall \vec{u} \in \mathbb{VL}$: $\vec{u} + \vec{0} = \vec{0} + \vec{u} = \vec{u}$.

Elemento opuesto: $\forall \vec{u} \in \mathbb{VL}, \exists -\vec{u} \in \mathbb{VL} : \vec{u} + (-\vec{u}) = (-\vec{u}) + \vec{u} = \vec{0}$.

Propiedad conmutativa: $\forall \vec{u}, \vec{v} \in \mathbb{VL} : \vec{u} + \vec{v} = \vec{v} + \vec{u}$.

Es decir, el conjunto de los vectores libres $\mathbb{VL}$ tiene estructura algebraica de grupo abeliano con respecto a la adición $+$: $(\mathbb{VL}, +)$.

Con respecto a la multiplicación de un número real por un vector libre se cumplen las siguientes propiedades:

Distributiva respecto a la suma de vectores: $\forall \lambda \in \mathbb{R}, \forall \vec{u}, \vec{v} \in \mathbb{VL} : \lambda \cdot (\vec{u} + \vec{v}) = \lambda \cdot \vec{u} + \lambda \cdot \vec{v}$.

Distributiva respecto a la suma de números reales: $\forall \lambda, \mu \in \mathbb{R}, \forall \vec{u} \in \mathbb{VL} : (\lambda + \mu) \cdot \vec{u} = \lambda \cdot \vec{u} + \mu \cdot \vec{u}$.

Pseudoasociativa[4]: $\forall \lambda, \mu \in \mathbb{R}, \forall \vec{u} \in \mathbb{VL} : (\lambda\mu) \cdot \vec{u} = \lambda \cdot (\mu \cdot \vec{u})$.

Modular[5]: $\forall \vec{u} \in \mathbb{VL} : 1 \cdot \vec{u} = \vec{u}$.

Estos dos ejemplos proporcionan **modelos geométricos** de situaciones más generales. Las estructuras algebraicas que se forman a partir de un conjunto no vacío, cuyos elementos por extensión se llamarán **vectores**, una adición y una multiplicación por escalares que satisfagan las propiedades mencionadas se llamará **espacio vectorial**.

Los modelos algebraicos $\mathbb{R}^2$ y $\mathbb{R}^3$

Ahora, a partir de estos modelos geométricos, se pretende encontrar **modelos algebraicos** y así poder intercambiar información entre uno y otro.

Para ello, se consideran dos vectores libres $\vec{e}_1$ y $\vec{e}_2$ de $\mathbb{VL}^2$, no nulos y con distinta dirección. Para cada vector $\vec{u} \in \mathbb{VL}^2$, es posible encontrar dos

[4] A veces se llama *asociatividad mixta*.

[5] A veces se llama *unimodular*.

números reales x, y, únicos, que cumplen $\vec{u} = x\vec{e}_1 + y\vec{e}_2$. Dicho de otro modo, y teniendo en cuenta que $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$, existe una biyección[6] entre $\mathbb{R}^2$ y $\mathbb{VL}^2$:

$$\begin{array}{rcccl} \mathcal{D} & : & \mathbb{R}^2 & \to & \mathbb{VL}^2 \\ & & (x, y) & \mapsto & \mathcal{D}(x, y) = x \cdot \vec{e}_1 + y \cdot \vec{e}_2 \end{array}$$

Dados $(x_1, y_1), (x_2, y_2) \in \mathbb{R}^2$, se tiene que

$$\mathcal{D}(x_1, y_1) = x_1 \cdot \vec{e}_1 + y_1 \cdot \vec{e}_2 \in \mathbb{VL}^2 \quad \text{y} \quad \mathcal{D}(x_2, y_2) = x_2 \cdot \vec{e}_1 + y_2 \cdot \vec{e}_2 \in \mathbb{VL}^2,$$

y en $\mathbb{VL}^2$ se pueden sumar estos vectores libres obteniendo el vector libre

$$\begin{aligned} \mathcal{D}(x_1, y_1) + \mathcal{D}(x_2, y_2) &= [x_1 \cdot \vec{e}_1 + y_1 \cdot \vec{e}_2] + [x_2 \cdot \vec{e}_1 + y_2 \cdot \vec{e}_2] \\ &= (x_1 + x_2) \cdot \vec{e}_1 + (y_1 + y_2) \cdot \vec{e}_2 \end{aligned}$$

(conmutando y sumando los vectores libres en las mismas direcciones). Al ser $\mathcal{D}$ una biyección, se calcula

$$\mathcal{D}^{-1}(\mathcal{D}(x_1, y_1) + \mathcal{D}(x_2, y_2))) = (x_1 + x_2, y_1 + y_2)$$

y así se puede definir

$$(x_1, y_1) + (x_2, y_2) := (x_1 + x_2, y_1 + y_2). \tag{8.1}$$

Del mismo modo, si $\lambda \in \mathbb{R}$, al calcular

$$\begin{aligned} \mathcal{D}^{-1}(\lambda \cdot \mathcal{D}((x_1, y_1))) &= \mathcal{D}^{-1}(\lambda \cdot (x_1 \cdot \vec{e}_1 + y_1 \cdot \vec{e}_2)) \\ &= \mathcal{D}^{-1}((\lambda x_1) \cdot \vec{e}_1 + (\lambda y_1) \cdot \vec{e}_2) \\ &= (\lambda x_1, \lambda y_1), \end{aligned}$$

se puede definir

$$\lambda \cdot (x_1, y_1) := (\lambda x_1, \lambda y_1). \tag{8.2}$$

Ahora, la suma de vectores libres de $\mathbb{VL}^2$ tiene su correspondencia en la suma de pares ordenados de $\mathbb{R}^2$ y el producto de un número real por un vector libre

[6]Más adelante, en el próximo capítulo, se realizará esta correspondencia con detalle y en general, por ahora es importante acercarse a la idea general al menos en $\mathbb{R}^2$.

de $\mathbb{VL}^2$ tiene su correspondencia en el producto de un número real por un par ordenado de $\mathbb{R}^2$.

Se ha encontrado un modelo algebraico formado por pares ordenados de $\mathbb{R}^2$ con la suma definida por (8.1) y la multiplicación por escalares por (8.2). Es importante observar que la suma y la multiplicación definidas en $\mathbb{R}^2$ no dependen de la elección realizada de los vectores no nulos $\vec{e}_1$ y $\vec{e}_2$ con diferentes direcciones. Se dice que el conjunto $\mathcal{B} = \{\vec{e}_1, \vec{e}_2\}$ es una **base** del plano.

Del mismo modo que se ha construido el modelo algebraico $\mathbb{R}^2$ a partir del modelo geométrico $\mathbb{VL}^2$, es posible construir un modelo algebraico $\mathbb{R}^3$ a partir del modelo geométrico $\mathbb{VL}^3$. Ambos tienen en común que se definen a partir de un conjunto no vacío $\mathbb{VL}$ y mediante dos operaciones, una interna $+$ y una externa $\cdot$ tal que $(V, +)$ es un grupo abeliano y $(V, +, \cdot)$ cumple las mismas cuatro propiedades que las indicadas en los modelos geométricos anteriores.

Estos hechos permitirán hacer un estudio abstracto y puramente algebraico, prescindiendo de la geometría subyacente, aunque volviendo a ella cuando se desee recurrir a la intuición.

Correspondencia entre bases y sistemas de coordenadas cartesianos

Al ser los elementos de $\mathbb{VL}^2$ de naturaleza geométrica, la idea ha sido encontrar, a partir del modelo algebraico $\mathbb{R}^2$, una expresión numérica para los vectores libres, de modo que sea posible operar con ellos de manera sencilla.

Para establecer una representación geométrica de la situación, al fijar un origen O en el plano, $\mathbb{VL}^2$ se convierte en espacio vectorial (pudiendo sumar vectores y multiplicar vectores por escalares).

Un **sistema coordenado cartesiano en el plano**, denotado por el par (O, XY), consiste de un par de ejes (distintos) OX y OY que se cortan en el punto O junto a otros dos puntos U_1 y U_2 que determinan las distancias unitarias (unidades de medida) desde el origen a lo largo de los ejes OX y

OY, respectivamente. Se observa que U_1 y U_2 imponen una orientación sobre cada eje.

Dado un sistema coordenado cartesiano en el plano, las coordenadas a y b de un punto P se obtienen geométricamente como sigue.

Se traza una recta por el punto P paralela a OY que intersecte a OX en el punto A y se define $a = OA$. De forma semejante, una recta por el punto P paralela a OX que intersecte a OY en el punto B determina $b = OB$. Como los puntos O, U_1 y U_2 no son colineales, los vectores $\vec{u}_1$ y $\vec{u}_2$ que representan los respectivos puntos U_1 y U_2, determinan una base de $\mathbb{R}^2$. El vector que representa a A es $a\vec{u}_1$ pues $OA/OU_1 = a$, y el que representa a B es $b\vec{u}_2$ pues $OB/OU_2 = b$. Como $PBOA$ es un paralelogramo, el vector $\vec{p}$ que representa a P es $a\vec{u}_1 + b\vec{u}_2$. Esta construcción se aprecia en la Figura 8.4.

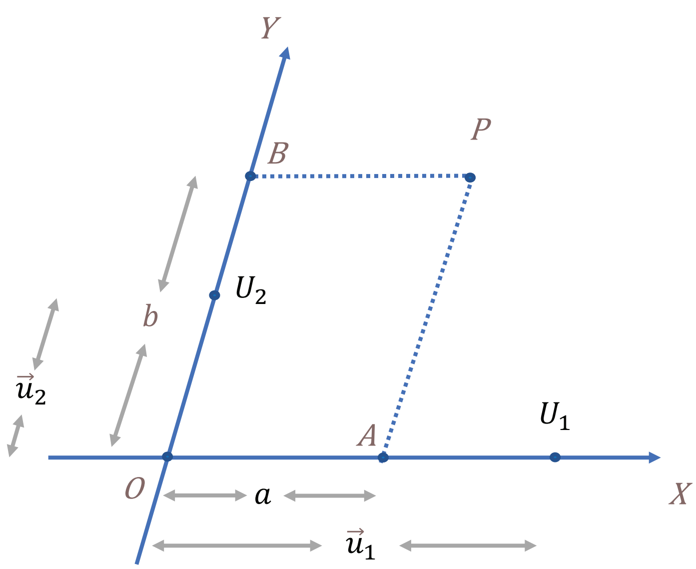


Figura 8.4: Sistema coordenado cartesiano y base en el plano.

Recíprocamente, dada una base $\{\vec{u}_1, \vec{u}_2\}$ del plano (donde el origen ya se ha fijado), se puede construir el sistema coordenado (O, U_1U_2) con U_1 correspondiente a $\vec{u}_1$ y U_2 correspondiente a $\vec{u}_2$. Como el vector del plano correspondiente a un punto P (cualquiera) se puede expresar de forma única como $a\vec{u}_1 + b\vec{u}_2$, en este sistema coordenado a y b son la primera y la segunda coordenadas del punto P.

En resumen, a un sistema coordenado cartesiano se le puede asociar una base y viceversa.

Mediante estas ideas, en el siglo XVII, Descartes y Fermat vincularon las formas y los números mediante la geometría de coordenadas, transformando la Geometría en Álgebra.

De este modo, mediante técnicas de la Geometría Analítica, se pueden abordar las dos cuestiones fundamentales siguientes:

- Dado un lugar geométrico (es decir, un conjunto formado por todos los puntos que cumplen una determinada condición) en un sistema de coordenadas, se pretende obtener su ecuación.

- Dada una ecuación en un sistema de coordenadas, se desea determinar la gráfica o lugar geométrico de todos los puntos que satisfacen dicha ecuación.

En el próximo capítulo se volverá sobre estas cuestiones.

8.3. Definición

Como hasta ahora, $\mathbb{K}$ representará un cuerpo arbitrario. Se trabajará básicamente con $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ o $\mathbb{Z}_p$ (p número primo positivo).

La siguiente definición axiomática de espacio vectorial se debe a G. Peano y es de finales del siglo XIX.

Definición 8.1. Sea $(\mathbb{K}, +, \cdot)$ un cuerpo. Un **espacio vectorial**[a] sobre $\mathbb{K}$ es una terna $(V, \oplus, \odot)$ formada por un conjunto V y dos operaciones definidas en V:

Operación interna: $\begin{array}{cccc} \oplus & : & V \times V & \to & V \\ & & (\vec{u}, \vec{v}) & \mapsto & \vec{u} \oplus \vec{v} \end{array}$ sujeta a los axiomas:

(V_1) $(\vec{u} \oplus \vec{v}) \oplus \vec{w} = \vec{u} \oplus (\vec{v} \oplus \vec{w}), \qquad \forall \vec{u}, \vec{v}, \vec{w} \in V.$

(V_2) $\exists \vec{0} \in V : \vec{u} \oplus \vec{0} = \vec{0} \oplus \vec{u} = \vec{u}, \qquad \forall \vec{u} \in V.$

(V_3) $\forall \vec{u} \in V : \exists \vec{u}' \in V : \vec{u} \oplus \vec{u}' = \vec{u}' \oplus \vec{u} = \vec{0}.$

(V_4) $\vec{u} \oplus \vec{v} = \vec{v} \oplus \vec{u}, \qquad \forall \vec{u}, \vec{v} \in V.$

Operación externa: $\begin{array}{cccc} \odot & : & \mathbb{K} \times V & \to & V \\ & & (\lambda, \vec{u}) & \mapsto & \lambda \odot \vec{u} \end{array}$ sujeta a los axiomas:

(V_5) $\lambda \odot (\vec{u} \oplus \vec{v}) = \lambda \odot \vec{u} \oplus \lambda \odot \vec{v}, \qquad \forall \lambda \in \mathbb{K}, \forall \vec{u}, \vec{v} \in V.$

(V_6) $(\lambda + \mu) \odot \vec{u} = \lambda \odot \vec{u} \oplus \mu \odot \vec{u}, \qquad \forall \lambda, \mu \in \mathbb{K}, \forall \vec{u} \in V.$

(V_7) $(\lambda \cdot \mu) \odot \vec{u} = \lambda \odot (\mu \odot \vec{u}), \qquad \forall \lambda, \mu \in \mathbb{K}, \forall \vec{u} \in V.$

(V_8) $1_{\mathbb{K}} \odot \vec{u} = \vec{u}, \qquad \forall \vec{u} \in V.$

[a] A veces se llama **espacio lineal**.

Al espacio vectorial V sobre $\mathbb{K}$ también se lo suele llamar **$\mathbb{K}$-espacio vectorial**.

A los elementos de V se los llama **vectores**, a los elementos de $\mathbb{K}$ se los denomina **escalares**, a la operación interna se la denomina **adición** y a la operación externa se la llama **multiplicación por escalares**. Al[7] elemento neutro de la adición se lo llama **vector nulo**.

[7] Se probará que es único.

Observación 8.1. Es importante remarcar que en los axiomas de la multiplicación por escalares de la Definición 8.1 aparecen las siguientes operaciones:

(a) $\oplus$: adición de vectores,

(b) $\odot$: multiplicación de un escalar por un vector,

(c) $+$: adición de escalares de $\mathbb{K}$,

(d) $\cdot$: multiplicación de escalares de $\mathbb{K}$.

Observación 8.2. La existencia del elemento neutro de la adición garantiza que $\{\vec{0}\} \subseteq V$, con lo cual todo espacio vectorial V satisface $V \neq \emptyset$.

De los requerimientos (axiomas) exigidos a la adición, queda claro que $(V, \oplus)$ es un grupo abeliano.

Cuando se utilice $\mathbb{K} = \mathbb{R}$, se hablará de **espacio vectorial real**; si $\mathbb{K} = \mathbb{C}$, de **espacio vectorial complejo**.

Observación 8.3. Los axiomas $(V_1) - (V_7)$ pedidos en la definición de espacio vectorial se aceptan como requerimientos naturales puesto que afirmaciones similares aparecen en las definiciones clásicas del concepto de cuerpo. Sin embargo, el axioma (V_8) es singular y, claramente, menos natural. Más adelante se probará que dicho axioma es equivalente a la propiedad "$\lambda\vec{u} = \vec{0}$ si y sólo si $\lambda = 0$ ó $\vec{u} = \vec{0}$", que es más natural puesto que recuerda a la no existencia de divisores de cero en un cuerpo.

Para simplificar la notación, cuando no lleve a confusión, se hablará del espacio vectorial $(V, +, \cdot)$. Más aún, cuando las operaciones utilizadas resulten claras del contexto, el espacio vectorial $(V, +, \cdot)$ se representará simplemente por su conjunto soporte V.

8.3.1. Ejemplos de espacios vectoriales

Ejemplo 8.1 (El espacio vectorial trivial)**.** *El conjunto $V = \{\vec{0}\}$ formado por un único elemento es un $\mathbb{K}$-espacio vectorial, llamado* **espacio vectorial trivial** *o* **espacio vectorial nulo**, *con las operaciones definidas mediante*

$$\begin{aligned} \vec{0} + \vec{0} &= \vec{0}, \\ \lambda \cdot \vec{0} &= \vec{0}, \qquad \forall \lambda \in \mathbb{K}. \end{aligned}$$

Ejemplo 8.2 (Los espacios vectoriales geométricos)**.** *Los conjuntos $V = \mathbb{VL}^2$ y $V = \mathbb{VL}^3$ de vectores libres en el plano $\mathcal{E}_2$ o espacio $\mathcal{E}_3$ ordinarios con las operaciones definidas en la sección anterior forman $\mathbb{R}$-espacios vectoriales, llamados* **espacios vectoriales geométricos**.

Ejemplo 8.3 (El $\mathbb{K}$-espacio vectorial $\mathbb{K}$)**.** *Un cuerpo $\mathbb{K}$ es un espacio vectorial sobre sí mismo, siendo la multiplicación por escalares la propia multiplicación del cuerpo $\mathbb{K}$. En particular,*

- *$\mathbb{Q}$ es un $\mathbb{Q}$-espacio vectorial.*
- *$\mathbb{R}$ es un $\mathbb{R}$-espacio vectorial.*
- *$\mathbb{C}$ es un $\mathbb{C}$-espacio vectorial.*
- *$\mathbb{Z}_p$ es un $\mathbb{Z}_p$-espacio vectorial (p número primo positivo).*

Considerando el producto cartesiano del cuerpo $\mathbb{R}$ por sí mismo se obtiene

$$\mathbb{R}^2 = \mathbb{R} \times \mathbb{R} = \{(x_1, x_2) : x_1, x_2 \in \mathbb{R}\},$$

que proporciona el siguiente ejemplo, el cual corresponde al modelo algebraico de la Sección 8.2.

Ejemplo 8.4 (El espacio vectorial $(\mathbb{R}^2, +, \cdot)$). *El conjunto*

$$\mathbb{R}^2 = \{(x_1, x_2) : x_1, x_2 \in \mathbb{R}\}$$

con las operaciones de adición $+ : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}^2$ *y multiplicación por escalares* $\cdot : \mathbb{R} \times \mathbb{R}^2 \to \mathbb{R}^2$ *definidas como*

$$\begin{aligned}(x_1, x_2) + (y_1, y_2) &= (x_1 + y_1, x_2 + y_2)\\ \lambda \cdot (x_1, x_2) &= (\lambda x_1, \lambda x_2)\end{aligned}$$

es un $\mathbb{R}$*-espacio vectorial.*

$\lhd$ Es claro que $+$ es una operación interna y que $\cdot$ es una operación externa. Se debe probar que se satisfacen los axiomas $(V_1) - (V_8)$. En efecto, sean $(x_1, x_2), (y_1, y_2), (z_1, z_2) \in \mathbb{R}^2$ y $\lambda, \mu \in \mathbb{R}$.

(V_1) Aplicando la definición de adición y utilizando la asociatividad de $\mathbb{R}$ se tiene

$$\begin{aligned}[(x_1, x_2) + (y_1, y_2)] + (z_1, z_2) &= (x_1 + y_1, x_2 + y_2) + (z_1, z_2)\\ &= ([x_1 + y_1] + z_1, [x_2 + y_2] + z_2)\\ &= (x_1 + [y_1 + z_1], x_2 + [y_2 + z_2])\\ &= (x_1, x_2) + (y_1 + z_1, y_2 + z_2)\\ &= (x_1, x_2) + [(y_1, y_2) + (z_1, z_2)].\end{aligned}$$

(V_2) Existe el vector $(0, 0) \in \mathbb{R}^2$ tal que $(x_1, x_2) + (0, 0) = (x_1 + 0, x_2 + 0) = (x_1, x_2)$ pues 0 es el elemento neutro de la adición en $\mathbb{R}$. De forma similar se prueba que $(0, 0) + (x_1, x_2) = (x_1, x_2)$.

(V_3) Para cada $(x_1, x_2) \in \mathbb{R}^2$, existe $(-x_1, -x_2) \in \mathbb{R}^2$ tal que $(x_1, x_2) + (-x_1, -x_2) = (x_1 + (-x_1), x_2 + (-x_2)) = (0, 0)$ pues $-x_1$ y $-x_2$ son los elementos opuestos de x_1 y x_2 en $\mathbb{R}$, respectivamente. De forma similar se prueba $(-x_1, -x_2) + (x_1, x_2) = (0, 0)$. Luego, el simétrico[8] de

[8]Se probará que es único.

(x_1, x_2) es $-(x_1, x_2) = (-x_1, -x_2)$.

(V_4) Aplicando la definición de adición y la propiedad conmutativa en $\mathbb{R}$ se tiene que $(x_1, x_2) + (y_1, y_2) = (x_1 + y_1, x_2 + y_2) = (y_1 + x_1, y_2 + x_2) = (y_1, y_2) + (x_1, x_2)$.

(V_5) Aplicando las definiciones de adición y multiplicación por escalares y la propiedad distributiva en $\mathbb{R}$ se tiene

$$\begin{aligned} \lambda \cdot [(x_1, x_2) + (y_1, y_2)] &= \lambda \cdot (x_1 + y_1, x_2 + y_2) \\ &= (\lambda(x_1 + y_1), \lambda(x_2 + y_2)) \\ &= (\lambda x_1 + \lambda y_1, \lambda x_2 + \lambda y_2) \\ &= (\lambda x_1, \lambda x_2) + (\lambda y_1, \lambda y_2) \\ &= \lambda \cdot (x_1, x_2) + \lambda \cdot (y_1, y_2). \end{aligned}$$

(V_6) Aplicando las definiciones de adición, multiplicación por escalares y distributiva en $\mathbb{R}$ se tiene que

$$\begin{aligned} (\lambda + \mu) \cdot (x_1, x_2) &= ((\lambda + \mu)x_1, (\lambda + \mu)x_2) \\ &= (\lambda x_1 + \mu x_1, \lambda x_2 + \mu x_2) \\ &= (\lambda x_1, \lambda x_2) + (\mu x_1, \mu x_2) \\ &= \lambda \cdot (x_1, x_2) + \mu \cdot (x_1, x_2). \end{aligned}$$

(V_7) Aplicando la definición de multiplicación por escalares y la propiedad asociativa en $\mathbb{R}$ se tiene que

$$\begin{aligned} (\lambda\mu) \cdot (x_1, x_2) &= ((\lambda\mu)x_1, (\lambda\mu)x_2) \\ &= (\lambda(\mu x_1), \lambda(\mu x_2)) \\ &= \lambda \cdot (\mu x_1, \mu x_2) \\ &= \lambda \cdot [\mu \cdot (x_1, x_2))]. \end{aligned}$$

(V_8) Aplicando la definición de multiplicación por escalares y el hecho que 1 es el elemento neutro para la multiplicación en el cuerpo $\mathbb{R}$ se tiene que $1 \cdot (x_1, x_2) = (1x_1, 1x_2) = (x_1, x_2)$.

Luego, $(\mathbb{R}^2, +, \cdot)$ es un $\mathbb{R}$-espacio vectorial. $\triangleright$

Considerando el producto cartesiano del cuerpo $\mathbb{K}$ por sí mismo n veces, con $n \in \mathbb{N}$, se obtiene

$$\mathbb{K}^n = \underbrace{\mathbb{K} \times \cdots \times \mathbb{K}}_{n \text{ veces}} = \{(x_1, x_2, \ldots, x_n) : x_i \in \mathbb{K}, i = 1, 2, \ldots, n\},$$

que proporciona el siguiente ejemplo, que es una generalización de los modelos algebraicos de la Sección 8.2 (y del ejemplo precedente).

Ejemplo 8.5 (El espacio vectorial de n-tuplas de $\mathbb{K}$)**.** *El conjunto*

$$\mathbb{K}^n = \{(x_1, x_2, \ldots, x_n) : x_i \in \mathbb{K}, i = 1, 2, \ldots, n\}$$

con las operaciones de adición $+ : \mathbb{K}^n \times \mathbb{K}^n \to \mathbb{K}^n$ *y multiplicación por escalares* $\cdot : \mathbb{K} \times \mathbb{K}^n \to \mathbb{K}^n$ *definidas como*

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

es un $\mathbb{K}$*-espacio vectorial, llamado* **espacio cartesiano**.

Ejemplo 8.6 (El espacio vectorial de las matrices)**.** *El conjunto* $\mathbb{K}^{m\times n}$ *de las matrices de tamaño* $m \times n$ *a coeficientes en* $\mathbb{K}$ *es un* $\mathbb{K}$*-espacio vectorial con las operaciones de adición de matrices*

$$+ : \mathbb{K}^{m\times n} \times \mathbb{K}^{m\times n} \to \mathbb{K}^{m\times n}$$

y multiplicación por un escalar

$$\cdot : \mathbb{K} \times \mathbb{K}^{m\times n} \to \mathbb{K}^{m\times n}$$

definidas en el Capítulo 2 (elemento a elemento).

Ejemplo 8.7 (El espacio vectorial de las aplicaciones de un conjunto a un cuerpo)**.** *Sean A un conjunto no vacío y $\mathbb{K}$ un cuerpo. El conjunto*

$$\mathbb{K}^A := \{f : A \to \mathbb{K} / \ f \text{ es una aplicación}\}$$

es un $\mathbb{K}$-espacio vectorial donde las operaciones de adición de aplicaciones $+ : \mathbb{K}^A \times \mathbb{K}^A \to \mathbb{K}^A$ y multiplicación por un escalar $\cdot : \mathbb{K} \times \mathbb{K}^A \to \mathbb{K}^A$ están definidas punto a punto de la siguiente forma

$$\begin{array}{rcl} (f+g)(a) & = & f(a)+g(a), \qquad \forall a \in A \\ (\lambda f)(a) & = & \lambda f(a), \qquad \forall a \in A, \qquad \forall \lambda \in \mathbb{K}. \end{array}$$

El caso particular $A = \mathbb{K} = \mathbb{R}$ corresponde al **$\mathbb{R}$-espacio vectorial de las funciones reales a valores reales**.

Es interesante comparar el ejemplo anterior con la Definición 2.1 de la página 107 de matriz sobre un cuerpo.

Ejemplo 8.8 (El espacio vectorial de las funciones polinomiales sobre un cuerpo)**.** *Sea V el conjunto de todas las funciones $f : \mathbb{K} \to \mathbb{K}$ definidas de la siguiente forma: para algún $n \in \{0\} \cup \mathbb{N}$,*

$$f(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n, \qquad \text{para } x \in \mathbb{K}$$

donde $a_0, a_1, a_2, \ldots, a_n$ son escalares fijos de $\mathbb{K}$ (independientes de x). Una función de este tipo se llama **función polinomial en la variable x** *y a los escalares se los llama* **coeficientes**. *Definiendo la adición y la multiplicación por escalares punto a punto (como en el Ejemplo 8.7), se tiene que V es un $\mathbb{K}$-espacio vectorial que se suele denotar por $\mathbb{K}[x]$. Si n es el mayor valor del subíndice tal que $a_n \neq 0$, se dice que la función polinomial es de* **grado** *n. A la función polinomial nula (que se obtiene cuando todos los coeficientes son nulos) no se le asigna grado.*

La definición de *función polinomial* del ejemplo anterior es de carácter analítico (es decir, utiliza funciones del Análisis Matemático). Cuando se define *polinomio* mediante un enfoque algebraico[9], este espacio vectorial se suele llamar el **espacio vectorial de los polinomios a coeficientes en** $\mathbb{K}$ **en la indeterminada** X y se suele denotar por $\mathbb{K}[X]$. Ambas denominaciones se utilizarán indistintamente pues si los coeficientes están en $\mathbb{Q}, \mathbb{R}$ o $\mathbb{C}$, es posible establecer[10] una correspondencia biunívoca entre polinomios y funciones polinomiales. Por tanto, $\mathbb{K}[X]$ y $\mathbb{K}[x]$ pueden identificarse (es decir, en cierto sentido, pueden considerarse iguales), al menos en estos tres casos.

Ejemplo 8.9 (Un espacio vectorial sobre diferentes cuerpos de escalares)**.** *Considerando las operaciones usuales, se tiene que:*

- *El cuerpo* $V = \mathbb{R}$ *es un* $\mathbb{Q}$*-espacio vectorial (es decir, los vectores se toman de* $V = \mathbb{R}$ *y los escalares de* $\mathbb{K} = \mathbb{Q}$*).*
- *El cuerpo* $V = \mathbb{C}$ *es un espacio vectorial sobre* $\mathbb{K} = \mathbb{Q}$ *y sobre* $\mathbb{K} = \mathbb{R}$*.*

De los Ejemplos 8.3 y 8.9 queda claro que el mismo conjunto de vectores V puede dar lugar a espacios vectoriales diferentes si es considerado sobre diferentes cuerpos de escalares.

Como $\mathbb{Q} \subset \mathbb{R}$ y $\mathbb{R} \subset \mathbb{C}$, se cumple que:

- todo $\mathbb{C}$-espacio vectorial es un $\mathbb{R}$-espacio vectorial,
- todo $\mathbb{R}$-espacio vectorial es un $\mathbb{Q}$-espacio vectorial,
- todo $\mathbb{C}$-espacio vectorial es un $\mathbb{Q}$-espacio vectorial.

[9] Para una definición algebraica de **polinomio** se puede consultar el Anexo C.

[10] Está probado en el Anexo C.

Ejemplo 8.10 (Un espacio vectorial sobre $\mathbb{Q}$). *El conjunto*

$$\mathbb{Q}[\sqrt{2}] := \{a + b\sqrt{2} : \ a, b \in \mathbb{Q}\}$$

es un $\mathbb{Q}$-espacio vectorial si se definen las operaciones mediante

$$\begin{aligned} (a_1 + b_1\sqrt{2}) + (a_2 + b_2\sqrt{2}) &= (a_1 + a_2) + (b_1 + b_2)\sqrt{2}, \ \forall a_1, a_2, b_1, b_2 \in \mathbb{Q} \\ \lambda(a + b\sqrt{2}) &= (\lambda a) + (\lambda b)\sqrt{2}, \qquad \forall a, b \in \mathbb{Q}, \qquad \forall \lambda \in \mathbb{Q}. \end{aligned}$$

Se recuerda que una **sucesión** es una función $x : \mathbb{N} \to A$, donde A es un conjunto no vacío. En lugar de escribir la imagen de n por x como $x(n)$ se escribe x_n y, aplicando x a cada número natural n, la sucesión x se representa como

$$(x_n)_{n\in\mathbb{N}} = (x_1, x_2, \ldots, x_m, \ldots).$$

Ejemplo 8.11 (El espacio vectorial de las sucesiones reales). *El conjunto de todas las sucesiones reales*

$$S := \{(a_n)_{n\in\mathbb{N}} : a_n \in \mathbb{R}, \ \text{para } n \in \mathbb{N}\}$$

es un $\mathbb{R}$-espacio vectorial donde la adición y la multiplicación por escalares se definen mediante

$$\begin{aligned} (a_n)_{n\in\mathbb{N}} + (b_n)_{n\in\mathbb{N}} &= (a_n + b_n)_{n\in\mathbb{N}}, \qquad \forall (a_n)_{n\in\mathbb{N}}, (b_n)_{n\in\mathbb{N}} \in S \\ \lambda(a_n)_{n\in\mathbb{N}} &= (\lambda a_n)_{n\in\mathbb{N}}, \qquad \forall (a_n)_{n\in\mathbb{N}} \in S, \qquad \forall \lambda \in \mathbb{R}. \end{aligned}$$

8.3.2. Consecuencias de la definición

Las siguientes propiedades algebraicas se cumplen en todo espacio vectorial.

Proposición 8.1 (Consecuencias de la definición)**.** Sea V un $\mathbb{K}$-espacio vectorial. Se cumplen las siguientes propiedades:

(a) Unicidad del neutro de la adición: El vector $\vec{0}$ es único.

(b) Unicidad del simétrico: Para cada $\vec{u} \in V$, el simétrico $-\vec{u} \in V$, es único.

(c) Multiplicación de 0 por un vector: $0\vec{u} = \vec{0}$, $\forall \vec{u} \in V$.

(d) Multiplicación de un escalar por el vector nulo: $\lambda\vec{0} = \vec{0}$, $\forall \lambda \in \mathbb{K}$.

(e) Productos nulos: Sean $\lambda \in \mathbb{K}$ y $\vec{u} \in V$. Entonces $\lambda\vec{u} = \vec{0}$ si y sólo si $\lambda = 0$ ó $\vec{u} = \vec{0}$.

(f) Ley de los signos: $-(\lambda\vec{u}) = (-\lambda)\vec{u} = \lambda(-\vec{u})$, $\forall \vec{u} \in V$, $\forall \lambda \in \mathbb{K}$. En particular, $(-1)\vec{u} = -\vec{u}$, $\forall \vec{u} \in V$.

Demostración. (a) Si se supone que existen dos vectores $\vec{0}, \vec{0}' \in V$ tales que $\vec{u} + \vec{0} = \vec{0} + \vec{u} = \vec{u}$ y $\vec{u} + \vec{0}' = \vec{0}' + \vec{u} = \vec{u}$ para todo $\vec{u} \in V$ entonces particularizando en las primeras para $\vec{u} = \vec{0}'$ y en las segundas para $\vec{u} = \vec{0}$ se tiene

$$\vec{0}' = \vec{0} + \vec{0}' = \vec{0}.$$

Luego, $\vec{0}' = \vec{0}$, lo que prueba la unicidad del elemento neutro para $+$.

(b) Sea $\vec{u} \in V$ y sean $\vec{u}_1, \vec{u}_2 \in V$ dos simétricos de $\vec{u}$, es decir, $\vec{u} + \vec{u}_1 = \vec{u}_1 + \vec{u} = \vec{0}$ y $\vec{u} + \vec{u}_2 = \vec{u}_2 + \vec{u} = \vec{0}$. Luego,

$$\vec{u}_1 = \vec{u}_1 + \vec{0} = \vec{u}_1 + (\vec{u} + \vec{u}_2) = (\vec{u}_1 + \vec{u}) + \vec{u}_2 = \vec{0} + \vec{u}_2 = \vec{u}_2.$$

Compárense las demostraciones de (a) y (b) con los Ejercicios 33 y 34 de la página 99.

(c) Sea $\vec{u} \in V$. Entonces $0\vec{u} = (0 + 0)\vec{u} = 0\vec{u} + 0\vec{u}$, donde se utiliza V_6 y que en $\mathbb{K}$ se cumple que $0 = 0 + 0$. Ahora, sumando a ambos miembros el opuesto del vector $0\vec{u}$, aplicando la propiedad asociativa y utilizando la existencia del elemento neutro de la adición, se tiene

$$\vec{0} = -(0\vec{u}) + 0\vec{u} = -(0\vec{u}) + (0\vec{u} + 0\vec{u}) = [-(0\vec{u}) + 0\vec{u}] + 0\vec{u} = \vec{0} + 0\vec{u} = 0\vec{u}.$$

(d) De forma similar a la prueba de la propiedad (c), utilizando V_5 se tiene que si $\lambda \in \mathbb{K}$ entonces

$$\lambda\vec{0} = \lambda(\vec{0} + \vec{0}) = \lambda\vec{0} + \lambda\vec{0}.$$

Ahora, sumando a ambos miembros el opuesto de $\lambda\vec{0}$, aplicando la propiedad asociativa y utilizando la existencia del elemento neutro de la adición se llega a $\lambda\vec{0} = \vec{0}$ (Se propone escribir los detalles como ejercicio).

(e) Se deben probar dos implicaciones.

($\Leftarrow$) Esta implicación se ha probado en los apartados (c) y (d) previos.

($\Rightarrow$) Se supone que $\lambda\vec{u} = \vec{0}$ y se debe probar que $\lambda = 0$ ó $\vec{u} = \vec{0}$. En efecto, si $\lambda = 0$, no hay nada que probar. Sea $\lambda \neq 0$, se debe probar que $\vec{u} = \vec{0}$. Para ello, sabiendo que existe $\lambda^{-1} \in \mathbb{K}$, multiplicando ambos miembros de $\lambda\vec{u} = \vec{0}$ por λ^{-1} y utilizando V_8, V_7 y (c) se tiene

$$\vec{u} = 1_{\mathbb{K}}\vec{u} = (\lambda^{-1}\lambda)\vec{u} = \lambda^{-1}(\lambda\vec{u}) = \lambda^{-1}\vec{0} = \vec{0}.$$

(f) Se probará que $-(\lambda\vec{u}) = \lambda(-\vec{u})$, $\forall\vec{u} \in V$, $\forall\lambda \in \mathbb{K}$ pues la otra igualdad es similar y se propone como ejercicio.

Sean $\vec{u} \in V$ y $\lambda \in \mathbb{K}$. Para probar que el opuesto de $\lambda\vec{u}$ es $\lambda(-\vec{u})$ es necesario ver que su suma (en los dos sentidos) es el vector nulo. En efecto, de los axiomas de espacio vectorial y (d) se tiene

$$\lambda\vec{u} + \lambda(-\vec{u}) = \lambda[\vec{u} + (-\vec{u})] = \lambda\vec{0} = \vec{0}.$$

Al cumplirse la conmutatividad de la adición, la suma en el otro sentido $\lambda(-\vec{u}) + \lambda\vec{u}$ también da el vector nulo. $\square$

El siguiente resultado prueba que, cuando se cumplen los restantes axiomas, es cierta la equivalencia:

$$\text{Axioma } (V_8) \qquad \Leftrightarrow \qquad [\lambda \odot \vec{v} = \vec{0} \text{ si y sólo si } \lambda = 0 \text{ ó } \vec{v} = \vec{0}],$$

lo que permite cambiar el axioma (V_8) de la definición de espacio vectorial por la propiedad equivalente indicada.

Proposición 8.2. Sea V un conjunto, $\oplus$ una operación interna sobre V y $\odot$ una operación externa con operadores en el cuerpo $(\mathbb{K}, +, \cdot)$ (con $1_{\mathbb{K}} \neq 0$). Si se cumplen los axiomas $(V_1) - (V_7)$ de la definición de espacio vectorial entonces

$$(V_8) \text{ es válido} \qquad \Leftrightarrow \qquad [\forall \lambda \in \mathbb{K}, \forall \vec{v} \in V,\ \lambda \odot \vec{v} = \vec{0} \Leftrightarrow \lambda = 0 \text{ ó } \vec{v} = \vec{0}].$$

Demostración. $(\Rightarrow)$ Se ha probado en el apartado (e) de la Proposición 8.1.

$(\Leftarrow)$ Se supone que $\forall \lambda \in \mathbb{K}, \forall \vec{v} \in V, \quad \lambda \odot \vec{v} = \vec{0} \Leftrightarrow \lambda = 0$ ó $\vec{v} = \vec{0}$.

Se debe probar que $1_{\mathbb{K}} \odot \vec{u} = \vec{u}$, $\forall \vec{u} \in V$, lo que es claramente equivalente a $1_{\mathbb{K}} \odot \vec{u} \oplus (-\vec{u}) = \vec{0}$, $\forall \vec{u} \in V$.

En efecto, sea $\vec{u} \in V$. Aplicando los axiomas $(V_1) - (V_7)$ y la hipótesis se tiene

$$\begin{aligned} 1_{\mathbb{K}} \odot [1_{\mathbb{K}} \odot \vec{u} \oplus (-\vec{u})] &= 1_{\mathbb{K}} \odot (1_{\mathbb{K}} \odot \vec{u}) \oplus 1_{\mathbb{K}} \odot (-\vec{u}) \\ &= (1_{\mathbb{K}} \cdot 1_{\mathbb{K}}) \odot \vec{u} \oplus 1_{\mathbb{K}} \odot (-\vec{u}) \\ &= 1_{\mathbb{K}} \odot \vec{u} \oplus 1_{\mathbb{K}} \odot (-\vec{u}) \\ &= 1_{\mathbb{K}} \odot [\vec{u} \oplus (-\vec{u})] \\ &= 1_{\mathbb{K}} \odot \vec{0} \\ &= \vec{0}. \end{aligned}$$

Por hipótesis, y del hecho que $1_{\mathbb{K}} \neq 0$, se tiene $1_{\mathbb{K}} \odot \vec{u} \oplus (-\vec{u}) = \vec{0}$. Sumando a ambos miembros a derecha el vector $\vec{u}$ y operando (a partir de los axiomas válidos), se llega a $1_{\mathbb{K}} \odot \vec{u} = \vec{u}$. □

Se observa que a lo largo de la demostración, de los axiomas $(V_1)-(V_7)$, no se utiliza (V_4). Se lo incluye en el enunciado por consistencia con la utilidad del resultado.

Notación. Con la intención de simplificar la notación, de ahora en más se prescindirá de la flecha encima del vector $\vec{u}$ y un vector se indicará simplemente mediante u.

8.4. Subespacios vectoriales

En esta sección se estudian los subconjuntos de un $\mathbb{K}$-espacio vectorial $(V, +, .)$ que heredan la estructura algebraica de V, esto es, que son espacios vectoriales con las "mismas" operaciones definidas en V; además, tendrán el mismo elemento neutro de la adición que V.

Definición 8.2. Sea $(V, +, .)$ un $\mathbb{K}$-espacio vectorial y sea $S \subseteq V$ un subconjunto no vacío. El conjunto S se llama **subespacio vectorial** o simplemente **subespacio** de V, y se denota $S \leq V$, si S es un $\mathbb{K}$-espacio vectorial con las operaciones de V restringidas a S.

En símbolos, para $\emptyset \neq S \subseteq V$,

$$S \text{ es subespacio vectorial de } (V, +, .) \Leftrightarrow (S, +|_S, .|_S) \text{ es } \mathbb{K}\text{-espacio vectorial.}$$

Se observa que, por definición, las restricciones de las operaciones de V a S cumplen

$$+|_S : S \times S \to V \qquad \text{y} \qquad .|_S : \mathbb{K} \times S \to V.$$

Para que S sea un $\mathbb{K}$-espacio vectorial debe cumplirse que sean leyes de composición interna y externa en S, respectivamente, es decir,

$$+|_S : S \times S \to S \qquad \text{y} \qquad .|_S : \mathbb{K} \times S \to S.$$

Ejemplo 8.12 (Subespacios vectoriales triviales)**.** *Los subconjuntos* $\{0\}$ *y* V *son subespacios vectoriales del* $\mathbb{K}$*-espacio vectorial* V*, llamados* **subespacios triviales** *o* **impropios***, en particular,* $\{0\}$ *se llama* **subespacio trivial nulo***. Los subespacios no triviales se llaman* **propios***.*

Observación 8.4. Si S es un subespacio vectorial de V entonces el elemento neutro 0_S de la adición de S coincide con el elemento neutro 0_V de la adición de V.
En efecto, al ser S (por sí mismo) un espacio vectorial, debe tener un elemento neutro 0_S de la suma en S. Puesto que todo elemento de S está en V, se cumple también que $u + 0_V = 0_V + u = u$ para todo $u \in S$. Como el neutro de la suma de S es único, debe ocurrir que $0_V \in S$ y $0_S = 0_V$. De ahora en adelante, el elemento neutro se denotará simplemente por 0.

Para demostrar que un subconjunto es un subespacio vectorial utilizando la definición anterior, se necesitan demostrar (además de que las operaciones interna y externa en V también lo son en S) los ocho axiomas que definen espacio vectorial. Sin embargo, al ser S un subconjunto de V y tener definidas en S las mismas operaciones que en V, la mayoría de las propiedades de V son válidas en S. Es decir, la propiedad asociativa (V_1) de V es cierta, en particular, para los elementos de S. Del mismo modo, las propiedades conmutativa (V_4), distributivas (V_5) y (V_6), pseudoasociativa (V_7) y la ley modular (V_8) son ciertas para los elementos de S. Por lo tanto, sólo se debe tener especial cuidado con las siguientes: que las operaciones interna y externa en V lo sean en S, la existencia del elemento neutro en S y que cada elemento de S posea simétrico en S.

A continuación, se presenta una caracterización de subespacio vectorial que requiere comprobar que se cumplen sólo tres propiedades.

Proposición 8.3 (Caracterización de subespacio vectorial)**.** Sea $(V, +, .)$ un $\mathbb{K}$-espacio vectorial y sea $S \subseteq V$. Las siguientes condiciones son equivalentes:

(a) S es un subespacio vectorial de V.

(b) Se cumple que:

   (S_1) $0 \in S$,

   (S_2) Si $u, v \in S$ entonces $u + v \in S$,

   (S_3) Si $u \in S$ y $\lambda \in \mathbb{K}$ entonces $\lambda.u \in S$.

Demostración. (a) $\Rightarrow$ (b) Sea S un subespacio vectorial.

De la Observación 8.4, se sabe que el vector nulo 0 de S coincide con el de V. Por tanto, $0 \in S$. Así, (S_1) se cumple.

Los apartados (S_2) y (S_3) son inmediatos, pues S es cerrado para la adición y la multiplicación por escalares de $\mathbb{K}$ ya que S es un $\mathbb{K}$-espacio vectorial.

(b) $\Rightarrow$ (a) Se supone que $(S_1), (S_2)$ y (S_3) son ciertas.

La hipótesis $0 \in S$ implica que $S \neq \emptyset$.

Las hipótesis (S_2) y (S_3) aseguran, respectivamente, que la operación adición es una ley de composición interna en S y la operación multiplicación por escalares de $\mathbb{K}$ es una ley de composición externa en S con operadores en $\mathbb{K}$. Faltan demostrar los axiomas $(V_1) - (V_8)$ de la definición de espacio vectorial.

Al ser V un $\mathbb{K}$-espacio vectorial, $u + 0 = 0 + u = u, \forall u \in V$. En particular, dicho axioma se cumple para todos los elementos de S. Como (por hipótesis) $0 \in S$, se tiene que el elemento 0 es el neutro de la suma en S, lo que asegura la validez de (V_2) en S.

Dado $u \in S$, de la Proposición 8.1, se tiene que $-u = (-1)u \in S$ por (S_3), lo que garantiza que se cumple (V_3) en S.

La validez de los restantes axiomas (en S) se siguen de particularizar a S su cumplimiento válido en todo V. $\square$

8.4.1. Ejemplos de subespacios vectoriales

Ejemplo 8.13. *En el $\mathbb{R}$-espacio vectorial $\mathbb{R}^4$ con la suma y el producto por escalares habituales, se considera el subconjunto*

$$S = \{(x_1, x_2, x_3, x_4) \in \mathbb{R}^4 : \ x_3 = 0 \ \wedge \ x_2 + x_4 = 0\}.$$

Probar que S es un subespacio vectorial de $\mathbb{R}^4$.

$\triangleleft$ Los elementos del conjunto S se pueden reescribir como

$$S = \{(x_1, x_2, 0, -x_2) : \ x_1, x_2 \in \mathbb{R}\}. \tag{8.3}$$

Es claro que $S \neq \emptyset$ pues el vector $(0, 0, 0, 0) \in S$ (tomando $x_1 = x_2 = 0$). Sean ahora $x = (x_1, x_2, x_3, x_4) \in S$, $y = (y_1, y_2, y_3, y_4) \in S$, es decir, $x = (x_1, x_2, 0, -x_2)$, $y = (y_1, y_2, 0, -y_2)$ con $x_1, x_2, y_1, y_2 \in \mathbb{R}$. Entonces

$$\begin{aligned} x + y &= (x_1 + y_1, x_2 + y_2, x_3 + y_3, x_4 + y_4) \\ &= (x_1 + y_1, x_2 + y_2, 0 + 0, -x_2 + (-y_2)) \\ &= (x_1 + y_1, x_2 + y_2, 0, -(x_2 + y_2)), \end{aligned}$$

lo que demuestra que $x + y \in S$ pues su tercera componente es 0 y la cuarta es la opuesta de la segunda, según se requiere en (8.3).

Ahora, si $\lambda \in \mathbb{R}$ y $x = (x_1, x_2, x_3, x_4) \in S$ entonces $x = (x_1, x_2, 0, -x_2)$ con $x_1, x_2 \in \mathbb{R}$. Luego,

$$\begin{aligned} \lambda x &= (\lambda x_1, \lambda x_2, \lambda x_3, \lambda x_4) \\ &= (\lambda x_1, \lambda x_2, 0, \lambda(-x_2)) \\ &= (\lambda x_1, \lambda x_2, 0, -(\lambda x_2)), \end{aligned}$$

lo que prueba que $\lambda x \in S$ puesto que su tercera componente es 0 y la cuarta es la opuesta de la segunda, según se requiere en (8.3). La Proposición 8.3 asegura que S es un subespacio vectorial de $\mathbb{R}^4$. $\triangleright$

Ejemplo 8.14 (Un conjunto que no es subespacio vectorial)**.** *Comprobar que el conjunto*

$$S = \{(x, y) \in \mathbb{R}^2 : \, xy = 0\}$$

no es un subespacio del $\mathbb{R}$-espacio vectorial $\mathbb{R}^2$ con las operaciones habituales.

$\triangleleft$ Se tiene que $(1, 0), (0, 1) \in S$ y, sin embargo, $(0, 1) + (1, 0) = (1, 1) \notin S$. Observar que se cumplen las condiciones (S_1) y (S_3) de la Proposición 8.3 pero no se cumple (S_2). Luego, S no es un subespacio[11] del espacio vectorial $\mathbb{R}^2$. $\triangleright$

Ejemplo 8.15 (Subespacio de funciones polinomiales)**.** *Probar que el conjunto $\mathbb{R}[x]$ de todas las funciones polinomiales a coeficientes en $\mathbb{R}$ es un subespacio del $\mathbb{R}$-espacio vectorial de todas las funciones reales a variable real del tipo $f : \mathbb{R} \to \mathbb{R}$ con las operaciones habituales.*

$\triangleleft$ Se propone comprobarlo como ejercicio. $\triangleright$

[11]Más preciso que la expresión “no es un subespacio del espacio vectorial” sería decir que la estructura algebraica del espacio vectorial $\mathbb{R}^2$ con la suma y el producto por escalares habituales, inducida en el subconjunto S, no lo convierte en subespacio vectorial, pero por sencillez se acepta la expresión entrecomillada.

Ejemplo 8.16 (Subespacio de funciones polinomiales de grado menor o igual que n)**.** *Para cada $n \in \{0\} \cup \mathbb{N}$, probar que el conjunto de todas las funciones polinomiales a coeficientes en $\mathbb{R}$ dado por*

$$\mathbb{R}_n[x] = \{P(x) \in \mathbb{R}[x] : P = 0 \ \vee \ \text{gr}(P) \leq n\}$$

es un subespacio del $\mathbb{R}$-espacio vectorial de todas las funciones reales a variable real del tipo $f : \mathbb{R} \to \mathbb{R}$ con las operaciones habituales.

$\lhd$ Se deben comprobar las propiedades de la Proposición 8.3. En efecto,

(S_1) La función polinomial nula pertenece a $\mathbb{R}_n[x]$.

(S_2) Sean $P, Q \in \mathbb{R}_n[x]$. Debido a la naturaleza de los elementos de $\mathbb{R}_n[x]$, se deben distinguir casos:

- Si $P = 0$ ó $Q = 0$ entonces $P + Q \in \{P, Q\} \subseteq \mathbb{R}_n[x]$.
- Sean $P \neq 0$ y $Q \neq 0$.
    - Si $P + Q = 0$ entonces $P + Q \in \mathbb{R}_n[x]$.
    - Si $P + Q \neq 0$ entonces $P(x) = a_0 + a_1 x + \cdots + a_m x^m$ y $Q(x) = b_0 + b_1 x + \cdots + b_s x^s$ para ciertos $m, s \in \{0, 1, 2, \ldots, n\}$ donde todos los coeficientes están en $\mathbb{R}$, con $a_m \neq 0 \neq b_s$. Luego, $\text{gr}(P + Q) \leq \text{mín}\{m, s\} \leq n$. Así, $P + Q \in \mathbb{R}_n[x]$.

(S_3) Sean $P \in \mathbb{R}_n[x]$ y $\lambda \in \mathbb{R}$.

- Si $\lambda = 0$ ó $P = 0$ entonces $\lambda P = 0 \in \mathbb{R}_n[x]$.
- Si $\lambda \neq 0$ y $P \neq 0$ entonces $\lambda P \neq 0$ y $\text{gr}(\lambda P) = \text{gr}(P) \leq n$. Así, $\lambda P \in \mathbb{R}_n[x]$.

Por lo tanto, $\mathbb{R}_n[x]$ es un subespacio vectorial de $\mathbb{R}^{\mathbb{R}}$. $\rhd$

Más aún, es fácil ver que se cumple que

$$\mathbb{R}_n[x] \leq \mathbb{R}[x] \leq \mathbb{R}^{\mathbb{R}}.$$

Ejemplo 8.17 (Subespacios vectoriales de $\mathbb{R}^2$)**.** *Probar que los subespacios de $\mathbb{R}^2$ son de una de las tres formas siguientes:*

(a) $\{(0,0)\}$,

(b) $\{\lambda u : \lambda \in \mathbb{R}\}$, *para* $u \in \mathbb{R}^2 - \{(0,0)\}$, *(es decir, los que geométricamente representan una recta que pasa por el origen),*

(c) $\mathbb{R}^2$.

$\triangleleft$ Según se observó en el Ejemplo 8.12, dos subespacios de $\mathbb{R}^2$ con las operaciones habituales son los triviales $\{(0,0)\}$ y $\mathbb{R}^2$. Se trata pues de encontrar la forma de los subespacios no triviales.

Si S es un subespacio no trivial de $\mathbb{R}^2$, existe un vector no nulo $u_0 = (x_0, y_0) \in S$. Por ser S un subespacio, también contiene a todos los vectores de la forma $\lambda u_0 = (\lambda x_0, \lambda y_0)$, para todo $\lambda \in \mathbb{R}$, es decir $\{\lambda u_0 : \lambda \in \mathbb{R}\} \subseteq S$. Como la suma de dos vectores del tipo λu_0 con $\lambda \in \mathbb{R}$ vuelve a ser del mismo tipo y el vector $(0,0)$ también lo es (pues corresponde a $\lambda = 0$), se tiene que

$$\{\lambda u_0 : \lambda \in \mathbb{R}\} \subseteq S$$

es un subespacio vectorial de $\mathbb{R}^2$. Geométricamente, dicho conjunto representa la recta L que pasa por el origen $(0,0)$ $(\lambda = 0)$ y por el punto de coordenadas (x_0, y_0) $(\lambda = 1)$.

Además, se cumple que $S \subseteq \{\lambda u_0 : \lambda \in \mathbb{R}\}$. En efecto, sea $v_0 \in S$.

- Si $v_0 = 0$, es claro que $v_0 = 0u_0 \in \{\lambda u_0 : \lambda \in \mathbb{R}\}$.
- Sea pues $v_0 \neq 0$. Si, por reducción al absurdo, $v_0 \notin \{\lambda u_0 : \lambda \in \mathbb{R}\}$ entonces v_0 tendría una dirección diferente a la de u_0 (es decir, $v_0 \neq \mu u_0$ para todo $\mu \in \mathbb{R}$). Dado que $u_0, v_0 \in S$ y $S \leq \mathbb{R}^2$, se tiene que $\alpha u_0 + \beta v_0 \in S$, para todo $\alpha, \beta \in \mathbb{R}$. Como se vio al estudiar el modelo algebraico $\mathbb{R}^2$, todo vector de $\mathbb{R}^2$ se puede escribir en términos de u_0 y v_0 como $\alpha u_0 + \beta v_0$, para ciertos $\alpha, \beta \in \mathbb{R}$. Entonces se ha probado

que $\mathbb{R}^2 \subseteq S$, lo que lleva a una contradicción pues S coincidiría con el subespacio trivial $\mathbb{R}^2$.

Por tanto, $S = \{\lambda u_0 : \lambda \in \mathbb{R}\}$. Luego, sólo hay un tipo de subespacios no triviales en $\mathbb{R}^2$: las rectas que pasan por el origen.

En resumen, los subespacios de $\mathbb{R}^2$ son de la forma: (a) $\{(0,0)\}$, (b) $\{\lambda u : \lambda \in \mathbb{R}\}$ para $0 \neq u \in \mathbb{R}^2$ o bien (c) $\mathbb{R}^2$. $\triangleright$

Ejemplo 8.18 (Subespacios vectoriales de $\mathbb{R}^3$)**.** *Probar que los subespacios vectoriales de $\mathbb{R}^3$ con las operaciones habituales son de una de las cuatro formas siguientes:*

(a) $\{(0,0,0)\}$,

(b) $\{\lambda u : \lambda \in \mathbb{R}\}$ *para* $(0,0,0) \neq u \in \mathbb{R}^3$ *(es decir, los que geométricamente representan una recta que pasa por el origen),*

(c) $\{\lambda u + \mu v : \lambda, \mu \in \mathbb{R}\}$ *para* $u, v \in \mathbb{R}^3 - \{(0,0,0)\}$, $u \neq \alpha v$, *para todo* $\alpha \in \mathbb{R}$ *(es decir, los que geométricamente representan un plano que pasa por el origen),*

(d) $\mathbb{R}^3$.

$\triangleleft$ Se propone como ejercicio. $\triangleright$

El próximo ejemplo es de especial relevancia.

Ejemplo 8.19 (Subespacio de soluciones de un sistema de ecuaciones lineales homogéneo)**.** *Sea* $A \in \mathbb{K}^{m\times n}$. *Probar que el conjunto*

$$S = \{x \in \mathbb{K}^{n\times 1} : \, Ax = 0_{m\times 1}\}$$

es un subespacio vectorial del $\mathbb{K}$*-espacio vectorial de* $\mathbb{K}^{n\times 1}$.

$\triangleleft$ De las propiedades de matrices es fácil ver que $\mathbb{K}^{n\times 1}$ es un $\mathbb{K}$-espacio vectorial.

Para probar que S es un subespacio del espacio vectorial de $\mathbb{K}^{n\times 1}$ se aplicará la Proposición 8.3.

En efecto, sean $x_1, x_2 \in S$ y sea $\lambda \in \mathbb{K}$. Entonces $Ax_1 = 0$ y $Ax_2 = 0$. Luego,

(S_1) $A0 = 0 \Rightarrow 0 \in S$.

(S_2) $A(x_1 + x_2) = Ax_1 + Ax_2 = 0 + 0 = 0 \Rightarrow x_1 + x_2 \in S$.

(S_3) $A(\lambda x_1) = \lambda(Ax_1) = \lambda 0 = 0 \Rightarrow \lambda x_1 \in S$.

Luego, S es subespacio de $\mathbb{K}^{n\times 1}$. $\triangleright$

Ejemplo 8.20 (Otro subconjunto que no es subespacio vectorial)**.** *Se considera el $\mathbb{R}$-espacio vectorial de todas las funciones reales definidas en el intervalo $[0, 1]$ con la adición y la multiplicación por escalares habituales (punto a punto), que se denota $\mathcal{F}_{[0,1]}$ o bien $\mathbb{R}^{[0,1]}$. Probar que el conjunto*

$$C = \{f \in \mathcal{F}_{[0,1]} : \ f(0) = 1\}$$

no es un subespacio de $\mathcal{F}_{[0,1]}$.

$\triangleleft$ Para probar que C no es un subespacio de $\mathcal{F}_{[0,1]}$ se aplicará la Proposición 8.3.

En primer lugar, la función nula O, evaluada en 0 toma el valor 0, y no 1. Luego, $O \notin C$.

Puesto que $f(x) = x+1$ y $g(x) = \cos(x)$ cumplen que $f(0) = 1 = g(0)$, se tiene que $f, g \in C$ pero $f+g \notin C$ pues $(f+g)(0) = f(0)+g(0) = 1+1 = 2 \neq 1$.

Además, si $\alpha \in \mathbb{R}$ entonces $(\alpha f)(0) = \alpha f(0) = \alpha 1 = \alpha$ y esto vale 1 si y sólo si $\alpha = 1$. Luego, $\alpha f \notin C$, $\forall \alpha \neq 1$. Luego, C no es un subespacio de $\mathcal{F}_{[0,1]}$ (no se cumple ninguna de las tres condiciones aunque, evidentemente, alcanza con que una no se cumpla para concluir). $\triangleright$

8.5. Combinaciones lineales. Subespacio generado

Las propiedades asociativa y conmutativa de la suma de vectores en un espacio vectorial permiten escribir, sin lugar a confusión, las siguientes igualdades: $u + (v + w) + x + y = u + v + w + (x + y) = u + [v + (x + w)] + y = u + v + w + x + y$, donde la última expresión carece de paréntesis.

La operación por excelencia que se puede realizar en un espacio vectorial es la que se define a continuación y es fundamental a lo largo del Álgebra Lineal.

Definición 8.3. Sea V un $\mathbb{K}$-espacio vectorial y sean $u_1, u_2, \ldots, u_n \in V$. Un vector $u \in V$ se dice que es **combinación lineal** de $u_1, u_2, \ldots, u_n$ si existen escalares $\alpha_1, \alpha_2, \ldots, \alpha_n \in \mathbb{K}$ tales que

$$u = \alpha_1 u_1 + \alpha_2 u_2 + \cdots + \alpha_n u_n = \sum_{k=1}^{n} \alpha_k u_k.$$

Ejemplo 8.21. *En el espacio vectorial $\mathbb{R}^4$ con la suma y el producto por escalares habituales, el vector $u = (2, 0, 3, 0)$ es combinación lineal de $u_1 = (1, 0, 0, 0)$ y $u_2 = (0, 0, 1, 0)$ pues $u = 2u_1 + 3u_2$.*

Ejemplo 8.22. *Comprobar que, en el espacio vectorial $\mathbb{R}^{2\times 2}$ con la adición y la multiplicación por escalares habituales, el vector $A = \begin{bmatrix} 1 & 0 \\ 5 & -6 \end{bmatrix}$ es combinación lineal de*

$$A_1 = \begin{bmatrix} 1 & 2 \\ -1 & 0 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 1 & 3 \\ -4 & 3 \end{bmatrix} \quad y \quad A_3 = \begin{bmatrix} 1 & 5 \\ 4 & -2 \end{bmatrix}$$

y que, sin embargo, A no es combinación lineal de A_1 y A_3.

$\triangleleft$ Para verlo, se deben buscar escalares $a, b, c \in \mathbb{R}$ tales que $A = aA_1 + bA_2 + cA_3$ en el primer caso y ver que no existen escalares a y c tales que $A = aA_1 + cA_3$ en el segundo. Resolviendo los sistemas de ecuaciones lineales que se obtienen al igualar matrices en cada caso se encuentra que $A = 3A_1 - 2A_2 + 0A_3$, mientras que A no es combinación lineal de A_1 y A_3 pues el sistema obtenido es incompatible. Se propone completar los detalles como ejercicio. $\triangleright$

El conjunto que contiene todas las expresiones de este tipo es especialmente importante.

Proposición 8.4 (Subespacio generado por $u_1, u_2, \ldots, u_n$)**.** Sea V un $\mathbb{K}$-espacio vectorial y sean $u_1, u_2, \ldots, u_n \in V$. El conjunto

$$\overline{\{u_1, u_2, \ldots, u_n\}} := \{\alpha_1 u_1 + \alpha_2 u_2 + \cdots + \alpha_n u_n : \alpha_1, \alpha_2, \ldots, \alpha_n \in \mathbb{K}\}$$

formado por todas las combinaciones lineales de $u_1, u_2, \ldots, u_n$ es un subespacio de V. Se llama **subespacio generado**[a] por los vectores $u_1, u_2, \ldots, u_n$.

[a] También se suele llamar la **envolvente lineal**, **envoltura lineal** o **clausura lineal** de $u_1, u_2, \ldots, u_n$.

Demostración. Se deben probar las tres propiedades de la Proposición 8.3.

(S_1) Tomando $\alpha_1 = \alpha_2 = \cdots = \alpha_n = 0$ se tiene que $0 = \alpha_1 u_1 + \alpha_2 u_2 + \cdots + \alpha_n u_n \in \overline{\{u_1, u_2, \ldots, u_n\}}$.

(S_2) Si $u, v \in \overline{\{u_1, u_2, \ldots, u_n\}}$ entonces existen escalares $\alpha_1, \alpha_2, \ldots, \alpha_n \in \mathbb{K}$ y $\beta_1, \beta_2, \ldots, \beta_n \in \mathbb{K}$ tales que

$$u = \alpha_1 u_1 + \alpha_2 u_2 + \cdots + \alpha_n u_n \quad \text{y} \quad v = \beta_1 u_1 + \beta_2 u_2 + \cdots + \beta_n u_n.$$

Luego, aplicando propiedades de espacio vectorial

$$\begin{aligned} u + v &= (\alpha_1 u_1 + \alpha_2 u_2 + \cdots + \alpha_n u_n) + (\beta_1 u_1 + \beta_2 u_2 + \cdots + \beta_n u_n) \\ &= (\alpha_1 u_1 + \beta_1 u_1) + (\alpha_2 u_2 + \beta_2 u_2) + \cdots + (\alpha_n u_n + \beta_n u_n) \\ &= (\alpha_1 + \beta_1) u_1 + (\alpha_2 + \beta_2) u_2 + \cdots + (\alpha_n + \beta_n) u_n. \end{aligned}$$

Luego, $u + v \in \overline{\{u_1, u_2, \ldots, u_n\}}$ pues $\alpha_i + \beta_i \in \mathbb{K}$ para todo $i = 1, 2, \ldots, n$.

(S_3) Sea $u \in \overline{\{u_1, u_2, \ldots, u_n\}}$ y $\lambda \in \mathbb{K}$. Entonces existen escalares α_1, $\alpha_2, \ldots, \alpha_n \in \mathbb{K}$ tales que $u = \alpha_1 u_1 + \alpha_2 u_2 + \cdots + \alpha_n u_n$. De nuevo, aplicando propiedades de espacio vectorial,

$$\begin{aligned} \lambda u &= \lambda(\alpha_1 u_1 + \alpha_2 u_2 + \cdots + \alpha_n u_n) \\ &= \lambda(\alpha_1 u_1) + \lambda(\alpha_2 u_2) + \cdots + \lambda(\alpha_n u_n) \\ &= (\lambda\alpha_1) u_1 + (\lambda\alpha_2) u_2 + \cdots + (\lambda\alpha_n) u_n \end{aligned}$$

Por lo tanto, $\lambda u \in \overline{\{u_1, u_2, \ldots, u_n\}}$ pues $\lambda\alpha_i \in \mathbb{K}$ para todo $i = 1, 2, \ldots, n$.

$\square$

El conjunto $\overline{\{u_1, u_2, \ldots, u_n\}}$, definido en la Proposición 8.4, es el conjunto de todas las combinaciones lineales obtenidas a partir de $u_1, u_2, \ldots, u_n$.

Ejemplo 8.23. *Hallar el subespacio generado por el vector $u = (1, 3)$ en el espacio vectorial $\mathbb{R}^2$ con la adición y multiplicación por escalares habituales.*

$\triangleleft$ Como no se indica otra cosa de forma explícita, se trata del $\mathbb{R}$-espacio vectorial $\mathbb{R}^2$. El subespacio generado por el vector $u = (1, 3)$ es

$$\begin{aligned} \overline{\{u\}} &= \{\alpha u : \alpha \in \mathbb{R}\} \\ &= \{\alpha(1, 3) : \alpha \in \mathbb{R}\} \\ &= \{(\alpha, 3\alpha) : \alpha \in \mathbb{R}\}. \end{aligned}$$

Geométricamente, representa la recta en el plano $\mathbb{R}^2$ que pasa por el origen en la dirección del vector $(1, 3)$. En efecto, llamando (x, y) a un elemento genérico de $\overline{\{u\}}$ se tiene que $(x, y) = (\alpha, 3\alpha)$, para cualquier $\alpha \in \mathbb{R}$, es decir, corresponde a la ecuación paramétrica de dicha recta:

$$\left\{ \begin{aligned} x &= \alpha \\ y &= 3\alpha \end{aligned} \right. , \quad \alpha \in \mathbb{R}.$$

$\triangleright$

Ejemplo 8.24. *Encontrar el subespacio generado por los vectores $u = (1,0,0)$ y $v = (0,1,0)$ en el espacio vectorial $\mathbb{R}^3$ con la adición y multiplicación por escalares habituales.*

◁ Como no se indica otra cosa de forma explícita, se trata del $\mathbb{R}$-espacio vectorial $\mathbb{R}^3$. El subespacio generado por los vectores $u = (1,0,0)$ y $v = (0,1,0)$ está dado por

$$\begin{aligned}\overline{\{u,v\}} &= \{\alpha u + \beta v : \alpha,\beta \in \mathbb{R}\}\\ &= \{\alpha(1,0,0) + \beta(0,1,0) : \alpha,\beta \in \mathbb{R}\}\\ &= \{(\alpha,\beta,0) : \alpha,\beta \in \mathbb{R}\}.\end{aligned}$$

Geométricamente, representa en $\mathbb{R}^3$ el plano de ecuación $z = 0$ (pues α y β varían arbitrariamente en $\mathbb{R}$). ▷

Ejemplo 8.25. *Encontrar el subespacio generado por las funciones polinomiales $P(x) = 1 + x$ y $Q(x) = x - 2$ en el espacio vectorial $\mathbb{R}_3[x]$ con la adición y multiplicación por escalares habituales.*

◁ El subespacio generado por las funciones polinomiales $P(x) = 1 + x$ y $Q(x) = x - 2$ en el espacio vectorial $\mathbb{R}_3[x]$ está dado por

$$\begin{aligned}\overline{\{P,Q\}} &= \{\alpha P(x) + \beta Q(x) \in \mathbb{R}_3[x] : \alpha,\beta \in \mathbb{R}\}\\ &= \{\alpha(1+x) + \beta(x-2) \in \mathbb{R}_3[x] : \alpha,\beta \in \mathbb{R}\}\\ &= \{(\alpha - 2\beta) + (\alpha+\beta)x \in \mathbb{R}_3[x] : \alpha,\beta \in \mathbb{R}\}\\ &= \{a + bx \in \mathbb{R}_3[x] : a,b \in \mathbb{R}\}\\ &= \mathbb{R}_1[x].\end{aligned}$$

Se ha probado que, dentro del espacio vectorial $\mathbb{R}_3[x]$, los polinomios $P(x) = 1 + x$ y $Q(x) = x - 2$ generan todos los polinomios de grado menor o igual que 1 junto al polinomio nulo.

Es importante responder claramente la siguiente pregunta (se propone como ejercicio): ¿Por qué es posible asegurar la igualdad de conjuntos

$$\{(\alpha - 2\beta) + (\alpha + \beta)x \in \mathbb{R}_3[x] : \ \alpha, \beta \in \mathbb{R}\} = \{a + bx \in \mathbb{R}_3[x] : \ a, b \in \mathbb{R}\}?$$

Una forma de responderla es planteando el sistema lineal

$$\begin{bmatrix} 1 & -2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix},$$

y estudiar la singularidad (o no) de la matriz de coeficientes. $\triangleright$

Ejemplo 8.26. *Hallar el subespacio generado por los vectores*

$$A_1 = \begin{bmatrix} 2 & 0 \\ 1 & 0 \end{bmatrix}, \qquad A_2 = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} \qquad y \qquad A_3 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

en el espacio vectorial $\mathbb{R}^{2\times 2}$ *con la adición y multiplicación por escalares habituales.*

$\triangleleft$ El subespacio generado por los vectores A_1, A_2 y A_3 está dado por

$$\begin{aligned}
\overline{\{A_1, A_2, A_3\}} &= \{\alpha A_1 + \beta A_2 + \gamma A_3 : \ \alpha, \beta, \gamma \in \mathbb{R}\} \\
&= \left\{ \alpha \begin{bmatrix} 2 & 0 \\ 1 & 0 \end{bmatrix} + \beta \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} + \gamma \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} : \ \alpha, \beta, \gamma \in \mathbb{R} \right\} \\
&= \left\{ \begin{bmatrix} 2\alpha + \beta + \gamma & 0 \\ \alpha + \beta & 0 \end{bmatrix} : \ \alpha, \beta, \gamma \in \mathbb{R} \right\} \\
&= \left\{ \begin{bmatrix} a & 0 \\ b & 0 \end{bmatrix} : \ a, b \in \mathbb{R} \right\},
\end{aligned}$$

es decir, se pueden generar todas las matrices con elementos arbitrarios en las posiciones $(1, 1)$ y $(2, 1)$ mientras que en las posiciones $(1, 2)$ y $(2, 2)$ siempre habrá un 0.

De nuevo, es importante tener clara la igualdad entre los últimos dos conjuntos anteriores; los detalles se proponen como ejercicio. $\triangleright$

Ejemplo 8.27. *Probar que el subespacio generado por el conjunto* $\{1\}$ *en el* $\mathbb{K}$*-espacio vectorial* $\mathbb{K}$ *es el propio* $\mathbb{K}$.

$\triangleleft$ Se debe probar que $\overline{\{1\}} = \mathbb{K}$.

En efecto, para probar esta igualdad de conjuntos, primero se observa que la inclusión $\overline{\{1\}} \subseteq \mathbb{K}$ es siempre válida. Para probar la otra inclusión de conjuntos basta observar que un elemento arbitrario $k \in \mathbb{K}$ se puede escribir como el producto $k = k1$, con el escalar $k \in \mathbb{K}$, por ser 1 el elemento neutro del cuerpo $\mathbb{K}$. Luego, $\mathbb{K} \subseteq \overline{\{1\}}$. $\triangleright$

Ejemplo 8.28. *Se considera el conjunto* $X = \{i\} \subseteq \mathbb{C}$. *Encontrar el subespacio generado por* X *en:*

(a) el $\mathbb{C}$*-espacio vectorial* $\mathbb{C}$.

(b) el $\mathbb{R}$*-espacio vectorial* $\mathbb{C}$.

$\triangleleft$ En el $\mathbb{C}$-espacio vectorial $\mathbb{C}$ se tiene que $\overline{X} = \mathbb{C}$ (Se propone comprobar como ejercicio la doble inclusión).

Sin embargo, en el $\mathbb{R}$-espacio vectorial $\mathbb{C}$ se tiene que

$$\overline{X} = \{b\,i \in \mathbb{C} : \, b \in \mathbb{R}\},$$

es decir, X genera el conjunto de los números imaginarios puros, que a veces se denota mediante $i\,\mathbb{R}$. $\triangleright$

Una consecuencia inmediata de la definición de subespacio generado es la siguiente caracterización.

Proposición 8.5 (Caracterización de subespacio generado por $\{u_1, u_2, \ldots, u_n\}$)**.** Sea V un $\mathbb{K}$-espacio vectorial y sean $\{u_1, u_2, \ldots, u_n\}$ y S dos subconjuntos de V. Las siguientes condiciones son equivalentes:

(a) $S = \overline{\{u_1, u_2, \ldots, u_n\}}$.

(b) S satisface las siguientes propiedades:

(I) S es un subespacio de V.

(II) $\{u_1, u_2, \ldots, u_n\} \subseteq S$.

(III) Si S' es un subespacio de V tal que $\{u_1, u_2, \ldots, u_n\} \subseteq S'$ entonces $S \subseteq S'$.

Demostración. (a) $\Rightarrow$ (b) Sea $S = \overline{\{u_1, u_2, \ldots, u_n\}}$. Se deben probar las 3 condiciones (I)-(III). En efecto,

(I) De la Proposición 8.4 es inmediato que S es un subespacio.

(II) Como $u_i = 0u_1 + \cdots + 0u_{i-1} + 1u_i + 0u_{i+1} + \cdots + 0u_n \in \overline{\{u_1, u_2, \ldots, u_n\}}$ para $i = 1, 2, \ldots, n$, es claro que $\{u_1, u_2, \ldots, u_n\} \subseteq S$.

(III) Sea S' un subespacio de V tal que $\{u_1, u_2, \ldots, u_n\} \subseteq S'$. Se debe probar que $S \subseteq S'$. En efecto, sea $x \in S$. Como por hipótesis $S = \overline{\{u_1, u_2, \ldots, u_n\}}$, existen escalares $\alpha_1, \alpha_2, \ldots, \alpha_n \in \mathbb{K}$ tales que

$$x = \alpha_1 u_1 + \alpha_2 u_2 + \cdots + \alpha_n u_n.$$

Como $\{u_1, u_2, \ldots, u_n\} \subseteq S'$ y $S' \leq V$ se tiene que $x \in S'$ (pues S' debe ser cerrado para las combinaciones lineales de elementos de $\{u_1, u_2, \ldots, u_n\}$). Luego, $S \subseteq S'$.

(b) $\Rightarrow$ (a) Se deben probar: $S \subseteq \overline{\{u_1, u_2, \ldots, u_n\}}$ y que $S \supseteq \overline{\{u_1, u_2, \ldots, u_n\}}$. En efecto,

($\subseteq$) Sea $S' := \overline{\{u_1, u_2, \ldots, u_n\}}$. Por Proposición 8.4, S' es un subespacio que, además, contiene a $\{u_1, u_2, \ldots, u_n\}$ (se demuestra como en el caso (II) de (a) $\Rightarrow$ (b)). El apartado (III) implica que $S \subseteq S' = \overline{\{u_1, u_2, \ldots, u_n\}}$.

($\supseteq$) Por (I) e (II), $S \leq V$ y $\{u_1, u_2, \ldots, u_n\} \subseteq S$. Se debe probar que $\overline{\{u_1, u_2, \ldots, u_n\}} \subseteq S$. En efecto, sea $s \in \overline{\{u_1, u_2, \ldots, u_n\}}$. Por definición, existen escalares $\alpha_1, \alpha_2, \ldots, \alpha_n \in \mathbb{K}$ tales que

$$x = \alpha_1 u_1 + \alpha_2 u_2 + \cdots + \alpha_n u_n.$$

Como $\{u_1, u_2, \ldots, u_n\} \subseteq S$ y $S \leq V$, se tiene que $x \in S$ (pues S debe ser cerrado para las combinaciones lineales de elementos de $\{u_1, u_2, \ldots, u_n\}$). Por lo tanto, $\overline{\{u_1, u_2, \ldots, u_n\}} \subseteq S$. $\square$

Este resultado se puede parafrasear diciendo que el subespacio generado por $u_1, u_2, \ldots, u_n$ es el menor (en el sentido de la inclusión) subespacio que contiene a $u_1, u_2, \ldots, u_n$.

Observación 8.5. De la Proposición 8.5 se observa que siempre se cumple que $\{u_1, u_2, \ldots, u_n\} \subseteq \overline{\{u_1, u_2, \ldots, u_n\}}$.

La definición de combinación lineal puede ampliarse a una familia de vectores (no necesariamente finita) de un espacio vectorial como sigue. Se debe recordar la Definición 1.27 en la que se observó que una familia de vectores también se suele llamar **sistema de vectores**.

Definición 8.4. Sea V un $\mathbb{K}$-espacio vectorial y sea $U = \{u_j\}_{j \in I} \subseteq V$ un sistema (o familia) de vectores de V. Un vector $u \in V$ se dice que es **combinación lineal** de U si, para cada $j \in I$ existe un escalar $\alpha_j \in \mathbb{K}$ y, además, existe un subconjunto finito $J_0 \subseteq I$ de forma que

$$u = \sum_{j \in I} \alpha_j u_j,$$

donde $\alpha_j = 0$ para todo $I - J_0$. Es decir, $u = \sum_{j \in J_0} \alpha_j u_j$.
Si $J_0 = \{j_1, j_2, \ldots, j_k\}$ entonces u se escribe como

$$u = \alpha_{j_1} u_{j_1} + \alpha_{j_2} u_{j_2} + \cdots + \alpha_{j_k} u_{j_k}.$$

Observar que, en definitiva, en los sumatorios sólo aparece un número finito de términos no nulos. De este modo, la definición anterior se puede parafrasear diciendo que: un elemento u es combinación lineal de un sistema de vectores U si u es combinación lineal de algún subconjunto finito de U.

Ejemplo 8.29. *Se considera la familia* $U = \{x^j\}_{j\in\{0\}\cup\mathbb{N}}$ *de funciones polinomiales en* $\mathbb{Q}[x]$. *Entonces las siguientes son combinaciones lineales de* U*:*

$$0, \qquad \frac{3}{5}x^2 + 7x^6 - x^{100}, \qquad x^{1000} - \frac{1}{3}x^{2000},$$

mientras que $\frac{1}{3} + \sqrt{2}x^2$ *no es combinación lineal de* U *en* $\mathbb{Q}[x]$ *aunque sí lo es en* $\mathbb{R}[x]$.

Del mismo modo, el concepto de subespacio generado puede ampliarse al caso de un sistema arbitrario (no necesariamente finito) de vectores.

Definición 8.5. Sea V un $\mathbb{K}$-espacio vectorial y sea $U = \{u_j\}_{j\in I} \subseteq V$ un sistema (o familia) de vectores de V. Se llama **subespacio generado** por U, y se denota $\overline{U}$, al conjunto de todas las combinaciones lineales de cualesquiera de los subconjuntos finitos de U. Es decir, el subespacio generado por U es el sistema de todos los vectores de V que son combinación lineal (finita) de U.

Observación 8.6. La expresión *combinación lineal* en la última oración de la Definición 8.5 es la introducida en la Definición 8.4 (que es más general que la considerada en la Definición 8.3). Aunque es redundante, se añade *(finita)* para enfatizar su importancia en dicha definición.

Por convención, $\overline{\emptyset} := \{0\}$.

Con la Definición 8.5 de *subespacio generado por una familia de vectores* se sigue cumpliendo que $\overline{U}$ es el menor subespacio de V que contiene a U.

Ejemplo 8.30. *Se considera la familia* $U = \{x^j\}_{j\in\{0\}\cup\mathbb{N}}$ *de funciones polinomiales en* $\mathbb{R}[x]$. *El subespacio generado por* U *es el sistema de todas las combinaciones lineales de subconjuntos finitos de* U *que da como resultado el conjunto de todas las funciones polinomiales de* $\mathbb{R}[x]$, *es decir,* $\overline{U} = \mathbb{R}[x]$.

De aquí en adelante, cuando aparezcan las expresiones *combinación lineal* y *subespacio generado* deberán interpretarse de acuerdo a las Definiciones 8.4 y 8.5, por ser más generales que la Definición 8.3 y el concepto definido en la Proposición 8.4.

8.6. Sistemas de generadores

Al estudiar el conjunto de todas las combinaciones lineales obtenidas a partir de un sistema de vectores $u_1, u_2, \ldots, u_n$ de un $\mathbb{K}$-espacio vectorial V, se probó que forman un subespacio $\overline{\{u_1, u_2, \ldots, u_n\}}$, llamado el subespacio generado por $u_1, u_2, \ldots, u_n$.

Ejemplo 8.31. *Se consideran los vectores* $u_1 = (1, 2, 2)$ *y* $u_2 = (2, 1, 1)$ *de* $\mathbb{R}^3$. *El subespacio generado por ambos está formado por vectores de la forma* $(x, y, z) = \alpha u_1 + \beta u_2$, *con* $\alpha, \beta \in \mathbb{R}$. *Es decir,*

$$(x, y, z) = \alpha(1, 2, 2) + \beta(2, 1, 1) = (\alpha + 2\beta, 2\alpha + \beta, 2\alpha + \beta),$$

de donde se obtiene que el sistema

$$\left\{\begin{array}{lcl} \alpha + 2\beta & = & x \\ 2\alpha + \beta & = & y \\ 2\alpha + \beta & = & z \end{array}\right.$$

debe ser compatible, para lo que claramente debe ser $y = z$. *Si bien, de las dos primeras ecuaciones, es posible obtener* α *y* β *en función de*

x e y, no es necesario para lo que sigue. Se ha obtenido que

$$\overline{\{u_1, u_2\}} = \{(x, y, y) \in \mathbb{R}^3\}.$$

Puesto que $(x, y, y) = x(1,0,0) + y(0,1,1)$, los vectores $(1,0,0)$ y $(0,1,1)$ generan en $\mathbb{R}^3$ el mismo subespacio que u_1 y u_2. Como casos particulares, se nota que $(3,9,9) \in \overline{\{u_1, u_2\}}$ y $(1,2,3) \notin \overline{\{u_1, u_2\}}$.

Del ejemplo se observa que tan sólo con dos vectores de $\mathbb{R}^3$ es posible obtener los demás vectores del subespacio $\overline{\{u_1, u_2\}}$ (¡que tiene infinitos vectores!). Además, es lo mismo utilizar los vectores u_1, u_2 dados originalmente, que los vectores $(1,0,0)$ y $(0,1,1)$, para obtener el subespacio $\overline{\{u_1, u_2\}}$.

Definición 8.6. Sea V un $\mathbb{K}$-espacio vectorial. Se llama **sistema de generadores**[a] de V a todo sistema de vectores $X \subseteq V$ tal que el subespacio generado por X es V. En símbolos,

$$X \text{ es un sistema de generadores de } V \qquad \Leftrightarrow \qquad \overline{X} = V.$$

Cuando X es un conjunto finito, se dice que V es **finitamente generado**.

[a] O también, **sistema generador**.

De la definición se tiene que:

X es un sistema de generadores de $V \Longleftrightarrow$ todo vector de V se puede escribir como combinación lineal de los elementos de X.

Ejemplo 8.32. *Al ser $\mathbb{R}^2 = \overline{\{e_1 := (1,0), e_2 := (0,1)\}}$ se tiene que $\{e_1, e_2\}$ es un sistema de generadores de $\mathbb{R}^2$. Luego, $\mathbb{R}^2$ es finitamente generado.*

$\lhd$ Se tiene que $\overline{\{e_1, e_2\}} = \mathbb{R}^2$ pues todo vector $(x, y) \in \mathbb{R}^2$ se puede escribir como $(x, y) = x(1, 0) + y(0, 1)$. $\triangleright$

Observación 8.7. Si bien se ha definido el concepto de sistema de generadores de un espacio vectorial, al ser un subespacio, un espacio vectorial por sí mismo, es claro que la misma noción es válida en subespacios.

Ejemplo 8.33. *En el $\mathbb{R}$-espacio vectorial $\mathbb{R}^2$,*

$$\{u = (1, 0), v = (0, 1), w = (0, 2)\}$$

es un sistema de generadores.

$\triangleleft$ Se tiene que $\overline{\{u, v, w\}} = \mathbb{R}^2$ pues todo vector $(x, y) \in \mathbb{R}^2$ se puede escribir como $(x, y) = x(1, 0) - y(0, 1) + y(0, 2)$. $\triangleright$

Ejemplo 8.34. *Los vectores $(1, 0, 0)$ y $(0, 1, 1)$ forman un sistema de generadores del subespacio $\{(x, y, y) : x, y \in \mathbb{R}\}$ de $\mathbb{R}^3$.*

$\triangleleft$ Se ha probado en el Ejemplo 8.31. $\triangleright$

Ejemplo 8.35. *Un sistema de generadores de $\mathbb{K}^n$ es*

$$\{e_1 := (1, 0, \dots, 0), e_2 := (0, 1, \dots, 0), \dots, e_n := (0, 0, \dots, 1)\}.$$

Luego, $\mathbb{K}^n$ es finitamente generado.

$\triangleleft$ Se prueba que $\mathbb{K}^n = \overline{\{e_1, e_2, \dots, e_n\}}$ pues todo vector $(x_1, x_2, \dots, x_n)$ de $\mathbb{K}^n$ puede escribirse como $(x_1, x_2, \dots, x_n) = x_1 e_1 + x_2 e_2 + \cdots + x_n e_n$. Luego, $\mathbb{K}^n$ es un espacio vectorial finitamente generado. $\triangleright$

Ejemplo 8.36. *En el espacio vectorial $\mathbb{K}[x]$ de todas las funciones polinomiales sobre $\mathbb{K}$ se considera el subconjunto*

$$X = \{p_0(x) = 1, p_1(x) = x, p_2(x) = x^2, \dots, p_n(x) = x^n, \dots\}.$$

Probar que $\overline{X} = \mathbb{K}[x]$. Luego, el espacio vectorial $\mathbb{K}[x]$ no es finitamente generado.

$\lhd$ Probar que $\overline{X} = \overline{\{x^k\}}_{k\in\{0\}\cup\mathbb{N}} = \mathbb{K}[x]$ significa ver que X es un sistema de generadores de $\mathbb{K}[x]$. Este hecho es válido pues toda función polinomial (nula o no) se puede escribir como combinación lineal (finita) de los elementos de X. $\rhd$

Ejemplo 8.37. *En el espacio vectorial $\mathbb{K}^{m\times n}$ de las matrices de tamaño $m\times n$ a coeficientes en $\mathbb{K}$, sea el subconjunto $X := \{E^{(i,j)}\}_{\substack{i=1,\dots,m\\ j=1,\dots,n}}$ donde*

$$(E^{(i,j)})_{rs} = \begin{cases} 1, & \text{si } r = i, s = j \\ 0, & \text{en otro caso} \end{cases} = [(\delta_{r,i}\delta_{s,j})_{rs}].$$

Probar que $\overline{X} = \mathbb{K}^{m\times n}$. Luego, el espacio vectorial $\mathbb{K}^{m\times n}$ es finitamente generado.

$\lhd$ Para probar que $\overline{X} = \mathbb{K}^{m\times n}$ basta observar que toda matriz de $\mathbb{K}^{m\times n}$ se puede escribir como combinación lineal de los elementos de X.

Por ejemplo, en el espacio $\mathbb{R}^{2\times 3}$, el elemento de la posición $(r, s) = (1, 2)$ de la matriz

$$E^{(1,3)} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} = [(\delta_{r,1}\delta_{s,3})_{rs}]$$

es 0 y el elemento de la posición $(r, s) = (1, 3)$ es 1. Además,

$$\begin{bmatrix} 0 & 2 & -1 \\ 4 & 0 & 0 \end{bmatrix} = 2E^{(1,2)} + (-1)E^{(1,3)} + 4E^{(2,1)}.$$

En general,

$$A = [a_{ij}]_{m\times n} = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}E^{(i,j)}.$$

Luego, X es un sistema de generadores de $\mathbb{K}^{m\times n}$. Por tanto, $\mathbb{K}^{m\times n}$ es un espacio vectorial finitamente generado. $\triangleright$

Ejemplo 8.38. *Averiguar si el conjunto*

$$\{u = (1, 1, 0), v = (1, 0, 1), w = (4, 2, 2)\}$$

es un sistema de generadores del espacio vectorial $\mathbb{R}^3$.

$\triangleleft$ Para que se cumpla que $\mathbb{R}^3 = \overline{\{u, v, w\}}$, todo vector de $\mathbb{R}^3$ se tiene que escribir como combinación lineal de u, v y w. En efecto, si $(x, y, z) \in \mathbb{R}^3$ satisface

$$(x, y, z) = a(1, 1, 0) + b(1, 0, 1) + c(4, 2, 2),$$

para ciertos escalares $a, b, c \in \mathbb{R}$ entonces el siguiente sistema de ecuaciones lineales debe ser compatible

$$\begin{cases} a + b + 4c = x \\ a + 2c = y \\ b + 2c = z \end{cases}$$

Para que esto ocurra, aplicando el método de eliminación de Gauss,

$$\left[\begin{array}{ccc|c} 1 & 1 & 4 & x \\ 1 & 0 & 2 & y \\ 0 & 1 & 2 & z \end{array}\right] \sim_f \left[\begin{array}{ccc|c} 1 & 0 & 2 & y \\ 0 & 1 & 2 & x-y \\ 0 & 0 & 0 & -x+y+z \end{array}\right],$$

por tanto, se debe satisfacer la relación (equivalente)

$$x - y - z = 0.$$

Es decir, no todo vector $(x, y, z) \in \mathbb{R}^3$ se puede generar con los vectores dados, sino sólo los que verifican la relación $x - y - z = 0$. Desde el punto de vista geométrico, la relación encontrada corresponde a la ecuación cartesiana de un plano que pasa por el origen.

En definitiva,

$$\overline{\{u, v, w\}} = \{(x, y, x - y) :\ x, y \in \mathbb{R}\} \neq \mathbb{R}^3.$$

Además, se observa que

$$\overline{\{u, v, w\}} = \overline{\{(1, 0, 1), (0, 1, -1)\}},$$

es decir, no son necesarios los tres vectores u, v y w de partida para conseguir el subespacio que ellos generan. ▷

El siguiente resultado muestra que es posible eliminar los vectores superfluos en un sistema de generadores.

A partir de este resultado será posible conseguir sistemas de generadores minimales con respecto a la inclusión de conjuntos, es decir, de forma tal que si se quita un elemento del sistema de generadores, deje de generar el espacio en cuestión (puesto que todos sus vectores son necesarios).

Lema 8.1 (Eliminación de vectores superfluos en un sistema generador)**.** Sea V un $\mathbb{K}$-espacio vectorial y sea el sistema de vectores $\{u_1, u_2, \ldots, u_n, u_{n+1}\} \subseteq V$. Entonces

$$u_{n+1} \in \overline{\{u_1, u_2, \ldots, u_n\}} \quad \Leftrightarrow \quad \overline{\{u_1, u_2, \ldots, u_n, u_{n+1}\}} = \overline{\{u_1, u_2, \ldots, u_n\}}.$$

Demostración. Al trararse de una caracterización se deberán probar las dos implicaciones ($\Rightarrow$) y ($\Leftarrow$).

($\Rightarrow$) Se deben demostrar dos inclusiones de conjuntos.

($\subseteq$) Sea $v \in \overline{\{u_1, u_2, \ldots, u_n, u_{n+1}\}}$. Existen escalares $\alpha_1, \alpha_2, \ldots, \alpha_n, \alpha_{n+1} \in \mathbb{K}$ tales que $v = \sum_{k=1}^{n+1} \alpha_k u_k$.

Por hipótesis, $u_{n+1} \in \overline{\{u_1, u_2, \ldots, u_n\}}$. Entonces u_{n+1} es combinación lineal de los vectores $u_1, u_2, \ldots, u_n$. Es decir, existen escalares $\beta_1, \beta_2, \ldots, \beta_n \in$

$\mathbb{K}$ tales que $u_{n+1} = \sum_{k=1}^{n} \beta_k u_k$. Luego, aplicando las propiedades distributivas, pseudoasociativa, asociativa y conmutativa,

$$\begin{aligned} v &= \sum_{k=1}^{n+1} \alpha_k u_k \\ &= \sum_{k=1}^{n} \alpha_k u_k + \alpha_{n+1} u_{n+1} \\ &= \sum_{k=1}^{n} \alpha_k u_k + \alpha_{n+1} \sum_{k=1}^{n} \beta_k u_k \\ &= \sum_{k=1}^{n} (\alpha_k + \alpha_{n+1}\beta_k) u_k. \end{aligned}$$

En definitiva, las combinaciones lineales obtenidas a partir de los vectores de $\{u_1, u_2, \ldots, u_n, u_{n+1}\}$ se han podido reescribir en términos de los vectores de $\{u_1, u_2, \ldots, u_n\}$ y, por lo tanto, $v \in \overline{\{u_1, u_2, \ldots, u_n\}}$. Así, se prueba que

$$\overline{\{u_1, u_2, \ldots, u_n, u_{n+1}\}} \subseteq \overline{\{u_1, u_2, \ldots, u_n\}}.$$

($\supseteq$) Por definición, si $v \in \overline{\{u_1, u_2, \ldots, u_n\}}$ entonces v se puede escribir como combinación lineal de los elementos de $\{u_1, u_2, \ldots, u_n\}$. Es claro que v se puede escribir, además, como combinación lineal de los elementos de $\{u_1, u_2, \ldots, u_n, u_{n+1}\}$, sin más que añadir a la combinación lineal anterior el término $0u_{n+1}$. Luego,

$$\overline{\{u_1, u_2, \ldots, u_n\}} \subseteq \overline{\{u_1, u_2, \ldots, u_n, u_{n+1}\}}.$$

($\Leftarrow$) Al ser $u_{n+1} = 0u_1 + 0u_2 + \cdots + 0u_n + 1u_{n+1}$, de

$$u_{n+1} \in \overline{\{u_1, u_2, \ldots, u_n, u_{n+1}\}} = \overline{\{u_1, u_2, \ldots, u_n\}}$$

se tiene que $u_{n+1} \in \overline{\{u_1, u_2, \ldots, u_n\}}$. $\square$

Es posible demostrar una versión más general del resultado anterior en que no se considere sólo un número finito de vectores, sino que se considere un sistema arbitrario de vectores.

Teorema 8.1 (Teorema de reducción de sistemas de generadores)**.** Sea V un $\mathbb{K}$-espacio vectorial. Si S es un sistema de generadores de V tal que $u \in S$ es combinación lineal de vectores de $S-\{u\}$ entonces $S-\{u\}$ es también un sistema de generadores de V.

Demostración. Por hipótesis se tiene que $\overline{S} = V$, $u \in S$ y $u \in \overline{S - \{u\}}$. Se debe probar que $\overline{S - \{u\}} = V$. Como la inclusión $\overline{S - \{u\}} \subseteq V$ es obvia, sólo se necesita probar la otra inclusión.

Sea $v \in V$. Por ser $V = \overline{S}$, existe un conjunto finito $\{u_1, u_2, \dots, u_m\} \subseteq S$ tal que v es combinación lineal de ellos, es decir, $v = \alpha_1 u_1 + \alpha_2 u_2 + \cdots + \alpha_m u_m$, para ciertos escalares $\alpha_1, \alpha_2, \dots, \alpha_m \in \mathbb{K}$. Ahora, pueden ocurrir dos casos:

- $u \in \{u_1, u_2, \dots, u_m\}$. En este caso, si fuese por ejemplo $u = u_1$ entonces $\{u_2, \dots, u_m\} \subseteq S - \{u\}$. Ahora, como $u \in \overline{S - \{u\}}$, u es combinación lineal de $S - \{u\}$, por tanto,

$$v \in \overline{\{u_1, u_2, \dots, u_m\}} = \overline{\{u, u_2, \dots, u_m\}} \subseteq \overline{S - \{u\}}.$$

- $u \notin \{u_1, u_2, \dots, u_m\}$. En este caso, $\{u_1, u_2, \dots, u_m\} \subseteq S - \{u\}$, y así, $\overline{\{u_1, u_2, \dots, u_m\}} \subseteq \overline{S - \{u\}}$, con lo cual $v \in \overline{S - \{u\}}$.

Luego, $V \subseteq \overline{S - \{u\}}$, y por lo tanto, $\overline{S - \{u\}} = V$. □

Ejemplo 8.39. *¿Son necesarios los tres vectores*

$$\{u_1 = (1,5), u_2 = (1,1), u_3 = (0,1)\}$$

para generar todo $\mathbb{R}^2$? Justificar.

◁ No. En efecto, al ser $u_1 = u_2 + 4u_3$ se tiene que $\overline{\{u_1, u_2, u_3\}} = \overline{\{u_2, u_3\}}$. Además, $\forall (x,y) \in \mathbb{R}^2$ se tiene que $(x,y) = x(1,1) + (y-x)(0,1)$, con lo cual $\overline{\{u_2, u_3\}} = \mathbb{R}^2$. ▷

8.7. Dependencia e independencia lineal

En el espacio vectorial $\mathbb{R}^2$ se consideran los vectores

$$u = (2,3), \qquad v = (1,0) \qquad y \qquad w = (0,1).$$

Se observa que $(2,3) = 2(1,0)+3(0,1)$, con lo que u es combinación lineal de v y w (se dirá que u depende linealmente de v y w). Esto se expresa diciendo que en la combinación lineal

$$0 = (-1)u + 2v + 3w,$$

no todos los escalares son nulos (en este caso, los tres son diferentes de cero).

Sin embargo, de la expresión

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

se obtiene una única posibilidad para los escalares: $\alpha = \beta = 0$ (se dirá que v y w son linealmente independientes).

Otra observación es que un vector arbitrario $(x,y) \in \mathbb{R}^2$ se puede escribir como

$$(x,y) = 1(2,3) + (x-2)(1,0) + (y-3)(0,1)$$

y también como

$$(x,y) = 2(2,3) + (x-4)(1,0) + (y-6)(0,1).$$

Es decir, está claro que $\overline{\{u,v,w\}} = \mathbb{R}^2$, pero también se observa que hay, al menos, dos formas de escribir a (x,y) como combinación lineal de u, v y w. El objetivo de esta sección es encontrar el menor (en el sentido de la inclusión) sistema de vectores que permita realizar este tipo de representaciones (es decir, que generen el espacio dado).

En el caso de que sólo se dispusiera de los vectores v y w, todo vector $(x,y) \in \mathbb{R}^2$ se puede escribir como combinación lineal de ellos de una única forma, a saber, $(x,y) = x(1,0) + y(0,1)$.

A continuación se abordan estos dos problemas de forma general. La idea de reducir la cantidad de vectores de un sistema de generadores de un espacio (finitamente generado), a la menor posible, se consigue añadiendo la condición de independencia lineal. Cuando ambas condiciones se cumplan, la cantidad de vectores necesaria será fija (y, por tanto, un invariante del espacio) y se denominará dimensión del espacio en cuestión.

Observar que siempre es posible escribir al vector nulo como combinación lineal de un sistema finito de vectores; basta tomar todos los escalares nulos. La distinción entre casos, según se aprecia en los ejemplos anteriores, viene cuando esa forma trivial de escritura es la única o no.

Definición 8.7. Sea V un $\mathbb{K}$-espacio vectorial. Se dice que un conjunto $\{u_1, u_2, \ldots, u_n\}$ de vectores de V es **linealmente independiente** si la única forma de escribir el vector nulo de V como combinación lineal de $u_1, u_2, \ldots, u_n$ es con todos los escalares nulos. Es decir, si para $\alpha_1, \alpha_2, \ldots, \alpha_n \in \mathbb{K}$:

$$0 = \alpha_1 u_1 + \alpha_2 u_2 + \cdots + \alpha_n u_n \qquad \Rightarrow \qquad \alpha_1 = \alpha_2 = \cdots = \alpha_n = 0. \tag{8.7}$$

Si un conjunto $\{u_1, u_2, \ldots, u_n\}$ no es linealmente independiente, se llama **linealmente dependiente**; es decir, si es posible encontrar una combinación lineal como la de (8.7) con no todos los escalares nulos.

Para decidir si un conjunto (finito) de vectores de un espacio vectorial es linealmente independiente o no, se plantea una combinación lineal de esos elementos y si es posible deducir que todos los escalares deben ser nulos se concluye que son linealmente independientes y si algún escalar puede ser no nulo, el conjunto es linealmente dependiente.

Ejemplo 8.40. *Analizar si el subconjunto*

$$\{u = (1, 1, 0), v = (1, 0, 1), w = (4, 2, 2)\}$$

del espacio vectorial $\mathbb{R}^3$ *es linealmente independiente o no. Justificar. ¿Y si se considera el subconjunto* $\{u = (1, 1, 0), v = (1, 0, 1)\}$*?*

$\lhd$ Para decidir si el subconjunto $\{u, v, w\}$ es linealmente independiente o no en $\mathbb{R}^3$ se debe plantear

$$(0, 0, 0) = a(1, 1, 0) + b(1, 0, 1) + c(4, 2, 2), \text{ con } a, b, c \in \mathbb{R}.$$

Realizando la multiplicación por escalares, sumando vectores e igualando componentes se obtiene el siguiente sistema de ecuaciones lineales homogéneo

$$\left\{\begin{array}{ccccccc} a & + & b & + & 4c & = & 0 \\ a & & & + & 2c & = & 0 \\ & & b & + & 2c & = & 0 \end{array}\right.$$

Resolviendo el sistema se obtiene $a = -2c$, $b = -2c$. Al no haber obtenido (únicamente) la solución trivial $(a, b, c) = (0, 0, 0)$, el sistema dado es linealmente dependiente. Por ejemplo, tomando $c = 1$ se tiene que $(0, 0, 0) = -2(1, 1, 0) - 2(1, 0, 1) + (4, 2, 2)$, o bien para resaltar la dependencia lineal se puede escribir, por ejemplo, $(4, 2, 2) = 2(1, 1, 0) + 2(1, 0, 1)$.

En cambio, si se considera el subconjunto

$$\{u = (1, 1, 0), v = (1, 0, 1)\},$$

resolviendo $(0, 0, 0) = a(1, 1, 0) + b(1, 0, 1)$, con $a, b \in \mathbb{R}$ se tiene $a = b = 0$, con lo que $\{u, v\}$ es linealmente independiente en $\mathbb{R}^3$.

Se plantea como ejercicio: ¿Qué ocurre con los subconjuntos $\{u, w\}$ y $\{v, w\}$? ¿Cuál es la conclusión? $\rhd$

Ejemplo 8.41. *Decidir si el subconjunto*

$$\{P_0(x) = 1,\ P_1(x) = x + 2,\ P_2(x) = 4x^2,\ P_3(x) = x^3 + 1\}$$

del espacio vectorial $\mathbb{R}_3[x]$ *es linealmente independiente o no.*

$\lhd$ Si existen escalares $a, b, c, d \in \mathbb{R}$ tales que

$$0 = O(x) = a1 + b(x + 2) + c(4x^2) + d(x^3 + 1),\ \text{para todo } x \in \mathbb{R}, \quad (8.8)$$

entonces operando se tiene

$$0 = (a + 2b + d) + bx + 4cx^2 + dx^3,\ \text{para todo } x \in \mathbb{R}. \quad (8.9)$$

En particular, la igualdad se cumple para $x = 0$, con lo que

$$a + 2b + d = 0. \quad (8.10)$$

Además, particularizando (8.9) en $x = 1$ y $x = -1$, se tiene

$$(a + 2b + d) + b + 4c + d = 0 \quad \text{y} \quad (a + 2b + d) - b + 4c - d = 0,$$

o, por (8.10), de forma simplificada

$$b + 4c + d = 0 \qquad \text{y} \qquad -b + 4c - d = 0.$$

De estas últimas dos ecuaciones se obtiene $c = 0$ y $d = -b$, y de (8.10) queda $a = -b$. En definitiva, se tiene $b(x - x^3) = 0$, $\forall x \in \mathbb{R}$. En particular, para $x = 2$ se tiene que $-6b = 0$ de donde $b = 0$. Así, $a = b = c = d = 0$, es decir, (8.8) sólo admite la solución trivial. Luego, el conjunto propuesto es linealmente independiente en $\mathbb{R}_3[x]$. $\rhd$

Si en el ejemplo anterior, en lugar de utilizar la igualdad de funciones polinomiales y particularizar en valores adecuados de $x \in \mathbb{R}$, se hubiese utilizado la definición de polinomio del Apéndice C, donde se indica que dos polinomios son iguales si lo son sus respectivos coeficientes, los valores de

$a, b, c, d \in \mathbb{R}$ se podrían haber hallado resolviendo el sistema de ecuaciones lineales

$$\left\{\begin{array}{ccccccccc} a & + & 2b & & + & d & = & 0 \\ & & b & & & & = & 0 \\ & & & & 4c & & = & 0 \\ & & & & & d & = & 0 \end{array}\right.$$

que, evidentemente, proporciona la misma solución, e incluso, de forma más rápida.

Ejemplo 8.42. *Decidir si el subconjunto*

$$\{f(x) = e^x, g(x) = e^{3x}\}$$

del $\mathbb{R}$-espacio vectorial de todas las funciones derivables V de $\mathbb{R}$ en $\mathbb{R}$ es linealmente independiente o no.

$\lhd$ Si existen escalares $a, b \in \mathbb{R}$ tales que

$$0 = O(x) = ae^x + be^{3x}, \text{ para todo } x \in \mathbb{R},$$

derivando con respecto a la variable x, se tiene $0 = O(x) = ae^x + 3be^{3x}$, $\forall x \in \mathbb{R}$. Particularizando en ambas igualdades en $x = 0$, se debe resolver el sistema lineal

$$\left\{\begin{array}{ccccc} a & + & b & = & 0 \\ a & + & 3b & = & 0 \end{array}\right.$$

cuya solución viene dada por $a = b = 0$. Por lo tanto, el conjunto propuesto es linealmente independiente en V. $\rhd$

Observación 8.8. El truco de considerar la derivada se ha realizado con la intención de conseguir una segunda ecuación lineal, y poder resolver así un sistema de 2 ecuaciones lineales con 2 incógnitas.

Ejemplo 8.43. *Decidir si el subconjunto*

$$\left\{A = \begin{bmatrix} 1 & 3 \\ 0 & 2 \end{bmatrix}, B = \begin{bmatrix} 0 & 1 \\ 0 & 4 \end{bmatrix}, C = \begin{bmatrix} 0 & 1 \\ 0 & 3 \end{bmatrix}\right\}$$

del $\mathbb{R}$-espacio vectorial $\mathbb{R}^{2\times 2}$ es linealmente independiente o no.

$\lhd$ Si existen escalares $a, b, c \in \mathbb{R}$ tales que

$$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = a \begin{bmatrix} 1 & 3 \\ 0 & 2 \end{bmatrix} + b \begin{bmatrix} 0 & 1 \\ 0 & 4 \end{bmatrix} + c \begin{bmatrix} 0 & 1 \\ 0 & 3 \end{bmatrix},$$

operando e igualando matrices se debe resolver el sistema lineal homogéneo

$$\begin{cases} a & & & & & = & 0 \\ 3a & + & b & + & c & = & 0 \\ 2a & + & 4b & + & 3c & = & 0 \end{cases}$$

cuya solución viene dada por $a = b = c = 0$. Por lo tanto, el conjunto propuesto es linealmente independiente en $\mathbb{R}^{2\times 2}$. $\rhd$

Ejemplo 8.44. *Analizar si el conjunto $X = \{1, i\} \subset \mathbb{C}$ es linealmente independiente en el $\mathbb{K}$-espacio vectorial $\mathbb{C}$ siendo: (a) $\mathbb{K} = \mathbb{R}$, (b) $\mathbb{K} = \mathbb{C}$.*

$\lhd$ Sea $X = \{1, i\}$. Si considera el $\mathbb{R}$-espacio vectorial $\mathbb{C}$ (es decir, $\mathbb{K} = \mathbb{R}$) entonces X es linealmente independiente pues si $0+0.i = a.1+b.i$, con $a, b \in \mathbb{R}$ entonces $a = b = 0$. Si $\mathbb{K} = \mathbb{C}$ entonces X es linealmente dependiente pues $0 = (-i).1 + 1.i$, con los escalares $-i, 1 \in \mathbb{C}$ no nulos. $\rhd$

El siguiente ejemplo es difícil y sólo se incluye a título informativo.

Ejemplo 8.45. *El conjunto $X = \{1, \pi, \pi^2, \ldots, \pi^n, \ldots\}$ es un conjunto linealmente independiente en el $\mathbb{Q}$-espacio vectorial $\mathbb{R}$.*

$\lhd$ Se analiza, por ejemplo, el subconjunto finito $Y = \{1, \pi, \pi^2, \ldots, \pi^n\} \subset X$. Si el polinomio nulo se escribe como combinación lineal de los elementos de Y con coeficientes en $\mathbb{Q}$ entonces

$$0 = q_0 + q_1\pi + q_2\pi^2 + \cdots + q_n\pi^n \text{ con } q_0, q_1, q_2, \ldots, q_n \in \mathbb{Q}$$

entonces π sería raíz del polinomio

$$p(x) = q_0 + q_1x + q_2x^2 + \cdots + q_nx^n \in \mathbb{Q}[x],$$

con lo cual π sería un número algebraico sobre $\mathbb{Q}$, y sin embargo, π es trascendente[12] sobre $\mathbb{Q}$. Por lo tanto, $q_0 = q_1 = q_2 = \cdots = q_n = 0$, de donde se prueba que X es un conjunto linealmente independiente en el $\mathbb{Q}$-espacio vectorial $\mathbb{R}$.

Los demás subconjuntos finitos se analizan de forma similar, aunque con una notación un tanto más tediosa. $\rhd$

Ejemplo 8.46. *Analizar si el conjunto*

$$X = \{(1+\sqrt{2}, -3+\sqrt{2}), (4+3\sqrt{2}, -4-\sqrt{2})\}$$

es linealmente independiente en el $\mathbb{K}$-espacio vectorial $\mathbb{R}^2$ para $\mathbb{K} = \mathbb{R}$ y para $\mathbb{K} = \mathbb{Q}$.

$\lhd$ Se considera una combinación lineal del tipo

$$(0,0) = a(1+\sqrt{2}, -3+\sqrt{2}) + b(4+3\sqrt{2}, -4-\sqrt{2}), \text{ con } a, b \in \mathbb{K}.$$

Si $\mathbb{K} = \mathbb{R}$, se debe resolver el sistema lineal homogéneo cuya matriz de coeficientes satisface[13]

$$\begin{bmatrix} 1+\sqrt{2} & 4+3\sqrt{2} \\ -3+\sqrt{2} & -4-\sqrt{2} \end{bmatrix} \sim_f \begin{bmatrix} 1 & 2+\sqrt{2} \\ 0 & 0 \end{bmatrix}.$$

[12]Un número complejo que satisface una ecuación del tipo $f(x) = 0$ con $f(x) = a_0 + a_1x + \cdots + a_nx^n \in \mathbb{Z}[x]$ se llama **número algebraico** sobre $\mathbb{Q}$; un número que no es algebraico se llama **trascendente** sobre $\mathbb{Q}$. La demostración de que π es trascendente sobre $\mathbb{Q}$ escapa el alcance de este libro.

[13]También podría comprobarse que su determinante es igual a cero.

Luego, el sistema correspondiente tiene soluciones no triviales y así X es linealmente dependiente en el $\mathbb{R}$-espacio vectorial $\mathbb{R}^2$. De hecho,

$$(2+\sqrt{2})(1+\sqrt{2}, -3+\sqrt{2}) = (4+3\sqrt{2}, -4-\sqrt{2}).$$

Si $\mathbb{K} = \mathbb{Q}$ entonces, de la primera componente de la combinación lineal, se tiene $a(1+\sqrt{2}) + b(4+3\sqrt{2}) = 0$, de donde

$$(a+4b) + (a+3b)\sqrt{2} = 0.$$

Se consideran dos casos:

- si $a+3b \neq 0$ entonces $\sqrt{2} = -\frac{a+4b}{a+3b} \in \mathbb{Q}$, lo que es una contradicción.

- si $a+3b = 0$ entonces $a+4b = 0$, de donde se obtiene $a = b = 0$, que también satisface la relación entre las segundas componentes de la combinación lineal.

Así, X es linealmente independiente en el $\mathbb{Q}$-espacio vectorial $\mathbb{R}^2$. $\triangleright$

La definición de dependencia e independencia lineal se puede extender a una familia de vectores (no necesariamente finita) de un espacio vectorial como sigue. En este punto es importante repasar la Definición 8.4, que dice que un elemento u es combinación lineal de un sistema de vectores U si u es combinación lineal de algún *subconjunto finito* de U. Un subconjunto U de un $\mathbb{K}$-espacio vectorial V se llamará *linealmente independiente* si $U = \emptyset$ o bien si todo subconjunto finito de U es linealmente independiente. Más formalmente, se tiene la siguiente definición.

Definición 8.8. Sea V un $\mathbb{K}$-espacio vectorial y $U = \{u_j\}_{j\in I}$ una familia de vectores de V. Se dice que la familia U es **linealmente independiente** si es vacía o si la única forma de escribir el vector nulo de V como combinación lineal de U es con todos los escalares nulos. En símbolos, esta segunda condición se escribe del siguiente modo: si

$$0 = \sum_{j\in I} \alpha_j u_j, \quad \text{con } \alpha_j \in \mathbb{K}, \forall j \in I \qquad \Rightarrow \qquad \alpha_j = 0, \forall j \in I,$$

o bien, para cada subconjunto finito $J_0 = \{j_1, j_2, \ldots, j_k\} \subseteq I$ se debe cumplir que

$$0 = \alpha_{j_1} u_{j_1} + \alpha_{j_2} u_{j_2} + \cdots + \alpha_{j_k} u_{j_k} \quad \Rightarrow \quad \alpha_{j_\ell} = 0, \text{ para } \ell = 1, 2, \ldots, k.$$

Si una familia no es linealmente independiente se llama **linealmente dependiente**. Es decir, una familia $U = \{u_j\}_{j\in I}$ de vectores de V es linealmente dependiente si es no vacía y existe algún subconjunto finito $J_0 = \{j_1, j_2, \ldots, j_k\} \subseteq I$ tal que $\{u_{j_1}, u_{j_2}, \ldots, u_{j_k}\}$ es linealmente dependiente.

Ejemplo 8.47. *Sea $\mathbb{K} \in \{\mathbb{Q}, \mathbb{R}, \mathbb{C}\}$. La familia*

$$\mathcal{F} = \left\{P_i(x) = x^i\right\}_{i\in\{0\}\cup\mathbb{N}}$$

del $\mathbb{K}$-espacio vectorial $\mathbb{K}[x]$ es linealmente independiente.

$\lhd$ Se debe probar que, toda combinación lineal formada por los elementos de cualquier subconjunto finito de la familia $\mathcal{F}$, debe tener todos los coeficientes nulos.

Se selecciona un subconjunto finito de elementos de $\mathcal{F}$. Sin pérdida de generalidad, se considerará que el subconjunto seleccionado contiene todos los elementos x^i para $i = 0, 1, 2, \ldots, n$, siendo $n \in \{0\} \cup \mathbb{N}$ el mayor subíndice tal que x^n está en dicha selección (de no ser así, pueden agregarse los x^i que falten, con coeficientes nulos), quedando así una expresión del tipo

$$a_0 + a_1 x + \cdots + a_n x^n = 0, \quad \forall x \in \mathbb{K}. \tag{8.11}$$

En este caso se está operando con la función polinomial $f(x) = a_0 + a_1 x + \cdots + a_n x^n$. Se eligen $n+1$ valores $x_1, x_2, \ldots, x_{n+1}$ de $\mathbb{K}$, distintos dos a dos[14], y se evalúa la identidad (8.11) en dichos valores. Se obtiene el sistema lineal homogéneo

$$\left\{ \begin{array}{ccccccccc} a_0 & + & a_1 & x_1 & + & \ldots & + & a_n & x_1^n & = & 0 \\ a_0 & + & a_1 & x_2 & + & \ldots & + & a_n & x_2^n & = & 0 \\ & & & & & \vdots & & & & \vdots & \vdots \\ a_0 & + & a_1 & x_{n+1} & + & \ldots & + & a_n & x_{n+1}^n & = & 0 \end{array} \right.$$

cuya matriz de coeficientes es la matriz de Vandermonde $V(x_1, x_2, \ldots, x_{n+1})$. Al haber elegido los escalares distintos dos a dos, su determinante es no nulo con lo cual el sistema lineal sólo admite la solución trivial $a_0 = a_1 = \cdots = a_n = 0$. Por lo tanto, $\mathcal{F}$ es linealmente independiente en $\mathbb{K}[x]$. $\triangleright$

Observación 8.9. Si en la demostración de este último ejemplo en lugar de tratar a $f(x) = a_0 + a_1 x + \cdots + a_n x^n$ como función polinomial (de variable x) en $\mathbb{K}[x]$ se la considera como un polinomio (en la indeterminada X) de $\mathbb{K}[X]$, se podría haber realizado simplemente la *identificación de coeficientes* obteniendo inmediatamente $a_0 = a_1 = \cdots = a_n = 0$. La prueba proporcionada en el ejemplo deja patente la importancia de la necesidad de trabajar en un cuerpo infinito. En efecto, la citada identificación de coeficientes en dos polinomios se puede realizar a partir de la definición de igualdad de *polinomios* introducida en el Anexo C y teniendo en cuenta el Teorema C.2 que proporciona la biyección (que respeta las operaciones) entre polinomios y funciones polinomiales en un cuerpo infinito.

[14]Observar que esta selección se puede realizar puesto que $\mathbb{K}$ tiene infinitos elementos.

Observación 8.10. Sea V un $\mathbb{K}$-espacio vectorial y $u \in V$. Se (propone como ejercicio comprobar que se) cumple que:

(a) Si $X \subseteq V$ es tal que $0 \in X$ entonces X es linealmente dependiente.

(b) $\{u\}$ es linealmente independiente si y sólo si $u \neq 0$.

Caracterización de conjuntos linealmente independientes

Proposición 8.6. Sea V un $\mathbb{K}$-espacio vectorial y sean $u_1, u_2, \ldots, u_n \in V$ con $n \geq 2$. Las siguientes condiciones son equivalentes:

(a) $\{u_1, u_2, \ldots, u_n\}$ es linealmente independiente.

(b) Ningún u_i, (para $i = 1, 2, \ldots, n$) es combinación lineal de los restantes.

Demostración. (a) $\Rightarrow$ (b) Si se supone, por reducción al absurdo, que algún u_i es combinación lineal de los restantes, se tiene que

$$u_i = \gamma_1 u_1 + \cdots + \gamma_{i-1} u_{i-1} + \gamma_{i+1} u_{i+1} + \cdots + \gamma_n u_n$$

de donde, al sumar $-u_i = (-1)u_i$ a ambos miembros y tras aplicar la propiedad conmutativa, se tiene

$$0 = \gamma_1 u_1 + \cdots + \gamma_{i-1} u_{i-1} + (-1)u_i + \gamma_{i+1} u_{i+1} + \cdots + \gamma_n u_n.$$

Este hecho contradice la independencia lineal pues el coeficiente de u_i es no nulo.

(b) $\Rightarrow$ (a) Suponer, por reducción al absurdo, que $\{u_1, u_2, \ldots, u_n\}$ no es linealmente independiente. Deben existir coeficientes $\gamma_1, \gamma_2, \ldots, \gamma_n \in \mathbb{K}$, no todos nulos, tales que

$$0 = \sum_{i=1}^{n} \gamma_i u_i.$$

Se supone que es $\gamma_i \neq 0$. Luego, existe $\gamma_i^{-1} \in \mathbb{K}$. Entonces de

$$0 = \gamma_1 u_1 + \cdots + \gamma_{i-1} u_{i-1} + \gamma_i u_i + \gamma_{i+1} u_{i+1} + \cdots + \gamma_n u_n$$

se tiene

$$u_i = -\gamma_i^{-1}\gamma_1 u_1 - \cdots - \gamma_i^{-1}\gamma_{i-1} u_{i-1} - \gamma_i^{-1}\gamma_{i+1} u_{i+1} - \cdots - \gamma_i^{-1}\gamma_n u_n,$$

con lo que se ha obtenido que u_i es combinación lineal de los restantes, en contra de la hipótesis. $\square$

Utilizando el método del contrarrecíproco, la Proposición 8.6 permite asegurar que las siguientes condiciones son equivalentes:

(a') $\{u_1, u_2, \ldots, u_n\}$ es linealmente dependiente.

(c') $\exists i_0 \in \{1, 2, \ldots, n\} : u_{i_0}$ es combinación lineal de los restantes.

Observación 8.11. Sea V un $\mathbb{K}$-espacio vectorial y $\{u_1, u_2, \ldots, u_n\}$ un conjunto linealmente independiente. Entonces (se propone como ejercicio probar que):

(a) Si un vector $u \in V$ se escribe como combinación lineal de $\{u_1, u_2, \ldots, u_n\}$ entonces los escalares de la combinación lineal son únicos.

(b) Si $X \subseteq \{u_1, u_2, \ldots, u_n\}$ entonces X es linealmente independiente.

(c) Si $u \notin \overline{\{u_1, u_2, \ldots, u_n\}}$ entonces $\{u_1, u_2, \ldots, u_n, u\}$ es linealmente independiente.

Además, cualquier subconjunto de V que contenga un subconjunto linealmente dependiente es linealmente dependiente.

El siguiente resultado relaciona la independencia lineal de un conjunto y las operaciones elementales (entre vectores).

Notación: Se abreviará "linealmente independiente" por LI y "linealmente dependiente" por LD.

Proposición 8.7. Sea V un $\mathbb{K}$-espacio vectorial y sea $\{u_1, \ldots, u_i, \ldots, u_j, \ldots, u_s\} \subseteq V$. Entonces

(a) $\{u_1, \ldots, u_i, \ldots, u_j, \ldots, u_s\}$ es LI $\Leftrightarrow$ $\{u_1, \ldots, u_j, \ldots, u_i, \ldots, u_s\}$ es LI.

(b) $\{u_1, \ldots, u_i, \ldots, u_s\}$ es LI $\Leftrightarrow$ $\{u_1, \ldots, ku_i, \ldots, u_s\}$ es LI, $\forall k \in \mathbb{K} - \{0\}$.

(c) $\{u_1, \ldots, u_i, \ldots, u_j, \ldots, u_s\}$ es LI $\Leftrightarrow$ $\{u_1, \ldots, u_i, \ldots, u_j + ku_i, \ldots, u_s\}$ es LI, $\forall k \in \mathbb{K}$.

Demostración. Se deja como ejercicio (comparar con la demostración del Teorema 3.1). $\square$

El resultado anterior asegura que la realización de operaciones elementales sobre vectores de un subconjunto dado en un espacio vectorial preservan tanto la dependencia como la independencia lineal de los mismos.

Ejemplo 8.48. *Sea* $S = \overline{\left\{ A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, B = \begin{bmatrix} 4 & 0 \\ -2 & 4 \end{bmatrix} \right\}}$. *Determinar si los vectores de* S *son linealmente independientes.*

$\triangleleft$ Se puede hacer por definición, o bien, observando que de la forma de los vectores dados (ambos tienen el mismo elemento en la diagonal y un 0 en la posición $(1, 2)$), se aprecia que $4A - B = 6C$ siendo $C := \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$. Luego, $S = \overline{\{A, B\}} = \overline{\{A, C\}}$. Como es inmediato ver que $\{A, C\}$ es LI, se tiene que $\{A, B\}$ es LI. $\triangleright$

El resultado anterior (Proposición 8.7) es especialmente útil cuando se quiere analizar si un subconjunto $\{u_1, u_2, \ldots, u_s\}$ de $\mathbb{K}^m$ es linealmente independiente o no[15].

[15] Más adelante, en la Proposición 9.3 de la página 474, se extenderá este procedimiento

Método para analizar si un subconjunto de $\mathbb{K}^n$ es linealmente independiente:

- Se colocan los vectores $\{u_1, u_2, \ldots, u_s\}$ en las columnas de una matriz $L \in \mathbb{K}^{n\times s}$.
- Se calcula la forma escalonada reducida por filas R_L.
- Los vectores $\{u_1, u_2, \ldots, u_s\}$ son LI si y sólo si los vectores columna de la matriz escalonada por filas R_L son LI.

Se observa que en el segundo paso anterior alcanza con realizar el método de Gauss hasta obtener una forma escalonada por filas aunque no sea reducida.)

Ejemplo 8.49. *Analizar la independencia lineal del subconjunto de $\mathbb{R}^4$ dado por*

$$\{u_1 = (1,0,3,0),\ u_2 = (1,0,1,1),\ u_3 = (1,0,0,3)\}.$$

◁ Si existen escalares $a, b, c \in \mathbb{R}$ tales que

$$(0,0,0,0) = a(1,0,3,0) + b(1,0,1,1) + c(1,0,0,3),$$

operando se obtiene el sistema de ecuaciones lineales

$$\left\{\begin{array}{rrrcl} a & +b & +c & = & 0 \\ 0a & +0b & +0c & = & 0 \\ 3a & +b & & = & 0 \\ & b & +3c & = & 0 \end{array}\right.,$$

a vectores de un espacio vectorial arbitrario, no necesariamente de $\mathbb{K}^m$.

que se reescribe matricialmente como

$$\begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 3 & 1 & 0 \\ 0 & 1 & 3 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

Llamando A a la matriz de coeficientes y x a la columna de las incógnitas, el sistema anterior $Ax = 0$ es equivalente al sistema $R_A x = 0$. Calculando R_A se obtiene

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

Puesto que las columnas de la matriz R_A son claramente linealmente independientes en $\mathbb{R}^4$, de la Observación 8.7 se concluye que el conjunto $\{u_1, u_2, u_3\}$ es linealmente independiente en $\mathbb{R}^4$. ▷

Observación 8.12. Se quiere analizar la independencia lineal del subconjunto de $\mathbb{R}^2$ dado por $\{u_1 = (1, 2),\ u_2 = (3, 6)\}$. Es claro que, aplicando el contrarrecíproco de la Proposición 8.7, el conjunto $\{u_1, u_2\}$ (formado por dos vectores distintos) es linealmente dependiente si y sólo si $\{u_1 = (1, 2), \frac{1}{3}u_2 = (1, 2)\}$ lo es. Se observa que, tras la aplicación de operaciones elementales se ha obtenido el mismo vector en ambos casos, con lo cual, en este segundo caso se debería hablar del *conjunto* $\{u_1 = (1, 2)\}$ y no del conjunto $\{u_1 = (1, 2), \frac{1}{3}u_2 = (1, 2)\}$ (por tener un vector repetido). Este hecho contribuye a que se hable de *sistemas de vectores*[a] cuando, a priori, no es posible saber si los elementos son todos distintos entre sí y, por tanto, si pueden formar parte de un conjunto. A pesar de ello, por abuso de lenguaje, muchas veces se habla de *conjunto de vectores*.

[a]Recuérdese lo comentado en la página 72 al respecto.

Si bien en algunos ejemplos sencillos es posible detectar, a simple vista, cuáles son las combinaciones lineales que se dan entre los vectores, el siguiente resultado proporciona un método de fácil aplicación para establecer, mediante la forma escalonada reducida por filas de una matriz, las relaciones de dependencia que existen entre las columnas de una matriz dada.

Proposición 8.8 (Correspondencia entre las columnas de matrices equivalentes por filas)**.** Sean $A, B \in \mathbb{K}^{m\times n}$ dos matrices particionadas según sus columnas como $A = \begin{bmatrix} a_1 & a_2 & \dots & a_n \end{bmatrix}$ y $B = \begin{bmatrix} b_1 & b_2 & \dots & b_n \end{bmatrix}$ tales que $A \sim_f B$. Entonces

$$a_i \in \overline{\{a_1, \dots, a_{i-1}, a_{i+1}, \dots, a_n\}} \quad \Leftrightarrow \quad b_i \in \overline{\{b_1, \dots, b_{i-1}, b_{i+1}, \dots, b_n\}}$$

y, además, se preservan los escalares $\gamma_j \in \mathbb{K}$ para $j = 1, \dots, i-1, i+1, \dots, n$ en las combinaciones lineales, en el sentido que

$$a_i = \sum_{\substack{j=1\\ j\neq i}}^{n} \gamma_j a_j \qquad \Leftrightarrow \qquad b_i = \sum_{\substack{j=1\\ j\neq i}}^{n} \gamma_j b_j.$$

Demostración. Como $A \sim_f B$, el Teorema 6.1 asegura que existe una matriz invertible $Q \in \mathbb{K}^{m\times m}$ tal que $QA = B$. Luego,

$$\begin{aligned} \begin{bmatrix} b_1 & b_2 & \dots & b_n \end{bmatrix} &= B \\ &= QA \\ &= Q\begin{bmatrix} a_1 & a_2 & \dots & a_n \end{bmatrix} \\ &= \begin{bmatrix} Qa_1 & Qa_2 & \dots & Qa_n \end{bmatrix}, \end{aligned}$$

de donde $b_i = Qa_i$ para todo $i = 1, 2, \dots, n$. Al ser Q invertible, es posible escribir $a_i = Q^{-1}b_i$ para todo $i = 1, 2, \dots, n$. Es claro que si

$$a_i = \sum_{\substack{j=1\\ j\neq i}}^{n} \gamma_j a_j,$$

multiplicando a izquierda ambos miembros por Q y distribuyendo se tiene

$$b_i = \sum_{\substack{j=1 \\ j \neq i}}^{n} \gamma_j b_j.$$

Y también si

$$b_i = \sum_{\substack{j=1 \\ j \neq i}}^{n} \gamma_j b_j,$$

multiplicando a izquierda ambos miembros por Q^{-1} se tiene

$$a_i = \sum_{\substack{j=1 \\ j \neq i}}^{n} \gamma_j a_j,$$

como se quería probar. □

Es decir, la equivalencia por filas preserva las relaciones de dependencia lineal entre las columnas de una matriz. Una consecuencia importante de la Proposición 8.7 y de la Proposición 8.8 es la siguiente.

Observación 8.13. Se consideran $A = \begin{bmatrix} a_1 & a_2 & \dots & a_n \end{bmatrix} \in \mathbb{K}^{m \times n}$ y $R_A = \begin{bmatrix} r_1 & r_2 & \dots & r_n \end{bmatrix}$ particionadas por columnas, siendo las columnas básicas de A las situadas en las posiciones $j_1 < j_2 < \ldots < j_r$. Se prueba fácilmente que las columnas de R_A correspondientes a las columnas básicas de A son LI. Teniendo en cuenta que siempre se cumple que $A \sim_f R_A$, es posible encontrar de manera inmediata las columnas LI de la matriz de partida (Proposición 8.7). Además, las mismas relaciones de linealidad que se encuentran entre las columnas de R_A (si las hay) también se dan entre las columnas de A (Proposición 8.8). Más aún, si en A una columna es no básica (es decir, la columna correspondiente en R_A no posee un uno principal), dicha columna será combinación lineal de las columnas básicas precedentes de A (preservando los escalares de la combinación lineal de las correspondientes columnas en R_A); es decir[a], si $n \geq 2$, para $i > 1$,

$$a_i \in \overline{\{a_1, \ldots, a_{i-1}\}} \quad \Leftrightarrow \quad b_i \in \overline{\{b_1, \ldots, b_{i-1}\}}$$

y

$$a_i = \sum_{j=1}^{i-1} \gamma_j a_j \qquad \Leftrightarrow \qquad b_i = \sum_{j=1}^{i-1} \gamma_j b_j.$$

Para demostrarlo, puesto que cada columna de R_A que no tenga un uno principal (si la hay) se escribe a partir de las precedentes que tienen un uno principal como

$$r_k := \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_s \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \gamma_1 \underbrace{\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_{r_{j_1}} + \gamma_2 \underbrace{\begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_{r_{j_2}} + \cdots + \gamma_s \underbrace{\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_{r_{j_s}},$$

entonces la columna no básica a_k en A también es combinación lineal de las columnas básicas precedentes a la s-ésima y preserva los escalares, es decir,

$$a_k = \gamma_1 a_{j_1} + \gamma_2 a_{j_2} + \cdots + \gamma_s a_{j_s}.$$

[a]Utilizando la notación de la Proposición 8.8.

Ejemplo 8.50. *Analizar si el subconjunto* $\{u_1, u_2, u_3, u_4\}$ *es linealmente independiente en* $\mathbb{Q}^4$ *siendo*

$$u_1 = (1, -1, 0, 1),\ u_2 = (3, -3, 0, 3),\ u_3 = (0, 1, 2, 1),\ u_4 = (4, -2, 4, 6).$$

Si fuese linealmente dependiente, encontrar las relaciones de linealidad entre dichos vectores.

$\triangleleft$ Se disponen los vectores dados como columnas de una matriz A y operando por filas se obtiene

$$A := \begin{bmatrix} 1 & 3 & 0 & 4 \\ -1 & -3 & 1 & -2 \\ 0 & 0 & 2 & 4 \\ 1 & 3 & 1 & 6 \end{bmatrix} \sim_f \begin{bmatrix} 1 & 3 & 0 & 4 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = R_A.$$

Si se denotan por b_1, b_2, b_3 y b_4 las columnas de R_A, se tiene que las columnas básicas son la primera b_1 y la tercera b_3 (que contienen los unos principales). Es inmediato comprobar que b_1 y b_3 son linealmente independientes en $\mathbb{Q}^4$. Además, de la forma de R_A se observa fácilmente que

$$b_2 = 3b_1 \qquad \text{y} \qquad b_4 = 4b_1 + 2b_3.$$

Entonces, la Proposición 8.8 establece que en la matriz A original, con columnas a_1, a_2, a_3 y a_4, se cumple que

$$a_2 = 3a_1 \qquad \text{y} \qquad a_4 = 4a_1 + 2a_3,$$

donde las combinaciones lineales preservan los escalares, hecho que no es inmediato a simple vista (para a_4).

Puesto que las columnas de la matriz R_A son LD, se puede afirmar que el conjunto de partida $\{u_1, u_2, u_3, u_4\}$ también es LD en $\mathbb{Q}^4$ y además, $u_2 = 3u_1$ y $u_4 = 4u_1 + 2u_3$. Por otra parte, $\{u_1, u_3\}$ es un conjunto linealmente independiente en $\mathbb{Q}^4$. ▷

8.8. Bases de un espacio vectorial

Definición 8.9. Sea V un $\mathbb{K}$-espacio vectorial. Un subconjunto $\mathcal{B}$ de vectores de V se llama **base** de V si se cumplen las dos condiciones siguientes:

(B1) $\mathcal{B}$ es un sistema de generadores de V,

(B2) $\mathcal{B}$ es un conjunto linealmente independiente de V.

Observación 8.14. Sea V un $\mathbb{K}$-espacio vectorial. Un subconjunto $\mathcal{B}$ de infinitos vectores de V se llama **base** de V si se cumplen las dos condiciones siguientes:

(B1) cualquier vector de V es combinación lineal de algún subconjunto finito de $\mathcal{B}$,

(B2) cualquier subconjunto finito (no vacío) de $\mathcal{B}$ linealmente independiente de V.

Ejemplo 8.51. *El conjunto $\mathcal{B} = \emptyset$ es una base de $V = \{0\}$.*

◁ Por definición, $\emptyset$ es linealmente independiente. Además, $\overline{\emptyset} = \{0\}$. Luego, el conjunto $\mathcal{B} = \emptyset$ es una base de $V = \{0\}$.
▷

Ejemplo 8.52. *El conjunto $\mathcal{C} = \{1\}$ es una base del $\mathbb{K}$-espacio vectorial $\mathbb{K}$. Se llama* **base canónica** *de $\mathbb{K}$.*

◁ Al ser $1 \neq 0$, $\{1\}$ es linealmente independiente en $\mathbb{K}$. Por otro lado, $\overline{\{1\}} = \{z \cdot 1 : z \in \mathbb{C}\} = \mathbb{C}$. Luego, el conjunto $\mathcal{C} = \{1\}$ es una base de $V = \mathbb{K}$.
▷

Ejemplo 8.53. *El conjunto*

$$\mathcal{C} = \{e_1 := (1, 0, \ldots, 0), e_2 := (0, 1, \ldots, 0), \ldots, e_n := (0, 0, \ldots, 1)\}$$

es una base de $\mathbb{K}^n$. Se llama **base canónica** *de $\mathbb{K}^n$.*

◁ Si se considera una combinación lineal del tipo

$$(0, 0, \ldots, 0) = a_1 e_1 + a_2 e_2 + \cdots + a_n e_n,$$

con $a_1, a_2, \ldots, a_n \in \mathbb{K}$, tras operar se obtiene $(0, 0, \ldots, 0) = (a_1, a_2, \ldots, a_n)$, de donde $a_1 = a_2 = \cdots = a_n$. Luego, $\mathcal{C}$ es linealmente independiente en $\mathbb{K}^n$.

Por otro lado, cualquier vector $(x_1, x_2, \dots, x_n) \in \mathbb{K}^n$ puede escribirse como

$$(x_1, x_2, \dots, x_n) = x_1 e_1 + x_2 e_2 + \cdots + x_n e_n \in \overline{\{x_1, x_2, \dots, x_n\}},$$

con lo que $\mathcal{C}$ es un sistema de generadores de $\mathbb{K}^n$.

En conclusión, $\mathcal{C}$ es una base de $\mathbb{K}^n$. $\triangleright$

Ejemplo 8.54. *El conjunto* $\mathcal{C} = \{E^{(i,j)}\}_{\substack{i=1,\dots,m \\ j=1,\dots,n}}$ *donde*

$$[E^{(i,j)}]_{rs} = \begin{cases} 1, & \text{si } r = i, s = j \\ 0, & \text{en otro caso} \end{cases} = [(\delta_{r,i}\delta_{s,j})_{rs}],$$

es una base de $\mathbb{K}^{m\times n}$*. Se llama* **base canónica** *de* $\mathbb{K}^{m\times n}$*.*

$\triangleleft$ Del Ejemplo 8.37 se tiene que $\mathcal{C}$ es un conjunto de generadores de $\mathbb{K}^{m\times n}$. La prueba de la independencia lineal es similar a la del Ejemplo 8.53 y se propone como ejercicio. $\triangleright$

Ejemplo 8.55. *El conjunto* $\mathcal{B} = \{1, i\}$ *es una base del* $\mathbb{R}$*-espacio vectorial* $\mathbb{C}$*.*

$\triangleleft$ Del Ejemplo 8.44 se tiene la independencia lineal. Se propone como ejercicio probar que $\mathcal{B} = \{1, i\}$ es un sistema de generadores del $\mathbb{R}$-espacio vectorial $\mathbb{C}$. $\triangleright$

Ejemplo 8.56. *Sea* $\mathbb{K} \in \{\mathbb{Q}, \mathbb{R}, \mathbb{C}\}$*. La familia*

$$\mathcal{C} = \left\{P_i(x) = x^i\right\}_{i\in\{0\}\cup\mathbb{N}}$$

es una base del $\mathbb{K}$*-espacio vectorial* $\mathbb{K}[x]$*. Se llama* **base canónica** *de* $\mathbb{K}[x]$*.*

$\triangleleft$ Es inmediato de los Ejemplos 8.36 y 8.47. $\triangleright$

Ejemplo 8.57. *El conjunto $\mathcal{B} = \{(2,3),(1,1)\}$ es una base del $\mathbb{R}$-espacio vectorial $\mathbb{R}^2$.*

$\lhd$ Si (x,y) es un vector arbitrario de $\mathbb{R}^2$ entonces resolviendo el sistema de ecuaciones lineales que se obtiene de

$$(x,y) = a(2,3) + b(1,1), \text{ con } a, b \in \mathbb{R},$$

se llega a $a = y - x$, $b = 3x - 2y$, con lo que

$$(x,y) = (y-x)(2,3) + (3x-2y)(1,1).$$

Por lo tanto, $\overline{\mathcal{B}} = \mathbb{R}^2$. Por otro lado, si

$$(0,0) = a(2,3) + b(1,1), \text{ con } a, b \in \mathbb{R}$$

se tiene que $a = b = 0$, con lo que $\mathcal{B}$ es un conjunto linealmente independiente. En definitiva, $\mathcal{B}$ es una base de $\mathbb{R}^2$. $\rhd$

Observar que se deben resolver dos sistemas de ecuaciones lineales con la misma matriz de coeficientes, uno de ellos homogéneo y el otro con un vector (términos independientes) genérico. En este caso, conviene estudiar primero qué espacio genera $\mathcal{B}$.

Ejemplo 8.58. *Los conjuntos*

$$\{u = (1,0), v = (0,1), w = (0,2)\} \qquad y \qquad \{u = (1,0)\}$$

no forman bases en el $\mathbb{R}$-espacio vectorial $\mathbb{R}^2$.

$\lhd$ Se tiene que $\{u = (1,0), v = (0,1), w = (0,2)\}$ es un sistema de generadores pero no es linealmente independiente en $\mathbb{R}^2$.

Por otro lado, el conjunto $\{u = (1,0)\}$ es linealmente independiente pero no es un sistema de generadores en $\mathbb{R}^2$. $\rhd$

Observación 8.15. Es interesante resaltar que si un espacio vectorial es finitamente generado, cualquier vector del espacio se escribe como combinación lineal utilizando un número finito de generadores. Aunque el espacio vectorial tenga infinitos elementos, como $\mathbb{R}^2$ por ejemplo, con tan solo 2 vectores y las operaciones del espacio quedarán completamente determinados todos los demás. Son, por lo tanto, de especial interés los espacios vectoriales finitamente generados. Sus bases tendrán, pues, un número finito de elementos.

Es importante responder las siguientes preguntas acerca de un $\mathbb{K}$-espacio vectorial finitamente generado V arbitrario:

- ¿Posee V una base?
- ¿Existe alguna relación entre la cantidad de elementos de dos bases distintas de V?
- ¿Existe alguna relación entre las bases de un subespacio $S \leq V$ y las bases de V?

Existencia de bases

Se verá más adelante que, en numerosas ocasiones, para un espacio vectorial dado, es necesario fijar una base con la que poder realizar operaciones. Para ello, es necesario asegurar previamente la existencia de una base para cada espacio vectorial. Si bien el resultado es válido en espacios vectoriales generales[16], se hará la demostración para los espacios finitamente generados.

Se ha indicado previamente que $\mathcal{B} = \emptyset$ es una base de $V = \{0\}$.

Teorema 8.2 (Existencia de bases)**.** Todo $\mathbb{K}$-espacio vectorial $V \neq \{0\}$ finitamente generado tiene una base (finita).

[16] Una demostración de este hecho puede encontrarse en [16, pág. 32]

Demostración. Sea $V \neq \{0\}$ un $\mathbb{K}$-espacio vectorial finitamente generado. Entonces, existe un sistema (finito) de generadores $X = \{u_1, u_2, \ldots, u_s\}$ de V, es decir $\overline{X} = V$. Es posible suponer que $0 \notin X$ (ya que si estuviese, se podría quitar y $\overline{X - \{0\}} = V$). Si X es linealmente independiente entonces es una base de V, y el teorema está demostrado.

Si no, por el comentario posterior a la Proposición 8.6, existe un vector $u_i \in X$ que es combinación lineal de los restantes. Del Lema 8.1,

$$V = \overline{\{u_1, \ldots, u_{i-1}, u_i, u_{i+1}, \ldots, u_s\}} = \overline{\{u_1, \ldots, u_{i-1}, u_{i+1}, \ldots, u_s\}}.$$

Si el conjunto $X_1 = \{u_1, \ldots, u_{i-1}, u_{i+1}, \ldots, u_s\}$ es linealmente independiente, forma una base de V. Si no, de nuevo, se quita uno de sus vectores que es combinación lineal de los restantes, y se analiza si el conjunto resultante es o no linealmente independiente. Se repite el proceso un número finito de veces hasta eliminar todos los vectores que son combinación lineal de los demás y se obtiene así una base de V. Observar que el conjunto más pequeño al que se puede llegar eliminando los vectores que dependen linealmente de los demás es de la forma $\{u\}$ con $u \neq 0$, que es linealmente independiente por la Observación 8.10 (b); de no ser así se habría llegado a $V = \{0\}$, que sería una contradicción. □

El siguiente resultado permite encontrar una base mediante la reducción de sistemas de generadores.

Teorema 8.3 (Teorema de reducción de sistemas generadores a bases)**.** De todo sistema de generadores de un $\mathbb{K}$-espacio vectorial $V \neq \{0\}$ finitamente generado es posible extraer una base (finita).

Demostración. Se ha realizado en la demostración del Teorema 8.2. □

Ejemplo 8.59. *En el $\mathbb{R}$-espacio vectorial $\mathbb{R}^3$, extraer una base a partir del siguiente sistema de generadores*

$$G = \{u = (1,1,0), v = (1,1,1), w = (2,2,1), x = (0,1,1)\}.$$

$\lhd$ Para comprobar que, efectivamente, $\{u, v, w, x\}$ es un sistema de generadores de $\mathbb{R}^3$, se parte de un vector genérico (α, β, γ) de $\mathbb{R}^3$ y realizando operaciones elementales se obtiene

$$\left[\begin{array}{cccc|c} 1 & 1 & 2 & 0 & \alpha \\ 1 & 1 & 2 & 1 & \beta \\ 0 & 1 & 1 & 1 & \gamma \end{array}\right] \sim_f \left[\begin{array}{cccc|c} 1 & 0 & 1 & 0 & \beta-\gamma \\ 0 & 1 & 1 & 0 & \gamma-\beta+\alpha \\ 0 & 0 & 0 & 1 & \beta-\alpha \end{array}\right].$$

Se propone como ejercicio completar los detalles e interpretar el resultado.

Para extraer una base del sistema de generadores G se utiliza la Observación 8.13 de la Proposición 8.8. En efecto, disponiendo por columnas en una matriz A los vectores dados y efectuando operaciones elementales por filas hasta obtener su forma escalonada reducida por filas se tiene

$$A = \begin{bmatrix} 1 & 1 & 2 & 0 \\ 1 & 1 & 2 & 1 \\ 0 & 1 & 1 & 1 \end{bmatrix} \sim_f \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

de donde resulta que la primera, la segunda y la cuarta columnas (de R_A, y por tanto de A) son LI (y generan) y la tercera es suma de las dos primeras.

Luego, una base de $\mathbb{R}^3$ obtenida de vectores de S es $\mathcal{B} = \{u = (1,1,0), v = (1,1,1), x = (0,1,1)\}$. $\rhd$

El método descrito en el ejemplo anterior es general y se sistematiza en el siguiente algoritmo:

Método para extraer una base de un subespacio vectorial S de $\mathbb{K}^n$ a partir de un sistema generador $X = \{u_1, u_2, \ldots, u_s\}$ de S:

- Se disponen los vectores de X por columnas en una matriz G.
- Se calcula la forma escalonada reducida por filas R_G de G.
- Las columnas de G correspondientes a las columnas básicas de R_G forman una base de S.

Equicardinalidad de las bases

Cuando un conjunto A es finito, se llama **cardinal** de A a la cantidad de elementos de dicho conjunto. La propiedad de dos conjuntos finitos de poseer la misma cantidad de elementos se llama **equicardinalidad** de los conjuntos.

El siguiente resultado es previo al teorema de equicardinalidad (o también llamado de coordinabilidad) de las bases y permitirá relacionar la cantidad de elementos de una base de un espacio vectorial (finitamente generado) con la cantidad de vectores de un subconjunto linealmente independiente de dicho espacio.

Lema 8.2. Sea $\mathcal{B} = \{u_1, u_2, \ldots, u_n\}$ una base de un $\mathbb{K}$-espacio vectorial V y sea $X = \{v_1, v_2, \ldots, v_{n+1}\} \subseteq V$. Entonces X es linealmente dependiente en V.

Demostración. Para demostrar que X es linealmente dependiente en V, se debe probar que existe alguna $(n+1)$-tupla $(\alpha_1, \alpha_2, \ldots, \alpha_{n+1})$ no trivial, con $\alpha_1, \alpha_2, \ldots, \alpha_{n+1} \in \mathbb{K}$, que soluciona

$$0_V = \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_{n+1} v_{n+1}. \tag{8.12}$$

Por ser $\mathcal{B}$ una base, es un sistema generador de V. Este hecho asegura que todo vector de V se puede escribir como combinación lineal de los elementos de $\mathcal{B}$; en particular, esto se puede hacer para $v_1, v_2, \ldots, v_{n+1}$. Es decir,

$$\begin{aligned} v_1 &= \beta_{11}u_1 + \beta_{21}u_2 + \cdots + \beta_{n1}u_n \\ v_2 &= \beta_{12}u_1 + \beta_{22}u_2 + \cdots + \beta_{n2}u_n \\ &\vdots \\ v_{n+1} &= \beta_{1,n+1}u_1 + \beta_{2,n+1}u_2 + \cdots + \beta_{n,n+1}u_n, \end{aligned}$$

para ciertos escalares $\beta_{ij} \in \mathbb{K}$, con $i = 1, \ldots, n$, $j = 1, \ldots, n+1$.

Sustituyendo en (8.12) y operando se tiene que

$$\begin{aligned} 0_V &= \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_{n+1} v_{n+1} \\ &= \alpha_1(\beta_{11}u_1 + \beta_{21}u_2 + \cdots + \beta_{n1}u_n) + \alpha_2(\beta_{12}u_1 + \beta_{22}u_2 + \cdots + \beta_{n2}u_n) + \\ &\quad + \cdots + \alpha_{n+1}(\beta_{1,n+1}u_1 + \beta_{2,n+1}u_2 + \cdots + \beta_{n,n+1}u_n) \\ &= (\alpha_1\beta_{11} + \alpha_2\beta_{12} + \cdots + \alpha_{n+1}\beta_{1,n+1})u_1 + \\ &\quad +(\alpha_1\beta_{21} + \alpha_2\beta_{22} + \cdots + \alpha_{n+1}\beta_{2,n+1})u_2 + \cdots + \\ &\quad +(\alpha_1\beta_{n1} + \alpha_2\beta_{n2} + \cdots + \alpha_{n+1}\beta_{n,n+1})u_n. \end{aligned}$$

Debido a la independencia lineal de $\{u_1, u_2, \ldots, u_n\}$, se debe cumplir que

$$\left\{ \begin{array}{rcl} \alpha_1\beta_{11} + \alpha_2\beta_{12} + \cdots + \alpha_{n+1}\beta_{1,n+1} &=& 0 \\ \alpha_1\beta_{21} + \alpha_2\beta_{22} + \cdots + \alpha_{n+1}\beta_{2,n+1} &=& 0 \\ \vdots & \vdots & \vdots \\ \alpha_1\beta_{n1} + \alpha_2\beta_{n2} + \cdots + \alpha_{n+1}\beta_{n,n+1} &=& 0 \end{array} \right. .$$

Al tener un sistema lineal homogéneo con n ecuaciones en las incógnitas $\alpha_1, \ldots, \alpha_{n+1}$, es decir, con más incógnitas que ecuaciones (pues $n+1 > n$), por el corolario del Teorema de Rouché-Frobenius, el sistema debe admitir al menos una solución no trivial $(c_1, c_2, \ldots, c_{n+1})$, que es la buscada para que el conjunto X sea linealmente dependiente. Como ejercicio se propone sustituir la $(n+1)$-tupla no trivial $(c_1, c_2, \ldots, c_{n+1})$ en $\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_{n+1} v_{n+1}$ y comprobar que se obtiene 0_V. $\square$

En el siguiente corolario se establece una versión más clara del Lema 8.2 que permitirá obtener de forma directa el Teorema de equicardinalidad de las bases.

Corolario 8.1. Sea V un $\mathbb{K}$-espacio vectorial. Si $\mathcal{B} = \{u_1, \ldots, u_n\}$ es una base de V y X es un subconjunto linealmente independiente de V entonces X tiene a lo sumo n vectores.

Demostración. Sigue de aplicar el contrarrecíproco al Lema 8.2. □

Teorema 8.4 (Teorema de equicardinalidad de las bases)**.** Sea V un $\mathbb{K}$-espacio vectorial finitamente generado. Dos bases cualesquiera de V tienen el mismo número de vectores.

Demostración. Por ser V un espacio finitamente generado admite, al menos, una base con un número finito de vectores.

Sean $\mathcal{B}_1 = \{u_1, u_2, \ldots, u_s\}$ y $\mathcal{B}_2 = \{v_1, v_2, \ldots, v_m\}$ dos bases de V.

- Por ser $\{u_1, u_2, \ldots, u_s\}$ un conjunto linealmente independiente y el conjunto $\{v_1, v_2, \ldots, v_m\}$ una base de V, del Corolario 8.1 se tiene que $s \leq m$.

- Por ser $\{v_1, v_2, \ldots, v_m\}$ un conjunto linealmente independiente y el conjunto $\{u_1, u_2, \ldots, u_s\}$ una base de V, del Corolario 8.1 se sigue $m \leq s$.

Por lo tanto, $s = m$. □

En el Anexo D se proporciona un resultado más general que el obtenido en el Lema 8.2 (con una prueba muy similar) suponiendo como hipótesis que $\mathcal{B}$ es sólo un conjunto linealmente independiente (y no una base). Además, se prueba el **Teorema de Steiniz**, también conocido como **Teorema del canje o del intercambio**.

8.9. Dimensión

El hecho fundamental que se acaba de demostrar en el Teorema de la equicardinalidad de las bases permite asignar a cada espacio vectorial finitamente generado $V \neq \{0\}$ un número natural, a saber, el número de vectores de cualesquiera de sus bases. Esto es posible realizarlo puesto que el número de vectores de una base no es una propiedad de dicha base, sino intrínseca del espacio vectorial.

Definición 8.10. Sea V un $\mathbb{K}$-espacio vectorial. Si $V \neq \{0\}$ es finitamente generado, se llama **dimensión** de V sobre $\mathbb{K}$, y se denota $\dim_{\mathbb{K}}(V)$, al número de vectores de una (cualquiera) de sus bases, y se dice que V es de **dimensión finita**[a] sobre $\mathbb{K}$. Si $V = \{0\}$, a su **dimensión** se le asigna el número entero 0. En caso contrario, se dice que V es de **dimensión infinita** sobre $\mathbb{K}$. En símbolos,

$$\dim_{\mathbb{K}}(V) = \begin{cases} 0, & \text{si } V = \{0\}, \\ n, & \text{si } \exists n \in \mathbb{N} \text{ tal que } V \text{ admite una base con } n \text{ vectores}, \\ \infty, & \text{en caso contrario.} \end{cases}$$

[a] A veces se expresa como **finito-dimensional**.

De ahora en más, finitamente generado y de dimensión finita se utilizarán como sinónimos.

Ejemplo 8.60. *A partir de las bases halladas para los espacios vectoriales de los ejemplos anteriores se tiene:*

- $\dim_{\mathbb{K}}(\mathbb{K}) = 1$.
- $\dim_{\mathbb{K}}(\mathbb{K}^n) = n$.
- $\dim_{\mathbb{K}}(\mathbb{K}^{m\times n}) = mn$.

- $\dim_{\mathbb{C}}(\mathbb{C}) = 1$ *y* $\dim_{\mathbb{R}}(\mathbb{C}) = 2$.

Ejemplo 8.61. *Si* $\mathbb{K} \in \{\mathbb{Q}, \mathbb{R}, \mathbb{C}\}$ *entonces*

- $\dim_{\mathbb{K}}(\mathbb{K}_n[x]) = n + 1$.
- $\dim_{\mathbb{K}}(\mathbb{K}[x]) = \infty$.

El siguiente resultado, consecuencia del Lema 8.2, relaciona tanto la cantidad de elementos de un subconjunto linealmente independiente como la cantidad de elementos de un sistema generador con la dimensión del espacio al que pertenecen.

Corolario 8.2. Sea V un $\mathbb{K}$-espacio vectorial de dimensión (finita) $\dim_{\mathbb{K}}(V) = n \geq 1$. Entonces:

(a) Si $\{u_1, u_2, \ldots, u_s\}$ es un subconjunto linealmente independiente de V entonces $s \leq n$.

(b) Si $\{u_1, u_2, \ldots, u_s\}$ es un sistema de generadores de V entonces $s \geq n$.

Demostración. Sea $\mathcal{B}$ una base de V. Como $\dim_{\mathbb{K}}(V) = n$, $\mathcal{B}$ debe estar formada por n vectores.

(a) Si $\{u_1, u_2, \ldots, u_s\}$ es un subconjunto linealmente independiente de V, al ser $\mathcal{B}$ una base de V, por el Corolario 8.1, $s \leq n$.

(b) Si $\{u_1, u_2, \ldots, u_s\}$ es un sistema de generadores de V, por el Teorema de reducción de sistemas generadores a bases (Teorema 8.3), existe una base $\mathcal{B}'$ de V contenida en $\{u_1, u_2, \ldots, u_s\}$, es decir, la base $\mathcal{B}'$ tiene r vectores con $r \leq s$. Al ser $\mathcal{B}$ una base de V, es un conjunto linealmente independiente con n vectores y, por el Corolario 8.1, $n \leq r$. Luego, $n \leq s$. $\square$

El siguiente es un lema técnico que se necesitará para establecer los resultados posteriores, en especial el que se conoce como *Teorema de ampliación de conjuntos linealmente independientes a una base*.

Lema 8.3. Sean V un $\mathbb{K}$-espacio vectorial, X un subconjunto linealmente independiente de V tal que $\overline{X} \neq V$ y sea $u \in V - \overline{X}$. Entonces $X \cup \{u\}$ es un subconjunto linealmente independiente de V.

Demostración. Sea $u \in V - \overline{X} \neq \emptyset$ y X un subconjunto linealmente independiente de V.

Si $\boxed{X = \emptyset}$ entonces $X \cup \{u\} = \{u\}$ que es linealmente independiente pues $u \neq 0$ (ya que u no está en el subespacio $\overline{X} = \{0\}$).

Si $\boxed{X \neq \emptyset}$, se debe probar que $X \cup \{u\}$ es un subconjunto linealmente independiente de V. Para ello, se considera el subconjunto finito

$$\{u_1, u_2, \ldots, u_s\} \subseteq X.$$

Es conocido que $\{u_1, u_2, \ldots, u_s\}$ es linealmente independiente; por lo tanto, se debe probar que $\{u_1, u_2, \ldots, u_s, u\}$ es un subconjunto linealmente independiente de V. En efecto, en la combinación lineal

$$0 = \alpha_1 u_1 + \alpha_2 u_2 + \cdots + \alpha_s u_s + \alpha_{s+1} u,$$

con $\alpha_1, \alpha_2, \ldots, \alpha_s, \alpha_{s+1} \in \mathbb{K}$ puede ocurrir que:

- $\alpha_{s+1} \neq 0$. En este caso, como $\alpha_{s+1}^{-1} \in \mathbb{K}$,

  $$u = -\alpha_{s+1}^{-1}\alpha_1 u_1 - \alpha_{s+1}^{-1}\alpha_2 u_2 - \cdots - \alpha_{s+1}^{-1}\alpha_s u_s \in \overline{X},$$

  que es una contradicción. Por lo tanto, esta posibilidad no puede ocurrir.

- $\alpha_{s+1} = 0$. En este caso, $0 = \alpha_1 u_1 + \alpha_2 u_2 + \cdots + \alpha_s u_s$, y al ser $\{u_1, u_2, \ldots, u_s\}$ un conjunto linealmente independiente, $\alpha_1 = \alpha_2 = \cdots = \alpha_s = 0$. Por lo tanto, $X \cup \{u\}$ es linealmente independiente.

$\square$

Se observa que la demostración del caso $X \neq \emptyset$ en el Lema anterior es la esencia de la demostración del apartado (c) de la Observación 8.11.

Teorema 8.5 (Teorema de ampliación de un conjunto linealmente independiente a una base)**.** Sea V un $\mathbb{K}$-espacio vectorial de dimensión (finita) $\dim_{\mathbb{K}}(V) = n \geq 1$ y sea $X = \{u_1, u_2, \ldots, u_s\}$ un conjunto con $s \geq 1$ vectores linealmente independientes de V. Existen $n-s$ vectores $u_{s+1}, \ldots, u_n \in V$ tales que $X \cup \{u_{s+1}, \ldots, u_n\}$ es una base de V.

Demostración. Sea $X = \{u_1, u_2, \ldots, u_s\}$ un conjunto linealmente independiente de V. Se consideran dos situaciones: $s = n$ y $s < n$.

Si $\boxed{s = n}$, debe ocurrir que

- $\overline{X} = V$.

En efecto, si por reducción al absurdo $\overline{X} \neq V$, existe un vector $u_{n+1} \in V$ tal que $u_{n+1} \notin \overline{X}$. Por el Lema 8.3, $X \cup \{u_{n+1}\}$ es un conjunto linealmente independiente en V con $n+1$ vectores, lo que contradice $\dim_{\mathbb{K}}(V) = n$. Se tiene que el resultado es válido con $n - s = 0$.

Sea $\boxed{s < n}$. Primero se debe observar que:

- $\overline{X} \neq V$.

En efecto, si fuese $\overline{X} = V$, como X es linealmente independiente en V, se tendría que X es una base de V con s vectores. Por hipótesis, $n = \dim_{\mathbb{K}}(V) = s < n$, que es una contradicción. Debe ocurrir entonces que $\overline{X} \subsetneq V$. En consecuencia, existe $u_{s+1} \in V$ tal que $u_{s+1} \notin \overline{X}$. Por el Lema 8.3, $X \cup \{u_{s+1}\}$ es linealmente independiente en V.

Si $\overline{X \cup \{u_{s+1}\}} = V$, se ha encontrado un conjunto de generadores con $s+1$ vectores linealmente independientes de V y, por tanto, es una base de V (en este caso es $s+1 = n$, es decir $n - s = 1$).

Si $\overline{X \cup \{u_{s+1}\}} \neq V$, se repite el razonamiento un número finito de veces hasta encontrar una base con $n - s$ vectores que contenga a X. $\square$

Observación 8.16. Si en la hipótesis del resultado anterior se tuviese que $X = \emptyset$, valdría la misma demostración que para el caso $s < n$, o bien, simplemente se podría añadir una base cualquiera y se obtendría el resultado.

En el caso concreto del espaacio vectorial $\mathbb{K}^m$, es posible realizar una ampliación de un subconjunto linealmente independiente de $\mathbb{K}^m$ a una base de $\mathbb{K}^m$ ejecutando los pasos del siguiente algoritmo:

Método de ampliación de un subconjunto LI de $\mathbb{K}^m$ a una base de $\mathbb{K}^n$:

- Se concatenan los vectores $\{e_1, e_2, \ldots, e_n\}$ de la base canónica de $\mathbb{K}^n$ a continuación de los vectores LI del conjunto $\{u_1, u_2, \ldots, u_s\}$ de $\mathbb{K}^m$ en las columnas de una matriz $B \in \mathbb{K}^{m\times(s+n)}$:

$$B = \left[\begin{array}{cccc|cccc} u_1 & u_2 & \ldots & u_s & e_1 & e_2 & \ldots & e_n \end{array}\right].$$

- Se calcula la forma escalonada reducida por filas R_B.

- Las columnas de B correspondientes a las columnas básicas de R_B determinan una base de $\mathbb{K}^n$ que contiene a $\{u_1, u_2, \ldots, u_s\}$.

Para garantizar que este método funciona no hay más que observar que las primeras s columnas de R_B contendrán pivotes (pues las operaciones elementales preservan la independencia lineal) y los restantes $n-s$ pivotes se encontrarán en las columnas posteriores a la s-ésima. Al unir las columnas de B correspondientes las columnas básicas de R_B se obtendrá una base de $\mathbb{K}^n$ (lo que sige de observar su forma escalonada reducida por filas) y contiene a $\{u_1, u_2, \ldots, u_s\}$ como se quería.

Suponiendo que se conoce la dimensión (finita) de un espacio vectorial, el resultado que sigue es útil en la práctica.

Corolario 8.3. Sea V un $\mathbb{K}$-espacio vectorial finito-dimensional con $\dim_{\mathbb{K}}(V) = n \geq 1$ y sea $X = \{u_1, u_2, \ldots, u_n\}$ un conjunto con n vectores de V. Las siguientes condiciones son equivalentes:

(a) X es una base de V.

(b) X es un conjunto linealmente independiente de V.

(c) X es un sistema de generadores de V.

Demostración. Sea $X = \{u_1, u_2, \ldots, u_n\} \subseteq V$.

(a) $\Rightarrow$ (b) Es evidente por definición.

(b) $\Rightarrow$ (c) Sea X un conjunto linealmente independiente de V. Por el Teorema de ampliación de conjuntos linealmente independientes a una base (caso $s = n$ de su demostración), se tiene que $\overline{X} = V$.

(c) $\Rightarrow$ (a) Como X es un sistema de generadores de V, del Teorema 8.3 se puede concluir que existe una base $\mathcal{B}$ de V tal que $\mathcal{B} \subseteq X$. De la hipótesis $\dim_{\mathbb{K}}(V) = n$ y por la equicardinalidad de las bases se tiene que dicha base $\mathcal{B}$ debe tener n vectores, la misma cantidad de elementos que X, lo que implica que $\mathcal{B} = X$. Luego, $X (= \mathcal{B})$ es una base de V. $\square$

Sea S un subconjunto no vacío del $\mathbb{K}$-espacio vectorial $V \neq \{0\}$. Se dice que:

- S es un **conjunto linealmente independiente maximal** de V si
  (a) S es un subconjunto linealmente independiente de V y
  (b) para todo $v \in V$ tal que $v \notin S$, se tiene que $S \cup \{v\}$ no es linealmente independiente.
- S es un **sistema de generadores minimal** de V si
  (a) S es un sistema de generadores de V y
  (b) para todo $v \in S$, se tiene que $S - \{v\}$ no es un sistema de generadores de V.

Utilizando la noción de dimensión en un espacio vectorial finitamente generado V, a partir del Corolario 8.2 y del Corolario 8.3, es posible interpretar el concepto de base de una de las dos formas siguientes:

- como un conjunto de vectores linealmente independiente maximal de V, o también
- como un sistema de generadores minimal de V.

Ejemplo 8.62. *En el $\mathbb{R}$-espacio vectorial $\mathbb{R}^4$, ampliar el conjunto linealmente independiente*

$$\mathcal{C} = \{u = (1,1,0,0), v = (1,1,1,1)\}$$

a una base de $\mathbb{R}^4$.

$\lhd$ Es inmediato comprobar que $\mathcal{C}$ es LI en $\mathbb{R}^4$. Para ello, basta disponer por columnas en una matriz A los vectores dados y efectuar operaciones elementales por filas hasta obtener su forma escalonada reducida por filas

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \sim_f \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} = R_A,$$

donde se observa que los vectores obtenidos tienen los unos principales en las posiciones $(1,1)$ y $(2,2)$ de la matriz, con lo que son LI. Completando el conjunto $\mathcal{C}$ con, por ejemplo, los vectores canónicos $e_3 = (0,0,1,0)$ y $e_3 = (0,0,0,1)$ de $\mathbb{R}^4$, se tiene que el conjunto $\{u, v, e_3, e_4\}$ es una base de $\mathbb{R}^4$.

Otra forma de resolver este ejemplo es mediante el método indicado en la página 417. En efecto, se construye la matriz que concatena los vectores dados y los de la base canónica de $\mathbb{R}^4$ y se calcula su formas escalonada reducida por filas:

$$A = \left[\begin{array}{cc|cccc} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \end{array}\right] \sim_f \left[\begin{array}{cc|cccc} 1 & 0 & 0 & 1 & 0 & -1 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{array}\right] = R_A.$$

La base pedida de viene dada por los vectores de A correspondientes a los pivotes, es decir, $\{u, v, (1,0,0,0), (0,0,1,0)\}$ es una base de $\mathbb{R}^4$ que contiene a $\{u, v\}$. ▷

Dimensión de un subespacio

Si V es un $\mathbb{K}$-espacio vectorial de dimensión finita, al ser $U \leq V$, un espacio vectorial por sí mismo, es posible considerar la dimensión de U y compararla con la de V.

Ejemplo 8.63. *Determinar la dimensión del $\mathbb{C}$-subespacio vectorial*

$$S = \{(x, y, z, t) \in \mathbb{C}^4 : x = (1+i)y + (1-i)t,\ z = it\}.$$

¿Cuánto vale $\dim_{\mathbb{R}}(S)$*?*

◁ Puesto que un elemento genérico de S tiene la forma

$$\begin{aligned}(x, y, z, t) &= ((1+i)y + (1-i)t, y, it, t) \\ &= y(1+i, 1, 0, 0) + t(1-i, 0, i, 1)\end{aligned}$$

con $y, t \in \mathbb{C}$, se tiene que $\{(1+i, 1, 0, 0), (1-i, 0, i, 1)\}$ es un sistema de generadores de S (se obtienen para $y = 1$, $t = 0$ y luego $y = 0$, $t = 1$). Por ser $\{(1+i, 1, 0, 0), (1-i, 0, i, 1)\}$ un subconjunto linealmente independiente de S en el $\mathbb{C}$-espacio vectorial $\mathbb{C}^4$, es una base de S. Luego,

$$\dim_{\mathbb{C}}(S) = 2 \leq 4 = \dim_{\mathbb{C}}(\mathbb{C}^4).$$

Ahora bien, al escribir $y = y_1 + iy_2$, $t = t_1 + it_2$ con $y_1, y_2, t_1, t_2 \in \mathbb{R}$ y operar con números complejos se obtiene

$$\begin{aligned}(x, y, z, t) &= (y_1 + iy_2)(1 + i, 1, 0, 0) + (t_1 + it_2)(1 - i, 0, i, 1) \\ &= y_1(1 + i, 1, 0, 0) + y_2(-1 + i, i, 0, 0) + t_1(1 - i, 0, i, 1) + \\ &\quad +t_2(1 + i, 0, -1, i).\end{aligned}$$

Es decir, se ha escrito un vector genérico de S como combinación lineal de vectores de $\mathbb{C}^4$ con *escalares reales*. Es posible garantizar que

$$\dim_{\mathbb{R}}(S) = 4 \leq 8 = \dim_{\mathbb{R}}(\mathbb{C}^4)$$

pues el sistema generador de S

$$\{(1 + i, 1, 0, 0), (-1 + i, i, 0, 0), (1 - i, 0, i, 1), (1 + i, 0, -1, i)\}$$

es un subconjunto linealmente independiente[17] (y, por tanto, una base) de S en el $\mathbb{R}$-espacio vectorial $\mathbb{C}^4$. ▷

Proposición 8.9. Sea V un $\mathbb{K}$-espacio vectorial finito-dimensional con $\dim_{\mathbb{K}}(V) = n \geq 1$ y sea $U \leq V$. Entonces:

(a) U es finito-dimensional y $\dim_{\mathbb{K}}(U) \leq n$.

(b) Si $\dim_{\mathbb{K}}(U) = n$ entonces $U = V$.

(c) Si U es subespacio propio de V entonces $1 \leq \dim_{\mathbb{K}}(U) < n$.

Demostración. Por ser el espacio vectorial V finitamente generado, es claro que el subespacio U de V también lo es.

Si $U = \{0\}$ entonces $\dim_{\mathbb{K}}(U) = 0 \leq n$ y el resultado se cumple.

Sea $U \neq \{0\}$. Luego, U tiene al menos un elemento linealmente independiente, por tanto, $\dim_{\mathbb{K}}(U) \geq 1$. Como $\dim_{\mathbb{K}}(V) = n$ y $U \subseteq V$, resulta por

[17]Se propone comprobarlo como ejercicio recordando que $y_1, y_2, t_1, t_2 \in \mathbb{R}$.

el Corolario 8.2 que U puede tener a lo sumo n vectores linealmente independientes. Por lo tanto, U tiene una base finita que contiene no más de n vectores, es decir, $\dim_{\mathbb{K}}(U) \leq n$.

Si $\dim_{\mathbb{K}}(U) = n$ entonces una base cualquiera $\mathcal{B}$ de U tiene n vectores. Al ser $\mathcal{B} \subseteq U \leq V$ (un subconjunto linealmente independiente) y $\dim_{\mathbb{K}}(V) = n$, de la implicación (b) $\Rightarrow$ (a) del Corolario 8.3, el conjunto $\mathcal{B}$ debe ser una base de V. Luego, $U = \overline{\mathcal{B}} = V$. Además, si $U \neq V$, se tiene que $\dim_{\mathbb{K}}(U) < n$. □

Observación 8.17. El apartado (b) de la Proposición 8.9 no es cierto si los espacios vectoriales son de dimensión infinita. En efecto, una base del $\mathbb{R}$-espacio vectorial $\mathbb{R}[x]$ es $\{x^i\}_{i\in\{0\}\cup\mathbb{N}}$ y una base del subespacio

$$U := \{p(x) \in \mathbb{R}[x] : p(0) = 0\}$$

es $\{x^i\}_{i\in\mathbb{N}}$. Si bien es posible definir una función biyectiva entre ambas bases (infinitas), lo que asegura que son conjuntos coordinables, se tiene que $U \neq \mathbb{R}[x]$.

Ejemplo 8.64. *En el $\mathbb{R}$-espacio vectorial $\mathbb{R}^3$, extraer una base del subespacio*

$$S = \overline{\{u = (1,1,0), v = (1,1,1), w = (2,2,1), x = (0,0,1)\}}$$

y hallar su dimensión.

◁ Se tiene que $\{u, v, w, x\}$ es un sistema de generadores de S, pero al tener 4 elementos no es linealmente independiente en $\mathbb{R}^3$ pues $\dim_{\mathbb{R}}(\mathbb{R}^3) = 3 < 4$.

Para extraer una base de dicho subespacio se utiliza la Observación 8.13 de la Proposición 8.8. En efecto, disponiendo por columnas en una matriz A los vectores dados y efectuando operaciones elementales por filas hasta obtener su forma escalonada reducida por filas se tiene

$$A = \begin{bmatrix} 1 & 1 & 2 & 0 \\ 1 & 1 & 2 & 0 \\ 0 & 1 & 1 & 1 \end{bmatrix} \sim_f \begin{bmatrix} 1 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \end{bmatrix} \sim_f \begin{bmatrix} 1 & 0 & 1 & -1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} = R_A,$$

de donde resulta que las dos primeras columnas (de R_A, y por tanto de A) son LI, la tercera es suma de las dos primeras y la cuarta es la segunda menos la primera.

Luego, una base de S es $\mathcal{B} = \{u = (1,1,0), v = (1,1,1)\}$. En este caso, la dimensión del subespacio S es $\dim_{\mathbb{R}}(S) = 2 \leq 3 = \dim_{\mathbb{R}}(\mathbb{R}^3)$. ▷

Observación 8.18. Para resolver el ejemplo anterior sería posible disponer los vectores dados por filas en una matriz B (en lugar de hacerlo por columnas) y efectuar operaciones elementales por filas hasta obtener su forma escalonada reducida por filas. Dado que las operaciones elementales por filas no alteran la independencia/dependencia lineal (véase la Proposición 8.7), se obtiene

$$B = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 2 & 2 & 1 \\ 0 & 0 & 1 \end{bmatrix} \sim_f \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \sim_f \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = R_A,$$

de donde las dos primeras filas de R_A y, por lo tanto de A, corresponden a vectores LI. La diferencia es que ahora no es posible reconocer cuáles son las relaciones de dependencia lineal con que se relacionan los vectores. Además, es importante notar que no se han realizado intercambios de filas al efectuar las operaciones elementales. De haberlo hecho, se debería tener en cuenta en el resultado final. Por ejemplo, si en lugar de eliminar los dos unos de la tercera y cuarta filas en el último paso, se hubieran eliminado los de las segunda y cuarta filas usando la tercera, las filas LI obtenidas serían la primera y la tercera (a pesar de llegar a la misma R_A). Al no ser tan transparente como el que usa las columnas en los cálculos, no es recomendable.

Un resultado similar al de la proposición anterior es válido si se consideran dos subespacios de un mismo espacio vectorial.

Proposición 8.10. Sean S_1 y S_2 dos subespacios de dimensión finita de un $\mathbb{K}$-espacio vectorial V. Se cumple que:

(a) Si $S_1 \subseteq S_2$ entonces $\dim_{\mathbb{K}}(S_1) \leq \dim_{\mathbb{K}}(S_2)$.

(b) Si $S_1 \subseteq S_2$ y $\dim_{\mathbb{K}}(S_1) = \dim_{\mathbb{K}}(S_2)$ entonces $S_1 = S_2$.

Demostración. Dado que S_1 tiene una base con un número finito de elementos y $S_1 \subseteq S_2$, dicha base es un conjunto linealmente independiente de S_2, y el mismo puede ser ampliado a una base, también con un número finito de elementos, de S_2.

(a) Es claro que $\dim_{\mathbb{K}}(S_1) \leq \dim_{\mathbb{K}}(S_2)$.

(b) La hipótesis $\dim_{\mathbb{K}}(S_1) = \dim_{\mathbb{K}}(S_2)$ indica que, al ampliar una base de S_1 a una base $\mathcal{B}$ de S_2, no ha sido necesario agregar elementos extra. Luego, $S_1 = \overline{\mathcal{B}} = S_2$.

□

Ejemplo 8.65. *Sea $S_1 \subseteq \mathbb{R}^3$ el subespacio de soluciones del sistema lineal homogéneo dado por*

$$\begin{cases} x + y - z &= 0 \\ x - y - z &= 0 \end{cases}.$$

¿Cúales son los valores posibles para la dimensión de un subespacio $S_2 \subseteq \mathbb{R}^3$ tal que S_1 sea un subespacio propio de S_2?

◁ Resolviendo el sistema lineal dado se obtiene que

$$S_1 = \{(x, 0, x) : x \in \mathbb{R}\}.$$

Puesto que un vector genérico de S_1 se puede escribir como

$$(x, 0, x) = x(1, 0, 1), \qquad \text{con } x \in \mathbb{R},$$

es claro que $\overline{\{(1, 0, 1)\}} = S_1$. Además, $\{(1, 0, 1)\}$ es linealmente independiente, con lo cual $\{(1, 0, 1)\}$ determina una base de S_1 y, por lo tanto, $\dim_{\mathbb{R}}(S_1) = 1$. Puesto que $S_1 \lneq S_2 \leq \mathbb{R}^3$ y $\dim_{\mathbb{R}}(\mathbb{R}^3) = 3$ se tiene que $\dim_{\mathbb{R}}(S_2) \in \{2, 3\}$. Si fuese $\dim_{\mathbb{R}}(S_2) = 2$, S_2 representaría geométricamente un plano que pasa por el origen y contiene a la recta determinada por S_1, y si $\dim_{\mathbb{R}}(S_2) = 3$ se tendría $S_2 = \mathbb{R}^3$. $\triangleright$

A continuación se estudian algunas formas en que se pueden relacionar dos o más subespacios para que el resultado sea de nuevo un subespacio vectorial.

8.10. Operaciones con subespacios

Las operaciones se basan en las definidas para conjuntos.

8.10.1. Intersección de subespacios

Dados dos subespacios S_1 y S_2 de un $\mathbb{K}$-espacio vectorial V, una pregunta natural es saber si el subconjunto $S_1 \cap S_2$ es un subespacio de V o no. La respuesta es afirmativa.

Proposición 8.11. Sean S_1 y S_2 dos subespacios de un $\mathbb{K}$-espacio vectorial V. Entonces la intersección $S_1 \cap S_2$ es un subespacio de V.

Demostración. Se deben probar las tres propiedades de la Proposición 8.3.

(S_1) Al ser S_1 y S_2 subespacios, $0 \in S_1$ y $0 \in S_2$. Luego $0 \in S_1 \cap S_2$.

(S_2) Si $u, v \in S_1 \cap S_2$ entonces $u, v \in S_1$ y $u, v \in S_2$. Por ser S_1 un subespacio, $u + v \in S_1$, y por ser S_2 un subespacio, $u + v \in S_2$. Luego, $u + v \in S_1 \cap S_2$.

(S_3) Sean $u \in S_1 \cap S_2$ y $\lambda \in \mathbb{K}$. Entonces $u \in S_1$ y $u \in S_2$. Por ser S_1 y S_2 subespacios, se tiene que $\lambda u \in S_1$ y $\lambda u \in S_2$. Por lo tanto, $\lambda u \in S_1 \cap S_2$.

En consecuencia, $S_1 \cap S_2$ es un subespacio de V. □

Ejemplo 8.66. *Encontrar la intersección de los subespacios de* $\mathbb{R}^3$ *dados por*

$$S_1 = \{(x_1, x_2, x_3) \in \mathbb{R}^3 : \, x_2 = 0\} \quad y \quad S_2 = \{(x_1, x_2, x_3) \in \mathbb{R}^3 : \, x_3 = 0\}.$$

◁ La intersección de los subespacios S_1 y S_2 de $\mathbb{R}^3$ es

$$S_1 \cap S_2 = \{(x_1, x_2, x_3) \in \mathbb{R}^3 : \, x_2 = 0 \wedge x_3 = 0\} = \{(x_1, 0, 0), \, x_1 \in \mathbb{R}\}$$

que, efectivamente, es un subespacio de $\mathbb{R}^3$. Geométricamente, S_1 y S_2 representan dos planos en $\mathbb{R}^3$ que pasan por el origen y su intersección es una recta que pasa por el origen. ▷

Puede ocurrir que la intersección de dos subespacios sea el subespacio trivial nulo.

Ejemplo 8.67. *La intersección de los subespacios de* $\mathbb{R}^2$ *dados por*

$$S_1 = \{(x_1, x_2) \in \mathbb{R}^2 : \, x_1 = 0\} \quad y \quad S_2 = \{(x_1, x_2) \in \mathbb{R}^2 : \, x_2 = 0\}$$

es el subespacio trivial nulo.

Proposición 8.12. Si $\{S_i\}_{i \in I}$ es una famila de subespacios de un $\mathbb{K}$-espacio vectorial V entonces la intersección $\cap_{i \in I} S_i$ es un subespacio de V.

Demostración. Véase el Ejercicio (35) del final del capítulo. □

Ejemplo 8.68. *Sea* $A = [a_{ij}]_{m\times n}$ *una matriz fija. Hallar la intersección de los subespacios* S_i *del espacio vectorial* $\mathbb{K}^{n\times 1}$ *dados por*

$$S_i := \left\{ \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix}^t \in \mathbb{K}^{n\times 1} : a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n = 0 \right\},$$

para cada $i = 1, 2, \dots, m$.

◁ Denotando por $S := \cap_{i=1}^m S_i$ la intersección de todas las ecuaciones lineales dadas por cada S_i $(i = 1, 2, \dots, m)$, de la Proposición 8.12 se deduce que S es un subespacio de $\mathbb{K}^{n\times 1}$ y viene dado el conjunto de todas las soluciones del sistema de ecuaciones lineales homogéneo

$$\left\{ \begin{array}{rcl} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n & = & 0 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n & = & 0 \\ & \vdots & \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n & = & 0 \end{array} \right.$$

Es decir, $S = \{x \in \mathbb{K}^{n\times 1} : Ax = 0\}$. ▷

Comparar este ejemplo con el Ejemplo 8.19.

Proposición 8.13. Sea V un $\mathbb{K}$-espacio vectorial. La intersección $S_1 \cap S_2$ de dos subespacios S_1 y S_2 de V es el mayor (en el sentido de la inclusión) subespacio de V incluido en S_1 y S_2.
En símbolos,

$$\text{Si } S \leq V : S \subseteq S_1 \wedge, S \subseteq S_2 \Rightarrow S \subseteq S_1 \cap S_2.$$

Demostración. Se ha probado que $S_1 \cap S_2$ es subespacio de V.

Es evidente que $S_1 \cap S_2 \subseteq S_1$ y $S_1 \cap S_2 \subseteq S_2$.

Ahora, sea $S \leq V$ tal que $S \subseteq S_1$ y $S \subseteq S_2$. Es claro que, como conjuntos, $S \subseteq S_1 \cap S_2$. Por tanto, se tiene que $S_1 \cap S_2$ es el mayor subespacio de V incluido en S_1 y S_2. □

Observación 8.19. En algunos libros se define el subespacio generado por un conjunto del siguiente modo. Sea X_0 subconjunto de un $\mathbb{K}$-espacio vectorial V. Se llama **subespacio generado** por X_0 a la intersección de todos los subespacios de V que contienen a X_0, y se denota $\overline{X_0}$. Es interesante comparar esta definición con la Proposición 8.5.

Observación 8.20. Se tiene que: Si V es un $\mathbb{K}$-espacio vectorial entonces $S \leq V \Leftrightarrow S = \overline{S}$. En particular, $V = \overline{V}$.

8.10.2. Suma de subespacios

Si se consideran los subespacios de $\mathbb{R}^3$ dados por

$$S_1 = \{(x_1, x_2, x_3) \in \mathbb{R}^3 : \ x_2 = 0\} \quad \text{y} \quad S_2 = \{(x_1, x_2, x_3) \in \mathbb{R}^3 : \ x_3 = 0\},$$

su unión es

$$S_1 \cup S_2 = \{(x_1, x_2, x_3) \in \mathbb{R}^3 : \ x_2 = 0 \ \vee \ x_3 = 0\},$$

es decir, está formada por ternas de $\mathbb{R}^3$ tales que la segunda componente es 0 o bien la tercera componente es 0 (pero no necesariamente ambas).

Al considerar los vectores

$$(1, 0, 1) \in S_1 \subseteq S_1 \cup S_2 \qquad \text{y} \qquad (1, 1, 0) \in S_2 \subseteq S_1 \cup S_2,$$

(es decir, ambos pertenecientes a $S_1 \cup S_2$), se cumple que

$$(1, 0, 1) + (1, 1, 0) = (2, 1, 1) \notin S_1 \cup S_2.$$

Luego, en general, $S_1 \cup S_2$ no es un subespacio vectorial.

Sin embargo, es posible construir el menor subespacio que contenga a S_1 y S_2 añadiendo los elementos que se obtienen al sumar los de ambos subespacios. Este nuevo subespacio coincidirá con $\overline{S_1 \cup S_2}$.

Proposición 8.14. Sea V un $\mathbb{K}$-espacio vectorial. Si S_1 y S_2 son dos subespacios de V entonces

$$S_1 + S_2 := \{s_1 + s_2 : \ s_1 \in S_1 \ \wedge \ s_2 \in S_2\}$$

es un subespacio vectorial de V. Este subespacio se llama **suma** de los subespacios S_1 y S_2.

Demostración. Se deben probar las 3 condiciones de la Proposición 8.3. En efecto,

(S_1) Por ser S_1 y S_2 subespacios, se tiene que $0 \in S_1$ y $0 \in S_2$. Luego $0 = 0 + 0 \in S_1 + S_2$.

(S_2) Si $s, s' \in S_1 + S_2$ entonces $s = s_1 + s_2$ y $s' = s_1' + s_2'$, con $s_1, s_1' \in S_1$ y $s_2, s_2' \in S_2$. Luego, $s + s' = (s_1 + s_2) + (s_1' + s_2') = (s_1 + s_1') + (s_2 + s_2') \in S_1 + S_2$ por ser S_1 y S_2 cerrados para la suma.

(S_3) Si $s \in S_1 + S_2$ y $\lambda \in \mathbb{K}$ entonces $s = s_1 + s_2$, con $s_1 \in S_1$ y $s_2 \in S_2$. Luego, $\lambda s = \lambda(s_1 + s_2) = (\lambda s_1) + (\lambda s_2) \in S_1 + S_2$ pues S_1 y S_2 son cerrados para la multiplicación por escalares.

Así, $S_1 + S_2$ es un subespacio vectorial de V. □

Proposición 8.15. Sea V un $\mathbb{K}$-espacio vectorial. Si S_1 y S_2 son dos subespacios de V entonces

$$S_1 + S_2 = \overline{S_1 \cup S_2}.$$

Demostración. Se debe probar que: $S_1 + S_2 \subseteq \overline{S_1 \cup S_2}$ y $S_1 + S_2 \supseteq \overline{S_1 \cup S_2}$.

($\subseteq$) Si $s \in S_1 + S_2$, existen elementos $s_1 \in S_1$ y $s_2 \in S_2$ tales que $s = s_1 + s_2$. Como $s_1 \in S_1 \subseteq S_1 \cup S_2$ y $s_2 \in S_2 \subseteq S_1 \cup S_2$ se tiene que $s_1, s_2 \in S_1 \cup S_2 \subseteq \overline{S_1 \cup S_2}$. Como $\overline{S_1 \cup S_2}$ es un subespacio, $s = s_1 + s_2 \in \overline{S_1 \cup S_2}$. Por lo tanto, $S_1 + S_2 \subseteq \overline{S_1 \cup S_2}$.

($\supseteq$) Es claro que $S_1 \subseteq S_1 + S_2$ pues si $s = s_1 \in S_1$ entonces $s = s_1 + 0 \in S_1 + S_2$. Análogamente, $S_2 \subseteq S_1 + S_2$ y, por tanto, $S_1 \cup S_2 \subseteq S_1 + S_2$.

Puesto que S_1+S_2 es un subespacio vectorial de V que contiene a $S_1\cup S_2$ y, por la Proposición 8.5, $\overline{S_1\cup S_2}$ es el menor subespacio[18] de V que contiene a $S_1\cup S_2$, resulta que $\overline{S_1\cup S_2}\subseteq S_1+S_2$. □

En definitiva,

$$S_1+S_2=\{s_1+s_2:\ s_1\in S_1\wedge s_2\in S_2\}=\overline{S_1\cup S_2}.$$

Ejemplo 8.69. *Se consideran los subespacios de* $\mathbb{R}^4$ *dados por*

$$S_1=\{(x_1,0,0,x_4)\in\mathbb{R}^4\}\quad y\quad S_2=\left\{(x_1,x_2,x_3,x_4)\in\mathbb{R}^4:\left\{\begin{smallmatrix}x_1+x_3=0\\x_1-x_3=0\end{smallmatrix}\right\}\right\}.$$

Calcular S_1+S_2.

◁ Los subespacios de $\mathbb{R}^4$ están dados por

$$S_1=\{(x_1,0,0,x_4)\in\mathbb{R}^4\}\quad \text{y}\quad S_2=\left\{(x_1,x_2,x_3,x_4)\in\mathbb{R}^4:x_1=x_3=0\right\}.$$

La forma de un vector genérico de S_2 es $(0,y_2,0,y_4)$ con $y_2,y_4\in\mathbb{R}$. La suma de un elemento genérico de cada subespacio es de la forma

$$(x_1,0,0,x_4)+(0,y_2,0,y_4)=(x_1,y_2,0,x_4+y_4)$$

con lo que la suma de ambos subespacios viene dada por

$$S_1+S_2=\{(z_1,z_2,0,z_4):\ z_1,z_2,z_4\in\mathbb{R}\}.$$

▷

Sistema generador de la suma de subespacios

A continuación se presenta una nueva caracterización para la suma de dos subespacios en términos de sistemas de generadores, que será de utilidad en la práctica.

[18]Se recuerda que el subespacio generado por $S_1\cup S_2$ está formado por todas las combinaciones lineales finitas de elementos de $S_1\cup S_2$.

Proposición 8.16. Sean S_1 y S_2 dos subespacios de un $\mathbb{K}$-espacio vectorial V tales que $\{u_j\}_{j\in I}$ es un sistema de generadores de S_1 y $\{v_k\}_{k\in J}$ es un sistema de generadores de S_2. Entonces $\{u_j\}_{j\in I}\cup\{v_k\}_{k\in J}$ es un sistema de generadores de S_1+S_2.

Demostración. Es claro que

$$\{u_j\}_{j\in I}\cup\{v_k\}_{k\in J}\subseteq S_1\cup S_2\subseteq \overline{S_1\cup S_2}=S_1+S_2.$$

Por ser S_1+S_2 un subespacio se tiene que $\overline{\{u_j\}_{j\in I}\cup\{v_k\}_{k\in J}}\subseteq S_1+S_2$.

Sea $s\in S_1+S_2$. Por definición, existen $s_1\in S_1$ y $s_2\in S_2$ tales que

$$s=s_1+s_2.$$

Al ser $\{u_j\}_{j\in I}$ un sistema de generadores de S_1 es posible escribir s_1 como combinación lineal finita de elementos de $\{u_j\}_{j\in I}$. De forma similar, s_2 se escribe como combinación lineal finita de elementos de $\{v_k\}_{k\in J}$. Luego, $s=s_1+s_2$ es combinación lineal finita de elementos de $\{u_j\}_{j\in I}\cup\{v_k\}_{k\in J}$, es decir, $s_1+s_2\in\overline{\{u_j\}_{j\in I}\cup\{v_k\}_{k\in J}}$.

Luego, $S_1+S_2=\overline{\{u_j\}_{j\in I}\cup\{v_k\}_{k\in J}}$, de donde resulta que $\{u_j\}_{j\in I}\cup\{v_k\}_{k\in J}$ es un sistema de generadores de S_1+S_2. □

En la práctica, al unir sistemas de generadores de cada uno de los subespacios S_1 y S_2, y extraer un sistema linealmente independiente maximal de dicha unión, se consigue una base para la suma S_1+S_2 de dichos subespacios.

Dimensión de la suma de dos subespacios

Al definir las operaciones entre subespacios se estudió la intersección y la suma de dos subespacios. Es posible establecer una relación entre las dimensiones de estos subespacios.

Proposición 8.17 (Fórmula de Grassmann)**.** Sean S_1 y S_2 dos subespacios de dimensión finita de un $\mathbb{K}$-espacio vectorial V. Entonces S_1+S_2 es finito-dimensional y

$$\dim_{\mathbb{K}}(S_1+S_2) = \dim_{\mathbb{K}}(S_1) + \dim_{\mathbb{K}}(S_2) - \dim_{\mathbb{K}}(S_1 \cap S_2).$$

Demostración. Al ser $S_1 \cap S_2 \leq S_1$, se tiene que $S_1 \cap S_2$ es finitamente generado pues S_1 lo es. Sea $s := \dim_{\mathbb{K}}(S_1 \cap S_2)$.

Si $\boxed{s > 0}$, se considera una base de $S_1 \cap S_2$:

$$\mathcal{B}_{S_1 \cap S_2} = \{x_1, \ldots, x_s\}.$$

Dado que $S_1 \cap S_2 \leq S_1$, por el teorema de ampliación de conjuntos linealmente independientes a una base, se pueden considerar vectores $\{y_{s+1}, \ldots, y_t\} \subseteq S_1$ tales que

$$\mathcal{B}_{S_1} = \{x_1, \ldots, x_s, y_{s+1}, \ldots, y_t\}$$

es una base de S_1. Análogamente, dado que $S_1 \cap S_2 \leq S_2$, existen vectores $\{z_{s+1}, \ldots, z_\ell\} \subseteq S_2$ tales que

$$\mathcal{B}_{S_2} = \{x_1, \ldots, x_s, z_{s+1}, \ldots, z_\ell\}$$

es una base de S_2. De este modo, $\dim_{\mathbb{K}}(S_1) = t$ y $\dim_{\mathbb{K}}(S_2) = \ell$.

La fórmula quedará probada si se encuentra una base para $S_1 + S_2$ que contenga $\dim(S_1) + \dim(S_2) - \dim(S_1 \cap S_2) = t + \ell - s$ elementos, que es justamente la cantidad de vectores de S_1 y de S_2 reunidos, quitando los que están repetidos. Se considera entonces el conjunto

$$\mathcal{B}_{S_1+S_2} := \{x_1, \ldots, x_s, y_{s+1}, \ldots, y_t, z_{s+1}, \ldots, z_\ell\} \subseteq V.$$

Se tiene que:

- $\mathcal{B}_{S_1+S_2}$ es un conjunto linealmente independiente de V. En efecto, se considera la combinación lineal

$$0 = \alpha_1 x_1 + \cdots + \alpha_s x_s + \alpha_{s+1} y_{s+1} + \cdots + \alpha_t y_t + \beta_{s+1} z_{s+1} + \cdots + \beta_\ell z_\ell$$

con todos los escalares en $\mathbb{K}$. Luego,

$$\begin{aligned}\beta_{s+1}z_{s+1} + \cdots + \beta_\ell z_\ell &= \\ &= -\alpha_1 x_1 - \cdots - \alpha_s x_s - \alpha_{s+1}y_{s+1} - \cdots - \alpha_t y_t, \qquad (8.13)\end{aligned}$$

de donde el primer miembro de esta última igualdad pertenece a $S_1 \cap S_2$. Por lo tanto, dicho elemento se puede escribir como combinación lineal de los elementos de la base $\mathcal{B}_{S_1\cap S_2}$, es decir, existen escalares $\gamma_1, \ldots, \gamma_s \in \mathbb{K}$ tales que

$$\beta_{s+1}z_{s+1} + \cdots + \beta_\ell z_\ell = \gamma_1 x_1 + \cdots + \gamma_s x_s.$$

En consecuencia, $0 = \gamma_1 x_1 + \cdots + \gamma_s x_s - \beta_{s+1}z_{s+1} + \cdots - \beta_\ell z_\ell$, y de la independencia lineal de los vectores de $\mathcal{B}_{S_2}$ se tiene que $\gamma_1 = \cdots = \gamma_s = \beta_{s+1} = \cdots = \beta_\ell = 0$. Sustituyendo en (8.13) se tiene que

$$0 = \alpha_1 x_1 + \cdots + \alpha_s x_s + \alpha_{s+1}y_{s+1} + \cdots + \alpha_t y_t,$$

y de la independencia lineal del conjunto $\mathcal{B}_{S_1}$ se concluye que $\alpha_j = 0$ para todo $j = 1, \ldots, s, s+1, \ldots, t$.

- $\mathcal{B}_{S_1+S_2}$ es un sistema de generadores de S_1+S_2. Se deduce directamente de la Proposición 8.16.

Se ha probado que $S_1 + S_2$ es finito-dimensional y $\dim(S_1 + S_2) = t + \ell - s = \dim(S_1) + \dim(S_2) - \dim(S_1 \cap S_2)$.

Si $\boxed{s = 0}$ entonces $S_1 \cap S_2 = \{0\}$ y el conjunto $\mathcal{B}_{S_1\cap S_2} = \emptyset$ es una base de $S_1 \cap S_2$. La demostración es similar procediendo como antes a partir (directamente) de bases de S_1 y S_2. □

Ejemplo 8.70. *Sean*

$$S_1 = \left\{ \begin{bmatrix} x & y \\ z & t \end{bmatrix} \in \mathbb{R}^{2\times 2} : z = 0 \wedge x - t = 0 \right\}$$

y

$$S_2 = \overline{\left\{ A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, B = \begin{bmatrix} 2 & 0 \\ 3 & 2 \end{bmatrix}, C = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \right\}}.$$

Calcular las dimensiones de los subespacios $S_1 + S_2$ y $S_1 \cap S_2$.

$\lhd$ Un vector genérico de S_1 tiene la forma

$$\begin{bmatrix} x & y \\ 0 & x \end{bmatrix} = x \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + y \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad x, y \in \mathbb{R}$$

luego $S_1 = \overline{\left\{ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \right\}}$. Al ser los dos vectores linealmente independientes (comprobarlo), forman base de S_1, con lo que $\dim_{\mathbb{R}}(S_1) = 2$.

Por otro lado, se observa que $B = 2A + C$ y que $\{A, C\}$ es linealmente independiente. Por tanto, $\{A, C\}$ es una base de S_2. Luego, $\dim_{\mathbb{R}}(S_2) = 2$.

Por la Proposición 8.16, un sistema de generadores de $S_1 + S_2$ se obtiene uniendo un sistema generador de cada subespacio. Por lo tanto,

$$S_1 + S_2 = \overline{\left\{ B_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, B_2 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, C = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \right\}}.$$

Para hallar su dimensión, de este sistema de generadores se debe extraer una base. Se observa que $B_1 = A - C$ y $\{B_2, A, C\}$ es un conjunto linealmente independiente de $\mathbb{R}^{2\times 2}$ (pues A y C lo son y B_2 no es combinación lineal de A y C). Luego, $\dim_{\mathbb{R}}(S_1 + S_2) = 3$ y

$$\dim(S_1 \cap S_2) = \dim(S_1) + \dim(S_2) - \dim(S_1 + S_2) = 2 + 2 - 3 = 1.$$

Se propone como ejercicio comprobar que una base de $S_1 \cap S_2$ es $\{I_2\}$. Por último, es posible observar que, en este ejemplo, el subespacio $S_1 \cap S_2$ está formado todas las matrices escalares de $\mathbb{R}^{2\times 2}$. $\rhd$

8.10.3. Suma directa de subespacios

Dados dos subespacios S_1 y S_2 de un $\mathbb{K}$-espacio vectorial V, se ha probado que $S_1 + S_2 = \overline{S_1 \cup S_2}$ y que para obtener un sistema de generadores de $S_1 + S_2$ alcanza con reunir un sistema de generadores de cada uno de los subespacios. Puede ocurrir que esta unión tenga vectores redundantes o bien resulte un conjunto que sea una base de $S_1 + S_2$. La siguiente definición permitirá distinguir estas dos situaciones. Para que no hayan vectores redundantes en dicha unión, el espacio vectorial se deberá poder descomponer en suma de (dos) subespacios completamente independientes (es decir, que sólo compartan el vector nulo).

Definición 8.11. Sea V un $\mathbb{K}$-espacio vectorial y sean S_1 y S_2 dos subespacios de V. Se dice que V es **suma directa** de S_1 y S_2, y se denota $V = S_1 \oplus S_2$[a], si se cumplen las dos condiciones siguientes:

(SD1) $V = S_1 + S_2$,

(SD2) $S_1 \cap S_2 = \{0\}$.

[a]La notación $\oplus$ utilizada para la suma directa de subespacios no tiene relación alguna con la misma notación utilizada para la ley de composición interna en la definición de espacio vectorial.

La condición (SD1) dice que el espacio vectorial V es suma de los subespacios S_1 y S_2. Cuando dos subespacios S_1 y S_2 cumplen la condición (SD2) se dice que son **subespacios independientes**.

Ejemplo 8.71. *El $\mathbb{R}$-espacio vectorial*

(a) $\mathbb{R}^2$ *es suma directa de los subespacios vectoriales* $S_1 = \overline{\{e_1 = (1,0)\}}$ *y* $S_2 = \overline{\{e_2 = (0,1)\}}$.

(b) $\mathbb{R}^3$ *es suma directa de los subespacios vectoriales* $S_1 =$

$\overline{\{e_1 = (1,0,0), e_2 = (0,1,0)\}}$ y $S_2 = \overline{\{e_3 = (0,0,1)\}}$.

$\lhd$ (a) Es claro que un vector genérico $(x, y) \in \mathbb{R}^2$ se puede escribir como $(x, y) = x(1, 0) + y(0, 1)$. Luego, $\mathbb{R}^2 = S_1 + S_2$. Además, si $(x, y) \in S_1 \cap S_2$ entonces $(x, y) = \alpha(1, 0)$ y $(x, y) = \beta(0, 1)$ para ciertos $\alpha, \beta \in \mathbb{R}$. Operando se tiene que $(x, y) = (0, 0)$. Luego, $S_1 \cap S_2 = \{(0, 0)\}$. En definitiva, $\mathbb{R}^2 = S_1 \oplus S_2$.

(b) Es similar al caso anterior y se propone como ejercicio. $\rhd$

El siguiente resultado formaliza los comentarios previos a la definición anterior.

Proposición 8.18. Sea V un $\mathbb{K}$-espacio vectorial y sean S_1 y S_2 dos subespacios de V. Las siguientes condiciones son equivalentes:

(a) $V = S_1 \oplus S_2$.

(b) Todo vector de V se escribe de forma única como suma de uno de S_1 y otro de S_2, es decir, $\forall u \in V$, existen vectores unívocamente determinados $s_1 \in S_1$ y $s_2 \in S_2$ tales que $u = s_1 + s_2$.

(c) Si $\mathcal{B}_{S_1}$ es una base de S_1 y $\mathcal{B}_{S_2}$ es una base de S_2 entonces $\mathcal{B}_{S_1} \cup \mathcal{B}_{S_2}$ es una base de V.

Demostración. (a) $\Rightarrow$ (b) Si $V = S_1 \oplus S_2$, por (SD1) se tiene que $V = S_1 + S_2$. Por lo tanto, si $u \in V$, existen vectores $s_1 \in S_1$ y $s_2 \in S_2$ tales que

$$u = s_1 + s_2.$$

Falta ver que dicha representación es única. En efecto, sean $s_1' \in S_1$ y $s_2' \in S_2$ tales que $u = s_1' + s_2'$. Entonces $s_1 + s_2 = s_1' + s_2'$, de donde $s_1 - s_1' = s_2' - s_2$. Al ser S_1 y S_2 subespacios, $s_1 - s_1' \in S_1$ y $s_2' - s_2 \in S_2$ y, por (SD2), se tiene que

$$s_1 - s_1' = s_2' - s_2 \in S_1 \cap S_2 = \{0\}.$$

Luego, $s_1 = s_1'$ y $s_2 = s_2'$, lo que prueba (b).

(b) $\Rightarrow$ (c) Sean $\mathcal{B}_{S_1} = \{u_i\}_{i\in I}$ y $\mathcal{B}_{S_2} = \{v_j\}_{j\in J}$ bases de S_1 y S_2, respectivamente. Se debe probar que $\mathcal{B} := \mathcal{B}_{S_1} \cup \mathcal{B}_{S_2}$ es una base de V. En efecto, si se considera el vector nulo escrito como combinación lineal finita formada por elementos de $\mathcal{B}_{S_1} \cup \mathcal{B}_{S_2}$ (es decir, de $\mathcal{B}_{S_1}$ y $\mathcal{B}_{S_2}$), y se aplican las propiedades conmutativa y asociativa se tiene

$$0 = \sum_{i\in I} \alpha_i u_i + \sum_{j\in J} \beta_j v_j \in S_1 + S_2,$$

donde, a lo sumo, hay un número finito de escalares α_i y β_j no nulos.

Por otro lado, es sabido que siempre se cumple que

$$0 = 0 + 0 \in S_1 + S_2.$$

Por la hipótesis (b),

$$0 = \sum_{i\in I} \alpha_i u_i \qquad \text{y} \qquad 0 = \sum_{j\in J} \beta_j v_j.$$

Utilizando que $\mathcal{B}_{S_1}$ es base de S_1 y $\mathcal{B}_{S_2}$ es base de S_2 se tiene que $\alpha_i = 0$, $\forall i \in I$ y $\beta_j = 0$, $\forall j \in J$. Por lo tanto, $\mathcal{B}$ es una familia linealmente independiente de V.

Ahora se prueba que $\overline{\mathcal{B}} = V$. Sea $u \in V$. Por (b), existen vectores $s_1 \in S_1$ y $s_2 \in S_2$ tales que $u = s_1 + s_2$. Como $\mathcal{B}_{S_1} = \{u_i\}_{i\in I}$ y $\mathcal{B}_{S_2} = \{v_j\}_{j\in J}$ son sistemas de generadores de S_1 y S_2, respectivamente, existen escalares $\alpha_i, \beta_j \in \mathbb{K}$ tales que

$$s_1 = \sum_{i\in I} \alpha_i u_i, \qquad s_2 = \sum_{j\in J} \beta_j v_j,$$

donde, a lo sumo, hay un número finito de escalares α_i y β_j no nulos. Luego,

$$u = s_1 + s_2 = \sum_{i\in I} \alpha_i u_i + \sum_{j\in J} \beta_j v_j \in \overline{\{u_i\}_{i\in I} \cup \{v_j\}_{j\in J}}.$$

Por lo tanto, $\mathcal{B}$ es un sistema de generadores de V.

En consecuencia, $\mathcal{B}$ es una base de V.

(c) $\Rightarrow$ (a) Fijadas[19]

$$\mathcal{B}_{S_1} = \{u_i\}_{i\in I} \text{ una base de } S_1 \quad \text{y} \quad \mathcal{B}_{S_2} = \{v_j\}_{j\in J} \text{ una base de } S_2,$$

la hipótesis (c) asegura que $\mathcal{B} := \mathcal{B}_{S_1} \cup \mathcal{B}_{S_2}$ es una base de V.

Se debe probar que $V = S_1 + S_2$. Sea $u \in V$. Es posible escribir

$$u = \sum_{i\in I} \alpha_i u_i + \sum_{j\in J} \beta_j v_j \in S_1 + S_2,$$

con, a lo sumo, un número finito de escalares no nulos. Luego, $V \subseteq S_1 + S_2$. Por otro lado, como siempre se cumple que $S_1 + S_2 \subseteq V$, se ha probado que $V = S_1 + S_2$.

Falta ver que dicha suma es directa, es decir: $S_1 \cap S_2 = \{0\}$. En efecto, si $s \in S_1 \cap S_2$ entonces $s \in S_1$ y $s \in S_2$. Luego, es posible escribirlo(s) como combinación lineal de los vectores de sus respectivas bases, es decir $s = \sum_{i\in I} \alpha_i u_i$ y $s = \sum_{j\in J} \beta_j v_j$, para (a lo sumo un número finito de) ciertos escalares $\alpha_i, \beta_j \in \mathbb{K}$. Luego, de $\sum_{i\in I} \alpha_i u_i = \sum_{j\in J} \beta_j v_j$ resulta que

$$0 = \sum_{i\in I} \alpha_i u_i - \sum_{j\in J} \beta_j v_j,$$

y al ser $\{u_i\}_{i\in I}, \{v_j\}_{j\in J} \subseteq \mathcal{B}$ y $\mathcal{B}$ una base de V, se tiene que $\alpha_i = 0$, $\forall i \in I$ y $\beta_j = 0$, $\forall j \in J$. Luego, $s = 0$. Así, $S_1 \cap S_2 \subseteq \{0\}$. Como siempre se cumple $\{0\} \subseteq S_1 \cap S_2$, se ha demostrado que $S_1 \cap S_2 = \{0\}$. Por lo tanto, $V = S_1 \oplus S_2$. $\square$

Conocer las dimensiones de dos subespacios S_1 y S_2 de un espacio vectorial V puede ayudar a determinar si el espacio es suma directa de dichos subespacios. Para ello será requisito conocer previamente que $V = S_1 + S_2$.

[19] Existen por el Teorema 8.2, aunque está probado para espacios finitamente generados.

Proposición 8.19. Sea V un $\mathbb{K}$-espacio vectorial con $\dim_{\mathbb{K}}(V) = n \geq 1$ y sean S_1 y S_2 dos subespacios de V. Las siguientes condiciones son equivalentes:

(a) $V = S_1 \oplus S_2$.

(b) $V = S_1 + S_2 \quad$ y $\quad \dim_{\mathbb{K}}(V) = \dim_{\mathbb{K}}(S_1) + \dim_{\mathbb{K}}(S_2)$.

Demostración. Los apartados (a) y (b) comparten la condición $V = S_1 + S_2$.

Para probar que (a) $\Leftrightarrow$ (b) basta ver que

$$S_1 \cap S_2 = \{0\} \qquad \Leftrightarrow \qquad \dim_{\mathbb{K}}(V) = \dim_{\mathbb{K}}(S_1) + \dim_{\mathbb{K}}(S_2).$$

En efecto, este hecho se deduce directamente de la Fórmula de Grassmann:

$$\begin{aligned} S_1 \cap S_2 = \{0\} \;&\Leftrightarrow\; \dim_{\mathbb{K}}(S_1 \cap S_2) = 0 \\ &\Leftrightarrow\; \dim_{\mathbb{K}}(S_1 + S_2) = \dim_{\mathbb{K}}(S_1) + \dim_{\mathbb{K}}(S_2) \\ &\Leftrightarrow\; \dim_{\mathbb{K}}(V) = \dim_{\mathbb{K}}(S_1) + \dim_{\mathbb{K}}(S_2). \end{aligned}$$

$\square$

Ejemplo 8.72. *Se consideran los subespacios de $\mathbb{R}^3$ dados por*

$$S_1 = \{(x, y, z) \in \mathbb{R}^3 : x - 2y + z = 0\},$$

$$S_2 = \overline{\{(0, 0, 1)\}} \qquad y \qquad S_3 = \overline{\{(0, 1, 2)\}}.$$

Estudiar si es posible escribir a $\mathbb{R}^3$ como suma directa de algunos de los siguientes pares de subespacios:

(a) S_1 *y* S_2,

(b) S_1 *y* S_3

(c) S_2 *y* S_3.

$\triangleleft$ Puesto que un elemento genérico de S_1 tiene la forma $(2y - z, y, z) = y(2, 1, 0) + z(-1, 0, 1)$ se tiene que

$$S_1 = \overline{\{(2, 1, 0), (-1, 0, 1)\}}.$$

Por ser ambos vectores linealmente independientes se tiene que forman una base de S_1, con lo que $\dim_{\mathbb{R}}(S_1) = 2$.

Debido a que S_2 y S_3 son subespacios generados por un único vector no nulo, cada uno determina sendas bases. Luego, $\dim_{\mathbb{R}}(S_2) = 1 = \dim_{\mathbb{R}}(S_3)$.

(c) Como $\dim_{\mathbb{R}}(\mathbb{R}^3) = 3 \neq 2 = \dim_{\mathbb{R}}(S_2) + \dim_{\mathbb{R}}(S_3)$, es posible asegurar que $\mathbb{R}^3 \neq S_2 \oplus S_3$ por la Proposición 8.19 (b).

(a) y (b) Al ser $\dim_{\mathbb{R}}(\mathbb{R}^3) = \dim_{\mathbb{R}}(S_1) + \dim_{\mathbb{R}}(S_i)$, para $i = 2, 3$, en lugar de analizar si se cumple que $\mathbb{R}^3 = S_1 + S_2$, $\mathbb{R}^3 = S_1 + S_3$ o ninguno de ellos es más directo aplicar la Proposición 8.18 (c). En efecto, uniendo bases de S_1 y S_2 se tiene que $\{(2, 1, 0), (-1, 0, 1), (0, 0, 1)\}$ es linealmente independiente (comprobarlo), y por tanto base de $\mathbb{R}^3$. Luego, $\mathbb{R}^3 = S_1 \oplus S_2$. Geométricamente, la recta que determina S_2 no está contenida en el plano que determina S_1.

Sin embargo, al unir bases de S_1 y S_3, $\{(2, 1, 0), (-1, 0, 1), (0, 1, 2)\}$ es un conjunto linealmente dependiente (comprobarlo), y por tanto no es base de $\mathbb{R}^3$. Luego, $\mathbb{R}^3 \neq S_1 \oplus S_3$. Geométricamente, la recta que determina S_3 está contenida en el plano que determina S_1. $\triangleright$

Se observa del ejemplo anterior que en un caso la suma directa de dos subespacios da como resultado todo el espacio vectorial y en los otros no. El siguiente resultado indica una forma de completar un subespacio (dado) con otro, de modo que sumen directamente todo el espacio en cuestión.

Proposición 8.20. Sea V un $\mathbb{K}$-espacio vectorial con $\dim_{\mathbb{K}}(V) = n \geq 1$ y sea S_1 un subespacio de V. Entonces existe un subespacio S_2 de V tal que $V = S_1 \oplus S_2$.

Demostración. Sea $\mathcal{B}_{S_1}$ una base de S_1. Por el teorema de ampliación de conjuntos linealmente independientes a una base, es posible considerar una base $\mathcal{B}$ de V tal que $\mathcal{B}_{S_1} \subseteq \mathcal{B}$.

El conjunto $\mathcal{B} - \mathcal{B}_{S_1}$ es linealmente independiente por ser un subconjunto de $\mathcal{B}$ (que es linealmente independiente). Luego, tomando $\mathcal{B}_{S_2} := \mathcal{B} - \mathcal{B}_{S_1}$ como base de $S_2 := \overline{\mathcal{B} - \mathcal{B}_{S_1}}$ se tiene que (se propone como ejercicio completar los detalles de este paso) $V = S_1 \oplus S_2$.

$\square$

Definición 8.12. Sea V un $\mathbb{K}$-espacio vectorial y sea S_1 un subespacio de V. Todo subespacio S_2 de V tal que $V = S_1 \oplus S_2$ se llama subespacio **complementario**[a] de S_1.

[a] Algunos autores lo llaman **subespacio sumplementario** de S_1.

Ejemplo 8.73. *Se considera el subespacio de* $\mathbb{R}^3$ *dado por*

$$S_1 = \{(x, y, z) \in \mathbb{R}^3 : x - 2y + z = 0\}.$$

Hallar un subespacio complementario de S_1*. ¿Es único dicho complemento? Justificar.*

$\triangleleft$ Del Ejemplo 8.72, se sabe que $\dim_{\mathbb{R}}(S_1) = 2$ y que $\{(2, 1, 0), (-1, 0, 1)\}$ forma una base de S_1. Además, se probó que, para $S_2 = \overline{\{(0, 0, 1)\}}$, se tiene $\mathbb{R}^3 = S_1 \oplus S_2$, con lo que S_2 es un subespacio complementario de S_1.

No es único puesto que cualquier otro subespacio de $\mathbb{R}^3$ con dimensión 1 generado por cualquier vector $u \in \mathbb{R}^3$ tal que $\{(2, 1, 0), (-1, 0, 1), u\}$ forme una base de $\mathbb{R}^3$ es complementario de S_1. Una opción es añadir $u = (0, 0, 1)$, pero hay infinitas: se puede elegir cualquier vector no nulo que no pertenezca al plano determinado por S_1. $\triangleright$

Es posible extender la definición de suma directa a un número finito de subespacios como sigue.

Definición 8.13. Sea V un $\mathbb{K}$-espacio vectorial y sean $S_1, S_2, \ldots, S_t$ subespacios de V, con $t \geq 2$. Se dice que V es **suma directa** de $S_1, S_2, \ldots, S_t$, y se denota $V = S_1 \oplus S_2 \oplus \cdots \oplus S_t$, si se cumplen las dos condiciones siguientes:

(SD1) $V = S_1 + S_2 + \cdots + S_t$,

(SD2) $S_i \cap (S_1 + \cdots + S_{i-1} + S_{i+1} + \cdots + S_t) = \{0\}$ para $i = 1, 2, \ldots, t$.

De nuevo, cuando los subespacios $S_1, S_2, \ldots, S_t$ de V cumplen la condición (SD2) se llaman **subespacios independientes**, que extiende claramente el caso $S_1 \cap S_2 = \{0\}$ correspondiente a $t = 2$.

Observación 8.21. Se observa que en la Definición 8.13 se busca reconstruir el espacio V a partir de los subespacios S_i, $i = 1, 2, \ldots, t$, de modo que cada uno de estos subespacios sea independiente del subespacio que genera la unión de todos los demás.

El siguiente resultado extiende el dado en la la Proposición 8.18. Su demostración se propone como ejercicio.

Proposición 8.21. Sea V un $\mathbb{K}$-espacio vectorial y sean $S_1, S_2, \ldots, S_t$ subespacios de V, con $t \geq 2$. Las siguientes condiciones son equivalentes:

(a) $V = S_1 \oplus S_2 \oplus \cdots \oplus S_t$.

(b) Todo vector u de V se escribe de forma única como $u = s_1 + s_2 + \cdots + s_t$ con $s_i \in S_i$ para $i = 1, 2, \ldots, t$.

(c) Si $\mathcal{B}_{S_i}$ es una base de S_i para $i = 1, 2, \ldots, t$ entonces $\mathcal{B}_{S_1} \cup \mathcal{B}_{S_2} \cup \cdots \cup \mathcal{B}_{S_t}$ es una base de V.

8.11. EJERCICIOS

(1) Comprobar que todos los ejemplos proporcionados en la Sección 8.3.1 determinan un espacio vectorial.

(2) Se considera el conjunto $V = \mathbb{Z}$ de los números enteros con la suma habitual y con la operación externa:

$$\text{Si } q \in \mathbb{Q} \text{ y } z \in \mathbb{Z}, \text{ entonces } q \cdot z = 0.$$

Comprobar que en $(V, +, \cdot)$ se cumplen todos los axiomas de espacio vectorial salvo V_8, y por tanto, $\mathbb{Z}$ no es un espacio vectorial sobre $\mathbb{Q}$ con estas operaciones.

(3) ¿Cómo se deberían reordenar los axiomas de la definición de espacio vectorial para necesitar sólo una de las dos igualdades en el axioma de existencia del elemento neutro de la suma y existencia del simétrico para cada elemento del espacio?

(4) Demostrar que, de los axiomas (V_1)-(V_8) que cumple un conjunto V para ser un espacio vectorial sobre un cuerpo $\mathbb{K}$, el axioma (V_4) puede deducirse de los axiomas (V_1), (V_2), (V_3), (V_5), (V_6) y (V_8).
(Ayuda: Probar primero que $\vec{u}+\vec{u}+\vec{v}+\vec{v} = \vec{u}+\vec{v}+\vec{u}+\vec{v}$ para todo $\vec{u}, \vec{v} \in V$ desarrollando de dos formas diferentes la expresión $(1+1)(\vec{u}+\vec{v})$.)

(5) Sea V un $\mathbb{K}$-espacio vectorial y sean $\vec{u}, \vec{v} \in V$, $\lambda, \mu \in \mathbb{K}$. Demostrar que:

(I) $-u = (-1)u$.

(II) $\lambda(\vec{u} - \vec{v}) = \lambda\vec{u} - \lambda\vec{v}$.

(III) $(\lambda - \mu)\vec{u} = \lambda\vec{u} - \mu\vec{u}$.

(6) Sea V un $\mathbb{K}$-espacio vectorial y sean $\vec{u}, \vec{v} \in V$, $\lambda, \mu \in \mathbb{K}$. Demostrar que son válidas las siguientes **leyes de cancelación** o **reglas de simplificación**:

(I) $\lambda\vec{u} = \lambda\vec{v}$ y $\lambda \neq 0 \Rightarrow \vec{u} = \vec{v}$.

(II) $\lambda\vec{u} = \mu\vec{u}$ y $\vec{u} \neq \vec{0} \Rightarrow \lambda = \mu$.

(7) Deducir que el espacio vectorial cartesiano $\mathbb{K}^n$, el espacio vectorial de las matrices $\mathbb{K}^{m\times n}$ y el espacio vectorial de las sucesiones son casos particulares del Ejemplo 8.7.

(8) Se considera el conjunto $\mathcal{C}^m([-1,1])$ de todas las funciones reales a valores en el intervalo $[-1,1]$ que poseen derivadas de orden m continuas. Definiendo las operaciones de suma $+$ y producto por escalares $\cdot$ punto a punto, probar que $(\mathcal{C}^m([-1,1]),+,\cdot)$ es un espacio vectorial.

(9) Sea $V = \mathbb{R}^+ = \{x \in \mathbb{R} : x > 0\}$. Se define la suma y el producto por escalares de $\mathbb{K} = \mathbb{Q}$ como

$$\begin{array}{rcl} x \oplus y &=& x \cdot y \quad (\text{donde } \cdot \text{ es el producto de } \mathbb{R}), \\ \dfrac{a}{b} \otimes x &=& x^{\frac{a}{b}} = \sqrt[b]{x^a}, \quad \left(\text{donde } \dfrac{a}{b} \in \mathbb{Q} \text{ es tal que } b > 0\right) \end{array}$$

para cualquier $x, y \in V$, $\frac{a}{b} \in \mathbb{Q}$. Probar que $(V,\oplus,\otimes)$ es un $\mathbb{Q}$-espacio vectorial.

(10) Sean V y W dos espacios vectoriales sobre el mismo cuerpo $\mathbb{K}$. Se considera como nuevo espacio el producto cartesiano $V \times W$ y sobre él las siguientes operaciones de suma $+ : (V \times W) \times (V \times W) \to V \times W$ y producto por escalares $\cdot : \mathbb{K} \times (V \times W) \to V \times W$ definidas como

$$\begin{array}{rcl} (v_1,w_1) + (v_2,w_2) &=& (v_1+v_2, w_1+w_2) \\ \lambda \cdot (v,w) &=& (\lambda v, \lambda w) \end{array}.$$

Demostrar que $(V \times W,+,\cdot)$ es un $\mathbb{K}$-espacio vectorial, llamado **espacio producto** $V \times W$.

(11) Sea $(V,+,.)$ un $\mathbb{K}$-espacio vectorial y sea $S \subseteq V$. Demostrar que las siguientes condiciones son equivalentes:

(a) S es un subespacio vectorial de V.

(b) Se cumple que:

(S_1') $S \neq \emptyset$,

(S_2) Si $u, v \in S$ entonces $u + v \in S$,

(S_3) Si $u \in S$ y $\lambda \in \mathbb{K}$ entonces $\lambda.u \in S$.

(12) Sea $(V, +, .)$ un $\mathbb{K}$-espacio vectorial y sea $S \subseteq V$. Demostrar que las siguientes condiciones son equivalentes:

(a) S es un subespacio vectorial de V.

(b) Se cumple que:

(S_1) $0 \in S$,

(S_4') Si $u, v \in S$ y $\lambda \in \mathbb{K}$ entonces $u + \lambda.v \in S$.

(13) Indicar, justificando la respuesta, si los siguientes subconjuntos de $\mathbb{R}^2$ son subespacios vectoriales:

(a) $S_1 = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^2 \ : \ x^2 = y \right\}.$

(b) $S_2 = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^2 \ : \ xy = 1 \right\}.$

(c) $S_3 = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^2 \ : \ x - 2y = 0 \right\}.$

Representar gráficamente cada subconjunto y analizar la pertenencia o no a cada S_i del vector nulo de $\mathbb{R}^2$, de la suma y del producto de un vector por un escalar para unos cuantos vectores.

(14) En este ejercicio se estudian matrices que conmutan, con respecto a la multiplicación de matrices, con una matriz fijada.

(a) Demostrar que el conjunto de matrices que conmutan con la matriz

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

determina un subespacio vectorial de $\mathbb{R}^{2\times 2}$ sobre $\mathbb{R}$. Hallar su dimensión y una base.

(b) Considerar ahora una matriz fija $A \in \mathbb{C}^{n\times n}$ arbitraria. Demostrar que el conjunto de matrices $X \in \mathbb{C}^{n\times n}$ tales que $AX = XA$ forma un subespacio vectorial de $\mathbb{C}^{n\times n}$.

(15) Analizar si los siguientes subconjuntos son subespacios de los espacios vectoriales a los que pertenecen:

(a) $S = \left\{ \begin{bmatrix} x & y \\ z & t \end{bmatrix} \in \mathbb{R}^{2\times 2} : x - z = 0, y = 2t \right\}$.

(b) $S = \{p(x) = a_0 + a_1 x + a_2 x^2 \in \mathbb{R}_2[x] : a_0 = 1 = a_1\}$.

(c) $S = \{(x, y) \in \mathbb{R}^2 : y = -2x\}$.

(d) $S = \{p(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 \in \mathbb{R}_3[x] : a_1 = a_0, a_2 = -a_1\}$.

(16) Para los apartados del ejercicio anterior que sean subespacios determinar una base y su dimensión.

(17) Indicar si el vector $\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \in \mathbb{R}^{2\times 2}$ pertenece al subespacio

$$\overline{\left\{ \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ -1 & -1 \end{bmatrix} \right\}}.$$

¿Cuál es la dimensión del subespacio anterior?

(18) Analizar la validez o falsedad de la siguiente afirmación: "el subconjunto $\mathbb{R}^2$ es un subespacio vectorial del espacio vectorial $\mathbb{C}^2$ con la suma y el producto por escalares habituales de $\mathbb{C}^2$".

(19) Mostrar (de dos maneras diferentes) que los vectores $1, x, x^2$ y x^3 son linealmente independientes en $\mathbb{R}_3[x]$.

(20) Determinar si cada uno de los sistemas

$S_1 = \{1, 1+x, -1+x\}$, $S_2 = \{1+x, -1+x\}$ y $S_3 = \{3, x-2, x+1, x^2\}$

constituye una base de $\mathbb{R}_2[x]$. En caso de no ser base, pero ser un conjunto linealmente independiente, extenderlo a una base de $\mathbb{R}_2[x]$; en caso de ser un conjunto de generadores, extraer de él un subconjunto que sea una base.

(21) Sea S el subconjunto de $\mathbb{R}_3[x]$ de los polinomios de grado menor o igual que 3 (junto al polinomio nulo) que se anulan en $x = 0$ y en $x = 1$.

(a) Mostrar que S es un subespacio vectorial de $\mathbb{R}_3[x]$.

(b) Hallar la forma general de los polinomios de S.

(c) Hallar una base de S y su dimensión.

(22) Determinar la dimensión y una base de los siguientes subespacios vectoriales:

(a) $S = \left\{ \begin{bmatrix} a - b \\ 2a + b \end{bmatrix} : \, a, b \in \mathbb{R} \right\} \subseteq \mathbb{R}^2.$

(b) $S = \left\{ \begin{bmatrix} a - c \\ b + c \\ 5c \end{bmatrix} : \, a, b, c \in \mathbb{R} \right\} \subseteq \mathbb{R}^3.$

(c) $S = \left\{ \begin{bmatrix} a \\ 0 \\ -2a \end{bmatrix} : \, a \in \mathbb{R} \right\} \subseteq \mathbb{R}^3.$

(d) $S = \left\{ \begin{bmatrix} a - c \\ 2a + b + c \end{bmatrix} : \, a, b, c \in \mathbb{R} \right\} \subseteq \mathbb{R}^2.$

(e) $S = \left\{ \begin{bmatrix} a - \frac{b}{2} + 2c \\ 2a - b + 4c \end{bmatrix} : \, a, b, c \in \mathbb{R} \right\} \subseteq \mathbb{R}^2.$

(f) $S = \left\{ \begin{bmatrix} 2x + y \\ x \\ -x - 2y \\ x + y + z \end{bmatrix} : \, x, y, z \in \mathbb{R} \right\} \subseteq \mathbb{R}^4.$

(g) $S = \left\{ \begin{bmatrix} a & b \\ c & -a \end{bmatrix} : a, b, c \in \mathbb{R} \right\} \subseteq \mathbb{R}^{2\times 2}$.

(23) Se considera el conjunto M_S de las matrices de tamaño $n \times n$ triangulares superiores a coeficientes reales con la adición y la multiplicación por escalares habituales.

(a) Probar que M_S es un subespacio vectorial de $\mathbb{R}^{n\times n}$. Calcular su dimensión y una base.

(b) Para $n = 4$, escribir el vector

$$A_0 = \begin{bmatrix} 1 & 4 & 9 & 7 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & 2 & 5 \\ 0 & 0 & 0 & -1 \end{bmatrix}$$

como combinación lineal de los vectores de la base encontrada.

(24) Sea V un $\mathbb{K}$-espacio vectorial y sean $u, v \in V$ dos vectores no nulos. Demostrar que $\{u, v\}$ es linealmente dependiente si y sólo si existe $\lambda \in \mathbb{K} - \{0\}$ tal que $u = \lambda v$.

(25) ¿Es el conjunto $\{x^2,\ x|x|\}$ linealmente independiente en el espacio vectorial de las funciones reales continuas definidas en el intervalo $[-1, 1]$? Justificar la respuesta. Interpretar el resultado en una gráfica. ¿Se obtiene el mismo resultado si se analizan en el intervalo $[-1, 0]$? ¿Se puede justificar esta pregunta mediante el ejercicio anterior?

(26) Probar que el conjunto $\{\cos(x), \operatorname{sen}(x)\}$ es linealmente independiente en el espacio vectorial de las funciones derivables de $\mathbb{R}$ en $\mathbb{R}$.

(27) Sea $\mathcal{C}^{(n-1)}_{[a,b]}$ el conjunto de funciones reales que poseen $n-1$ derivadas continuas en el intervalo $[a, b]$. Si $\{f_1(x), f_2(x), \ldots, f_n(x)\}$ es un subconjunto

de $\mathcal{C}^{(n-1)}_{[a,b]}$ y existe un punto $x_0 \in [a,b]$ tal que[20]

$$\det\left(\begin{bmatrix} f_1(x_0) & f_2(x_0) & \dots & f_n(x_0) \\ f_1'(x_0) & f_2'(x_0) & \dots & f_n'(x_0) \\ \vdots & \vdots & & \vdots \\ f_1^{(n-1)}(x_0) & f_2^{(n-1)}(x_0) & \dots & f_n^{(n-1)}(x_0) \end{bmatrix}\right) \neq 0$$

entonces $\{f_1(x), f_2(x), \dots, f_n(x)\}$ es linealmente independiente en $\mathcal{C}^{(n-1)}_{[a,b]}$. Deducir de este resultado, si es posible, los dos Ejercicios anteriores. (Ayuda: Plantear la combinación lineal nula de las funciones dadas y derivar $n-1$ veces. Luego, analizar el sistema lineal resultante).

(28) Usando como argumento su dimensión, demostrar que los subespacios vectoriales:

(a) de $\mathbb{R}^1$ son sólo los subespacios triviales.

(b) no triviales de $\mathbb{R}^2$ son las rectas que pasan por $(0,0)$.

(c) no triviales de $\mathbb{R}^3$ son las rectas que pasan por $(0,0,0)$ y los planos que pasan por $(0,0,0)$.

(29) ¿Para qué valores de $a \in \mathbb{R}$ es linealmente dependiente el conjunto $S = \{1 + ax, a + (a+2)x\} \subseteq \mathbb{R}_2[x]$? Justificar la respuesta.

(30) ¿Bajo qué restricciones sobre los valores de a y b en $\mathbb{R}$ puede garantizarse que el conjunto $S = \{a + ax + ax^2, bx^2, 1\}$ genera el espacio vectorial $\mathbb{R}_2[x]$? Justificar la respuesta.

(31) Hallar el valor de $\alpha \in \mathbb{R}$ para que existan matrices no nulas $B \in \mathbb{R}^{2\times 2}$ tales que

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

[20]El determinante considerado se llama **wronskiano** de $f_1(x), f_2(x), \dots, f_n(x)$. Se utiliza para resolver ecuaciones diferenciales lineales y fue utilizado en 1812 por el matemático polaco Józef Hoene-Wroński (1776-1853) llamado de este modo en 1882 por el matemático escocés Thomas Muir (1844-1934).

Para el valor de α encontrado probar que el conjunto

$$S = \left\{ B \in \mathbb{R}^{2\times 2} : \begin{bmatrix} 1 & 2 \\ 3 & \alpha \end{bmatrix} B = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \right\}$$

es un subespacio vectorial de $\mathbb{R}^{2\times 2}$. Hallar una base de S y su dimensión.

(32) Sea $\{A_1, A_2, \ldots, A_p\}$ un conjunto linealmente independiente de matrices de tamaño $m \times n$ a coeficientes reales y sean B y C dos matrices invertibles reales de tamaños $m \times m$ y $n \times n$, respectivamente. Demostrar que

$$\{BA_1C, BA_2C, \ldots, BA_pC\}$$

es un conjunto linealmente independiente en $\mathbb{R}^{m\times n}$. ¿Se debe cumplir alguna relación entre m, n y p?

(33) Sea V un $\mathbb{K}$-espacio vectorial y sean $u_1, u_2, \ldots, u_n \in V$. Demostrar que la suma de dos combinaciones lineales de $u_1, u_2, \ldots, u_n$ es otra combinación lineal de $u_1, u_2, \ldots, u_n$ y el producto de un escalar $\lambda \in \mathbb{K}$ por una combinación lineal de $u_1, u_2, \ldots, u_n$ es otra combinación lineal de $u_1, u_2, \ldots, u_n$.

(34) Sea V un $\mathbb{K}$-espacio vectorial, $S \leq V$ y $\{u_1, u_2, \ldots, u_n\} \subseteq S$. Probar que $\overline{\{u_1, u_2, \ldots, u_n\}} \subseteq S$.

(35) Se considera una famila de subespacios $\{S_i\}_{i\in I}$ de un $\mathbb{K}$-espacio vectorial V. Demostrar que la intersección $\cap_{i\in I} S_i$ es un subespacio de V.

(36) Dar un ejemplo de dos subespacios de $\mathbb{R}^2$ tales que $S_1 \cup S_2$ no sea un subespacio vectorial de $\mathbb{R}^2$.

(37) Sea V un $\mathbb{K}$-espacio vectorial y sean $S_1, S_2 \leq V$. Demostrar que

$$S_1 \cup S_2 \leq V \qquad \Longleftrightarrow \qquad S_1 \subseteq S_2 \;\text{ ó }\; S_2 \subseteq S_1.$$

(38) Hallar el valor de $k \in \mathbb{R}$ para que el conjunto

$$S_k = \{ax + b \in \mathbb{R}_1[x] : a + b = k\}$$

sea un subespacio vectorial de $\mathbb{R}_1[x]$. Para dicho valor de k, determinar la dimensión y una base de S_k.

(39) Si V es un $\mathbb{K}$-espacio vectorial. Demostrar que $S \leq V \Leftrightarrow S = \overline{S}$.

(40) Comprobar que el cuerpo (finito) $\mathbb{Z}_2$ definido en los Ejemplos 1.18 y 1.39 puede considerarse como $\mathbb{Z}_2$-espacio vectorial. Hallar una base del mismo.

(41) Sea V un $\mathbb{K}$-espacio vectorial y $\{u_1, \ldots, u_n\} \subseteq V$. Demostrar que:

(I) $\overline{\{u_1, \ldots, u_i, \ldots, u_j, \ldots, u_n\}} = \overline{\{u_1, \ldots, u_j, \ldots, u_i, \ldots, u_n\}}$.

(II) $\overline{\{u_1, \ldots, u_i, \ldots, u_n\}} = \overline{\{u_1, \ldots, cu_i, \ldots, u_n\}}$, para todo $c \in \mathbb{K} - \{0\}$.

(III) $\overline{\{u_1, \ldots, u_i, \ldots, u_j, \ldots, u_n\}} = \overline{\{u_1, \ldots, u_i, \ldots, u_j + cu_i, \ldots, u_n\}}$ para todo $c \in \mathbb{K}$ y $j \neq i$.

Deducir que las operaciones elementales de vectores preservan las combinaciones lineales. Comparar este resultado con el obtenido en la Proposición 8.7.

(42) Sea V un $\mathbb{C}$-espacio vectorial finitamente generado. Demostrar que V considerado como $\mathbb{R}$-espacio vectorial también es finitamente generado.

(43) Probar que si el sistema lineal formado por las ecuaciones $ax + by = 0$ y $cx + dy = 0$ (con $a, b, c, d \in \mathbb{K}$) tiene solución no trivial entonces los vectores: (a) (a, c) y (b, d) de $\mathbb{K}^2$ son linealmente dependientes; (b) (a, b) y (c, d) de $\mathbb{K}^2$ son linealmente dependientes.

(44) Encontrar una matriz $A = \begin{bmatrix} a_1 & a_2 & a_3 & a_4 \end{bmatrix}$ particionada según sus columnas para la cual

$$a_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \qquad a_3 = \begin{bmatrix} -1 \\ 0 \\ 4 \end{bmatrix}$$

y la forma escalonada reducida por filas de A es

$$R_A = \begin{bmatrix} 1 & -6 & 0 & -2 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

(45) Probar que la familia

$$\mathcal{C} = \left\{ P_i(x) = x^i \right\}_{i=0}^{n}$$

es un conjunto linealmente independiente del $\mathbb{R}$-espacio vectorial $\mathbb{R}_n[x]$. Para ello, escribir al vector nulo como una combinación lineal de $\mathcal{C}$ y derivarlo n veces. Comparar con el Ejemplo 8.56.

(46) Probar que los subespacios $S_1 = \overline{\{(1,0)\}}$ y $S_2 = \overline{\{(2,3)\}}$ satisfacen $S_1 \oplus S_2 = \mathbb{R}^2$. Nótese que, en definitiva, se trata de resolver sistemas de ecuaciones lineales.

(47) Se utilizará la notación $2\mathbb{N} = \{2, 4, 6, \dots\}$ para indicar el conjunto de los números naturales pares. Se considera el espacio vectorial $\mathbb{R}[x]$ de las funciones polinomiales en la variable x. Sean

$$S_1 = \overline{\{1, x^2, x^4, \dots\}} = \overline{\{x^j\}_{j \in \{0\} \cup 2\mathbb{N}}}$$

y

$$S_2 = \overline{\{x, x^3, x^5, \dots\}} = \overline{\{x^j\}_{j \in \mathbb{N} - 2\mathbb{N}}}.$$

Probar que $\mathbb{R}[x] = S_1 \oplus S_2$.

(48) Demostrar que las funciones

$$f(x) = \begin{cases} 1 & x \in [0, \frac{1}{3}] \\ 0 & x \in]\frac{1}{3}, 1] \end{cases}, \quad g(x) = \begin{cases} 1 & x \in [0, \frac{2}{3}] \\ 0 & x \in]\frac{2}{3}, 1] \end{cases} \quad \text{y} \quad h(x) = 1 - x,$$

son linealmente independientes en el espacio de todas las funciones reales definidas sobre el intervalo $[0, 1]$. Indicar si la función k definida por

$$k(x) = \begin{cases} 2 & x \in [0, \frac{1}{3}] \\ 1 & x \in]\frac{1}{3}, \frac{2}{3}] \\ 0 & x \in]\frac{2}{3}, 1] \end{cases}$$

cumple que $k \in \overline{\{f, g\}}$. Representar gráficamente.

(49) Demostrar que las funciones continuas e^x, $\operatorname{sen}(x)$ y $\cos(x)$ no pertenecen a $\overline{\{1, x, x^2, x^3, \dots\}}$. En consecuencia, estas funciones trascendentes familiares no pueden ser expresadas como combinaciones lineales (finitas) de polinomios.

(50) Siguiendo la idea de la demostración de la Proposición 8.8, enunciar y probar una proposición similar que establezca que la equivalencia por columnas preserva las relaciones de dependencia lineal entre las filas de una matriz.

(51) Sea V un espacio vectorial. Es conocido que al unir un sistema de generadores de un subespacio S_1 de V con un sistema de generadores de un subespacio S_2 de V, se obtiene un sistema generador de $S_1 + S_2$. Para ver que este hecho no se cumple para bases, encontrar dos subespacios S_1 y S_2 del $\mathbb{R}$-espacio vectorial $\mathbb{R}^3$ tales que al unir una base de S_1 con una de S_2 se obtiene un sistema linealmente dependiente de $S_1 + S_2$. ¿Contradice este hecho el apartado (c) de la Proposición 8.18? Justificar.

Capítulo 9

Coordenadas en espacios vectoriales

Índice

9.1. Introducción

Al comenzar el capítulo anterior se introdujo el modelo geométrico de los vectores libres $\mathbb{VL}^2$ y su correspondiente modelo algebraico $\mathbb{R}^2$. Con esta representación algebraica del modelo de un plano (físico), se requiere que se fijen dos vectores libres en $\mathbb{VL}^2$ para establecer tal correspondencia. Más precisamente, fijado un par de vectores libres $\vec{e}_1$ y $\vec{e}_2$ no nulos y con distinta dirección, a cada vector $\vec{u} \in \mathbb{VL}^2$ se le asocia los escalares reales x e y que permiten escribir de forma única a $\vec{u}$ como combinación lineal de ellos, es decir, $\vec{u} = x\vec{e}_1 + y\vec{e}_2$.

Se indicó que del mismo modo se procede con el modelo geométrico $\mathbb{VL}^3$ (del espacio físico) y el modelo algebraico $\mathbb{R}^3$.

Luego, se desarrolló el tema de espacios vectoriales abstractos haciendo especial hincapié en los de dimensión finita.

El objetivo de este capítulo es establecer que la correspondencia entre un espacio vectorial abstracto de dimensión finita y el modelo algebraico $\mathbb{R}^n$ (para un $n \in \mathbb{N}$ adecuado) se puede extender al caso general y se mostrará cómo trabajar con ellos. De este modo, estos modelos permiten introducir coordenadas para operar en espacios vectoriales. Este hecho permitió establecer el gran salto que dieron las Matemáticas al pasar de los métodos geométricos de Euclides a considerar la idea de coordenadas introducida por René Descartes (1596-1650) y Pierre de Fermat (1601-1665). A partir del uso de coordenadas será posible pensar a cualquier $\mathbb{K}$-espacio vectorial de dimensión finita $n \geq 1$ como el propio $\mathbb{K}$-espacio vectorial $\mathbb{K}^n$ (el espacio cartesiano), que es el ejemplo más sencillo de entre todos los espacios vectoriales de dimensión finita.

Una anécdota curiosa es que, al parecer, Descartes se encontraba con gripe en la cama, cuando se sintió molesto por una incordiosa mosca que de repente se posó en el techo de su dormitorio. Gracias a ese episodio se le ocurrió que calculando la distancia a las dos paredes adyacentes, se podría establecer la posición de la mosca utilizando números. Según cuenta la historia, en la misma época, Pierre de Fermat estaba trabajando con una idea similar

(a) René Descartes

(b) Pierre de Fermat

Figura 9.1: René Descartes y Pierre de Fermat.

aunque no publicó sus resultados y fue quien propuso ampliar las coordenadas a la tercera dimensión. Esta anécdota acerca de Descartes puede haber sido cierta o no, lo que está claro es que ha revolucionado la Geometría de tal modo que pasó a llamarse **Geometría Sintética** a la estudiada mediante los métodos de la época de Euclides, para llamarse **Geometría Analítica** a la abordada mediante métodos que se basan en coordenadas. Disponer de coordenadas permitió expresar ecuaciones de curvas y superficies, por ejemplo, mediante fórmulas donde se relacionan dichas coordenadas, lo que impulsó el desarrollo del Análisis Matemático y su conexión con el Álgebra, en especial, con el Álgebra Lineal.

En la Figura 9.2 se aprecia a Descartes explicando su método a la reina Cristina de Suecia.

La *Geometría* (*Sintética*[1]) tuvo su origen con los *Elementos*, que es una colección de trece libros escrita en torno al año 300 a.C. por el matemático griego Euclides, quien desarrolló su magnífica labor en Alejandría (Egipto). La obra cubre tanto geometría elemental (geometría plana y geometría de los

[1]Llamada así en oposición a la posterior Geometría Analítica, debido a la dualidad análisis-síntesis.

Figura 9.2: Descartes en la Corte de la reina Cristina de Suecia (detalle), Louis-Michel Dumesnil, Museo Nacional del Palacio de Versalles.

cuerpos sólidos) como aritmética (razones, proporciones y teoría de números).

La importancia de la Geometría (por su estrecha relación con la Filosofía) en la época de la Antigua Grecia era tal, que en la Academia de Platón, delante del templo de las Musas, estaba escrito:

> "No entre nadie que no conozca la Geometría".

En el siglo XVII, a partir del trabajo de René Descartes (Discurso del método, 1637) y Pierre de Fermat, se produjo una revolución científica al poder conectar el Álgebra con la Geometría (mediante la introducción de los sistemas de coordenadas), lo que originó lo que el propio Descartes llamó *Geometría Analítica*, que hoy se conoce como *Geometría Cartesiana*[2]. Informalmente, la idea básica fue asociar una ecuación de dos incógnitas con una *curva* (figura geométrica) a partir del conjunto de todas sus soluciones

[2]En la actualidad, el nombre *Geometría Analítica* comprende no sólo la Geometría Cartesiana sino también la Geometría Diferencial (iniciada con Gauss) y la más reciente Geometría Algebraica.

y del establecimiento de un sistema de coordenadas en el plano. Además, esta misma idea permitía que dos variables se pudieran relacionar a través de una *función*, que se representaría gráficamente como una curva, donde el concepto de *variable* permitía tratar cantidades que no eran fijas. De este modo, sería ahora posible resolver el (antiguo) problema geométrico de la intersección de curvas mediante el problema algebraico de la resolución de sistemas de ecuaciones. Para indicar que se resuelve un problema de un área mediante técnicas de la otra se dice que *se algebriza la Geometría o bien que se geometriza el Álgebra*.

9.2. Coordenadas de un vector respecto de una base

El siguiente resultado permite caracterizar el concepto de base de un espacio vectorial poniendo el énfasis en la existencia y unicidad de los coeficientes de las combinaciones lineales.

Proposición 9.1 (Representación única en una base)**.** Sea V un $\mathbb{K}$-espacio vectorial y $\{u_1, u_2, \dots, u_n\} \subseteq V$. Las siguientes condiciones son equivalentes:

(a) $\{u_1, u_2, \dots, u_n\}$ es una base de V.

(b) Se cumplen las dos condiciones:

(I) Todo vector $u \in V$ se escribe como combinación lineal de los vectores de $\{u_1, u_2, \dots, u_n\}$.

(II) Si $\alpha_1 u_1 + \alpha_2 u_2 + \cdots + \alpha_n u_n = \beta_1 u_1 + \beta_2 u_2 + \cdots + \beta_n u_n$ con $\alpha_i, \beta_i \in \mathbb{K}$, $i = 1, 2, \dots, n$, entonces $\alpha_i = \beta_i$, $\forall i = 1, 2, \dots, n$.

Demostración. (a) $\Rightarrow$ (b) Al ser $\{u_1, u_2, \dots, u_n\}$ un sistema de generadores de V, el apartado (I) se deduce inmediatamnete.

Para probar el apartado (II), se supone que

$$\alpha_1 u_1 + \alpha_2 u_2 + \cdots + \alpha_n u_n = \beta_1 u_1 + \beta_2 u_2 + \cdots + \beta_n u_n,$$

con $\alpha_i, \beta_i \in \mathbb{K}$, $i = 1, 2, \ldots, n$. Entonces

$$(\alpha_1 - \beta_1)u_1 + (\alpha_2 - \beta_2)u_2 + \cdots + (\alpha_n - \beta_n)u_n = 0.$$

Por ser $\{u_1, u_2, \ldots, u_n\}$ un conjunto linealmente independiente se tiene que $\alpha_1 - \beta_1 = \alpha_2 - \beta_2 = \cdots = \alpha_n - \beta_n = 0$. Luego, $\alpha_i = \beta_i$, $\forall i = 1, 2, \ldots, n$.

(b) $\Rightarrow$ (a) Del apartado (I) se tiene que

$$\overline{\{u_1, u_2, \ldots, u_n\}} = V.$$

Falta probar que $\{u_1, u_2, \ldots, u_n\}$ es linealmente independiente. En efecto, sea $\alpha_1 u_1 + \alpha_2 u_2 + \cdots + \alpha_n u_n = 0$ con $\alpha_1, \alpha_2, \ldots, \alpha_n \in \mathbb{K}$. Como siempre es posible escribir $0u_1 + 0u_2 + \cdots + 0u_n = 0$, de la hipótesis se tiene que $\alpha_i = 0$, para todo $i = 1, 2, \ldots, n$, lo que prueba dicha independencia lineal. $\square$

Al escribir "el conjunto de vectores $\{u_1, u_2, \ldots, u_n\}$" se entiende que dicho conjunto está formado por elementos diferentes; sin embargo, no se hace alusión alguna al orden en que dichos vectores están situados. Por otro lado, una sucesión de vectores[3] indica que los elementos están dispuestos en cierto orden pero no se dice nada acerca de la igualdad o no de sus elementos, de hecho podría haber repeticiones. Al tratar con coordenadas será necesario tener en cuenta ambos hechos de forma simultánea.

Al hablar de una base $\mathcal{B} = \{u_1, u_2, \ldots, u_n\}$ de un $\mathbb{K}$-espacio vectorial V de dimensión $n \geq 1$ se deberá respetar el hecho que los vectores de dicha base estén dispuestos en ese orden. Se suele hablar entonces de *base ordenada*, para remarcar la condición sobre el orden que se le requiere a los elementos del conjunto $\mathcal{B}$. Se observa que en el caso de la base canónica $\{e_1, e_2, \ldots, e_n\}$ de $\mathbb{K}^n$ es inmediato indicar cuál es su i-ésimo vector, pues se tiene un orden

[3] Que también puede verse como una familia indexada, según lo indicado en la página 71.

natural; sin embargo, en el caso de una base arbitraria, este orden debe ser fijado de antemano. Con más precisión, es posible establecer la siguiente definición.

Definición 9.1. Sea V un $\mathbb{K}$-espacio vectorial de dimensión $n \geq 1$. Se llama **base ordenada** de V a toda sucesión $u_1, u_2, \ldots, u_n$ de vectores linealmente independientes que generan V.

De ahora en adelante, si bien la mayoría de las veces no se indicará de forma explícita el adjetivo "ordenada" de una base, se la presupondrá ordenada.

Definición 9.2. Sea V un $\mathbb{K}$-espacio vectorial de dimensión $n \geq 1$ y sea $\mathcal{B} = \{u_1, u_2, \ldots, u_n\}$ una base ordenada de V. Dado $u \in V$, los escalares $\alpha_1, \alpha_2, \ldots, \alpha_n \in \mathbb{K}$, unívocamente determinados, tales que

$$u = \alpha_1 u_1 + \alpha_2 u_2 + \cdots + \alpha_n u_n$$

se llaman las **coordenadas de u en la base $\mathcal{B}$**, y se denota

$$[u]_{\mathcal{B}} = (\alpha_1, \alpha_2, \ldots, \alpha_n).$$

Es claro que $[u]_{\mathcal{B}} = (\alpha_1, \alpha_2, \ldots, \alpha_n) \in \mathbb{K}^n$.

Observación 9.1. En el espacio vectorial $\mathbb{K}^n$, de la igualdad

$$x = (x_1, x_2, \ldots, x_n) = x_1 e_1 + x_2 e_2 + \cdots + x_n e_n,$$

se tiene que las **componentes** de una n-tupla ordenada coindicen con las **coordenadas** del vector en la base canónica $\{e_1, e_2, \ldots, e_n\}$.

Ejemplo 9.1. *Hallar las coordenadas del vector $u = (3, -4, 12)$ en la base $\{u_1 = (1, 0, 2), u_2 = (2, 2, 0), u_3 = (1, -1, 3)\}$ de $\mathbb{R}^3$.*

◁ Se debe escribir u como combinación lineal de u_1, u_2, u_3 (en ese orden), es decir se deben buscar $\alpha_1, \alpha_2, \alpha_3 \in \mathbb{R}$ tales que

$$(3, -4, 12) = \alpha_1(1, 0, 2) + \alpha_2(2, 2, 0) + \alpha_3(1, -1, 3).$$

Tras operar y resolver el sistema de ecuaciones lineales resultante

$$\left\{\begin{array}{rcrcrcr} \alpha_1 & + & 2\alpha_2 & + & \alpha_3 & = & 3 \\ & & 2\alpha_2 & - & \alpha_3 & = & -4 \\ 2\alpha_1 & & & + & 3\alpha_3 & = & 12 \end{array}\right.$$

se obtiene $\alpha_1 = 3$, $\alpha_2 = -1$ y $\alpha_3 = 2$. Las coordenadas de u en la base $\mathcal{B}$ son los escalares encontrados, en ese orden. Es decir,

$$[u]_{\mathcal{B}} = (3, -1, 2).$$

▷

Ejemplo 9.2. *Hallar las coordenadas del vector* $P(x) = (x-1)^2 + 3$ *en la base*

$$\mathcal{C} = \{P_1(x) = 1, P_2(x) = x, P_3(x) = x^2\}$$

de $\mathbb{R}_2[x]$*. Repetir con el mismo vector* P *en la base*

$$\mathcal{B} = \{P_1(x) = 1, P_2(x) = x - 1, P_3(x) = (x-1)^2\}$$

de $\mathbb{R}_2[x]$*.*

◁ Se debe escribir P como combinación lineal de P_1, P_2, P_3 (en ese orden). En este caso, desarrollando el cuadrado se obtiene

$$P(x) = (x-1)^2 + 3 = x^2 - 2x + 4,$$

y teniendo en cuenta el orden de los vectores de $\mathcal{C}$, las coordenadas de P en la base $\mathcal{B}$ de $\mathbb{R}_2[x]$ son $4, -2, 1$. Luego,

$$[P]_{\mathcal{C}} = (4, -2, 1).$$

Dado que $P(x) = 3 + 0(x-1) + (x-1)^2$, las coordenadas de P en la base $\mathcal{B}$ son inmediatas:

$$[P]_{\mathcal{B}} = (3, 0, 1).$$

$\triangleright$

9.3. Isomorfismo de Descartes

Definición 9.3 (Espacios vectoriales isomorfos)**.** Dos espacios vectoriales $(V_1, \oplus, \odot)$ y $(V_2, +, \cdot)$ sobre el mismo cuerpo $\mathbb{K}$ se dicen **isomorfos**, y se denota $V_1 \cong V_2$, si existe una aplicación biyectiva

$$f : V_1 \to V_2$$

que respeta las operaciones de los espacios, es decir,

$$\begin{aligned} f(u_1 \oplus u_2) &= f(u_1) + f(u_2), & \forall u_1, u_2 \in V_1, \\ f(\lambda \odot u) &= \lambda \cdot f(u), & \forall u \in V_1, \forall \lambda \in \mathbb{K}. \end{aligned}$$

Se dice que la aplicación f es un **isomorfismo** entre los espacios vectoriales V_1 y V_2.

Además de establecer una correspondencia biunívoca entre los elementos de V_1 y V_2, un isomorfirmo respeta la estructura algebraica de los espacios vectoriales involucrados, en el sentido que es lo mismo operar en el espacio V_1 (con su adición o su multiplicación por escalares) y calcular la imagen en V_2 de el/los vector/es obtenido/s, que calcular primero la/las imagen/es en V_2 de el/los vector/es de V_1 y luego operar en V_2 (con su adición o su multiplicación por escalares).

Ejemplo 9.3. *Establecer un isomorfismo entre los espacios vectoriales* $(\mathbb{R}^2, \oplus, \odot)$ *y* $(\mathbb{R}_1[x], +, \cdot)$*, donde las operaciones son la adición y la multiplicación por escalares habituales en ambos espacios.*

$\triangleleft$ Se debe definir una aplicación $f : \mathbb{R}^2 \to \mathbb{R}_1[x]$ biyectiva que, además, respete las operaciones. Para cada $(a, b) \in \mathbb{R}^2$ se define[4]

$$f((a, b)) = a + bx.$$

Es claro que f está bien definida (pues f se aplica en un vector de $\mathbb{R}^2$ y su imagen es un polinomio de $\mathbb{R}_1[x]$).

Para probar que f es inyectiva, se consideran $(a_1, b_1), (a_2, b_2) \in \mathbb{R}^2$ tales que $f((a_1, b_1)) = f((a_2, b_2))$. Luego, $a_1 + b_1 x = a_2 + b_2 x$, de donde se tiene que $a_1 = a_2$ y $b_1 = b_2$. Luego $(a_1, b_1) = (a_2, b_2)$, es decir, f es inyectiva.

Para probar que f es sobreyectiva, se considera un elemento $p(x) \in \mathbb{R}_1[x]$, es decir, $p(x) = a + bx$ para algunos escalares $a, b \in \mathbb{R}$. Tomando $(a, b) \in \mathbb{R}^2$ se tiene que $f((a, b)) = a + bx = p(x)$, con lo cual f es sobreyectiva.

Para probar que f respeta las operaciones de los espacios vectoriales se consideran $(a_1, b_1), (a_2, b_2) \in \mathbb{R}^2$ y $\lambda \in \mathbb{R}$. Se cumple que

$$\begin{aligned} f((a_1, b_1) \oplus (a_2, b_2)) &= f((a_1 + a_2, b_1 + b_2)) \\ &= (a_1 + a_2) + (b_1 + b_2)x \\ &= (a_1 + b_1 x) + (a_2 + b_2 x) \\ &= f((a_1, b_1)) + f((a_2, b_2)) \end{aligned}$$

y

$$\begin{aligned} f(\lambda \odot (a_1, b_1)) &= f((\lambda a_1, \lambda b_1)) \\ &= (\lambda a_1) + (\lambda b_1)x \\ &= \lambda \cdot (a_1 + b_1 x) \\ &= \lambda \cdot f((a_1, b_1)). \end{aligned}$$

Luego, f es un isomorfismo de espacios vectoriales entre $\mathbb{R}^2$ y $\mathbb{R}_1[x]$, es decir $\mathbb{R}^2 \cong \mathbb{R}_1[x]$. $\triangleright$

El próximo resultado establece que todo $\mathbb{K}$-espacio vectorial finito-dimensional V con $\dim_{\mathbb{K}}(V) = n \geq 1$ es isomorfo al $\mathbb{K}$-espacio vectorial $\mathbb{K}^n$ algebrizado con la adición y la multiplicación por escalares habituales. En lugar

[4]Implícitamente, se están pensando los vectores de $\mathbb{R}^2$ escritos en su base canónica.

de operar con elementos en V, este resultado permitirá operar directamente con sus coordenadas en $\mathbb{K}^n$.

Teorema 9.1 (Isomorfismo de Descartes para espacios vectoriales)**.** Sea V un $\mathbb{K}$-espacio vectorial finito-dimensional con $\dim_{\mathbb{K}}(V) = n \geq 1$. Para cada base $\mathcal{B}$ de V, existe una aplicación biyectiva

$$\mathcal{D}_{\mathcal{B}} : V \to \mathbb{K}^n$$

que satisface[a] $\forall v_1, v_2 \in V, \forall \lambda \in \mathbb{K}$:

$$\begin{array}{rcl} \mathcal{D}_{\mathcal{B}}(v_1 + v_2) & = & \mathcal{D}_{\mathcal{B}}(v_1) + \mathcal{D}_{\mathcal{B}}(v_2) \\ \mathcal{D}_{\mathcal{B}}(\lambda v_1) & = & \lambda \mathcal{D}_{\mathcal{B}}(v_1). \end{array} \tag{9.1}$$

En consecuencia, V y $\mathbb{K}^n$ son espacios vectoriales isomorfos sobre $\mathbb{K}$, es decir

$$V \cong \mathbb{K}^n.$$

La aplicación $\mathcal{D}_{\mathcal{B}}$ se llama **isomorfismo de Descartes**.

[a]Una aplicación definida entre espacios vectoriales sobre un mismo cuerpo y que respeta la adición y la multiplicación por escalares, se llama **aplicación lineal**.

Demostración. Si bien se ha hecho mención exclusiva al (soporte) $\mathbb{K}$-espacio vectorial $\mathbb{K}^n$, en realidad, también se deberá tener en cuenta que la adición y la multiplicación por escalares en $\mathbb{K}^n$ son las habituales. Se utilizará el mismo símbolo para las operaciones en V y en $\mathbb{K}^n$.

Sea $\mathcal{B} = \{u_1, u_2, \ldots, u_n\}$ una base de V. Su existencia está garantizada por el Teorema 8.2. Sea $u \in V$. Existen escalares $\alpha_1, \alpha_2, \ldots, \alpha_n \in \mathbb{K}$, unívocamente determinados, tales que $u = \alpha_1 u_1 + \alpha_2 u_2 + \cdots + \alpha_n u_n$. Luego, las coordenadas de u en la base $\mathcal{B}$ son $[u]_{\mathcal{B}} = (\alpha_1, \alpha_2, \ldots, \alpha_n)$.

Se considera la aplicación

$$\begin{array}{rcccc} \mathcal{D}_{\mathcal{B}} & : & V & \to & \mathbb{K}^n \\ & & u & \mapsto & \mathcal{D}_{\mathcal{B}}(u) = [u]_{\mathcal{B}} = (\alpha_1, \alpha_2, \ldots, \alpha_n). \end{array}$$

Para evitar demasiados subíndices se indicará $\mathcal{D}_{\mathcal{B}}$ únicamente mediante $\mathcal{D}$. Se cumple que:

- $\mathcal{D}$ es inyectiva: Sean $u, v \in V$ tales que $\mathcal{D}(u) = \mathcal{D}(v)$. Escribiendo a u y v en la base $\mathcal{B}$ como

  $$u = \alpha_1 u_1 + \alpha_2 u_2 + \cdots + \alpha_n u_n \quad \text{y} \quad v = \beta_1 u_1 + \beta_2 u_2 + \cdots + \beta_n u_n,$$

  resulta que $[u]_{\mathcal{B}} = [v]_{\mathcal{B}}$, es decir $(\alpha_1, \alpha_2, \ldots, \alpha_n) = (\beta_1, \beta_2, \ldots, \beta_n)$. Luego, $\alpha_j = \beta_j$ para $j = 1, 2, \ldots, n$. Por lo tanto, $u = \alpha_1 u_1 + \alpha_2 u_2 + \cdots + \alpha_n u_n = v$.

- $\mathcal{D}$ es sobreyectiva. Dado $(\alpha_1, \alpha_2, \ldots, \alpha_n) \in \mathbb{K}^n$, es claro que el vector $u := \alpha_1 u_1 + \alpha_2 u_2 + \cdots + \alpha_n u_n \in V$ es tal que $\mathcal{D}(u) = [u]_{\mathcal{B}} = (\alpha_1, \alpha_2, \ldots, \alpha_n)$ debido a la unicidad de la representación como combinación lineal de los elementos de $\mathcal{B}$.

- $\mathcal{D}$ respeta la adición y la multiplicación por escalares: Si $u, v \in V$ están escritos en la base $\mathcal{B}$ como $u = \alpha_1 u_1 + \alpha_2 u_2 + \cdots + \alpha_n u_n$ y $v = \beta_1 u_1 + \beta_2 u_2 + \cdots + \beta_n u_n$ entonces, aplicando propiedades de espacio vectorial, se tiene

  $$\begin{aligned} u + v &= (\alpha_1 u_1 + \alpha_2 u_2 + \cdots + \alpha_n u_n) + (\beta_1 u_1 + \beta_2 u_2 + \cdots + \beta_n u_n) \\ &= (\alpha_1 + \beta_1) u_1 + (\alpha_2 + \beta_2) u_2 + \cdots + (\alpha_n + \beta_n) u_n. \end{aligned}$$

  Luego, por la unicidad de la representación como combinación lineal de los elementos de $\mathcal{B}$ y la definición de la adición en $\mathbb{K}^n$, se tiene

  $$\begin{aligned} \mathcal{D}(u + v) &= [u + v]_{\mathcal{B}} \\ &= (\alpha_1 + \beta_1, \alpha_2 + \beta_2, \ldots, \alpha_n + \beta_n) \\ &= (\alpha_1, \alpha_2, \ldots, \alpha_n) + (\beta_1, \beta_2, \ldots, \beta_n) \\ &= [u]_{\mathcal{B}} + [v]_{\mathcal{B}} \\ &= \mathcal{D}(u) + \mathcal{D}(v). \end{aligned}$$

Si $\lambda \in \mathbb{K}$ entonces, aplicando propiedades de espacio vectorial, se tiene

$$\begin{aligned} \lambda u &= \lambda(\alpha_1 u_1 + \alpha_2 u_2 + \cdots + \alpha_n u_n) \\ &= (\lambda\alpha_1)u_1 + (\lambda\alpha_2)u_2 + \cdots + (\lambda\alpha_n)u_n. \end{aligned}$$

De nuevo, por la unicidad de la representación y la definición de la multiplicación por escalares en $\mathbb{K}^n$,

$$\begin{aligned} \mathcal{D}(\lambda u) &= [\lambda u]_{\mathcal{B}} \\ &= (\lambda\alpha_1, \lambda\alpha_2, \ldots, \lambda\alpha_n) \\ &= \lambda(\alpha_1, \alpha_2, \ldots, \alpha_n) \\ &= \lambda[u]_{\mathcal{B}} \\ &= \lambda\mathcal{D}(u). \end{aligned}$$

□

Observación 9.2. El resultado del Teorema 9.1 permite llamar **espacio cartesiano** al $\mathbb{K}$-espacio vectorial $\mathbb{K}^n$ (de hecho se ha constituido como $\mathbb{K}^n = \mathbb{K} \times \mathbb{K} \times \cdots \times \mathbb{K}$, es decir, a partir del producto cartesiano de n veces por sí mismo el conjunto $\mathbb{K}$).

Observación 9.3. Las propiedades (9.1) escritas en notación de coordenadas quedan

$$\begin{aligned} [u_1 + u_2]_{\mathcal{B}} &= [u_1]_{\mathcal{B}} + [u_2]_{\mathcal{B}}, && \forall u_1, u_2 \in V, \\ [\lambda u]_{\mathcal{B}} &= \lambda[u]_{\mathcal{B}}, && \forall u \in V, \forall \lambda \in \mathbb{K}. \end{aligned}$$

La importancia de estas igualdades radica en la forma en que se puede operar con vectores en un cierto $\mathbb{K}$-espacio vectorial V con $\dim_{\mathbb{K}}(V) = n \geq 1$. Si $\mathcal{B} = \{u_1, u_2, \ldots, u_n\}$ es una base de V, basta tomar las coordenadas de cada vector en la base $\mathcal{B}$ y sumar o multiplicar por escalares dichos vectores coordenados, teniendo en cuenta que son vectores del espacio cartesiano $\mathbb{K}^n$.

Ejemplo 9.4. *Se considera el $\mathbb{R}$-espacio vectorial*

$$\mathbb{R}_2[x] = \{P \in \mathbb{R}[x] : \; P(x) = 0 \vee \mathrm{gr}(P(x)) \leq 2\}$$

y la base $\mathcal{B} = \{1, x-5, (x-5)^2\}$. Se pide:

(a) Escribir un vector arbitrario $P(x) = a_0 + a_1x + a_2x^2 \in \mathbb{R}_2[x]$ en la base $\mathcal{B}$.

(b) Calcular $2Q(x)+R(x)$ para los vectores $Q(x) = -6x+x^2$ y $R(x) = 4 + x - x^2$ sumando las correspondientes coordenadas en $\mathbb{R}^3$.

$\lhd$ (a) Una forma de escribir el vector $P(x) = a_0 + a_1x + a_2x^2 \in \mathbb{R}_2[x]$ en la base $\mathcal{B}$ es expresando previamente cada uno de los vectores 1, x y x^2 (que generan el polinomio $P(x)$) en dicha base (es decir, escribiéndolos como pontencias de $x - 5$). Observando que

$$x = x - 5 + 5 = (x-5) + 5$$

y que

$$x^2 = (x-5+5)^2 = (x-5)^2 + 10(x-5) + 25,$$

se tiene que

$$\begin{aligned}
P(x) &= a_0 + a_1x + a_2x^2 \\
&= a_0 + a_1(x-5+5) + a_2(x-5+5)^2 \\
&= a_0 + a_1(x-5) + 5a_1 + a_2(x-5)^2 + 10a_2(x-5) + 25a_2 \\
&= (a_0 + 5a_1 + 25a_2) + (a_1 + 10a_2)(x-5) + a_2(x-5)^2
\end{aligned}$$

con lo que $[P]_{\mathcal{B}} = (a_0 + 5a_1 + 25a_2, a_1 + 10a_2, a_2)$.

(b) Utilizando (a), es fácil ver que

$$[Q]_{\mathcal{B}} = (-5, 4, 1) \qquad \text{y} \qquad [R]_{\mathcal{B}} = (-16, -9, -1).$$

Por la observación anterior,

$$\begin{aligned}[2Q+R]_{\mathcal{B}} &= 2[Q]_{\mathcal{B}}+[R]_{\mathcal{B}}\\ &= 2(-5,4,1)+(-16,-9,-1)\\ &= (-10,8,2)+(-16,-9,-1)\\ &= (-26,-1,1).\end{aligned}$$

De la definición de coordenadas de un vector en una base,

$$2Q(x)+R(x) = -26-(x-5)+(x-5)^2.$$

Es interesante observar que si primero se hubiese operado, quedaría

$$2Q(x)+R(x) = 2(-6x+x^2)+(4+x-x^2) = 4-11x+x^2.$$

Usando (a), las coordenadas de este vector en la base $\mathcal{B}$ son $(-26,-1,1)$, que evidentemente, coinciden. $\triangleright$

Observación 9.4. En definitiva, dado un $\mathbb{K}$-espacio vectorial V de dimensión n y una base $\mathcal{B}$ fija de V, el concepto de coordenadas de un vector en la base $\mathcal{B}$ permite identificar (que es, en realidad, lo que establece el isomorfismo de Descartes) cada vector de V con un vector de $\mathbb{K}^n$ y operar entonces con elementos de $\mathbb{K}^n$ que, en general, resulta más sencillo.

Observación 9.5. Es crucial observar que el isomorfismo de Descartes $\mathcal{D}_{\mathcal{B}}$, definido en el Teorema 9.1, *depende de la base* $\mathcal{B}$ de V elegida para establecerlo[a] o dicho de otro modo, **no** se trata de un **isomorfismo natural** o **canónico**.

[a]Si bien escapa el alcance de este libro, el lector interesado en profundizar en la expresión *depende de la base* puede consultar el artículo [20] para hacerse una idea general.

9.4. Caracterización de espacios vectoriales isomorfos

Proposición 9.2. Sean V_1 y V_2 dos espacios vectoriales sobre el mismo cuerpo $\mathbb{K}$. Si $f : V_1 \to V_2$ es un isomorfismo de espacios vectoriales entonces $f^{-1} : V_2 \to V_1$ es un isomorfismo de espacios vectoriales.

Demostración. Por ser $f : V_1 \to V_2$ un isomorfismo de espacios vectoriales se tiene que f es biyectiva y respeta las operaciones de los espacios. De los Teoremas 1.3 y 1.4 se tiene que existe

$$f^{-1} : V_2 \to V_1$$

y que f^{-1} también es biyectiva.

Falta probar que f^{-1} respeta las operaciones. En efecto, sean $u, v \in V_2$ y $\lambda \in \mathbb{K}$. Denotando por $x := f^{-1}(u)$, $y := f^{-1}(v)$ se tiene que $f(x) = u$ y $f(y) = v$. Entonces

- $\underline{f^{-1}(u+v) = f^{-1}(u) + f^{-1}(v)}$: Dado que f respeta las operaciones, se tiene $u+v = f(x)+f(y) = f(x+y)$, de donde $x+y = f^{-1}(u+v)$. Por otro lado, de $x = f^{-1}(u)$, $y = f^{-1}(v)$ resulta $x+y = f^{-1}(u) + f^{-1}(v)$, de donde se tiene la igualdad buscada.

- $\underline{f^{-1}(\lambda u) = \lambda f^{-1}(u)}$: Dado que f respeta las operaciones, se tiene $\lambda u = \lambda f(x) = f(\lambda x)$, de donde $\lambda x = f^{-1}(\lambda u)$. Por otro lado, de $x = f^{-1}(u)$ resulta $\lambda x = \lambda f^{-1}(u)$, de donde se tiene la igualdad buscada.

□

Sea $\mathbb{K}$ un cuerpo fijo. Se define la relación binaria, llamada **relación de isomorfía vectorial**, siguiente:

$$V_1 \,\mathcal{R}\, V_2 \qquad \Leftrightarrow \qquad V_1 \cong V_2,$$

donde V_1 y V_2 son dos $\mathbb{K}$-espacios vectoriales.

Teorema 9.2. La relación de isomorfía vectorial es una relación de equivalencia sobre un conjunto de $\mathbb{K}$-espacios vectoriales.

Demostración. Se deben probar las propiedades reflexiva, simétrica y transitiva. En efecto,

- $V \cong V$. Para verlo basta definir la aplicación identidad en V.
- $V_1 \cong V_2 \Rightarrow V_2 \cong V_1$. Para verlo basta utilizar la aplicación inversa del isomorfismo $f : V_1 \to V_2$ (que existe pues $V_1 \cong V_2$) de la Proposición 9.2.
- $V_1 \cong V_2$ y $V_2 \cong V_3 \Rightarrow V_1 \cong V_3$. Para verlo basta utilizar la composición de los dos isomorfismos que existen por las hipótesis y el Ejercicio 4 de la página 515.

Los detalles se proponen como ejercicio. □

Lema 9.1. Sean V_1 y V_2 dos $\mathbb{K}$-espacios vectoriales isomorfos de dimensión finita. Entonces $\dim_{\mathbb{K}}(V_1) = \dim_{\mathbb{K}}(V_2)$.

Demostración. Sea $f : V_1 \to V_2$ un isomorfismo de espacios vectoriales. Si $\mathcal{B}_1 = \{u_1, u_2, \ldots, u_n\}$ es una base de V_1 entonces

$$\mathcal{B}_2 = \{f(u_1), f(u_2), \ldots, f(u_n)\}$$

es una base de V_2, con lo que $\dim_{\mathbb{K}}(V_2) = \dim_{\mathbb{K}}(V_1) = n$. En efecto,

- $\underline{\mathcal{B}_2 \text{ es linealmente independiente}}$: Observar, en primer lugar que, si f respeta la multiplicación por escalares, entonces

  $$f(\vec{0}) = f(0\,\vec{0}) = 0f(\vec{0}) = \vec{0}.$$

  Ahora, sea $0 = \alpha_1 f(u_1) + \alpha_2 f(u_2) + \ldots \alpha_n f(u_n)$. Dado que f respeta las operaciones, $f(0) = 0 = f(\alpha_1 u_1 + \alpha_2 u_2 + \ldots \alpha_n u_n)$. Por ser f inyectiva se tiene que $\alpha_1 u_1 + \alpha_2 u_2 + \ldots \alpha_n u_n = 0$. Utilizando que $\mathcal{B}_1$ es un conjunto linealmente independiente se llega a que $\alpha_j = 0$, para todo $j = 1, 2, \ldots, n$.

- $\mathcal{B}_2$ es un sistema de generadores de V_2: Sea $v \in V_2$. Por ser f sobreyectiva, existe $u \in V_1$ tal que $f(u) = v$. Puesto que $u \in V_1 = \overline{\mathcal{B}_1}$, es posible escribir $u = \alpha_1 u_1 + \alpha_2 u_2 + \ldots \alpha_n u_n$, para ciertos escalares $\alpha_i \in \mathbb{K}$. Luego, utilizando que f respeta las operaciones se tiene que $v = f(u) = f(\alpha_1 u_1 + \alpha_2 u_2 + \ldots \alpha_n u_n) = \alpha_1 f(u_1) + \alpha_2 f(u_2) + \ldots \alpha_n f(u_n) \in \overline{\mathcal{B}_2}$. Recordando que se cumple la inclusión trivial $\overline{\mathcal{B}_2} \subseteq V_2$, se tiene que $V_2 = \overline{\mathcal{B}_2}$.

□

Lema 9.2. Sean V_1 y V_2 dos $\mathbb{K}$-espacios vectoriales de dimensión finita tales que $\dim_{\mathbb{K}}(V_1) = \dim_{\mathbb{K}}(V_2) \geq 1$. Entonces $V_1 \cong V_2$.

Demostración. Se consideran los espacios V_1 y V_2 tales que $\dim_{\mathbb{K}}(V_1) = \dim_{\mathbb{K}}(V_2) = n \geq 1$. Del isomorfismo de Descartes se puede asegurar que $V_1 \cong \mathbb{K}^n$ y $V_2 \cong \mathbb{K}^n$. Puesto que la isomorfía es una relación de equivalencia, de $V_1 \cong \mathbb{K}^n$ y $\mathbb{K}^n \cong V_2$ se llega a $V_1 \cong V_2$. □

Teorema 9.3 (Caracterización de espacios vectoriales isomorfos)**.** Sean V_1 y V_2 dos $\mathbb{K}$-espacios vectoriales de dimensión finita. Entonces

$$V_1 \cong V_2 \qquad \Leftrightarrow \qquad \dim_{\mathbb{K}}(V_1) = \dim_{\mathbb{K}}(V_2).$$

Demostración. Se deduce de los Lemas 9.1 y 9.2. □

Se habla entonces de **unicidad** en el sentido de que para cada $n \in \{0\} \cup \mathbb{N}$, existe un único $\mathbb{K}$-espacio vectorial (salvo isomorfismos) de dimensión n.

Además, si $\mathbb{K}^n$ (con las operaciones habituales) es un elemento del conjunto[5] de espacios vectoriales considerado en el Teorema 9.2, $\mathbb{K}^n$ puede tomarse como el **representante** *más sencillo* de la clase a la que pertenece.

Se ha construido un **invariante** en la teoría de espacios vectoriales (de dimensión finita): la dimensión del $\mathbb{K}$-espacio vectorial. Más aún, determina un **conjunto completo de invariantes**. Es claro entonces que la dimensión es una característica intrínseca de cada espacio vectorial (independientemente de la base considerada). A partir de la estructura fundamental de un espacio vectorial, que es la construcción de bases, se pueden establecer sistemas coordenados (dados por los espacios cartesianos) y, gracias al concepto de dimensión de un espacio vectorial, se pueden establecer las cuestiones geométricas elementales, a partir de una estructura puramente algebraica.

Ejemplo 9.5. *Los siguientes espacios vectoriales son isomorfos:*

- $\mathbb{VL}^2 \cong \mathbb{R}^2$ *y* $\mathbb{VL}^3 \cong \mathbb{R}^3$.
- $\mathbb{R}_n[x] \cong \mathbb{R}^{n+1}$.
- $\mathbb{K}^{m\times n} \cong \mathbb{K}^{mn}$. *En particular,* $\mathbb{K}^{m\times 1} \cong \mathbb{K}^m \cong \mathbb{K}^{1\times m}$.

En la Definición 9.2 se denotó mediante $[u]_{\mathcal{B}} = (\alpha_1, \alpha_2, \ldots, \alpha_n)$ las coordenadas del vector u de un espacio vectorial V en una base ordenada $\mathcal{B} = \{u_1, u_2, \ldots, u_n\}$. En muchas ocasiones será necesario multiplicar una matriz por el vector $[u]_{\mathcal{B}}$ y, para que el producto esté bien definido, $[u]_{\mathcal{B}}$ deberá aparecer como un vector columna. Esta identificación queda justificada por el isomorfismo entre el espacio cartesiano $\mathbb{K}^n$ y el espacio de las matrices

[5] Debe considerarse *un conjunto* de $\mathbb{K}$-espacios vectoriales puesto que la clase de *todos* los $\mathbb{K}$-espacios vectoriales (es una clase propia pero) no es un conjunto. Si bien en este libro no se entrará en detalles, este hecho puede deducirse de que la clase de todos los conjuntos (es una clase propia pero) no es un conjunto (véanse la paradoja de Russell y la página 41).

columnas $\mathbb{K}^{n\times 1}$. Luego, se utilizarán indistintamente

$$[u]_{\mathcal{B}} = (\alpha_1, \alpha_2, \ldots, \alpha_n) \in \mathbb{K}^n \qquad \text{o bien} \qquad [u]_{\mathcal{B}} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} \in \mathbb{K}^{n\times 1}.$$

9.4.1. Dimensión, coordenadas y rango

En esta sección, en primer lugar, se probará que es posible determinar si un conjunto de vectores de un $\mathbb{K}$-espacio vectorial cualquiera (de dimensión finita) es o no linealmente independiente, a partir del análisis de una matriz formada con las coordenadas de sus vectores en una base de $\mathbb{K}^n$, lo que resulta ventajoso a la hora de realizar los cálculos.

Obsérvese que $[0]_{\mathcal{B}} = 0$ cualquiera que sea la base utilizada de un espacio vectorial. Este hecho se debe a la unicidad de la representación y a que $0 = 0u_1 + 0u_2 + \cdots + 0u_s$ para cualquier base $\{u_1, u_2, \ldots, u_n\}$ de un $\mathbb{K}$-espacio vectorial V.

Proposición 9.3. Sean V un $\mathbb{K}$-espacio vectorial de dimensión $n \geq 1$, $\mathcal{B}$ una base de V, $\{u_1, u_2, \ldots, u_s\} \subseteq V$ y A la matriz cuyas columnas son las coordenadas de $u_1, u_2, \ldots, u_s$ respecto de la base $\mathcal{B}$, es decir

$$A = \begin{bmatrix} [u_1]_{\mathcal{B}} & [u_2]_{\mathcal{B}} & \ldots & [u_s]_{\mathcal{B}} \end{bmatrix} \in \mathbb{K}^{n\times s}.$$

Las siguientes condiciones son equivalentes:

(a) El conjunto $\{u_1, u_2, \ldots, u_s\}$ es linealmente independiente en V.

(b) $\operatorname{rg}(A) = s$.

Demostración. En este caso no hace falta separar las demostraciones en (a) $\Rightarrow$ (b) y (b) $\Rightarrow$ (a), se pueden realizar simultáneamente puesto que en todos los pasos se obtendrá una equivalencia.

El conjunto $\{u_1, u_2, \ldots, u_s\} \subseteq V$ es linealmente independiente en V si y sólo si una combinación lineal del tipo $0 = \alpha_1 u_1 + \alpha_2 u_2 + \cdots + \alpha_s u_s$ (con $\alpha_i \in \mathbb{K}$, para $i = 1, 2, \ldots, s$) implica $\alpha_1 = \alpha_2 = \cdots = \alpha_s = 0$. Utilizando la Observación 9.3, esto último equivale a que

$$\begin{aligned} 0_{\mathbb{K}^{n\times 1}} &= [0]_{\mathcal{B}} \\ &= [\alpha_1 u_1 + \alpha_2 u_2 + \cdots + \alpha_s u_s]_{\mathcal{B}} \\ &= \alpha_1 [u_1]_{\mathcal{B}} + \alpha_2 [u_2]_{\mathcal{B}} + \cdots + \alpha_s [u_s]_{\mathcal{B}} \\ &= \begin{bmatrix} [u_1]_{\mathcal{B}} & [u_2]_{\mathcal{B}} & \ldots & [u_s]_{\mathcal{B}} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_s \end{bmatrix} \end{aligned}$$

implica $\alpha_1 = \alpha_2 = \cdots = \alpha_s = 0$. Para expresar la combinación lineal anterior como el producto de una matriz por un vector columna se ha utilizado la Proposición 2.8. Llamando

$$A := \begin{bmatrix} [u_1]_{\mathcal{B}} & [u_2]_{\mathcal{B}} & \ldots & [u_s]_{\mathcal{B}} \end{bmatrix},$$

por el Corolario 4.1, un sistema lineal homogéneo

$$A \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_s \end{bmatrix} = 0$$

tiene únicamente la solución trivial si y sólo si $\operatorname{rg}(A) = s$. $\square$

Observación 9.6. Nótese que es equivalente formar la matriz A con las coordenadas de los vectores $\{[u_1]_{\mathcal{B}}, [u_2]_{\mathcal{B}}, \ldots, [u_s]_{\mathcal{B}}\}$ en sus columnas o bien en sus filas puesto que $\operatorname{rg}(A) = \operatorname{rg}(A^t)$.

Ejemplo 9.6. *Extraer una base del subespacio de $\mathbb{R}^{2\times 2}$ dado por*

$$S = \overline{\left\{ A = \begin{bmatrix} -3 & 6 \\ 4 & 2 \end{bmatrix}, B = \begin{bmatrix} 6 & -11 \\ -5 & 2 \end{bmatrix}, C = \begin{bmatrix} 0 & 1 \\ 3 & 6 \end{bmatrix} \right\}}.$$

$\lhd$ Las coordenadas de los vectores (que son matrices) A, B y C en la base canónica

$$\begin{aligned} \mathcal{C} &= \left\{ E^{(1,1)}, E^{(1,2)}, E^{(2,1)}, E^{(2,2)} \right\} \\ &= \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\} \end{aligned}$$

son $[A]_{\mathcal{C}} = (-3, 6, 4, 2)$, $[B]_{\mathcal{C}} = (6, -11, -5, 2)$, $[C]_{\mathcal{C}} = (0, 1, 3, 6)$. Se disponen estos vectores de $\mathbb{R}^4$ en las columnas de una matriz y se calcula su forma escalonada reducida por filas para obtener su rango

$$M = \begin{bmatrix} [A]_{\mathcal{C}} & [B]_{\mathcal{C}} & [C]_{\mathcal{C}} \end{bmatrix} = \begin{bmatrix} -3 & 6 & 0 \\ 6 & -11 & 1 \\ 4 & -5 & 3 \\ 2 & 2 & 6 \end{bmatrix} \sim_f \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = R_M.$$

Al tener la última matriz R_M su rango igual a 2, los vectores $[A]_{\mathcal{C}}, [B]_{\mathcal{C}}$ y $[C]_{\mathcal{C}}$ son linealmente dependientes en $\mathbb{R}^4$.

Observando la forma escalonada reducida por filas en R_M, se tiene que las dos primeras columnas son linealmente independientes y la tercera depende linealmente de las dos anteriores. Por la Proposición 8.8, la misma relación de dependencia/independencia se cumple con las columnas de la matriz de partida. Es decir, $[C]_{\mathcal{C}} = 2[A]_{\mathcal{C}} + [B]_{\mathcal{C}}$, como vectores de $\mathbb{R}^4$. La independencia lineal de $\{[A]_{\mathcal{C}}, [B]_{\mathcal{C}}\}$ asegura la de $\{A, B\}$ en $\mathbb{R}^{2\times 2}$, y por tanto, forman una base de S. Se deduce que $\dim_{\mathbb{R}}(S) = 2$ y, además, $S \cong \mathbb{R}^2$.

Es fácil ver que $\{A, C\}$ es otra base de S (y es posible escribir $[B]_{\mathcal{C}} = -2[A]_{\mathcal{C}} + [C]_{\mathcal{C}}$). En efecto, es evidente que $\overline{\{A, C\}} = S$ y si $\{A, C\}$ fuese linealmente dependiente, uno de sus vectores sería combinación lineal del otro y puesto que B también es combinación lineal de A y C, se llegaría a $\dim_{\mathbb{R}}(S) = 1$, que es una contradiccón pues se ha probado que $\dim_{\mathbb{R}}(S) = 2$.

Un análisis similar prueba que $\{B, C\}$ es otra base de S. $\triangleright$

La consecuencia siguiente es especialmente útil en la práctica puesto que establece la relación entre la dependencia/independencia lineal en $\mathbb{K}^n$ con el rango de una matriz.

Corolario 9.1. Un conjunto de vectores columnas $\{u_1, u_2, \dots, u_s\} \subseteq \mathbb{K}^n$ es linealmente independiente en $\mathbb{K}^n$ si y sólo si

$$\operatorname{rg}\left(\begin{bmatrix} u_1 & u_2 & \dots & u_s \end{bmatrix}\right) = s.$$

Demostración. Se observa que

$$\begin{bmatrix} u_1 & u_2 & \dots & u_s \end{bmatrix} \in \mathbb{K}^{n\times s}.$$

En la Proposición 9.3 se considera como $\mathcal{B}$ la base canónica $\mathcal{C} = \{e_1, e_2, \dots, e_n\}$ de $\mathbb{K}^n$. De este modo,

$$[u_j]_{\mathcal{C}} = u_j, \qquad \text{para todo } j = 1, 2, \dots, s.$$

El resultado ahora sigue de forma inmediata. $\square$

Ejemplo 9.7. *Analizar si los vectores*

$$u_1 = (1,1,3,4), \qquad u_2 = (-1,2,3,2) \qquad y \qquad u_3 = (0,0,1,3)$$

son linealmente independientes en $\mathbb{R}^4$.

$\triangleleft$ Se disponen los vectores u_1, u_2, u_3 como las columnas de una matriz A y se calcula su rango. En efecto,

$$A = \begin{bmatrix} [u_1]_{\mathcal{C}} & [u_2]_{\mathcal{C}} & [u_3]_{\mathcal{C}} \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 2 & 0 \\ 3 & 3 & 1 \\ 4 & 2 & 3 \end{bmatrix} \sim_f \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} = R_A.$$

Dado que $\operatorname{rg}(A) = 3$, los vectores u_1, u_2, u_3 son linealmente independientes en $\mathbb{R}^4$. ▷

Es interesante comparar el método empleado en este último ejemplo con el de la página 398.

Cuando se necesita analizar la independencia lineal de n vectores en un espacio vectorial de dimensión n se puede utilizar el determinante de acuerdo al siguiente resultado.

Corolario 9.2. Un conjunto $\{u_1, u_2, \ldots, u_n\} \subseteq \mathbb{K}^n$ de n vectores columnas es linealmente independiente en $\mathbb{K}^n$ si y sólo si

$$\det\left(\begin{bmatrix} u_1 & u_2 & \ldots & u_n \end{bmatrix}\right) \neq 0.$$

Demostración. Se deduce del Corolario 9.1 y de la Proposición 7.5 de donde

$$\operatorname{rg}\left(\begin{bmatrix} u_1 & u_2 & \ldots & u_n \end{bmatrix}\right) = n \quad \Leftrightarrow \quad \det\left(\begin{bmatrix} u_1 & u_2 & \ldots & u_n \end{bmatrix}\right) \neq 0.$$

□

Sea V un $\mathbb{K}$-espacio vectorial de dimensión $n \geq 1$. Aplicando la definición de independencia lineal y el isomorfismo de Descartes es fácil probar (se propone como ejercicio) que

$$\{u_1, u_2, \ldots, u_s\} \text{ es LI en } V \;\Leftrightarrow\; \{[u_1]_{\mathcal{B}}, [u_2]_{\mathcal{B}}, \ldots, [u_s]_{\mathcal{B}}\} \text{ es LI en } \mathbb{K}^n. \quad (9.2)$$

Este hecho permite dar un resultado más general que la Proposición 9.3.

Proposición 9.4. Sean V un $\mathbb{K}$-espacio vectorial de dimensión $n \geq 1$, $\mathcal{B}$ una base de V, $S = \overline{\{u_1, u_2, \ldots, u_s\}} \leq V$ y A la matriz cuyas columnas son las coordenadas de $u_1, u_2, \ldots, u_s$ respecto de la base $\mathcal{B}$, es decir

$$A = \begin{bmatrix} [u_1]_{\mathcal{B}} & [u_2]_{\mathcal{B}} & \ldots & [u_s]_{\mathcal{B}} \end{bmatrix} \in \mathbb{K}^{n\times s}.$$

Entonces

$$\dim_{\mathbb{K}}(S) = \operatorname{rg}(A).$$

Demostración. Por hipótesis, $S = \overline{\{u_1, u_2, \ldots, u_s\}}$, con lo cual $\{u_1, u_2, \ldots, u_s\}$ es un sistema generador de S. Para determinar la cantidad de vectores de una base de S falta ver cuántos de estos s vectores de V son linealmente independientes.

Para ello, utilizando (9.2), se tiene que $\dim_{\mathbb{K}}(S)$ es la cantidad máxima de columnas linealmente independientes de A como vectores de $\mathbb{K}^n$. Como $A \sim_f R_A$, las columnas básicas de R_A son linealmente independientes y las columnas no básicas dependen linealmente de las columnas básicas precedentes (Observación 8.13); las correspondientes columnas de A se comportan de la misma forma. Entonces

$$\begin{aligned} \dim_{\mathbb{K}}(S) &= \text{número máximo de columnas LI de } A \\ &= \text{número máximo de columnas LI de } R_A \\ &= \text{número de columnas básicas de } R_A \\ &= \operatorname{rg}(A). \end{aligned}$$

□

Observación 9.7. Observando la forma escalonada reducida por filas en la última matriz del Ejemplo 9.6, se tiene que la tercera columna es el duplo de la primera más la segunda. Por la Proposición 8.8 se sabe que la misma relación de dependencia/independencia se cumple con la matriz de partida. Es decir, $[C]_{\mathcal{C}} = 2[A]_{\mathcal{C}} + [B]_{\mathcal{C}}$, como vectores de $\mathbb{R}^4$. Se observa que, en este caso, también es posible asegurar que las propias matrices A, B y C dadas en el subespacio S cumplen dicha relación:

$$C = 2A + B,$$

como vectores de $\mathbb{R}^{2\times 2}$.

¿Se mantienen las mismas relaciones de dependencia encontradas en la Observación 9.7 si en lugar de utilizar la base canónica se considera otra base en el espacio vectorial $\mathbb{R}^{2\times 2}$? Una primera aproximación al problema puede ser considerar otra base. Por ejemplo, se toma la base de $\mathbb{R}^{2\times 2}$ dada por

$$\mathcal{B} = \{B_1, B_2, B_3, B_4\} = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 3 & 0 \\ 1 & -1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} \right\}.$$

Escribiendo los vectores A, B y C como combinación lineal de los de la base $\mathcal{B}$ se obtiene

$$\begin{array}{rcrlrlrlrl} A & = & -15 & B_1 & +0 & B_2 & +4 & B_3 & +6 & B_4, \\ B & = & 21 & B_1 & -8 & B_2 & -5 & B_3 & -3 & B_4, \\ C & = & -9 & B_1 & -8 & B_2 & +3 & B_3 & +9 & B_4, \end{array}$$

con lo que

$$[A]_{\mathcal{B}} = (-15, 0, 4, 6), \quad [B]_{\mathcal{B}} = (21, -8, -5, -3) \quad \text{y} \quad [C]_{\mathcal{B}} = (-9, -8, 3, 9).$$

Dispuestos en las columnas de una matriz y tras calcular su forma escalonada reducida por filas se obtiene

$$\begin{bmatrix} -15 & 21 & -9 \\ 0 & -8 & -8 \\ 4 & -5 & 3 \\ 6 & -3 & 9 \end{bmatrix} \sim_f \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Luego, $\{[A]_{\mathcal{B}}, [B]_{\mathcal{B}}\}$ es un conjunto linealmente independiente y

$$[C]_{\mathcal{B}} = 2[A]_{\mathcal{B}} + [B]_{\mathcal{B}}.$$

Es curioso haber obtenido la misma forma escalonada reducida en ambos casos. ¿Será posible deducir, a partir de esta última igualdad en $\mathbb{R}^4$, la relación $[C]_{\mathcal{C}} = 2[A]_{\mathcal{C}} + [B]_{\mathcal{C}}$ en $\mathbb{R}^4$ y, por lo tanto, la igualdad $C = 2A + B$ en $\mathbb{R}^{2\times 2}$?

En la próxima sección se estudia este y otros problemas, en general, que surgen al relacionar distintas bases.

9.5. Matriz de cambio de base

Se considera ahora un $\mathbb{K}$-espacio vectorial V y dos bases de V:

$$\mathcal{B} = \{u_1, u_2, \ldots, u_n\} \qquad \text{y} \qquad \mathcal{B}' = \{v_1, v_2, \ldots, v_n\}.$$

Es claro que un mismo vector $u \in V$ (fijo pero arbitrario) se puede escribir de forma única como combinación lineal de los vectores de $\mathcal{B}$ como

$$u = \alpha_1 u_1 + \alpha_2 u_2 + \cdots + \alpha_n u_n \tag{9.3}$$

y también de forma única como combinación lineal de los vectores de $\mathcal{B}'$ como

$$u = \beta_1 v_1 + \beta_2 v_2 + \cdots + \beta_n v_n. \tag{9.4}$$

Esto permite considerar sus coordenadas en sendas bases

$$[u]_{\mathcal{B}} = (\alpha_1, \alpha_2, \ldots, \alpha_n) \qquad \text{y} \qquad [u]_{\mathcal{B}'} = (\beta_1, \beta_2, \ldots, \beta_n). \tag{9.5}$$

Es importante observar que mientras que el vector u pertenece al $\mathbb{K}$-espacio vectorial V, los vectores $[u]_{\mathcal{B}}$ y $[u]_{\mathcal{B}'}$ son vectores de $\mathbb{K}^n$. Se quiere conocer la relación que existe entre $[u]_{\mathcal{B}}$ y $[u]_{\mathcal{B}'}$. Para ello, es necesario conocer los vectores de la base $\mathcal{B}$ escritos (de forma única) como combinación lineal de los vectores de la base $\mathcal{B}'$:

$$\begin{array}{rcl} u_1 & = & a_{11}v_1 + a_{21}v_2 + \cdots + a_{n1}v_n \\ u_2 & = & a_{12}v_1 + a_{22}v_2 + \cdots + a_{n2}v_n \\ & \vdots & \\ u_n & = & a_{1n}v_1 + a_{2n}v_2 + \cdots + a_{nn}v_n \end{array} \tag{9.6}$$

Sustituyendo los vectores de (9.6) en (9.3) se tiene:

$$\begin{aligned} u = {} & \alpha_1 u_1 + \alpha_2 u_2 + \cdots + \alpha_n u_n \\ = {} & \alpha_1(a_{11}v_1 + a_{21}v_2 + \cdots + a_{n1}v_n) + \alpha_2(a_{12}v_1 + a_{22}v_2 + \cdots + a_{n2}v_n) + \\ & + \cdots + \alpha_n(a_{1n}v_1 + a_{2n}v_2 + \cdots + a_{nn}v_n) \\ = {} & (\alpha_1 a_{11} + \alpha_2 a_{12} + \cdots + \alpha_n a_{1n})v_1 + (\alpha_1 a_{21} + \alpha_2 a_{22} + \cdots + \alpha_n a_{2n})v_2 + \\ & + \cdots + (\alpha_1 a_{n1} + \alpha_2 a_{n2} + \cdots + \alpha_n a_{nn})v_n. \end{aligned}$$

En definitiva, en este último cálculo se ha escrito el vector u en la base $\mathcal{B}'$. Comparándolo con (9.4), y recordando que la representación es única puesto que $\{v_1, v_2, \ldots, v_n\}$ es un conjunto linealmente independiente, se tiene que

$$\begin{array}{rcl} \beta_1 & = & \alpha_1 a_{11} + \alpha_2 a_{12} + \cdots + \alpha_n a_{1n} \\ \beta_2 & = & \alpha_1 a_{21} + \alpha_2 a_{22} + \cdots + \alpha_n a_{2n} \\ & \vdots & \\ \beta_n & = & \alpha_1 a_{n1} + \alpha_2 a_{n2} + \cdots + \alpha_n a_{nn} \end{array}. \tag{9.7}$$

Se recuerda que se escribirán indistintamente

$$[u]_{\mathbb{B}} = (\alpha_1, \alpha_2, \ldots, \alpha_n) \qquad \text{ó} \qquad [u]_{\mathbb{B}} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix},$$

por el isomorfismo entre las n-tuplas del espacio cartesiano $\mathbb{K}^n$ y el espacio de las matrices columna $\mathbb{K}^{n\times 1}$. Escribiendo las ecuaciones (9.7) mediante la notación de coordenadas (9.5) se tiene que

$$\begin{array}{rcl} [u]_{\mathbb{B}'} & = & \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix} \\ & = & \begin{bmatrix} \alpha_1 a_{11} + \alpha_2 a_{12} + \cdots + \alpha_n a_{1n} \\ \alpha_1 a_{21} + \alpha_2 a_{22} + \cdots + \alpha_n a_{2n} \\ \vdots \\ \alpha_1 a_{n1} + \alpha_2 a_{n2} + \cdots + \alpha_n a_{nn} \end{bmatrix}. \end{array}$$

Ahora, utilizando la definición de suma y multiplicación por escalares de vectores y la Proposición 2.8,

$$\begin{aligned}
[u]_{\mathcal{B}'} &= \alpha_1 \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{n1} \end{bmatrix} + \alpha_2 \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{n2} \end{bmatrix} + \cdots + \alpha_n \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{nn} \end{bmatrix} \\
&= \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} \\
&= A[u]_{\mathcal{B}},
\end{aligned}$$

donde la matriz A se denotará $[\mathcal{B}]_{\mathcal{B}'}$, sus columnas son las coordenadas de los vectores de la base $\mathcal{B}$ escritos en la base $\mathcal{B}'$, y se llama **matriz de cambio de base** o **matriz de paso** que permite pasar de la base $\mathcal{B}$ a la base $\mathcal{B}'$.

Nótese que, en el cálculo anterior, el paso en el que se resalta la combinación lineal de las columnas de A podría evitarse, y utilizarse directamente el producto de matrices, sin embargo, su inclusión es para resaltar la importancia de que la información está escrita por columnas.

El siguiente resultado formaliza el razonamiento anterior.

Proposición 9.5. Sean $\mathcal{B} = \{u_1, u_2, \dots, u_n\}$ y $\mathcal{B}'$ dos bases ordenadas del $\mathbb{K}$-espacio vectorial V de dimensión finita. Existe una única matriz, la matriz de cambio de base

$$[\mathcal{B}]_{\mathcal{B}'} = \begin{bmatrix} [u_1]_{\mathcal{B}'} & [u_2]_{\mathcal{B}'} & \cdots & [u_n]_{\mathcal{B}'} \end{bmatrix},$$

que permite pasar de la base $\mathcal{B}$ a la base $\mathcal{B}'$, es decir, es tal que

$$[u]_{\mathcal{B}'} = [\mathcal{B}]_{\mathcal{B}'}[u]_{\mathcal{B}}, \qquad \text{para todo } u \in V.$$

Demostración. La existencia de la matriz de cambio de base $A := [\mathcal{B}]_{\mathcal{B}'}$ se ha demostrado previamente. Para demostrar su unicidad, si $Q \in \mathbb{K}^{n\times n}$ fuese otra matriz tal que

$$[u]_{\mathcal{B}'} = Q[u]_{\mathcal{B}}, \qquad \text{para todo } u \in V,$$

se tendría que

$$Q[u]_{\mathcal{B}} = A[u]_{\mathcal{B}}, \qquad \text{para todo } u \in V.$$

Tomando $u \in \{u_1, u_2, \ldots, u_n\}$ y teniendo en cuenta que $[u_i]_{\mathcal{B}} = e_i$, es decir, es el i-ésimo vector de la base canónica de $\mathbb{K}^n$, se tendría que $(Q - A)e_i = 0$ para todo $i = 1, 2, \ldots, n$. Al multiplicar se obtiene que cada columna de la resta $Q - A$ es el vector nulo, con lo que $Q = A$. □

Cambiando los roles de $\mathcal{B}$ y $\mathcal{B}'$ se tiene que

$$[u]_{\mathcal{B}} = [\mathcal{B}']_{\mathcal{B}}[u]_{\mathcal{B}'}.$$

Ejemplo 9.8. *Hallar* $[u]_{\mathcal{C}}$ *siendo* $[u]_{\mathcal{B}} = (0, 1, 3)$ *y*

$$\mathcal{B} = \{u_1 = (0, 1, 2), u_2 = (1, 1, -1), u_3 = (1, 0, 0)\}$$

una base de $\mathbb{R}^3$.

◁ Es necesario escribir los vectores de $\mathcal{B}$ como combinación lineal de la base canónica de $\mathbb{R}^3$, lo cual es evidente pues basta ponerlos como columnas. Luego,

$$[u]_{\mathcal{C}} = [\mathcal{B}]_{\mathcal{C}}[u]_{\mathcal{B}} = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \\ 2 & -1 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 4 \\ 1 \\ -1 \end{bmatrix}.$$

▷

Ejemplo 9.9. *Hallar* $[u]_{\mathcal{B}'}$ *siendo* $[u]_{\mathcal{B}} = (0,1,3)$, *y además*

$$\mathcal{B} = \{u_1 = (0,1,2), u_2 = (1,1,-1), u_3 = (1,0,0)\}$$

y

$$\mathcal{B}' = \{v_1 = (1,1,2), v_2 = (1,1,-1), v_3 = (1,2,2)\}$$

dos bases de $\mathbb{R}^3$.

$\lhd$ Primero se debe escribir cada vector de $\mathcal{B}$ como combinación lineal de los vectores de $\mathcal{B}'$. En efecto, dado $(x,y,z) \in \mathbb{R}^3$ se deben encontrar los escalares $\alpha, \beta, \gamma \in \mathbb{R}$ tales que

$$(x,y,z) = \alpha(1,1,2) + \beta(1,1,-1) + \gamma(1,2,2),$$

que es equivalente a resolver el sistema lineal no homogéneo

$$\left\{\begin{array}{ccccccc} \alpha & + & \beta & + & \gamma & = & x \\ \alpha & + & \beta & + & 2\gamma & = & y \\ 2\alpha & - & \beta & + & 2\gamma & = & z \end{array}\right. .$$

Calculando la forma escalonada reducida por filas se tiene

$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & x \\ 1 & 1 & 2 & y \\ 2 & -1 & 2 & z \end{array}\right] \sim_f \left[\begin{array}{ccc|c} 1 & 0 & 0 & \frac{4}{3}x - y + \frac{1}{3}z \\ 0 & 1 & 0 & \frac{2}{3}x - \frac{1}{3}z \\ 0 & 0 & 1 & -x + y \end{array}\right].$$

Luego, $\alpha = \frac{4}{3}x - y + \frac{1}{3}z$, $\beta = \frac{2}{3}x - \frac{1}{3}z$, $\gamma = -x + y$. Sustituyendo en estas expresiones las coordenadas de los vectores de $\mathcal{B}$,

$$\left\{\begin{array}{cccccc} u_1 & = & -\frac{1}{3}v_1 & -\frac{2}{3}v_2 & +v_3 \\ u_2 & = & & v_2 & \\ u_3 & = & \frac{4}{3}v_1 & +\frac{2}{3}v_2 & -v_3 \end{array}\right. .$$

Así,

$$[\mathcal{B}]_{\mathcal{B}'} = \left[\begin{array}{ccc} [u_1]_{\mathcal{B}'} & [u_2]_{\mathcal{B}'} & [u_3]_{\mathcal{B}'} \end{array}\right] = \begin{bmatrix} -\frac{1}{3} & 0 & \frac{4}{3} \\ -\frac{2}{3} & 1 & \frac{2}{3} \\ 1 & 0 & -1 \end{bmatrix}.$$

Finalmente,

$$[u]_{\mathcal{B}'} = [\mathcal{B}]_{\mathcal{B}'}[u]_{\mathcal{B}} = \begin{bmatrix} -\frac{1}{3} & 0 & \frac{4}{3} \\ -\frac{2}{3} & 1 & \frac{2}{3} \\ 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 4 \\ 3 \\ -3 \end{bmatrix}.$$

▷

En el Ejemplo 9.8 se ha calculado $[u]_{\mathcal{C}}$ a partir de $[u]_{\mathcal{B}}$ y en el Ejemplo 9.9 se ha calculado $[u]_{\mathcal{B}'}$ a partir de $[u]_{\mathcal{B}}$. Podría calcularse directamente $[u]_{\mathcal{B}'}$ a partir de $[u]_{\mathcal{C}}$ mediante la expresión

$$[u]_{\mathcal{B}'} = [\mathcal{C}]_{\mathcal{B}'}[u]_{\mathcal{C}}, \tag{9.8}$$

para lo cual se requiere expresar los vectores de la base canónica de $\mathbb{R}^3$ en función de la base $\mathcal{B}'$ (se propone hacerlo como ejercicio). El siguiente resultado se proporciona otra forma de calcularlo.

Proposición 9.6. Sean $\mathcal{B}$ y $\mathcal{B}'$ dos bases ordenadas del $\mathbb{K}$-espacio vectorial V de dimensión finita $n \geq 1$. Entonces la matriz de cambio de base $[\mathcal{B}]_{\mathcal{B}'}$ de $\mathcal{B}$ a $\mathcal{B}'$ es invertible y, además,

$$([\mathcal{B}]_{\mathcal{B}'})^{-1} = [\mathcal{B}']_{\mathcal{B}}.$$

Demostración. Las coordenadas de cada vector de la base

$$\mathcal{B} = \{u_1, u_2, \ldots, u_n\}$$

están relacionadas en las bases $\mathcal{B}$ y $\mathcal{B}'$ mediante

$$[u_i]_{\mathcal{B}} = [\mathcal{B}']_{\mathcal{B}}[u_i]_{\mathcal{B}'}, \qquad \text{para } i = 1, 2, \ldots, n.$$

Teniendo en cuenta que $[u_i]_{\mathcal{B}} = e_i$ (es decir, el vector de coordenadas de u_i en la base $\mathcal{B}$ es el i-ésimo vector de la base canónica de $\mathbb{K}^n$), aplicando la

Proposición 2.8 se tiene que

$$\begin{aligned}
[\mathcal{B}']_{\mathcal{B}}[\mathcal{B}]_{\mathcal{B}'} &= [\mathcal{B}']_{\mathcal{B}}\begin{bmatrix} [u_1]_{\mathcal{B}'} & [u_2]_{\mathcal{B}'} & \ldots & [u_n]_{\mathcal{B}'} \end{bmatrix} \\
&= \begin{bmatrix} [\mathcal{B}']_{\mathcal{B}}[u_1]_{\mathcal{B}'} & [\mathcal{B}']_{\mathcal{B}}[u_2]_{\mathcal{B}'} & \ldots & [\mathcal{B}']_{\mathcal{B}}[u_n]_{\mathcal{B}'} \end{bmatrix} \\
&= \begin{bmatrix} [u_1]_{\mathcal{B}} & [u_2]_{\mathcal{B}} & \ldots & [u_n]_{\mathcal{B}} \end{bmatrix} \\
&= \begin{bmatrix} e_1 & e_2 & \ldots & e_n \end{bmatrix} \\
&= I_n.
\end{aligned}$$

Puesto que el producto de dos matrices cuadradas da la identidad, se tiene que ambas son invertibles y una es la inversa de la otra. □

Ahora, la expresión (9.8) se puede calcular mediante

$$[u]_{\mathcal{B}'} = [\mathcal{C}]_{\mathcal{B}'}[u]_{\mathcal{C}} = ([\mathcal{B}']_{\mathcal{C}})^{-1}[u]_{\mathcal{C}}.$$

Ejemplo 9.10. *Hallar $[u]_{\mathcal{B}'}$ a partir del cálculo de $[u]_{\mathcal{C}}$ realizado en el Ejemplo 9.8 para la misma base de $\mathbb{R}^3$*

$$\mathcal{B}' = \{v_1 = (1,1,2), v_2 = (1,1,-1), v_3 = (1,2,2)\}.$$

◁ Se tiene que

$$\begin{aligned}
[u]_{\mathcal{B}'} &= [\mathcal{C}]_{\mathcal{B}'}[u]_{\mathcal{C}} \\
&= ([\mathcal{B}']_{\mathcal{C}})^{-1}[u]_{\mathcal{C}} \\
&= \left(\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 2 \\ 2 & -1 & 2 \end{bmatrix}\right)^{-1}\begin{bmatrix} 4 \\ 1 \\ -1 \end{bmatrix} \\
&= \begin{bmatrix} \frac{4}{3} & -1 & \frac{1}{3} \\ \frac{2}{3} & 0 & -\frac{1}{3} \\ -1 & 1 & 0 \end{bmatrix}\begin{bmatrix} 4 \\ 1 \\ -1 \end{bmatrix} \\
&= \begin{bmatrix} 4 \\ 3 \\ -3 \end{bmatrix}.
\end{aligned}$$

Es evidente que el vector encontrado coincide con el resultado obtenido en el ejemplo anterior. ▷

Utilizando matrices de cambio de bases es posible dar una nueva caracterizción de matrices invertibles.

Corolario 9.3. Sea $A \in \mathbb{K}^{n\times n}$. Entonces A es una matriz invertible si y sólo si A es una matriz de cambio de base.

Demostración. ($\Rightarrow$) Se considera la matriz A particionada según sus columnas como

$$A = \begin{bmatrix} a_1 & a_2 & \dots & a_n \end{bmatrix}.$$

Los vectores $u_i := a_i$ de $\mathbb{K}^n$ cumplen que $[u_i]_{\mathcal{C}} = a_i$, donde $\mathcal{C}$ es la base canónica de $\mathbb{K}^n$. Dado que A es invertible, $\text{rg}(A) = n$ y el Corolario 9.1 asegura que $\mathcal{B} = \{u_1, u_2, \dots, u_n\}$ es una base de $\mathbb{K}^n$ (por ser $\mathcal{B}$ un conjunto linealmente independiente y $\dim_{\mathbb{K}}(\mathbb{K}^n) = n$) y cumple que

$$[\mathcal{B}]_{\mathcal{C}} = A.$$

($\Leftarrow$) Se ha probado en la Proposición 9.6. □

9.5.1. Relaciones entre las matrices de cambio de base

En algunas ocasiones se dispone de tres bases ordenadas y es necesario establecer relaciones entre ellas. El siguiente resultado da una fórmula para vincularlas.

Proposición 9.7. Sean $\mathcal{B}$, $\mathcal{B}'$ y $\mathcal{B}''$ tres bases ordenadas del $\mathbb{K}$-espacio vectorial V de dimensión finita $n \geq 1$. Entonces la matriz de cambio de base $[\mathcal{B}]_{\mathcal{B}''}$ de $\mathcal{B}$ a $\mathcal{B}''$ se obtiene mediante

$$[\mathcal{B}]_{\mathcal{B}''} = [\mathcal{B}']_{\mathcal{B}''}[\mathcal{B}]_{\mathcal{B}'}.$$

Demostración. Las coordenadas de cada vector de la base

$$\mathcal{B} = \{u_1, u_2, \ldots, u_n\}$$

están relacionadas en las bases $\mathcal{B}'$ y $\mathcal{B}''$ mediante

$$[u_i]_{\mathcal{B}''} = [\mathcal{B}']_{\mathcal{B}''}[u_i]_{\mathcal{B}'}, \qquad \text{para } i = 1, 2, \ldots, n.$$

Aplicando la Proposición 2.8 se tiene que

$$\begin{aligned}
[\mathcal{B}]_{\mathcal{B}''} &= \begin{bmatrix} [u_1]_{\mathcal{B}''} & [u_2]_{\mathcal{B}''} & \ldots & [u_n]_{\mathcal{B}''} \end{bmatrix} \\
&= \begin{bmatrix} [\mathcal{B}']_{\mathcal{B}''}[u_1]_{\mathcal{B}'} & [\mathcal{B}']_{\mathcal{B}''}[u_2]_{\mathcal{B}'} & \ldots & [\mathcal{B}']_{\mathcal{B}''}[u_n]_{\mathcal{B}'} \end{bmatrix} \\
&= [\mathcal{B}']_{\mathcal{B}''} \begin{bmatrix} [u_1]_{\mathcal{B}'} & [u_2]_{\mathcal{B}'} & \ldots & [u_n]_{\mathcal{B}'} \end{bmatrix} \\
&= [\mathcal{B}']_{\mathcal{B}''}[\mathcal{B}]_{\mathcal{B}'}.
\end{aligned}$$

□

Un caso particular importante es cuando se pasa a través de la base canónica, como base intermedia, en $\mathbb{K}^n$.

Corolario 9.4. Sean $\mathcal{B}$ y $\mathcal{B}'$ dos bases ordenadas del $\mathbb{K}$-espacio vectorial $\mathbb{K}^n$ y sea $\mathcal{C}$ su base canónica. Entonces

$$[\mathcal{B}]_{\mathcal{B}'} = [\mathcal{C}]_{\mathcal{B}'}[\mathcal{B}]_{\mathcal{C}} = ([\mathcal{B}']_{\mathcal{C}})^{-1}[\mathcal{B}]_{\mathcal{C}}.$$

Demostración. Basta particularizar en la Proposición 9.7 a la base canónica. Para la segunda igualdad utilizar la Proposición 9.6. □

Ejemplo 9.11. *Resolver de nuevo el Ejemplo 9.9 utilizando el Corolario 9.4.*

◁ Aplicando la Proposición 9.5 y el Corolario 9.4 se tiene que

$$\begin{aligned}
[u]_{\mathcal{B}'} &= [\mathcal{B}]_{\mathcal{B}'}[u]_{\mathcal{B}} \\
&= ([\mathcal{B}']_{\mathcal{C}})^{-1}[\mathcal{B}]_{\mathcal{C}}[u]_{\mathcal{B}} \\
&= \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 2 \\ 2 & -1 & 2 \end{bmatrix}^{-1} \begin{bmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \\ 2 & -1 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 3 \end{bmatrix} \\
&= \begin{bmatrix} \frac{4}{3} & -1 & \frac{1}{3} \\ \frac{2}{3} & 0 & -\frac{1}{3} \\ -1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \\ 2 & -1 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 3 \end{bmatrix} \\
&= \begin{bmatrix} 4 \\ 3 \\ -3 \end{bmatrix}.
\end{aligned}$$

▷

Ejemplo 9.12. *Se considera de nuevo el Ejemplo 9.4. Se pide calcular* $[2Q + R]_{\mathcal{B}}$ *para los vectores* $Q(x) = -6x + x^2$ *y* $R(x) = 4 + x - x^2$ *siendo* $\mathcal{B} = \{1, x-5, (x-5)^2\}$ *una base del* $\mathbb{R}$*-espacio vectorial* $\mathbb{R}_2[x] = \{P \in \mathbb{R}[x] : P(x) = 0 \vee \mathrm{gr}(P(x)) \leq 2\}$.

◁ Puesto que $2Q(x) + R(x) = 2(-6x + x^2) + (4 + x - x^2) = 4 - 11x + x^2$, sus coordenadas en la base canónica $\mathcal{C} = \{1, x, x^2\}$ vienen dadas por el vector

$$[2Q + R]_{\mathcal{C}} = \begin{bmatrix} 4 \\ -11 \\ 1 \end{bmatrix}.$$

A partir de la base $\mathcal{B}$ formada por

$$1, \qquad x - 5 = -5 + x \qquad \text{y} \qquad (x-5)^2 = 25 - 10x + x^2,$$

si $\mathcal{C} = \{1, x, x^2\}$, se tiene que

$$[\mathcal{B}]_{\mathcal{C}} = \begin{bmatrix} 1 & -5 & 25 \\ 0 & 1 & -10 \\ 0 & 0 & 1 \end{bmatrix}.$$

Luego,

$$\begin{aligned} [2Q + R]_{\mathcal{B}} &= [\mathcal{C}]_{\mathcal{B}}[2Q + R]_{\mathcal{C}} \\ &= ([\mathcal{B}]_{\mathcal{C}})^{-1}[2Q + R]_{\mathcal{C}} \\ &= \left(\begin{bmatrix} 1 & -5 & 25 \\ 0 & 1 & -10 \\ 0 & 0 & 1 \end{bmatrix} \right)^{-1} \begin{bmatrix} 4 \\ -11 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 5 & 25 \\ 0 & 1 & 10 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 4 \\ -11 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} -26 \\ -1 \\ 1 \end{bmatrix}. \end{aligned}$$

Es decir,

$$2Q(x) + R(x) = -26 - (x - 5) + (x - 5)^2.$$

$\triangleright$

Observación 9.8. Ahora se obtienen fácilmente las relaciones entre las combinaciones lineales del comentario posterior a la Observacion 9.7. La relación

$$[C]_{\mathcal{C}} = 2[A]_{\mathcal{C}} + [B]_{\mathcal{C}}$$

se consigue multiplicando a izquierda $[C]_{\mathcal{B}} = 2[A]_{\mathcal{B}} + [B]_{\mathcal{B}}$ por $[\mathcal{B}]_{\mathcal{C}}$ y la relación

$$[C]_{\mathcal{B}} = 2[A]_{\mathcal{B}} + [B]_{\mathcal{B}}$$

se encuentra multiplicando a izquierda $[C]_{\mathcal{C}} = 2[A]_{\mathcal{C}} + [B]_{\mathcal{C}}$ por $[\mathcal{C}]_{\mathcal{B}}$.

9.6. Subespacios y sistemas homogéneos

Se recuerda la definición de sistema de ecuaciones lineales.

Definición 9.4. Sean m y n dos números naturales. Al conjunto de m ecuaciones lineales con n incógnitas

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

se lo llama **sistema de m ecuaciones lineales con n incógnitas**. Si $b_1 = b_2 = \cdots = b_m = 0$, el sistema se llama **homogéneo** y si para algún $i_0 \in \{1, 2, \ldots, m\}$ se tiene $b_{i_0} \neq 0$, el sistema se llama **no homogéneo**.

Matricialmente un sistema de ecuaciones lineales se representa mediante

$$Ax = b, \qquad \text{donde} \quad A \in \mathbb{K}^{m\times n},\ b \in \mathbb{K}^{m\times 1}$$

y el sistema es **homogéneo** si $b = 0$.

Si se tiene un sistema homogéneo $Ax = 0$ y se utiliza la notación

$$S_i = \{(x_1, x_2, \ldots, x_n) \in \mathbb{K}^n : \ a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n = 0\},$$

para cada $i = 1, 2, \ldots, m$, se tiene que el conjunto solución del sistema $Ax = 0$ es el subespacio vectorial $\bigcap_{i=1}^{m} S_i$ de $\mathbb{K}^n$. Más adelante, se probará el interesante hecho de que todo subespacio vectorial de $\mathbb{K}^n$ es intersección de subespacios de este tipo.

Caracterización del conjunto solución de los sistemas lineales no homogéneos

Si S es un subespacio de un espacio vectorial V y $x_0 \in V$, se utilizará la notación:

$$x_0 + S = \{x_0 + s : s \in S\}$$

para denotar al conjunto de vectores que son suma de uno fijo x_0 más uno arbitrario de S. Debido a las connotaciones geométricas[6] que se observan en $\mathbb{R}^2$ y $\mathbb{R}^3$ el conjunto $x_0 + S$ se suele llamar **variedad lineal vectorial** que pasa por x_0 y es paralela al subespacio vectorial S.

El siguiente resultado muestra que la estructura de las soluciones de un sistema no homogéneo se deduce de la estructura de las soluciones de un sistema homogéneo junto a un vector particular.

Teorema 9.4. Sean $A \in \mathbb{K}^{m\times n}$, $b \in \mathbb{K}^{m\times 1}$ y sea x_p una solución particular del sistema $Ax = b$. Entonces el conjunto solución de $Ax = b$ dado por

$$S(A, b) = \{x \in \mathbb{K}^{n\times 1} : Ax = b\}$$

se obtiene como

$$S(A, b) = x_p + S(A, 0).$$

Demostración. Esta igualdad de conjuntos se probará por doble inclusión. Como x_p es una solución particular del sistema $Ax = b$, se tiene que $Ax_p = b$.

($\subseteq$) Si $x \in S(A, b)$ entonces $Ax = b$. Luego, de $Ax_p = b$ el vector $z := x - x_p$ cumple que

$$Az = A(x - x_p) = Ax - Ax_p = b - b = 0.$$

Luego $x = x_p + z$ con $z \in S(A, 0)$. Se ha probado que $S(A, b) \subseteq x_p + S(A, 0)$.

[6]Para verlo, se propone como ejercicio pintar la gráfica de una recta en $\mathbb{R}^2$ que pasa por el origen a la que se le suma un vector fijo mediante la definición citada, y repetir con un plano en $\mathbb{R}^3$ por el origen al que se le suma un vector fijo.

($\supseteq$) Si $x \in x_p + S(A, 0)$ entonces $x = x_p + z$, para algún $z \in S(A, 0)$. Luego, de $Ax_p = b$ y $Az = 0$ se tiene que

$$Ax = A(x_p + z) = Ax_p + Az = b + 0 = b.$$

Así, $x \in S(A, b)$. Por lo tanto, $x_p + S(A, 0) \subseteq S(A, b)$. □

Por lo tanto, para obtener la solución general de un sistema lineal $Ax = b$ es necesario conocer una solución particular x_p de dicho sistema y la solución general del sistema lineal homogéneo $Ax = 0$. En lo que sigue se analizará la estructura algebraica de este último.

El conjunto $S(A, 0)$ es crucial en Álgebra Lineal y se denomina espacio nulo de A.

Definición 9.5. Sea $A \in \mathbb{K}^{m\times n}$. El conjunto solución del sistema lineal homogéneo $Ax = 0$ se llama **espacio nulo** de A y se denota $\mathcal{N}(A)$. En símbolos,

$$\mathcal{N}(A) = \{x \in \mathbb{K}^{n\times 1} : \ Ax = 0\}.$$

Se recuerda que, al estudiar ejemplos de subespacios vectoriales, se probó que $\mathcal{N}(A)$ es un subespacio del espacio vectorial $\mathbb{K}^{n\times 1}$.

Caracterización del conjunto solución de los sistemas lineales homogéneos

El concepto de independencia lineal y el concepto de base permiten analizar con más profundidad la estructura del conjunto solución de un sistema homogéneo $Ax = 0$ para $A \in \mathbb{K}^{m\times n}$.

Teorema 9.5. Sea $A \in \mathbb{K}^{m\times n}$ con $\text{rg}(A) = r$. Entonces el conjunto solución

$$\mathcal{N}(A) = S(A,0) = \{x \in \mathbb{K}^{n\times 1} : Ax = 0\}$$

del sistema homogéneo $Ax = 0$ es un subespacio vectorial de $\mathbb{K}^{n\times 1}$ de dimensión $n - r$. En consecuencia,

- si $r = n$ entonces $\dim_{\mathbb{K}}(S(A,0)) = 0$.
- si $r < n$ entonces $\dim_{\mathbb{K}}(S(A,0)) = n - r > 0$. Por lo tanto, existen vectores $h_{r+1}, h_{r+2}, \ldots, h_n \in \mathbb{K}^{n\times 1}$ linealmente independientes tales que cualquier vector $x \in S(A,0)$ es de la forma

$$x = \alpha_{r+1}h_{r+1} + \alpha_{r+2}h_{r+2} + \cdots + \alpha_n h_n,$$

para ciertos $\alpha_{r+1}, \alpha_{r+2}, \ldots, \alpha_n \in \mathbb{K}$.

Demostración. Es conocido que $S(A,0) \leq \mathbb{K}^{n\times 1}$. Ahora se debe determinar su dimensión. Se analizan 3 casos:

- Si $r = n$, por el Teorema de Rouché-Frobenius, el sistema $Ax = 0$ es compatible determinado. Luego, $S(A,0) = \{0\}$.
- Si $r = 0$, el conjunto solución de $Ax = 0$ es todo $\mathbb{K}^{n\times 1}$ y cualquier base de $\mathbb{K}^{n\times 1}$ resuelve el problema. En este caso, $\dim_{\mathbb{K}}(S(A,0)) = n$.
- Sea $0 < r < n$. Se debe hallar una base para $S(A,0)$. Dado que $\text{rg}(A) = r > 0$, el Teorema 6.4 asegura que $A \sim \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$, es decir, existen matrices invertibles $P \in \mathbb{K}^{n\times n}$ y $Q \in \mathbb{K}^{m\times m}$ tales que

$$QAP = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} \qquad \text{o bien} \qquad QA = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} P^{-1}.$$

Una solución x del sistema homogéneo $Ax = 0$ cumple que

$$Ax = 0 \qquad \Leftrightarrow \qquad QAx = 0 \qquad \Leftrightarrow \qquad \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} P^{-1}x = 0$$

$$\Leftrightarrow \qquad \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = 0 \qquad \Leftrightarrow \qquad z_1 = 0,$$

donde se ha denotado $P^{-1}x = y = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \in \mathbb{K}^{(r+(n-r))\times 1}$.

Es decir, una solución de $Ax = 0$ es de la forma

$$x = P \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = P \begin{bmatrix} 0 \\ z_2 \end{bmatrix} \in \mathbb{K}^{n\times 1}.$$

Si la matriz P se particiona según sus columnas y se escriben las componentes del vector z_2 como

$$P = \begin{bmatrix} p_1 & \ldots & p_r & h_{r+1} & \ldots & h_n \end{bmatrix} \qquad \text{y} \qquad z_2 = \begin{bmatrix} \alpha_{r+1} \\ \vdots \\ \alpha_n \end{bmatrix},$$

utilizando la Proposición 2.8, se tiene que

$$\begin{aligned} x &= \begin{bmatrix} p_1 & \ldots & p_r & h_{r+1} & \ldots & h_n \end{bmatrix} \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \alpha_{r+1} \\ \ldots \\ \alpha_n \end{bmatrix} \\ &= \alpha_{r+1}h_{r+1} + \alpha_{r+2}h_{r+2} + \cdots + \alpha_n h_n, \end{aligned}$$

con lo cual $\{h_{r+1}, h_{r+2}, \ldots, h_n\}$ es un sistema generador de $S(A, 0)$.

Además, es claro que el conjunto $\{h_{r+1}, h_{r+2}, \ldots, h_n\}$ es linealmente independiente por ser un subconjunto del conjunto linealmente independiente formado por las columnas de P, que es invertible (véase el Corolario 9.1). Por lo tanto, $\{h_{r+1}, h_{r+2}, \ldots, h_n\}$ es una base de $S(A, 0)$. En este caso, se tiene que $\dim_{\mathbb{K}}(S(A, 0)) = n - r$.

$\square$

Según este resultado la dimensión del subespacio $\mathcal{N}(A)$ coincide con el número de grados de libertad del sistema (es decir, el número de incógnitas menos el número de ecuaciones relevantes).

Observación 9.9. Si $A \in \mathbb{K}^{m\times n}$ es de rango r, en el Teorema 9.5 se ha probado la siguiente importante fórmula[a]:

$$\dim_{\mathbb{K}}(\mathcal{N}(A)) + \mathrm{rg}(A) = n.$$

Pensada a partir del sistema lineal homogéneo $Ax = 0$, la fórmula anterior expresa que:

$$\left\{\begin{array}{c}\text{número de}\\ \text{columnas no básicas}\end{array}\right\} + \left\{\begin{array}{c}\text{número de}\\ \text{columnas básicas}\end{array}\right\} = \left\{\begin{array}{c}\text{número total}\\ \text{de columnas}\end{array}\right\}.$$

[a] En la teoría de Aplicaciones Lineales se conoce como **Teorema de la dimensión**.

En la demostración del Teorema 9.5 se utilizó la equivalencia entre las matrices

$$A \qquad \text{y} \qquad \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$$

garantizada por la hipótesis $\mathrm{rg}(A) = r > 0$. Se observa que a pesar de ello, el paso crucial es que los sistemas $Ax = 0$ y $QAx = 0$ son equivalentes (tratando la matriz P^{-1} conjuntamente con las incógnitas). Esto facilita la demostración de la existencia de la base buscada. Sin embargo, como se aprecia en el siguiente ejemplo, en la práctica no hace falta realizar las operaciones elementales por columnas indicadas mediante la matriz P para hallar una base del conjunto solución de $Ax = 0$, sino que alcanza con hallar la forma escalonada reducida por filas de A. Al resolver un sistema lineal homogéneo $Ax = 0$ con este procedimiento se encontrará una base del subespacio solución del mismo.

Ejemplo 9.13. *Resolver el sistema lineal homogéneo*

$$\left\{\begin{array}{rcrcrcrcrcr} x_1 & + & 3x_2 & + & 4x_3 & - & 2x_4 & + & 7x_5 & = & 0\\ 3x_1 & + & 9x_2 & - & 4x_3 & + & 2x_4 & + & 13x_5 & = & 0\\ x_1 & + & 3x_2 & - & 4x_3 & + & 2x_4 & + & 3x_5 & = & 0 \end{array}\right.$$

y hallar una base en $\mathbb{R}^5$ *del subespacio solución.*

$\triangleleft$ Puesto que $A \sim_f R_A$, los sistemas lineales homogéneos $Ax = 0$ y $R_A x = 0$ son sistemas equivalentes (Lema 4.1), y por tanto se resolverá este último. Realizando operaciones elementales por filas se tiene

$$A = \begin{bmatrix} 1 & 3 & 4 & -2 & 7\\ 3 & 9 & -4 & 2 & 13\\ 1 & 3 & -4 & 2 & 3 \end{bmatrix} \sim_f \begin{bmatrix} 1 & 3 & 0 & 0 & 5\\ 0 & 0 & 1 & -1/2 & 1/2\\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} = R_A.$$

Mediante sustitución regresiva se despejan las incógnitas principales x_1 y x_3, asociadas a las columnas básicas, en función de las incógnitas libres x_2, x_4 y x_5, obteniendo $x_1 = -3x_2 - 5x_5$ y $x_3 = \frac{1}{2}x_4 - \frac{1}{2}x_5$. Es posible escribir el vector que proporciona la solución general como sigue

$$\begin{bmatrix} x_1\\ x_2\\ x_3\\ x_4\\ x_5 \end{bmatrix} = \begin{bmatrix} -3x_2 - 5x_5\\ x_2\\ \frac{1}{2}x_4 - \frac{1}{2}x_5\\ x_4\\ x_5 \end{bmatrix} = x_2 \begin{bmatrix} -3\\ 1\\ 0\\ 0\\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 0\\ 0\\ \frac{1}{2}\\ 1\\ 0 \end{bmatrix} + x_5 \begin{bmatrix} -5\\ 0\\ -\frac{1}{2}\\ 0\\ 1 \end{bmatrix} \tag{9.9}$$

con $x_2, x_3, x_5 \in \mathbb{R}$. De este modo, los vectores

$$\begin{bmatrix} -3\\ 1\\ 0\\ 0\\ 0 \end{bmatrix}, \begin{bmatrix} 0\\ 0\\ \frac{1}{2}\\ 1\\ 0 \end{bmatrix}, \begin{bmatrix} -5\\ 0\\ -\frac{1}{2}\\ 0\\ 1 \end{bmatrix}$$

forman un sistema de generadores del conjunto solución del sistema y, por la disposición de los unos y los ceros en el vector que acompaña a cada incógnita libre, es claro que forman un conjunto linealmente independiente de $\mathbb{R}^5$. Observando la segunda, cuarta y quinta filas, tienen un uno en la componente que ocupaba en el vector correspondiente a la incógnita libre y ceros en las componentes correspondientes a esa incógnita libre en los vectores restantes. Luego, determinan una base del conjunto solución del sistema original, subespacio que tiene dimensión $n - \text{rg}(A) = 5 - 2 = 3$. $\triangleright$

Puesto que la solución de $Ax = 0$ se expresa a partir de los parámetros $x_2, x_3, x_5 \in \mathbb{R}$, la ecuación (9.9) suelen llamarse *ecuaciones paramétricas* del subespacio solución. En la sección próxima se estudiará este tipo de ecuaciones para un subespacio cualquiera de un espacio vectorial abstracto V.

Se ha probado que el conjunto solución de un sistema lineal homogéneo es un subespacio vectorial de $\mathbb{K}^n$ de dimensión $d := n - \text{rg}(A)$. Ahora se probará que todo subespacio de $\mathbb{K}^n$ tiene esa forma, es decir, que es el conjunto solución de algún sistema homogéneo. De este modo, dichos conjuntos solución quedarán completamente caracterizados.

Es claro que si un subespacio S de dimensión $d \geq 1$ del $\mathbb{K}$-espacio vectorial $\mathbb{K}^n$ de dimensión $n \geq 1$ cumple $d = n$ entonces $S = \mathbb{K}^n$, con lo que S se puede representar como el conjunto solución $(x_1, x_2, \ldots, x_n)$ de la (única) ecuación lineal

$$0x_1 + 0x_2 + \cdots + 0x_n = 0.$$

Entonces, el caso interesante es el de un subespacio S de dimensión $1 \leq d < n$.

En realidad, se demostrará un resultado más general: todo subespacio de un espacio vectorial de dimensión finita (arbitrario, no necesariamente $\mathbb{K}^n$) "es"[7] solución de un sistema de ecuaciones lineales homogéneo.

[7]Podría ocurrir que no sean exactamente iguales por tener, quizás, elementos de diferente naturaleza, sin embargo será isomorfo a un subespacio de $\mathbb{K}^n$ que es solución de un sistema homogéneo.

Teorema 9.6. Todo subespacio de dimensión $1 \leq d < n$ de un $\mathbb{K}$-espacio vectorial V de dimensión $n \geq 1$ es isomorfo al conjunto solución de un sistema lineal homogéneo de $n - d$ ecuaciones (linealmente) independientes.

Demostración. Puesto que el $\mathbb{K}$-espacio vectorial tiene dimensión $n \geq 1$ es posible considerar una base $\mathcal{B} = \{v_1, v_2, \dots, v_n\}$ de V, con al menos un vector.

Sea S un subespacio vectorial de V de dimensión $1 \leq d < n$ y sea $\mathcal{B}_S = \{u_1, u_2, \dots, u_d\}$ una base de S tal que

$$\begin{array}{rcl} [u_1]_{\mathcal{B}} & = & (a_{11}, a_{21}, \dots, a_{n1}), \\ [u_2]_{\mathcal{B}} & = & (a_{12}, a_{22}, \dots, a_{n2}), \\ & \vdots & \\ [u_d]_{\mathcal{B}} & = & (a_{1d}, a_{2d}, \dots, a_{nd}). \end{array}$$

Sea x un vector genérico de S. Existen escalares $x_1, x_2, \dots, x_n \in \mathbb{K}$ tales que

$$[x]_{\mathcal{B}} = (x_1, x_2, \dots, x_n).$$

Ahora se tiene que $x \in S$ si y sólo si x es combinación lineal de $u_1, u_2, \dots, u_d$, es decir existen escalares $\lambda_1, \lambda_2, \dots, \lambda_d \in \mathbb{K}$ tales que

$$x = \lambda_1 u_1 + \lambda_2 u_2 + \cdots + \lambda_d u_d.$$

Tomando coordenadas en la base $\mathcal{B}$ (es decir, aplicando el isomorfismo de Descartes y, por lo tanto, pasando a un espacio isomorfo, eventualmente diferente) y utilizando la Observación 9.3 se tiene

$$[x]_{\mathcal{B}} = \lambda_1 [u_1]_{\mathcal{B}} + \lambda_2 [u_2]_{\mathcal{B}} + \cdots + \lambda_d [u_d]_{\mathcal{B}}.$$

Escribiendo esta última igualdad en términos de sus componentes se obtiene el sistema de ecuaciones lineales

$$\begin{cases} x_1 &= \lambda_1 a_{11} + \lambda_2 a_{12} + \cdots + \lambda_d a_{1d} \\ x_2 &= \lambda_1 a_{21} + \lambda_2 a_{22} + \cdots + \lambda_d a_{2d} \\ &\vdots \\ x_n &= \lambda_1 a_{n1} + \lambda_2 a_{n2} + \cdots + \lambda_d a_{nd} \end{cases} \tag{9.10}$$

que matricialmente se escribe como $A\lambda = x$ siendo

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1d} \\ a_{21} & a_{22} & \dots & a_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nd} \end{bmatrix} \in \mathbb{K}^{n\times d}, \quad \lambda = \begin{bmatrix} \lambda_1 \\ \lambda_n \\ \vdots \\ \lambda_d \end{bmatrix}, \quad x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

Dado que el sistema $A\lambda = x$ tiene solución, por el Teorema de Rouché-Frobenius, $\mathrm{rg}(A) = \mathrm{rg}\left(\begin{bmatrix} A & x \end{bmatrix}\right)$. Por la independencia lineal de los vectores de $\mathcal{B}_S$, debe ser $\mathrm{rg}(A) = d$ (Proposición 9.3), por lo tanto

$$\mathrm{rg}\left(\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1d} & x_1 \\ a_{21} & a_{22} & \dots & a_{2d} & x_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nd} & x_n \end{bmatrix}\right) = d.$$

Por la Proposición 7.5 es conocido que existe un menor no nulo de orden d de A tal que todos los menores que se obtienen orlándolo son nulos. Como el reordenamiento de filas no altera el rango de la matriz es posible suponer, sin pérdida de generalidad, que el menor no nulo de orden d se encuentra en las primeras d filas (y d columnas) de A. Luego, orlándolo con la columna $(d+1)$-ésima y cualesquiera de las $n-d$ filas restantes, el menor obtenido será nulo. Por lo tanto,

$$\begin{vmatrix} a_{11} & a_{12} & \dots & a_{1d} & x_1 \\ a_{21} & a_{22} & \dots & a_{2d} & x_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{d1} & a_{d2} & \dots & a_{dd} & x_d \\ a_{d+1,1} & a_{d+1,2} & \dots & a_{d+1,d} & x_{d+1} \end{vmatrix} = 0, \tag{9.11}$$

$$\begin{vmatrix} a_{11} & a_{12} & \dots & a_{1d} & x_1 \\ a_{21} & a_{22} & \dots & a_{2d} & x_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{d1} & a_{d2} & \dots & a_{dd} & x_d \\ a_{d+2,1} & a_{d+2,2} & \dots & a_{d+2,d} & x_{d+2} \end{vmatrix} = 0, \tag{9.12}$$

y así hasta

$$\begin{vmatrix} a_{11} & a_{12} & \dots & a_{1d} & x_1 \\ a_{21} & a_{22} & \dots & a_{2d} & x_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{d1} & a_{d2} & \dots & a_{dd} & x_d \\ a_{n,1} & a_{n,2} & \dots & a_{n,d} & x_n \end{vmatrix} = 0. \tag{9.13}$$

Es claro que desarrollando cada uno de estos determinantes (por su última columna) se obtiene una ecuación lineal en las incógnitas x_1, x_2, $\dots$, x_d, x_j para cada $j = d{+}1, d{+}2, \dots, n$, respectivamente. Su intersección proporciona un sistema homogéneo de $n-d$ ecuaciones lineales con n incógnitas del tipo

$$\left\{ \begin{array}{cccccccccccccc} \alpha_{11}x_1 & + & \dots & + & \alpha_{1d}x_d+ & \alpha_1 x_{d+1} & + & 0 & + & \dots & + & 0 & = & 0 \\ \alpha_{21}x_1 & + & \dots & + & \alpha_{2d}x_d+ & 0 & + & \alpha_2 x_{d+2} & + & \dots & + & 0 & = & 0 \\ & & & & & & & & & & & & \vdots & \\ \alpha_{n-d,1}x_1 & + & \dots & + & \alpha_{n-d,d}x_d+ & 0 & + & 0 & + & \dots & + & \alpha_{n-d}x_n & = & 0 \end{array} \right.$$

Estas ecuaciones son (linealmente) independientes pues se observa que las últimas $n-d$ columnas son básicas (es decir, las correspondientes a las incógnitas x_{d+1}, x_{d+2}, $\dots$, x_n. Esto se debe a que $\alpha_1 = \alpha_2 = \cdots = \alpha_{n-d} \neq 0$ por coincidir con el menor no nulo de orden d anterior, como se aprecia al desarrollar por la última columna cada uno de los determinantes).

Utilizando propiedades de determinantes, es fácil ver que las coordenadas (9.10) de los vectores de S son soluciones del sistema obtenido mediante las expresiones (9.11)-(9.13). $\square$

Observación 9.10. Observar que la esencia de la demostración del Teorema 9.6 ha sido encontrar una relación entre las coordenadas $[x]_{\mathcal{B}} = (x_1, x_2, \ldots, x_n)$ de un vector genérico $x \in S$ en la base $\mathcal{B}$ de V con las coordenadas $[x]_{\mathcal{B}_S} = (\lambda_1, \lambda_2, \ldots, \lambda_d)$ del mismo vector en la base $\mathcal{B}_S$ de S.

Al particularizar el resultado a subespacios del espacio vectorial $\mathbb{K}^n$, se consigue la igualdad entre subespacios en lugar de un subesapcio isomorfo al dado.

Corolario 9.5. Todo subespacio de dimensión $1 \leq d < n$ del $\mathbb{K}$-espacio vectorial $\mathbb{K}^n$ es igual al conjunto solución de un sistema lineal homogéneo de $n - d$ ecuaciones (linealmente) independientes.

Para $1 \leq d < n$, este corolario se puede expresar como sigue:

$$S \leq \mathbb{K}^n : \dim_{\mathbb{K}}(S) = d \quad \Leftrightarrow \quad \exists A \in \mathbb{K}^{(n-d)\times n} : S = \{x \in \mathbb{K}^{n\times 1} : Ax = 0\},$$

donde las filas de A son vectores linealmente independientes en $\mathbb{K}^n$.

9.6.1. Ecuaciones paramétricas y ecuaciones cartesianas de subespacios

De la demostración del Teorema 9.6 se obtienen dos tipos de ecuaciones que son básicas para describir (o representar) a los subespacios en la práctica. En este punto es importante recordar la notación utilizada:

- V un $\mathbb{K}$-espacio vectorial de dimensión n y $\mathcal{B} = \{v_1, v_2, \ldots, v_n\}$ una base de V.

- S un subespacio vectorial de V de dimensión $1 \leq d < n$ y $\mathcal{B}_S = \{u_1, u_2, \ldots, u_d\}$ una base de S.

- Los vectores u_i escritos en la base $\mathcal{B}$ son

$$[u_1]_{\mathcal{B}} = \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{n1} \end{bmatrix}, \quad [u_2]_{\mathcal{B}} = \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{n2} \end{bmatrix}, \quad \ldots, \quad [u_d]_{\mathcal{B}} = \begin{bmatrix} a_{1d} \\ a_{2d} \\ \vdots \\ a_{nd} \end{bmatrix}.$$

- x un vector genérico de S con coordenadas $[x]_{\mathcal{B}} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$.

- En la base $\mathcal{B}_S$ se tiene $x = \lambda_1 u_1 + \lambda_2 u_2 + \cdots + \lambda_d u_d$.

- $[x]_{\mathcal{B}} = \lambda_1 [u_1]_{\mathcal{B}} + \lambda_2 [u_2]_{\mathcal{B}} + \cdots + \lambda_d [u_d]_{\mathcal{B}}$.

Definición 9.6. Sea V un $\mathbb{K}$-espacio vectorial de dimensión n y sea S un subespacio de V de dimensión $1 \leq d < n$. Las ecuaciones (9.10) del subespacio S se pueden escribir como

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \lambda_1 \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{n1} \end{bmatrix} + \lambda_2 \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{n2} \end{bmatrix} + \cdots + \lambda_d \begin{bmatrix} a_{1d} \\ a_{2d} \\ \vdots \\ a_{nd} \end{bmatrix}, \lambda_1, \lambda_2, \ldots, \lambda_d \in \mathbb{K}$$

y se llaman **ecuaciones paramétricas** de S respecto de la base $\mathcal{B}$ de V, siendo $\lambda_1, \lambda_2, \ldots, \lambda_d \in \mathbb{K}$ los parámetros[a].

[a] Estos parámetros deben variar libremente en $\mathbb{K}$ para obtener todos los vectores $x \in S$.

Es decir, las n ecuaciones anteriores dan las coordenadas de cada $x \in S$ en $\mathcal{B}$ en función de las coordenadas de x en la base $\mathcal{B}_S$ y de los vectores de $\mathcal{B}_S$ en $\mathcal{B}$. La cantidad de parámetros coincide con la dimensión de S.

En el caso extremo $d = n$, la expresión de la Definición 9.6 puede reescribirse como

$$[x]_{\mathcal{B}} = [\mathcal{B}_S]_{\mathcal{B}}[x]_{\mathcal{B}_S},$$

y proporciona las ecuaciones de cambio de base de la base $\mathcal{B}_S$ a la base $\mathcal{B}$ de V.

Definición 9.7. Sea V un $\mathbb{K}$-espacio vectorial de dimensión n y sea S un subespacio de V de dimensión $1 \leq d < n$. Las $n - d$ ecuaciones independientes obtenidas en las expresiones (9.11)-(9.13) se llaman **ecuaciones cartesianas** o **implícitas** del subespacio S respecto de la base $\mathcal{B}$ de V.

Ejemplo 9.14. *Se considera el subespacio S de $\mathbb{R}^3$ dado por las soluciones (x, y, z) del sistema lineal homogéneo*

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

Hallar unas ecuaciones paramétricas y unas cartesianas de S respecto de:

(a) la base canónica de $\mathbb{R}^3$.

(b) la base $\mathcal{B} = \{v_1 = (1, 0, 0), v_2 = (1, 1, 0), v_3 = (0, 0, 1)\}$ de $\mathbb{R}^3$.

$\triangleleft$ Las soluciones del sistema planteado son de la forma $(\alpha, \alpha, 0)$ con $\alpha \in \mathbb{R}$. Una base de S es $\mathcal{B}_S = \{u = (1, 1, 0)\}$.

(a) El propio sistema de ecuaciones lineales del enunciado que determina a S se corresponde con unas ecuaciones cartesianas con respecto a la base canónica de $\mathbb{R}^3$. Para hallar unas ecuaciones paramétricas se debe tener en cuenta que $s \in S$ si y sólo si $s = \lambda u$, con $\lambda \in \mathbb{R}$. Tomando coordenadas en la base canónica $\mathcal{C} = \{e_1 = (1, 0, 0), e_2 = (0, 1, 0), e_3 = (0, 0, 1)\}$ de $\mathbb{R}^3$

se tiene que $[s]_{\mathcal{C}} = \lambda[u]_{\mathcal{C}}$. Considerando que $[s]_{\mathcal{C}} = (x, y, z)$, se llega a que unas ecuaciones paramétricas, en la base canónica de $\mathbb{R}^3$, son

$$[S]_{\mathcal{C}} : \begin{cases} x &= \lambda \\ y &= \lambda \\ z &= 0 \end{cases}, \quad \lambda \in \mathbb{R}.$$

Geométricamente, el subespacio S se corresponde con la recta que pasa por el origen de $\mathbb{R}^3$ y tiene vector director $(1, 1, 0)$.

(b) Ahora se encontrarán ambas ecuaciones de S con respecto a la base $\mathcal{B}$ de $\mathbb{R}^3$. En primer lugar, se escribe un vector $s \in S$ en la base $\mathcal{B}$. Considerando que un vector genérico de S es de la forma $s = (\alpha, \alpha, 0)$ con $\alpha \in \mathbb{R}$, se tiene $s = 0v_1 + \alpha v_2 + 0v_3$. Si $[s]_{\mathcal{B}} = (x', y', z')$, tomando coordenadas en la base $\mathcal{B}$ se tiene

$$[s]_{\mathcal{B}} = 0[v_1]_{\mathcal{B}} + \alpha[v_2]_{\mathcal{B}} + 0[v_3]_{\mathcal{B}} = \alpha(0, 1, 0),$$

de donde se deduce que unas ecuaciones paramétricas de S en la base $\mathcal{B}$ son

$$[S]_{\mathcal{B}} : \begin{cases} x' &= 0 \\ y' &= \alpha \\ z' &= 0 \end{cases}, \quad \alpha \in \mathbb{R}. \tag{9.14}$$

Otra forma de resolver este ejemplo es la siguiente. Dado que la forma genérica de un punto cualquiera de S es $[P]_{\mathcal{C}} = \begin{bmatrix} \lambda \\ \lambda \\ 0 \end{bmatrix}$, haciendo un cambio de base, se tiene que un punto genérico de S en la base $\mathcal{B}$ se obtiene mediante

$$\begin{aligned}
[P]_{\mathcal{B}} &= ([\mathcal{B}]_{\mathcal{C}})^{-1}\,[P]_{\mathcal{C}} \\
&= \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} \lambda \\ \lambda \\ 0 \end{bmatrix} \\
&= \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \lambda \\ \lambda \\ 0 \end{bmatrix} \\
&= \begin{bmatrix} 0 \\ \lambda \\ 0 \end{bmatrix}.
\end{aligned}$$

Poniendo $[P]_{\mathcal{B}} = \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix}$ se obtiene el mismo resultado que en (9.14).

Por último, falta encontrar unas ecuaciones cartesianas de S en la base $\mathcal{B}$. Siguiendo el procedimiento de la demostración del Teorema 9.6, al ser $d = \dim(S) = 1$, el sistema buscado contendrá $n - d = 3 - 1 = 2$ ecuaciones cartesianas (linealmente) independientes. Dado que, para cada $\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} \in S$, el sistema lineal

$$\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \alpha = \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix}$$

debe tener solución (en α), por el Teorema de Rouché-Frobenius debe ser

$$\operatorname{rg}\left(\begin{bmatrix} 0 & x' \\ 1 & y' \\ 0 & z' \end{bmatrix} \right) = 1.$$

Dado que el único menor no nulo de la matriz $\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$ es $\begin{bmatrix} 1 \end{bmatrix}$ entonces, orlándolo, unas ecuaciones cartesianas de S en $\mathcal{B}$ se encuentran resolviendo los determinantes

$$\begin{vmatrix} 0 & x' \\ 1 & y' \end{vmatrix} = 0 \qquad \text{y} \qquad \begin{vmatrix} 1 & y' \\ 0 & z' \end{vmatrix} = 0,$$

obteniendo

$$[S]_{\mathcal{B}} : \begin{cases} x' & = & 0 \\ z' & = & 0 \end{cases}.$$

Si bien se ha seguido el procedimiento general para obtenerlas, estas ecuaciones cartesianas se podrían haber deducido, por simple observación, de la expresión (9.14). ▷

Ejemplo 9.15. *En el $\mathbb{R}$-espacio vectorial $\mathbb{R}^{2\times 2}$ se considera el subespacio*

$$S = \left\{ \begin{bmatrix} x & y \\ z & t \end{bmatrix} : \begin{bmatrix} 1 & 0 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} x & y \\ z & t \end{bmatrix} = \begin{bmatrix} x & y \\ z & t \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 2 & -1 \end{bmatrix} \right\}.$$

Hallar unas ecuaciones paramétricas y unas cartesianas de S respecto:

(a) de la base canónica de $\mathbb{R}^{2\times 2}$.

(b) de la base

$$\mathcal{B} = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix} \right\}.$$

◁ Multiplicando las matrices de S y resolviendo el sistema lineal resultante se llega a

$$S = \left\{ \begin{bmatrix} z+t & 0 \\ z & t \end{bmatrix} : z, t \in \mathbb{R} \right\}$$

con lo que $d = \dim_{\mathbb{R}}(S) = 2$ y una base de S es

$$\mathcal{B}_S = \left\{ \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right\}.$$

(a) Las coordenadas de un vector genérico $X = \begin{bmatrix} z+t & 0 \\ z & t \end{bmatrix}$ de S en la base ordenada $\mathcal{C}$ son $[X]_{\mathcal{C}} = \begin{bmatrix} z+t \\ 0 \\ z \\ t \end{bmatrix}$. Si se denota $[X]_{\mathcal{C}} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$, unas ecuaciones paramétricas de S respecto a la base $\mathcal{C}$ vienen dadas por

$$[S]_{\mathcal{C}} : \begin{cases} x_1 &= z+t \\ x_2 &= 0 \\ x_3 &= z \\ x_4 &= t \end{cases}, \quad z, t \in \mathbb{R}. \tag{9.15}$$

Puesto que, para cada $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \in S$, el sistema lineal

$$\begin{bmatrix} 1 & 1 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} z \\ t \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \tag{9.16}$$

debe tener solución (en z y t), por el Teorema de Rouché-Frobenius debe ser

$$\operatorname{rg}\left(\begin{bmatrix} 1 & 1 & x_1 \\ 0 & 0 & x_2 \\ 1 & 0 & x_3 \\ 0 & 1 & x_4 \end{bmatrix} \right) = \operatorname{rg}\left(\begin{bmatrix} 1 & 1 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \right) = 2.$$

Dado que un menor no nulo de la primera matriz anterior es $\begin{vmatrix} 1 & 1 \\ 1 & 0 \end{vmatrix}$, entonces unas ecuaciones cartesianas se encuentran orlándolo y resolviendo los determinantes

$$\begin{vmatrix} 1 & 1 & x_1 \\ 1 & 0 & x_3 \\ 0 & 0 & x_2 \end{vmatrix} = 0 \qquad \text{y} \qquad \begin{vmatrix} 1 & 1 & x_1 \\ 1 & 0 & x_3 \\ 0 & 1 & x_4 \end{vmatrix} = 0,$$

obteniendo

$$[S]_{\mathcal{C}} : \left\{ \begin{array}{rcl} x_2 & = & 0 \\ x_1 - x_3 - x_4 & = & 0 \end{array} \right. .$$

Otra forma de encontrar unas ecuaciones cartesianas es eliminando los parámetros en las ecuaciones paramétricas anteriores. Es decir, de (9.15), es claro que $x_2 = 0$ y sumando la tercera y la cuarta ecuaciones se obtiene $x_3 + x_4 = z + t$ y puesto que la primera ecuación es $x_1 = z + t$, se llega a $x_1 = x_3 + x_4$ o equivalentemente $x_1 - x_3 - x_4 = 0$, que coinciden con las ecuaciones cartesianas halladas anteriormente.

Una tercera forma de encontrar unas ecuaciones cartesianas a partir de unas ecuaciones paramétricas consiste en aplicar el método de Gauss-Jordan para resolver el sistema lineal (9.16). En efecto,

$$\left[\begin{array}{cc|c} 1 & 1 & x_1 \\ 0 & 0 & x_2 \\ 1 & 0 & x_3 \\ 0 & 1 & x_4 \end{array}\right] \sim_f \left[\begin{array}{cc|c} 1 & 1 & x_1 \\ 1 & 0 & x_3 \\ 0 & 1 & x_4 \\ 0 & 0 & x_2 \end{array}\right] \sim_f \left[\begin{array}{cc|c} 1 & 1 & x_1 \\ 0 & -1 & x_3 - x_1 \\ 0 & 1 & x_4 \\ 0 & 0 & x_2 \end{array}\right] \sim_f$$

$$\sim_f \left[\begin{array}{cc|c} 1 & 1 & x_1 \\ 0 & 1 & x_1 - x_3 \\ 0 & 1 & x_4 \\ 0 & 0 & x_2 \end{array}\right] \sim_f \left[\begin{array}{cc|c} 1 & 0 & x_3 \\ 0 & 1 & x_1 - x_3 \\ 0 & 0 & -x_1 + x_3 + x_4 \\ 0 & 0 & x_2 \end{array}\right].$$

Es claro que al ser las dos columnas de la matriz de coeficientes linealmente independientes, debía aparecer un pivote en cada una de las columnas transformadas al calcular su forma escalonada reducida por filas. Para que el último sistema tenga solución, debe cumplirse que

$$[S]_{\mathcal{C}} : \left\{ \begin{array}{rcl} x_2 & = & 0 \\ x_1 - x_3 - x_4 & = & 0 \end{array} \right. ,$$

que son dos ecuaciones lineales independientes y proporcionan unas ecuaciones cartesianas para $[S]_{\mathcal{C}}$.

(b) Las coordenadas de un vector genérico $X = \begin{bmatrix} z+t & 0 \\ z & t \end{bmatrix}$ de S en la base ordenada $\mathcal{B}$ son $[X]_{\mathcal{B}} = \begin{bmatrix} z+t \\ 0 \\ z-t \\ t \end{bmatrix}$. Si se denota $[X]_{\mathcal{B}} = \begin{bmatrix} x'_1 \\ x'_2 \\ x'_3 \\ x'_4 \end{bmatrix}$, unas ecuaciones paramétricas de S respecto a la base $\mathcal{B}$ vienen dadas por

$$[S]_{\mathcal{B}} : \left\{ \begin{array}{rcl} x'_1 & = & z+t \\ x'_2 & = & 0 \\ x'_3 & = & z-t \\ x'_4 & = & t \end{array} \right. , \quad z, t \in \mathbb{R}. \tag{9.17}$$

Puesto que, para cada $\begin{bmatrix} x'_1 \\ x'_2 \\ x'_3 \\ x'_4 \end{bmatrix} \in S$, el sistema lineal

$$\begin{bmatrix} 1 & 1 \\ 0 & 0 \\ 1 & -1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} z \\ t \end{bmatrix} = \begin{bmatrix} x'_1 \\ x'_2 \\ x'_3 \\ x'_4 \end{bmatrix}$$

debe tener solución (en z y t), por el Teorema de Rouché-Frobenius debe ser

$$\operatorname{rg}\left(\begin{bmatrix} 1 & 1 & x_1' \\ 0 & 0 & x_2' \\ 1 & -1 & x_3' \\ 0 & 1 & x_4' \end{bmatrix}\right) = \operatorname{rg}\left(\begin{bmatrix} 1 & 1 \\ 0 & 0 \\ 1 & -1 \\ 0 & 1 \end{bmatrix}\right) = 2.$$

Al ser un menor no nulo de la primera matriz anterior el dado por

$$\begin{vmatrix} 1 & 1 \\ 1 & -1 \end{vmatrix} = -2,$$

unas ecuaciones cartesianas se encuentran orlándolo y resolviendo los determinantes

$$\begin{vmatrix} 1 & 1 & x_1' \\ 1 & -1 & x_3' \\ 0 & 0 & x_2' \end{vmatrix} = 0 \qquad \text{y} \qquad \begin{vmatrix} 1 & 1 & x_1' \\ 1 & -1 & x_3' \\ 0 & 1 & x_4' \end{vmatrix} = 0.$$

Se obtiene

$$[S]_{\mathcal{B}} : \left\{ \begin{array}{rcl} x_2' & = & 0 \\ x_1' - x_3' - 2x_4' & = & 0 \end{array} \right. .$$

Otra forma de encontrar unas ecuaciones cartesianas es eliminando los parámetros de las ecuaciones paramétricas anteriores. Es decir, de (9.17) es claro que $x_2' = 0$, y sustituyendo $x_4' = t$ en la primera y la tercera ecuaciones se tiene $x_1' = z + x_4'$ y $x_3' = z - x_4'$. Despejando z de cada ecuación, igualando y reordenando se llega a $x_1' - x_3' - 2x_4' = 0$.

▷

A continuación se prueba que la tercera forma de resolver el apartado (a) del ejemplo anterior es válida en general.

Otro método para pasar de unas ecuaciones paramétricas a unas cartesianas: Una vez obtenidas unas ecuaciones paramétricas (9.10) en la demostración del Teorema 9.6, y escritas matricialmente como

$$A\lambda = x,$$

este último sistema lineal tendrá solución (en λ) si y sólo si al aplicar el método de Gauss-Jordan no aparecen ecuaciones de la forma

$$0\lambda_1 + \cdots + 0\lambda_d = z \neq 0.$$

Dado que la matriz A tiene sus columnas linealmente independientes, se tiene que $\operatorname{rg}(A) = d$. Al calcular la forma escalonada reducida por filas R_A de A, se encuentra una matriz invertible $Q \in \mathbb{K}^{n\times n}$ tal que

$$QA = R_A = \begin{bmatrix} I_d \\ O_{(n-d)\times d} \end{bmatrix},$$

puesto que las d columnas de R_A deben contener un pivote. Luego, se llega al sistema lineal equivalente

$$R_A\lambda = QA\lambda = Qx.$$

Particionando Q en bloques de tamaños adecuados y operando se tiene

$$\begin{bmatrix} \lambda \\ O_{(n-d)\times 1} \end{bmatrix} = \begin{bmatrix} I_d \\ O_{(n-d)\times d} \end{bmatrix} \lambda = R_A\lambda = Qx = \begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix} x = \begin{bmatrix} Q_1x \\ Q_2x \end{bmatrix},$$

de donde la condición necesaria y suficiente para que el sistema $A\lambda = x$ tenga solución es

$$Q_2x = O_{(n-d)\times 1}. \tag{9.18}$$

Al ser $Q_2 \in \mathbb{K}^{(n-d)\times n}$, si se considera $x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ y se realiza el producto en (9.18), se obtienen $n-d$ ecuaciones con n incógnitas $x_1, x_2, \ldots, x_n$. Dado que estas ecuaciones surgen de las filas de la matriz invertible Q, deben ser independientes.

Este razonamiento sirve tanto como método para obtener unas ecuaciones cartesianas a partir de unas paramétricas como demostración de la segunda parte del Teorema 9.6.

En resumen, para pasar de unas ecuaciones:

- paramétricas a unas cartesianas se procede, como en los ejemplos anteriores, calculando ciertos menores, eliminando parámetros y obteniendo la cantidad requerida de ecuaciones o bien aplicando el método de Gauss-Jordan.

- cartesianas a unas paramétricas se resuelve el sistema lineal homogéneo dado. Al tratarse de un sistema compatible indeterminado ($d < n$), aparecen de forma natural los parámetros calculando la expresión de la solución general.

9.7. EJERCICIOS

(1) Probar que el conjunto $S = \{a_0 + a_1 x \in \mathbb{R}_1[x] : a_0 + 2a_1 = 0\}$ es subespacio de $\mathbb{R}_1[x]$ y hallar su dimensión. Demostrar que, como espacios vectoriales, S y $\mathbb{R}_0[x]$ son isomorfos.

(2) Establecer isomorfismos de espacios vectoriales sobre el cuerpo $\mathbb{K}$ entre los siguientes espacios: el espacio vectorial de las matrices columnas $\mathbb{K}^{m\times 1}$, el espacio cartesiano $\mathbb{K}^m$ y espacio vectorial de las matrices fila $\mathbb{K}^{1\times m}$.

(3) Encontrar espacios isomorfos a los siguientes, indicando su dimensión:

   (I) Subespacio de matrices diagonales a coeficientes reales de tamaño $n \times n$.

   (II) Subespacio de matrices triangulares superiores a coeficientes reales de tamaño $n \times n$.

   (III) Subespacio de los polinomios p a coeficientes reales de grado menor o igual que 5 (junto al polinomio nulo) tales que $p(0) = 0$.

(4) Demostrar que la composición de dos isomorfismos de $\mathbb{K}$-espacios vectoriales es un isomorfismo de $\mathbb{K}$-espacio vectorial.

(5) Indicar, justificando la respuesta, cuáles de los siguientes vectores son miembros de $S = \overline{\{\text{sen}^2(x), \cos^2(x)\}}$, considerando a S como subespacio del espacio de todas las funciones de $\mathbb{R}$ en $\mathbb{R}$:

   (a) 1.

   (b) $x + 1$.

   (c) $\cos(2x)$.

   (d) $\cos(x)$.

   Para los que lo sean, escribir sus coordenadas respecto de la base dada de S.

(6) Para el sistema

$$\begin{cases} 2x + y &= 6 \\ -x - \frac{1}{2}y &= -3 \end{cases}$$

comprobar que se verifica el teorema que permite expresar su solución general como suma de una solución particular del sistema no homogéneo más la solución general del sistema homogéneo asociado. Representar gráficamente.

(7) Resolver el siguiente sistema lineal (no homogéneo)

$$\begin{cases} x_1 + 2x_2 + 5x_3 + 2x_4 = 3 \\ \phantom{x_1 + {}} x_2 + x_3 + x_4 = 3 \\ 2x_1 + 5x_2 + 11x_3 + 5x_4 = 9 \end{cases}$$

y expresar su solución general como suma de una solución particular del sistema no homogéneo más la solución general del homogéneo asociado (es decir, esta última como una combinación lineal de elementos de una base del sistema homogéneo).

(8) Demostrar que si x_0 es una solución no trivial del sistema homogéneo $Ax = 0$ con $A \in \mathbb{R}^{n\times n}$ entonces el sistema admite infinitas soluciones no triviales. ¿Es cierto el mismo resultado si A tiene coeficientes complejos? ¿Y si los coeficientes de A son elementos de $\mathbb{Z}_2$?

(9) Considerar el vector $u = (1, 3, -2)^t$ y las bases de $\mathbb{R}^3$ siguientes:

$$\mathcal{B} = \{b_1 = (1, 0, 0)^t,\, b_2 = (1, 1, 0)^t,\, b_3 = (1, 1, 1)^t\}$$

y

$$\mathcal{B}' = \{b'_1 = (-1, 1, 2)^t,\, b'_2 = (1, 1, 0)^t,\, b'_3 = (1, 1, 1)^t\}.$$

(a) Hallar las coordenadas del vector u con respecto a la base $\mathcal{B}$.

(b) Calcular las coordenadas del vector u con respecto a la base $\mathcal{B}'$ usando:

(I) las coordenadas del vector u en la base canónica.

(II) las coordenadas del vector u en la base $\mathcal{B}$.

(10) Sean $V = \mathbb{R}_2[x]$ y $a \in \mathbb{R}$ un número fijo no nulo.

(a) Probar que $\mathcal{B} = \{1, x+a, (x+a)^2\}$ es una base de V.

(b) Hallar $[\mathcal{B}]_{\mathcal{C}}$ siendo $\mathcal{C}$ la base canónica de V.

(c) Escribir el vector $p(x) = 1 + 2x + 3x^2$ en la base $\mathcal{B}$. Realizarlo de dos formas distintas.

(d) Para $a = 0$, ¿coinciden los resultados obtenidos en los apartados anteriores?

(e) Calcular el polinomio de Taylor[8] de p de grado 2 alrededor del punto $x_0 = 1$ y comparar con el resultado del apartado (c) para el caso $a = -1$.

(11) Sea S el subespacio de $\mathbb{R}^3$ dado por las soluciones (x, y, z) del sistema homogéneo $x + y - z = 0$. Hallar unas ecuaciones paramétricas y unas cartesianas de S respecto de la base:

(a) canónica de $\mathbb{R}^3$.

(b) $\mathcal{B} = \{v_1 = (1, 0, 0), v_2 = (1, 1, 0), v_3 = (1, 1, 1)\}$.

(12) Se considera el espacio vectorial $\mathbb{R}_4[x]$ de los polinomios a coeficientes reales de grado menor o igual que 4 junto al polinomio nulo y el subespacio $S = \{p(x) \in \mathbb{R}_4[x] : p(x) = p(-x)\}$. Hallar unas ecuaciones paramétricas y unas cartesianas de S:

(a) en la base $\mathcal{C} = \{1, x, x^2, x^3, x^4\}$.

(b) en la base $\mathcal{B} = \{1, x-a, (x-a)^2, (x-a)^3, (x-a)^4\}$ para $a \in \mathbb{R} - \{0\}$.

[8] Se recuerda que el polinomio de Taylor $p_2(x)$ de una función f (suficientemente derivable) alrededor de un punto x_0 viene dado por la expresión $p_2(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2}(x - x_0)^2$.

(13) Hallar unas ecuaciones paramétricas y unas implícitas del subespacio de $\mathbb{R}^5$ definido por

$$S = \overline{\{e_2 + e_4, e_5\}}$$

en la base canónica $\mathcal{C} = \{e_1, e_2, e_3, e_4, e_5\}$ de $\mathbb{R}^5$. Repetir el ejercicio en la base $\mathcal{B}$ dada por

$$\{(1,1,1,1,1), (0,1,1,1,1), (0,0,1,1,1), (0,0,0,1,1), (0,0,0,0,1)\}.$$

(14) En el $\mathbb{R}$-espacio vectorial $\mathbb{R}^5$, encontrar unas ecuaciones paramétricas y unas ecuaciones cartesianas en la base canónica de $\mathbb{R}^5$ del siguiente subespacio

$$U = \overline{\{(1,1,1,1,1), (0,2,1,0,2), (-1,3,1,-1,3)\}}.$$

(15) En el espacio

$$\mathbb{R}_3[x] = \{p(x) = a_0 + a_1x + a_2x^2 + a_3x^3 : a_0, a_1, a_2, a_3 \in \mathbb{R}\},$$

encontrar unas ecuaciones paramétricas del siguiente subespacio, dado mediante unas ecuaciones cartesianas en la base canónica por

$$[S]_{\mathcal{C}} : \left\{ \begin{array}{ccccc} a_0 & + & 3a_2 & = & 0 \\ a_1 & - & 2a_3 & = & 0 \end{array} \right. .$$

(16) En el $\mathbb{R}$-espacio vectorial $\mathbb{R}^4$ encontrar, de tres formas diferentes, unas ecuaciones cartesianas en la base canónica del subespacio dado según las siguientes ecuaciones paramétricas en la base canónica

$$[S]_{\mathcal{C}} : \left\{ \begin{array}{ccl} x_1 & = & 0 \\ x_2 & = & \lambda - \mu \\ x_3 & = & \lambda + \mu - \gamma \\ x_4 & = & 2\lambda + \gamma \end{array} \right. , \quad \lambda, \mu, \gamma \in \mathbb{R}.$$

(17) En el $\mathbb{R}$-espacio vectorial V de dimensión 4 se considera la base $\mathcal{B} = \{u_1, u_2, u_3, u_4\}$ y los subespacios vectoriales

$$U = \overline{\{u_1 - 2u_2 - u_3, 3u_1 + u_3, 2u_2\}}$$

y W de ecuaciones cartesianas

$$[W]_{\mathcal{B}} : \left\{ \begin{array}{ccccc} a_1 & + & a_3 & = & 0 \\ a_2 & - & a_4 & = & 0 \end{array} \right. .$$

Encontrar unas ecuaciones paramétricas y unas ecuaciones cartesianas de los subespacios $U \cap W$ y $U + W$ en la base $\mathcal{B}$.

Capítulo 10

Espacios euclídeos

Índice

10.1. Introducción

La teoría de espacios vectoriales permite estudiar las relaciones de dependencia (o independencia) lineal de un conjunto de vectores no sólo en los modelos geométricos $\mathbb{VL}^2$ (espacio de vectores libres de $\mathcal{E}_2$) y $\mathbb{VL}^3$ (espacio de vectores libres de $\mathcal{E}_3$), sino en un espacio vectorial abstracto en general.

Sin embargo, en estos espacios abstractos, no se ha visto cómo abordar conceptos relativos a longitudes, distancias, ángulos y perpendicularidad de vectores, conceptos conocidos de geometría elemental en el plano y en el espacio de vectores libres desde la época de Euclides. Para ello, se introducirá una herramienta (llamada producto escalar) que permitirá analizar desde un punto de vista algebraico las construcciones que son posibles de realizar con la regla graduada y el compás desde un punto de vista geométrico.

En este capítulo se establece el concepto de espacio euclídeo que permitirá abordar formalmente dichas ideas en $\mathbb{R}$-espacios vectoriales (abstractos), principalmente de dimensión finita. En resumen, el propósito es enriquecer los espacios vectoriales añadiendo a ellos los conceptos geométricos mencionados.

A partir del módulo de un vector en el espacio de vectores libres $\mathbb{VL}^2$ (o de su espacio vectorial isomorfo $\mathbb{R}^2$) y del ángulo α que forman dos vectores (no nulos) $u, v \in \mathbb{R}^2$, se conoce la *definición geométrica* de producto escalar entre los vectores u y v, utilizada en geometría elemental y en Física, dada por:

$$\langle u, v\rangle = \|u\|\|v\| \cos(\text{áng}(u, v)).$$

De esta definición y de la existencia de raíces cuadradas en $\{0\} \cup \mathbb{R}^+$ no es difícil probar que, $\forall u, v, w \in \mathbb{R}^2 - \{(0, 0)\}$, $\forall \alpha \in \mathbb{R}$, se tiene:

- $\langle u + v, w\rangle = \langle u, w\rangle + \langle v, w\rangle$,
- $\langle \alpha u, v\rangle = \alpha\langle u, v\rangle$, para $\alpha \in \mathbb{R}$,
- $\langle u, v\rangle = \langle v, u\rangle$,

- $\langle u, u \rangle = \|u\|^2 > 0$[1], o equivalentemente $\|u\| = +\sqrt{\langle u, u \rangle}$,
- u y v son perpendiculares (es decir, forman un ángulo de $\frac{\pi}{2}$ radianes) si y sólo si $\langle u, v \rangle = 0$.

Por lo tanto, el producto escalar definido de esta forma permite caracterizar la perpendicularidad de dos vectores y definir el módulo de un vector en $\mathbb{R}^2$.

Figura 10.1: Pintura de Euclides (Fundación Cariplo).

A continuación se presentará una *definición algebraica* (axiomática) de producto escalar, es decir, mediante un conjunto y una aplicación que cumplan ciertos axiomas, como una abstracción de las cuatro primeras propiedades mencionadas y, posteriormente, la última proporcionará el concepto de ortogonalidad.

De esta forma, se podrá recuperar la fórmula geométrica presentada anteriormente y permitirá desarrollar la teoría de espacios con producto escalar sobre $\mathbb{R}$-espacios vectoriales (es decir, de forma más general que únicamente en $\mathbb{R}^2$ o $\mathbb{R}^3$).

Se generalizarán las nociones geométricas de módulo, distancia, ángulo y perpendicularidad (conocidas en $\mathbb{R}^2$ y $\mathbb{R}^3$) a espacios vectoriales como el

[1]Si se considera que $\|u\| = 0$ si y sólo si $u = 0$ se tiene que, además, $\langle u, u \rangle = 0$ si y sólo si $u = 0$.

de los polinomios, el de las matrices, el de funciones continuas, etc. aunque en algunos de ellos, estos conceptos, ya no se puedan interpretar geométricamente.

En este capítulo, el cuerpo en el que se consideran los escalares es $\mathbb{K} = \mathbb{R}$ puesto que se requiere la existencia de la raíz cuadrada de los elementos de $\mathbb{K}$. Es posible desarrollar una teoría similar sobre el cuerpo de los números complejos (es decir, utilizando $\mathbb{K} = \mathbb{C}$), que dará lugar a los *espacios unitarios* (en lugar de a los *espacios euclídeos*), pero no se realizará en este libro.

10.2. Espacios vectoriales con producto interno

Se comienza con la definición axiomática de producto escalar que extiende el concepto conocido en $\mathbb{R}^2$ y $\mathbb{R}^3$ a $\mathbb{R}$-espacios vectoriales arbitrarios. Como se ha comentado, para ello, se realiza un proceso de abstracción de las propiedades válidas en los dos casos particulares mencionados.

Definición 10.1. Sea E un $\mathbb{R}$-espacio vectorial. Un **producto escalar** o **producto interno** sobre E es una aplicación

$$\begin{array}{rcccc} \langle\cdot,\cdot\rangle & : & E\times E & \to & \mathbb{R} \\ & & (u,v) & \mapsto & \langle u,v\rangle \end{array}$$

que satisface los siguientes axiomas: $\forall u, v, w \in E$ y $\forall \alpha \in \mathbb{R}$,

(PE1) <u>Aditividad en el primer argumento</u>: $\langle u+v, w\rangle = \langle u, w\rangle + \langle v, w\rangle$,

(PE2) <u>Homogeneidad en el primer argumento</u>: $\langle \alpha u, v\rangle = \alpha\langle u, v\rangle$,

(PE3) <u>Conmutatividad o simetría</u>: $\langle u, v\rangle = \langle v, u\rangle$,

(PE4) <u>Positividad</u>: $\langle u, u\rangle > 0$ para todo $u \in E - \{0\}$.

El último axioma también se suele nombrar diciendo que el producto escalar es *estrictamente definido positivo*.

Observación 10.1. Algunos autores reservan la expresión producto escalar para el espacio vectorial $\mathbb{R}^n$ y no para un $\mathbb{R}$-espacio vectorial abstracto, otros utilizan la expresión producto interno cuando se considera el cuerpo de los números complejos. En este libro se utiliza producto escalar o producto interno de manera indistinta.

Los axiomas (PE1) y (PE2) indican que el producto escalar respeta parte de la estructura algebraica del espacio vectorial: la linealidad en el primer argumento. Más adelante, se deducirá que también respeta la linealidad en el segundo argumento, con lo cual no se requiere que ésta sea una exigencia en la definición. La validez de los axiomas (PE3) y (PE4) ha sido observada previamente en los casos de $\mathbb{R}^2$ y $\mathbb{R}^3$.

Observación 10.2. Se observa que, en el axioma relativo a la positividad del producto escalar, se está utilizando la relación de orden conocida de los números reales. Más adelante, se verá cómo este axioma resulta fundamental para definir la norma de un vector.

Definición 10.2. Se llama **espacio vectorial con producto interno**, **espacio euclídeo**[a] o **espacio vectorial euclídeo** a un par $(E, \langle \cdot, \cdot \rangle)$ donde E es un $\mathbb{R}$-espacio vectorial y $\langle \cdot, \cdot \rangle$ es un producto escalar sobre E.

[a] Algunos autores llaman espacio euclídeo a todo $\mathbb{R}$-espacio vectorial provisto de un producto escalar únicamente cuando tiene dimensión finita.

Cuando se quiere poner énfasis en que el cuerpo utilizado para los escalares es $\mathbb{R}$ se suele llamar **espacio euclídeo real**.

Observación 10.3. Es importante remarcar que, si sobre un mismo espacio vectorial se definen dos productos escalares diferentes, se obtendrán dos espacios euclídeos distintos.

Observación 10.4. Si se conoce que la aplicación

$$\begin{array}{rccc} \langle \cdot, \cdot \rangle & : & E \times E & \rightarrow & \mathbb{R} \\ & & (u, v) & \mapsto & \langle u, v \rangle \end{array}$$

es un producto escalar sobre E y que S es un subespacio del espacio vectorial E entonces la restricción

$$\begin{array}{rccc} \langle \cdot, \cdot \rangle|_S & : & S \times S & \rightarrow & \mathbb{R} \\ & & (u, v) & \mapsto & \langle u, v \rangle \end{array}$$

define un producto escalar sobre S. Su comprobación se propone como ejercicio.

10.2.1. Consecuencias de la definición

Las siguientes propiedades algebraicas se cumplen en todo espacio euclídeo (real).

Proposición 10.1. Sea $(E, \langle \cdot, \cdot \rangle)$ un espacio euclídeo real. Se cumplen las siguientes propiedades:

(a) Linealidad en el segundo argumento: $\langle u, \alpha v + \beta w \rangle = \alpha \langle u, v \rangle + \beta \langle u, w \rangle$ para todo $u, v, w \in E$ y para todo $\alpha, \beta \in \mathbb{R}$.

(b) Producto escalar con el vector nulo: $\langle 0, u \rangle = 0$, para todo $u \in E$.

(c) Definición positiva: Sea $u \in E$. Entonces $\langle u, u \rangle = 0$ si y sólo si $u = 0$.

Demostración. Sean $u, v, w \in E$ y $\alpha, \beta \in \mathbb{R}$. La demostración sigue de aplicar los axiomas del producto escalar. En efecto,

(a) Aplicando adecuadamente (PE1), (PE2) y (PE3) se obtiene

$$\begin{aligned}\langle u, \alpha v + \beta w\rangle &= \langle \alpha v + \beta w, u\rangle \\ &= \alpha\langle v, u\rangle + \beta\langle w, u\rangle \\ &= \alpha\langle u, v\rangle + \beta\langle u, w\rangle.\end{aligned}$$

(b) Teniendo en cuenta que 0 es el elemento neutro de la adición de vectores,

$$\begin{aligned}\langle 0, u\rangle &= \langle 0 + 0, u\rangle \\ &= \langle 0, u\rangle + \langle 0, u\rangle.\end{aligned}$$

Sumando a ambos miembros de la igualdad el escalar $-\langle 0, u\rangle$, asociando y cancelando, se tiene que $\langle 0, u\rangle = 0$.

(c) Si $u = 0$, por la propiedad (b) anterior, es claro que $\langle u, u\rangle = 0$. Recíprocamente, si $\langle u, u\rangle = 0$ se debe probar que $u = 0$. En efecto, si fuese $u \neq 0$, al ser la aplicación $\langle \cdot, \cdot\rangle$ positiva, se tendría que $\langle u, u\rangle > 0$, en contra de la hipótesis. Luego, debe ser $u = 0$.

$\square$

Observación 10.5. De la positividad y de, su consecuencia, la definición positiva, se tiene que un producto escalar cumple que:

- $\langle u, u\rangle \geq 0$, $\forall u \in E$ y, además,
- si $\langle u, u\rangle = 0$ entonces $u = 0$.

Es decir, $\langle u, u\rangle$ toma siempre un valor no negativo, y la única posibilidad de que sea igual a cero es cuando $u = 0$.

Toda aplicación $\langle \cdot, \cdot\rangle$ que es aditiva y homogénea en cada uno de los (dos) argumentos se llama **aplicación bilineal**.

Teniendo en cuenta la Definición 10.1 y las consecuencias dadas en la Proposición 10.1, algunos autores definen el **producto interior** sobre un $\mathbb{R}$-espacio vectorial E como toda aplicación bilineal, simétrica y definida positiva (recordando que, para ser definida positiva, debe ser previamente una aplicación positiva).

10.2.2. Ejemplos de espacios euclídeos

Ejemplo 10.1 (Producto escalar canónico en $\mathbb{R}^n$)**.** *La aplicación*

$$\langle \cdot, \cdot \rangle : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$$

definida por

$$\langle (x_1, x_2, \ldots, x_n), (y_1, y_2, \ldots, y_n) \rangle = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n = \sum_{k=1}^{n} x_k y_k$$

es un producto escalar sobre $\mathbb{R}^n$ *llamado* **producto escalar canónico de** $\mathbb{R}^n$. *El espacio vectorial* $\mathbb{R}^n$ *con el producto escalar canónico se denomina* **espacio euclídeo usual** $\mathbb{R}^n$.

$\lhd$ Si bien es posible comprobar la validez de los axiomas utilizando la definición anterior, se observa que si se denota[2] por

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \qquad \text{e} \qquad y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

entonces el producto escalar definido en el enunciado se escribe de forma más

[2] Es un abuso en la notación; en realidad las matrices columna consideradas son las coordenadas de los vectores x e y en la base canónica de $\mathbb{R}^n$. Además, mientras que $\langle x, y \rangle \in \mathbb{R}$, se tiene que $x^t y \in \mathbb{R}^{1\times 1}$, y puesto que los espacios vectoriales $\mathbb{R}$ y $\mathbb{R}^{1\times 1}$ son isomorfos, se permite realizar dicha identificación escribiéndola como una igualdad.

compacta como un producto matricial mediante

$$\langle x, y \rangle = x^t y.$$

Ahora resulta más sencillo realizar la demostración de los axiomas (PE1)-(PE3); para probar (PE4) será necesario recurrir a las componentes del vector. En efecto, sean $x, y, z \in \mathbb{R}^n$ y $\alpha \in \mathbb{R}$. Utilizando propiedades de la traspuesta de una matriz se tiene:

(PE1) $\langle x + y, z \rangle = (x + y)^t z = (x^t + y^t)z = x^t z + y^t z = \langle x, z \rangle + \langle y, z \rangle$.

(PE2) $\langle \alpha x, y \rangle = (\alpha x)^t y = (\alpha x^t)y = \alpha(x^t y) = \alpha \langle x, y \rangle$.

(PE3) $\langle x, y \rangle = x^t y = (x^t y)^t = y^t (x^t)^t = y^t x = \langle y, x \rangle$, donde se ha utilizado que el escalar real $\langle x, y \rangle$ coincide con su traspuesta.

(PE4) $\langle x, x \rangle = x^t x = x_1^2 + x_2^2 + \cdots + x_n^2 \geq 0$ para todo $x \in \mathbb{R}^n$ y es claro que si $x \neq 0$ entonces debe ser $x_i \neq 0$ para algún $i \in \{1, 2, \ldots, n\}$. Luego, $x_i^2 > 0$ para ese valor del subíndice i y $x_j^2 \geq 0$ para los restantes valores del subíndice j; y así, $x^t x > 0$, para todo $x \in \mathbb{R}^n - \{0\}$.

Por lo tanto, $\langle \cdot, \cdot \rangle$ es un producto escalar sobre $\mathbb{R}^n$. $\triangleright$

Ejemplo 10.2 (Otro producto escalar en $\mathbb{R}^2$)**.** *La aplicación*

$$\langle \cdot, \cdot \rangle : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$$

definida por

$$\langle (x_1, x_2), (y_1, y_2) \rangle = 2x_1 y_1 + 3x_2 y_2$$

es un producto escalar sobre $\mathbb{R}^2$.

$\triangleleft$ Sean $x, y, z \in \mathbb{R}^2$ y $\alpha \in \mathbb{R}$. Se tiene que:

(PE1) $\langle x + y, z \rangle = \langle (x_1 + y_1, x_2 + y_2), (z_1, z_2) \rangle = 2(x_1 + y_1)z_1 + 3(x_2 + y_2)z_2 = 2x_1 z_1 + 2y_1 z_1 + 3x_2 z_2 + 3y_2 z_2 = (2x_1 z_1 + 3x_2 z_2) + (2y_1 z_1 + 3y_2 z_2) = \langle x, z \rangle + \langle y, z \rangle$.

(PE2) $\langle \alpha x, y\rangle = \langle(\alpha x_1, \alpha x_2), (y_1, y_2)\rangle = 2(\alpha x_1)y_1 + 3(\alpha x_2)y_2 = \alpha(2x_1y_1 + 3x_2y_2) = \alpha\langle x, y\rangle$.

(PE3) $\langle x, y\rangle = 2x_1y_1 + 3x_2y_2 = 2y_1x_1 + 3y_2x_2 = \langle y, x\rangle$.

(PE4) $\langle x, x\rangle = 2x_1^2 + 3x_2^2 \geq 0$ para todo $x \in \mathbb{R}^2$. Si $x \neq 0$ entonces $x_1 \neq 0$ ó $x_2 \neq 0$. Si fuese $x_1 \neq 0$ entonces $x_1^2 > 0$ y $x_2^2 \geq 0$, con lo que $\langle x, x\rangle = 2x_1^2 + 3x_2^2 > 0$. Un razonamiento similar es válido si $x_2 \neq 0$.

Por lo tanto, $\langle\cdot,\cdot\rangle$ es un producto escalar sobre $\mathbb{R}^2$. ▷

Observar que si se denota

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \qquad \text{e} \qquad y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

entonces el producto escalar definido en el ejemplo anterior se puede escribir de forma más compacta como una multiplicación matricial mediante

$$\langle x, y\rangle = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}.$$

Como ejercicio, se propone repetir la demostración utilizando esta expresión. Más adelante, se volverá sobre esta forma matricial para representar los productos escalares.

Ejemplo 10.3 (Un producto escalar en el espacio vectorial de las funciones continuas)**.** *Sea $a < b$. La aplicación*

$$\langle\cdot,\cdot\rangle : \mathcal{C}[a,b] \times \mathcal{C}[a,b] \to \mathbb{R}$$

definida por

$$\langle f, g\rangle = \int_a^b f(x)g(x)\,\mathrm{d}x$$

es un producto escalar sobre el $\mathbb{R}$-espacio vectorial $\mathcal{C}[a,b]$ definido por

$$\mathcal{C}[a,b] = \{f / \ f : [a,b] \to \mathbb{R} \text{ es una función continua en } [a,b]\}$$

con la adición y multiplicación de un escalar (real) por una función definidos punto a punto como en el Ejemplo 8.7 de la página 353.

$\lhd$ Para la demostración son necesarias las propiedades elementales de la integral definida. Sean $f, g, h \in \mathcal{C}[a, b]$ y $\alpha \in \mathbb{R}$. Se tiene que:

(PE1) $\langle f+g, h\rangle = \int_a^b (f+g)(x)h(x)\,\mathrm{d}x = \int_a^b [f(x)+g(x)]h(x)\,\mathrm{d}x = \int_a^b [f(x)h(x) + g(x)h(x)]\,\mathrm{d}x = \int_a^b f(x)h(x)\,\mathrm{d}x + \int_a^b g(x)h(x)\,\mathrm{d}x = \langle f, h\rangle + \langle g, h\rangle$.

(PE2) $\langle \alpha f, g\rangle = \int_a^b (\alpha f)(x)g(x)\,\mathrm{d}x = \int_a^b [\alpha f(x)]g(x)\,\mathrm{d}x = \int_a^b \alpha[f(x)g(x)]\,\mathrm{d}x = \alpha \int_a^b f(x)g(x)\,\mathrm{d}x = \alpha\langle f, g\rangle$.

(PE3) $\langle f, g\rangle = \int_a^b f(x)g(x)\,\mathrm{d}x = \int_a^b g(x)f(x)\,\mathrm{d}x = \langle g, f\rangle$.

(PE4) De la propiedad de monotonía de la integral definida, es claro que $\langle f, f\rangle = \int_a^b [f(x)]^2\,\mathrm{d}x \geq 0$ pues $[f(x)]^2 \geq 0$ para todo $x \in [a, b]$.

Falta probar que si $f \not\equiv 0$ entonces $\langle f, f\rangle \neq 0$.

En efecto, dado que la función f no es idénticamente nula, existe $x_0 \in [a, b]$ tal que $f(x_0) \neq 0$. Se supondrá, en primer lugar, que $x_0 \neq a$ y $x_0 \neq b$. Al ser $f(x_0) \neq 0$, por la ley de tricotomía, debe ocurrir que $f(x_0) > 0$ o bien $f(x_0) < 0$. Se analizará el caso $f(x_0) > 0$ puesto que el análisis del otro caso es similar y se propone como ejercicio. De la definición de función continua[3], tomando $\epsilon = \frac{f(x_0)}{2} > 0$, existe $\delta > 0$ tal que $\forall x \in [a, b]$: $|x - x_0| < \delta$ debe ser $|f(x) - f(x_0)| < \epsilon$, es decir,

$$-\frac{f(x_0)}{2} < f(x) - f(x_0) < \frac{f(x_0)}{2}, \quad \forall x \in]x_0 - \delta, x_0 + \delta[,$$

o de forma equivalente

$$0 < \frac{f(x_0)}{2} < f(x) < \frac{3}{2}f(x_0), \quad \forall x \in]x_0 - \delta, x_0 + \delta[.$$

[3]Se ha probado una propiedad de Cálculo Diferencial que se suele llamar **Teorema de permanencia de signo**: "Si $f : [a, b] \to \mathbb{R}$ es una función continua y existe $x_0 \in]a, b[$ tal que $f(x_0) > 0$ entonces existe un entorno $E(x_0)$ de x_0 tal que $f(x) > 0$, para todo $x \in E(x_0)$".

Teniendo en cuenta que $p(x) = x^2$ es una función creciente en $[0, +\infty[$ y recordando la propiedad de monotonía de la integral deinida, las desigualdades anteriores se preservan al tomar integral definida.

Utilizando nuevamente monotonía de la integral definida[4],

$$
\begin{aligned}
\langle f, f\rangle &= \int_a^b [f(x)]^2 \,\mathrm{d}x \\
&= \int_a^{x_0-\delta} [f(x)]^2 \,\mathrm{d}x + \int_{x_0-\delta}^{x_0+\delta} [f(x)]^2 \,\mathrm{d}x + \int_{x_0+\delta}^{b} [f(x)]^2 \,\mathrm{d}x \\
&\geq \int_{x_0-\delta}^{x_0+\delta} [f(x)]^2 \,\mathrm{d}x \\
&> \int_{x_0-\delta}^{x_0+\delta} \left[\frac{f(x_0)}{2}\right]^2 \,\mathrm{d}x \\
&= \frac{[f(x_0)]^2}{4} \int_{x_0-\delta}^{x_0+\delta} \,\mathrm{d}x \\
&= \frac{[f(x_0)]^2}{4} 2\delta \\
&= \frac{\delta[f(x_0)]^2}{2} \\
&> 0.
\end{aligned}
$$

Se ha probado que si $f \not\equiv 0$ entonces $\langle f, f\rangle > 0$.

Como ejercicio, se propone modificar ligeramente la demostración para el caso en que sea $x_0 = a$ o bien $x_0 = b$.

De este modo, la aplicación dada por la integral definida proporciona un producto escalar sobre el espacio $\mathcal{C}[a, b]$, y lo convierte en un espacio euclídeo de gran importancia en las Matemáticas. $\triangleright$

[4] El valor de δ encontrado, eventualmente, se debería modificar de modo que $x_0 - \delta > a$ y $x_0 + \delta < b$.

Observación 10.6. Observar la analogía que existe entre el producto escalar canónico definido sobre el espacio vectorial $\mathbb{R}^n$ de dimensión *finita* (Ejemplo 10.1) con el producto escalar definido sobre el espacio vectorial $\mathcal{C}[a,b]$ de dimensión *infinita* (Ejemplo 10.3). Recordando la definición de integral como el límite de las sumas de Riemann, podría decirse que el segundo ejemplo se obtiene, en cierto modo, mediante un paso al límite del primero.

Utilizando una notación más adecuada ($x(i)$ en lugar de x_i), informalmente, y a falta de rigor, puede pensarse que si $a = 1$, $b = n$ y en la expresión

$$\begin{aligned} &\langle (x(1), x(2), \ldots, x(n)), (y(1), y(2), \ldots, y(n)) \rangle = \\ &= \quad x(1)y(1) + x(2)y(2) + \cdots + x(n)y(n) \\ &= \quad \sum_{k=1}^{n} x(k)y(k) \end{aligned}$$

se añaden cada vez más términos intermedios entre a y b hasta confundirse con el intervalo continuo $[a,b]$, la suma tiende a la integral

$$\langle x, y \rangle = \int_a^b x(t)y(t)\,\mathrm{d}t,$$

siendo $x, y : [a,b] \to \mathbb{R}$ funciones continuas.

Ejemplo 10.4 (El producto escalar de Frobenius en $\mathbb{R}^{n\times n}$)**.** *La aplicación*

$$\langle \cdot, \cdot \rangle_F : \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n} \to \mathbb{R}$$

definida por

$$\langle A, B \rangle_F = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} b_{ij} = \operatorname{tr}(A^t B)$$

es un producto escalar sobre $\mathbb{R}^{n\times n}$*. Se conoce como* **producto escalar de Frobenius**[a]*.*

[a] Ferdinand Georg Frobenius (1849-1917).

◁ Sean $A, B, C \in \mathbb{R}^{n\times n}$ y $\alpha \in \mathbb{R}$. Utilizando propiedades de la traza y de la traspuesta de una matriz se tiene que:

(PE1) $\langle A+B, C\rangle_F = \text{tr}((A+B)^t C) = \text{tr}((A^t+B^t)C) = \text{tr}(A^tC + B^tC) = \text{tr}(A^tC) + \text{tr}(B^tC) = \langle A, C\rangle_F + \langle B, C\rangle_F$.

(PE2) $\langle \alpha A, B\rangle_F = \text{tr}((\alpha A)^t B) = \text{tr}(\alpha(A^tB)) = \alpha\text{tr}(A^tB) = \alpha\langle A, B\rangle_F$.

(PE3) $\langle A, B\rangle_F = \text{tr}(A^tB) = \text{tr}((A^tB)^t) = \text{tr}(B^t(A^t)^t) = \text{tr}(B^tA) = \langle B, A\rangle_F$.

(PE4) $\langle A, A\rangle_F = \text{tr}(A^tA) = \sum_{i=1}^n \sum_{j=1}^n [a_{ij}]^2 \geq 0$ para toda $A \in \mathbb{R}^{n\times n}$. De esta expresión también se deduce que si $\langle A, A\rangle_F = 0$ entonces $a_{ij} = 0$ para todo i y para todo j. Luego, $A = O_{n\times n}$.

Por lo tanto, $\langle\cdot,\cdot\rangle_F$ es un producto escalar sobre $\mathbb{R}^{n\times n}$. ▷

Observación 10.7. Observar que la idea de la definición del producto escalar de Frobenius surge de forma natural del Ejemplo 10.1 y del hecho que $\mathbb{R}^{n\times n} \cong \mathbb{R}^{n^2}$ (a veces nombrado como **vectorización** de una matriz).

Se propone como ejercicio comprobar que este producto escalar se puede extender a matrices rectangulares.

10.3. Norma y distancia

A continuación se extiende el concepto conocido de *módulo* o *longitud* de un vector libre del plano ordinario $\mathcal{E}_2$ o del espacio ordinario $\mathcal{E}_3$ a un espacio euclídeo arbitrario. Posteriormente, el concepto de distancia se introducirá a partir de la definicón de norma.

10.3.1. Norma de un vector

Más precisamente, en un espacio euclídeo, se definirá la norma de un vector inducida por el producto interno definido en dicho espacio.

Definición 10.3. Sea $(E, \langle \cdot, \cdot \rangle)$ un espacio euclídeo. Se llama **norma** del vector $u \in E$ **inducida por el producto escalar** $\langle \cdot, \cdot \rangle$ a la raíz cuadrada aritmética (es decir, positiva) del producto escalar de dicho vector por sí mismo. En símbolos,

$$\|u\| = +\sqrt{\langle u, u \rangle}. \tag{10.1}$$

En estas condiciones, es claro que se puede definir la siguiente aplicación

$$\begin{array}{rcccc} \| \cdot \| & : & E & \to & \mathbb{R} \\ & & u & \mapsto & \|u\| = +\sqrt{\langle u, u \rangle}, \end{array}$$

y se llama **aplicación norma** inducida por el producto interno $\langle \cdot, \cdot \rangle$.

La expresión (10.1) está bien definida puesto que la condición[5]

$$\langle u, u \rangle \geq 0, \qquad \text{para todo } u \in E,$$

garantiza la existencia de la raíz cuadrada[6]. Además, (10.1) implica

$$\|u\|^2 = \langle u, u \rangle.$$

También, para cualquier producto escalar, del apartado (c) de la Proposición 10.1, se cumple que

$$\|0\| = 0.$$

Ejemplo 10.5. *Calcular la norma del vector $u = (3, 4)$ en el espacio euclídeo $\mathbb{R}^2$ dotado del producto escalar canónico.*

$\triangleleft$ Para un vector genérico $u = (u_1, u_2) \in \mathbb{R}^2$, se tiene que

$$\|u\| = +\sqrt{\langle u, u \rangle} = +\sqrt{\langle (u_1, u_2), (u_1, u_2) \rangle} = +\sqrt{u_1^2 + u_2^2}.$$

[5]Esta condición se deduce del axioma (PE4) y de la Proposición 10.1 (c).

[6]Se observa que el orden utilizado es el orden conocido sobre el cuerpo de los números reales.

En particular,

$$\|u\| = +\sqrt{3^2+4^2} = +\sqrt{25} = 5,$$

que es el caso en que la norma coincide con el módulo del vector conocido para vectores geométricos. ▷

De forma más general, cuando se considera el espacio euclídeo $\mathbb{R}^n$ con el *producto escalar canónico* dado por

$$\langle u, v\rangle = u_1v_1 + u_2v_2 + \cdots + u_nv_n,$$

siendo $u = (u_1, u_2, \ldots, u_n)$ y $v = (v_1, v_2, \ldots, v_n)$, la norma inducida es

$$\|u\|_2 = +\sqrt{\langle u, u\rangle} = +\sqrt{u_1^2 + u_2^2 + \cdots + u_n^2} = +\sqrt{\sum_{k=1}^{n} u_k^2},$$

y se llama **norma euclídea** o **norma 2**.

Ejemplo 10.6. *Calcular la norma del vector $u = (1, 1, 1)$ en el espacio euclídeo $\mathbb{R}^3$ provisto del producto escalar*

$$\langle u, v\rangle = \begin{bmatrix} u_1 & u_2 & u_3 \end{bmatrix} \begin{bmatrix} 6 & -1 & 1 \\ -1 & 2 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$$

donde las matrices (columna)

$$\begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} \qquad y \qquad \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$$

son, respectivamente, las coordenadas de los vectores $u = (u_1, u_2, u_3)$ y $v = (v_1, v_2, v_3)$ en la base canónica del $\mathbb{R}$-espacio vectorial $\mathbb{R}^3$.

◁ Para un vector genérico $u = (u_1, u_2, u_3) \in \mathbb{R}^3$, se tiene que

$$\begin{aligned}
\|u\| &= +\sqrt{\langle u, u\rangle} \\
&= +\sqrt{\langle (u_1, u_2, u_3), (u_1, u_2, u_3)\rangle} \\
&= +\sqrt{\begin{bmatrix} u_1 & u_2 & u_3 \end{bmatrix} \begin{bmatrix} 6 & -1 & 1 \\ -1 & 2 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}} \\
&= +\sqrt{6u_1^2 - 2u_1u_2 + 2u_1u_3 + 2u_2^2 + u_3^2}.
\end{aligned}$$

En particular, si $u = (1, 1, 1)$ se tiene

$$\|u\| = +\sqrt{6 - 2 + 2 + 2 + 1} = +\sqrt{9} = 3.$$

Se observa que la norma de este vector calculada mediante el producto escalar canónico es $\|u\| = \sqrt{3} \neq 3$. ▷

Para probar que la aplicación del ejemplo anterior es, efectivamente, un producto escalar sobre $\mathbb{R}^3$, en la página 607 se proporciona el **método de Gauss de descomposición en cuadrados** que permite analizar el axioma (PE4).

Ejemplo 10.7. *Calcular la norma del vector* $p(x) = 3x - 1$ *en el espacio euclídeo* $\mathbb{R}_1[x]$ *dotado del producto interno*

$$\langle p, q\rangle = \int_0^1 p(x)q(x)\,\mathrm{d}x. \tag{10.2}$$

◁ De la Observación 10.4, es fácil justificar que la aplicación (10.2) proporciona un producto escalar sobre el espacio vectorial $\mathbb{R}_1[x]$ puesto que la misma aplicación define un producto escalar sobre el espacio $\mathcal{C}[0, 1]$, del cual $\mathbb{R}_1[x]$ es un subespacio vectorial.

Para un vector genérico $p(x) = ax + b \in \mathbb{R}_1[x]$, se tiene que

$$\begin{aligned} \|p\| &= +\sqrt{\langle p, p\rangle} \\ &= +\sqrt{\int_0^1 [p(x)]^2 \, \mathrm{d}x} \\ &= +\sqrt{\int_0^1 (ax+b)^2 \, \mathrm{d}x} \\ &= +\sqrt{\int_0^1 (a^2x^2 + 2abx + b^2) \, \mathrm{d}x} \\ &= +\sqrt{\left[\frac{a^2}{3}x^3 + abx^2 + b^2x\right]_0^1} \\ &= +\sqrt{\frac{a^2}{3} + ab + b^2}. \end{aligned}$$

En particular, si $p(x) = 3x - 1$ se tiene

$$\|p\| = +\sqrt{\frac{3^2}{3} + 3(-1) + (-1)^2} = +\sqrt{1} = 1.$$

▷

10.3.2. Propiedades de la norma de un vector

Proposición 10.2 (Propiedades de la norma de un vector)**.** Sea $(E, \langle \cdot, \cdot \rangle)$ un espacio euclídeo. Para todo $u, v \in E$ y para todo $\alpha \in \mathbb{R}$ se cumple que:

(a) $\|u\| \geq 0$. Además, $\|u\| = 0$ si y sólo si $u = 0$.

(b) $\|\alpha u\| = |\alpha| \|u\|$.

(c) **Desigualdad de Cauchy-Schwarz**: $|\langle u, v\rangle| \leq \|u\| \|v\|$.

(d) **Desigualdad triangular**[a]: $\|u + v\| \leq \|u\| + \|v\|$.

(e) **Ley del paralelogramo**: $\|u + v\|^2 + \|u - v\|^2 = 2(\|u\|^2 + \|v\|^2)$.

(f) **Identidad de polarización**: $\langle u, v\rangle = \frac{1}{4}(\|u+v\|^2 - \|u-v\|^2)$.

[a] O bien, desigualdad de Minkowski, en honor al matemático alemán Hermann Minkowski (1864-1909).

Demostración. Sean $u, v \in E$ y $\alpha \in \mathbb{R}$.

(a) $\|u\| = +\sqrt{\langle u, u\rangle} \geq 0$. Además, teniendo en cuenta la propiedad (c) de la Proposición 10.1, $\|u\| = 0 \iff \langle u, u\rangle = 0 \iff u = 0$.

(b) De las propiedades (PE2) y el apartado (a) de la Proposición 10.1 se tiene[7]

$$\|\alpha u\| = +\sqrt{\langle \alpha u, \alpha u\rangle} = +\sqrt{\alpha^2\langle u, u\rangle} = |\alpha|\sqrt{\langle u, u\rangle} = |\alpha|\|u\|.$$

(c) Si $v = 0$, se tiene que $|\langle u, v\rangle| = 0$ y $\|v\| = 0$, de donde se sigue la igualdad $|\langle u, v\rangle| = \|u\|\|v\|$.

Para $v \neq 0$ se calcula la expresión $\|u + \lambda v\|^2$ siendo λ un parámetro real. Utilizando los axiomas del producto escalar se tiene que

$$\begin{aligned} 0 &\leq \|u + \lambda v\|^2 \\ &= \langle u + \lambda v, u + \lambda v\rangle \\ &= \langle u, u\rangle + \lambda\langle u, v\rangle + \lambda\langle v, u\rangle + \lambda^2\langle v, v\rangle \\ &= \langle u, u\rangle + 2\lambda\langle u, v\rangle + \lambda^2\langle v, v\rangle. \end{aligned} \tag{10.3}$$

Se obtiene un polinomio de segundo grado en λ (con coeficiente principal $\langle v, v\rangle$ estrictamente positivo), que debe ser no negativo, para todo $\lambda \in \mathbb{R}$. Este hecho obliga a que la ecuación

$$\langle v, v\rangle\lambda^2 + 2\langle u, v\rangle\lambda + \langle u, u\rangle = 0$$

[7] El valor absoluto que aparece se debe a la propiedad $\sqrt{x^2} = |x|$, para todo $x \in \mathbb{R}$.

tenga una raíz real doble o bien dos raíces complejas (pero no dos raíces reales distintas). Luego, su discriminante es nulo o negativo[8], de donde se tiene que

$$(2\langle u, v\rangle)^2 - 4\langle u, u\rangle\langle v, v\rangle \leq 0.$$

Despejando se llega a $(\langle u, v\rangle)^2 \leq \langle u, u\rangle\langle v, v\rangle$ y tomando raíces cuadradas en ambos miembros de la desigualdad,

$$|\langle u, v\rangle| \leq \sqrt{\langle u, u\rangle\langle v, v\rangle} = \sqrt{\langle u, u\rangle}\sqrt{\langle v, v\rangle} = \|u\|\|v\|.$$

(d) Primero se observa que al tratarse de cantidades todas no negativas, del apartado (a) de la Proposición 10.1, se cumple la equivalencia:

$$\|u + v\| \leq \|u\| + \|v\| \qquad \Longleftrightarrow \qquad \|u + v\|^2 \leq (\|u\| + \|v\|)^2.$$

Ahora, utilizando la desigualdad de Cauchy-Schwarz se tiene que[9]

$$\begin{aligned} \|u + v\|^2 &= \langle u + v, u + v\rangle \\ &= \langle u, u\rangle + \langle u, v\rangle + \langle v, u\rangle + \langle v, v\rangle \\ &= \|u\|^2 + 2\langle u, v\rangle + \|v\|^2 \\ &\leq \|u\|^2 + 2|\langle u, v\rangle| + \|v\|^2 \\ &\leq \|u\|^2 + 2\|u\|\|v\| + \|v\|^2 \\ &= (\|u\| + \|v\|)^2. \end{aligned} \tag{10.4}$$

Como se ha mencionado, teniendo en cuenta que la norma es no negativa, se obtiene la desigualdad buscada.

(e) Se propone como ejercicio (en la página 624).

(f) Se propone como ejercicio (en la página 624).

□

[8] Véase el Ejercicio 24 de la página 622.

[9] Se utiliza, además, la propiedad de valor absoluto: $x \leq |x|$, $\forall x \in \mathbb{R}$.

Es importante remarcar que, en un espacio euclídeo, la identidad de polarización permite recuperar el producto escalar a partir de su norma inducida.

Observación 10.8. En la desigualdad de Cauchy-Schwarz, la igualdad se cumple si y sólo si u y v son linealmente dependientes (Ejercicio de la página 623), mientras que en la desigualdad triangular, la igualdad se cumple si y sólo si u y v son linealmente dependientes y la constante de proporcionalidad es no negativa (Ejercicio de la página 623).

10.3.3. Distancia

Definición 10.4. Sea $(E, \langle \cdot, \cdot \rangle)$ un espacio euclídeo y sean $u, v \in E$ dos vectores. Se llama **distancia** entre[a] los vectores u y v, y se denota $d(u, v)$, a la norma de la diferencia $u - v$ (siendo dicha norma la inducida por el producto escalar de E). En símbolos,

$$d(u, v) = \|u - v\|. \tag{10.5}$$

[a] Sería más preciso llamar **distancia del vector** u **al vector** v, pero la propiedad de simetría (Proposición 10.3 (c)) de la distancia permite asegurar que coincide con la distancia del vector v al vector u. En definitiva, es correcto hablar de la distancia entre los vectores u y v.

Es claro que $d(u, 0) = \|u - 0\| = \|u\|$ para todo $u \in E$. Este hecho permite interpretar la norma del vector u como su distancia al origen.

En estas condiciones, se define la **aplicación distancia**

$$\begin{array}{rcl} d \; : \; E \times E & \rightarrow & \mathbb{R} \\ (u, v) & \mapsto & d(u, v) = \|u - v\| = +\sqrt{\langle u - v, u - v \rangle}, \end{array}$$

donde la norma es la inducida por el producto interno $\langle \cdot, \cdot \rangle$.

Ejemplo 10.8. *Determinar la expresión de la distancia en el espacio* $\mathbb{R}^n$ *con el producto escalar canónico.*

◁ En este caso se obtiene una generalización de la distancia conocida en los espacios euclídeos $\mathbb{R}^2$ y $\mathbb{R}^3$:

$$\begin{aligned} & d((u_1, u_2, \dots, u_n), (v_1, v_2, \dots, v_n)) = \\ & \quad = \quad \|(u_1, u_2, \dots, u_n) - (v_1, v_2, \dots, v_n)\| \\ & \quad = \quad \|(u_1 - v_1, u_2 - v_2, \dots, u_n - v_n)\| \\ & \quad = \quad \sqrt{(u_1 - v_1)^2 + (u_2 - v_2)^2 + \cdots + (u_n - v_n)^2}. \end{aligned}$$

▷

Ejemplo 10.9. *Encontrar la expresión de la distancia en el espacio vectorial $\mathcal{C}[a, b]$ con el producto escalar $\langle f, g\rangle = \int_a^b f(x)g(x)\,\mathrm{d}x$ y hallar $d(\operatorname{sen}(x), 0)$ en $\mathcal{C}[-\pi, \pi]$.*

◁ En este ejemplo se tiene que la distancia toma la expresión

$$d(f, g) = \|f - g\| = \sqrt{\int_a^b [f(x) - g(x)]^2\,\mathrm{d}x}.$$

En particular, por ser $\operatorname{sen}^2(x)$ una función par, en $\mathcal{C}[-\pi, \pi]$:

$$\begin{aligned} d(\operatorname{sen}(x), 0) &= \|\operatorname{sen}(x) - 0\| \\ &= \sqrt{\int_{-\pi}^{\pi} \operatorname{sen}^2(x)\,\mathrm{d}x} \\ &= \sqrt{2\int_0^{\pi} \frac{1}{2}[1 - \cos(2x)]\,\mathrm{d}x} \\ &= \sqrt{\left[x - \frac{\operatorname{sen}(2x)}{2}\right]_0^{\pi}} \\ &= \sqrt{\pi}. \end{aligned}$$

▷

10.3.4. Propiedades de la distancia

Proposición 10.3 (Propiedades de la distancia entre dos vectores)**.** Sea $(E, \langle \cdot, \cdot \rangle)$ un espacio euclídeo real. Para todo $u, v, w \in E$ se cumple que:

(a) $d(u, v) \geq 0$.

(b) $d(u, v) = 0 \Longleftrightarrow u = v$.

(c) $d(u, v) = d(v, u)$ (Propiedad de simetría).

(d) $d(u, w) \leq d(u, v) + d(v, w)$ (Desigualdad triangular).

Demostración. Los dos primeros apartados siguen directamente de la definición de distancia y las correspondientes propiedades para la norma.

El tercer apartado es inmediato de

$$\begin{aligned} d(u, v) &= \|u - v\| \\ &= \|(-1)(v - u)\| \\ &= |-1|\|v - u\| \\ &= 1\|v - u\| \\ &= \|v - u\| \\ &= d(v, u). \end{aligned}$$

Por último, por la desigualdad triangular para normas, se tiene

$$\begin{aligned} d(u, w) &= \|u - w\| \\ &= \|(u - v) + (v - w)\| \\ &\leq \|u - v\| + \|v - w\| \\ &= d(u, v) + d(v, w). \end{aligned}$$

□

10.3.5. Relación entre producto escalar, norma y distancia

Si bien la norma se ha definido a partir de un producto escalar, hay otras funciones que poseen las mismas propiedades (que la norma) y no provienen de un producto escalar.

De forma más general que la Definición 10.3, es posible definir una aplicación, llamada **norma**, a partir de un $\mathbb{R}$-espacio vectorial V

$$\begin{array}{rcccc} \|\cdot\| & : & V & \to & \mathbb{R} \\ & & u & \mapsto & \|u\| \end{array}$$

donde $\|\cdot\|$ satisface los tres axiomas siguientes (véase la Proposición 10.2):

(a) $\|u\| \geq 0$ para todo $u \in V$. Además, $\|u\| = 0$ si y sólo si $u = 0$.

(b) $\|\alpha u\| = |\alpha|\|u\|$ para todo $u \in V$ y para todo $\alpha \in \mathbb{R}$.

(c) $\|u + v\| \leq \|u\| + \|v\|$, para todo $u, v \in V$.

Un espacio vectorial provisto de una norma se llama **espacio normado**[10] sobre V.

Ejemplo 10.10 (Norma 1 o métrica del taxista)**.** *Sea*

$$\|\cdot\|_1 : \mathbb{R}^2 \to \mathbb{R}$$

la aplicación definida mediante

$$\|(x, y)\|_1 = |x| + |y|.$$

Comprobar que la aplicación $\|\cdot\|_1$ es una norma y no puede ser definida a partir de un producto escalar sobre $\mathbb{R}^2$. Esta aplicación recibe el nombre de **norma 1** *o* **métrica del taxista**.

[10] Observar que estos espacios deben definirse sobre espacios vectoriales pues hace falta la adición de vectores y la multiplicación por escalares.

$\triangleleft$ Es claro que $\|(x,y)\|_1 = |x| + |y| \geq 0$, para todo $x, y \in \mathbb{R}$. Además,

$$\|(x,y)\|_1 = 0 \Leftrightarrow |x| + |y| = 0 \Leftrightarrow |x| = |y| = 0 \Leftrightarrow x = y = 0 \Leftrightarrow (x,y) = (0,0).$$

Por otra parte, si $x, y \in \mathbb{R}^2, \alpha \in \mathbb{R}$, utilizando que la función valor absoluto distribuye con respecto al producto, se tiene que

$$\begin{aligned}
\|\alpha(x,y)\|_1 &= \|(\alpha x, \alpha y)\|_1 \\
&= |\alpha x| + |\alpha y| \\
&= |\alpha||x| + |\alpha||y| \\
&= |\alpha|[|x| + |y|] \\
&= |\alpha|\|(x,y)\|_1.
\end{aligned}$$

Por último, utilizando que la función valor absoluto cumple la desigualdad triangular, se tiene que

$$\begin{aligned}
\|(x_1,y_1) + (x_2,y_2)\|_1 &= \|(x_1 + x_2, y_1 + y_2)\|_1 \\
&= |x_1 + x_2| + |y_1 + y_2| \\
&\leq |x_1| + |x_2| + |y_1| + |y_2| \\
&= [|x_1| + |y_1|] + [|x_2| + |y_2|] \\
&= \|(x_1,y_1)\|_1 + \|(x_2,y_2)\|_1,
\end{aligned}$$

para todo $(x_1,y_1), (x_2,y_2) \in \mathbb{R}^2$.

Se ha probado así que $\|\cdot\|_1$ es una norma sobre $\mathbb{R}^2$.

Es conocido que las normas inducidas por un producto escalar cumplen la identidad de polarización. Supóngase, por reducción al absurdo, que la norma 1 se define a partir de un producto escalar, es decir, $\|u\|_1 = +\sqrt{\langle u, u\rangle}$. Basta tomar los vectores $u = (1,0)$ y $v = (1,1)$ y, utilizando la identidad de polarización, se tiene que

$$\begin{aligned}
\langle u, v\rangle &= \frac{1}{4}(\|(1,0)+(1,1)\|_1^2 - \|(1,0)-(1,1)\|_1^2)\\
&= \frac{1}{4}(\|(2,1)\|_1^2 - \|(0,-1)\|_1^2)\\
&= \frac{1}{4}((|2|+|1|)^2 - (|0|+|-1|)^2)\\
&= \frac{1}{4}(9-1)\\
&= 2,
\end{aligned}$$

mientras que, procediendo de manera similar, se llega a

$$\langle 2u, v\rangle = \langle (2,0),(1,1)\rangle = 3 \neq 4 = 2\langle u, v\rangle.$$

Luego, esta norma no proviene de un producto escalar por no satisfacer la propiedad de homogeneidad.

Como ejercicio se propone comprobar, de otra manera, que esta norma no proviene de un producto escalar. Para ello, utilizar que si así fuese se debería cumplir la ley del paralelogramo y, ver que ésta no es válida en general. $\triangleright$

Si bien, en un espacio euclídeo, la distancia se ha definido a partir de una norma (y, a su vez, esta proviene de un producto escalar), hay otras funciones que poseen las mismas propiedades (que la distancia) y no provienen de una norma.

De forma más general que en la Definición 10.4, es posible definir una aplicación, llamada **distancia** o **métrica**, a partir de un conjunto X

$$\begin{array}{rccc}
d\ : & X\times X & \to & \mathbb{R}\\
 & (u,v) & \mapsto & d(u,v)
\end{array}$$

donde d satisface los cuatro axiomas (véase la Proposición 10.3):

(a) $d(u,v)\geq 0$, para todo $u,v\in X$.

(b) $d(u,v)=0 \iff u=v$, para todo $u,v\in X$.

(c) $d(u,v) = d(v,u)$, para todo $u, v \in X$.

(d) $d(u,w) \leq d(u,v) + d(v,w)$, para todo $u, v, w \in X$.

Un conjunto provisto de una distancia se llama **espacio métrico**[11] sobre X.

Ejemplo 10.11 (Distancia discreta)**.** *Sea E un espacio vectorial real y se considera la aplicación $d : E \times E \to \mathbb{R}$ definida por*

$$d(u,v) = \begin{cases} 0, & \textit{si } u = v \\ 1, & \textit{si } u \neq v \end{cases}$$

Comprobar que d es una aplicación distancia y no puede ser definida a partir de una norma sobre E.

$\lhd$ De la propia definición es fácil ver que $\mathrm{d}(u,v) \geq 0$ (pues sólo toma los valores 0 y 1) para todo $u, v \in E$ y que $d(u,v) = 0 \iff u = v$.

Además, si $u = v$ se tiene que $d(u,v) = 0 = d(v,u)$ y si $u \neq v$ se tiene que $d(u,v) = 1 = d(v,u)$.

Por último, se analizan dos casos:

- Si $u = w$ entonces $d(u,w) = 0 \leq d(u,v) + d(v,w)$, para todo $v \in E$.
- Si $u \neq w$, se tiene que $d(u,w) = 1$ y, en este caso, debe ser $u \neq v$ o $v \neq w$ (pues si fuesen $u = v$ y $v = w$ se tendría la contradicción $u = w$); así es válida al menos una de las dos expresiones $d(u,v) = 1$ o bien $d(v,w) = 1$. Luego, $d(u,w) = 1 \leq d(u,v) + d(v,w)$.

Se ha probado que la aplicación d es una distancia[12].

Es sabido que si una distancia proviene de una norma $\|\cdot\|$, debe satisfacer la igualdad $d(u,v) = \|u - v\|$. Sin embargo, si se toman $u, v \in E$ con $u \neq v$

[11] Observar la generalidad de estos espacios, tan sólo basta un conjunto X para definirlos, ni siquiera es necesario un espacio vectorial. Fue el matemático francés Maurice Fréchet (1878-1973), quien en 1905 introdujo el concepto de espacio métrico.

[12] Se llama **distancia discreta** sobre E.

se tiene que $d(u,v)=1$ y, además,

$$1=d(2u,2v)=\|2u-2v\|=2\|u-v\|=2d(u,v)=2,$$

pues $2u\neq 2v$. Este absurdo proviene de suponer que la distancia es inducida por una norma; por lo tanto, no lo es. $\triangleright$

10.4. Ángulo de dos vectores

Sea $(E,\langle\cdot,\cdot\rangle)$ un espacio euclídeo real. Para todo $u,v\in E$, de las propiedades del valor absoluto[13], la desigualdad de Cauchy-Schwarz

$$|\langle u,v\rangle|\leq\|u\|\|v\|$$

se puede escribir como

$$-\|u\|\|v\|\leq\langle u,v\rangle\leq\|u\|\|v\|,$$

y, por lo tanto, si $u\neq 0\neq v$, se tiene que

$$-1\leq\frac{\langle u,v\rangle}{\|u\|\|v\|}\leq 1,$$

debido a que ambas normas son estrictamente positivas. En Cálculo Diferencial se prueba que la función

$$\cos:[0,\pi]\to[-1,1]$$

es biyectiva, por tanto, admite función inversa. Este hecho asegura que para cada par de vectores no nulos u y v, existe un único valor $\theta\in[0,\pi]$ tal que

$$\cos(\theta)=\frac{\langle u,v\rangle}{\|u\|\|v\|}.$$

[13] Se utiliza la propiedad: "Si $k\in\mathbb{R}$ satisface que $k\geq 0$ entonces $|x|\leq k\iff -k\leq x\leq k$", particularizada al caso $k:=\|u\|\|v\|$ y $x:=\langle u,v\rangle$.

Definición 10.5. Sea $(E, \langle\cdot,\cdot\rangle)$ un espacio euclídeo y sean $u, v \in E$ dos vectores no nulos. Se llama **ángulo**[a] entre[b] los vectores u y v, y se denota $\theta = \text{áng}(u, v)$, al único número real θ que satisface:

$$0 \leq \theta \leq \pi \qquad \text{y} \qquad \cos(\theta) = \frac{\langle u, v\rangle}{\|u\|\|v\|}. \tag{10.6}$$

[a]Es un abuso del lenguaje llamarlo ángulo. Al calcular θ, lo que realmente se quiere expresar es la medida de dicho ángulo.

[b]Sería más preciso llamar **ángulo del vector** u **con el vector** v, pero la propiedad de simetría del producto escalar permite asegurar que coincide con el ángulo entre el vector v con el vector u. En definitiva, es correcto hablar del ángulo entre los vectores u y v; se trata de ángulos no orientados.

Observación 10.9. Como era de esperar, esta definición de ángulo, considerada en un espacio euclídeo real abstracto, coincide con la noción de ángulo conocida en el espacio euclídeo $\mathbb{R}^2$ con el producto escalar canónico. En efecto, dibujando un par de ejes perpendiculares OX y OY y vectores $u = (u_1, u_2)$ y $v = (v_1, v_2)$, y marcando los ángulos $\alpha = \text{áng}(u, OX)$ y $\beta = \text{áng}(v, OX)$ con $\alpha < \beta$, se aprecia que

$$u_1 = \|u\|\cos(\alpha), \; u_2 = \|u\|\operatorname{sen}(\alpha), \; v_1 = \|v\|\cos(\beta), \; v_2 = \|v\|\operatorname{sen}(\beta).$$

Recordando la expresión del producto escalar canónico de $\mathbb{R}^2$ en términos de sus componentes, al sustituir estos valores en (10.6), se tiene que $\cos(\theta)$ es igual a

$$\frac{\langle u, v\rangle}{\|u\|\|v\|} = \frac{u_1v_1 + u_2v_2}{\|u\|\|v\|} = \cos(\alpha)\cos(\beta) + \operatorname{sen}(\alpha)\operatorname{sen}(\beta) = \cos(\beta - \alpha),$$

con lo que el valor θ definido en (10.6) coincide con $\beta - \alpha$, que es el ángulo (más pequeño de los dos posibles, es decir $0 \leq \beta - \alpha \leq \pi$) formado por las representaciones geométricas de los vectores u y v.

Utilizando el concepto de ángulo, el producto escalar de los vectores u y

v se puede escribir como

$$\langle u, v\rangle = \|u\|\|v\|\cos(\theta),$$

en función del módulo de dichos vectores y el ángulo que ellos forman. Se observa que, además, esta última fórmula se cumple vacuamente, para cualquier valor de θ, cuando alguno de los vectores es nulo. En este caso, la arbitrariedad de θ no permite definir (unívocamente) el ángulo entre dos vectores cuando al menos uno de ellos es nulo.

10.4.1. Propiedades del ángulo de dos vectores

Proposición 10.4 (Propiedades del ángulo entre dos vectores)**.** Sea $(E, \langle\cdot,\cdot\rangle)$ un espacio euclídeo. Para todo $u, v \in E$ no nulos se tiene:

(a) $\text{áng}(u, v) = \text{áng}(v, u)$ (Propiedad de simetría).

(b) Si $\alpha, \beta \in \mathbb{R}$ son tales que $\alpha\beta > 0$ entonces $\text{áng}(\alpha u, \beta v) = \text{áng}(u, v)$.

(c) $\text{áng}(-u, -v) = \text{áng}(u, v)$.

(d) $\text{áng}(u, v) + \text{áng}(-u, v) = \pi$.

Demostración. Las tres primeras propiedades se deducen directamente de la definición de ángulo, de las propiedades del producto escalar y de las propiedades de la norma. Se proponen como ejercicio.

Para demostrar la propiedad (d), si se llama $\beta = \text{áng}(u, v)$ y $\gamma = \text{áng}(-u, v)$ se tiene que $0 \leq \beta, \gamma \leq \pi$ y, además,

$$\cos(\beta) + \cos(\gamma) = \frac{\langle u, v\rangle}{\|u\|\|v\|} + \frac{\langle -u, v\rangle}{\|u\|\|v\|} = \frac{\langle u, v\rangle}{\|u\|\|v\|} - \frac{\langle u, v\rangle}{\|u\|\|v\|} = 0.$$

Luego, $0 = \cos(\beta) + \cos(\gamma) = 2\cos\left(\frac{\beta+\gamma}{2}\right)\cos\left(\frac{\beta-\gamma}{2}\right)$. De aquí se tienen dos posibilidades:

- $\cos\left(\frac{\beta+\gamma}{2}\right) = 0$. Como $0 \leq \frac{\beta+\gamma}{2} \leq \pi$, en este caso se tiene $\frac{\beta+\gamma}{2} = \frac{\pi}{2}$, es decir, $\beta + \gamma = \pi$. De esta forma, $\text{áng}(u, v) + \text{áng}(-u, v) = \pi$.

- $\cos\left(\frac{\beta-\gamma}{2}\right) = 0$. Como $-\frac{\pi}{2} \leq \frac{\beta-\gamma}{2} \leq \frac{\pi}{2}$, en este caso se tiene $\frac{\beta-\gamma}{2} = \pm\frac{\pi}{2}$, es decir, $\beta - \gamma = \pm\pi$. De nuevo, hay dos posibilidades:
  - Si $\beta - \gamma = \pi$ (o bien $\beta = \gamma + \pi$), debe ser $\gamma = 0$ y $\beta = \pi$, con lo cual $\text{áng}(u, v) + \text{áng}(-u, v) = \pi$.
  - Si $\beta - \gamma = -\pi$ (es decir, $\beta = \gamma - \pi$), debe ser $\gamma = \pi$ y $\beta = 0$, con lo cual $\text{áng}(u, v) + \text{áng}(-u, v) = \pi$.

□

Ejemplo 10.12. *Calcular (la medida de) el ángulo entre los vectores*

$$p(x) = 3x \qquad y \qquad q(x) = x^2 + 1$$

en el espacio euclídeo $\mathcal{C}[-1, 1]$ *dotado del producto interno*

$$\langle f, g\rangle = \int_{-1}^{1} f(x)g(x)\,\mathrm{d}x.$$

◁ Primero se calculan $\|p\|$, $\|q\|$ y $\langle p, q\rangle$:

$$\|p\| = +\sqrt{\int_{-1}^{1}[3x]^2\,\mathrm{d}x} = +\sqrt{\int_{-1}^{1} 9x^2\,\mathrm{d}x} = +\sqrt{[3x^3]_{-1}^{1}} = +\sqrt{6},$$

$$\begin{aligned}
\|q\| &= +\sqrt{\int_{-1}^{1}[x^2+1]^2\,\mathrm{d}x} \\
&= +\sqrt{\int_{-1}^{1}[x^4+2x^2+1]\,\mathrm{d}x} \\
&= +\sqrt{\left[\frac{1}{5}x^5 + \frac{2}{3}x^3 + x\right]_{-1}^{1}} \\
&= +\sqrt{\left[\left(\frac{1}{5} + \frac{2}{3} + 1\right) - \left(-\frac{1}{5} - \frac{2}{3} - 1\right)\right]_{-1}^{1}} \\
&= +\sqrt{\frac{56}{15}},
\end{aligned}$$

y

$$\langle p, q\rangle = \int_{-1}^{1} 3x(x^2+1)\,\mathrm{d}x = 0,$$

puesto que el producto de una función par por una impar es una función impar y el intervalo donde se evalúa la integral es simétrico. Observar que también se podría utilizar simetría para calcular las dos primeras integrales y operar en el intervalo $[0, 1]$, lo que simplificaría los cálculos.

De la fórmula (10.6) se tiene que

$$\cos(\theta) = \frac{\langle p, q\rangle}{\|p\|\|q\|} = 0,$$

de donde, el ángulo buscado es $\theta = \frac{\pi}{2}$. $\triangleright$

Observando el ejemplo anterior se aprecia que es posible introducir, de forma semejante a lo conocido en el plano y espacio ordinarios, el concepto de vectores perpendiculares en un espacio euclídeo arbitrario. En esta situación más general dicho concepto se denominará ortogonalidad.

10.5. Ortogonalidad

Definición 10.6. Sea $(E, \langle\cdot,\cdot\rangle)$ un espacio euclídeo y sean $u, v \in E$. Se dice que u es **ortogonal** a v, y se denota $u \perp v$, si $\langle u, v\rangle = 0$. En símbolos,

$$u \perp v \qquad \Longleftrightarrow \qquad \langle u, v\rangle = 0.$$

Ejemplo 10.13. *Indicar si en el espacio euclídeo $\mathbb{R}^2$ con el producto escalar canónico el vector $u = (1, 1)$ es ortogonal al vector $v = (0, 1)$.*

$\triangleleft$ Se debe calcular el producto escalar entre ambos vectores. Puesto que $\langle u, v\rangle = 1 \cdot 0 + 1 \cdot 1 = 1 \neq 0$, se tiene que u no es ortogonal a v (o bien, como se verá enseguida, u y v no son ortogonales). De hecho,

$$\cos(\theta) = \frac{\langle u, v\rangle}{\|u\|\|v\|} = \frac{1}{\sqrt{2}\cdot 1} = \frac{1}{\sqrt{2}},$$

de donde u y v forman un ángulo de $\theta = \frac{\pi}{4}$ radianes.

$\triangleright$

Ejemplo 10.14. *Indicar si el polinomio $p(x) = 3x$ es ortogonal al polinomio $q(x) = x^2 + 1$ en el espacio euclídeo $\mathcal{C}[-1, 1]$ con el producto interno*

$$\langle f, g \rangle = \int_{-1}^{1} f(x)g(x)\,\mathrm{d}x.$$

$\triangleleft$ Se ha visto en el Ejemplo 10.12 que su producto escalar es 0, con lo cual $p \perp q$. $\triangleright$

10.5.1. Propiedades de la ortogonalidad

Proposición 10.5 (Propiedades de la ortogonalidad de dos vectores)**.** Sea $(E, \langle \cdot, \cdot \rangle)$ un espacio euclídeo. Se cumplen las siguientes propiedades:

(a) Para todo $u, v \in E$: $u \perp v \iff v \perp u$.

(b) $0 \perp u$, para todo $u \in E$.

(c) Si $v \perp u$, para todo $u \in E$ entonces $v = 0$.

Demostración. Las pruebas de los dos primeros apartados son inmediatas de las propiedades del producto escalar y se proponen como ejercicio.

Para demostrar el tercero, si $v \perp u$ para todo $u \in E$ entonces $\langle v, u \rangle = 0$ para todo $u \in E$. En particular, se cumple para $u = v$, es decir, $\langle v, v \rangle = 0$, con lo cual $v = 0$. $\square$

De la Definición 10.6 y de la propiedad (a) de la Proposición anterior, se suele decir que u y v **son ortogonales** cuando $\langle u, v \rangle = 0$. El apartado (b) de la Proposición 10.5 dice que el vector nulo es ortogonal a todos los vectores del espacio euclídeo (incluso a sí mismo). Y el apartado (c) de la misma Proposición indica que el vector nulo es el único vector ortogonal a todos los vectores del espacio euclídeo.

El siguiente teorema es seguramente el más conocido por la mayor parte de la sociedad. Se trata del célebre Teorema de Pitágoras. Este filósofo y matemático griego (c. 569-c. 475 a. C.) se dedicó a la Geometría y la Aritmética como así también a la Música y a la Astronomía, entre otras cosas. Fundó la Escuela Pitagórica, una sociedad religiosa que se interesaba también por la Medicina y el estilo de vida mediante una visión holística, tanto, que filósofos de la actualidad lo consideran el primer vegano de la historia.

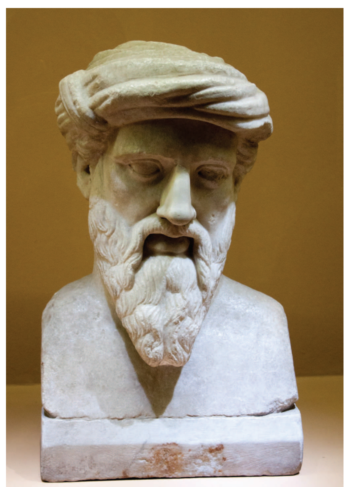


Figura 10.2: Busto de Pitágoras en los Museos Capitolinos en Roma.

Como se puede observar, impacta la sencillez de la prueba cuando se consideran los vectores como elementos de un espacio euclídeo.

Teorema 10.1 (Teorema de Pitágoras)**.** Sea $(E, \langle\cdot,\cdot\rangle)$ un espacio euclídeo real. Si $u, v \in E$ son ortogonales entonces

$$\|u+v\|^2 = \|u\|^2 + \|v\|^2.$$

Demostración. Si $u, v \in E$ son vectores ortogonales se tiene que $\langle u, v\rangle =$

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

$$\begin{aligned} \|u+v\|^2 &= \langle u+v, u+v\rangle \\ &= \langle u,u\rangle + \langle u,v\rangle + \langle v,u\rangle + \langle v,v\rangle \\ &= \|u\|^2 + \|v\|^2. \end{aligned}$$

□

10.5.2. Sistema ortogonal de vectores

Definición 10.7. Sea $(E, \langle\cdot,\cdot\rangle)$ un espacio euclídeo. Un **sistema** de vectores $S = \{u_i\}_{i\in I} \subseteq E$ no vacío (finito o infinito) se llama **ortogonal** si toda pareja de vectores distintos son ortogonales entre sí. En símbolos,

$$\langle u_i, u_j\rangle = 0 \text{ para todo } i, j \in I \text{ tales que } u_i \neq u_j.$$

De la definición anterior, es importante observar que puede ocurrir que $0 \in S$ para un sistema ortogonal S dado. Además, se cumple vacuamente que todo conjunto unitario (es decir, que contenga un sólo vector, nulo o no) es ortogonal.

Ejemplo 10.15. *El sistema*

$$\{u_1 = (0,1,1), u_2 = (3,1,-2)\}$$

es ortogonal con respecto al producto interno del Ejemplo 10.6 pero no lo es con respecto al producto escalar canónico de $\mathbb{R}^3$.

$\lhd$ Se debe probar que $\langle u_1, u_2\rangle = 0$. En efecto, con respecto a ese producto interno,

$$\langle u_1, u_2\rangle = \begin{bmatrix} 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 6 & -1 & 1 \\ -1 & 2 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \\ -2 \end{bmatrix} = 0,$$

luego $\{u_1, u_2\}$ es un sistema (o conjunto) ortogonal.

Sin embargo, si se denota $[\cdot, \cdot]$ al producto escalar canónico de $\mathbb{R}^3$, se tiene que

$$[u_1, u_2] = 0 \cdot 3 + 1 \cdot 1 + 1 \cdot (-2) = -1 \neq 0,$$

con lo cual el sistema $\{u_1, u_2\}$ no es ortogonal con respecto a este producto escalar. $\triangleright$

Ejemplo 10.16. *El sistema trigonométrico*

$$\{1, \cos(x), \operatorname{sen}(x), \cos(2x), \operatorname{sen}(2x), \dots, \cos(nx), \operatorname{sen}(nx), \dots\}$$

es un sistema ortogonal infinito con respecto al producto interno

$$\langle f, g\rangle = \int_{-\pi}^{\pi} f(x)g(x)\,\mathrm{d}x$$

definido sobre el espacio vectorial $\mathcal{C}[-\pi, \pi]$.

$\triangleleft$ Se debe comprobar que el producto escalar de todo par de vectores distintos es 0. En efecto, sean $r, s \in \mathbb{N}$. Luego,

$$\begin{aligned}
\langle 1, \cos(rx)\rangle &= \int_{-\pi}^{\pi} 1 \cdot \cos(rx)\,\mathrm{d}x \\
&= \left[\frac{\operatorname{sen}(rx)}{r}\right]_{-\pi}^{\pi} \\
&= \frac{1}{r}[\operatorname{sen}(\pi r) - \operatorname{sen}(-\pi r)] \\
&= \frac{1}{r}[0 - 0] \\
&= 0,
\end{aligned}$$

de este modo, $1 \perp \cos(rx)$, para todo $r \in \mathbb{N}$. De forma similar, se prueba que $\langle 1, \operatorname{sen}(rx)\rangle = 0$ con lo que $1 \perp \operatorname{sen}(rx)$, para todo $r \in \mathbb{N}$.

Falta ver que $\operatorname{sen}(rx) \perp \operatorname{sen}(sx)$, para todo $r \neq s$, $\cos(rx) \perp \cos(sx)$, para todo $r \neq s$ y $\operatorname{sen}(rx) \perp \cos(sx)$, para todo r, s. Estas integrales se pueden resolver recurriendo a las identidades trigonométricas

$$\operatorname{sen}(rx)\operatorname{sen}(sx) = \frac{1}{2}[\cos((r-s)x) - \cos((r+s)x)],$$

$$\cos(rx)\cos(sx) = \frac{1}{2}[\cos((r+s)x) + \cos((r-s)x)],$$

$$\operatorname{sen}(rx)\cos(sx) = \frac{1}{2}[\operatorname{sen}((r+s)x) + \operatorname{sen}((r-s)x)],$$

y los detalles de la demostración se proponen como ejercicio. $\triangleright$

Proposición 10.6 (Ortogonalidad implica independencia lineal)**.** Sea $(E, \langle\cdot,\cdot\rangle)$ un espacio euclídeo. Todo sistema ortogonal (no vacío) de vectores no nulos de E es linealmente independiente.

Demostración. Sea S un sistema de vectores no nulos (no vacío, finito o no) ortogonal con respecto al producto escalar $\langle\cdot,\cdot\rangle$. Para probar que S es linealmente independiente, se debe demostrar que cualquier subconjunto finito $\{u_1, u_2, \ldots, u_m\}$ de elementos de S (que, por ser ortogonales dos a dos, son distintos) para el cual

$$0 = \alpha_1 u_1 + \alpha_2 u_2 + \cdots + \alpha_m u_m \tag{10.7}$$

(para escalares $\alpha_j \in \mathbb{R}$ para todo $j = 1, 2, \ldots, m$) debe cumplir

$$\alpha_1 = \alpha_2 = \cdots = \alpha_m = 0.$$

En efecto, realizando el producto escalar de ambos miembros de (10.7) por uno cualquiera de ellos u_j (para cualquier $j = 1, 2, \ldots, m$) se tiene que

$$\begin{aligned}
0 &= \langle 0, u_j\rangle \\
&= \langle \alpha_1 u_1 + \cdots + \alpha_{j-1}u_{j-1} + \alpha_j u_j + \alpha_{j+1}u_{j+1} + \cdots + \alpha_m u_m, u_j\rangle \\
&= \alpha_1\langle u_1, u_j\rangle + \cdots + \alpha_{j-1}\langle u_{j-1}, u_j\rangle + \alpha_j\langle u_j, u_j\rangle + \alpha_{j+1}\langle u_{j+1}, u_j\rangle + \\
&\quad + \cdots + \alpha_m\langle u_m, u_j\rangle \\
&= \alpha_j\langle u_j, u_j\rangle,
\end{aligned}$$

puesto que, por la ortogonalidad, $\langle u_k, u_j\rangle = 0$ para todo $k \neq j$. Debido a que $\langle u_j, u_j\rangle > 0$, para todo $j = 1, 2, \dots, m$, se llega a $\alpha_j = 0$. Puesto que u_j es arbitrario, se ha probado que $\alpha_j = 0$ para todo $j = 1, 2, \dots, m$. □

Si bien el resultado anterior es evidente desde un punto de vista geométrico para espacios vectoriales como $\mathbb{R}^2$ y $\mathbb{R}^3$ con el producto escalar canónico, no es evidente en espacios euclídeos en general. Como caso concreto se tiene el siguiente ejemplo.

Ejemplo 10.17. *El $\mathbb{R}$-espacio vectorial $\mathcal{C}[-\pi, \pi]$ tiene dimensión infinita.*

◁ Aplicando la Proposición 10.6, el sistema trigonométrico del Ejemplo 10.16 proporciona un sistema con un número infinito de vectores linealmente independientes en el espacio vectorial $\mathcal{C}[-\pi, \pi]$, de donde se concluye que el espacio vectorial $\mathcal{C}[-\pi, \pi]$ tiene dimensión infinita (es decir, su dimensión no es finita). ▷

Observación 10.10. Una observación importante es que el recíproco del resultado anterior no es cierto. Por ejemplo, con respecto al producto escalar canónico de $\mathbb{R}^2$, es fácil comprobar que $u_1 = (1, 0)$ y $u_2 = (1, 1)$ son vectores linealmente independientes pero no son ortogonales.

Corolario 10.1. Sea $(E, \langle\cdot,\cdot\rangle)$ un espacio euclídeo tal que $\dim_{\mathbb{R}}(E) = n \geq 1$. Si $\{u_1, u_2, \dots, u_m\}$ es un sistema ortogonal de vectores no nulos de E entonces $m \leq n$.

Demostración. De la Proposición 10.6, se tiene que $\{u_1, u_2, \dots, u_m\}$ es linealmente independiente. El Corolario 8.2 de la página 414 asegura que $m \leq n = \dim_{\mathbb{R}}(E)$. □

Este corolario indica que, en un espacio euclídeo E de dimensión finita, el número de direcciones mutuamente ortogonales (a veces llamado **dimen-**

sión geométrica) no puede exceder la **dimensión algebraica** (número de vectores de una base cualquiera).

Se ha visto en la página 344 que a toda base del plano $\mathbb{R}^2$ se le puede asociar un sistema de coordenadas. Ese procedimiento se puede generalizar a un espacio vectorial V de dimensión finita n. Sea $\{u_1, u_2, \ldots, u_n\}$ una base de V y $P_1, P_2, \ldots, P_n$ sus puntos correspondientes (es decir, los extremos de los vectores u_i, $i = 1, 2, \ldots, n$). Si se toman las rectas $OP_1, OP_2, \ldots, OP_n$ (extendidas indefinidamente en ambos sentidos) como los ejes coordenados y $P_1, P_2, \ldots, P_n$ como los puntos de referencia a lo largo de esos ejes, queda establecido un **sistema de coordenadas cartesiano** de V.

10.5.3. Bases ortonormales

Definición 10.8. Sea $(E, \langle\cdot,\cdot\rangle)$ un espacio euclídeo. Un **vector** $u \in E$ se llama **unitario** o **normalizado** si $\|u\| = 1$, o equivalentemente $\langle u, u\rangle = 1$. Un sistema ortogonal formado por vectores unitarios se llama **sistema ortonormal**.

En general, dado un vector no nulo $x \in E$, es posible construir un vector normalizado multiplicándolo por el inverso (multiplicativo) de su norma:

$$\left\| \frac{1}{\|x\|} x \right\| = \left| \frac{1}{\|x\|} \right| \|x\| = \frac{1}{\|x\|} \|x\| = 1.$$

Definición 10.9. Sea $(E, \langle\cdot,\cdot\rangle)$ un espacio euclídeo de dimensión finita. Se llama **base ortogonal** o **base ortonormal** a toda base $\mathcal{B}$ de E que sea un sistema ortogonal u ortonormal de vectores, respectivamente.

Recordando que la delta de Kronecker se define como

$$\delta_{ij} = \begin{cases} 1, & \text{si } i = j \\ 0, & \text{si } i \neq j \end{cases},$$

si $\dim_{\mathbb{R}}(E) = n \geq 1$ y $\mathcal{B} = \{u_1, u_2, \ldots, u_n\}$ es una base de E, en símbolos se tiene que

- $\mathcal{B}$ es base ortogonal $\iff$ $\langle u_i, u_j \rangle = 0, \forall i, j \in \{1, 2, \ldots, n\} : i \neq j$.
- $\mathcal{B}$ es base ortonormal $\iff$ $\langle u_i, u_j \rangle = \delta_{ij}, \forall i, j \in \{1, 2, \ldots, n\}$.

Ejemplo 10.18. *La base canónica de $\mathbb{R}^n$ es una base ortonormal con respecto al producto escalar canónico.*

$\lhd$ Es inmediato comprobar que $\langle e_i, e_j \rangle = 0$ para todo $i, j = 1, 2, \ldots, n$ con $i \neq j$ y que $\langle e_i, e_i \rangle = 1$ para todo $i = 1, 2, \ldots, n$. $\rhd$

Ejemplo 10.19. *En el espacio vectorial $\mathbb{R}^{n\times n}$ se considera el producto interno de Frobenius (véase Ejemplo 10.4 de la página 534). El conjunto $\{E^{(i,j)}\}_{\substack{1 \leq i \leq n \\ 1 \leq j \leq n}}$ es una base ortonormal de $\mathbb{R}^{n\times n}$, siendo $E^{(i,j)}$ la matriz cuyo único elemento no nulo es un 1 en la fila i-ésima y la columna j-ésima.*

$\lhd$ Es conocido que el conjunto dado por $\{E^{(i,j)}\}_{\substack{1 \leq i \leq n \\ 1 \leq j \leq n}}$ es una base de $\mathbb{R}^{n\times n}$. Para que sea base ortonormal se debe comprobar que

$$\langle E^{(i,j)}, E^{(r,s)} \rangle = 1, \qquad \text{cuando } i = r \text{ y } j = s$$

y

$$\langle E^{(i,j)}, E^{(r,s)} \rangle = 0, \qquad \text{en el resto de los casos.}$$

Ambas situaciones se pueden escribir en una única expresión mediante

$$\langle E^{(i,j)}, E^{(r,s)} \rangle = \delta_{ir}\delta_{js}.$$

Para demostrarlo se aplica la definición del producto escalar de Frobenius obteniendo

$$\langle E^{(i,j)}, E^{(r,s)} \rangle = \operatorname{tr}((E^{(i,j)})^t E^{(r,s)}) = \operatorname{tr}(E^{(j,i)} E^{(r,s)}).$$

Ahora, es fácil ver que

$$E^{(j,i)}E^{(r,s)} = \begin{cases} E^{(j,s)}, & \text{si } i = r \\ O, & \text{si } i \neq r \end{cases} = \delta_{ir}E^{(j,s)}.$$

Luego,

$$\langle E^{(i,j)}, E^{(r,s)}\rangle = \operatorname{tr}(E^{(j,i)}E^{(r,s)}) = \operatorname{tr}(\delta_{ir}E^{(j,s)}) = \delta_{ir}\operatorname{tr}(E^{(j,s)}) = \delta_{ir}\delta_{js},$$

pues $\operatorname{tr}(E^{(j,s)}) = 1$ si $j = s$ y $\operatorname{tr}(E^{(j,s)}) = 0$ si $j \neq s$. ▷

Ahora se observará el significado de un sencillo cálculo. En el espacio vectorial $\mathbb{R}^2$ con el producto escalar canónico se tiene que cualquier vector $(x, y) \in \mathbb{R}^2$ se puede escribir como

$$\begin{aligned} (x, y) &= x(1, 0) + y(0, 1) \\ &= \langle (x, y), (1, 0)\rangle(1, 0) + \langle (x, y), (0, 1)\rangle(0, 1), \end{aligned}$$

lo que indica que un vector arbitrario de $\mathbb{R}^2$ se puede escribir (de manera única) como combinación lineal de los vectores canónicos $e_1 = (1, 0)$ y $e_2 = (0, 1)$ siendo los escalares de dicha combinación los productos escalares de dicho vector con cada uno de los vectores de la base canónica. A continuación, se generalizan estas ideas a un espacio euclídeo de dimensión finita. Se observa, además, que

$$\langle (x, y), (1, 0)\rangle = x \qquad \text{y} \qquad \langle (x, y), (0, 1)\rangle = y,$$

que es lo que se conoce como la proyección ortogonal del vector (x, y) sobre los vectores canónicos e_1 y e_2, respectivamente (compárense con la fórmula (10.12) de la Subsección 10.5.6).

Utilizando un enfoque geométrico, como se ha indicado en la página 560, a una base de un espacio vectorial E de dimensión n le corresponde un sistema de coordenadas cartesiano. Si el espacio E es euclídeo y dicha base es ortonormal, le corresponde un **sistema de coordenadas rectangulares** (es decir, con ejes ortogonales dos a dos) donde los puntos de referencia sobre cada eje coordenado está a una distancia unidad desde el origen. Encontrar las coordenadas de un vector con respecto a este tipo de sistemas es sencillo.

Proposición 10.7. Sea $(E, \langle\cdot,\cdot\rangle)$ un espacio euclídeo de dimensión finita y $\mathcal{B} = \{u_1, u_2, \ldots, u_n\}$ una base ortogonal de E. Entonces cualquier vector $u \in E$ se escribe, de forma única, en la base $\mathcal{B}$ como

$$u = \frac{\langle u, u_1\rangle}{\|u_1\|^2}u_1 + \frac{\langle u, u_2\rangle}{\|u_2\|^2}u_2 + \cdots + \frac{\langle u, u_n\rangle}{\|u_n\|^2}u_n. \tag{10.8}$$

Demostración. Por ser $\mathcal{B} = \{u_1, u_2, \ldots, u_n\}$ una base de E, un vector $u \in E$ se escribe de forma única como

$$u = \alpha_1 u_1 + \alpha_2 u_2 + \cdots + \alpha_n u_n = \sum_{i=1}^{n} \alpha_i u_i, \tag{10.9}$$

para ciertos $\alpha_1, \alpha_2, \ldots, \alpha_n \in \mathbb{R}$. Realizando el producto escalar de ambos miembros de la igualdad anterior por uno cualquiera de ellos u_j (para cualquier $j = 1, 2, \ldots, n$) y teniendo en cuenta que $\langle u_i, u_j\rangle = 0$ para todo $i \neq j$, se tiene que

$$\begin{aligned} \langle u, u_j\rangle &= \left\langle \sum_{i=1}^{n} \alpha_i u_i, u_j \right\rangle \\ &= \sum_{i=1}^{n} \alpha_i \langle u_i, u_j\rangle \\ &= \alpha_j \langle u_j, u_j\rangle \\ &= \alpha_j \|u_j\|^2, \end{aligned}$$

de donde $\alpha_j = \frac{\langle u, u_j\rangle}{\|u_j\|^2}$ por ser $u_j \neq 0$. Puesto que u_j es arbitrario, la expresión obtenida es válida para todo $j = 1, 2, \ldots, n$, y sustituyendo la expresión de α_j en (10.9) se obtiene el resultado deseado. □

Es claro que si la base $\mathcal{B}$ fuese (no sólo ortogonal sino) ortonormal, la expresión (10.8) adoptaría la forma más simplificada:

$$u = \langle x, u_1\rangle u_1 + \langle x, u_2\rangle u_2 + \cdots + \langle x, u_n\rangle u_n.$$

Los coeficientes de esta expresión se interpretarán en la Subsección 10.5.6 utilizando el concepto de proyección.

Observación 10.11. Los coeficientes

$$\frac{\langle x, u_i\rangle}{\|u_i\|^2}, \qquad \text{para } i = 1, 2, \ldots, n$$

de la Proposición 10.7 se conocen como **coeficientes generalizados de Fourier** del vector u en la base ortogonal $\mathcal{B}$.

Se observa que si se dispone de una base ortogonal, para expresar un vector arbitrario como combinación lineal de los elementos de esa base, tan sólo se deben calcular $2n$ productos escalares (sin necesidad de resolver sistemas de ecuaciones lineales).

El adjetivo *generalizados* permite distinguirlos de los **coeficientes de Fourier** que fueron introducidos por primera vez para expresar una función[a] como combinación lineal de los vectores de un sistema específico: el sistema trigonométrico.

En este libro se tratarán sólo los *polinomios trigonométricos* al estudiar el *problema de la mejor aproximación* en la página 585.

[a] Para que una función se pueda escribir de este modo se requieren propiedades bastante generales que se estudian en la teoría denominada **Análisis de Fourier**, lo que proporciona un potente método ampliamente utilizado en Ingeniería.

El matemático y físico francés Jean-Baptiste Joseph Fourier (1768-1830) analizó el desarrollo de funciones periódicas en series trigonométricas llamadas **Series de Fourier**. Con la necesidad de justificar rigurosamente sus estudios, considerados poco rigurosos por los matemáticos de la Academia Francesa, fue necesario desarrollar posteriormente conceptos potentes del Análisis Matemático como clarificar el propio concepto de función, introducir el concepto de convegencia puntual y convergencia uniforme, etc., todos ellos necesarios para realizar un estudio teórico de la ecuación del calor, de la ecuación de onda y de la ecuación de Laplace (o del potencial). Cabe mencionar, como curiosidad, que Fourier fue el primero en explicar el efecto invernadero.

Figura 10.3: Retrato de Jean-Baptiste Joseph Fourier realizado por Boilly.

Ejemplo 10.20. *La base*

$$\left\{u_1 = \left(\frac{\sqrt{3}}{2}, \frac{1}{2}\right), u_2 = \left(-\frac{1}{2}, \frac{\sqrt{3}}{2}\right)\right\}$$

es una base ortonormal en $\mathbb{R}^2$ con el producto escalar canónico. Expresar el vector $u = (1, \sqrt{3})$ como combinación lineal de dicha base.

$\lhd$ Es fácil comprobar que $\langle u_1, u_2\rangle = 0$, $\langle u_1, u_1\rangle = 1$ y $\langle u_2, u_2\rangle = 1$.

De la Proposición 10.7,

$$u = (1, \sqrt{3}) = \langle u, u_1\rangle u_1 + \langle u, u_2\rangle u_2 = \sqrt{3}u_1 + 1u_2.$$

$\rhd$

Se observa que estos ejemplos son relativamente fáciles de determinar dado que se utiliza el producto escalar canónico en el espacio vectorial $\mathbb{R}^n$.

En la siguiente subsección se estudia un método que permite encontrar bases ortogonales en espacios euclídeos abstractos en general.

10.5.4. Ortonormalización de una base

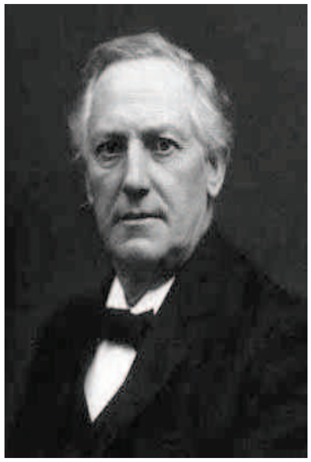
Si en un espacio euclídeo $(E, \langle \cdot, \cdot \rangle)$ de dimensión finita $n \geq 1$ se dispone de una base ortogonal $\mathcal{B} = \{u_1, u_2, \ldots, u_n\}$, se cumple que:

- Cualquier vector $x \in E$ se escribe, de forma única, en la base $\mathcal{B}$ como
$$x = \frac{\langle x, u_1 \rangle}{\|u_1\|^2} u_1 + \frac{\langle x, u_2 \rangle}{\|u_2\|^2} u_2 + \cdots + \frac{\langle x, u_n \rangle}{\|u_n\|^2} u_n,$$
donde los escalares de la combinación lineal se encuentran a partir del cálculo de productos escalares.

- Si, además, $\mathcal{B}$ es ortonormal, para dos vectores $x, y \in E$ que se expresan unívocamente como
$$x = x_1 u_1 + x_2 u_2 + \cdots + x_n u_n \quad \text{e} \quad y = y_1 u_1 + y_2 u_2 + \cdots + y_n u_n,$$
se puede ver (se propone como ejercicio) que su producto escalar es
$$\langle x, y \rangle = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n = ([x]_{\mathcal{B}})^t [y]_{\mathcal{B}}$$
y la norma de x viene dada por
$$\|x\| = +\sqrt{([x]_{\mathcal{B}})^t [x]_{\mathcal{B}}}.$$

La sencillez con la que se realizan los cálculos, a partir de una base ortonormal, ha motivado la búsqueda de un procedimiento para encontrar una base ortogonal a partir de una base arbitraria.

Si bien se ha visto que el resultado recíproco de la Proposición 10.6 es falso, en lo que sigue se aborda una suerte de recíproco, en el sentido que, a partir de un conjunto linealmente independiente en un espacio euclídeo de dimensión finita, se podrá construir otro que es ortogonal (y generan el mismo subespacio).



(a) Jørgen Pedresen Gram

(b) Erhard Schmidt

Figura 10.4: Matemáticos danés y alemán que estudiaron el proceso de ortogonalización.

Idea intuitiva del proceso de ortogonalización: A continuación se realizan algunas observaciones que harán más intuitiva la demostración del proceso de ortogonalización de Gram-Schmidt[14] del Teorema 10.2.

Para ello, se considera un espacio euclídeo $(E, \langle\cdot,\cdot\rangle)$ de dimensión $n \geq 1$ y se supone que $\{u_1, u_2, \ldots, u_r\}$ es un conjunto linealmente independiente $(r \leq n)$ conocido de E. Se analiza, para los casos $r = 1, 2, 3$, la forma de construir r vectores ortogonales $\{v_1, v_2, \ldots, v_r\}$ que cumplan

$$\overline{\{u_1, u_2, \ldots, u_r\}} = \overline{\{v_1, v_2, \ldots, v_r\}}$$

[14]Si bien el método se conoce con el nombre del matemático danés Jørgen Pedresen Gram (1850-1916) y del matemático alemán Erhard Schmidt (1876-1959), al parecer Pierre-Simon de Laplace, lo había utilizado previamente en sus estudios casi un siglo antes. Como curiosidades, Gram murió a los 61 años arrollado por una bicicleta y Schmidt fue discípulo de David Hilbert (1862-1943) y fundó el primer Instituto de Matemáticas Aplicadas de Berlín. El matemático alemán David Hilbert fue uno de los más influyentes del siglo XIX y principios del XX. De hecho, a los espacios vectoriales con producto interno se los suele llamar **espacios pre-hilbertianos** en su honor, por ser el primero en considerar el conjunto de funciones como espacio vectorial; se nombran de este modo sobre todo cuando el espacio vectorial es de dimensión infinita.

(es decir, que generen el mismo subespacio vectorial de E). En efecto,

- Sea $r = 1$. Tomando $v_1 := u_1$ es claro que $\{v_1\}$ es un conjunto ortogonal (por ser un conjunto unitario) y $\overline{\{v_1\}} = \overline{\{u_1\}}$.

- Sea $r = 2$. Construido v_1 como antes, ahora se trata de buscar un vector $v_2 \in E$ de modo que:

  - $0 \neq v_2 \perp \{v_1\}$,
  - $\overline{\{v_1, v_2\}} = \overline{\{u_1, u_2\}}$.

  Para que generen el mismo subespacio, puesto que $\overline{\{v_1\}} = \overline{\{u_1\}}$, el vector v_2 debe ser combinación lineal de $u_1 (= v_1)$ y de u_2 (¡este último necesariamente debe encontrarse en la combinación lineal!), es decir:

  $$v_2 := \beta u_2 + \alpha_1 v_1, \qquad \text{con} \qquad \beta \neq 0.$$

  El valor de α_1 se debe buscar de modo que $\langle v_2, v_1 \rangle = 0$. Realizando el producto escalar

  $$0 = \langle v_2, v_1 \rangle = \langle \beta u_2 + \alpha_1 v_1, v_1 \rangle = \beta \langle u_2, v_1 \rangle + \alpha_1 \langle v_1, v_1 \rangle$$

  de donde $\alpha_1 = -\beta \frac{\langle u_2, v_1 \rangle}{\langle v_1, v_1 \rangle}$ (pues $v_1 \neq 0$) y, así,

  $$v_2 := \beta u_2 - \beta \frac{\langle u_2, v_1 \rangle}{\|v_1\|^2} v_1 = \beta \left[u_2 - \frac{\langle u_2, v_1 \rangle}{\|v_1\|^2} v_1 \right].$$

  Como β únicamente modifica la norma de v_2, alcanza con tomar $\beta = 1$. De este modo, el vector

  $$v_2 := u_2 - \frac{\langle u_2, v_1 \rangle}{\|v_1\|^2} v_1$$

  cumple las condiciones requeridas.

  Es claro que $v_2 \neq 0$, pues si lo fuera, se tendría $u_2 = \frac{\langle u_2, v_1 \rangle}{\|v_1\|^2} v_1 \in \overline{\{v_1\}} = \overline{\{u_1\}}$, que contradice la independencia lineal de $\{u_1, u_2\}$. Además, los vectores no nulos v_1 y v_2 son ortogonales por construcción, de donde se deduce que $\{v_1, v_2\}$ es linealmente independiente (Proposición 10.6). Luego, $\{v_1, v_2\}$ es una base de $\overline{\{u_1, u_2\}}$ y, por lo tanto, se tiene que

$$\overline{\{v_1, v_2\}} = \overline{\{u_1, u_2\}}.$$

Observar que el coeficiente α_1 (para $\beta = 1$) es el opuesto del coeficiente de Fourier de u_2 correspondiente a v_1 escrito en la base $\{v_1, v_2\}$.

- Sea $r = 3$. Construidos $\{v_1, v_2\}$ como antes, ahora se trata de buscar un vector $v_3 \in E$ de modo que:

  - $0 \neq v_3 \perp \{v_1, v_2\}$,
  - $\overline{\{v_1, v_2, v_3\}} = \overline{\{u_1, u_2, u_3\}}$.

  Para que generen el mismo subespacio, puesto que $\overline{\{v_1, v_2\}} = \overline{\{u_1, u_2\}}$, el vector v_3 debe ser combinación lineal de u_1, u_2 (estos dos vectores generan el mismo subespacio que $\overline{\{v_1, v_2\}}$) y de u_3 (este necesariamente debe figurar en la combinación lineal y, como antes, puede tomarse con coeficiente 1), es decir:

  $$v_3 := u_3 + \alpha_1 v_1 + \alpha_2 v_2.$$

  Los valores de α_1 y α_2 se deben buscar de modo que $\langle v_3, v_1\rangle = 0$ y $\langle v_3, v_2\rangle = 0$. Realizando los productos escalares y utilizando la relación de ortogonalidad conocida se tiene

  $$\begin{array}{rcl} 0 & = & \langle v_3, v_1\rangle \\ & = & \langle u_3 + \alpha_1 v_1 + \alpha_2 v_2, v_1\rangle \\ & = & \langle u_3, v_1\rangle + \alpha_1\langle v_1, v_1\rangle + \alpha_2\langle v_2, v_1\rangle \\ & = & \langle u_3, v_1\rangle + \alpha_1\langle v_1, v_1\rangle \end{array}$$

  se tiene $\alpha_1 = -\frac{\langle u_3, v_1\rangle}{\langle v_1, v_1\rangle}$.

  De modo similar se obtiene $\alpha_2 = -\frac{\langle u_3, v_2\rangle}{\langle v_2, v_2\rangle}$.

Sustituyendo se llega a la expresión

$$v_3 := u_3 - \frac{\langle u_3, v_1\rangle}{\|v_1\|^2} v_1 - \frac{\langle u_3, v_2\rangle}{\|v_2\|^2} v_2.$$

De nuevo, $v_3 \neq 0$ pues en caso contrario sería $u_3 \in \overline{\{v_1, v_2\}} = \overline{\{u_1, u_2\}}$, lo que contradice la independencia lineal de $\{u_1, u_2, u_3\}$. Luego, el conjunto $\{v_1, v_2, v_3\}$ es un subconjunto de $\overline{\{u_1, u_2, u_3\}}$ que, por la Proposición 10.6, es linealmente independiente, con lo cual $\{v_1, v_2, v_3\}$ es una base de $\overline{\{u_1, u_2, u_3\}}$ y, por lo tanto,

$$\overline{\{v_1, v_2, v_3\}} = \overline{\{u_1, u_2, u_3\}}.$$

A partir de estas observaciones, la construcción general parece clara y el resultado se podrá probar por inducción.

Teorema 10.2 (Proceso de ortogonalización de Gram-Schmidt)**.** Sea $(E, \langle\cdot,\cdot\rangle)$ un espacio euclídeo y $\{u_1, u_2, \ldots, u_r\}$ un conjunto linealmente independiente (finito, con $r \geq 1$) de E. Entonces el conjunto $\{v_1, v_2, \ldots, v_r\}$ construido de forma recursiva a partir de las expresiones

$$v_1 \quad := \quad u_1,$$

y, para $k = 2, 3, \ldots, r$,

$$v_k \quad := \quad u_k - \frac{\langle u_k, v_1\rangle}{\|v_1\|^2} v_1 - \frac{\langle u_k, v_2\rangle}{\|v_2\|^2} v_2 - \cdots - \frac{\langle u_k, v_{k-1}\rangle}{\|v_{k-1}\|^2} v_{k-1},$$

es ortogonal (formado por vectores no nulos) y cumple que

$$\overline{\{u_1, u_2, \ldots, u_k\}} = \overline{\{v_1, v_2, \ldots, v_k\}}, \text{ para cada } k = 1, 2, \ldots, r.$$

Demostración. Se utiliza la notación

$$U_j := \{u_1, u_2, \ldots, u_j\} \text{ y } V_j := \{v_1, v_2, \ldots, v_j\} \text{ para } j = 1, 2, \ldots, r,$$

donde el subíndice j indica la cantidad de vectores consecutivos que se seleccionan en el conjunto (siempre comenzando a partir del primero). La demostración se realiza por inducción sobre r.

Si $r = 1$, es claro que haciendo $v_1 := u_1$, se tiene que el conjunto $V_1 := \{v_1\}$ es ortogonal y $v_1 \neq 0$ (pues $U_1 = \{u_1\}$ es linealmente independiente) y, además, $\overline{V_1} = \overline{U_1}$.

Sea $r > 1$ y, por hipótesis de inducción, supóngase que para el conjunto linealmente independiente $U_{r-1} := \{u_1, u_2, \ldots, u_{r-1}\}$, se ha construido un conjunto ortogonal $V_{r-1} = \{v_1, v_2, \ldots, v_{r-1}\}$ de vectores no nulos, según las fórmulas de recurrencia del enunciado, tal que $\overline{U_{r-1}} = \overline{V_{r-1}}$. Se debe probar que si $U_r := \{u_1, u_2, \ldots, u_{r-1}, u_r\}$ es linealmente independiente entonces se puede construir, mediante las expresiones del enunciado, un conjunto ortogonal $V_r = \{v_1, v_2, \ldots, v_{r-1}, v_r\}$ de vectores no nulos tal que $\overline{U_r} = \overline{V_r}$. En efecto, de la expresión

$$v_r := u_r - \frac{\langle u_r, v_1\rangle}{\|v_1\|^2} v_1 - \frac{\langle u_r, v_2\rangle}{\|v_2\|^2} v_2 - \cdots - \frac{\langle u_r, v_{r-1}\rangle}{\|v_{r-1}\|^2} v_{r-1} \tag{10.10}$$

se satisface que $v_r \neq 0$ pues, en caso contrario, sería $u_r \in \overline{V_{r-1}} = \overline{U_{r-1}}$, lo que lleva a una contradicción con la independencia lineal de U_r.

Para $j = 1, 2, \ldots, r-1$, de la expresión (10.10) se tiene que

$$\begin{aligned}
\langle v_r, v_j\rangle &= \left\langle u_r - \sum_{i=1}^{r-1} \frac{\langle u_r, v_i\rangle}{\|v_i\|^2} v_i, v_j \right\rangle \\
&= \langle u_r, v_j\rangle - \sum_{i=1}^{r-1} \frac{\langle u_r, v_i\rangle}{\|v_i\|^2} \langle v_i, v_j\rangle \\
&= \langle u_r, v_j\rangle - \frac{\langle u_r, v_j\rangle}{\|v_j\|^2} \|v_j\|^2 \\
&= 0,
\end{aligned}$$

donde se ha utilizado que V_{r-1} es un conjunto ortogonal. Esto prueba que V_r es un conjunto ortogonal de vectores no nulos.

Por otro lado, de la igualdad $\overline{U_{r-1}} = \overline{V_{r-1}}$ y de la expresión (10.10) se llega a $\overline{V_r} \subseteq \overline{U_r}$ (se propone completar los detalles como ejercicio). De la

Proposición 10.6 se puede garantizar que V_r es linealmente independiente. Luego, $\dim_{\mathbb{R}}(\overline{V_r}) = r = \dim_{\mathbb{R}}(\overline{U_r})$. Por lo tanto, $\overline{U_r} = \overline{V_r}$. □

El proceso de ortogonalización de Gram-Schmidt permite establecer el siguiente importante resultado.

Teorema 10.3 (Existencia de bases ortonormales)**.** Todo espacio euclídeo de dimensión finita $n \geq 1$ admite una base ortonormal $\{b_1, b_2, \ldots, b_n\}$. Además, si $u \in E$ entonces

$$u = \langle u, b_1\rangle b_1 + \langle u, b_2\rangle b_2 + \cdots + \langle u, b_n\rangle b_n.$$

Demostración. Basta aplicar el Teorema 10.2 a una base $\{u_1, u_2, \ldots, u_n\}$ de E (cuya existencia está garantizada por el Teorema 8.2) para encontrar una base ortogonal $\{v_1, v_2, \ldots, v_n\}$ de E. Luego, se deben normalizar todos los vectores para obtener una base ortonormal

$$\left\{\frac{1}{\|v_1\|}v_1, \frac{1}{\|v_2\|}v_2, \ldots, \frac{1}{\|v_n\|}v_n\right\}$$

de E. La representación de u como combinación lineal de los elementos de la base ortonormal es consecuencia de la Proposición 10.7. □

Observación 10.12. En el Corolario 10.1 se estableció que el número de direcciones mutuamente ortogonales (dimensión geométrica) no puede exceder la dimensión algebraica del espacio. Una consecuencia del Teorema 10.3 (que garantiza la existencia de bases ortonormales) es que ambas dimensiones coinciden.

Ejemplo 10.21. *Obtener una base ortonormal de $\mathbb{R}^2$ con el producto escalar canónico a partir de la base*

$$\{u_1 = (1,1)\,, u_2 = (1,0)\}\,.$$

$\triangleleft$ Se aplica el proceso de Gram-Schmidt al conjunto linealmente independiente $\{u_1, u_2\}$ utilizando el producto escalar canónico de $\mathbb{R}^2$. Se toma $v_1 = u_1 = (1,1)$ y se construye v_2 a partir de

$$\begin{aligned} v_2 &= u_2 - \frac{\langle u_2, v_1 \rangle}{\|v_1\|^2} v_1 \\ &= (1,0) - \frac{\langle (1,0),(1,1) \rangle}{\|(1,1)\|^2}(1,1) \\ &= (1,0) - \frac{1}{2}(1,1) \\ &= \left(\frac{1}{2}, -\frac{1}{2}\right). \end{aligned}$$

Luego, el conjunto $\left\{v_1 = (1,1), v_2 = \left(\frac{1}{2}, -\frac{1}{2}\right)\right\}$ es una base ortogonal de $\mathbb{R}^2$ y, multiplicando cada vector por el inverso multiplicativo de su norma (y racionalizando), el conjunto

$$\left\{\left(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right), \left(\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}\right)\right\}$$

es una base ortonormal de $\mathbb{R}^2$, hechos que se comprueban fácilmente. $\triangleright$

Ejemplo 10.22. *Obtener un conjunto ortonormal del espacio vectorial* $\mathbb{R}_2[x]$ *con el producto escalar dado por*

$$\langle p, q \rangle = \int_{-1}^{1} p(x)q(x)\,\mathrm{d}x$$

a partir del conjunto

$$\left\{1, x, x^2\right\}.$$

$\triangleleft$ Se aplica el proceso de Gram-Schmidt al conjunto linealmente independiente $\{u_1 = 1, u_2 = x, u_3 = x^2\}$ de $\mathbb{R}_2[x]$ utilizando el producto interno indicado.

Se toma $v_1 = u_1 = 1$. Se construye v_2 mediante

$$v_2 = u_2 - \frac{\langle u_2, v_1 \rangle}{\|v_1\|^2} v_1 = x - \frac{\langle x, 1 \rangle}{\|1\|^2} 1 = x - \frac{\int_{-1}^{1} x \cdot 1\,\mathrm{d}x}{\int_{-1}^{1} 1 \cdot 1\,\mathrm{d}x} 1 = x.$$

Este resultado indica que el conjunto $\{1, x\}$ ya era ortogonal. Ahora, se construye v_3 mediante

$$\begin{aligned} v_3 &= u_3 - \frac{\langle u_3, v_1\rangle}{\|v_1\|^2} v_1 - \frac{\langle u_3, v_2\rangle}{\|v_2\|^2} v_2 \\ &= x^2 - \frac{\langle x^2, 1\rangle}{\|1\|^2} 1 - \frac{\langle x^2, x\rangle}{\|x\|^2} x \\ &= x^2 - \frac{\int_{-1}^{1} x^2 \cdot 1 \, \mathrm{d}x}{\int_{-1}^{1} 1 \cdot 1 \, \mathrm{d}x} 1 - \frac{\int_{-1}^{1} x^2 \cdot x \, \mathrm{d}x}{\int_{-1}^{1} 1 \cdot 1 \, \mathrm{d}x} x \\ &= x^2 - \frac{1}{3}. \end{aligned}$$

Luego, el conjunto $\left\{v_1 = 1, v_2 = x, v_3 = x^2 - \frac{1}{3}\right\}$ es una base ortogonal de $\mathbb{R}_2[x]$ y, multiplicando cada vector por el inverso multiplicativo de su norma, el conjunto

$$\left\{ \frac{1}{\sqrt{2}}, \frac{\sqrt{3}}{\sqrt{2}} x, \frac{\sqrt{5}}{\sqrt{2}} \left(\frac{3}{2} x^2 - \frac{1}{2} \right) \right\}$$

es una base ortonormal de $\mathbb{R}_2[x]$. $\triangleright$

Los vectores obtenidos en el ejemplo anterior se conocen como los **Polinomios de Legendre** de grado 1, 2 y 3. En este caso, el proceso se podría continuar de manera indefinida.

10.5.5. Complemento ortogonal

Hasta ahora se han estudiado sistemas ortogonales y bases ortogonales/ortonormales. Ahora se aborda el caso en que el sistema de partida sea un subespacio, aunque en realidad, para esta definición alcanzaría con considerar un subconjunto de E.

Definición 10.10. Sea $(E, \langle \cdot, \cdot \rangle)$ un espacio euclídeo y S un subespacio vectorial de E. Se llama **complemento ortogonal** de S, y se denota $S^{\perp}$, al conjunto de los vectores de E que son ortogonales a todos los vectores de S. En símbolos,

$$S^{\perp} = \{u \in E : u \perp s, \forall s \in S\}.$$

Ejemplo 10.23. *Obtener el complemento ortogonal del subespacio*

$$S = \{(x, x) : x \in \mathbb{R}\}$$

de $\mathbb{R}^2$ *con el producto escalar canónico.*

$\lhd$ Por definición,

$$S^{\perp} = \{(u, v) \in \mathbb{R}^2 : (u, v) \perp (x, x), \forall x \in \mathbb{R}\}.$$

Dado que $0 = \langle (u, v), (x, x) \rangle = ux + vx$ para todo $x \in \mathbb{R}$, en particular, debe satisfacerse, por ejemplo, para $x = 1$. Es decir $0 = \langle (u, v), (1, 1) \rangle = u + v$, de donde $v = -u$. Luego,

$$S^{\perp} = \{(u, -u) : u \in \mathbb{R}\}.$$

El subespacio S puede representarse geométricamente por la recta que pasa por los puntos $(0, 0)$ y $(1, 1)$ y $S^{\perp}$ por la recta que pasa por los puntos $(0, 0)$ y $(1, -1)$ (es decir, su recta perpendicular que pasa por el origen). $\rhd$

Observación 10.13. En todo espacio euclídeo $(E, \langle \cdot, \cdot \rangle)$ se cumple que:

- $V^{\perp} = \{0\}$.
- $\{0\}^{\perp} = V$.

Su demostración se propone como ejercicio.

De acuerdo a esta observación, sólo es necesario estudiar los subespacios no triviales.

Propiedades del complemento ortogonal

Proposición 10.8. Sea $(E, \langle\cdot,\cdot\rangle)$ un espacio euclídeo y sea S un subespacio vectorial de E. Entonces el complemento ortogonal $S^\perp$ de S es un subespacio vectorial de E.

Demostración. Se recuerda que $S^\perp = \{u \in E : \langle u, s\rangle = 0, \forall s \in S\}$.

Es claro que $0 \in S^\perp$ pues $\langle 0, s\rangle = 0$, para todo $s \in S$.

Sean $u_1, u_2 \in S^\perp$ y $\alpha \in \mathbb{R}$. Entonces $\langle u_1, s\rangle = \langle u_2, s\rangle = 0$, para todo $s \in S$. Luego, $\langle \alpha u_1 + u_2, s\rangle = \alpha\langle u_1, s\rangle + \langle u_2, s\rangle = 0$, para todo $s \in S$. De aquí, $\alpha u_1 + u_2 \in S^\perp$.

Por lo tanto, $S^\perp$ es un subespacio vectorial de E. $\square$

Observar que a la hora de resolver el Ejemplo 10.23 no ha sido necesario utilizar todos los vectores de S sino que se utilizó sólo uno: el vector $(1, 1)$, que es un conjunto linealmente independiente y forma una base de S. Este hecho es general, como se prueba en el siguiente resultado.

Proposición 10.9. Sea $(E, \langle\cdot,\cdot\rangle)$ un espacio euclídeo de dimensión finita, S un subespacio vectorial de E, $\mathcal{B} = \{u_1, u_2, \ldots, u_r\}$ una base de S y $u \in E$. Entonces

$$u \in S^\perp \qquad \Longleftrightarrow \qquad u \perp u_i, \quad \text{para todo } i = 1, 2, \ldots, r.$$

Demostración. Se debe probar la doble implicación.

($\Longrightarrow$) Si $u \in S^\perp$ entonces $u \perp s$ para todo $s \in S$. En particular, se debe cumplir para los vectores de la base $\mathcal{B}$ de S, es decir, $u \perp u_i$, para todo $i = 1, 2, \ldots, r$.

($\Longleftarrow$) Supóngase que $u \perp u_i$, para todo $i = 1, 2, \ldots, r$, es decir

$$\langle u, u_i\rangle = 0, \text{ para todo } i = 1, 2, \ldots, r.$$

Se debe probar que $u \in S^{\perp}$, o equivalentemente $\langle u, s\rangle = 0$, para todo $s \in S$. En efecto, si $s \in S$, entonces existen escalares (reales) $\alpha_1, \alpha_2, \ldots, \alpha_r$ unívocamente determinados tales que $s = \alpha_1 u_1 + \alpha_2 u_2 + \cdots + \alpha_r u_r$. Luego,

$$\begin{aligned} \langle u, s\rangle &= \langle u, \alpha_1 u_1 + \alpha_2 u_2 + \cdots + \alpha_r u_r\rangle \\ &= \alpha_1\langle u, u_1\rangle + \alpha_2\langle u, u_2\rangle + \cdots + \alpha_r\langle u, u_r\rangle \\ &= 0. \end{aligned}$$

Por lo tanto, $u \perp s$, con lo cual $u \in S^{\perp}$. □

Definición 10.11. Sean $(E, \langle\cdot,\cdot\rangle)$ un espacio euclídeo y S y T dos subespacios vectoriales de E. Se dice que E es **suma directa ortogonal** de los subespacios S y T, y se denota $E = S \oplus^{\perp} T$, si se cumplen las dos condiciones siguientes:

- $E = S \oplus T$ (es decir, S y T son subespacios complementarios en el sentido de la Definición 8.12),
- cada elemento de S es ortogonal a cada elemento de T (en este caso se dice que S y T son **subespacios ortogonales**).

En el siguiente resultado se prueba que en un espacio euclídeo, todo subespacio de dimensión finita es siempre completementario, en el sentido de la suma directa ortogonal, a su complemento ortogonal.

El título del próximo resultado se justificará en la Subsección 10.5.6, donde se definirá la *proyección ortogonal* de un vector $u \in E$ sobre un subespacio S a partir del Teorema 10.4.

Teorema 10.4 (Teorema de la proyección ortogonal)**.** Sea $(E, \langle\cdot,\cdot\rangle)$ un espacio euclídeo. Para cualquier subespacio vectorial S de dimensión finita de E, se cumple que

$$E = S \oplus^{\perp} S^{\perp}.$$

Demostración. Si $S = \{0\}$ o $S = E$, el resultado es evidente de la Observación 10.13.

Sea S un subespacio no trivial de E con $1 \leq \dim_{\mathbb{R}}(S) = r < n$.

Por el Teorema 10.3 es posible tomar una base ortonormal $\{b_1, b_2, \ldots, b_r\}$ de S.

Ahora, para cada $u \in E$, se definen:

$$u_S := \langle u, b_1\rangle b_1 + \langle u, b_2\rangle b_2 + \cdots + \langle u, b_r\rangle b_r = \sum_{j=1}^{r} \langle u, b_j\rangle b_j$$

y

$$u_{S^\perp} := u - u_S.$$

Es claro que $u = u_S + u_{S^\perp}$ con $u_S \in S$. Para ver que $E = S + S^\perp$ se debe probar que $u_{S^\perp} \in S^\perp$. Por la Proposición 10.9 alcanza con probar que $u_{S^\perp} \perp b_i$ para todo $i = 1, 2, \ldots, r$. En efecto,

$$\begin{aligned}
\langle u_{S^\perp}, b_i\rangle &= \langle u - u_S, b_i\rangle \\
&= \langle u, b_i\rangle - \langle u_S, b_i\rangle \\
&= \langle u, b_i\rangle - \left\langle \sum_{j=1}^{r} \langle u, b_j\rangle b_j, b_i \right\rangle \\
&= \langle u, b_i\rangle - \sum_{j=1}^{r} \langle u, b_j\rangle \langle b_j, b_i\rangle \\
&= \langle u, b_i\rangle - \sum_{j=1}^{r} \langle u, b_j\rangle \delta_{ji} \\
&= \langle u, b_i\rangle - \langle u, b_i\rangle \\
&= 0,
\end{aligned}$$

para todo $i = 1, 2, \ldots, r$.

Por último, se debe probar (por doble inclusión) que $S \cap S^\perp = \{0\}$.

($\supseteq$) Es claro que $\{0\} \subseteq S \cap S^\perp$ pues S y $S^\perp$ son subespacios vectoriales (véase la Proposición 10.8).

($\subseteq$) Si $u \in S \cap S^\perp$ entonces $u \in S$ y $u \in S^\perp$. Por definición de complemento ortogonal, $\langle u, u\rangle = 0$, de donde se tiene $u = 0$.

Por lo tanto, se ha probado que $E = S \oplus S^\perp$.

Ahora, la fórmula $E = S \oplus^\perp S^\perp$ es inmediata de la propia definición de $S^\perp$. $\square$

Observación 10.14. Si en el Teorema de la proyección ortogonal se supone como hipótesis que E es un espacio euclídeo de dimensión finita, dicho teorema se puede demostrar a partir de la ampliación de una base ortonormal de S a todo E, aplicando el método de ortonormalización de Gram-Schmidt para construir una base ortonormal de E y probando que los vectores añadidos forman una base ortonormal de $S^\perp$. Se propone completar los detalles como ejercicio.

Del Teorema de la proyección ortogonal se obtienen las siguientes consecuencias inmediatas.

Corolario 10.2. Sea $(E, \langle\cdot,\cdot\rangle)$ un espacio euclídeo y S un subespacio vectorial de E con $\dim_{\mathbb{R}}(S) = r \geq 1$. Se cumple que:

(a) Si $\{b_1, b_2, \ldots, b_r\}$ es una base ortonormal de S entonces cualquier vector $u \in E$ se puede escribir como

$$u = \sum_{j=1}^{r} \langle u, b_j\rangle b_j + u_{S^\perp},$$

donde $u_{S^\perp} \in S^\perp$ está unívocamente determinado.

(b) Si E es de dimensión finita entonces

$$\dim_{\mathbb{R}}(E) = \dim_{\mathbb{R}}(S) + \dim_{\mathbb{R}}(S^\perp).$$

Ejemplo 10.24. *Resolver el Ejemplo 10.23 utilizando el Corolario 10.2.*

◁ Puesto que S es un subespacio de $\mathbb{R}^2$ de dimensión 1, del Corolario 10.2 se sabe que $\dim_{\mathbb{R}}(S^{\perp}) = \dim_{\mathbb{R}}(\mathbb{R}^2) - \dim_{\mathbb{R}}(S) = 2 - 1 = 1$. Es fácil ver que $(-1, 1) \in S^{\perp}$ y $\{(-1, 1)\}$ es linealmente independiente. Luego, $S^{\perp} = \overline{\{(-1, 1)\}} = \{(-u, u) : u \in \mathbb{R}\} = \{(u, -u) : u \in \mathbb{R}\}$. ▷

Ejemplo 10.25. *Sea* $S = \overline{\{u_1 = (1, 2, 3), u_2 = (0, 0, 1)\}}$ *un subespacio del espacio euclídeo* $\mathbb{R}^3$ *con el producto escalar canónico. Encontrar los posibles valores de* $a \in \mathbb{R}$ *tales que* $\{u_3 = (a, 1, 0)\}$ *sea una base de* $S^{\perp}$*.*

◁ Dado que S es un subespacio de $\mathbb{R}^3$ de dimensión 2, del Corolario 10.2 se sabe que $\dim_{\mathbb{R}}(S^{\perp}) = \dim_{\mathbb{R}}(\mathbb{R}^3) - \dim_{\mathbb{R}}(S) = 3 - 2 = 1$. Al ser $u_3 \neq (0, 0, 0)$, será una base de $S^{\perp}$ siempre que $\langle (1, 2, 3), (a, 1, 0) \rangle = 0$ y $\langle (0, 0, 1), (a, 1, 0) \rangle = 0$. Luego, $a = -2$. ▷

Las siguientes propiedades son válidas para subespacios de dimensión arbitraria (finita o infinita).

Proposición 10.10. Sea $(E, \langle \cdot, \cdot \rangle)$ un espacio euclídeo y sean S, T dos subespacios vectoriales de E. Entonces:

(a) Si $S \subseteq T$ entonces $T^{\perp} \subseteq S^{\perp}$.

(b) $(S + T)^{\perp} = S^{\perp} \cap T^{\perp}$.

Demostración. (a) Supóngase que $S \subseteq T$. Si $u \in T^{\perp}$ entonces $\langle u, t \rangle = 0$ para toda $t \in T$. En particular, $\langle u, s \rangle = 0$ para toda $s \in S$ (pues $S \subseteq T$). Por lo tanto, $u \in S^{\perp}$.

(b) Siempre se cumple que $S \subseteq S + T$ y $T \subseteq S + T$. Por el apartado (a),

$$(S + T)^{\perp} \subseteq S^{\perp} \qquad \text{y} \qquad (S + T)^{\perp} \subseteq T^{\perp}.$$

Luego, $(S+T)^{\perp} \subseteq S^{\perp} \cap T^{\perp}$.

Ahora se debe probar que $S^{\perp} \cap T^{\perp} \subseteq (S+T)^{\perp}$.

En efecto, sea $u \in S^{\perp} \cap T^{\perp}$. Entonces $u \in S^{\perp}$ y $u \in T^{\perp}$. Luego, $\langle u, s\rangle = 0$ para toda $s \in S$ y $\langle u, t\rangle = 0$ para toda $t \in T$. Si $v \in S+T$, existen vectores $s_0 \in S$ y $t_0 \in T$ unívocamente determinados tales que $v = s_0 + t_0$. De las igualdades anteriores,

$$\langle u, s_0 + t_0\rangle = \langle u, s_0\rangle + \langle u, t_0\rangle = 0,$$

es decir, $\langle u, v\rangle = 0$; y esta igualdad es válida para toda $v \in S+T$. Luego, $u \in (S+T)^{\perp}$.

Por lo tanto, se ha probado la igualdad $(S+T)^{\perp} = S^{\perp} \cap T^{\perp}$.

□

Sin embargo, en las siguientes propiedades es necesario suponer la hipótesis sobre la finitud en la dimensión de los subespacios.

Proposición 10.11. Sea $(E, \langle \cdot, \cdot \rangle)$ un espacio euclídeo de dimensión finita y sean S, T dos subespacios vectoriales de E. Entonces:

(a) $(S^{\perp})^{\perp} = S$.

(b) $(S \cap T)^{\perp} = S^{\perp} + T^{\perp}$.

Demostración. (a) Sea $s \in S$. Para que sea cierta la inclusión $S \subseteq (S^{\perp})^{\perp}$, se debe probar que $s \in (S^{\perp})^{\perp}$, es decir, que $\langle s, y\rangle = 0$ para todo $y \in S^{\perp}$, lo cual se cumple por la definición de complemento ortogonal:

$$S^{\perp} = \{y \in E : \langle y, u\rangle = 0, \text{ para todo } u \in S\}.$$

Por lo tanto, $S \subseteq (S^{\perp})^{\perp}$.

Ahora, aplicando dos veces el Corolario 10.2 (téngase en cuenta que S y $S^{\perp}$ son subespacios de un espacio finito-dimensional y, por lo tanto, S y

$S^\perp$ también lo son), se tiene que

$$\dim_{\mathbb{R}}((S^\perp)^\perp) = \dim_{\mathbb{R}}(E) - \dim_{\mathbb{R}}(S^\perp) = \dim_{\mathbb{R}}(S).$$

Luego, $(S^\perp)^\perp = S$.

(b) Utilizando el apartado (a) anterior y el apartado (b) de la Proposición 10.10 se probará la igualdad de sus complementos ortogonales. En efecto,

$$(S^\perp + T^\perp)^\perp = (S^\perp)^\perp \cap (T^\perp)^\perp = S \cap T$$

Luego, se tiene que $S^\perp + T^\perp = ((S^\perp + T^\perp)^\perp)^\perp = (S \cap T)^\perp$.

□

10.5.6. Proyección ortogonal y mejor aproximación

Como se ha visto en el Teorema de la proyección ortogonal (Teorema 10.4), dado un espacio euclídeo E y un subespacio vectorial S de dimensión finita de E, se tiene que $E = S \oplus^\perp S^\perp$. Es decir, si $u \in E$, es posible escribir

$$u = u_S + u_{S^\perp}, \qquad \text{con } \; u_S \in S, \; u_{S^\perp} \in S^\perp \tag{10.11}$$

donde u_S y $u_{S^\perp}$ están unívocamente determinados.

La siguiente definición otorga el nombre al Teorema 10.4.

Definición 10.12. Sea $(E, \langle \cdot, \cdot \rangle)$ un espacio euclídeo, S un subespacio vectorial de dimensión finita de E y $u \in E$. Se llama **proyección ortogonal** de u sobre S, y se denota $\mathrm{proy}_S(u)$, al (único) vector u_S de la descomposición (10.11) y se llama **componente** de u **ortogonal** a S, y se denota $\mathrm{proy}_{S^\perp}(u)$, al (único) vector $u_{S^\perp}$ de dicha descomposición. En símbolos,

$$\mathrm{proy}_S(u) = u_S \qquad \text{y} \qquad \mathrm{proy}_{S^\perp}(u) = u_{S^\perp} = u - u_S = u - \mathrm{proy}_S(u).$$

Además, si $\{b_1, b_2, \ldots, b_r\}$ es una base ortogonal de S, se tiene que

$$\mathrm{proy}_S(u) = \frac{\langle u, b_1 \rangle}{\langle b_1, b_1 \rangle} b_1 + \frac{\langle u, b_2 \rangle}{\langle b_2, b_2 \rangle} b_2 + \cdots + \frac{\langle u, b_r \rangle}{\langle b_r, b_r \rangle} b_r.$$

Observación 10.15. Las siguientes son dos observaciones acerca de la proyección ortogonal:

- $\text{proy}_{S^\perp}(u_S) = 0$ y $\text{proy}_S(u_{S^\perp}) = 0$.
- De la definición se puede caracterizar la **proyección de** u **sobre** S como el único vector $u_S \in S$ tal que $u - u_S \perp S$.

Ejemplo 10.26. *Se considera el espacio euclídeo* $\mathbb{R}^3$ *con el producto escalar canónico. Calcular, de dos formas diferentes, la proyección ortogonal del vector* $u = (2, 0, -3)$ *sobre el subespacio* $S = \overline{\{(1,1,1)\}}$ *y hallar también su componente ortogonal.*

$\lhd$ Primera forma: Unas ecuaciones paramétricas de S (en la base canónica de $\mathbb{R}^3$) vienen dadas por

$$[S]_{\mathcal{C}} : \begin{cases} x &= \lambda \\ y &= \lambda \\ z &= \lambda \end{cases}, \text{ para todo } \lambda \in \mathbb{R}.$$

Un vector satisface que $u_{S^\perp} = (a, b, c) \in S^\perp \Leftrightarrow \langle (a, b, c), (1, 1, 1)\rangle = 0$ por ser $\{(1, 1, 1)\}$ una base de S. Luego, $a + b + c = 0$, de donde $c = -a - b$. Por lo tanto, $S^\perp = \{(a, b, c) \in \mathbb{R}^3 : c = -a - b\}$.

Se debe escribir $u = u_S + u_{S^\perp}$ siendo $u_S \in S$ y $u_{S^\perp} \in S^\perp$. Es decir,

$$u = (2, 0, -3) = (\lambda, \lambda, \lambda) + (a, b, -a - b).$$

De este modo, resolviendo el sistema de ecuaciones lineales que resulta de igualar ambos vectores se tiene que $\lambda = -\frac{1}{3}$, $a = \frac{7}{3}$ y $b = \frac{1}{3}$. Es decir, $u = (-\frac{1}{3}, -\frac{1}{3}, -\frac{1}{3}) + (\frac{7}{3}, \frac{1}{3}, -\frac{8}{3}) \in S \oplus S^\perp$. Por definición,

$$\text{proy}_S(u) = \left(-\frac{1}{3}, -\frac{1}{3}, -\frac{1}{3}\right) \qquad \text{y} \qquad \text{proy}_{S^\perp}(u) = \left(\frac{7}{3}, \frac{1}{3}, -\frac{8}{3}\right).$$

Segunda forma: Por la Observación 10.15, basta conocer una base de S, a saber $\{u_S = (1,1,1)\}$. Luego,

$$\begin{aligned}
\mathrm{proy}_S(u) &= \frac{\langle u, u_S\rangle}{\|u_S\|^2} u_S \\
&= \frac{\langle (2,0,-3),(1,1,1)\rangle}{\|(1,1,1)\|^2} u_S \\
&= \frac{-1}{3}(1,1,1) \\
&= \left(-\frac{1}{3}, -\frac{1}{3}, -\frac{1}{3}\right).
\end{aligned}$$

En este segundo caso, la componente ortogonal se obtiene como

$$\mathrm{proy}_{S^\perp}(u) = u - \mathrm{proy}_S(u) = (2,0,-3) - \left(-\frac{1}{3}, -\frac{1}{3}, -\frac{1}{3}\right) = \left(\frac{7}{3}, \frac{1}{3}, -\frac{8}{3}\right).$$

$\triangleright$

Un caso particular importante: cuando $\dim_{\mathbb{R}}(S) = 1$.

Sea $(E, \langle\cdot,\cdot\rangle)$ un espacio euclídeo y $u, v \in E$, con $v \neq 0$. Se llama **proyección ortogonal** de u sobre v, y se denota $\mathrm{proy}_v(u)$, al vector

$$\mathrm{proy}_v(u) = \frac{\langle u, v\rangle}{\|v\|^2} v. \tag{10.12}$$

Otra forma directa de establecer la fórmula (10.12) es buscando el escalar $\alpha \in \mathbb{R}$ que cumple $\langle u - \alpha v, v\rangle = 0$, hecho geométricamente intuitivo en $\mathbb{R}^2$ con el producto escalar canónico. Despejando queda $\alpha = \frac{\langle u,v\rangle}{\|v\|^2}$ con lo cual $\mathrm{proy}_v(u) = \alpha v = \frac{\langle u,v\rangle}{\|v\|^2} v$. De la igualdad (10.12),

$$\mathrm{proy}_v(u) = \frac{\langle u, v\rangle}{\|v\|} \frac{v}{\|v\|},$$

a veces se llama **proyección ortogonal (escalar)** al número real $\frac{\langle u,v\rangle}{\|v\|}$ (puesto que el vector unitario $\frac{v}{\|v\|}$ indica la dirección sobre la que se proyecta). Este número puede ser positivo, negativo o nulo. Por este hecho, se dice que la proyección ortogonal (escalar) de un vector sobre un vector unitario es el producto escalar entre ambos vectores. Se llama **vector proyección ortogonal** al vector $\mathrm{proy}_v(u)$ definido en (10.12). La norma del vector proyección ortogonal es:

$$\|\mathrm{proy}_v(u)\| = \left\| \frac{\langle u, v\rangle}{\|v\|^2} v \right\| = \left\| \frac{\langle u, v\rangle}{\|v\|} \frac{v}{\|v\|} \right\| = \left| \frac{\langle u, v\rangle}{\|v\|} \right| \left\| \frac{v}{\|v\|} \right\| = \frac{|\langle u, v\rangle|}{\|v\|}.$$

Observación 10.16. En la Proposición 10.7 se ha visto que los coeficientes generalizados de Fourier de un vector $u \in E$ con respecto a una base ortogonal $\mathcal{B} = \{u_1, u_2, \ldots, u_n\}$ de E tienen un aspecto concreto en términos del producto escalar. Ahora se pueden reescribir en términos de las proyecciones ortogonales del vector u sobre cada uno de los ejes que proporciona la base $\mathcal{B}$ como sigue:

$$\begin{aligned} u &= \frac{\langle u, u_1\rangle}{\langle u_1, u_1\rangle} u_1 + \frac{\langle u, u_2\rangle}{\langle u_2, u_2\rangle} u_2 + \cdots + \frac{\langle u, u_n\rangle}{\langle u_n, u_n\rangle} u_n \\ &= \mathrm{proy}_{u_1}(u) u_1 + \mathrm{proy}_{u_2}(u) u_2 + \cdots + \mathrm{proy}_{u_n}(u) u_n. \end{aligned}$$

El problema de la mejor aproximación: Sea $u \in E$. De entre todos los vectores de S, se trata de encontrar el vector $z \in S$ que sea el "más próximo" a u , es decir z debe cumplir

$$\|u - z\| \leq \|u - v\|, \ \forall v \in S.$$

La solución la proporciona el vector proyección ortogonal: $z = \mathrm{proy}_S(u)$.

Teorema 10.5 (Teorema de la mejor aproximación)**.** Sea $(E, \langle \cdot, \cdot \rangle)$ un espacio euclídeo y S un subespacio vectorial de dimensión finita. Dado $u \in E$, se cumple que

$$\|u - \mathrm{proy}_S(u)\| \leq \|u - v\|, \quad \forall v \in S.$$

Demostración. Es claro que, para todo $v \in S$, se cumple que

$$u - v = [u - \mathrm{proy}_S(u)] + [\mathrm{proy}_S(u) - v],$$

donde $\mathrm{proy}_S(u) - v \in S$ (pues $\mathrm{proy}_S(u) \in S$, $v \in S$ y S es un subespacio) y $u - \mathrm{proy}_S(u) \in S^\perp$ (de acuerdo con la Observación 10.15). Por el Teorema de Pitágoras,

$$\begin{aligned} \|u - v\|^2 &= \|u - \mathrm{proy}_S(u)\|^2 + \|\mathrm{proy}_S(u) - v\|^2 \\ &\geq \|u - \mathrm{proy}_S(u)\|^2. \end{aligned}$$

Tomando raíces cuadradas se tiene $\|u - \mathrm{proy}_S(u)\| \leq \|u - v\|$, para todo $v \in S$, como se quería probar. □

Si $d(\cdot,\cdot)$ denota la aplicación distancia inducida por el producto escalar $\langle\cdot,\cdot\rangle$ en un espacio euclídeo E, el resultado del teorema anterior se puede escribir como

$$d(u, \mathrm{proy}_S(u)) \leq d(u, v), \quad \forall v \in S.$$

Ejemplo 10.27. *Encontrar la mejor aproximación de la función $f(x) = |x| \in \mathcal{C}_{[-\pi,\pi]}$ en el subespacio*

$$S = \overline{\{1, \cos(x), \mathrm{sen}(x)\}}$$

mediante el producto escalar

$$\langle g, h\rangle = \int_{-\pi}^{\pi} g(x)h(x)\,\mathrm{d}x.$$

◁ En el Ejemplo 10.16 se ha probado que el sistema trigonométrico es un sistema ortogonal en $[-\pi, \pi]$. En particular, el subconjunto

$$\{1, \cos(x), \mathrm{sen}(x)\}$$

también lo es; es decir, forma una base ortogonal del subespacio S con respecto al producto escalar dado.

Por el Teorema 10.5, la mejor aproximación de f por un vector del subespacio S viene dada por

$$\begin{aligned}
&\operatorname{proy}_S(f) = \\
&= \frac{\langle f(x), 1\rangle}{\langle 1, 1\rangle}1 + \frac{\langle f(x), \cos(x)\rangle}{\langle \cos(x), \cos(x)\rangle}\cos(x) + \frac{\langle f(x), \operatorname{sen}(x)\rangle}{\langle \operatorname{sen}(x), \operatorname{sen}(x)\rangle}\operatorname{sen}(x) \\
&= \frac{\pi^2}{2\pi}1 + \frac{-4}{\pi}\cos(x) + \frac{0}{\pi}\operatorname{sen}(x) \\
&= \frac{\pi}{2} - \frac{4}{\pi}\cos(x).
\end{aligned}$$

Para obtener los productos escalares anteriores se requieren calcular 6 integrales y tener en cuenta la definición de valor absoluto. Por ejemplo, utilizando la paridad de la función $|x|\cos(x)$,

$$\langle f(x), \cos(x)\rangle = \int_{-\pi}^{\pi} |x|\cos(x)\,\mathrm{d}x = 2\int_0^{\pi} x\cos(x)\,\mathrm{d}x,$$

donde en la última integral se debe aplicar el método integración por partes. El resto de detalles se proponen como ejercicio.

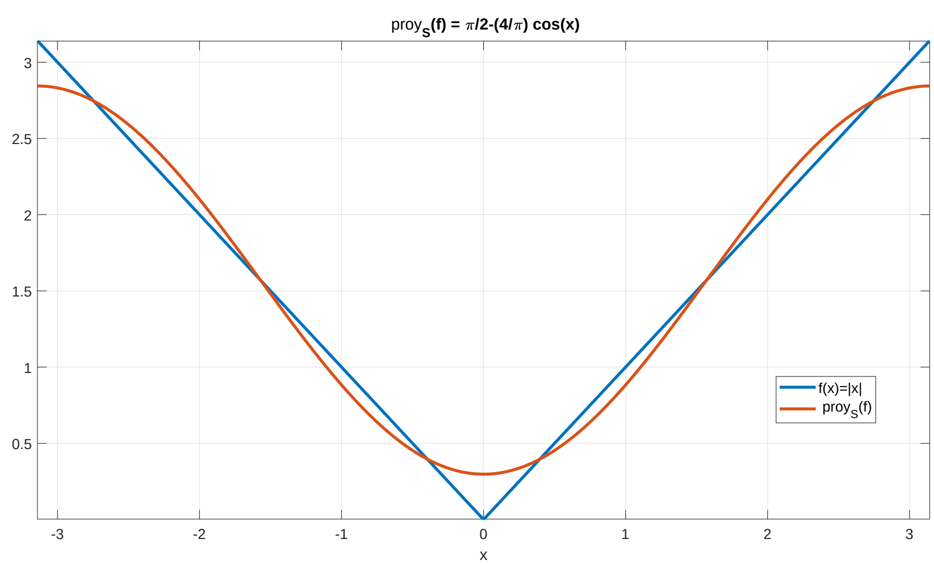


Figura 10.5: Gráfica de $f(x) = |x|$ y su mejor aproximación $\operatorname{proy}_S(f)$.

En la Figura 10.5 se pueden apreciar la gráfica de la función $f(x) = |x|$ y de su mejor aproximación $\mathrm{proy}_S(f) = \frac{\pi}{2} - \frac{4}{\pi}\cos(x)$. $\triangleright$

Aproximación por polinomios trigonométricos de grado n: En general, si $f \in \mathcal{C}_{[-\pi,\pi]}$, se encontrará su mejor aproximación por vectores del subespacio vectorial $S := \overline{S_n}$ donde

$$S_n := \{1, \cos(x), \mathrm{sen}(x), \cos(2x), \mathrm{sen}(2x), \ldots, \cos(nx), \mathrm{sen}(nx)\}$$

para algún $n \in \{0\} \cup \mathbb{N}$. Es claro que S_n es una base ortogonal de S (pues se trata de un subconjunto del sistema trigonométrico, que es ortogonal en $\mathcal{C}_{[-\pi,\pi]}$, y por tanto, S_n es linealmente independiente). Cuando $n = 0$ se debe interpretar que es $S_0 := \{1\}$. Es claro que $\dim_{\mathbb{R}}(S) = 2n + 1$.

Luego, la mejor aproximación de f por vectores del subespacio S viene dada por

$$\begin{aligned} f(x) \approx \mathrm{proy}_S(f) = & \\ = \; & \frac{\langle f(x), 1\rangle}{\langle 1, 1\rangle}1 + \frac{\langle f(x), \cos(x)\rangle}{\langle \cos(x), \cos(x)\rangle}\cos(x) + \frac{\langle f(x), \mathrm{sen}(x)\rangle}{\langle \mathrm{sen}(x), \mathrm{sen}(x)\rangle}\mathrm{sen}(x) + \cdots + \\ & + \frac{\langle f(x), \cos(nx)\rangle}{\langle \cos(nx), \cos(nx)\rangle}\cos(nx) + \frac{\langle f(x), \mathrm{sen}(nx)\rangle}{\langle \mathrm{sen}(nx), \mathrm{sen}(nx)\rangle}\mathrm{sen}(nx) \\ = \; & \frac{\langle f(x), 1\rangle}{2\pi}1 + \frac{\langle f(x), \cos(x)\rangle}{\pi}\cos(x) + \frac{\langle f(x), \mathrm{sen}(x)\rangle}{\pi}\mathrm{sen}(x) + \cdots + \\ & + \frac{\langle f(x), \cos(nx)\rangle}{\pi}\cos(nx) + \frac{\langle f(x), \mathrm{sen}(nx)\rangle}{\pi}\mathrm{sen}(nx), \end{aligned} \tag{10.13}$$

donde, para establecer la última igualdad, se han calculado las integrales que aparecen en los denominadores. Ahora, teniendo en cuenta las expresiones de los numeradores, se introduce, para $k = 1, 2, 3, \ldots$, la siguiente notación para los coeficientes

$$a_0 := \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\,\mathrm{d}x, \qquad a_k := \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos(kx)\,\mathrm{d}x$$

y

$$b_k := \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\,\mathrm{sen}(kx)\,\mathrm{d}x,$$

con lo que la última expresión obtenida en (10.13) se expresa de forma más sencilla mediante

$$\begin{aligned} f(x) &= \frac{a_0}{2} + a_1 \cos(x) + b_1 \operatorname{sen}(x) + \cdots + a_n \cos(nx) + b_n \operatorname{sen}(nx) \\ &= \frac{a_0}{2} + \sum_{k=1}^{n} [a_k \cos(kx) + b_k \operatorname{sen}(kx)], \end{aligned}$$

expresión que se denomina **polinomio trigonométrico** de grado n.

En Análisis de Fourier se estudian condiciones bajo las cuales al hacer que $n \to +\infty$ se tiene que

$$SP_n(f)(x) := \frac{a_0}{2} + \sum_{k=1}^{n} [a_k \cos(kx) + b_k \operatorname{sen}(kx)] \quad \to \quad f(x),$$

en el sentido que

$$\lim_{n \to +\infty} \|f - SP_n(f)\| = 0.$$

Se obtiene la **Serie de Fourier**

$$f(x) \approx \frac{a_0}{2} + \sum_{k=1}^{+\infty} [a_k \cos(kx) + b_k \operatorname{sen}(kx)],$$

en la cual el polinomio trigonométrico $SP_n(f)$ es la suma parcial n-ésima. Se deberá estudiar para qué valores de x se puede cambiar la aproximación por una igualdad.

10.6. Isomorfismo de espacios euclídeos

Definición 10.13 (Espacios euclídeos isomorfos). Dos espacios euclídeos $(E_1, \oplus, \odot, [\cdot,\cdot])$ y $(E_2, +, \cdot, \langle\cdot,\cdot\rangle)$ sobre el cuerpo $\mathbb{R}$ se dicen **isomorfos**, y se denota $E_1 \cong E_2$, si existe una aplicación biyectiva

$$f : E_1 \to E_2$$

que respeta las operaciones de los espacios como espacios vectoriales y el producto escalar, es decir,

$$\begin{array}{rcll} f(u_1 \oplus u_2) &=& f(u_1) + f(u_2), & \forall u_1, u_2 \in E_1, \\ f(\lambda \odot u) &=& \lambda \cdot f(u), & \forall u \in E_1, \forall \lambda \in \mathbb{R}, \\ [u_1, u_2] &=& \langle f(u_1), f(u_2)\rangle, & \forall u_1, u_2 \in E_1. \end{array} \tag{10.14}$$

La aplicación f se llama **isomorfismo** entre los espacios euclídeos E_1 y E_2.

Teorema 10.6 (Isomorfismo de Descartes para espacios euclídeos). Sea $(E, [\cdot,\cdot])$ un espacio euclídeo real con $\dim_{\mathbb{R}}(E) = n \geq 1$ y sea $\mathbb{R}^n$ el $\mathbb{R}$-espacio vectorial con la adición y la multiplicación de un escalar por un vector habituales y el producto escalar canónico $\langle\cdot,\cdot\rangle$. Entonces existe una aplicación biyectiva

$$\mathcal{D} : E \to \mathbb{R}^n$$

que satisface:

$$\begin{array}{rcll} \mathcal{D}(u_1 + u_2) &=& \mathcal{D}(u_1) + \mathcal{D}(u_2), & \forall u_1, u_2 \in E_1, \\ \mathcal{D}(\lambda u_1) &=& \lambda \mathcal{D}(u_1), & \forall u_1 \in E_1, \forall \lambda \in \mathbb{R}, \\ [u_1, u_2] &=& \langle \mathcal{D}(u_1), \mathcal{D}(u_2)\rangle, & \forall u_1, u_2 \in E_1. \end{array} \tag{10.15}$$

En consecuencia, E y $\mathbb{R}^n$ son espacios euclídeos isomorfos (sobre $\mathbb{R}$), es decir:

$$E \cong \mathbb{R}^n \text{ (el segundo con el producto escalar canónico).}$$

Demostración. Por el Teorema 10.3, es posible considerar una base ortonormal $\mathcal{B} = \{b_1, b_2, \ldots, b_n\}$ del espacio euclídeo E, es decir se cumple que

$$[b_i, b_j] = \delta_{ij} \qquad \text{para todo} \qquad i, j \in \{1, 2, \ldots, n\}.$$

Se considera ahora el isomorfismo de Descartes

$$\begin{array}{rcccc} \mathcal{D} & : & E & \to & \mathbb{R}^n \\ & & u & \mapsto & \mathcal{D}(u) = [u]_{\mathcal{B}} = (\alpha_1, \alpha_2, \ldots, \alpha_n) \end{array}$$

que a cada vector $u \in E$ escrito en la base ortonormal $\mathcal{B}$ mediante

$$u = \alpha_1 b_1 + \alpha_2 b_2 + \cdots + \alpha_n b_n$$

(con $\alpha_1, \alpha_2, \ldots, \alpha_n \in \mathbb{R}$ unívocamente determinados) le asocia el vector de $\mathbb{R}^n$ formado por dichas coordenadas, es decir: $[u]_{\mathcal{B}} = (\alpha_1, \alpha_2, \ldots, \alpha_n)$.

En el Teorema 9.1 se ha probado que $\mathcal{D}$ es biyectiva y respeta la suma y el producto de un escalar por un vector.

Para que $\mathcal{D}$ sea un isomorfismo de espacios euclídeos falta probar que $\mathcal{D}$ respeta el producto escalar, es decir

$$[u_1, u_2] = \langle \mathcal{D}(u_1), \mathcal{D}(u_2) \rangle, \qquad \forall u_1, u_2 \in E_1.$$

En efecto, para $u_1, u_2 \in E_1$ se tiene que

$$u_1 = \alpha_1 b_1 + \alpha_2 b_2 + \cdots + \alpha_n b_n = \sum_{i=1}^{n} \alpha_i b_i$$

y

$$u_2 = \beta_1 b_1 + \beta_2 b_2 + \cdots + \beta_n b_n = \sum_{j=1}^{n} \beta_j b_j.$$

Luego, al calcular el producto escalar en E_1 se obtiene

$$\begin{aligned}
[u_1, u_2] &= \left[\sum_{i=1}^{n} \alpha_i b_i, \sum_{j=1}^{n} \beta_j b_j\right] \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \beta_j [b_i, b_j] \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \beta_j \delta_{ij} \\
&= \sum_{i=1}^{n} \alpha_i \beta_i \\
&= \langle (\alpha_1, \alpha_2, \ldots, \alpha_n), (\beta_1, \beta_2, \ldots, \beta_n) \rangle \\
&= \langle \mathcal{D}(u_1), \mathcal{D}(u_2) \rangle.
\end{aligned}$$

Por lo tanto, $\mathcal{D}$ es un isomorfismo de espacios euclídeos y $E \cong \mathbb{R}^n$. □

Del mismo modo que en el Teorema 9.1, la aplicación $\mathcal{D}$ depende de la base ortonormal seleccionada en la cual se tomarán sus coordenadas y, si se quisiera remarcar dicha dependencia, se podría escribir $\mathcal{D}_{\mathcal{B}}$.

A continuación se deben establecer resultados similares a los de la Sección 9.4, donde ahora los espacios que se consideran son (no sólo vectoriales sino además) euclídeos reales. Para establecer dichos resultados la única novedad será que se deben realizar las comprobaciones relativas al producto escalar, que se proponen como ejercicio. Se enuncian los resultados correspondientes:

- Sean $(E_1, [\cdot, \cdot])$ y $(E_2, \langle \cdot, \cdot \rangle)$ dos espacios euclídeos reales. Si $f : E_1 \to E_2$ es un isomorfismo de espacios euclídeos entonces $f^{-1} : E_2 \to E_1$ es un isomorfismo de espacios euclídeos (véase la Proposición 9.2).

- Se considera un conjunto de espacios euclídeos reales. Se define la relación binaria, llamada **relación de isomorfía euclídea**, dada por

$$E_1 \, \mathcal{R} \, E_2 \qquad \Leftrightarrow \qquad E_1 \cong E_2,$$

donde $(E_1, [\cdot,\cdot])$ y $(E_2, \langle\cdot,\cdot\rangle)$ son dos espacios euclídeos en dicho conjunto. La relación de isomorfía euclídea así definida es una relación de equivalencia (véase el Teorema 9.2).

- Sean E_1 y E_2 dos $\mathbb{R}$-espacios euclídeos isomorfos de dimensión finita. Entonces
$$\dim_{\mathbb{R}}(E_1) = \dim_{\mathbb{R}}(E_2)$$
(véase el Lema 9.1).

- Sean E_1 y E_2 dos $\mathbb{R}$-espacios euclídeos isomorfos de dimensión finita tales que $\dim_{\mathbb{R}}(E_1) = \dim_{\mathbb{R}}(E_2) \geq 1$. Entonces
$$E_1 \cong E_2$$
(véase el Lema 9.2).

Para la prueba del siguiente resultado véase el Teorema 9.3.

Teorema 10.7 (Caracterización de espacios euclídeos isomorfos)**.** Sean $(E_1, [\cdot,\cdot])$ y $(E_2, \langle\cdot,\cdot\rangle)$ dos $\mathbb{R}$-espacios euclídeos de dimensión finita. Entonces
$$E_1 \cong E_2 \qquad \Leftrightarrow \qquad \dim_{\mathbb{R}}(E_1) = \dim_{\mathbb{R}}(E_2).$$

Se habla entonces de **unicidad** en el sentido de que para cada $n \in \mathbb{N}$, existe un único $\mathbb{R}$-espacio euclídeo (salvo isomorfismos) de dimensión n. Además, si $\mathbb{R}^n$ (con las operaciones habituales y el producto escalar canónico) es un elemento de un conjunto de espacios euclídeos considerado, $\mathbb{R}^n$ puede tomarse como el **representante** *más sencillo* de la clase a la que pertenece.

Se ha construido un **invariante** en la teoría de espacios euclídeos (de dimensión finita): la dimensión del $\mathbb{R}$-espacio vectorial asociado. Del mismo modo que para espacios vectoriales de dimensión finita, resulta que la dimensión es, además, un **conjunto completo de invariantes**. Es claro entonces que la dimensión es una característica intrínseca de cada espacio euclídeo (independientemente de la base ortonormal considerada).

10.7. Matriz de Gram

Problema: ¿Es posible encontrar una expresión del producto escalar a partir de su cálculo en los elementos de una base?

Esta pregunta ha sido respondida previamente para el caso de una base ortonormal en la página 566. Se verá que, en general, la respuesta sigue siendo afirmativa, se hallará una matriz (llamada matriz de Gram) que contiene toda la información necesaria para ello y se estudiarán algunas de sus propiedades.

10.7.1. Definición de matriz de Gram

Sea $(E, \langle\cdot,\cdot\rangle)$ un espacio euclídeo de dimensión finita y

$$\mathcal{B} = \{u_1, u_2, \ldots, u_n\}$$

una base ordenada de E.

Dados dos vectores $x, y \in E$, existen escalares únicos $x_1, x_2, \ldots, x_n \in \mathbb{R}$ e $y_1, y_2, \ldots, y_n \in \mathbb{R}$ tales que

$$x = x_1u_1 + x_2u_2 + \cdots + x_nu_n, \qquad y = y_1u_1 + y_2u_2 + \cdots + y_nu_n.$$

Por las propiedades de linealidad del producto escalar se tiene

$$\begin{aligned}
\langle x, y\rangle &= \left\langle \sum_{i=1}^{n} x_iu_i, \sum_{j=1}^{n} y_ju_j \right\rangle \\
&= \sum_{i=1}^{n} x_i \left\langle u_i, \sum_{j=1}^{n} y_ju_j \right\rangle \\
&= \sum_{i=1}^{n} x_i \sum_{j=1}^{n} y_j \langle u_i, u_j\rangle \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} x_iy_j \langle u_i, u_j\rangle \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} x_i \langle u_i, u_j\rangle y_j.
\end{aligned}$$

Si se define la matriz

$$G_{\mathcal{B}} := \begin{bmatrix} \langle u_1, u_1\rangle & \langle u_1, u_2\rangle & \dots & \langle u_1, u_n\rangle \\ \langle u_2, u_1\rangle & \langle u_2, u_2\rangle & \dots & \langle u_2, u_n\rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle u_n, u_1\rangle & \langle u_n, u_2\rangle & \dots & \langle u_n, u_n\rangle \end{bmatrix},$$

conocida como **matriz de Gram** o **matriz métrica** del producto escalar $\langle\cdot,\cdot\rangle$ con respecto a la base $\mathcal{B}$, es posible escribir

$$\langle x, y\rangle = \begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix} G_{\mathcal{B}} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = ([x]_{\mathcal{B}})^t G_{\mathcal{B}} [y]_{\mathcal{B}}.$$

En el razonamiento previo se ha probado el siguiente resultado:

Teorema 10.8 (Cálculo del producto escalar de dos vectores mediante la matriz de Gram)**.** Sea $(E, \langle\cdot,\cdot\rangle)$ un espacio euclídeo de dimensión $n \geq 1$, $\mathcal{B} = \{u_1, u_2, \dots, u_n\}$ una base ordenada de E y $G_{\mathcal{B}}$ la matriz de Gram del producto escalar $\langle\cdot,\cdot\rangle$ en la base $\mathcal{B}$. Si $x, y \in E$ se tiene que

$$\langle x, y\rangle = ([x]_{\mathcal{B}})^t G_{\mathcal{B}} [y]_{\mathcal{B}}. \tag{10.16}$$

Observación 10.17. La expresión (10.16) permite calcular el producto escalar de dos vectores conociendo la matriz de Gram en una base ordenada y las coordenadas de ambos vectores en dicha base. Es importante tener presente que el resultado es *independiente de la base* utilizada para el cálculo (sólo es necesario que todos, las coordenadas de los vectores y la matriz de Gram, estén referidos a la misma base).

Ejemplo 10.28. *Calcular la matriz de Gram del producto escalar canónico de $\mathbb{R}^2$ con respecto a la base canónica. Generalizar a $\mathbb{R}^n$ para todo $n \geq 2$.*

◁ La base canónica $\mathcal{C} = \{e_1 = (1,0), e_2 = (0,1)\}$ de $\mathbb{R}^2$ satisface que

$$G_{\mathcal{C}} = \begin{bmatrix} \langle e_1, e_1 \rangle & \langle e_1, e_2 \rangle \\ \langle e_2, e_1 \rangle & \langle e_2, e_2 \rangle \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I_2.$$

Es claro que en $\mathbb{R}^n$ se tiene que $\langle e_i, e_j \rangle = \delta_{ij}$ para todo $1 \leq i, j \leq n$. Luego, $G_{\mathcal{C}} = I_n$. ▷

Ejemplo 10.29. *Calcular la matriz de Gram del producto escalar*

$$\langle (x_1, y_1, z_1), (x_2, y_2, z_2) \rangle = \begin{bmatrix} x_1 & y_1 & z_1 \end{bmatrix} \begin{bmatrix} 2 & -1 & 0 \\ -1 & 3 & 0 \\ 0 & 0 & 6 \end{bmatrix} \begin{bmatrix} x_2 \\ y_2 \\ z_2 \end{bmatrix}$$

de $\mathbb{R}^3$ con respecto a:

(a) la base canónica de $\mathbb{R}^3$.

(b) la base $\mathcal{B} = \{u_1 = (1,1,1), u_2 = (1,1,0), u_3 = (1,0,0)\}$ de $\mathbb{R}^3$.

◁ Se observa que se pide calcular la matriz de Gram con respecto a dos bases diferentes de $\mathbb{R}^3$.

(a) El producto escalar de los dos primeros vectores de la base canónica da

$$\langle e_1, e_2 \rangle = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 2 & -1 & 0 \\ -1 & 3 & 0 \\ 0 & 0 & 6 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = -1.$$

Calculando los restantes productos escalares, la matriz de Gram en dicha base es

$$G_{\mathcal{C}} = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 3 & 0 \\ 0 & 0 & 6 \end{bmatrix}.$$

Se observa que se obtiene exactamente la misma matriz que se utiliza en la definición del producto escalar del enunciado.

(b) El producto escalar de los dos primeros vectores de la base $\mathcal{B}$ da

$$\langle u_1, u_2 \rangle = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 2 & -1 & 0 \\ -1 & 3 & 0 \\ 0 & 0 & 6 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = 3.$$

Calculando los restantes productos escalares, la matriz de Gram en la base $\mathcal{B}$ es

$$G_{\mathcal{B}} = \begin{bmatrix} 9 & 3 & 1 \\ 3 & 3 & 1 \\ 1 & 1 & 2 \end{bmatrix}.$$

Como era de esperar, las matrices de Gram en distintas bases han resultado ser distintas entre sí. $\triangleright$

10.7.2. Propiedades de la matriz de Gram

El primer resultado presenta un criterio general para la independencia lineal (de un conjunto finito de vectores) en un espacio vectorial dotado de un producto escalar. Para ello, se considerarán matrices, que llamaremos **matrices de tipo Gram**, por estar construidas de forma similar a las mismas.

Proposición 10.12 (Matriz de tipo Gram vs. independencia lineal)**.** Sea $(E, \langle\cdot,\cdot\rangle)$ un espacio euclídeo real de dimensión finita $n \geq 1$ y $\{v_1, v_2, \ldots, v_s\} \subseteq E$. Si se calculan los s^2 productos escalares dispuestos en la **matriz de tipo Gram**

$$G(v_1, v_2, \ldots, v_s) := \begin{bmatrix} \langle v_1, v_1\rangle & \langle v_1, v_2\rangle & \ldots & \langle v_1, v_s\rangle \\ \langle v_2, v_1\rangle & \langle v_2, v_2\rangle & \ldots & \langle v_2, v_s\rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle v_s, v_1\rangle & \langle v_s, v_2\rangle & \ldots & \langle v_s, v_s\rangle \end{bmatrix} \in \mathbb{R}^{s\times s}$$

entonces

$$\det(G(v_1, v_2, \ldots, v_s)) = 0 \quad \Leftrightarrow \quad \{v_1, v_2, \ldots, v_s\} \text{ es LD.}$$

En consecuencia, si $\{v_1, v_2, \ldots, v_n\}$ es una base de E entonces $\det(G(v_1, v_2, \ldots, v_n)) \neq 0$.

Demostración. Se debe probar la doble implicación.

($\Rightarrow$) Si $\det(G(v_1, v_2, \ldots, v_s)) = 0$ entonces $\mathrm{rg}(G(v_1, v_2, \ldots, v_s)) < s$, donde la matriz está formada por s vectores columna c_i, $i = 1, 2, \ldots, s$, de $\mathbb{R}^s$. Por la Proposición 9.3, $\{c_1, c_2, \ldots, c_s\}$ es linealmente dependiente. Luego, existe $i \in \{1, 2, \ldots, s\}$ tal que la columna i-ésima es combinación lineal de las restantes, es decir, para ciertos escalares $a_i \in \mathbb{R}$ se tiene

$$\begin{bmatrix} \langle v_1, v_i\rangle \\ \langle v_2, v_i\rangle \\ \vdots \\ \langle v_s, v_i\rangle \end{bmatrix} = c_i = \sum_{\substack{j=1 \\ j\neq i}}^{s} a_j c_j = \sum_{\substack{j=1 \\ j\neq i}}^{s} a_j \begin{bmatrix} \langle v_1, v_j\rangle \\ \langle v_2, v_j\rangle \\ \vdots \\ \langle v_s, v_j\rangle \end{bmatrix}.$$

Igualando componentes en cada fila, para cada $k = 1, 2, \ldots, s$, se tiene que

$$\langle v_k, v_i\rangle = \sum_{\substack{j=1 \\ j\neq i}}^{s} a_j \langle v_k, v_j\rangle,$$

que, por la linealidad del producto escalar, se puede reescribir como

$$\left\langle v_k, v_i - \sum_{\substack{j=1 \\ j \neq i}}^{s} a_j v_j \right\rangle = 0, \qquad \text{para todo } k = 1, 2, \ldots, s.$$

La última expresión indica que el vector $v_i - \sum^{s}_{\substack{j=1 \\ j \neq i}} a_j v_j$ es combinación lineal de $v_1, v_2, \ldots, v_s$ y, además, es ortogonal a todos ellos. Es fácil probar (se propone hacerlo como ejercicio) que dicho vector debe coincidir con el vector nulo. Por lo tanto,

$$v_i = \sum_{\substack{j=1 \\ j \neq i}}^{s} a_j v_j,$$

lo que indica que $\{v_1, v_2, \ldots, v_s\}$ es un conjunto linealmente dependiente.

($\Leftarrow$) Si $\{v_1, v_2, \ldots, v_s\}$ es linealmente dependiente, existe algún vector v_i de dicho conjunto que es combinación lineal de los demás, es decir

$$v_i = \sum_{j=1, j \neq i}^{s} a_j v_j.$$

Teniendo en cuenta la linealidad del producto escalar se llega a que, para cada $k = 1, 2, \ldots, s$ se tiene que

$$\langle v_k, v_i \rangle = \left\langle v_k, \sum_{\substack{j=1 \\ j \neq i}}^{s} a_j v_j \right\rangle = \sum_{\substack{j=1 \\ j \neq i}}^{s} a_j \langle v_k, v_j \rangle.$$

Al sustituir en la matriz $G(v_1, v_2, \ldots, v_s))$ resulta que la columna i-ésima es combinación lineal de las restantes. Luego, de las Proposiciones 7.1, 7.2 y 7.3 (para columnas), el determinante de $G(v_1, v_2, \ldots, v_s))$ debe valer cero. $\square$

El método proporcionado por el resultado anterior es útil a la hora de querer demostrar la dependencia o independencia lineal de un conjunto de s vectores en un espacio euclídeo de dimensión n cuando $s < n$.

Ejemplo 10.30. *Probar que*

$$\{v_1 = (3, 2, 0, 0, -1, 1), v_2 = (1, 2, 0, 0, 2, 1)\}$$

es un conjunto linealmente independiente en $\mathbb{R}^6$.

$\triangleleft$ Como es posible pensar al espacio vectorial $\mathbb{R}^6$ como un espacio euclídeo con el producto escalar canónico, y la matriz

$$G(v_1, v_2) = \begin{bmatrix} \langle v_1, v_1\rangle & \langle v_1, v_2\rangle \\ \langle v_2, v_1\rangle & \langle v_2, v_2\rangle \end{bmatrix} = \begin{bmatrix} 15 & 6 \\ 6 & 10 \end{bmatrix}$$

tiene claramente determinante no nulo, $\{v_1, v_2\}$ es un conjunto linealmente independiente en $\mathbb{R}^6$. $\triangleright$

Al demostrar que una aplicación definida como $\langle x, y\rangle = x^t Ay$, para $A \in \mathbb{R}^{n\times n}$ fija y $x, y \in \mathbb{R}^{n\times 1}$ se cumplen las propiedades (PE1)-(PE3) siempre que A sea una matriz simétrica. Sin embargo, es necesario alguna condición extra para que se cumpla (PE4). Para ello, antes de continuar con más propiedades, se define un nuevo tipo de matriz.

Definición 10.14 (Matriz definida positiva)**.** Una matriz simétrica $A \in \mathbb{R}^{n\times n}$ se llama **definida positiva** si $x^t Ax > 0$, para todo $x \in \mathbb{R}^{n\times 1} - \{0\}$.

Ahora sí se presentan propiedades que cumple toda matriz de Gram.

Proposición 10.13 (Propiedades de la matriz de Gram)**.** Sea $(E, \langle \cdot, \cdot \rangle)$ un espacio euclídeo real de dimensión finita $n \geq 1$ y sea $\mathcal{B} = \{u_1, u_2, \ldots, u_n\}$ una base ordenada de E. Se cumple que:

(a) $G_{\mathcal{B}}$ es una matriz simétrica.

(b) $G_{\mathcal{B}}$ es una matriz definida positiva (es decir, $([x]_{\mathcal{B}})^t G_{\mathcal{B}} [x]_{\mathcal{B}} > 0$, para todo $x \in E - \{0\}$[a]).

(c) $G_{\mathcal{B}}$ es una matriz invertible.

[a] En realidad, en lugar de "para todo $x \in E - \{0\}$" debería decir "para todo $[x]_{\mathcal{B}} \in \mathbb{R}^{n\times 1} - \{0\}$", lo cual es equivalente pues por el isomorfismo de Descartes, para cada $x \in E$, existe un único $[x]_{\mathcal{B}} \in \mathbb{R}^{n\times 1}$ tal que $\mathcal{D}(x) = [x]_{\mathcal{B}}$.

Demostración. (a) Es inmediato que la matriz $G_{\mathcal{B}}$ es simétrica puesto que $\langle u_i, u_j \rangle = \langle u_j, u_i \rangle$ para todo $1 \leq i, j \leq n$ debido a la propiedad de simetría del producto escalar.

(b) De la expresión (10.16) y de la propiedad de positividad del producto escalar se tiene que

$$([x]_{\mathcal{B}})^t G_{\mathcal{B}} [x]_{\mathcal{B}} = \langle x, x \rangle > 0, \qquad \text{para todo } x \in E - \{0\}.$$

(c) Si $\mathcal{B} = \{u_1, u_2, \ldots, u_n\}$ es una base de E, de la Proposición 10.12 se tiene que $\det(G_{\mathcal{B}}) \neq 0$, y por lo tanto, $G_{\mathcal{B}}$ es una matriz invertible.

$\square$

Observación 10.18. En la Proposición anterior, el apartado (c) se puede deducir sin acudir a la Proposición 10.12. Es importante recordar previamente que un vector z es nulo si y sólo si sus coordenadas respecto de una base cualquiera son todas nulas.
En efecto, sea $y \in E$ un vector tal que $G_{\mathcal{B}}[y]_{\mathcal{B}} = 0$. Multiplicando a izquierda dicha igualdad por $([y]_{\mathcal{B}})^t$ se tiene que

$$0 = ([y]_{\mathcal{B}})^t G_{\mathcal{B}}[y]_{\mathcal{B}} = \langle y, y \rangle.$$

De la definición positiva del producto escalar se deduce que $y = 0$. De este modo, $G_{\mathcal{B}}$ es invertible.

Las propiedades de la Proposición 10.13 permiten establecer que, fijada una base, la matriz de Gram relativa a dicha base caracteriza completamente el producto escalar, es decir un producto interno en un espacio euclídeo E está determinado por los escalares $\langle u_i, u_j \rangle$ para todo $i, j \in \{1, 2, \ldots, n\}$ de una base ordenada $\mathcal{B} = \{u_1, u_2, \ldots, u_n\}$ de E.

Teorema 10.9 (La matriz de Gram caracteriza el producto escalar)**.** Sea $(E, \langle \cdot, \cdot \rangle)$ un espacio euclídeo real de dimensión finita y sea $\mathcal{B} = \{u_1, u_2, \ldots, u_n\}$ una base ordenada de E. Entonces la matriz de Gram del producto escalar $\langle \cdot, \cdot \rangle$ con respecto a dicha base es simétrica, definida positiva y satisface que

$$\langle x, y \rangle = ([x]_{\mathcal{B}})^t G_{\mathcal{B}}[y]_{\mathcal{B}} \quad \text{para todo } x, y \in E.$$

Además, si $G \in \mathbb{R}^{n \times n}$ es una matriz simétrica y definida positiva entonces la expresión

$$\langle x, y \rangle := x^t G y, \text{ para todo } x, y \in \mathbb{R}^n \tag{10.17}$$

define un producto escalar en $\mathbb{R}^n$.

Demostración. La primera parte del enunciado ha sido probada en la Proposición 10.13 y la fórmula (10.16). Para demostrar la afirmación restante, se debe comprobar que la expresión (10.17) satisface los axiomas de un producto escalar; se propone como ejercicio. $\square$

Ejemplo 10.31. *Averiguar si la aplicación dada por*

$$\langle (x_1, x_2), (y_1, y_2) \rangle = 3x_1y_1 + 3x_1y_2 - 3x_2y_2, \forall (x_1, x_2), (y_1, y_2) \in \mathbb{R}^2$$

define un producto escalar sobre $\mathbb{R}^2$.

$\lhd$ Se considera la base canónica $\mathcal{C} = \{e_1 = (1,0), e_2 = (0,1)\}$ de $\mathbb{R}^2$ y respecto de dicha base se tiene la matriz

$$\begin{bmatrix} \langle e_1, e_1 \rangle & \langle e_1, e_2 \rangle \\ \langle e_2, e_1 \rangle & \langle e_2, e_2 \rangle \end{bmatrix} = \begin{bmatrix} 3 & 3 \\ 0 & -3 \end{bmatrix}.$$

Dado que dicha matriz no es simétrica, la aplicación $\langle \cdot, \cdot \rangle$ proporcionada no define un producto escalar sobre $\mathbb{R}^2$. $\rhd$

Observación 10.19. Los elementos de la diagonal (principal) de la matriz de Gram de un producto escalar $\langle \cdot, \cdot \rangle$ respecto a una base $\mathcal{B} = \{u_1, u_2, \ldots, u_n\}$ son estrictamente positivos pues son de la forma $\langle u_i, u_i \rangle$, para todo $i = 1, 2, \ldots, n$ y $\langle u_i, u_i \rangle > 0$ por la condición de positividad de la definición de producto escalar.

Ejemplo 10.32. *Repetir el Ejercicio 10.31 utilizando la Observación 10.19.*

$\lhd$ De la Observación 10.19 se puede concluir que la aplicación $\langle \cdot, \cdot \rangle$ del Ejemplo 10.31 no define un producto escalar sobre $\mathbb{R}^2$ puesto que el elemento diagonal $\langle e_2, e_2 \rangle = -3$ no es estrictamente positivo. $\rhd$

Observación 10.20. La expresión (10.17) define un producto escalar en $\mathbb{R}^n$. Como el espacio de partida E es isomorfo a $\mathbb{R}^n$ (como espacios vectoriales, por tener la misma dimensión), es posible dotar al espacio E de un producto escalar, a partir del producto escalar de $\mathbb{R}^n$, mediante la expresión

$$[x, y] := (\mathcal{D}(x))^t G\mathcal{D}(y) = ([x]_{\mathcal{B}})^t G[y]_{\mathcal{B}}, \qquad \text{para todo } x, y \in E$$

siendo $\mathcal{D}$ el isomorfismo de Descartes del Teorema 10.6 de la página 590, y por tanto $[x]_{\mathcal{B}}$ e $[y]_{\mathcal{B}}$ son las imágenes de x e y por $\mathcal{D}$ para la base $\mathcal{B}$ del Teorema 10.9 y $G_{\mathcal{B}}$ (la matriz de Gram con respecto al producto escalar $[\cdot, \cdot]$) cumple que $G_{\mathcal{B}} = G$.
Esta última igualdad sale directa de multiplicar las matrices teniendo en cuenta que $[u_i]_{\mathcal{B}} = e_i$, el i-ésimo vector canónico de $\mathbb{R}^n$. De este modo se obtiene la implicación recíproca de la primera parte del enunciado del Teorema 10.9 (pues todo está referido a la misma base).

En definitiva, con la notación del Teorema 10.9, definiendo los conjuntos

$$\mathcal{PE}_E := \{\langle \cdot, \cdot \rangle : \langle \cdot, \cdot \rangle \text{ es un producto escalar en } E\}$$

y

$$\mathcal{MSDP} := \{G \in \mathbb{R}^{n\times n} : G \text{ es simétrica y definida positiva}\}$$

se tiene que la aplicación

$$\begin{array}{rccc} \varphi_{\mathcal{B}} & : & \mathcal{PE}_E & \rightarrow & \mathcal{MSDP} \\ & & \langle \cdot, \cdot \rangle & \mapsto & G_{\mathcal{B}} \end{array},$$

siendo $G_{\mathcal{B}}$ la matriz de Gram de $\langle \cdot, \cdot \rangle$ en la base $\mathcal{B}$, es biyectiva.

Este hecho garantiza que todo producto escalar tiene una única matriz simétrica y definida positiva asociada y viceversa. Por tanto, es equivalente hablar de un producto escalar o de su matriz simétrica y definida positiva asociada puesto que con cualquiera de los dos conceptos se obtendrá la misma información sobre el espacio euclídeo.

Relación entre matrices de Gram en diferentes bases

En el próximo resultado se analiza la relación que verifican las matrices de Gram en dos bases diferentes.

Proposición 10.14 (Matrices de Gram en diferentes bases)**.** Sea $(E, \langle\cdot,\cdot\rangle)$ un espacio euclídeo real de dimensión finita y sean $\mathcal{B}_1$ y $\mathcal{B}_2$ dos bases ordenadas de E. Entonces

$$G_{\mathcal{B}_2} = ([\mathcal{B}_2]_{\mathcal{B}_1})^t G_{\mathcal{B}_1} [\mathcal{B}_2]_{\mathcal{B}_1}.$$

Demostración. Sean $x, y \in E$. De la teoría de cambio de base en espacios vectoriales se sabe que

$$[x]_{\mathcal{B}_1} = [\mathcal{B}_2]_{\mathcal{B}_1}[x]_{\mathcal{B}_2}.$$

Del mismo modo,

$$[y]_{\mathcal{B}_1} = [\mathcal{B}_2]_{\mathcal{B}_1}[y]_{\mathcal{B}_2}$$

De la expresión (10.16), la matriz de Gram del producto escalar satisface

$$\begin{aligned} \langle x, y\rangle &= [x]^t_{\mathcal{B}_1} G_{\mathcal{B}_1}[y]_{\mathcal{B}_1} \\ &= ([\mathcal{B}_2]_{\mathcal{B}_1}[x]_{\mathcal{B}_2})^t G_{\mathcal{B}_1}[\mathcal{B}_2]_{\mathcal{B}_1}[y]_{\mathcal{B}_2} \\ &= [x]^t_{\mathcal{B}_2}[\mathcal{B}_2]^t_{\mathcal{B}_1} G_{\mathcal{B}_1}[\mathcal{B}_2]_{\mathcal{B}_1}[y]_{\mathcal{B}_2}. \end{aligned}$$

Por otro lado, de nuevo utilizando (10.16), se tiene que

$$\langle x, y\rangle = [x]^t_{\mathcal{B}_2} G_{\mathcal{B}_2}[y]_{\mathcal{B}_2}.$$

Igualando ambas expresiones,

$$[x]^t_{\mathcal{B}_2} G_{\mathcal{B}_2}[y]_{\mathcal{B}_2} = [x]^t_{\mathcal{B}_2}[\mathcal{B}_2]^t_{\mathcal{B}_1} G_{\mathcal{B}_1}[\mathcal{B}_2]_{\mathcal{B}_1}[y]_{\mathcal{B}_2}, \quad \text{para todo } x, y \in E.$$

Sea $\mathcal{B}_2 = \{v_1, v_2, \ldots, v_n\}$. La igualdad anterior es válida, en particular, para $x = v_i$, $y = v_j$, para todo $i, j = 1, 2, \ldots, n$. Realizando los productos indicados para todos los posibles pares ordenados $(i, j) \in \{1, 2, \ldots, n\} \times \{1, 2, \ldots, n\}$ se llega a

$$G_{\mathcal{B}_2} = [\mathcal{B}_2]^t_{\mathcal{B}_1} G_{\mathcal{B}_1} [\mathcal{B}_2]_{\mathcal{B}_1},$$

como se quería probar. $\square$

Se recuerda que la matriz de cambio de base $[\mathcal{B}_2]_{\mathcal{B}_1}$ es invertible.

Observación 10.21. Si $A, B \in \mathbb{R}^{n\times n}$ satisfacen la relación

$$B = P^t AP,$$

para cierta matriz invertible $P \in \mathbb{R}^{n\times n}$ se dice que B es una **matriz congruente** a A.
La congruencia de matrices (que es un caso particular de la equivalencia de matrices) es una relación de equivalencia sobre $\mathbb{R}^{n\times n}$. El resultado anterior se puede parafrasear diciendo que matrices de Gram correspondientes a bases diferentes son congruentes[a].

[a] Es posible realizar un estudio detallado de la congruencia de matrices, pero excede el alcance de este libro.

El siguiente ejemplo permite ver las relación entre las matrices de Gram en diferentes bases.

Ejemplo 10.33. *Utilizando la matriz de Gram hallada en el Ejemplo 10.29 en la base canónica, hallar la matriz de Gram en la base $\mathcal{B}$ del apartado (b) y comprobar que coincide con la obtenida previamente. Verificar que ambas matrices son simétricas y definidas positivas.*

$\triangleleft$ A partir de la matriz

$$G_{\mathcal{C}} = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 3 & 0 \\ 0 & 0 & 6 \end{bmatrix}$$

se puede obtener la matriz de Gram $G_{\mathcal{B}}$ utilizando la Proposición 10.14. Para ello es necesario conocer previamente la matriz de cambio de base dada por

$$[\mathcal{B}]_{\mathcal{C}} = \begin{bmatrix} [u_1]_{\mathcal{C}} & [u_2]_{\mathcal{C}} & [u_3]_{\mathcal{C}} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}.$$

Luego,

$$\begin{aligned} G_{\mathcal{B}} &= [\mathcal{B}]_{\mathcal{C}}^t G_{\mathcal{C}} [\mathcal{B}]_{\mathcal{C}} \\ &= \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 2 & -1 & 0 \\ -1 & 3 & 0 \\ 0 & 0 & 6 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \\ &= \begin{bmatrix} 9 & 3 & 1 \\ 3 & 3 & 1 \\ 1 & 1 & 2 \end{bmatrix}, \end{aligned}$$

que coincide con el resultado obtenido en el Ejercicio 10.29.

Es evidente que $G_{\mathcal{C}}$ es simétrica.

Para verificar que $G_{\mathcal{C}}$ es definida positiva se aplica el conocido como **método de Gauss de descomposición en cuadrados**, que consiste en construir sumas de cuadrados de combinaciones lineales de monomios lineales en una sola variable.

A continuación se describe dicho método.

Multiplicando las matrices, primero se obtiene una función polinomial de segundo grado en las 3 variables x, y y z del tipo

$$p(x) = a_{11}x^2 + a_{22}y^2 + a_{33}z^2 + 2a_{12}xy + 2a_{13}xz + 2a_{23}yz,$$

donde se supone que $a_{11} \neq 0$. En este caso, si en primer lugar se fija la atención en la variable x, se tiene

$$
\begin{aligned}
p(x) &= a_{11}x^2 + 2x(a_{12}y + a_{13}z) + (a_{22}y^2 + a_{33}z^2 + 2a_{23}yz) \\
&= a_{11}\left[x^2 + 2x\frac{a_{12}y + a_{13}z}{a_{11}} + \left(\frac{a_{12}y + a_{13}z}{a_{11}}\right)^2 - \left(\frac{a_{12}y + a_{13}z}{a_{11}}\right)^2\right] + \\
&\quad + (a_{22}y^2 + a_{33}z^2 + 2a_{23}yz) \\
&= a_{11}\left[x + \frac{a_{12}y + a_{13}z}{a_{11}}\right]^2 + a_{22}y^2 + a_{33}z^2 + 2a_{23}yz - \frac{(a_{12}y + a_{13}z)^2}{a_{11}}
\end{aligned}
$$

donde se ha completado el primer cuadrado y para los dos cuadrados restantes se procede de modo similar; se observa que ahora queda una expresión en 2 variables.

En el caso del ejemplo se procede como sigue,

$$
\begin{aligned}
u^t G_{\mathcal{C}} u &= \begin{bmatrix} x & y & z \end{bmatrix} \begin{bmatrix} 2 & -1 & 0 \\ -1 & 3 & 0 \\ 0 & 0 & 6 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} \\
&= 2x^2 - 2xy + 3y^2 + 6z^2 \\
&= 2\left[x^2 - 2x\left(\frac{1}{2}y\right) + \left(\frac{1}{2}y\right)^2 - \left(\frac{1}{2}y\right)^2\right] + 3y^2 + 6z^2 \\
&= 2\left(x - \frac{1}{2}y\right)^2 - \frac{1}{2}y^2 + 3y^2 + 6z^2 \\
&= 2\left(x - \frac{1}{2}y\right)^2 + \frac{5}{2}y^2 + 6z^2 \\
&\geq 0,
\end{aligned}
$$

para todo $x, y, z \in \mathbb{R}$. Además, si $u^t G_{\mathcal{C}} u = 0$ entonces $x - \frac{1}{2}y = 0$, $y = 0$ y $z = 0$, de donde $x = y = z = 0$. Luego, $u = 0$.

También es evidente que $G_{\mathcal{B}}$ es simétrica. Para ver que $G_{\mathcal{B}}$ es definida positiva se aplica el método de Gauss de descomposición en cuadrados calculando

$$
\begin{aligned}
u^t G_{\mathcal{B}} u &= \\
&= \begin{bmatrix} x & y & z \end{bmatrix} \begin{bmatrix} 9 & 3 & 1 \\ 3 & 3 & 1 \\ 1 & 1 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} \\
&= 9x^2 + 6xy + 2xz + 3y^2 + 2yz + 2z^2 \\
&= 9\left[x^2 + x\left(\frac{6y+2z}{9}\right)\right] + 3y^2 + 2yz + 2z^2 \\
&= 9\left[x^2 + 2x\left(\frac{1}{3}y + \frac{1}{9}z\right) + \left(\frac{1}{3}y + \frac{1}{9}z\right)^2 - \left(\frac{1}{3}y + \frac{1}{9}z\right)^2\right] + \\
&\quad +3y^2 + 2yz + 2z^2 \\
&= 9\left[x + \left(\frac{1}{3}y + \frac{1}{9}z\right)\right]^2 - 9\left(\frac{1}{9}y^2 + \frac{2}{27}yz + \frac{1}{81}z^2\right) + 3y^2 + 2yz + 2z^2 \\
&= 9\left(x + \frac{1}{3}y + \frac{1}{9}z\right)^2 + 2y^2 + \frac{4}{3}yz + \frac{17}{9}z^2 \\
&= 9\left(x + \frac{1}{3}y + \frac{1}{9}z\right)^2 + 2\left[y^2 + 2y\left(\frac{1}{3}z\right) + \left(\frac{1}{3}z\right)^2 - \left(\frac{1}{3}z\right)^2\right] + \frac{17}{9}z^2 \\
&= 9\left(x + \frac{1}{3}y + \frac{1}{9}z\right)^2 + 2\left(y + \frac{1}{3}z\right)^2 - \frac{2}{9}z^2 + \frac{17}{9}z^2 \\
&= 9\left(x + \frac{1}{3}y + \frac{1}{9}z\right)^2 + 2\left(y + \frac{1}{3}z\right)^2 + \frac{15}{9}z^2 \\
&\geq 0,
\end{aligned}
$$

para todo $x, y, z \in \mathbb{R}$. Además, si $u^t G_{\mathcal{B}} u = 0$ entonces $x + \frac{1}{3}y + \frac{1}{9}z = 0$, $y + \frac{1}{3}z = 0$ y $z = 0$, de donde $x = y = z = 0$. Luego, $u = 0$.

Es posible evitar realizar estos últimos cálculos pues utilizando que $G_{\mathcal{C}}$ es definida positiva y que $G_{\mathcal{B}} = P^t G_{\mathcal{C}} P$, con $P = [\mathcal{B}]_{\mathcal{C}}$ invertible, se tiene que

$$x^t G_{\mathcal{B}} x = x^t P^t G_{\mathcal{C}} P x = (Px)^t G_{\mathcal{C}} (Px) > 0,$$

para todo vector $y := Px \neq 0$. Al ser P invertible, la desigualdad se cumple para todo vector $x \neq 0$. $\triangleright$

El siguiente ejemplo es una interesante aplicación.

Ejemplo 10.34. *Calcular la siguiente integral*

$$\int_{-\pi}^{\pi} (3 - 5\cos(x) + 7\cos(2x) - \cos(6x))^2 \, \mathrm{d}x$$

utilizando teoría de espacios euclídeos.

$\lhd$ Al tratarse del cálculo de una integral, parece apropiado utilizar el espacio de las funciones continuas del Ejemplo 10.16 correspondiente al sistema trigonométrico en el intervalo $[-\pi, \pi]$.

Puesto que el integrando es el cuadrado de una función, se debe calcular el producto escalar de una función por sí misma, es decir, una expresión del tipo

$$\langle f, f \rangle$$

donde

$$f(x) = 3 - 5\cos(x) + 7\cos(2x) - \cos(6x).$$

Se observa también que la función f es una combinación lineal de las funciones 1, $\cos(x)$, $\cos(2x)$ y $\cos(6x)$. Esto sugiere considerar la base

$$\mathcal{B} = \{f_1(x) = 1, f_2(x) = \cos(x), f_3(x) = \cos(2x), f_4(x) = \cos(6x)\}$$

de subespacio que ellas generan. Recordando que el sistema trigonométrico es un sistema ortogonal en el intervalo $[-\pi, \pi]$ con respecto al producto escalar

$$\langle g, h \rangle = \int_{-\pi}^{\pi} g(x)h(x) \, \mathrm{d}x, \qquad g, h \in \mathcal{C}[-\pi, \pi],$$

se observa inmediatamente que, en particular, $\mathcal{B}$ es también un sistema ortogonal.

Este hecho garantiza que todos los elementos de la matriz de Gram que se encuentran fuera de la diagonal son iguales a 0.

Faltan calcular los elementos de la diagonal de $G_{\mathcal{C}}$, es decir los elementos de la forma $\langle h, h\rangle$ para todos los vectores h de la base $\mathcal{B}$. En efecto,

$$\langle 1, 1\rangle = \int_{-\pi}^{\pi} \mathrm{d}x = 2\pi,$$

y el resto de funciones son del tipo $h_m(x) = \cos(mx)$, con lo cual se pueden calcular todas las normas restantes de forma simultánea:

$$\begin{aligned}
\langle \cos(mx), \cos(mx)\rangle &= \int_{-\pi}^{\pi} \cos^2(mx)\,\mathrm{d}x \\
&= \int_{-\pi}^{\pi} \frac{1}{2}[1+\cos(2mx)]\,\mathrm{d}x \\
&= \frac{1}{2}\int_{-\pi}^{\pi} \mathrm{d}x + \frac{1}{2}\int_{-\pi}^{\pi} \cos(2mx)\,\mathrm{d}x \\
&= \pi + \frac{1}{2}\langle 1, \cos(2mx)\rangle \\
&= \pi,
\end{aligned}$$

donde se ha utilizado que la función 1 es ortogonal a $\cos(2mx)$.

De este modo,

$$G_{\mathcal{C}} = \begin{bmatrix} 2\pi & 0 & 0 & 0 \\ 0 & \pi & 0 & 0 \\ 0 & 0 & \pi & 0 \\ 0 & 0 & 0 & \pi \end{bmatrix}.$$

Por último, las coordenadas del vector f en la base $\mathcal{B}$ son

$$[f]_{\mathcal{B}} = \begin{bmatrix} 3 \\ -5 \\ 7 \\ -1 \end{bmatrix}.$$

Luego,

$$\begin{aligned}
&\int_{-\pi}^{\pi} (3 - 5\cos(x) + 7\cos(2x) - \cos(6x))^2 \,\mathrm{d}x = \\
&= \langle f, f \rangle \\
&= [f]_{\mathcal{B}}^t G_{\mathcal{B}} [f]_{\mathcal{B}} \\
&= \begin{bmatrix} 3 & -5 & 7 & -1 \end{bmatrix} \begin{bmatrix} 2\pi & 0 & 0 & 0 \\ 0 & \pi & 0 & 0 \\ 0 & 0 & \pi & 0 \\ 0 & 0 & 0 & \pi \end{bmatrix} \begin{bmatrix} 3 \\ -5 \\ 7 \\ -1 \end{bmatrix} \\
&= 93\pi.
\end{aligned}$$

▷

Se observa en el ejemplo anterior que la matriz de Gram toma una forma especialmente sencilla cuando se la calcula en una base ortogonal. Ahora se analiza la forma de la matriz de Gram en una base ortogonal y en una ortonormal para el caso general.

Proposición 10.15 (Matriz de Gram en una base ortonormal)**.** Sea $(E, \langle \cdot, \cdot \rangle)$ un espacio euclídeo real de dimensión finita y sea $\mathcal{B} = \{u_1, u_2, \ldots, u_n\}$ una base ordenada de E. Entonces

(a) $\mathcal{B}$ es una base ortogonal $\Leftrightarrow$ $G_{\mathcal{B}}$ es una matriz diagonal con los elementos de la diagonal principal estrictamente positivos.

(b) $\mathcal{B}$ es una base ortonormal $\Leftrightarrow$ $G_{\mathcal{B}} = I_n$.

Demostración. El resultado se sigue de forma evidente del hecho de que los productos escalares de elementos de la base diferentes entre sí son cero (de ahí el carácter diagonal) y son estrictamente positivos si se multiplican escalarmente por sí mismos pues coinciden con el cuadrado de su norma. En particular, estos últimos valen la unidad cuando la base es ortonormal. □

El apartado (b) puede escribirse en símbolos como sigue:

$$\mathcal{B} \text{ es una base ortonormal } \Leftrightarrow \langle u_i, u_j \rangle = \delta_{ij}, 1 \leq i, j \leq n \Leftrightarrow G_{\mathcal{B}} = I_n.$$

A continuación, se analiza una propiedad de las matrices de cambio de base cuando ambas bases son ortonormales.

Proposición 10.16 (Matriz de cambio de base para bases ortonormales)**.** Sea $(E, \langle \cdot, \cdot \rangle)$ un espacio euclídeo real de dimensión finita y sean $\mathcal{B}$ y $\mathcal{B}_1$ dos bases ortonormales de E. Entonces la matriz de cambio de base $[\mathcal{B}]_{\mathcal{B}_1}$ satisface que[a]

$$([\mathcal{B}]_{\mathcal{B}_1})^{-1} = ([\mathcal{B}]_{\mathcal{B}_1})^t.$$

Es decir, las matrices de cambio de base que permiten pasar de una base ortonormal a otra base ortonormal es una matriz ortogonal.

[a]Las matrices cuadradas reales (invertibles) tales que su inversa coincide con su traspuesta se llaman **matrices ortogonales**.

Demostración. Puesto que la matriz de Gram en una base ortonormal es la identidad se tiene que $G_{\mathcal{B}} = I_n = G_{\mathcal{B}_1}$. Además, dos matrices de Gram del mismo producto escalar correspondientes a dos bases diferentes son congruentes. Más precisamente, de la Proposición 10.14 se tiene que

$$I_n = G_{\mathcal{B}} = ([\mathcal{B}]_{\mathcal{B}_1})^t G_{\mathcal{B}_1} [\mathcal{B}]_{\mathcal{B}_1} = ([\mathcal{B}]_{\mathcal{B}_1})^t [\mathcal{B}]_{\mathcal{B}_1}.$$

De este modo se ha llegado a que $[\mathcal{B}]_{\mathcal{B}_1}$ es invertible y $([\mathcal{B}]_{\mathcal{B}_1})^{-1} = ([\mathcal{B}]_{\mathcal{B}_1})^t$. □

Ejemplo 10.35. *Encontrar la matriz de cambio de base de la base canónica $\mathcal{C}$ de $\mathbb{R}^2$ a la base*

$$\mathcal{B} = \left\{ u_1 = \left(\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2} \right), u_2 = \left(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2} \right) \right\}.$$

$\triangleleft$ Es claro que

$$[\mathcal{B}]_{\mathcal{C}} = \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix}.$$

Como $\mathcal{B}$ y $\mathcal{C}$ son ambas bases ortonormales, se tiene que

$$[\mathcal{C}]_{\mathcal{B}} = ([\mathcal{B}]_{\mathcal{C}})^{-1} = ([\mathcal{B}]_{\mathcal{C}})^t = \begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix}.$$

$\triangleright$

Esta capítulo finaliza con un resultado que expresa de manera sencilla la forma de todos los productos escalares sobre el espacio vectorial $\mathbb{R}^n$.

Teorema 10.10 (Caracterización de una matriz simétrica y definida positiva)**.** Sea $A \in \mathbb{R}^{n\times n}$. Entonces la matriz A es simétrica y definida positiva si y sólo si existe una matriz invertible $P \in \mathbb{R}^{n\times n}$ tal que

$$A = P^t P.$$

Demostración. Si $A \in \mathbb{R}^{n\times n}$ es simétrica y definida positiva, el Teorema 10.9 asegura que la expresión

$$\langle x, y\rangle = x^t A y$$

define un producto escalar sobre $\mathbb{R}^n$. Es fácil ver que su matriz de Gram cumple que $G_{\mathcal{C}} = A$, siendo $\mathcal{C}$ la base canónica de $\mathbb{R}^n$.

Sea $\mathcal{B}$ una base ortonormal de $\mathbb{R}^n$. Por la Proposición 10.15, $G_{\mathcal{B}} = I_n$. Luego, la Proposición 10.14 garantiza que

$$A = G_{\mathcal{C}} = ([\mathcal{C}]_{\mathcal{B}})^t G_{\mathcal{B}} [\mathcal{C}]_{\mathcal{B}} = ([\mathcal{C}]_{\mathcal{B}})^t [\mathcal{C}]_{\mathcal{B}}.$$

Denotando por $P := [\mathcal{C}]_{\mathcal{B}}$ se tiene que $A = P^t P$, siendo $P \in \mathbb{R}^{n\times n}$ una matriz invertible por tratarse de una matriz de cambio de base.

Recíprocamente, si existe una matriz invertible $P \in \mathbb{R}^{n\times n}$ tal que $A = P^t P$, es conocido que $A^t = A$. Además, se tiene que $x^t A x \geq 0$, $\forall x \in \mathbb{R}^n$ puesto que

$$x^t A x = (x^t P^t)(Px) = (Px)^t Px = y^t y = \sum_{i=1}^{n} y_i^2 \geq 0,$$

donde se ha utilizado que $y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$. Finalmente, se prueba que $x^t A x \neq 0$ si $x \neq 0$, con lo cual A es definida positiva. En efecto, si fuese $x^t A x = 0$ entonces $x^t P^t P x = 0$ de donde de nuevo se puede escribir $y^t y = \sum_{i=1}^n y_i^2 = 0$ siendo $y = Px$. De esta última igualdad se deduce que $y_i = 0$ para todo $i = 1, \ldots, n$. Es decir, $Px = y = 0$ y al ser P invertible se obtiene que $x = 0$. $\square$

Aplicando el Teorema 10.9 y, posteriormente, el Teorema 10.10 se deduce inmediatamente el resultado final.

Teorema 10.11 (Expresión de un producto escalar arbitrario en $\mathbb{R}^n$)**.** Sea $(E, \langle \cdot, \cdot \rangle)$ un espacio euclídeo real de dimensión finita y sea $\mathcal{B}$ una base ordenada de E. Entonces existe una matriz invertible $P \in \mathbb{R}^{n \times n}$ tal que

$$\langle x, y \rangle = ([x]_{\mathcal{B}})^t P^t P [y]_{\mathcal{B}} \quad \text{para todo } x, y \in E.$$

Por este resultado, algunos autores llaman matriz de Gram a toda matriz del tipo $P^t P$, aunque P no sea invertible[15].

[15] Este hecho se justifica al estudiar formas bilineales y sesquilineales.

10.8. EJERCICIOS

(1) Comprobar que en el cuerpo $(\mathbb{Z}_3, +, \cdot)$, la ecuación $x^2 = 2$ no tiene solución[16].

(2) Sea $(E, \langle \cdot, \cdot \rangle)$ un espacio euclídeo. Demostrar que las siguientes condiciones son equivalentes:

(a) Condición de **positividad**: $\langle u, u \rangle > 0$ para todo $u \in E - \{0\}$.

(b) Condición (de **no negatividad**): $\langle u, u \rangle \geq 0$ para todo $u \in E$ y la condición (de **definición positiva**): $\langle u, u \rangle = 0$ si y sólo si $u = 0$.

(3) Probar que si E es un espacio euclídeo real con producto escalar $\langle \cdot, \cdot \rangle$ entonces las siguientes condiciones son equivalentes:

(a) $\langle \alpha u + v, w \rangle = \alpha \langle u, w \rangle + \langle v, w \rangle$ para todo $u, v, w \in E$ y para todo $\alpha \in \mathbb{R}$,

(b) $\langle u + v, w \rangle = \langle u, w \rangle + \langle v, w \rangle$ y $\langle \alpha u, v \rangle = \alpha \langle u, v \rangle$ para todo $u, v, w \in E$ y $\alpha \in \mathbb{R}$.

Enunciar un resultado similar para la linealidad en el segundo argumento.

(4) Probar que en un espacio euclídeo, el vector nulo es el único vector que tiene norma igual a 0.

(5) Sea $(E, \langle \cdot, \cdot \rangle)$ un espacio euclídeo real, $n, m \in \mathbb{N}$, $a_1, \ldots, a_n \in \mathbb{R}$, $b_1, \ldots, b_m \in \mathbb{R}$ y $u, u_1, \ldots, u_n \in E$, $v, v_1, \ldots, v_m \in E$. Demostrar que se cumplen las siguientes propiedades:

(a) $\langle 0, u \rangle = 0$ para todo $u \in E$ utilizando que el vector nulo es producto del escalar cero por el vector u. Comparar esta demostración con la dada en la Proposición 10.1.

[16] Este ejercicio justifica que, para definir un producto escalar, es necesario considerar $\mathbb{K}$-espacios vectoriales donde los elementos del cuerpo $\mathbb{K}$ admitan raíz cuadrada.

(b) $\langle \sum_{i=1}^{n} a_i u_i, \sum_{j=1}^{m} b_j v_j \rangle = \sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j \langle u_i, v_j \rangle$.

(c) Si $\langle u, w \rangle = 0$, para todo $w \in E$ entonces $u = 0$.

(d) Si $\langle w, v \rangle = 0$, para todo $w \in E$ entonces $v = 0$.

(e) Si $\langle u, w \rangle = \langle v, w \rangle$, para todo $w \in E$ entonces $u = v$.

(f) Si $\langle w, u \rangle = \langle w, v \rangle$, para todo $w \in E$ entonces $u = v$.

(Sugerencia: en el apartado (a) utilizar el método de inducción doble, sobre n y sobre m.)

(6) Sea E un espacio euclídeo y $\{v_1, v_2, \ldots, v_s\} \subseteq E$. Probar que si para algún $i \in \{1, 2, \ldots, s\}$, v_i es combinación lineal de los restantes y ortogonal a todos ellos entonces $v_i = 0$.

(7) Comprobar la validez de los axiomas (PE1)-(PE3) en el Ejemplo 10.1 utilizando las componentes de los vectores y comparar con la resolución utilizando el producto matricial.

(8) Indicar, justificando la respuesta, si en el $\mathbb{R}$-espacio vectorial $\mathbb{R}^{2\times 2}$ la aplicación definida por

$$\langle A, B \rangle = a_1 b_1 + a_2 b_3 + a_3 b_2 + a_4 b_4,$$

siendo

$$A = \begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix} \qquad \text{y} \qquad B = \begin{bmatrix} b_1 & b_2 \\ b_3 & b_4 \end{bmatrix},$$

define un producto escalar.

(9) Se considera la matriz $A = \begin{bmatrix} 1 & -2 \\ -2 & 6 \end{bmatrix}$. Probar que la aplicación

$$\langle x, y \rangle = x^t A y$$

define un producto escalar en el $\mathbb{R}$-espacio vectorial $\mathbb{R}^2$. ¿Ocurre lo mismo si se considera la matriz $A = \begin{bmatrix} 1 & 3 \\ 3 & 6 \end{bmatrix}$? ¿Y con la matriz $A = \begin{bmatrix} 1 & 1 \\ 2 & 6 \end{bmatrix}$? Justificar.

(10) Comprobar que la aplicación definida en el Ejemplo 10.6 es, efectivamente, un producto escalar sobre el $\mathbb{R}$-espacio vectorial $\mathbb{R}^3$.

(11) (a) Demostrar que

$$\langle x, y\rangle = \left[\begin{array}{ccc} x_1 & x_2 & x_3 \end{array}\right] \left[\begin{array}{rrr} 2 & 1 & -1 \\ 1 & 1 & 0 \\ -1 & 0 & 2 \end{array}\right] \left[\begin{array}{c} y_1 \\ y_2 \\ y_3 \end{array}\right]$$

define un producto escalar en $\mathbb{R}^3$ siendo $x = (x_1, x_2, x_3)^t$ e $y = (y_1, y_2, y_3)^t$ vectores de $\mathbb{R}^3$.

(b) Si se cambia la matriz del apartado anterior por

$$\left[\begin{array}{rrr} 1 & 0 & 0 \\ 2 & -3 & 0 \\ 0 & 0 & 1 \end{array}\right],$$

¿sigue siendo un producto escalar la aplicación considerada?

(c) ¿Qué producto escalar se obtiene en el primer apartado si se cambia la matriz dada por la matriz identidad de tamaño 3×3?

(12) En el espacio euclídeo $\mathbb{R}_2[x]$ con el producto escalar

$$\langle p, q\rangle = \int_0^1 p(x)q(x)\mathrm{d}x \qquad \text{para } p, q \in \mathbb{R}_2[x]$$

calcular la norma del vector $r(x) = x - 1$.

(13) Sean x e y dos vectores de un espacio vectorial real con producto interior y sea $y \neq 0$. Hallar el valor de $a \in \mathbb{R}$ para que $x - ay$ sea ortogonal a y.

(14) Los vectores x e y de un espacio euclídeo real forman un ángulo de $\pi/3$ radianes y la norma de x es 4. Determimar la norma de y para que $y - x$ sea ortogonal a x.

(15) Sea $(E, \langle\cdot,\cdot\rangle)$ un espacio euclídeo real de dimensión finita $n \geq 1$ y sea $\mathcal{B} = \{u_1, u_2, \ldots, u_n\}$ una base ortonormal de E. Para dos vectores $x, y \in E$ tales que

$$[x]_{\mathcal{B}} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \qquad \text{e} \qquad [y]_{\mathcal{B}} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix},$$

probar que:

(a) su producto escalar es

$$\langle x, y\rangle = x_1y_1 + x_2y_2 + \cdots + x_ny_n = ([x]_{\mathcal{B}})^t[y]_{\mathcal{B}}.$$

(b) la norma de x viene dada por

$$\|x\| = +\sqrt{([x]_{\mathcal{B}})^t[x]_{\mathcal{B}}}.$$

Este ejercicio asegura que el producto escalar y la norma en un espacio euclídeo E cualquiera de dimensión finita n se calculan como el producto escalar y la norma de $\mathbb{R}^n$ con el producto escalar canónico siempre que se utilice una base ortonormal de E para dicho cálculo.

(16) Construir una base ortonormal de $\mathbb{R}^3$ con el producto escalar canónico a partir de la base

$$\mathcal{B} = \{u_1 = (1,1,0)^t,\ u_2 = (1,2,0)^t,\ u_3 = (0,1,2)^t\}.$$

(17) En el $\mathbb{R}$-espacio vectorial $\mathbb{R}_2[x]$ se considera la aplicación definida por

$$\langle p, q\rangle = p(0)q(0) + p\left(\frac{1}{2}\right)q\left(\frac{1}{2}\right) + p(1)q(1), \qquad p, q \in \mathbb{R}_2[x].$$

(a) Probar que $\langle\cdot,\cdot\rangle$ define un producto escalar en $\mathbb{R}_2[x]$ y, por tanto, $(\mathbb{R}_2[x], \langle\cdot,\cdot\rangle)$ es un espacio euclídeo.

(b) Hallar el complemento ortogonal de $U = \overline{\{x\}}$ respecto del producto interior indicado y una base de $U^{\perp}$.

(c) Verificar que se cumple la propiedad que relaciona las dimensiones de los subespacios vectoriales U y $U^{\perp}$.

(d) Calcular la proyección ortogonal de $p(x) = 1 + 3x - x^2$ sobre U.

(18) Sea $\langle .,. \rangle$ la aplicación real

$$\langle \cdot, \cdot \rangle : \mathbb{R}^3 \times \mathbb{R}^3 \longrightarrow \mathbb{R},$$

definida como sigue

$$(u, v) \longmapsto \langle u, v \rangle := u_1 v_1 + 2u_2 v_2 + 3u_3 v_3$$

siendo $u = (u_1, u_2, u_3)^t$ y $v = (v_1, v_2, v_3)^t$ vectores de $\mathbb{R}^3$.

(a) Probar que $\langle \cdot, \cdot \rangle$ define un producto escalar sobre $\mathbb{R}^3$.

(b) ¿Cuál es la norma inducida por el producto escalar considerado?

(c) ¿Cuáles son todos los vectores de $\mathbb{R}^3$ de norma 1? Realizar una interpretación geométrica.

(d) Calcular el ángulo y la distancia entre los vectores $u = (1, 1, -1)^t$ y $v = (1, 1, 0)^t$.

(e) Hallar $U^{\perp}$, una base y su dimensión siendo $U = \overline{\{u_0\}}$ para $u_0 = (1, 1, -1)^t$.

(f) Aplicando el método de Gram-Schmidt construir una base ortonormal para $U^{\perp}$ a partir de la base hallada en el apartado anterior. Comprobar que los vectores hallados son ortogonales a u_0 con respecto a este producto escalar.

(g) Expresar $w = (7, 1, -1)^t$ como suma de un vector de U y otro de $U^{\perp}$. ¿Es única dicha representación? Justificar.

(h) Hallar la proyección ortogonal de $w = (7, 1, -1)^t$ sobre U y su componente ortogonal.

(i) Comprobar el teorema de Pitágoras para los vectores hallados en el apartado anterior.

(19) Obtener, si es posible, un conjunto ortogonal a partir de las columnas de la matriz

$$A = \begin{bmatrix} 1 & -1 & 4 \\ 1 & 4 & -2 \\ 1 & 4 & 2 \\ 1 & -1 & 0 \end{bmatrix}$$

utilizando el producto escalar canónico de $\mathbb{R}^4$.

(20) En $\mathbb{R}^3$ con el producto escalar canónico considerar los vectores $u_1 = (4, -3, 0)^t$ y $u_2 = (1, 2, 0)^t$.

(a) Hallar un vector $u_3 \in \mathbb{R}^3$ ortogonal a u_1 y u_2 de modo que $||u_3|| = 4$ y tal que su tercera componente sea positiva. ¿Cuántos vectores hay en estas condiciones? Intentarlo primero de manera intuitiva y luego analíticamente.

(b) A partir de la base $\{u_1, u_2, u_3\}$, hallar una base ortonormal para $\mathbb{R}^3$. ¿Podrían haberse determinado a priori y de manera intuitiva las direcciones de los vectores de la base ortonormal pedida?

(c) Expresar $v = (5, 10, 15)^t$ como combinación lineal de los vectores de la base ortonormal hallada en el apartado anterior (¡sin resolver sistemas de ecuaciones lineales!).

(21) Las *funciones de Walsh* son funciones ortogonales[17] convenientes para usar en *sistemas digitales*. Las tres primeras funciones de Walsh se definen a continuación:

$$f_1(x) = \alpha \text{ si } x \in [0, 1]; \qquad f_2(x) = \begin{cases} -\beta, & 0 \leq x \leq \frac{1}{2} \\ \beta, & \frac{1}{2} < x \leq 1 \end{cases}$$

[17] Respecto del mismo producto escalar que se ha definido sobre las funciones continuas. En realidad, para que sea producto escalar debería definirse sobre un conjunto que escapa el nivel de este libro y aquí se prescindirá de estas cuestiones técnicas.

y

$$f_3(x) = \begin{cases} \gamma, & 0 \leq x \leq \frac{1}{4} \\ -\gamma, & \frac{1}{4} < x \leq \frac{3}{4} \\ \gamma, & \frac{3}{4} < x \leq 1 \end{cases}$$

siendo α, β y γ números reales positivos.

(a) Comprobar que las funciones f_1, f_2 y f_3 son ortogonales sobre el intervalo $[0, 1]$.

(b) Determinar los valores de α, β y γ para que las funciones f_1, f_2 y f_3 sean ortonormales sobre el intervalo $[0, 1]$.

(c) Calcular la proyección ortogonal de la función $f(x) = x$, $x \in [0, 1]$, sobre el subespacio generado por las funciones f_1, f_2 y f_3.

(22) A partir de la base canónica encontrar una base ortonormal de $\mathbb{R}_2[x]$ si el producto escalar está definido por

$$\langle p,\, q \rangle = p(x_1)q(x_1) + p(x_2)q(x_2) + p(x_3)q(x_3) \qquad \text{para } p,\, q \in \mathbb{R}_2[x]$$

donde $x_1 = -1$, $x_2 = 0$ y $x_3 = 1$. ¿A qué es igual la matriz de Gram de este producto escalar respecto de la base encontrada? ¿Y respecto de la base $\mathcal{B} = \{x + 1, x - 1, x^2\}$? ¿Son $x + 1$ y $x - 1$ ortogonales? Justificar.

(23) Se considera la matriz $A \in \mathbb{R}^{3\times 3}$ siguiente:

$$A = \begin{bmatrix} 8 & -16 & 24 \\ 4 & -8 & 12 \\ 1 & -2 & 3 \end{bmatrix}.$$

¿Se puede aplicar el método de Gram-Schmidt para hallar una base ortonormal a partir de las columnas de la matriz A? ¿Por qué? Intentarlo e indicar a qué resultado se llega.

(24) Este problema de Cálculo Diferencial será necesario para el resolver el próximo ejercicio. Probar que si $a, b, c \in \mathbb{R}$ con $a > 0$ satisfacen que

$$ax^2 + bx + c \geq 0, \qquad \text{para todo } x \in \mathbb{R}$$

entonces $b^2 - 4ac \leq 0$. (Ayuda: Encontrar el valor mínimo que alcanza el polinomio en la variable x.)

(25) Proporcionar tres demostraciones alternativas de la **desigualdad de Cauchy-Schwarz**:

(a) Para la primera, sustituir el valor de λ por aquel para el cual la expresión (10.3) alcance el mínimo absoluto y razonar por qué dicho valor es el adecuado.

(b) Al definir la proyección ortogonal de un vector v sobre un vector no nulo u, ambos de un espacio euclídeo, se encuentra que $\mathrm{proy}_u v = \lambda u$ siendo $\lambda = \frac{\langle v,u\rangle}{\langle u,u\rangle}$. Razonando geométricamente en una gráfica de $\mathbb{R}^2$ con el producto escalar canónico, se observa que $0 \leq \|\alpha u - v\| = d(\alpha u, v)$ obtiene su valor mínimo cuando se toma $\alpha = \lambda$. Ahora, desarrollar el cálculo de $0 \leq \|\lambda u - v\|^2$ para llegar a la conclusión en un espacio euclídeo arbitrario.

(c) Para la tercera, supóngase el caso (no trivial) en que ambos vectores sean no nulos. Ahora, la desigualdad es equivalente a la expresión $\left|\left\langle \frac{1}{\|u\|}u, \frac{1}{\|v\|}v\right\rangle\right| \leq 1$. Para demostrarla, considerar la suma y la resta de los vectores $\frac{1}{\|u\|}u$ y $\frac{1}{\|v\|}v$ y realizar el producto escalar de cada uno consigo mismo.

(26) Probar que, en la desigualdad de Cauchy-Swcharz, la igualdad se obtiene si y sólo si los vectores son linealmente dependientes.

(27) Probar que, en la desigualdad triangular, la igualdad se obtiene si y sólo si los vectores son linealmente dependientes y un vector se obtiene como un múltiplo no negativo del otro.

(28) Demostrar que en un espacio euclídeo E se cumple:

(a) $\|x \pm y\|^2 = \|x\|^2 \pm 2\langle x, y\rangle + \|y\|^2$, para todo $x, y \in E$.

(b) $\|-x\| = \|x\|$, para todo $x \in E$.

(c) $|\|x\| - \|y\|| \leq \|x - y\|$, para todo $x, y \in E$.

(29) Demostrar la ley del paralelogramo de la Proposición 10.2.

(30) Demostrar la identidad de polarización de la Proposición 10.2.

(31) Realizar una interpretación geométrica de la ley del paralelogramo y de la desigualdad triangular en el espacio euclídeo $\mathbb{R}^2$ con el producto escalar canónico.

(32) Sea E un espacio euclídeo real y $\alpha, \beta \in \mathbb{R}$ dos escalares no nulos con el mismo signo. Probar que $\text{áng}(u, v) = \text{áng}(\alpha u, \beta v)$, para todo $u, v \in E$.

(33) **Teorema del coseno**: Sea E un espacio euclídeo real y sean $u, v \in E$ dos vectores no nulos. Probar que

$$\|u - v\|^2 = \|u\|^2 + \|v\|^2 - 2\|u\|\|v\| \cos(\text{áng}(u, v)).$$

Realizar una interpretación geométrica en $\mathbb{R}^2$ con el producto escalar canónico.

(34) Estudiar en qué caso se da la igualdad en el Teorema de la mejor aproximación. Realizar una interpretación geométrica en $\mathbb{R}^3$ con el producto escalar canónico siendo el subespacio S un plano que pasa por el origen. Considerar los casos $u \in S$ y $u \notin S$.

(35) Comprobar que de la desigualdad de Cauchy-Schwarz particularizada al espacio:

(a) $\mathbb{R}^n$ con el producto escalar canónico se obtiene

$$\begin{aligned} &(u_1 v_1 + u_2 v_2 + \cdots + u_n v_n)^2 \leq \\ &\leq \quad (u_1^2 + u_2^2 + \cdots + u_n^2)(v_1^2 + v_2^2 + \cdots + v_n^2). \end{aligned}$$

(b) $\mathcal{C}[a, b]$ con el producto escalar del Ejemplo 10.3 se obtiene

$$\left(\int_a^b f(x) g(x) \, \mathrm{d}x \right)^2 \leq \left(\int_a^b [f(x)]^2 \, \mathrm{d}x \right) \left(\int_a^b [g(x)]^2 \, \mathrm{d}x \right).$$

(c) $\mathbb{R}^{n\times n}$ con el producto escalar de Frobenius del Ejemplo 10.4 se obtiene

$$\left(\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}b_{ij}\right)^2 \leq \left(\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}^2\right)\left(\sum_{i=1}^{n}\sum_{j=1}^{n} b_{ij}^2\right).$$

(d) $\mathbb{R}^3$ con el producto escalar del Ejemplo 10.6 se obtiene

$$\begin{aligned}(6u_1v_1 - u_2v_1 + u_3v_1 - u_1v_2 + 2u_2v_2 + u_1v_3 + u_3v_3)^2 \leq \\ \leq \quad \left(6u_1^2 + 2u_2^2 + u_3^2 - 2u_1u_2 + 2u_1u_3\right)\cdot \\ \left(6v_1^2 + 2v_2^2 + v_3^2 - 2v_1v_2 + 2v_1v_3\right).\end{aligned}$$

(36) Comprobar que las funciones $f(x) = \cos(x)$ y $g(x) = \mathrm{sen}(x)$ satisfacen el teorema de Pitágoras con el producto escalar de la integral en el espacio de las funciones continuas definidas en $[-\pi, \pi]$.

(37) **Teorema de Pitágonas generalizado**: Sea $(E, \langle\cdot,\cdot,\rangle)$ un espacio euclídeo real. Si $\{u_1, u_2, \ldots, u_s\} \subseteq E$ es un conjunto ortogonal con $s \in \mathbb{N}$, $s \geq 2$, entonces

$$\|u_1 + u_2 + \cdots + u_s\|^2 = \|u_1\|^2 + \|u_2\|^2 + \cdots + \|u_s\|^2.$$

(38) Probar que en todo espacio euclídeo real es válida la implicación recíproca del Teorema de Pitágoras.

(39) **Ley del rombo**: Sea $(E, \langle\cdot,\cdot,\rangle)$ un espacio euclídeo real y sean $x, y \in E$. Entonces

$$\|x\| = \|y\| \qquad \Longleftrightarrow \qquad \langle x+y, x-y\rangle = 0.$$

Realizar una interpretación geométrica en el caso del espacio euclídeo $\mathbb{R}^2$ con el producto escalar canónico y $\{x, y\}$ un conjunto linealmente independiente.

(40) La norma del Ejemplo 10.10 de la página 545 se conoce como **norma 1**. Hallar la expresión de la **distancia inducida por la norma 1** entre dos puntos $(x_1, x_2), (y_1, y_2) \in \mathbb{R}^2$. Esta distancia recibe el nombre de

métrica del taxista y fue desarrollada por el matemático alemán Herman Minkowski (1864-1909). Partiendo ahora de la expresión encontrada, comprobar que esta función satisface los axiomas de distancia. Interpretar geométricamente el significado de esta distancia en $\mathbb{R}^2$ calculándola entre los puntos $(0,0)$ y $(4,4)$; esto permitirá ver por qué también se la conoce como **distancia Manhattan**. Para ello, pintar los puntos dados y dibujar, desde el origen, 4 unidades sobre el eje horizontal y desde el punto obtenido, 4 unidades sobre el eje vertical hasta llegar al punto $(4,4)$. Por otro lado, pintar sucesivamente escalones horizontales y verticales de una unidad cada uno para llegar desde $(0,0)$ hasta $(4,4)$ y contar el número total de escalones.

(41) Hallar el complemento ortogonal del subespacio

$$S = \overline{\{(1,0,0),(0,1,0)\}}$$

del espacio euclídeo $\mathbb{R}^3$ con el producto escalar canónico. Luego, calcular $(S^{\perp})^{\perp}$ y comparar con S.

(42) En el espacio euclídeo $\mathbb{R}^{2\times 2}$ con el producto escalar de Frobenius, se considera el subespacio

$$S = \overline{\left\{ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix} \right\}}.$$

Se pide:

(a) Hallar la dimensión del $\mathbb{R}$-subespacio vectorial S.

(b) Calcular la dimensión de $S^{\perp}$.

(c) Encontrar una base ortonormal de S y una de $S^{\perp}$.

(d) Si $A = \begin{bmatrix} 1 & 2 \\ 3 & 0 \end{bmatrix}$ y $B = \begin{bmatrix} 3 & 2 \\ 0 & 1 \end{bmatrix}$, calcular $\mathrm{proy}_S(A)$ y $\mathrm{proy}_S(B)$.

(e) Para las matrices A y B del apartado anterior, calcular las distancias (inducida por la norma de Frobenius) de A y de B a sus respectivas proyecciones ortogonales sobre S y proporcionar un significado geométrico para dichos resultados.

(43) En el $\mathbb{R}$-espacio vectorial $\mathbb{R}^7$ se considera el producto escalar canónico, el vector $u = (1, 1, 1, 1, 1, 1, 1)$ y el subespacio

$$S = \left\{ (x_1, x_2, x_3, x_4, x_5, x_6, x_7) \in \mathbb{R}^7 : \sum_{i=1}^{7} i x_i = 0 \right\}.$$

Calcular la mejor aproximación de u en el subespacio S hallando previamente $\text{proy}_{S^\perp}(u)$. Observar que se ahorran numerosos cálculos si se procede de esta forma indirecta. Generalizar este ejercicio a $\mathbb{R}^n$ para $n \in \mathbb{N}$ arbitrario (Ayuda: Recordar las fórmulas de la suma de los n primeros números naturales y la de sus cuadrados.)

(44) **Desigualdad de Bessel**: Sea $(E, \langle \cdot, \cdot, \rangle)$ un espacio euclídeo real y sea $S = \{u_1, u_2, \ldots, u_s\} \subseteq E$ un conjunto ortogonal de vectores no nulos. Probar que para cualquier vector $u \in E$ se cumple que

$$\sum_{k=1}^{s} \frac{\langle u, u_k \rangle^2}{\|u_k\|^2} \leq \|u\|^2.$$

Deducir que $\|\text{proy}_{\overline{S}}(u)\| \leq \|u\|$. Además, la igualdad se cumple si y sólo si $u \in \overline{S}$.

(Ayuda: Considerar $u = \text{proy}_{\overline{S}}(u) + \widetilde{u_2}$ con $\widetilde{u_2} \perp \overline{S}$. Aplicar el Teorema de Pitágoras y luego desarrollar $\|\text{proy}_{\overline{S}}(u)\|^2$ escribiendo la proyección ortogonal como una combinación lineal de los elementos de $\overline{S}$ mediante los coeficientes de Fourier).

(45) Sea $Q \in \mathbb{R}^{m \times n}$ una matriz cuyas columnas son ortonormales respecto del producto escalar canónico de $\mathbb{R}^m$.

(a) Demostrar que $Q^t Q = I_n$.

(b) ¿Podría suceder que $QQ^t \neq I_m$? ¿Y si $m = n$? Justificar la respuesta.

Capítulo 11

Subespacios asociados a una matriz

Índice

11.1. Introducción

Este capítulo tiene un objetivo doble. Por un lado, se complementará el tema de equivalencia de matrices mediante la teoría de espacios vectoriales, donde se probará que el subespacio que generan las filas de una matriz rectangular tiene la misma dimensión que el que generan sus columnas y ambos números coinciden con el rango de la matriz (*Teorema fundamental del rango*). Por otro lado, se analizarán los cuatros subespacios fundamentales de una matriz y se establecerán sus relaciones a partir de la teoría de espacios euclídeos (*Teorema fundamental del Álgebra Lineal*).

Más concretamene, la importancia de los subespacios fundamentales de una matriz $A \in \mathbb{R}^{m\times n}$ radica en el hecho que presentan relaciones de ortogonalidad tanto en $\mathbb{R}^n$ como en $\mathbb{R}^m$ (considerando ambos espacios vectoriales con el producto escalar canónico) y, además, cada uno de estos espacios euclídeos se pueden expresar como suma directa (ortogonal) de dos de los subespacios fundamentales.

En definitiva, los cuatro *subespacios* que se introducirán proporcionan toda la información importante sobre una matriz rectangular dada, por lo que se llaman *fundamentales*.

A lo largo de este capítulo se considerarán los vectores que forman las filas o las columnas de una matriz $A \in \mathbb{K}^{m\times n}$ y se estudiarán relaciones de dependencia o independencia lineal entre las filas o bien entre las columnas de A. Para este estudio únicamente es necesaria la estructura algebraica de espacio vectorial para su análisis.

Además, se abordarán relaciones de ortogonalidad entre los subespacios fundamentales. En este último caso, será necesario utilizar el correspondiente $\mathbb{R}$-espacio vectorial $\mathbb{R}^m$ o $\mathbb{R}^n$ provisto de su producto escalar canónico, puesto que el producto escalar se estudió para espacios euclídeos reales.

11.2. Los cuatro subespacios fundamentales

Definición 11.1. Sea $A \in \mathbb{K}^{m\times n}$ una matriz particionada según sus filas o sus columnas de la siguiente forma

$$A = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_m \end{bmatrix} = \begin{bmatrix} c_1 & c_2 & \dots & c_n \end{bmatrix}.$$

Se llama

- **espacio fila** de A al subespacio de $\mathbb{K}^n$, denotado $\mathcal{F}(A)$, generado por las filas de A. En símbolos,

$$\mathcal{F}(A) = \overline{\{f_1, f_2, \dots, f_m\}}.$$

- **espacio columna** de A al subespacio de $\mathbb{K}^m$, denotado $\mathcal{C}(A)$, generado por las columnas de A. En símbolos,

$$\mathcal{C}(A) = \overline{\{c_1, c_2, \dots, c_n\}}.$$

Dado que toda matriz $A \in \mathbb{K}^{m\times n}$ tiene asociada su matriz traspuesta $A^t \in \mathbb{K}^{n\times m}$ se define como

- **espacio fila** de A^t al espacio columna de A, es decir $\mathcal{F}(A^t) = \mathcal{C}(A)$, teniendo en cuenta que $\mathbb{K}^{1\times m} \cong \mathbb{K}^{m\times 1}$.

- **espacio columna** de A^t al espacio fila de A, es decir $\mathcal{C}(A^t) = \mathcal{F}(A)$, teniendo en cuenta que $\mathbb{K}^{n\times 1} \cong \mathbb{K}^{1\times n}$.

Para mayor precisión, sería necesario decir que $\mathcal{F}(A^t) \cong \mathcal{C}(A)$ y $\mathcal{C}(A^t) \cong \mathcal{F}(A)$, pero se acepta este abuso de lenguaje. Estos hechos se clarifican en el siguiente ejemplo.

Ejemplo 11.1. *Hallar el espacio fila y el espacio columna de la matriz*

$$A = \begin{bmatrix} 1 & 4 & 0 & -3 \\ 3 & -1 & 2 & 0 \\ 5 & 7 & 2 & -6 \end{bmatrix} \in \mathbb{R}^{3\times 4}$$

y de su traspuesta A^t.

◁ Aplicando la definición y eliminando los vectores que son combinación lineal de los restantes se tiene que

$$\begin{aligned} \mathcal{F}(A) &= \overline{\{(1,4,0,-3),(3,-1,2,0),(5,7,2,-6)\}} \\ &= \overline{\{(1,4,0,-3),(3,-1,2,0)\}} \end{aligned}$$

y

$$\mathcal{C}(A) = \overline{\left\{ \begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix}, \begin{bmatrix} 4 \\ -1 \\ 7 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \\ 2 \end{bmatrix}, \begin{bmatrix} -3 \\ 0 \\ -6 \end{bmatrix} \right\}} = \overline{\left\{ \begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix}, \begin{bmatrix} 4 \\ -1 \\ 7 \end{bmatrix} \right\}}.$$

Dado que

$$A^t = \begin{bmatrix} 1 & 3 & 5 \\ 4 & -1 & 7 \\ 0 & 2 & 2 \\ -3 & 0 & -6 \end{bmatrix} \in \mathbb{R}^{4\times 3},$$

se tiene que

$$\begin{aligned} \mathcal{F}(A^t) &= \overline{\{(1,3,5),(4,-1,7),(0,2,2),(-3,0,-6)\}} \\ &= \overline{\{(1,3,5),(4,-1,7)\}} \end{aligned}$$

y

$$\mathcal{C}(A^t) = \overline{\left\{ \begin{bmatrix} 1 \\ 4 \\ 0 \\ -3 \end{bmatrix}, \begin{bmatrix} 3 \\ -1 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 5 \\ 7 \\ 2 \\ -6 \end{bmatrix} \right\}} = \overline{\left\{ \begin{bmatrix} 1 \\ 4 \\ 0 \\ -3 \end{bmatrix}, \begin{bmatrix} 3 \\ -1 \\ 2 \\ 0 \end{bmatrix} \right\}}.$$

Notar que $\mathcal{F}(A)$ y $\mathcal{C}(A^t)$ están generados por los mismos vectores con diferente posición (unos por filas y otros por columnas). Lo mismo ocurre con $\mathcal{C}(A)$ y $\mathcal{F}(A^t)$. ▷

Los espacios fila y columna se pueden describir de forma compacta como se muestra a continuación.

Lema 11.1. Sea $A \in \mathbb{K}^{m\times n}$. Las siguientes propiedades son válidas:

(a) $\mathcal{F}(A) = \{x^t A : x \in \mathbb{K}^{m\times 1}\}$.

(b) $\mathcal{C}(A) = \{Ax : x \in \mathbb{K}^{n\times 1}\} =: \mathcal{R}(A)$ (el conjunto $\mathcal{R}(A)$ se llama **espacio imagen** de A).

Demostración. (a) Si $A \in \mathbb{K}^{m\times n}$ se particiona según sus filas como

$$A = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_m \end{bmatrix},$$

entonces, por el Ejercicio 6 de la página 148, para $y \in \mathbb{K}^{1\times n}$,

$$\begin{aligned} y \in \mathcal{F}(A) \quad &\Leftrightarrow \quad y = \sum_{i=1}^{m} \alpha_i f_i \text{ para ciertos } \alpha_i \in \mathbb{K} \quad \Leftrightarrow \\ &\Leftrightarrow \quad y = \begin{bmatrix} \alpha_1 & \alpha_2 & \dots & \alpha_m \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_m \end{bmatrix} : \alpha_i \in \mathbb{K} \\ &\Leftrightarrow \quad y \in \{x^t A : x \in \mathbb{K}^{m\times 1}\}. \end{aligned}$$

(b) Si $A \in \mathbb{K}^{m\times n}$ se particiona según sus columnas como

$$A = \begin{bmatrix} c_1 & c_2 & \dots & c_n \end{bmatrix},$$

entonces, por la Proposición 2.8, para $y \in \mathbb{K}^{m\times 1}$,

$$y \in \mathcal{C}(A) \quad \Leftrightarrow \quad y = \sum_{i=1}^{n} \beta_i c_i \text{ para ciertos } \beta_i \in \mathbb{K} \quad \Leftrightarrow$$

$$y = \begin{bmatrix} c_1 & c_2 & \dots & c_n \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix} : \beta_i \in \mathbb{K} \quad \Leftrightarrow \quad y \in \{Ax : x \in \mathbb{K}^{n\times 1}\}.$$

$\square$

Observación 11.1. Las propias definiciones de espacios fila y columna de una matriz $A \in \mathbb{K}^{m\times n}$ aseguran que dichos espacios están estrechamente vinculados con los *sistemas de generadores* de cada uno de ellos. Además, del Lema 11.1, se observa que las expresiones $x^t A$ y Ax no representan más que el conjunto de todas las combinaciones lineales de las filas o columnas de A, respectivamente.

¿Cómo se relacionan el espacio fila y columna de un producto de matrices y los subespacios de alguno de sus factores?

Lema 11.2. Sean $A \in \mathbb{K}^{m\times n}$ y $B \in \mathbb{K}^{n\times p}$. Las siguientes propiedades son válidas:

(a) $\mathcal{C}(AB) \subseteq \mathcal{C}(A)$.

(b) $\mathcal{F}(AB) \subseteq \mathcal{F}(B)$.

Demostración. (a) Si $B \in \mathbb{K}^{n\times p}$ se particiona según sus columnas como

$$B = \begin{bmatrix} c_1 & c_2 & \dots & c_p \end{bmatrix},$$

entonces, por la Proposición 2.9,

$$AB = A\begin{bmatrix} c_1 & c_2 & \dots & c_p \end{bmatrix} = \begin{bmatrix} Ac_1 & Ac_2 & \dots & Ac_p \end{bmatrix}.$$

Puesto que $Ac_1, Ac_2, \ldots, Ac_p \in \mathcal{R}(A)$, es claro que

$$\mathcal{C}(AB) = \overline{\{Ac_1, Ac_2, \ldots, Ac_p\}} \subseteq \mathcal{R}(A) = \mathcal{C}(A).$$

(b) Aplicando el apartado (a),

$$\mathcal{F}(AB) = \mathcal{C}((AB)^t) = \mathcal{C}(B^t A^t) \subseteq \mathcal{C}(B^t) = \mathcal{F}(B).$$

$\square$

Otros dos espacios que complementan a los anteriores, como se verá más adelante, son los que se introducen a continuación[1].

Definición 11.2. Sea $A \in \mathbb{K}^{m\times n}$. Se llama

- **espacio nulo** de A al subespacio de $\mathbb{K}^n$, denotado $\mathcal{N}(A)$, de todos los vectores $x \in \mathbb{K}^{n\times 1}$ que son solución del sistema lineal homogéneo $Ax = 0$. En símbolos,

$$\mathcal{N}(A) = \{x \in \mathbb{K}^{n\times 1} : Ax = 0\}.$$

- **espacio nulo izquierdo** de A al subespacio de $\mathbb{K}^m$, denotado $\mathcal{N}(A^t)$, de todos los vectores $x \in \mathbb{K}^{m\times 1}$ que son solución del sistema lineal homogéneo $A^t x = 0$. En símbolos,

$$\mathcal{N}(A^t) = \{x \in \mathbb{K}^{m\times 1} : A^t x = 0\}.$$

El nombre del espacio nulo izquierdo se puede justificar a partir de la propiedad: $A^t x = 0 \Leftrightarrow x^t A = 0$, expresado en términos de A y no de A^t, donde el vector x^t aparece a la izquierda de la matriz A.

[1] En realidad, el espacio nulo se introdujo en la Definición 9.5 de la página 494.

Ejemplo 11.2. *Hallar el espacio nulo y el espacio nulo izquierdo de la matriz*

$$A = \begin{bmatrix} 1 & 4 & 0 & -3 \\ 3 & -1 & 2 & 0 \\ 5 & 7 & 2 & -6 \end{bmatrix} \in \mathbb{R}^{3\times 4}.$$

◁ Realizando operaciones elementales por filas se obtiene

$$A = \begin{bmatrix} 1 & 4 & 0 & -3 \\ 3 & -1 & 2 & 0 \\ 5 & 7 & 2 & -6 \end{bmatrix} \sim_f \begin{bmatrix} 1 & 0 & \frac{8}{13} & -\frac{3}{13} \\ 0 & 1 & -\frac{2}{13} & -\frac{9}{13} \\ 0 & 0 & 0 & 0 \end{bmatrix} = R_A.$$

Ahora, la solución de $Ax = 0$ se obtiene resolviendo el sistema de ecuaciones equivalente $R_A x = 0$ a partir de

$$\mathcal{N}(A) = \left\{ x_3 \begin{bmatrix} -\frac{8}{13} \\ \frac{2}{13} \\ 1 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} \frac{3}{13} \\ \frac{9}{13} \\ 0 \\ 1 \end{bmatrix} : x_3, x_4 \in \mathbb{R} \right\}.$$

Del mismo modo, al resolver el sistema $A^t x = 0$ se obtiene

$$\mathcal{N}(A^t) = \left\{ x_3 \begin{bmatrix} -2 \\ -1 \\ 1 \end{bmatrix} : x_3 \in \mathbb{R} \right\}.$$

▷

Observación 11.2. Las propias definiciones de espacios nulo y nulo izquierdo de una matriz $A \in \mathbb{K}^{m\times n}$ aseguran que dichos espacios están estrechamente vinculados con la *independencia lineal* de las filas o de las columnas de la matriz A. En efecto, del Lema 11.1, se observa que las expresiones $x^t A = 0$ y $Ax = 0$ representan al vector nulo escrito como combinación lineal de las filas o de las columnas de A, respectivamente.

11.3. Propiedades de los subespacios fundamentales

Ahora se establece que si dos matrices son equivalentes por filas (columnas) entonces tienen el mismo espacio fila (columna).

Lema 11.3. Sean $A, B \in \mathbb{K}^{m\times n}$. Se cumple que:

(a) Si $A \sim_f B$ entonces $\mathcal{F}(A) = \mathcal{F}(B)$.

(b) Si $A \sim_c B$ entonces $\mathcal{C}(A) = \mathcal{C}(B)$.

Demostración. (a) Si $A \sim_f B$ entonces existe una matriz invertible $Q \in \mathbb{K}^{m\times m}$ tal que $B = QA$. Por el Lema 11.2, $\mathcal{F}(B) \subseteq \mathcal{F}(A)$. Al ser Q invertible, $A = Q^{-1}B$, con lo que $\mathcal{F}(A) \subseteq \mathcal{F}(B)$, de donde se tiene $\mathcal{F}(A) = \mathcal{F}(B)$.

(b) Se demuestra de forma semejante al apartado (a).

$\square$

El resultado anterior dice que la equivalencia por filas no cambia el espacio fila y la equivalencia por columnas no varía el espacio columna.

Un caso particular importante se tiene al utilizar las matrices escalonadas reducidas por filas o por columnas.

Corolario 11.1. Sean $A \in \mathbb{K}^{m\times n}$. Se cumple que:

(a) $\mathcal{F}(A) = \mathcal{F}(R_A)$ donde R_A es la forma escalonada reducida por filas de A.

(b) $\mathcal{C}(A) = \mathcal{C}({}_AR)$ donde ${}_AR$ es la forma escalonada reducida por columnas de A.

Demostración. (a) Sigue directamente del hecho que $A \sim_f R_A$.

(b) Sigue directamente del hecho que $A \sim_{c\ A} R$.

$\square$

En el siguiente resultado se prueba que una relación de inclusión entre subespacios se puede trasladar a una relación de una igualdad puramente matricial.

Proposición 11.1. Sean $A, B \in \mathbb{K}^{m\times n}$. Se cumplen las siguientes propiedades:

(a) $\mathcal{F}(A) \subseteq \mathcal{F}(B)$ si y sólo si existe una matriz $C \in \mathbb{K}^{m\times m}$ tal que $A = CB$.

(b) $\mathcal{C}(A) \subseteq \mathcal{C}(B)$ si y sólo si existe una matriz $C \in \mathbb{K}^{n\times n}$ tal que $A = BC$.

Demostración. (a) La inclusión $\mathcal{F}(A) = \mathcal{F}(CB) \subseteq \mathcal{F}(B)$ se ha probado en el Lema 11.2.

Recíprocamente, si se cumple que $\mathcal{F}(A) \subseteq \mathcal{F}(B)$, se debe probar que $A = CB$ para alguna matriz C.

En efecto[2], se particionan A y B según sus filas mediante

$$A = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix} \qquad \text{y} \qquad B = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}.$$

Como cada fila de A pertenece a $\mathcal{F}(A)$, por hipótesis, también pertenece al subespacio $\mathcal{F}(B)$. Luego, cada fila de A es combinación lineal de las filas de B, es decir, existen ciertos escalares $c_{ij} \in \mathbb{K}$ con $1 \leq i \leq m$ y

[2] Como ejercicio se propone demostrarlo utilizando el Lema 11.1 evitando así la notación de los subíndices en c_{ij}.

$1 \leq j \leq m$ tales que

$$a_i = \sum_{j=1}^{m} c_{ij} b_j, \quad \text{para} \quad i = 1, 2, \ldots, m.$$

Por lo tanto,

$$\begin{aligned} A &= \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix} \\ &= \begin{bmatrix} \sum_{j=1}^{m} c_{1j} b_j \\ \sum_{j=1}^{m} c_{2j} b_j \\ \vdots \\ \sum_{j=1}^{m} c_{mj} b_j \end{bmatrix} \\ &= \begin{bmatrix} c_{11} & c_{12} & \ldots & c_{1m} \\ c_{21} & c_{22} & \ldots & c_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ c_{m1} & c_{m2} & \ldots & c_{mm} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} \\ &= CB, \end{aligned}$$

donde se ha llamado C a la matriz de tamaño $m \times m$ con coeficientes $[c_{ij}]$.

(b) Por el apartado (a), se tiene que $\mathcal{C}(A) \subseteq \mathcal{C}(B)$ si y sólo si $\mathcal{F}(A^t) \subseteq \mathcal{F}(B^t)$ si y sólo si existe una matriz $C \in \mathbb{K}^{n \times n}$ tal que $A^t = C^t B^t$, lo que equivale a $A = BC$.

Observar que $C \in \mathbb{K}^{n \times n}$ puesto que $A^t, B^t \in \mathbb{K}^{n \times m}$.

$\square$

11.4. Dimensión de los subespacios fundamentales

Definición 11.3. Sea $A \in \mathbb{K}^{m\times n}$. Se llama

- **rango fila** de A, y se denota $\mathrm{rg}_f(A)$ a la dimensión del espacio fila de A. En símbolos,

$$\mathrm{rg}_f(A) = \dim_{\mathbb{K}}(\mathcal{F}(A)).$$

- **rango columna** de A, y se denota $\mathrm{rg}_c(A)$, a la dimensión del espacio columna de A. En símbolos,

$$\mathrm{rg}_c(A) = \dim_{\mathbb{K}}(\mathcal{C}(A)).$$

Es claro que:

- $\mathrm{rg}_f(A)$ coincide con el máximo número de filas linealmente independientes de A.
- $\mathrm{rg}_c(A)$ coincide con el máximo número de columnas linealmente independientes de A.

11.4.1. Teorema fundamental del rango

Antes de abordar el Teorema fundamental del rango, será necesario probar un resultado que es intuitivamente claro.

Lema 11.4. Sea $A \in \mathbb{K}^{m\times n}$ una matriz no nula y R_A su forma escalonada reducida por filas. Entonces el conjunto determinado por las filas no nulas de R_A forma una base del subespacio $\mathcal{F}(A)$.

Demostración. Sean $f_1, f_2, \ldots, f_m$ las filas de R_A de las cuales las no nulas corresponden a las r primeras: $f_1, f_2, \ldots, f_r$. Por el Corolario 11.1 es claro que

$$\mathcal{F}(A) = \mathcal{F}(R_A) = \overline{\{f_1, f_2, \ldots, f_r\}}.$$

De este modo, el conjunto formado por las filas no nulas de R_A determinan un sistema de generadores de $\mathcal{F}(A)$.

Falta ver que el conjunto $\{f_1, f_2, \ldots, f_r\}$ es linealmente independiente en $\mathbb{K}^n$. En efecto, se considera una combinación lineal del tipo

$$0 = \alpha_1 f_1 + \alpha_2 f_2 + \cdots + \alpha_r f_r, \qquad \text{con } \alpha_1, \alpha_2, \ldots, \alpha_r \in \mathbb{K}. \tag{11.1}$$

Se supone que los vectores $f_1, f_2, \ldots, f_r$ tienen sus unos principales en las posiciones $k_1, k_2, \ldots, k_r$ con $1 \leq k_1 < k_2 < \ldots < k_r \leq n$. Al operar en (11.1), se observa inmediatamente que $\alpha_1 = 0$ pues el primer elemento no nulo del vector f_1, que se halla en la posición k_1, es un 1 y los demás vectores fila poseen un 0 en esa posición k_1. Sustituyendo en (11.1),

$$0 = \alpha_2 f_2 + \cdots + \alpha_r f_r.$$

Por la misma razón, es inmediato que $\alpha_2 = 0$. Siguiendo de este modo, se llega a $\alpha_1 = \alpha_2 = \cdots = \alpha_r = 0$.

Por lo tanto, $\overline{\{f_1, f_2, \ldots, f_r\}}$ es una base de $\mathcal{F}(A)$. □

Ejemplo 11.3. *Comprobar, para el subconjunto de $\mathbb{R}^4$ dado por*

$$u_1 = (1, 0, 3, 0),\ u_2 = (0, 0, 1, 1),\ u_3 = (0, 0, 0, 3),$$

la propiedad que dice: "Los vectores no nulos de una matriz escalonada de $\mathbb{K}^{m\times n}$ son linealmente independientes en $\mathbb{K}^n$" (véase el Lema 11.4).

◁ Se observa que si los vectores dados se disponen como filas de una matriz, dicha matriz está en forma escalonada por filas (aunque no reducida). Ahora, si

$$(0, 0, 0, 0) = a(1, 0, 3, 0) + b(0, 0, 1, 1) + c(0, 0, 0, 4),$$

comparando la primera componente en ambos miembros se obtiene $a = 0$. Queda entonces

$$(0,0,0,0) = b(0,0,1,1) + c(0,0,0,4).$$

Comparando la tercera componente se obtiene $b = 0$. Finalmente, de la expresión simplificada $(0,0,0,0) = c(0,0,0,4)$, se deduce $c = 0$. $\triangleright$

Teorema 11.1 (Teorema fundamental del rango)**.** Sea $A \in \mathbb{K}^{m\times n}$. Entonces el rango fila de A es igual al rango columna de A y coinciden con $\operatorname{rg}(A)$, es decir

$$\operatorname{rg}_f(A) = \operatorname{rg}_c(A) = \operatorname{rg}(A),$$

donde $\operatorname{rg}(A)$ representa el número de unos principales en la forma escalonada reducida por filas de A.

Demostración. Si $A = O$ entonces es claro que

$$\operatorname{rg}_f(A) = \operatorname{rg}_c(A) = \operatorname{rg}(A) = 0.$$

Sean $A \neq O$ y R_A la forma escalonada reducida por filas de A. Se supone que el número de filas no nulas de R_A es r, es decir, por la Definición 4.5 de rango de A es $r := \operatorname{rg}(A)$.

Por el Lema 11.4, el conjunto de las r filas no nulas de R_A forman una base de $\mathcal{F}(A)$. Por definición de rango fila,

$$\operatorname{rg}_f(A) = \dim_{\mathbb{K}}(\mathcal{F}(A)) = r.$$

Por lo tanto,

$$\operatorname{rg}_f(A) = \operatorname{rg}(A).$$

Por otro lado, aplicando el resultado demostrado a la matriz A^t y utilizando la Proposición 6.5 de la página 257 se tiene que

$$\operatorname{rg}_c(A) = \dim_{\mathbb{K}}(\mathcal{C}(A)) = \dim_{\mathbb{K}}(\mathcal{F}(A^t)) = \operatorname{rg}_f(A^t) = \operatorname{rg}(A^t) = \operatorname{rg}(A),$$

lo que termina la demostración.

Demostración alternativa: De la definición de rango columna de A se tiene que $\mathrm{rg}_c(A) = \dim_{\mathbb{K}}(\mathcal{C}(A))$. Ahora, particionando la matriz A por columnas como $A = \begin{bmatrix} c_1 & c_2 & \dots & c_n \end{bmatrix}$ y utilizando la Proposición 9.4 se tiene que

$$\begin{aligned}
\mathrm{rg}_c(A) &= \dim_{\mathbb{K}}(\mathcal{C}(A)) \\
&= \dim_{\mathbb{K}}(\overline{\{c_1, c_2, \dots, c_n\}}) \\
&= \mathrm{rg}(\begin{bmatrix} [c_1]_{\mathcal{C}} & [c_2]_{\mathcal{C}} & \dots & [c_n]_{\mathcal{C}} \end{bmatrix}) \qquad (11.2) \\
&= \mathrm{rg}(\begin{bmatrix} c_1 & c_2 & \dots & c_n \end{bmatrix}) \\
&= \mathrm{rg}(A), \qquad (11.3)
\end{aligned}$$

puesto que cada columna de A coincide con el vector formado por sus propias coordenadas en la base canónica $\mathcal{C}$ de $\mathbb{K}^m$ (es decir, $[c_i]_{\mathcal{C}} = c_i$ para $i = 1, 2, \dots, n$). De este modo, se ha probado que $\mathrm{rg}_c(A) = \mathrm{rg}(A)$.

Ahora bien, de la Proposición 6.5, $\mathrm{rg}(A^t) = \mathrm{rg}(A)$, por lo tanto, aplicando la definición de rango fila de A y el resultado probado en (11.2) a la matriz A^t se tiene que

$$\mathrm{rg}_f(A) = \dim_{\mathbb{K}}(\mathcal{F}(A)) = \dim_{\mathbb{K}}(\mathcal{C}(A^t)) = \mathrm{rg}(A^t) = \mathrm{rg}(A).$$

Se ha llegado entonces a $\mathrm{rg}_f(A) = \mathrm{rg}_c(A) = \mathrm{rg}(A)$, como se quería demostrar.

$\square$

Teorema 11.2 (Dimensión de los subespacios fundamentales)**.** Sea $A \in \mathbb{K}^{m\times n}$ una matriz con $\mathrm{rg}(A) = r$. Se cumplen las siguientes igualdades:

(a) $\dim_{\mathbb{K}}(\mathcal{F}(A)) = r$.

(b) $\dim_{\mathbb{K}}(\mathcal{C}(A)) = r$.

(c) $\dim_{\mathbb{K}}(\mathcal{N}(A)) = n - r$.

(d) $\dim_{\mathbb{K}}(\mathcal{N}(A^t)) = m - r$.

Demostración. Los apartados (a) y (b) han sido probados en el Teorema 11.1.

El apartado (c) ha sido probado en el Teorema 9.5. De este mismo Teorema se deduce el apartado (d), aplicándolo a A^t en lugar de a la matriz A y recordando que $\mathrm{rg}(A^t) = \mathrm{rg}(A)$. □

11.5. Relación con la equivalencia de matrices

En el Lema 11.3 se ha probado que si dos matrices $A, B \in \mathbb{K}^{m\times n}$ son tales que $A \sim_f B$ entonces $\mathcal{F}(A) = \mathcal{F}(B)$. Un resultado similar se ha presentado para las columnas. ¿Son válidos los recíprocos de los enunciados de dicho lema? La respuesta es afirmativa y se prueba en el siguiente teorema. Nótese que se obtienen nuevas caracterizaciones para la equivalencia por filas y para equivalencia por columnas de dos matrices, esta vez a partir de los subespacios fila y columna.

Teorema 11.3. Sean $A, B \in \mathbb{K}^{m\times n}$. Se cumple que:

(a) $A \sim_f B$ si y sólo si $\mathcal{F}(A) = \mathcal{F}(B)$.

(b) $A \sim_c B$ si y sólo si $\mathcal{C}(A) = \mathcal{C}(B)$.

Demostración. (a) Que la condición $A \sim_f B$ es suficiente para que se cumpla $\mathcal{F}(A) = \mathcal{F}(B)$ se ha probado en el Lema 11.3.

Ahora se probará que la condición $A \sim_f B$ es necesaria para $\mathcal{F}(A) = \mathcal{F}(B)$, suponiendo que se cumple $\mathcal{F}(A) = \mathcal{F}(B)$ y probando que $A \sim_f B$.

En efecto, si R_A y R_B denotan las formas escalonadas reducidas por filas de A y B con unos principales en las columnas $a_1 < a_2 < \ldots < a_s$ y $b_1 < b_2 < \ldots < b_t$, respectivamente, del Corolario 11.1 se puede asegurar que

$$\mathcal{F}(R_A) = \mathcal{F}(A) = \mathcal{F}(B) = \mathcal{F}(R_B).$$

Por definición de rango, $s = \mathrm{rg}(A)$ y $t = \mathrm{rg}(B)$. Del Teorema fundamental del rango se tiene que

$$\begin{aligned} s &= \mathrm{rg}(A) \\ &= \mathrm{rg}_f(A) \\ &= \dim_{\mathbb{K}}(\mathcal{F}(A)) \\ &= \dim_{\mathbb{K}}(\mathcal{F}(B)) \\ &= \mathrm{rg}_f(B) \\ &= \mathrm{rg}(B) \\ &= t, \end{aligned}$$

con lo que $s = t$.

Si se prueba que $R_A = R_B$ entonces se obtiene el resultado deseado pues, al ser $\sim_f$ una relación de equivalencia,

$$A \sim_f R_A = R_B \sim_f B.$$

En efecto, de $\mathcal{F}(R_A) \subseteq \mathcal{F}(R_B)$ y la Proposición 11.1, existe una matriz $S' \in \mathbb{K}^{m\times m}$ tal que

$$R_A = S'R_B.$$

Por estar en forma escalonada reducida por filas, las últimas $m-s$ $(= m-t)$ filas de R_A y de R_B son nulas. Si se particionan R_A y R_B en dos bloques filas con s y $m-s$ filas, respectivamente, y S' en una matriz de tipo $[s + (m-s)] \times [s + (m-s)]$ (es decir, se particiona en cuatro bloques, siendo el primero de tamaño $s \times s$) se tiene

$$R_A = \begin{bmatrix} R'_A \\ O \end{bmatrix}, \qquad R_B = \begin{bmatrix} R'_B \\ O \end{bmatrix} \qquad \text{y} \qquad S' = \begin{bmatrix} S & S_2 \\ S_3 & S_4 \end{bmatrix}$$

con $R'_A, R'_B \in \mathbb{K}^{s\times n}$ y $S \in \mathbb{K}^{s\times s}$. En realidad, los bloques S_2 y S_4 no intervienen en la operación y, sin pérdida de generalidad, pueden considerarse nulos.

Multiplicando por bloques en $R_A = S'R_B$ e igualando se tiene

$$R'_A = SR'_B \qquad \text{y} \qquad O = S_3R'_B. \tag{11.4}$$

De la segunda igualdad es fácil probar (se propone como ejercicio) que $\mathcal{R}(R'_B) \subseteq \mathcal{N}(S_3)$. Puesto que un sistema de generadores del espacio imagen de R'_B está compuesto por los vectores canónicos $e_1, e_2, \ldots, e_s \in \mathbb{K}^{s\times 1}$, de $\mathcal{R}(R'_B) = \overline{\{e_1, e_2, \ldots, e_s\}}$ se tiene que $S_3e_j = 0$ para todo $j = 1, 2, \ldots, s$. De este modo, $S_3 = O$ y

$$S' = \begin{bmatrix} S & O \\ O & O \end{bmatrix}.$$

Ahora alcanza con probar que $S = I_s$, puesto que por (11.4) equivale a decir que $R'_A = R'_B$, y así, $R_A = R_B$.

De $\mathcal{F}(R_B) \subseteq \mathcal{F}(R_A)$, un razonamiento similar permite asegurar la existencia de una matriz $T \in \mathbb{K}^{s\times s}$ tal que

$$R'_B = TR'_A, \qquad \text{con } T \in \mathbb{K}^{s\times s}. \tag{11.5}$$

Si las primeras columnas $1, 2, \ldots, b_1 - 1$ de R'_B (si las hay, es decir si $b_1 > 1$) son nulas entonces, por (11.4), las columnas correspondientes de R'_A también lo son (pues cada columna de R'_A se obtiene de multiplicar S por la correspondiente columna de R'_B). Luego, $a_1 \geq b_1$.

Si $b_1 = 1$ entonces es claro que $a_1 \geq 1 = b_1$. Si las primeras columnas $1, 2, \ldots, a_1 - 1$ de R'_A son nulas (si las hay, es decir si $a_1 > 1$) entonces, por (11.5), las columnas correspondientes de R'_B también lo son (pues cada columna de R'_B se obtiene de multiplicar T por la correspondiente columna de R'_A). Luego, $b_1 \geq a_1$. Si $a_1 = 1$ entonces es claro que $b_1 \geq 1 = a_1$. Por lo tanto, de ambas desigualdades, se obtiene $a_1 = b_1$.

Ahora, como las columnas a_1-ésima de R'_A y b_1-ésima de R'_B coinciden con el vector canónico $e_1 \in \mathbb{K}^{s\times 1}$, multiplicando por bloques, de (11.4) se deduce que $Se_1 = e_1$.

Las columnas de R'_B posteriores a la b_1-ésima y previas a la b_2-ésima (si las hay) son combinación lineal de la columna b_1-ésima de R'_B. Multiplicando por bloques, por (11.4), lo mismo ocurre con las columnas de R'_A posteriores a la a_1-ésima y previas a la b_2-ésima. Hasta ahora R'_A y R'_B coinciden en las primeras $b_2 - 1$ columnas. Por lo tanto, $a_2 \geq b_2$.

Como las columnas posteriores a la a_1-ésima y previas a la a_2-ésima (si las hay) son combinación lineal de la columna a_1-ésima, utilizando (11.5), se tiene que lo mismo ocurre con las columnas de R'_B posteriores a la b_1-ésima y previas a la a_2-ésima. Entonces $b_2 \geq a_2$. Así, $a_2 = b_2$.

Ahora, como las columnas a_2-ésima de R'_A y b_2-ésima de R'_B coinciden con el vector canónico e_2, multiplicando por bloques, de (11.4) se deduce que $Se_2 = e_2$.

Continuando con un razonamiento similar se obtiene que $a_j = b_j$ y $Se_j = e_j$ para todo $j = 3, \ldots, s$. En definitiva, $S = I_s$.

(b) Se deduce del apartado anterior

$$A \sim_c B \quad \Leftrightarrow \quad A^t \sim_f B^t \ \Leftrightarrow\ \mathcal{F}(A^t) = \mathcal{F}(B^t) \ \Leftrightarrow\ \mathcal{C}(A) = \mathcal{C}(B).$$

$\square$

11.6. Bases de los subespacios fundamentales

Sea $A \in \mathbb{K}^{m\times n}$ una matriz no nula de rango $r > 0$ y R_A su forma escalonada reducida por filas. Se sabe que existe una matriz invertible $Q \in \mathbb{K}^{m\times m}$ tal que

$$QA = R_A = \begin{bmatrix} Q' \\ O \end{bmatrix}, \qquad \text{con} \qquad Q' \in \mathbb{K}^{r\times n}. \tag{11.6}$$

Luego,

$$Q \begin{bmatrix} A & I_m \end{bmatrix} = \begin{bmatrix} R_A & Q \end{bmatrix}. \tag{11.7}$$

A continuación se indican métodos para determinar bases para cada uno de los cuatro subespacios fundamentales: $\mathcal{F}(A)$, $\mathcal{C}(A)$, $\mathcal{N}(A)$ y $\mathcal{N}(A^t)$.

Base del espacio fila $\mathcal{F}(A)$: En el Lema 11.4 se ha probado que una base de $\mathcal{F}(A)$ viene dada por el conjunto determinado por las filas no nulas de R_A, es decir,

$$\mathcal{B}_{\mathcal{F}(A)} = \{\text{Filas no nulas de } R_A\}.$$

Base del espacio columna $\mathcal{C}(A)$: La Observación 8.13 asegura que las columnas de R_A que no contienen los unos principales se pueden escribir como combinación lineal de las columnas precedentes que sí los tienen. Estas últimas forman un sistema generador del espacio columna de R_A y, por ser vectores de la base canónica de $\mathbb{K}^m$, son linealmente independientes. Por la Proposición 8.8, las columnas de A verifican las mismas relaciones de dependencia lineal que las columnas de R_A. Por tanto, para obtener una base del espacio columna de A basta considerar las columnas de A correspondientes a las columnas de R_A que contienen los unos principales, es decir,

$$\mathcal{B}_{\mathcal{C}(A)} = \{\text{Columnas de } A \text{ correspondientes a las columnas básicas de } R_A\}.$$

Base del espacio nulo $\mathcal{N}(A)$: El sistema lineal homogéneo $Ax = 0$ es equivalente al sistema $R_Ax = 0$. Se resuelve este último y se asignan parámetros a las variables libres (que son las que no corresponden a los unos principales). La cantidad de estos parámetros (que es $n - \operatorname{rg}(A)$) proporciona la dimensión del espacio nulo de A y cada vector de una base de $\mathcal{N}(A)$ se obtiene asignando el valor 1 a un parámetro (por vez) y cero a todos los demás, recorriendo todos los parámetros (véanse Teorema 9.5 y Ejemplo 9.13). Resumiendo,

$$\mathcal{B}_{\mathcal{N}(A)} = \mathcal{B}_{\mathcal{N}(R_A)} : \text{ se obtiene resolviendo } R_Ax = 0.$$

Base del espacio nulo izquierdo $\mathcal{N}(A^t)$: Particionando Q por filas, por (11.6) y teniendo en cuenta que $\operatorname{rg}(A) = r$ se tiene que

$$\begin{bmatrix} q_1^t A \\ q_2^t A \\ \vdots \\ q_m^t A \end{bmatrix} = \begin{bmatrix} q_1^t \\ q_2^t \\ \vdots \\ q_m^t \end{bmatrix} A = QA = R_A = \begin{bmatrix} Q' \\ O \end{bmatrix},$$

de donde $q_i^t A = 0$ para $i = r+1, \ldots, m$. Luego, $A^t q_i = 0$ para $i = r+1, \ldots, m$. Al ser $Q \in \mathbb{K}^{m\times m}$ una matriz invertible, sus filas q_i^t, $i = 1, 2, \ldots, m$ forman una base de $\mathbb{K}^{1\times m}$ y, por lo tanto $\{q_{r+1}, \ldots, q_m\} \subseteq \mathcal{N}(A^t)$ es un conjunto linealmente independiente $\mathbb{K}^{m\times 1}$. Por el Teorema 11.2, $\dim(\mathcal{N}(A^t)) = m - r$, con lo que $\{q_{r+1}, \ldots, q_m\}$ es una base de $\mathcal{N}(A^t)$. Resumiendo,

$$\mathcal{B}_{\mathcal{N}(A^t)} = \left\{ q_{r+1}, \ldots, q_m \in \mathbb{K}^{m\times 1} : \begin{bmatrix} q_1^t \\ \vdots \\ q_r^t \\ q_{r+1}^t \\ \vdots \\ q_m^t \end{bmatrix} A = R_A \right\}.$$

El siguiente ejemplo ha sido resuelto anteriormente aplicando la definición de cada uno de los subespacios. Ahora se resolverá aplicando los resultados obtenidos, lo que permite ahorrar algunos cálculos.

Ejemplo 11.4. *Encontrar una base para cada uno de los cuatro subespacios fundamentales asociados a la matriz*

$$A = \begin{bmatrix} 1 & 4 & 0 & -3 \\ 3 & -1 & 2 & 0 \\ 5 & 7 & 2 & -6 \end{bmatrix} \in \mathbb{R}^{3\times 4}.$$

$\lhd$ Se realizarán operaciones elementales por filas a la matriz $\begin{bmatrix} A & I_3 \end{bmatrix}$ hasta obtener la forma escalonada reducida por filas R_A de A. Se obtiene

$$\left[\begin{array}{cccc|ccc} 1 & 4 & 0 & -3 & 1 & 0 & 0 \\ 3 & -1 & 2 & 0 & 0 & 1 & 0 \\ 5 & 7 & 2 & -6 & 0 & 0 & 1 \end{array}\right] \sim_f \left[\begin{array}{cccc|ccc} 1 & 0 & \frac{8}{13} & -\frac{3}{13} & \frac{1}{13} & \frac{4}{13} & 0 \\ 0 & 1 & -\frac{2}{13} & -\frac{9}{13} & \frac{3}{13} & -\frac{1}{13} & 0 \\ \hline 0 & 0 & 0 & 0 & -2 & -1 & 1 \end{array}\right]. \tag{11.8}$$

Una base de $\mathcal{F}(A)$ es

$$\left\{ \left(1, 0, \frac{8}{13}, -\frac{3}{13}\right), \left(0, 1, -\frac{2}{13}, -\frac{9}{13}\right) \right\}.$$

Puesto que $\mathcal{F}(A) = \mathcal{F}(R_A)$ y como al realizar operaciones elementales no se han producido intercambios de filas, otra base de $\mathcal{F}(A)$ es

$$\{(1,4,0,-3),(3,-1,2,0)\}.$$

Es claro que $(13,13,6,-12) \in \mathcal{F}(A)$ y que $(13,13,6,12) \notin \mathcal{F}(A)$.

Debido a que los unos principales se encuentran en las dos primeras columnas de R_A, las dos primeras columnas de A proporcionan una base de $\mathcal{C}(A)$, es decir

$$\left\{ \begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix}, \begin{bmatrix} 4 \\ -1 \\ 7 \end{bmatrix} \right\}$$

es una base de $\mathcal{C}(A)$. Sin embargo, es importante observar que las dos primeras columnas de R_A no determinan una base de $\mathcal{C}(A)$.

Una base del espacio nulo de A se encuentra con el mismo procedimiento que en el Ejemplo 11.2, es decir, resolviendo el sistema asociado $R_Ax = 0$. En este caso, una base de $\mathcal{N}(A)$ es

$$\left\{ \begin{bmatrix} -\frac{8}{13} \\ \frac{2}{13} \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} \frac{3}{13} \\ \frac{9}{13} \\ 0 \\ 1 \end{bmatrix} \right\},$$

que resulta muy sencillo tras despejar en la expresión (11.8). Observar que basta asignar adecuadamente ceros y unos a las variables libres y aparecerán los elementos de las filas no nulas y de las columnas no básicas de R_A cambiados de signo.

Por último, debido a que $\mathrm{rg}(A) = 2$ se tiene que $\dim(\mathcal{N}(A^t)) = m - r = 3 - 2 = 1$, y una base del espacio nulo izquierdo $\mathcal{N}(A^t)$ la proporciona el vector de la última fila de (11.8) en que se ha transformado I_3, es decir, una base del espacio nulo izquierdo $\mathcal{N}(A^t)$ es $\left\{ \begin{bmatrix} -2 \\ -1 \\ 1 \end{bmatrix} \right\}$. ▷

Ejemplo 11.5. *A partir de la igualdad $\mathcal{C}(A) = \mathcal{F}(A^t)$ encontrar una base del espacio columna y una base del espacio nulo asociado a la matriz*

$$A = \begin{bmatrix} 1 & 4 & 0 & -3 \\ 3 & -1 & 2 & 0 \\ 5 & 7 & 2 & -6 \end{bmatrix} \in \mathbb{R}^{3\times 4}$$

y la forma general de los vectores de $\mathcal{C}(A)$. Decidir si $\begin{bmatrix} -2 \\ 3 \\ -1 \end{bmatrix} \in \mathcal{C}(A)$ o no.

$\triangleleft$ Se realizarán operaciones elementales por filas a la matriz $\begin{bmatrix} A^t & I_4 \end{bmatrix}$ hasta obtener su forma escalonada reducida por filas. Se obtiene

$$\left[\begin{array}{ccc|cccc} 1 & 3 & 5 & 1 & 0 & 0 & 0 \\ 4 & -1 & 7 & 0 & 1 & 0 & 0 \\ 0 & 2 & 2 & 0 & 0 & 1 & 0 \\ -3 & 0 & -6 & 0 & 0 & 0 & 1 \end{array}\right] \sim_f \left[\begin{array}{ccc|cccc} 1 & 0 & 2 & \frac{1}{13} & \frac{3}{13} & 0 & 0 \\ 0 & 1 & 1 & \frac{4}{13} & -\frac{1}{13} & 0 & 0 \\ \hline 0 & 0 & 0 & -\frac{8}{13} & \frac{2}{13} & 1 & 0 \\ 0 & 0 & 0 & \frac{3}{13} & \frac{9}{13} & 0 & 1 \end{array}\right].$$

Al ser $A^t \sim_f R_{A^t}$, una base de $\mathcal{C}(A) = \mathcal{F}(A^t) = \mathcal{F}(R_{A^t})$ es

$$\left\{ \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \right\}.$$

Ahora, un vector arbitrario $\begin{bmatrix} a \\ b \\ c \end{bmatrix} \in \mathcal{C}(A)$ si y sólo si

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} = \alpha \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} + \beta \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix},$$

para ciertos $\alpha, \beta \in \mathbb{R}$. Luego, $a = \alpha$, $b = \beta$, $c = 2\alpha + \beta$, con lo que se debe cumplir que $c = 2a + b$. Es decir, $\mathcal{C}(A)$ está formado por los vectores de la forma $\begin{bmatrix} a \\ b \\ 2a+b \end{bmatrix}$ con $a, b \in \mathbb{R}$. Además, $\begin{bmatrix} -2 \\ 3 \\ -1 \end{bmatrix} \in \mathcal{C}(A)$ pues $-1 = 2(-2) + 3$.

Procediendo como en el ejemplo anterior, una base de $\mathcal{N}((A^t)^t) = \mathcal{N}(A)$ viene dada por

$$\left\{ \begin{bmatrix} -\frac{8}{13} \\ \frac{2}{13} \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} \frac{3}{13} \\ \frac{9}{13} \\ 0 \\ 1 \end{bmatrix} \right\}.$$

▷

Observación 11.3. Nótese que tanto la base de $\mathcal{F}(A)$ hallada en el Ejemplo 11.4 como la base de $\mathcal{C}(A)$ hallada en el Ejemplo 11.5, permiten escribir fácilmente un vector arbitratio del espacio como combinación lineal de los básicos puesto que varias componentes vienen determinadas por unos y ceros. No ocurre lo mismo con la base de $\mathcal{C}(A)$ hallada en el Ejemplo 11.4.

Observación 11.4. Al definir la matriz escalonada reducida por filas de una matriz dada (Definición 4.1), se indicó que más adelante se justificaría el nombre dado a las *columnas básicas*. Ahora es posible realizarlo. Las columnas que contienen los unos principales se llaman columnas básicas puesto que forman una base del espacio columna de la matriz R_A.

11.7. Teorema fundamental del Álgebra Lineal

Ahora se establecerán importantes relaciones de ortogonalidad (en los espacios euclídeos cartesianos $\mathbb{R}^n$ y $\mathbb{R}^m$) de los subespacios fundamentales de una matriz $A \in \mathbb{R}^{m \times n}$. También, se mostrará cómo $\mathbb{R}^n$ y $\mathbb{R}^m$ se pueden expresar como suma directa (ortogonal) de dos subespacios fundamentales, hecho que suele conocerse como el **Teorema fundamental del Álgebra Lineal**. Para establecer estos resultados es crucial la utilización del producto escalar canónico de $\mathbb{R}^n$ y $\mathbb{R}^m$.

Los subespacios fundamentales $\mathcal{N}(A)$ y $\mathcal{C}(A^t)$ son subespacios del mismo espacio vectorial, a saber de $\mathbb{R}^n$. Del mismo modo, $\mathcal{N}(A^t)$ y $\mathcal{C}(A)$ son subespacios de $\mathbb{R}^m$. El próximo resultado muestra la relación de ortogonalidad entre cada uno de estos pares.

Lema 11.5 (Ortogonalidad de los subespacios fundamentales)**.** Sea $A \in \mathbb{R}^{m\times n}$. Se cumple que:

(a) $[\mathcal{N}(A)]^\perp = \mathcal{C}(A^t)$.

(b) $[\mathcal{C}(A)]^\perp = \mathcal{N}(A^t)$.

Demostración. (a) Primero se probará la inclusión:

$$\mathcal{C}(A^t) \subseteq [\mathcal{N}(A)]^\perp.$$

En efecto, sea $x \in \mathcal{C}(A^t)$. Por el Lema 11.1, existe $y \in \mathbb{R}^m$ tal que $x = A^t y$.

Sea, ahora, $z \in \mathcal{N}(A)$. Se debe probar que $\langle x, z\rangle = 0$. En efecto, recordando que el producto escalar canónico de $\mathbb{R}^n$ se puede escribir como $\langle u, v\rangle = u^t v$ se tiene que

$$\langle x, z\rangle = \langle A^t y, z\rangle = (A^t y)^t z = y^t (A^t)^t z = y^t A z = y^t 0 = 0,$$

pues de $z \in \mathcal{N}(A)$ se cumple que $Az = 0$. Luego, $x \in [\mathcal{N}(A)]^\perp$, de donde resulta la inclusión requerida.

Ahora se probará que ambos subespacios tienen la misma dimensión, de donde resultará que

$$[\mathcal{N}(A)]^\perp = \mathcal{C}(A^t).$$

En efecto, por el Teorema 11.2,

$$\dim_{\mathbb{R}}(\mathcal{C}(A^t)) = \dim_{\mathbb{R}}(\mathcal{F}(A)) = \operatorname{rg}(A).$$

Del Corolario 10.2 y del Teorema 11.2 se tiene que

$$\dim_{\mathbb{R}}([\mathcal{N}(A)]^{\perp}) = n - \dim_{\mathbb{R}}(\mathcal{N}(A)) = \operatorname{rg}(A).$$

Luego, $\dim_{\mathbb{R}}(\mathcal{C}(A^t)) = \dim_{\mathbb{R}}([\mathcal{N}(A)]^{\perp})$.

(b) Ahora bien, el apartado (a) aplicado a la matriz A^t establece que

$$[\mathcal{N}(A^t)]^{\perp} = \mathcal{C}((A^t)^t) = \mathcal{C}(A).$$

Tomando ortogonales y recordando que $(S^{\perp})^{\perp} = S$ para todo subespacio S, se tiene

$$\mathcal{N}(A^t) = ([\mathcal{N}(A^t)]^{\perp})^{\perp} = [\mathcal{C}(A)]^{\perp}.$$

□

En este momento, se tienen todos los elementos necesarios para establecer el resultado que proporciona la estructura de los espacios $\mathbb{R}^n$ y $\mathbb{R}^m$ a partir de los cuatro subespacios fundamentales asociados a una matriz $A \in \mathbb{R}^{m\times n}$.

Teorema 11.4 (Teorema fundamental del Álgebra Lineal)**.** Sea $A \in \mathbb{R}^{m\times n}$. Se cumplen las siguientes relaciones:

(a) $\mathbb{R}^n = \mathcal{N}(A) \oplus^{\perp} \mathcal{C}(A^t)$.

(b) $\mathbb{R}^m = \mathcal{N}(A^t) \oplus^{\perp} \mathcal{C}(A)$.

Demostración. Por el Teorema de la proyección ortogonal (Teorema 10.4) y del Lema 11.5 se puede establecer que

$$\mathbb{R}^n = \mathcal{N}(A) \oplus^{\perp} [\mathcal{N}(A)]^{\perp} = \mathcal{N}(A) \oplus^{\perp} \mathcal{C}(A^t).$$

Ahora, tomando ortogonales en $[\mathcal{C}(A)]^{\perp} = \mathcal{N}(A^t)$ se tiene $\mathcal{C}(A) = [\mathcal{N}(A^t)]^{\perp}$. Del mismo modo que antes,

$$\mathbb{R}^m = \mathcal{N}(A^t) \oplus^{\perp} [\mathcal{N}(A^t)]^{\perp} = \mathcal{N}(A^t) \oplus^{\perp} \mathcal{C}(A),$$

lo que termina la demostración. □

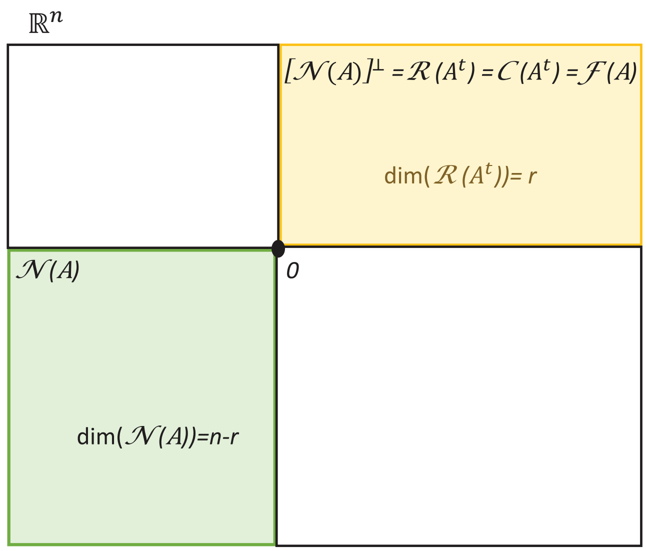


Figura 11.1: Los subespacios fundamentales $\mathcal{N}(A)$ y $\mathcal{C}(A^t)$ de $\mathbb{R}^n$.

Una representación geométrica de cada apartado del Teorema fundamental del Álgebra Lineal se aprecia en las Figuras 11.1 y 11.2, respectivamente.

Ejemplo 11.6. *Para la matriz A del Ejemplo 11.4, descomponer adecuadamente los espacios $\mathbb{R}^4$ y $\mathbb{R}^3$ como suma directa ortogonal de los cuatro subespacios fundamentales.*

$\lhd$ Puesto que $A \in \mathbb{R}^{3\times 4}$, es posible descomponer

$$\mathbb{R}^4 = \mathcal{N}(A) \oplus^\perp \mathcal{C}(A^t) \quad \text{y} \quad \mathbb{R}^3 = \mathcal{N}(A^t) \oplus^\perp \mathcal{C}(A).$$

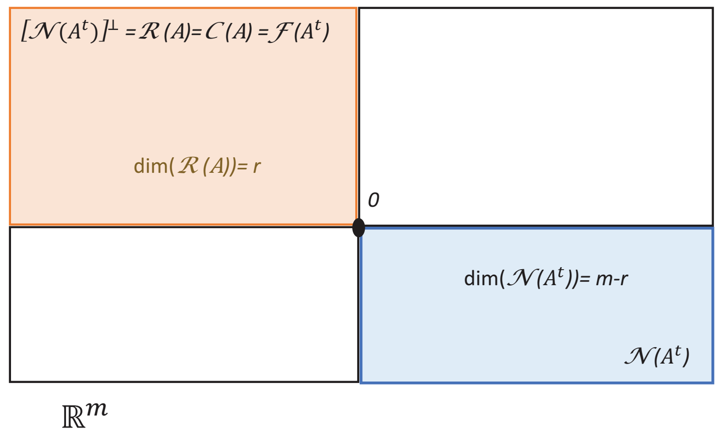


Figura 11.2: Los subespacios fundamentales $\mathcal{C}(A)$ y $\mathcal{N}(A^t)$ de $\mathbb{R}^m$.

De acuerdo a los cálculos realizados en el Ejemplo 11.4 se tiene que

$$
\begin{aligned}
\mathbb{R}^4 &= \mathcal{N}(A) \oplus^\perp \mathcal{C}(A^t) \\
&= \overline{\left\{ \begin{bmatrix} -\frac{8}{13} \\ \frac{2}{13} \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} \frac{3}{13} \\ \frac{9}{13} \\ 0 \\ 1 \end{bmatrix} \right\}} \oplus^\perp \overline{\left\{ \begin{bmatrix} 1 \\ 0 \\ \frac{8}{13} \\ -\frac{3}{13} \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ -\frac{2}{13} \\ -\frac{9}{13} \end{bmatrix} \right\}}
\end{aligned}
$$

y

$$
\begin{aligned}
\mathbb{R}^3 &= \mathcal{N}(A^t) \oplus^\perp \mathcal{C}(A) \\
&= \overline{\left\{ \begin{bmatrix} -2 \\ -1 \\ 1 \end{bmatrix} \right\}} \oplus^\perp \overline{\left\{ \begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix}, \begin{bmatrix} 4 \\ -1 \\ 7 \end{bmatrix} \right\}}.
\end{aligned}
$$

Las dimensiones de cada uno de los subespacios fundamentales son ahora inmediatas.

Observar que los espacios euclídeos $\mathbb{R}^4$ y $\mathbb{R}^4$ se han descompuesto como suma directa de subespacios ortogonales. Sin embargo, excepto para $\mathcal{N}(A^t)$ (cuya base tiene un único vector, y por tanto es ortogonal), las bases halladas para los restantes subespacios fundamentales no son bases ortogonales, aunque se las podría convertir mediante el método de ortogonalización de Gram-Schmidt. ▷

Para terminar esta subsección se analizan los subespacios fundamentales de la matriz de Gram A^tA. Más específicamente, se determinan sus espacios nulo y columna. Además, se calcula el rango de A^tA y el de AA^t.

Proposición 11.2. Sea $A \in \mathbb{R}^{m\times n}$. Entonces

(a) $\mathcal{N}(A^tA) = \mathcal{N}(A)$.

(b) $\mathcal{C}(A^tA) = \mathcal{C}(A^t)$. En consecuencia,

$$\operatorname{rg}(A^tA) = \operatorname{rg}(A) = \operatorname{rg}(A^t) = \operatorname{rg}(AA^t).$$

Demostración. El apartado (a) se demostrará por doble inclusión.

(a) ($\subseteq$) $x \in \mathcal{N}(A^tA) \Rightarrow A^tAx = 0 \Rightarrow x^tA^tAx = 0 \Rightarrow (Ax)^tAx = 0 \Rightarrow \langle Ax, Ax\rangle = 0 \Rightarrow Ax = 0 \Rightarrow x \in \mathcal{N}(A)$ donde $\langle\cdot,\cdot\rangle$ es el producto escalar canónico de $\mathbb{R}^m$.

($\supseteq$) $x \in \mathcal{N}(A) \Rightarrow Ax = 0 \Rightarrow A^tAx = 0 \Rightarrow x \in \mathcal{N}(A^tA)$.

(b) Calculando complementos ortogonales en el apartado (a) y utilizando el Lema 11.5 se tiene que

$$\mathcal{R}(A^tA) = \mathcal{R}((A^tA)^t) = [\mathcal{N}(A^tA)]^\perp = [\mathcal{N}(A)]^\perp = \mathcal{R}(A^t).$$

Ahora basta recordar que el espacio imagen de una matriz coincide con su espacio columna.

En consecuencia,

$$\begin{aligned} \mathrm{rg}(A^tA) &= \dim_{\mathbb{R}}(\mathcal{C}(A^tA)) \\ &= \dim_{\mathbb{R}}(\mathcal{C}(A^t)) \\ &= \mathrm{rg}(A^t) \\ &= \mathrm{rg}(A). \end{aligned}$$

Aplicando este resultado a A^t se obtiene

$$\begin{aligned} \mathrm{rg}(A) &= \mathrm{rg}(A^t) \\ &= \mathrm{rg}((A^t)^tA^t) \\ &= \mathrm{rg}(AA^t). \end{aligned}$$

$\square$

Utilizando resultados previos, a continuación se presenta una aplicación de gran importancia, sobre todo, en las Ingenierías.

11.7.1. Método de los mínimos cuadrados

Sean $A \in \mathbb{R}^{m\times n}$ y $b \in \mathbb{R}^{m\times 1}$. Se trata de estudiar el sistema lineal $Ax = b$ siendo

$$A := \begin{bmatrix} c_1 & \dots & c_n \end{bmatrix} \in \mathbb{R}^{m\times(1+\cdots+1)} \qquad \text{y} \qquad x := \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix},$$

en caso de que sea un sistema incompatible.

Planteamiento del problema: En primer lugar se observa que

$$Ax = b \text{ es incompatible} \qquad \Leftrightarrow \qquad b \notin \mathcal{R}(A).$$

En efecto,

$$\begin{aligned}
\exists x \in \mathbb{R}^{n\times 1} : Ax = b \quad &\Leftrightarrow \quad \begin{bmatrix} c_1 & \dots & c_n \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = b \\
&\Leftrightarrow \quad x_1c_1 + \cdots + x_nc_n = b \\
&\Leftrightarrow \quad b \in \overline{\{c_1, \dots, c_n\}} \\
&\Leftrightarrow \quad b \in \mathcal{R}(A).
\end{aligned}$$

Si bien el sistema planteado carece de solución pues $Ax - b \neq 0$ para todo $x \in \mathbb{R}^{n\times 1}$, se busca alguna solución aproximada $x \in \mathbb{R}^{n\times 1}$ tal que

$$\|Ax - b\|_2 \text{ sea mínima} \tag{11.8}$$

en el sentido que se define a continuación.

Es importante destacar que para establecer este método se utilizará la norma 2.

Definición 11.4. Sean $A \in \mathbb{R}^{m\times n}$ y $b \in \mathbb{R}^{m\times 1}$. Se llama **solución de mínimos cuadrados** de $Ax = b$ a todo vector $v \in \mathbb{R}^{n\times 1}$ tal que

$$\|Av - b\|_2 \leq \|Au - b\|_2, \forall u \in \mathbb{R}^{n\times 1}, \tag{11.9}$$

es decir,

$$\|Av - b\|_2 = \min_{u\in\mathbb{R}^{n\times 1}} \|Au - b\|_2.$$

El nombre del método queda justificado pues al calcular $\|Av - b\|_2$ mediante los elementos de la matriz y las coordendas de los vectores aparece una suma de cuadrados donde se busca que cada uno de los términos tome el menor valor posible.

Soluciones de mínimos cuadrados: El problema de la mejor aproximación proporciona la solución al problema de los mínimos cuadrados.

Proposición 11.3. Sean $A \in \mathbb{R}^{m\times n}$ y $b \in \mathbb{R}^{m\times 1}$. El conjunto S_{MC} de todas las soluciones de mínimos cuadrados del sistema $Ax = b$ es

$$\begin{aligned} S_{MC} &= \{v \in \mathbb{R}^{n\times 1} : Av = \operatorname{proy}_{\mathcal{R}(A)}(b)\} \ (\textbf{solución geométrica}) \\ &= \{v \in \mathbb{R}^{n\times 1} : A^t Av = A^t b\}, \qquad (\textbf{solución algebraica}) \end{aligned}$$

donde la última se conoce como las **ecuaciones normales** asociadas al sistema $Ax = b$.

Demostración. Dado que

$$\mathcal{R}(A) = \{Au : u \in \mathbb{R}^{n\times 1}\},$$

la desigualdad (11.9) se puede expresar mediante

$$\|Av - b\|_2 \leq \|y - b\|_2, \forall y \in \mathcal{R}(A). \tag{11.10}$$

Al tratarse de la solución del problema de la mejor aproximación, los vectores $v \in \mathbb{R}^{n\times 1}$ que satisfacen[3] (11.10) son los mismos que los que satisfacen

$$Av = \operatorname{proy}_{\mathcal{R}(A)}(b),$$

lo que permite establecer la solución geométrica (puesto que se ha determinado a través de la proyección ortogonal).

Ahora, por definición de proyección ortogonal de b sobre $\mathcal{R}(A)$, es conocido que $\operatorname{proy}_{\mathcal{R}(A)}(b)$ es el único vector de $\mathcal{R}(A)$ tal que

$$b - \operatorname{proy}_{\mathcal{R}(A)}(b) \perp \mathcal{R}(A).$$

De nuevo, teniendo en cuenta la definición del conjunto $\mathcal{R}(A)$, un vector

[3] Es decir, los vectores v que son las soluciones de mínimos cuadrados de $Ax = b$.

$v \in \mathbb{R}^{n\times 1}$ satisface

$$\begin{aligned} Av = \text{proy}_{\mathcal{R}(A)}(b) &\Leftrightarrow Au \perp b - Av, \text{ para todo } u \in \mathbb{R}^{n\times 1}, \\ &\Leftrightarrow \langle Au, b - Av\rangle = 0, \text{ para todo } u \in \mathbb{R}^{n\times 1}, \\ &\Leftrightarrow (Au)^t(b - Av) = 0, \text{ para todo } u \in \mathbb{R}^{n\times 1}, \\ &\Leftrightarrow u^tA^t(b - Av) = 0, \text{ para todo } u \in \mathbb{R}^{n\times 1}, \\ &\Leftrightarrow \langle u, A^t(b - Av)\rangle = 0, \text{ para todo } u \in \mathbb{R}^{n\times 1}, \end{aligned}$$

es decir, $A^t(b-Av)$ debe ser ortogonal a todos los vectores del espacio $\mathbb{R}^{n\times 1}$. Como el único vector ortogonal a todos los del espacio al que pertenece es el vector nulo, se tiene que

$$A^t(b - Av) = 0.$$

Así, los vectores v deben ser solución del sistema de ecuaciones lineales

$$A^tAv = A^tb,$$

lo que permite establecer la solución algebraica (puesto que se ha determinado a través de la solución de un sistema de ecuaciones lineales).

Las ecuaciones del sistema obtenido, que tiene n ecuaciones lineales con n incógnitas, se conocen como las **ecuaciones normales** asociadas al sistema $Ax = b$. $\square$

Existencia y unicidad de la solución de mínimos cuadrados: Antes de continuar es importante tener claro si las ecuaciones normales determinan un sistema lineal compatible o no. La respuesta es afirmativa, como se prueba en el siguiente resultado. Además, se establece una condición suficiente para que la solución de mínimos cuadrados sea única.

Proposición 11.4. Sean $A \in \mathbb{R}^{m\times n}$ y $b \in \mathbb{R}^{m\times 1}$. Se cumple que:

(a) Las ecuaciones normales

$$A^tAv = A^tb$$

son siempre compatibles.

(b) Si $\text{rg}(A) = n$ entonces existe una única solución de mínimos cuadrados y está dada por[a]

$$v = A_Ib = (A^tA)^{-1}A^tb.$$

[a] La notación A_I se introduce para hacer hincapié en que la matriz

$$A_I := (A^tA)^{-1}A^t$$

es una inversa a izquierda de A, hecho que se comprueba de forma inmediata.

Demostración. (a) Las ecuaciones normales determinan un sistema compatible si y sólo si $A^tb \in \mathcal{R}(A^tA)$, hecho que sigue inmediatamente de la igualdad $\mathcal{R}(A^tA) = \mathcal{R}(A^t)$ y por ser $A^tb \in \mathcal{R}(A^t)$.

(b) Si $\text{rg}(A) = n$ entonces $A^tA \in \mathbb{R}^{n\times n}$ es invertible pues

$$\text{rg}(A^tA) = \text{rg}(A) = n.$$

Así, la única solución de $A^tAv = A^tb$ es

$$v = (A^tA)^{-1}A^tb = A_Ib,$$

lo que termina la demostración. □

Ejemplo 11.7. *Encontrar las soluciones de mínimos cuadrados de los sistemas $Ax = b$ siendo:*

(a) $A = \begin{bmatrix} 1 & -1 \\ 1 & 1 \\ 0 & 0 \end{bmatrix}$ *y* $b = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$.

(b) $A = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$ *y* $b = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$.

$\triangleleft$ Se aplicará la Proposición 11.4.

(a) Es fácil ver que $b \notin \mathcal{R}(A)$. Como $\operatorname{rg}(A) = 2$, la solución de mínimos cuadrados será única. Puesto que

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

y

$$A^t b = \begin{bmatrix} 1 & 1 & 0 \\ -1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix},$$

la solución de mínimos cuadrados coincide con la única solución de las ecaciones normales $A^t A v = A^t b$, que es

$$v = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \end{bmatrix}.$$

(b) De nuevo, $b \notin \mathcal{R}(A)$. En este caso, $\operatorname{rg}(A) = 2 \neq 3$. Puesto que

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

y

$$A^t b = \begin{bmatrix} 1 & 1 & 0 \\ -1 & 1 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix},$$

las soluciones de mínimos cuadrados coinciden con las soluciones de las ecuaciones normales $A^tAv = A^tb$ (en este caso, no se obtiene solución única) y su conjunto solución viene dado por los vectores de la forma

$$v = \begin{bmatrix} \frac{1}{2} - \frac{1}{2}\alpha \\ \frac{1}{2} - \frac{1}{2}\alpha \\ \alpha \end{bmatrix}, \qquad \alpha \in \mathbb{R}.$$

$\triangleright$

Solución de norma mínima: Con la intención de buscar unicidad en las soluciones de mínimos cuadrados (que, como se ha visto puede tener más de una solución) el objetivo final es buscar, de entre todas ellas, una solución de **norma mínima**, y por tanto una **solución óptima** del problema de mínimos cuadrados. Es decir, se analizará el problema de buscar los vectores

$$v \in S_{MC} \qquad \text{tales que} \qquad \|v\|_2 \text{ sea mínima.}$$

Ejemplo 11.8. *Encontrar la solución de mínimos cuadrados de norma mínima del apartado (b) del Ejemplo 11.8.*

$\triangleleft$ De todas las soluciones del apartado (b) del Ejemplo 11.8, es posible determinar la(s) de norma mínima mediante métodos de Cálculo Diferencial como sigue. Se debe encontrar el mínimo de la función

$$\left\| \begin{bmatrix} \frac{1}{2} - \frac{1}{2}\alpha \\ \frac{1}{2} - \frac{1}{2}\alpha \\ \alpha \end{bmatrix} \right\|_2^2 = \left[\frac{1}{2} - \frac{1}{2}\alpha\right]^2 + \left[\frac{1}{2} - \frac{1}{2}\alpha\right]^2 + \alpha^2 = \frac{1}{2}(1-\alpha)^2 + \alpha^2 := f(\alpha),$$

que derivando queda $f'(\alpha) = 3\alpha - 1$, e igualando a 0, $\alpha = \frac{1}{3}$. Es evidente que se debe tratar de un valor mínimo pues la gráfica de la función a optimizar es una parábola con coeficiente principal positivo. Luego, la solución en mínimos cuadrados de norma mínima del sistema

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

es

$$\begin{bmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{bmatrix}.$$

$\triangleright$

¿Qué ocurre si la función a optimizar tuviese más variables? Está claro que el método aplicado en el último ejemplo es limitado, a menos que se tengan los conocimientos suficientes de Cálculo Diferencial para estudiar el problema de optimización de una función de varias variables.

A continuación, se presenta una condición suficiente que garantiza la existencia y unicidad de solución de norma mínima a través de un método algebraico.

Proposición 11.5. Sean $A \in \mathbb{R}^{m\times n}$ y $b \in \mathbb{R}^{m\times 1}$. Si $\operatorname{rg}(A) = m$ entonces el sistema $Ax = b$ es compatible y la única solución de norma mínima es[a]

$$v = A^t(AA^t)^{-1}b = A_D b.$$

[a] La notación A_D se introduce para hacer hincapié en que la matriz $A_D := A^t(AA^t)^{-1}$ es una inversa a derecha de A. Observar que su existencia queda garantizada si $\operatorname{rg}(A) = m$.

Demostración. En primer lugar se probará que el sistema $Ax = b$ es compatible.

De la hipótesis $\operatorname{rg}(A) = m$, la matriz $AA^t \in \mathbb{R}^{m\times m}$ es invertible (puesto que $\operatorname{rg}(AA^t) = \operatorname{rg}(A) = m$), con lo cual el sistema $Ax = b$ es compatible pues

$$v := A_D b = A^t(AA^t)^{-1}b$$

satisface

$$Av = A[A^t(AA^t)^{-1}b] = (AA^t)(AA^t)^{-1}b = I_n b = b.$$

En segundo lugar se probará que la solución de norma mínima es $v = A_D b$.

En efecto, si $s \in \mathbb{R}^{n\times 1}$ es otra solución del sistema $Ax = b$ (i.e., $s \neq v$) entonces $As = b = Av$, de donde $A(s - v) = As - Av = 0$, y así

$$s - v \in \mathcal{N}(A).$$

Luego, existe $n \in \mathcal{N}(A)$ tal que $s - v = n$. Además,

$$v = A^t(AA^t)^{-1}b = A^t[(AA^t)^{-1}b] \in \mathcal{R}(A^t) = [\mathcal{N}(A)]^\perp.$$

Luego, $s = n + v \in \mathcal{N}(A) \oplus^\perp [\mathcal{N}(A)]^\perp$. Aplicando el Teorema de Pitágoras se tiene que

$$\|s\|_2^2 = \|n + v\|_2^2 = \|n\|_2^2 + \|v\|_2^2 \geq \|v\|_2^2. \tag{11.11}$$

Por tanto, $\|v\|_2 \leq \|s\|_2$, con lo cual v tiene norma mínima de entre todos los vectores solución del sistema $Ax = b$.

Por último, $\|v\|_2 < \|s\|_2$ pues si fuese $\|v\|_2 = \|s\|_2$, de (11.11) se tendría que $\|n\|_2 = 0$. Es decir, $n = 0$, o equivalentemente $s = v$, lo que es una contradición (pues $s \neq v$). Este hecho prueba que v es el único vector solución de $Ax = b$ de norma mínima. □

Solución de mínimos cuadrados de norma mínima: Una forma de establecer el resultado general es mediante una factorización de la matriz de coeficientes del sistema. Dicha factorización se deduce fácilmente de los resultados conocidos de equivalencia de matrices.

Factorización de rango completo: Si $O \neq A \in \mathbb{R}^{m\times n}$ con $r := \operatorname{rg}(A)$, existen matrices invertibles $\widetilde{Q} \in \mathbb{R}^{m\times m}$ y $\widetilde{P} \in \mathbb{R}^{n\times n}$ tales que

$$A = \widetilde{Q}\begin{bmatrix} I_r & O \\ O & O \end{bmatrix}\widetilde{P} = \begin{bmatrix} Q & R \end{bmatrix}\begin{bmatrix} I_r & O \\ O & O \end{bmatrix}\begin{bmatrix} P \\ S \end{bmatrix} = QP, \tag{11.12}$$

donde las matrices $Q \in \mathbb{R}^{m\times r}$ y $P \in \mathbb{R}^{r\times n}$ satisfacen

$$\operatorname{rg}(Q) = \operatorname{rg}(P) = \operatorname{rg}(A) = r.$$

Como $\mathrm{rg}(Q) = r$, es decir, coincide con el número de columnas de Q, se dice que Q es una matriz de **rango completo por columnas**. Análogamente, al ser $\mathrm{rg}(P) = r$, es decir, el número de filas de la matriz P, se dice que P tiene **rango completo por filas**. Una factorización de una matriz $A \in \mathbb{R}^{m\times n}$, de rango $r > 0$, del tipo

$$A = QP, \qquad \text{con } \mathrm{rg}(Q) = \mathrm{rg}(P) = \mathrm{rg}(A) = r$$

se conoce como **factorización de rango completo** de A.

Teorema 11.5 (Solución del problema de mínimos cuadrados de norma mínima)**.** Sean $O \neq A \in \mathbb{R}^{m\times n}$ y $b \in \mathbb{R}^{m\times 1}$. Si $A = QP$ es una factorización de rango completo de A, entonces la (única) solución de mínimos cuadrados de norma mínima de $Ax = b$ es

$$x_* = P_D Q_I b,$$

donde $P_D = P^t(PP^t)^{-1}$ y $Q_I = (Q^tQ)^{-1}Q^t$.

Demostración. Tras encontrar una factorización de rango completo

$$A = QP$$

de A como en (11.12), el sistema $Ax = b$ se puede escribir como

$$Q(Px) = Ax = b,$$

que se puede resolver mediante dos sistemas de ecuaciones lineales por separado[4]

$$Qy = b, \qquad Px = y.$$

Como $\mathrm{rg}(Q) = r$ y $Q \in \mathbb{R}^{m\times r}$, por el apartado (b) de la Proposición 11.4, la única solución de mínimos cuadrados de $Qy = b$ es $y = Q_I b$.

[4]Se observa que primero se debe resolver el sistema $Qy = b$ y, posteriormente, $Px = y$.

Ahora, por la Proposición 11.5, el sistema $Px = Q_I b$, con $P \in \mathbb{R}^{r\times n}$ y $\mathrm{rg}(P) = r$, es compatible y admite una única solución de norma mínima que viene dada por

$$x_* := P_D Q_I b. \tag{11.13}$$

Falta comprobar que x_* es la solución deseada del problema original $Ax = b$.

- x_* es una solución de mínimos cuadrados de $Ax = b$. En efecto, teniendo en cuenta las expresiones de A y x_* y que $PP_D = I_r$ se tiene que

$$\begin{aligned} \|Ax_* - b\|_2 &= \|(QP)(P_D Q_I b) - b\|_2 \\ &= \|Q(PP_D)Q_I b - b\|_2 \\ &= \|Q(Q_I b) - b\|_2 \\ &\leq \|Qz - b\|_2, \quad \text{para todo } z \in \mathbb{R}^{r\times 1}, \end{aligned}$$

  donde el último paso se deduce del hecho que $Q_I b$ es (la) solución de mínimos cuadrados de $Qy = b$. Dada la arbitrariedad de z, particularizando a $z = Pu$, con $u \in \mathbb{R}^{n\times 1}$, se tiene

$$\begin{aligned} \|Ax_* - b\|_2 &\leq \|Qz - b\|_2, \quad \text{para todo } z \in \mathbb{R}^{r\times 1} \\ &\leq \|QPu - b\|_2, \quad \text{para todo } u \in \mathbb{R}^{n\times 1} \\ &= \|Au - b\|_2, \quad \text{para todo } u \in \mathbb{R}^{n\times 1}. \end{aligned}$$

- x_* es una solución de norma mínima de $Ax = b$.

  Para ello, previamente se probará, por doble inclusión, el siguiente resultado:

> Si S_{MC} denota el conjunto de soluciones de mínimos cuadrados de $Ax = b$, entonces
>
> $$S_{MC} = \{z \in \mathbb{R}^{n\times 1} : Pz = Q_I b\}. \tag{11.14}$$

En efecto,

($\subseteq$) Si $z \in S_{MC}$ entonces, por la Proposición 11.3, $A^t Az = A^t b$. Luego,

$$P^t Q^t QPz = P^t Q^t b.$$

Tomando traspuestas en $PP_D = I_r$ se tiene $P_D^t P^t = I_r$ con lo cual, de

$$P_D^t(P^t Q^t QPz) = P_D^t(P^t Q^t b)$$

se obtiene $Q^t Q(Pz) = Q^t b$, es decir, Pz es una solución del sistema $Q^t Qw = Q^t b$, y por la Proposición 11.3 se llega a que Pz es una solución de mínimos cuadrados del sistema $Qy = b$. Como se ha probado que el sistema $Qy = b$ posee una única solución de mínimos cuadrados (que es $Q_I b$), debe ser $Pz = Q_I b$.

($\supseteq$) Si $z \in \mathbb{R}^{n\times 1}$ satisface $Pz = Q_I b$ entonces

$$Az = Q(Pz) = Q(Q_I b).$$

Premultiplicando por A^t y recordando que $PP_D = I_r$ y la expresión obtenida de la solución x_* de mínimos cuadrados de $Ax = b$, se tiene que

$$A^t Az = A^t QQ_I b = A^t Q(PP_D)Q_I b = A^t Ax_* = A^t b.$$

Luego, z es solución de mínimos cuadrados de $Ax = b$, es decir, $z \in S_{MC}$.

De este modo, se ha probado la igualdad (11.14). $\square$

Ahora, continuando con la demostración del Teorema, como se sabe que $P_D Q_I b$ es la solución de norma mínima de $Px = Q_I b$, por (11.14) se tiene que

$$\begin{aligned} \|x_*\|_2 &= \|P_D Q_I b\|_2 \\ &\leq \|z\|_2, \text{ para todo } z \text{ tal que } Px = Q_I b \\ &= \|z\|_2, \text{ para todo } z \in S_{MC}. \end{aligned}$$

Se ha probado que

$$x_* \in \{x \in \mathbb{R}^{n\times 1} : x \in S_{MC}, \|x\|_2 \text{ mínima}\}.$$

Ahora se probará que:

$$\{x \in \mathbb{R}^{n\times 1} : x \in S_{MC}, \|x\|_2 \text{ mínima}\} = \{x_*\}.$$

- Unicidad de x_*. En efecto, se ha probado que $A^tAx_* = A^tb$ y $\|x_*\|_2$ es mínima de entre todas las soluciones de mínimos cuadrados de $Ax = b$. Si se supone que existe un vector $x_0 \in \mathbb{R}^{n\times 1}$ en estas mismas condiciones, es decir, $A^tAx_0 = A^tb$ y con $\|x_0\|_2$ mínima, se debe probar que $x_0 = x_*$.

  Como $x_0 \in S_{MC}$, de (11.14) se tiene que

  $$x_0 \text{ es una solución de la ecuación } Px = Q_Ib. \tag{11.15}$$

  Utilizando la minimalidad de x_* se tiene que $\|x_*\|_2 \leq \|x_0\|_2$, por ser x_0 otra solución de mínimos cuadrados de $Ax = b$. Intercambiando los papeles de x_* con x_0 se tiene que $\|x_0\|_2 \leq \|x_*\|_2$. En definitiva,

  $$\|x_0\|_2 = \|x_*\|_2. \tag{11.16}$$

  En (11.13) se ha probado que la ecuación $Px = Q_Ib$ tiene una única solución de norma mínima, que es x_* (cuya expresión es P_DQ_Ib).

  De la unicidad en esta última afirmación, (11.15) y (11.16) se concluye que $x_0 = x_*$.

De este modo, el Teorema queda probado. □

Ejemplo 11.9. *Resolver los dos sistemas lineales del Ejemplo 11.8 utilizando el Teorema 11.5.*

◁ En primer lugar se calcula el rango de la matriz A.

(a) Como $\mathrm{rg}(A) = 2$, existen matrices invertibles $Q \in \mathbb{R}^{3\times 2}$ y $P \in \mathbb{R}^{2\times 2}$ tales que $A = QP$. En este caso se observa que no es necesario calcular dichas matrices mediante procedimiento alguno puesto que A es de rango completo por columnas, con lo que una factorización de rango completo inmediata de A se obtiene tomando $Q = A$ y $P = I_2$. De este modo, es claro que $P_D = I_2$ y

$$Q_I = (Q^tQ)^{-1}Q^t = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 1 & 0 \\ -1 & 1 & 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ -\frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}.$$

Del Teorema 11.5, la solución de mínimos cuadrados de norma mínima del sistema $Ax = b$ es

$$x_* = P_DQ_Ib = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ -\frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \end{bmatrix}.$$

(b) En este caso, $\mathrm{rg}(A) = 2$. Realizando operaciones elementales por filas y por columnas se puede obtener que

$$\left[\begin{array}{ccc|ccc} 1 & -1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ \hline 1 & 0 & 0 & & & \\ 0 & 1 & 0 & & & \\ 0 & 0 & 1 & & & \end{array}\right] \sim \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 1 & 0 & -\frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ \hline 1 & 0 & -\frac{1}{2} & & & \\ 0 & 1 & -\frac{1}{2} & & & \\ 0 & 0 & 1 & & & \end{array}\right].$$

Siguiendo la notación previa, $A = Q \left[\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \hline 0 & 0 & 0 \end{array}\right] P$ y se ha obtenido que

$$Q^{-1} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ -\frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad \Rightarrow \qquad Q = \left[\begin{array}{cc|c} 1 & -1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right]$$

y

$$P^{-1} = \begin{bmatrix} 1 & 0 & -\frac{1}{2} \\ 0 & 1 & -\frac{1}{2} \\ 0 & 0 & 1 \end{bmatrix} \qquad \Rightarrow \qquad P = \left[\begin{array}{ccc} 1 & 0 & \frac{1}{2} \\ 0 & 1 & \frac{1}{2} \\ \hline 0 & 0 & 1 \end{array}\right].$$

Entonces

$$A = QP = \begin{bmatrix} 1 & -1 \\ 1 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & \frac{1}{2} \\ 0 & 1 & \frac{1}{2} \end{bmatrix}.$$

Ahora, realizando los cálculos intermedios se tiene que

$$P_D = P^t(PP^t)^{-1} = \frac{1}{6} \begin{bmatrix} 5 & -1 \\ -1 & 5 \\ 2 & 2 \end{bmatrix}$$

y

$$Q_I = (Q^tQ)^{-1}Q^t = \frac{1}{2} \begin{bmatrix} 1 & 1 & 0 \\ -1 & 1 & 0 \end{bmatrix}.$$

Finalmente, aplicando el Teorema 11.5 se tiene

$$x_* = P_D Q_I b = \frac{1}{3} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix},$$

que es la solución de mínimos cuadrados de norma mínima de $Ax = b$.

$\triangleright$

Pseudoinversa o inversa de Moore-Penrose.: Antes de abordar el tema es interesante indicar que Roger Penrose es físico y matemático y nació en Reino Unido en 1931. En la actualidad es profesor emérito de Matemáticas de la Universidad de Oxford. Recibió (de forma compartida) el Premio Nobel en 2020 por concluir que la formación de agujeros negros es una predicción robusta de la Teoría General de la Relatividad.

Figura 11.3: Roger Penrose

Con la intención de escribir la solución obtenida de mínimos cuadrados de norma mínima en términos de los datos A y b (y no a partir de P_D, Q_I y b como en el Teorema 11.5) se introduce la siguiente definición.

Definición 11.5. Sea $A \in \mathbb{R}^{m\times n}$. Una matriz $X \in \mathbb{R}^{n\times m}$ se llama **inversa de Moore-Penrose** si satisface las **ecuaciones de Penrose**

$$AXA = A, \qquad XAX = X, \qquad (AX)^t = AX, \qquad (XA)^t = XA.$$

Lema 11.6 (Existencia y unicidad de la inversa de Moore-Penrose)**.** Si QP es una factorización de rango completo de una matriz no nula $A \in \mathbb{R}^{m\times n}$ entonces la única matriz $X \in \mathbb{R}^{n\times m}$ que satisface las ecuaciones de Penrose es

$$X = P_D Q_I,$$

siendo $P_D = P^t(PP^t)^{-1}$ y $Q_I = (Q^tQ)^{-1}Q^t$. Dicha matriz X se denota $A^\dagger$.

Demostración. Primero se probará que si dicha matriz existe entonces es única. Luego se comprobará que la matriz indicada satisface las ecuaciones de Penrose. De este modo, se podrá afirmar que $A^\dagger = P_D Q_I$ siempre existe y es, efectivamente, la inversa de Moore-Penrose de A.

Unicidad: Si X_1 y X_2 satisfacen las ecuaciones de Penrose entonces se tiene que

$$AX_1A = A, \quad X_1AX_1 = X_1, \quad (AX_1)^t = AX_1, \quad (X_1A)^t = X_1A$$

y

$$AX_2A = A, \quad X_2AX_2 = X_2, \quad (AX_2)^t = AX_2, \quad (X_2A)^t = X_2A.$$

Luego,

$$\begin{aligned}
X_1 &= X_1AX_1 \\
&= X_1AX_2AX_1 \\
&= (X_1A)^t(X_2A)^tX_1 \\
&= A^tX_1^tA^tX_2^tX_1 \\
&= (AX_1A)^tX_2^tX_1 \\
&= A^tX_2^tX_1 \\
&= (X_2A)^tX_1 \\
&= X_2AX_1 \\
&= X_2(AX_1)^t \\
&= X_2X_1^tA^t \\
&= X_2X_1^t(AX_2A)^t \\
&= X_2X_1^tA^tX_2^tA^t \\
&= X_2(AX_1)^t(AX_2)^t \\
&= X_2AX_1AX_2 \\
&= X_2AX_2 \\
&= X_2,
\end{aligned}$$

con lo cual $X_1 = X_2$.

Existencia: Como $PP_D = I_r = Q_IQ$,

- $AXA = QPP_DQ_IQP = QP = A$.
- $XAX = P_DQ_IQPP_DQ_I = P_DQ_I = X$.
- $(AX)^t = AX$ pues $P_D = P^t(PP^t)^{-1} \Rightarrow P_DP = P^t(PP^t)^{-1}P$ es simétrica.
- $(XA)^t = XA$ pues $Q_I = (Q^tQ)^{-1}Q^t \Rightarrow QQ_I = Q(Q^tQ)^{-1}Q^t$ es simétrica.

$\square$

Finalmente, el Teorema de Penrose expresa la solución óptima en términos de A y b.

Teorema 11.6 (Teorema de Penrose)**.** Sean $A \in \mathbb{R}^{m\times n}$ y $b \in \mathbb{R}^{m\times 1}$. La (única) solución de **mínimos cuadrados** de **norma mínima** de $Ax = b$ es

$$x_* = A^\dagger b.$$

Observación 11.5. Es verdad que al realizar diferentes operaciones elementales es posible encontrar distintas factorizaciones de rango completo para una matriz (no nula) A, es decir, dicha factorización no es única.
Podría plantearse la cuestión siguiente: En el Teorema 11.5, ¿la solución x_* depende de la factorización utilizada para A? La respuesta es claramente: no. Se deduce de la unicidad probada para la inversa de Moore-Penrose y, por tanto, de acuerdo al Teorema de Penrose, dicha solución es única y viene dada en términos de A (e independientemente de Q y P). Es decir, al realizar los cálculos con diferentes factorizaciones de rango completo de A

$$A = Q\,P \qquad \text{y} \qquad A = \widetilde{Q}\,\widetilde{P},$$

se obtendrá la (única) inversa de Moore-Penrose

$$A^\dagger = P_DQ_I = \widetilde{P}_D\widetilde{Q}_I$$

y, por tanto, una única solución

$$x_* = P_D Q_I b = \widetilde{P}_D \widetilde{Q}_I b = A^{\dagger} b,$$

donde las inversas laterales se calculan con las fórmulas indicadas en el Teorema 11.5.

11.8. Resumen de resultados importantes

Simplificando mucho, es posible decir que a lo largo de este libro se han resuelto completamente tres problemas del ámbito del Álgebra Lineal y la Geometría:

- La caracterización de la equivalencia de matrices desde diferentes puntos de vista.
- La caracterización de la no singularidad (invertibilidad) de matrices desde diferentes puntos de vista.
- La resolución de sistemas de ecuaciones lineales y su relación con los subespacios fundamentales asociados a la matriz de coeficientes y al vector de términos independientes.

Para finalizar, se establecen relaciones destacables entre los resultados estudiados previamente y se hace un resumen de los resultados presentados a lo largo del libro que permitirá tener una visión de conjunto de los mismos.

Equivalencia de matrices:

Teorema 11.7 (Equivalencia de matrices por filas)**.** Sean $A, B \in \mathbb{K}^{m\times n}$. Son equivalentes:

(i) $A \sim_f B$.

(ii) Existen matrices elementales por filas $E_1, E_2, \ldots, E_s \in \mathbb{K}^{m\times m}$ tales que $B = E_1E_2 \ldots E_sA$.

(iii) Existe una matriz invertible $Q \in \mathbb{K}^{m\times m}$ tal que $B = QA$.

(iv) $R_A = R_B$, con R_C la matriz escalonada reducida por filas de C.

(v) $\mathcal{F}(A) = \mathcal{F}(B)$ ($\Leftrightarrow \mathcal{C}(A^t) = \mathcal{C}(B^t) \Leftrightarrow \mathcal{R}(A^t) = \mathcal{R}(B^t)$).

(vi) $\mathcal{N}(A) = \mathcal{N}(B)$.

(vii) $Ax = 0$ y $Bx = 0$ tienen el mismo conjunto solución.

Demostración. Sigue inmediatamente del Teorema 6.1, del Teorema 11.3 y del Lema 11.5. □

Teorema 11.8 (Equivalencia de matrices por columnas)**.** Sean $A, B \in \mathbb{K}^{m\times n}$. Son equivalentes:

(i) $A \sim_c B$.

(ii) Existen matrices elementales por columnas $F_1, F_2, \ldots, F_\ell \in \mathbb{K}^{n\times n}$ tales que $B = AF_1F_2 \ldots F_\ell$.

(iii) Existe una matriz invertible $P \in \mathbb{K}^{n\times n}$ tal que $B = AP$.

(iv) $R_{A^t} = R_{B^t}$, con R_C la matriz escalonada reducida por filas de C.

(v) $\mathcal{C}(A) = \mathcal{C}(B)$ ($\Leftrightarrow \mathcal{R}(A) = \mathcal{R}(B) \Leftrightarrow \mathcal{F}(A^t) = \mathcal{F}(B^t)$).

(vi) $\mathcal{N}(A^t) = \mathcal{N}(B^t)$.

(vii) $A^tx = 0$ y $B^tx = 0$ tienen el mismo conjunto solución.

Demostración. Sigue inmediatamente del Teorema 6.2, del Teorema 11.3 y del Lema 11.5. □

Es posible dar una nueva caracterización de matrices equivalentes (y, en particular, por filas y por columnas) al interpretar las matrices a partir de aplicaciones lineales definidas entre espacios vectoriales, pero este tema escapa el alcance de este libro.

Matrices invertibles:

A continuación se presenta un resultado que recoge todas las caracterizaciones que indican cuándo una matriz es invertible y se deducen de los temas estudiados previamente.

Teorema 11.9 (Teorema fundamental de matrices invertibles)**.** Sea $A \in \mathbb{K}^{n\times n}$. Las siguientes condiciones son equivalentes:

(a) A es invertible.

(b) A admite una inversa a izquierda.

(c) A admite una inversa a derecha.

(d) El sistema homogéneo $Ax = 0$ sólo admite la solución trivial $x = 0$.

(e) Para cada $b \in \mathbb{K}^{n\times 1}$, el sistema $Ax = b$ tiene solución única.

(f) $R_A = I_n$, con R_A la forma escalonada reducida por filas de A.

(g) $A \sim_f I_n$.

(h) $\operatorname{rg}(A) = n$.

(i) A se escribe como producto de matrices elementales por filas de $\mathbb{K}^{n\times n}$.

(j) $\det(A) \neq 0$ (es decir, A es regular).

(k) A es una matriz de cambio de base.

(l) $\mathcal{F}(A) = \mathbb{K}^n$.

(m) las filas de A determinan un conjunto linealmente independiente en $\mathbb{K}^n$.

(n) las filas de A determinan un sistema generador en $\mathbb{K}^n$.

(ñ) las filas de A determinan una base de $\mathbb{K}^n$.

(o) $\dim_{\mathbb{K}}(\mathcal{F}(A)) = n$.

(p) $\dim_{\mathbb{K}}(\mathcal{N}(A)) = 0$.

(q) $\mathcal{C}(A) = \mathbb{K}^n$.

(r) las columnas de A determinan un conjunto linealmente independiente en $\mathbb{K}^n$.

(s) las columnas de A determinan un sistema generador en $\mathbb{K}^n$.

(t) las columnas de A determinan una base de $\mathbb{K}^n$.

(u) $\dim_{\mathbb{K}}(\mathcal{C}(A)) = n$.

(v) $\dim_{\mathbb{K}}(\mathcal{N}(A^t)) = 0$.

Si $\mathbb{K} = \mathbb{R}$, también son equivalentes:

(w) A^tA es invertible.

(x) AA^t es invertible.

(y) $[\mathcal{N}(A)]^\perp = \mathbb{R}^n$.

(z) $[\mathcal{F}(A)]^\perp = \{0\}$.

Demostración. Se pueden demostrar las equivalencias a partir del Teorema 5.1, el Teorema 7.1, el Corolario 8.3, el Corolario 9.3, el Lema 11.4, el Teorema

11.2, el Lema 11.5 y la Proposición 11.2.

Los detalles se proponen como ejercicio. □

Sistemas de ecuaciones lineales:

A la hora de resolver un sistema de ecuaciones lineales del tipo

$$Ax = b, \qquad \text{con} \quad A \in \mathbb{R}^{m\times n} \quad \text{y} \quad b \in \mathbb{R}^{m\times 1},$$

aparecen involucrados los espacios euclídeos $\mathbb{R}^n$ y $\mathbb{R}^m$. Los subespacios fundamentales asociados a la matriz A permiten analizar, en el caso de:

- sistemas compatibles: la existencia y unicidad de sus soluciones,
- sistemas incompatibles: las soluciones aproximadas de mínimos cuadrados.

Sistemas compatibles:

Dado que el espacio fila $\mathcal{F}(A)$ y el espacio nulo $\mathcal{N}(A)$ son subespacios vectoriales complementarios (y ortogonales) de $\mathbb{R}^n$ y el espacio columna $\mathcal{C}(A)$ y el espacio nulo izquierdo $\mathcal{N}(A^t)$ son subespacios vectoriales complementarios (y ortogonales) de $\mathbb{R}^m$, a partir del Teorema fundamental del Álgebra Lineal, los espacios se pueden distribuir como los diagramas de las Figuras 11.1 y 11.2, gráficas que se suelen atribuir a G. Strang.

Se trata de analizar el conjunto solución S del sistema de ecuaciones lineales $Ax = b$ para una matriz $A \in \mathbb{R}^{m\times n}$ y un vector (columna) $b \in \mathbb{R}^m$ dados.

Dado $b \in \mathbb{R}^m = \mathcal{C}(A) \oplus^{\perp} \mathcal{N}(A^t)$, existen vectores unívocamente determinados $b' \in \mathcal{C}(A)$ y $w \in \mathcal{N}(A^t)$ tales que

$$b = b' + w.$$

Si existe $x \in \mathbb{R}^n = \mathcal{C}(A^t) \oplus^{\perp} \mathcal{N}(A)$ tal que $Ax = b$, sería posible escribir

$$x = y + z,$$

donde $y \in \mathcal{C}(A^t)$ y $z \in \mathcal{N}(A)$ son vectores unívocamente determinados. Luego,

$$b' + w = b = Ax = A(y + z) = Ay + Az = Ay + 0 = Ay,$$

de donde, por ser $\mathcal{C}(A)$ un subespacio vectorial,

$$w = Ay - b' \in \mathcal{C}(A) \cap \mathcal{N}(A^t) = \{0\},$$

es decir, $w = 0$. Luego, la única posibilidad para que exista solución es que $w = 0$, i.e.,

$$b = b' \in \mathcal{C}(A).$$

Es claro que el recíproco de este resultado también se cumple pues si

$$b \in \mathcal{C}(A) = \mathcal{R}(A) = \{Ax : x \in \mathbb{R}^{n\times 1}\},$$

existe $x \in \mathbb{R}^{n\times 1}$ tal que $b = Ax$.

De este modo, se ha probado[5] el siguiente resultado.

Proposición 11.6 (La existencia de la solución depende del espacio columna)**.** Sean $A \in \mathbb{R}^{m\times n}$ y $b \in \mathbb{R}^{m\times 1}$. Las siguientes condiciones son equivalentes:

(a) Existe $x \in \mathbb{R}^{n\times 1}$ tal que $Ax = b$.

(b) $b \in \mathcal{C}(A)$.

Observación 11.6. El resultado de la Proposición 11.6 se obtiene de forma más directa de la Proposición 2.8 particionando A según sus columnas. Al realizar el producto Ax es posible escribir al vector b como combinación lineal de las columnas de A. La demostración previa, sin embargo, puede tener un alcance mayor al no utilizar coordenadas.

[5] Si bien este hecho fue probado al comenzar el apartado “Planteamiento del problema” de la página 658, en aquella ocasión se utilizó un procedimiento puramente matricial y en esta el enfoque es más bien funcional, a partir de la aplicación $x \mapsto Ax$, utilizando la descomposición de los espacios como suma directa.

De esta manera, una primera aproximación al problema permite utilizar:

- el espacio columna para saber si el sistema $Ax = b$ tiene solución. La tendrá únicamente en caso que $b \in \mathcal{C}(A)$.

- el espacio nulo ayuda a saber cómo será el conjunto solución. Por ejemplo, si $A \in \mathbb{R}^{3\times 3}$ con $b \in \mathcal{C}(A)$ y $\dim(\mathcal{N}(A)) = 1$, el conjunto solución se formará sumando una solución particular de $Ax = b$ más un subespacio de dimensión 1; si fuese $\dim(\mathcal{N}(A)) = 2$, el conjunto solución se formará sumando una solución particular de $Ax = b$ más un subespacio de dimensión 2, etc.

Más específicamente,

- si $b \notin \mathcal{C}(A)$, no existe $x \in \mathbb{R}^n$ tal que $Ax = b$. Luego, el sistema es incompatible y $S = \emptyset$. En este caso, se encontró la solución del problema de mínimos cuadrados de norma mínima siendo la solución $x = A^\dagger b$.

- si $b \in \mathcal{C}(A)$, existe $x \in \mathbb{R}^n$ tal que $Ax = b$. Razonando como antes, x se escribe de forma única como $x = y + z$ con $y \in \mathcal{C}(A^t)$ y $z \in \mathcal{N}(A)$. Luego,

  $$b = Ax = A(y + z) = Ay + Az = Ay + 0 = Ay,$$

  es decir, $y \in \mathcal{C}(A^t)$ es una solución particular de $Ax = b$.

  Es interesante observar que existe un único $y \in \mathcal{C}(A^t)$ tal que $Ay = b$. En efecto, si existiese otro $y' \in \mathcal{C}(A^t)$ tal que $Ay' = b$ entonces $Ay = Ay'$ de donde $A(y - y') = 0$, es decir, $y - y' \in \mathcal{N}(A)$, pero además, $y - y' \in \mathcal{C}(A^t)$, por ser $\mathcal{C}(A^t)$ un subespacio. Luego, $y - y' \in \mathcal{N}(A) \cap \mathcal{C}(A^t) = \{0\}$, es decir, $y = y'$.

  Ahora, por el Teorema 9.4, el conjunto solución viene dado por

  $$S = \{x \in \mathbb{R}^n : Ax = b\} = y + \{u \in \mathbb{R}^n : Au = 0\} = y + \mathcal{N}(A). \quad (11.17)$$

  La Figura 11.4 proporciona una idea gráfica de lo que ocurre (los vectores y y z son ortogonales por pertenecer a subespacios ortogonales entre sí).

Además,

- Si $\underline{\dim(\mathcal{N}(A)) = 0}$ entonces $Ax = b$ es un sistema compatible determinado y $S = \{y\}$. En el caso en que $m = n$ se tiene que $m = n = \mathrm{rg}(A)$ con lo cual $y = A^{-1}b$, es decir, $S = \{A^{-1}b\}$.
- Si $\underline{\dim(\mathcal{N}(A)) > 0}$ entonces $Ax = b$ es un sistema compatible indeterminado y (11.17) proporciona el conjunto solución.

Se ha probado el siguiente resultado.

Proposición 11.7 (La unicidad de la solución depende del espacio nulo)**.** Sean $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^{m\times 1}$ y $Ax = b$ un sistema de ecuaciones lineales compatible. Entonces el sistema $Ax = b$ es compatible determinado o indeterminado según sea $\dim(\mathcal{N}(A)) = 0$ o $\dim(\mathcal{N}(A)) > 0$, respectivamente.

Dos representaciones para el conjunto solución S:

Como se mostró en el Teorema 9.4, para obtener el conjunto solución (11.17) no es necesario elegir el único vector $y \in \mathcal{C}(A^t)$ tal que $Ay = b$, sino que alcanza con encontrar una solución particular cualquiera $p \in \mathbb{R}^n$ tal que $Ap = b$ y así el conjunto solución queda:

$$S = \{x \in \mathbb{R}^n : Ax = b\} = p + \mathcal{N}(A).$$

En este caso, fijado p tal que $Ap = b$, una solución arbitraria $x \in S$ se podrá representar de dos formas como

$$x = y + z = p + (x - p),$$

donde, para cada $x \in S$, los vectores $z, x - p \in \mathcal{N}(A)$ son únicos (pues si un vector q satisface $x = p + q$, debe ser $q = x - p$ y, además, $A(x - p) = Ax - Ap = b - b = 0$, con lo cual $x - p \in \mathcal{N}(A)$).

En definitiva, en estas condiciones,

$$\boxed{S = y + \mathcal{N}(A) = p + \mathcal{N}(A)}.$$

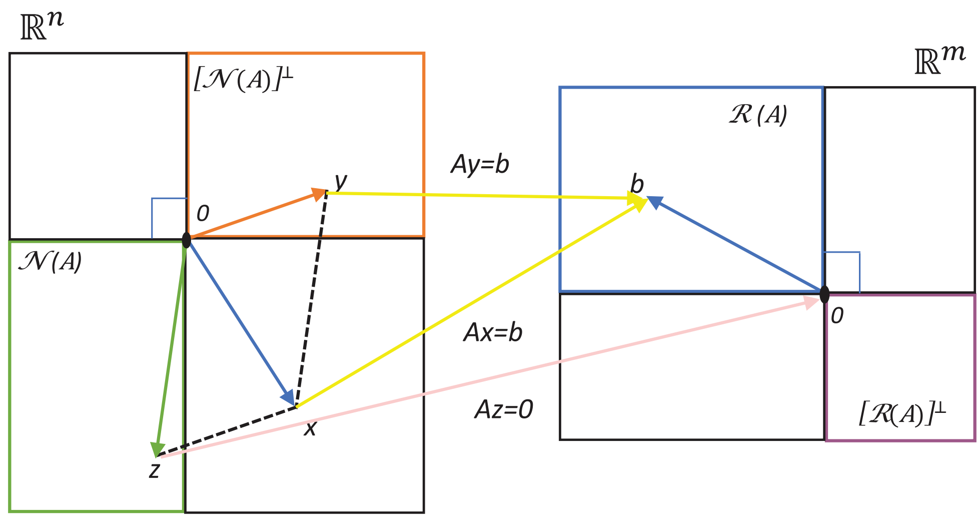


Figura 11.4: Conjunto solución $S = \{x : Ax = b\} = y + \{z : Az = 0\}$ de un sistema compatible $Ax = b$.

Sistemas incompatibles:

Dado que los subespacios $\mathcal{R}(A^t)$ y $\mathcal{R}(A)$ tienen la misma dimensión, se trata de espacios vectoriales isomorfos sobre $\mathbb{R}$, es decir,

$$\mathcal{R}(A^t) \cong \mathcal{R}(A).$$

Luego, si se descartan los dos espacios nulos $\mathcal{N}(A)$ y $\mathcal{N}(A^t)$ (es decir, si se analiza sólo el efecto de A sobre los vectores de $\mathcal{R}(A^t)$ en $\mathcal{R}(A)$, que podría representarse como $A|_{\mathcal{R}(A^t)} : \mathcal{R}(A^t) \to \mathcal{R}(A)$), se puede afirmar que hay una matriz invertible de tamaño $r \times r$ "escondida" dentro de A, lo que corresponde al concepto de **matriz pseudoinversa** en $\mathbb{R}^{n\times m}$.

Esta matriz que permite establecer un proceso de "inversión parcial" (que podría representarse como $\left(A|_{\mathcal{R}(A^t)}\right)^{-1} : \mathcal{R}(A) \to \mathcal{R}(A^t)$) dio origen el tema de

Matrices Inversas Generalizadas, dentro del cual, la inversa de Moore-Penrose es sólo una de ellas.

Solución aproximada de un sistema incompatible:

Si no existe $x \in \mathbb{R}^n$ tal que $Ax = b$, el conjunto de soluciones de mínimos cuadrados es

$$\boxed{S_{MC} = A^\dagger b + \mathcal{N}(A)}.$$

En efecto, esta igualdad se deduce de la representación obtenida en la Proposición 11.3 y del Teorema 9.5:

$$S_{MC} = \{v \in \mathbb{R}^{n\times 1} : A^t A v = A^t b\} = A^\dagger b + \mathcal{N}(A^t A),$$

y de los hechos siguientes:

- $\underline{A^\dagger b \text{ es una solución particular de } A^t A v = A^t b}$: En efecto, de la tercera y primera ecuación de Penrose,

$$A^t A(A^\dagger b) = A^t (AA^\dagger)^t b = (AA^\dagger A)^t b = A^t b.$$

- $\underline{\mathcal{N}(A^t A) = \mathcal{N}(A)}$: Se ha probado en la Proposición 11.2.

Además, se cumple que $A^\dagger b$ es una solución de mínimos cuadrados de $Ax = b$ en el subespacio $[\mathcal{N}(A)]^\perp = \mathcal{R}(A^t)$, es decir,

$$A^\dagger b \in [\mathcal{N}(A)]^\perp.$$

En efecto,

$$A^\dagger b \in \mathcal{R}(A^\dagger)$$

y

$$\mathcal{R}(A^\dagger) = \mathcal{R}(A^t)$$

puesto que, de la segunda, cuarta y primera ecuación de Penrose,

$$\begin{array}{rcl} \mathcal{R}(A^\dagger) & = & \mathcal{R}(A^\dagger A A^\dagger) \\ & = & \mathcal{R}((A^\dagger A)^t A^\dagger) \\ & = & \mathcal{R}(A^t (A^\dagger)^t A^\dagger) \\ & \subseteq & \mathcal{R}(A^t) \end{array}$$

y

$$\begin{aligned}
\mathcal{R}(A^t) &= \mathcal{R}((AA^\dagger A)^t) \\
&= \mathcal{R}(A^t(A^\dagger)^t A^t) \\
&= \mathcal{R}((A^\dagger A)^t A^t) \\
&\subseteq \mathcal{R}(A^\dagger).
\end{aligned}$$

Más aún, $A^\dagger b$ es la <u>única</u> solución de mínimos cuadrados de $Ax = b$ en $[\mathcal{N}(A)]^\perp$. La prueba de la unicidad es similar a la del caso de sistemas compatibles (y se propone como ejercicio).

Por lo tanto,

- si $\underline{\dim_{\mathbb{R}}(\mathcal{N}(A)) = 0}$ entonces $A^t A v = A^t b$ es un sistema compatible determinado y $S_{MC} = \{A^\dagger b\}$.
- si $\underline{\dim_{\mathbb{R}}(\mathcal{N}(A)) > 0}$ entonces $A^t A v = A^t b$ es un sistema compatible indeterminado, y de todas las soluciones de mínimos cuadrados, $A^\dagger b$ es la única solución de norma mínima de $Ax = b$ y $A^\dagger b \in [\mathcal{N}(A)]^\perp$.

En definitiva,

$$\boxed{S_{MC} \cap [\mathcal{N}(A)]^\perp = \{A^\dagger b\}}.$$

11.9. EJERCICIOS

(1) Si $A \in \mathbb{K}^{2\times 3}$ y $B \in \mathbb{K}^{3\times 2}$, utilizando los elementos de las matrices probar que $\mathcal{C}(AB) \subseteq \mathcal{C}(A)$ y $\mathcal{F}(AB) \subseteq \mathcal{F}(B)$.

(2) Sean $A, B \in \mathbb{K}^{m\times n}$. Analizar la validez de la siguiente afirmación:

$$\text{“}A \sim B \text{ si y sólo si } \mathcal{F}(A) = \mathcal{F}(B) \text{ y } \mathcal{C}(A) = \mathcal{C}(B).\text{”}$$

(3) Demostrar el apartado (b) del Lema 11.1 a partir del apartado (a) por doble inclusión. (Ayuda: Si $y \in \mathcal{C}(A)$, ¿dónde está y^t?)

(4) Para la matriz $A \in \mathbb{R}^{m\times n}$ se considera el conjunto:

$$\mathcal{N}(A) = \{x \in \mathbb{R}^n : \ Ax = 0\},$$

de todas las soluciones del sistema homogéneo $Ax = 0$, llamado el *espacio nulo* de la matriz A.

(a) Demostrar que $\mathcal{N}(A)$ es un subespacio vectorial de $\mathbb{R}^n$.

En los apartados subsiguientes se considerará la matriz

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

(b) Hallar una base $\mathcal{B}$ y la dimensión de $\mathcal{N}(A)$.

(c) Hallar un sistema de generadores de $\mathcal{N}(A)$ que no formen una base de $\mathcal{N}(A)$ y, por otro lado, hallar un conjunto linealmente independiente de vectores de $\mathcal{N}(A)$ que no formen una base de $\mathcal{N}(A)$.

(d) ¿Son vectores linealmente independientes en $\mathbb{R}^4$ las filas de la matriz A? Justificar.

(e) Encontrar el subespacio vectorial de $\mathbb{R}^4$ generado por las filas de la matriz A. ¿Generan las filas de la matriz A todo $\mathbb{R}^4$? Justificar.

(f) Usando la base hallada en el apartado (b), responder:

(I) ¿Es el vector $u = (0, -1, 1, 2)^t$ un elemento de $\mathcal{N}(A)$?

(II) ¿Y el vector $v = (-1, 0, 1, 2)^t$?

(III) Sabiendo que $z_p = (0, 1, 0, 0)^t$ es una solución particular del sistema lineal $Az = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, escribir su solución general.

(IV) Hallar el único vector $y \in \mathcal{R}(A^t)$ que resuelve (es decir, es solución particular) el sistema del apartado anterior y escribir su solución general.

(5) Sea $A \in \mathbb{K}^{m\times n}$ una matriz no nula con rango fila f y rango columna c (definidos como las dimensiones de los correspondientes subespacios fundamentales). Demostrar que:

(a) Existe una matriz $B \in \mathbb{K}^{m\times c}$ y una matriz $C \in \mathbb{K}^{c\times n}$ tal que $A = BC$.

(b) Existe una matriz $M \in \mathbb{K}^{m\times f}$ y una matriz $N \in \mathbb{K}^{f\times n}$ tal que $A = MN$.

(c) $\dim(\mathcal{F}(A)) \leq \dim(\mathcal{F}(C)) \leq c$ y $\dim(\mathcal{C}(A)) \leq \dim(\mathcal{C}(M)) \leq f$.

Deducir que $f = c$.

(6) Utilizar la Proposición 2.8 para probar que las columnas de una matriz (cuadrada) invertible a coeficientes en $\mathbb{K}$ forman una base del espacio columna de dicha matriz.

(7) Sean $A \in \mathbb{R}^{m\times n}$ y $b \in \mathbb{R}^{m\times 1}$. Demostrar el **Teorema de Rouché-Frobenius**:

$$Ax = b \text{ tiene solución} \qquad \Leftrightarrow \qquad \operatorname{rg}([\ A \quad b\ ]) = \operatorname{rg}([\ A\ ])$$

utilizando el espacio columna de A, teoría de espacios vectoriales y el teorema fundamental del rango.

(8) Sea $A \in \mathbb{R}^{m\times n}$. Demostrar que:

(a) $\mathcal{N}(A^tA) = \mathcal{N}(A)$.

(b) $\mathcal{N}(AA^t) = \mathcal{N}(A^t)$.

(c) $\operatorname{rg}(A^tA) = \operatorname{rg}(A)$.

(d) $\operatorname{rg}(AA^t) = \operatorname{rg}(A^t)$.

(e) $\mathcal{R}(A^tA) = \mathcal{R}(A^t)$.

(f) $\mathcal{R}(AA^t) = \mathcal{R}(A)$.

(g) Si se denota por $c_1, c_2, \ldots, c_n$ las columnas de la matriz A, utilizando la notación de la Proposición 10.12 probar que $A^tA = G(c_1, c_2, \ldots, c_n)$, la matriz de Gram formada con los productos escalares de las columnas $c_1, c_2, \ldots, c_n$.

(9) Resolver el sistema de ecuaciones lineales $Ax = b$ siendo

$$\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \end{bmatrix}.$$

Para ello se pide:

(a) Hallar un vector solución particular p de $Ax = b$ con tercera componente no nula y encontrar $S = p + \mathcal{N}(A)$.

(b) Hallar el único vector $\alpha \in [\mathcal{N}(A)]^\perp$ tal que $A\alpha = b$ y encontrar $S = \alpha + \mathcal{N}(A)$.

(c) Comparar ambas soluciones eligiendo una solución arbitraria de S y expresarla según cada uno de los conjuntos de los apartados anteriores.

(d) Calcular los restantes subespacios fundamentales y sus dimensiones.

(e) Realizar una interpretación geométrica en $\mathbb{R}^3$ y $\mathbb{R}^2$ situando los subespacios $\mathcal{N}(A)$, $\mathcal{N}(A^t)$, $\mathcal{R}(A)$ y $\mathcal{R}(A^t)$.

(10) Probar que las soluciones de un sistema lineal compatible coinciden con sus soluciones de mínimos cuadrados.

(11) Sea $O \neq A \in \mathbb{R}^{m\times n}$. Si $A = QP$ es una factorización de rango completo de A, probar que $\mathcal{N}(A) = \mathcal{N}(P)$.

(12) Encontrar la recta de ecuación $y = ax+b$ que mejor aproxima a los puntos $(0, 0)$, $(1, 1)$ y $(2, 3)$ en el sentido de los mínimos cuadrados. Representar gráficamente los puntos y la recta obtenida. En Estadística se conoce como la **recta de regresión**.

(13) Encontrar las soluciones de mínimos cuadrados de los siguientes sistemas $Ax = b$. Analizar, en cada caso, si es única o no. Hallar la solución de mínimos cuadrados de norma mínima para cada uno de ellos (procediendo mediante un método geométrico, mediante uno algebraico (las ecuaciones normales) y a partir de una factorización de rango completo.

(a) $A = \begin{bmatrix} 2 & 1 \\ 2 & 2 \\ 0 & 0 \end{bmatrix}$, $b = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$.

(b) $A = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}$, $b = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}$.

(c) $A = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}$, $b = \begin{bmatrix} 3 \\ 1 \\ 0 \end{bmatrix}$.

Anexos

Anexo A

Sumatorios

A.1. Definición de sumatorio

Se considera un conjunto $A = \{a_1, a_2, \ldots, a_n\} \subseteq \mathbb{R}$. Se conoce que dichos elementos se pueden sumar y multiplicar en el cuerpo $\mathbb{R}$ de los números reales. Se utiliza la letra griega sigma mayúscula

$$\sum$$

para expresar la suma $a_1 + a_2 + \cdots + a_n$; es decir,

$$\boxed{\sum_{i=1}^{n} a_i := a_1 + a_2 + \cdots + a_n},$$

donde el primer miembro se lee: "sumatorio de los elementos a_i desde $i = 1$ hasta $i = n$" y la letra i se utiliza como índice que aumenta de uno en uno.

Por ejemplo,

(a) $\sum_{i=1}^{3} a_i = a_1 + a_2 + a_3$,

(b) $\sum_{j=2}^{5} b_j = b_2 + b_3 + b_4 + b_5$.

A.2. Propiedades del sumatorio

Las siguientes son algunas propiedades del sumatorio, donde $m, n \in \mathbb{N}$ y los restanes números considerados son reales.

(1) $\sum_{i=1}^{n} \alpha a_i = \alpha \sum_{i=1}^{n} a_i$.

(2) $\sum_{i=1}^{n} \alpha = \alpha n$.

(3) $\sum_{i=1}^{n}(a_i + b_i) = \sum_{i=1}^{n} a_i + \sum_{i=1}^{n} b_i$.

(4) $\sum_{i=1}^{n} \sum_{j=1}^{m} a_{ij} b_{ij} = \sum_{j=1}^{m} \sum_{i=1}^{n} a_{ij} b_{ij}$. En particular,

$$\sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j = \left[\sum_{i=1}^{n} a_i\right] \left[\sum_{j=1}^{m} b_j\right].$$

Se estudian con detalle en un curso de Matemática Discreta.

Por ejemplo, la propiedad (4) para $n = 2$ y $m = 3$ se justifica como sigue:

$$\begin{aligned}
\sum_{i=1}^{2} \sum_{j=1}^{3} a_{ij} b_{ij} &= \underbrace{\sum_{j=1}^{3} a_{1j} b_{1j}}_{i=1} + \underbrace{\sum_{j=1}^{3} a_{2j} b_{2j}}_{i=2} \\
&= [a_{11}b_{11} + a_{12}b_{12} + a_{13}b_{13}] + [a_{21}b_{21} + a_{22}b_{22} + a_{23}b_{23}] \\
&= [a_{11}b_{11} + a_{21}b_{21}] + [a_{12}b_{12} + a_{22}b_{22}] + [a_{13}b_{13} + a_{23}b_{23}] \\
&= \underbrace{\sum_{i=1}^{2} a_{i1} b_{i1}}_{j=1} + \underbrace{\sum_{i=1}^{2} a_{i2} b_{i2}}_{j=2} + \underbrace{\sum_{i=1}^{2} a_{i3} b_{i3}}_{j=3} \\
&= \sum_{j=1}^{3} \sum_{i=1}^{2} a_{ij} b_{ij},
\end{aligned}$$

lo que permite intercambiar estos sumatorios entre sí.

Anexo B

Números complejos

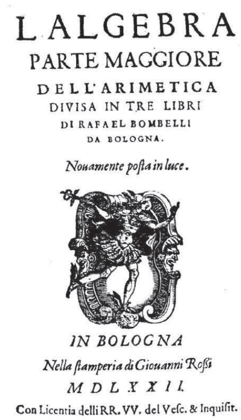
B.1. Introducción

En el libro *Ars Magna* del matemático italiano Gerolamo Cardano (1501-1576) aparecía la fórmula de resolución de la ecuación de tercer grado y una traducción completa de la *Arithmetica* de Diofanto de Alejandría (alrededor del 200/214-alrededor del 284/298 d.C.). Después de haberlo estudiado detalladamente, el ingeniero hidráulico italiano Rafael Bombelli (1526-1572) escribió el libro *Álgebra* donde introdujo las bases de los números complejos.

Las raíces de números negativos que aparecían en los estudios de Cardano, llevaron a Bombelli a desarrollar la aritmética de los números complejos, introduciendo las definiciones de suma y multiplicación. Bombelli se refería a los números imaginarios $+\sqrt{-1}$ y $-\sqrt{-1}$ como "più di meno" y "meno di meno" y fue Descartes quien posteriormente los llamaría números imaginarios. Por su parte, Leonhard Euler fue el primero en denotar, en 1777, a la raíz cuadrada de -1 como i.

Tras la definición algebraica de los números complejos, fue el matemático autodidacta francés (nacido en Suiza) Jean-Robert Argand (1768-1822)

(a) Rafael Bombelli



L ALGEBRA
PARTE MAGGIORE
DELL'ARIMETICA
DIVISA IN TRE LIBRI
DI RAFAEL BOMBELLI
DA BOLOGNA.
Nouamente posta in luce.
IN BOLOGNA
Nella stamperia di Giouanni Rossi
M D L X X I I.
Con Licentia delli RR. VV. del Vesc. & Inquisit.

(b) Libro Álgebra

Figura B.1: Rafael Bombelli y su libro Álgebra.

quien realizó el estudio de su representación geométrica. Publicó en 1806 el *Ensayo sobre una forma de representar las cantidades imaginarias mediante construcciones geométricas* y en 1813 una nueva edición del mismo apareció en la revista francesa Annales de Mathématiques. Argand basó su estudio de la representación gráfica de los números complejos en la geometría analítica, para lo que consideró a la unidad imaginaria i como una rotación de 90 grados en el plano cartesiano, a partir de lo cual es llamado *plano de Argand*.

Matemáticos como Gauss y Euler estudiaron propiedades de los números complejos y los utilizaron en sus trabajos. El mismo Euler fue quien combinó la unidad imaginaria con los números e y π en su célebre fórmula

$$e^{i\pi} + 1 = 0.$$

Sin embargo, fue un ingeniero quien, ante una necesidad práctica en sus trabajos de ingeniería, no tuvo complejos a la hora de trabajar ellos a pesar de las reticencias de los matemáticos de la época ante unos números que les resultaban tan extraños.

Hoy en día, el cuerpo de los números complejos es una herramienta básica en las matemáticas. Surgen en numerosas ramas tanto de las matemáticas como en las ingenierías (análisis complejo, teoría de las funciones analíticas, geometría compleja, fractales, teoría de control, circuitos eléctricos, etc.).

B.2. Conjuntos numéricos

Se comienza repasando brevemente los diferentes conjuntos numéricos.

B.2.1. De los números naturales a los números reales

A continuación se indicará la notación utilizada para cada uno de los conjuntos numéricos.

- *Números naturales*: $\mathbb{N} = \{1, 2, 3, \dots\} = \mathbb{Z}^+$.

  En $\mathbb{N}$ no tiene solución la ecuación

  $$n + x = m, \quad n, m \in \mathbb{N} \quad \text{si } m \leq n$$

  pues $x = m - n \notin \mathbb{N}$.

- *Números enteros negativos*: $\mathbb{Z}^- = \{\dots, -3, -2, -1\}$.

- *Números enteros*: $\mathbb{Z} = \mathbb{Z}^- \cup \{0\} \cup \mathbb{Z}^+$.

  En $\mathbb{Z}$ no tiene solución la ecuación

  $$nx = m, \quad n, m \in \mathbb{Z},\ n \neq 0,\ n \neq 1, \quad m \neq \dot{n}$$

  pues $x = \frac{m}{n} \notin \mathbb{Z}$ ($m \neq \dot{n}$ se lee: m no es un múltiplo de n).

- *Números racionales*: $\mathbb{Q} = \left\{\frac{m}{n} :\ m, n \in \mathbb{Z}, n \neq 0\right\}$.

  En $\mathbb{Q}$ no tiene solución la ecuación $x^2 = 2$: se puede probar que $\sqrt{2} \notin \mathbb{Q}$.

- *Números irracionales*: $\mathbb{I}$. Ejemplos: $\sqrt{2}$, e, $\pi \notin \mathbb{Q}$.
- *Números reales*: $\mathbb{R} = \mathbb{Q} \cup \mathbb{I}$.

  En $\mathbb{R}$ no tiene solución la ecuación $x^2 + 1 = 0$ ya que $x^2 + 1 > 0$ para todo $x \in \mathbb{R}$ con lo cual $\sqrt{-1}$ no tiene sentido en $\mathbb{R}$.

Por tanto, será necesario ampliar al conjunto de los números complejos.

B.3. Definición y propiedades de los números complejos

De teoría de conjuntos es conocido que el *producto cartesiano* $\mathbb{R} \times \mathbb{R}$ se define como el conjunto de todos los pares ordenados (a, b) de números reales, es decir:

$$\mathbb{R} \times \mathbb{R} = \{(a, b) : \ a, b \in \mathbb{R}\}.$$

B.3.1. Definición de número complejo

Definición B.1. El conjunto $\mathbb{C} = \mathbb{R} \times \mathbb{R}$ junto con las operaciones binarias definidas en $\mathbb{C}$ como sigue:

- Adición: $(a, b) + (c, d) = (a + c, b + d)$,
- Multiplicación: $(a, b) \cdot (c, d) = (ac - bd, ad + bc)$,

se llama *sistema de los números complejos*.

Escrito de manera compacta, el sistema de los números complejos consta de un conjunto y dos operaciones binarias, es decir:

$$(\mathbb{C}, +, \cdot) = (\{(a, b) : \ a, b \in \mathbb{R}\}, +, \cdot).$$

Algunas observaciones que se pueden realizar sobre los números complejos son las siguientes:

(a) Si se representa por $i = (0, 1)$ entonces

$$i^2 = (0, 1) \cdot (0, 1) = (0 \cdot 0 - 1 \cdot 1, 0 \cdot 1 + 1 \cdot 0) = (-1, 0).$$

(b) Los números complejos de la forma $(a, 0)$ se identificarán con el número real a. Por tanto se llama **unidad imaginaria** al número complejo i que satisface $i^2 = -1$.

(c) Todo número complejo de la forma $z = (a, b) \in \mathbb{C}$ se puede escribir como

$$z = (a, b) = (a, 0) + (0, b) = (a, 0) + (b, 0) \cdot (0, 1) = a + bi$$

A $a + bi$ se la llama la **forma binómica** de $z = (a, b) \in \mathbb{C}$. Se llama **parte real** de z a a y se denota $\mathrm{Re}(z)$ y **parte imaginaria** de z a b y se denota $\mathrm{Im}(z)$. Se debe observar que se llama parte imaginaria de z sólo a b (y no a bi).

La forma binómica junto a la propiedad $i^2 = -1$ de la unidad imaginaria permiten sumar y multiplicar números complejos como si fueran números reales. Se pueden recordar fácilmente pues para la regla de la suma:

$$\begin{aligned}(a + bi) + (c + di) &= (a + c) + (b + d)i \\ &= (a + c, b + d) \\ &= (a, b) + (c, d)\end{aligned}$$

y para la regla de la multiplicación:

$$\begin{aligned}(a + bi) \cdot (c + di) &= ac + adi + bci + bdi^2 \\ &= (ac - bd) + (ad + bc)i \\ &= (ac - bd, ad + bc) \\ &= (a, b) \cdot (c, d).\end{aligned}$$

Definición B.2. Se dice que los *números complejos* en forma binómica $z = a + bi$ y $w = c + di$ son *iguales* si $a = c$ y $b = d$.

B.3.2. Propiedades de los números complejos

En el siguiente resultado se presentan algunas propiedades básicas de los números complejos.

Proposición B.1. El sistema de los números complejos satisface las siguientes propiedades:

(a) $z + w = w + z, \quad \forall\, z, w \in \mathbb{C}$.

(b) $(z + v) + w = z + (v + w), \quad \forall\, z, v, w \in \mathbb{C}$.

(c) Existe $n \in \mathbb{C}$ tal que $z + n = z$, $\forall\, z \in \mathbb{C}$. Además, $n = (0, 0)$ y se llama *elemento neutro*.

(d) Para todo $z \in \mathbb{C}$, existe $w \in \mathbb{C}$ tal que $z + w = n = (0, 0)$. Se denota: $w = -z$ y se llama *simétrico* u *opuesto* de z.

(e) $zw = wz, \quad \forall\, z, w \in \mathbb{C}$.

(f) $(zv)w = z(vw), \quad \forall\, z, v, w \in \mathbb{C}$.

(g) Existe $u \in \mathbb{C}$ tal que $zu = z$, $\forall\, z \in \mathbb{C}$. Además, $u = (1, 0)$ y se llama *elemento unitario*.

(h) Para todo $z \in \mathbb{C} - \{(0, 0)\}$, existe $w \in \mathbb{C}$ tal que $zw = u = (1, 0)$. Se denota $w = z^{-1}$ y se llama *inverso* de z.

(i) $z(v + w) = zv + zw, \quad \forall\, z, v, w \in \mathbb{C}$.

Demostración. Se deduce de las correspondientes propiedades de los números reales. Se propone como ejercicio. □

Corolario B.1. $(\mathbb{C}, +, \cdot)$ es un cuerpo.

Se puede comprobar (se propone hacerlo como ejercicio) que a partir de la forma binómica de un número complejo no nulo $z = a + bi$ se obtiene su inverso

$$z^{-1} = \frac{a}{a^2 + b^2} - \frac{b}{a^2 + b^2}i. \tag{B.1}$$

A partir del simétrico y del inverso de un número complejo se pueden definir la *sustracción* y la *división* de los números complejos z y w como sigue:

- *Sustracción*: $z - w = z + (-w)$.
- *División*: $\dfrac{z}{w} = z.w^{-1}$ siempre que $w \neq (0, 0)$.

B.4. Conjugado y módulo de un número complejo

B.4.1. Conjugado de un número complejo

Definición B.3. Se llama *conjugado* de un número complejo $z = a + bi$ al número complejo

$$\overline{z} = a - bi.$$

Se llama **imaginario puro** a todo número complejo z tal que $\mathrm{Re}(z) = 0$ que suele denotarse por $z \in i\mathbb{R}$.

Proposición B.2 (Propiedades del conjugado de un número complejo)**.** Se cumplen las siguiente propiedades:

(a) $\overline{\overline{z}} = z, \quad \forall z \in \mathbb{C}$.

(b) $\overline{z+w} = \overline{z} + \overline{w}, \quad \forall z, w \in \mathbb{C}$. En consecuencia, $\overline{-z} = -\overline{z}$, $\forall z \in \mathbb{C}$.

(c) $\overline{zw} = \overline{z}\,\overline{w}, \quad \forall z, w \in \mathbb{C}$.

(d) $\overline{\left(\dfrac{z}{w}\right)} = \dfrac{\overline{z}}{\overline{w}}, \quad \forall z, w \in \mathbb{C}$ tales que $w \neq (0,0)$.

(e) $z + \overline{z} = 2\operatorname{Re}(z), \quad \forall z \in \mathbb{C}$.

(f) $z - \overline{z} = 2\operatorname{Im}(z)i, \quad \forall z \in \mathbb{C}$.

(g) $\overline{z} = z \iff z \in \mathbb{R}$.

(h) $\overline{z} = -z \iff z \in i\mathbb{R}$ (es decir, z es imaginario puro).

Demostración. Se propone como ejercicio. □

Representación geométrica: Como los números complejos se definen mediante el conjunto $\mathbb{C} = \mathbb{R} \times \mathbb{R}$, es natural representarlos geométricamente en el plano $\mathbb{R}^2$, en el que se haya fijado un sistema de coordenadas cartesianas ortogonales. Puesto que todo punto $P = (a, b)$ del plano está completamente determinado por su abscisa a y su ordenada b, es posible establecer la correspondencia biunívoca entre los puntos del plano y los números complejos de modo que a cada punto $P = (a, b)$ le corresponda el número complejo $z = a + bi$. De este modo, las partes real e imaginaria de un número complejo z son las *coordenadas cartesianas* del punto P del plano asociado a dicho número z. Así, los números complejos se representan en el llamado **plano complejo** o también **plano de Argand**. Al punto del plano que representa a un número complejo se lo llama **afijo** de dicho número. Se llama *eje real* al eje de las X (sobre el cual se representan los números complejos reales, es decir los que tienen parte imaginaria nula) y *eje imaginario* al eje de las Y (sobre el cual se representan los números imaginarios puros).

Como ejercicio se propone interpretar en el plano complejo las propiedades enumeradas en el teorema anterior (excepto (c) y (d)) considerando números complejos arbitrarios.

B.4.2. Módulo de un número complejo

Teniendo en cuenta la interpretación geométrica de los números complejos aparece de manera natural otro elemento que será su distancia al origen de coordenadas.

Definición B.4. Se llama *módulo* de un número complejo $z = a + bi$ al número real positivo o nulo

$$|z| = +\sqrt{a^2 + b^2}.$$

Proposición B.3 (Propiedades del módulo de un número complejo)**.** Se cumplen las siguientes propiedades:

(a) $|\overline{z}| = |-z| = |z|, \quad \forall z \in \mathbb{C}.$

(b) $z\overline{z} = |z|^2, \quad \forall z \in \mathbb{C}.$

(c) $|z| \geq 0, \quad \forall z \in \mathbb{C}.$

(d) $|z| = 0 \Longleftrightarrow z = 0.$

(e) $|\mathrm{Re}(z)| \leq |z|, \quad \forall z \in \mathbb{C}.$

(f) $|\mathrm{Im}(z)| \leq |z|, \quad \forall z \in \mathbb{C}.$

(g) $|z_1 z_2| = |z_1||z_2|, \ \forall z_1, z_2 \in \mathbb{C}$.

(h) $\left|\dfrac{z_1}{z_2}\right| = \dfrac{|z_1|}{|z_2|}, \ \forall z_1, z_2 \in \mathbb{C}, \ z_2 \neq 0$.

(i) $|z_1 + z_2| \leq |z_1| + |z_2|, \quad \forall z_1, z_2 \in \mathbb{C}$ (Desigualdad triangular).

(j) $||z_1| - |z_2|| \leq |z_1 - z_2|, \quad \forall z_1, z_2 \in \mathbb{C}$.

Demostración. La demostración de los apartados (a)-(f) siguen directamente de la definición. Se proponen como ejercicios.

(g) Probar $|z_1 z_2| = |z_1||z_2|$ equivale a probar $|z_1 z_2|^2 = |z_1|^2 |z_2|^2$ y esta última igualdad se deduce fácilmente del apartado (b) y de la propiedad (c) de la Proposición B.2. Se proponen los detalles como ejercicio.

(h) Se deduce de aplicar (g) a la expresión $z_1 = \frac{z_1}{z_2} z_2$. Se proponen los detalles como ejercicio.

(i) Probar $|z_1 + z_2| \leq |z_1| + |z_2|$ es equivalente a demostrar que se cumple $|z_1 + z_2|^2 \leq (|z_1| + |z_2|)^2$. En efecto,

$$\begin{aligned}
|z_1 + z_2|^2 &= (z_1 + z_2)\overline{z_1 + z_2} \\
&= (z_1 + z_2)(\overline{z_1} + \overline{z_2}) \\
&= z_1\overline{z_1} + z_1\overline{z_2} + z_2\overline{z_1} + z_2\overline{z_2} \\
&= |z_1|^2 + z_1\overline{z_2} + z_2\overline{z_1} + |z_2|^2.
\end{aligned}$$

Ahora se analiza la expresión $z_1\overline{z_2} + z_2\overline{z_1}$:

$$\begin{aligned}
z_1\overline{z_2} + z_2\overline{z_1} &= z_1\overline{z_2} + \overline{z_1\overline{z_2}} \\
&= 2\mathrm{Re}(z_1\overline{z_2}) \\
&\leq 2|\mathrm{Re}(z_1\overline{z_2})| \\
&\leq 2|z_1\overline{z_2}| \\
&= 2|z_1||\overline{z_2}| \\
&= 2|z_1||z_2|.
\end{aligned}$$

Luego,

$$|z_1+z_2|^2 = |z_1|^2+z_1\overline{z_2}+z_2\overline{z_1}+|z_2|^2 \leq |z_1|^2+2|z_1||z_2|+|z_2|^2 = (|z_1|+|z_2|)^2.$$

(j) Se deduce de aplicar (i) a las expresiones $z_1 = (z_1 - z_2) + z_2$ y $z_2 = (z_2 - z_1) + z_1$. Se proponen los detalles como ejercicio. $\square$

Con las nociones de conjugado y módulo de un número complejo se puede reescribir la definición de división de la siguiente manera.

Sean $z \in \mathbb{C}$ y $v = a + bi \in \mathbb{C} - \{(0, 0)\}$. De (B.1) se tiene que

$$v^{-1} = \frac{a}{a^2 + b^2} - \frac{b}{a^2 + b^2}i = \frac{a - bi}{a^2 + b^2} = \frac{\overline{v}}{|v|^2} = \frac{\overline{v}}{v \cdot \overline{v}}$$

entonces la división entre z y v se puede calcular como

$$\frac{z}{v} = z.v^{-1} = \frac{z \cdot \overline{v}}{v \cdot \overline{v}}.$$

De aquí, que una regla para dividir un número complejo por otro (no nulo) puede ser la de multiplicar dividendo y divisor por el conjugado del divisor.

B.5. Forma trigonométrica o polar de un número complejo

Siguiendo con la geometría de un número complejo, así como se ha definido su distancia al origen de coordenadas, también se puede definir el ángulo que forma dicho número con el eje real (positivo).

B.5.1. Argumento de un número complejo

Por el Teorema de Pitágoras es fácil ver que $[\mathrm{Re}(z)]^2 + [\mathrm{Im}(z)]^2 = |z|^2$. Ahora bien, si $z \neq 0$ se tiene que

$$\left[\frac{\mathrm{Re}(z)}{|z|}\right]^2 + \left[\frac{\mathrm{Im}(z)}{|z|}\right]^2 = 1.$$

Luego, existe $\theta \in \mathbb{R}$ tal que

$$\cos(\theta) = \frac{\operatorname{Re}(z)}{|z|} \qquad \text{y} \qquad \operatorname{sen}(\theta) = \frac{\operatorname{Im}(z)}{|z|}.$$

Uno cualesquiera de estos valores de $\theta \in \mathbb{R}$ se llama **argumento** de z y se denota $\arg(z)$. Es claro que hay infinitos valores en estas condiciones.

Definición B.5. Al ángulo $\theta \in [0, 2\pi[$, único en las condiciones anteriores, se lo llama **argumento principal** de z y se lo denota $\theta = \operatorname{Arg}(z)$.

Se observa que si $z \neq 0$ y P es el afijo de z, el argumento principal de z es el ángulo θ formado por el semieje real positivo y la semirrecta OP, medida en radianes (con valor positivo o nulo y menor de 2π), donde se toma como sentido positivo de giro en el plano el que lleva al semieje real positivo sobre el semieje imaginario positivo recorriendo un ángulo de $\frac{\pi}{2}$ radianes.

El número complejo $z = 0 = 0 + 0i$ se representa por $|z| = 0$ (no está definido su argumento).

B.5.2. Forma trigonométrica o polar de un número complejo

Si en la expresión binómica de un número complejo no nulo $z = a + bi$ se reemplazan a y b por sus valores en función del módulo y del argumento se obtiene

$$z = |z|\cos(\theta) + |z|\operatorname{sen}(\theta)i = |z|(\cos(\theta) + i\operatorname{sen}(\theta)).$$

Definición B.6. La expresión $z = |z|(\cos(\theta) + i\operatorname{sen}(\theta))$ se llama *forma trigonométrica* o *polar* de z y se la denota $z = |z|_\theta$.

B.6. Operaciones con números complejos en forma polar

Es posible realizar la *multiplicación* y la *división* de números complejos dados en forma polar.

Sean
$$z = |z|(\cos(\theta) + i\mathrm{sen}(\theta)) = |z|_\theta$$
y
$$w = |w|(\cos(\varphi) + i\mathrm{sen}(\varphi)) = |w|_\varphi,$$
donde θ y φ representan sus argumentos principales. Utilizando identidades trigonométricas no es difícil demostrar que:

- *Multiplicación*: $zw = (|z||w|)_{\theta+\varphi}$.
- *División*: $\dfrac{z}{w} = \left(\dfrac{|z|}{|w|}\right)_{\theta-\varphi}$ si $w \neq 0$.

En particular, si $z = |z|_\theta \in \mathbb{C} - 0 + 0i$ entonces $z^{-1} = \dfrac{1}{z} = \left(\dfrac{1}{|z|}\right)_{-\theta}$.

También es posible definir la *potenciación* de números complejos.

Si $z \in \mathbb{C}$ y $n \in \mathbb{N}$ se define:
$$\begin{aligned} z^1 &:= z, \\ z^{n+1} &:= z^n z, \end{aligned}$$
y se extiende a un exponente entero negativo como sigue:
$$\begin{aligned} z^0 &:= 1, \\ z^n &:= (z^{-1})^{-n}, \text{ si } z \neq 0, n \in \mathbb{Z}^-. \end{aligned}$$

El siguiente resultado se conoce con el nombre de Fórmula de de Moivre[1].

[1] El matemático francés Abraham de Moivre (1667-1754) demostró esta fórmula en 1749 que relaciona números complejos y trigonometría.

Teorema B.1 (Fórmula de de Moivre)**.** Si $z \in \mathbb{C}$ y $n \in \mathbb{Z}$ entonces

$$z^n = (|z|_\theta)^n = (|z|^n)_{n\theta}\,,$$

con $z \neq 0$ si $n < 0$.

Demostración. La prueba se divide en tres casos:

(a) $\underline{n \in \mathbb{N}}$: Se procederá por inducción sobre n. En efecto, la igualdad es evidente para $n = 1$. Sea $n > 1$ y suponer que lgualdad se cumple para $n - 1$, es decir, $z^{n-1} = (|z|^{n-1})_{(n-1)\theta}$. Se debe probar que se cumple la fórmula para n. Para ello, utilizando la regla del producto en forma polar se tiene

$$\begin{aligned} z^n &= z^{n-1}z \\ &= (|z|_\theta)^{n-1}|z|_\theta \\ &= \left(|z|^{n-1}\right)_{(n-1)\theta}|z|_\theta \\ &= \left(|z|^{n-1}|z|\right)_{[(n-1)\theta+\theta]} \\ &= (|z|^n)_{n\theta}. \end{aligned}$$

(b) $\underline{n = 0}$: es evidente de la definición.

(c) $\underline{n \in \mathbb{Z}^- \text{ y } z \neq 0}$: En este caso, si $n < 0$ entonces $-n \in \mathbb{N}$ y es posible aplicar el caso anterior:

$$\begin{aligned} z^n &= (z^{-n})^{-1} \\ &= \left[(|z|^{-n})_{(-n)\theta}\right]^{-1} \\ &= \left(\frac{1}{|z|^{-n}}\right)_{-(-n)\theta} \\ &= (|z|^n)_{n\theta}\,. \end{aligned}$$

□

Al realizar la multiplicación, división y potenciación de números complejos se han obtenido los argumentos $\theta + \varphi$, $\theta - \varphi$ y $n\theta$, respectivamente. Para que todos ellos correspondan a argumentos principales se deberán representar en el intervalo $[0, 2\pi[$.

Por otra parte, la *radicación* de números complejos presentará novedades respecto a la radicación estudiada en números reales.

Definición B.7. Dado un número complejo z y un número natural n se llama *raíz n-ésima* de z a todo número complejo w tal que

$$w^n = z.$$

El conjunto de tales números w se denotará por $\sqrt[n]{z}$.

El siguiente resultado presenta la regla de cálculo de las raíces n-ésimas de un número complejo.

Teorema B.2 (Raíces n-ésimas de un número complejo)**.** Las n raíces n-ésimas distintas de un número complejo no nulo $z = |z|_\theta$ están dadas por

$$\sqrt[n]{z} = \sqrt[n]{|z|}\left[\cos\left(\frac{\theta + 2k\pi}{n}\right) + i \operatorname{sen}\left(\frac{\theta + 2k\pi}{n}\right)\right],\ k \in \{0, 1, \ldots, n-1\}.$$

Demostración. Sea $w := \sqrt[n]{z}$. Por definición de raíz n-ésima es $w^n = z$. Escribiendo los números complejos z y w en sus formas polares $z = |z|_\theta$ y $w = |w|_\alpha$ y aplicando la fórmula de de Moivre,

$$(|w|^n)_{n\alpha} = |z|_\theta.$$

Luego,

$$|w|^n = |z| \qquad \text{y} \qquad n\alpha = \theta + 2k\pi, \qquad \text{con } k \in \mathbb{Z}.$$

Como todo número real positivo admite una única raíz real positiva, se tiene que

$$|w| = +\sqrt[n]{|z|} \qquad \text{y} \qquad \alpha = \frac{\theta + 2k\pi}{n}, \qquad \text{con } k \in \mathbb{Z}.$$

De aquí se aprecia que el módulo de w está unívocamente determinado por la raíz n-ésima aritmética del módulo de z. Sin embargo, se tiene un valor de α para cada valor que se asigne a $k \in \mathbb{Z}$. Se observa que para los valores $k = 0, 1, 2, \ldots, n-1$ se obtienen los siguientes argumentos principales

$$\frac{\theta}{n}, \quad \frac{\theta}{n} + 1\frac{2\pi}{n}, \quad \frac{\theta}{n} + 2\frac{2\pi}{n}, \quad \ldots, \quad \frac{\theta}{n} + (n-1)\frac{2\pi}{n}. \tag{B.3}$$

Hasta ahora se ha probado que si w es raíz n-ésima de z entonces existen n números complejos $w_0, w_1, \ldots, w_{n-1}$,

$$w_k = \left(\sqrt[n]{|z|}\right)_{\frac{\theta+2k\pi}{n}}, \qquad \text{con } k = 0, 1, , \ldots, n-1,$$

todos de módulo $\sqrt[n]{|z|}$ y con argumentos dados por las expresiones (B.3). Recíprocamente, es inmediato ver que los números complejos dados por $w_0, w_1, \ldots, w_{n-1}$ son raíces n-ésimas de z (por la fórmula de de Moivre y la periodicidad de las funciones seno y coseno).

Se debe probar que si k toma valores diferentes de $0, 1, \ldots, n-1$ entonces se obtiene una de las raíces n-ésimas de z dada por w_k para $k = 0, 1, , \ldots, n-1$ y que estos n números complejos son diferentes entre sí.

- *Si se consideran valores de k distintos de $0, 1, \ldots, n-1$, las raíces se repiten*: En efecto, para un valor cualquiera de $k \in \mathbb{Z}$, por el algoritmo de la división entera, se tiene que existen enteros únicos, el cociente q y el resto r, de modo que

$$k = nq + r, \qquad \text{con} \qquad 0 \leq r < n.$$

Entonces

$$\alpha = \frac{\theta + 2k\pi}{n} = \frac{\theta + 2(nq + r)\pi}{n} = q(2\pi) + \frac{\theta + 2r\pi}{n}.$$

Se ha obtenido que los argumentos $\alpha = \frac{\theta+2k\pi}{n}$ y $\frac{\theta+2r\pi}{n}$ difieren en un múltiplo entero de 2π siendo r uno de los enteros $0, 1, 2, \ldots, n-1$. De este modo, la raíz que se obtiene dando un valor k arbitrario coincide con una de las obtenidas dando el valor r para un $r \in \mathbb{Z}$ entre 0 y $n-1$, es decir coincide con uno de los números $w_0, w_1, \ldots, w_{n-1}$.

- *Las raíces $w_0, w_1, \ldots, w_{n-1}$ son todas distintas entre sí*: En efecto, si para $k \neq k'$ fuesen $w_k = w'_k$ entonces $\frac{w_k}{w'_k} = 1$ (ya que $w_{k'} \neq 0$ por ser $z \neq 0$) y, por tanto, $\arg\left(\frac{w_k}{w'_k}\right) = \arg(1)$, es decir sus argumentos cumplen que existe $t \in \mathbb{Z}$ tal que

$$(k - k')\frac{2\pi}{n} = \left(\theta + k\frac{2\pi}{n}\right) - \left(\theta + k'\frac{2\pi}{n}\right) = 0 + 2t\pi,$$

de donde $k - k' = nt$. Al ser $0 \leq k, k' \leq n-1$ se tiene que$-(n-1) \leq -k' \leq 0$ y por tanto $-(n-1) < k - k' < n-1$, luego $0 < |k - k'| < n-1$. Tomando valor absoluto, $n|t| = |nt| = |k - k'| < n - 1$ con lo que sólo cabe la posibilidad $t = 0$, es decir, $k = k'$ que es una contradicción.

Se ha probado entonces que $w_0, w_1, \ldots, w_{n-1}$ son todas las raíces n-ésimas de z (diferentes). □

Es posible realizar una **interpretación geométrica** de las raíces complejas. En efecto, como las n raíces n-ésimas de un número complejo $z = |z|_\theta \neq 0$ tienen todas módulo igual a $\sqrt[n]{|z|}$, sus afijos están sobre la circunferencia de centro $(0, 0)$ y radio $\sqrt[n]{|z|}$.

Como, además, los argumentos de las raíces son respectivamente:

$$\frac{\theta}{n}, \quad \frac{\theta}{n} + 1\frac{2\pi}{n}, \quad \frac{\theta}{n} + 2\frac{2\pi}{n}, \quad \ldots, \quad \frac{\theta}{n} + (n-1)\frac{2\pi}{n}$$

es fácil ver que los argumentos de dos raíces consecutivas difieren en $\frac{2\pi}{n}$. Por tanto, se tiene que las n raíces n-ésimas de un número complejo $z \neq 0$ son los vértices de un polígono regular de n lados inscrito en la circunferencia de centro $(0,0)$ y radio $\sqrt[n]{|z|}$.

Ejemplo B.1. Las 3 raíces cúbicas del número complejo $z = -8i$ se representan gráficamente en la Figura B.2.

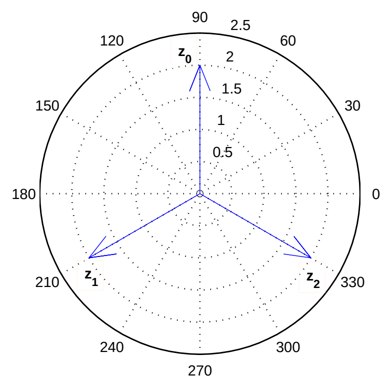


Figura B.2: Raíces complejas del número $z = -8i$.

B.7. Exponencial compleja

En la teoría de variable compleja una de las principales funciones definidas de $\mathbb{C} \to \mathbb{C}$ es la *función exponencial*. Para introducirla se comienza definiendo la *fórmula de Euler*.

Definición B.8. Dado $y \in \mathbb{R}$, se define la llamada **fórmula de Euler** como

$$e^{iy} = \cos(y) + i\operatorname{sen}(y).$$

De esta manera es posible representar un número complejo mediante la fórmula de Euler. En efecto, cualquier número complejo z se puede escribir en la forma

$$z = x + iy = |z|(\cos(\theta) + i\mathrm{sen}(\theta)) = |z|e^{i\theta}.$$

En particular, la fórmula de las raíces complejas indicada en el Teorema B.2 toma la forma más compacta siguiente:

$$\sqrt[n]{z} = \sqrt[n]{|z|}\, e^{\frac{\theta+2k\pi}{n}i}, \qquad k \in \{0, 1, \ldots, n-1\}.$$

Ahora es posible definir la función exponencial compleja del siguiente modo.

Definición B.9. Dado un número complejo $z = x + iy$, se define la **exponencial compleja** e^z como

$$e^z = e^{x+iy} = e^x(\cos(y) + i\mathrm{sen}(y)).$$

Se observa que si $y = 0$ entonces

$$e^z = e^{x+0i} = e^x(\cos(0) + i\mathrm{sen}(0)) = e^x.$$

Es decir, como era de esperar, la restricción de la exponencial compleja a $\mathbb{R}$ coincide con la exponencial real.

Como curiosidad se puede particularizar poniendo $z = \pi i$ obteniendo la siguiente fórmula:

$$e^{\pi i} + 1 = 0,$$

que está considerada como la fórmula más bella de las Matemáticas por involucrar los elementos básicos de las diferentes áreas de la misma.

En el siguiente resultado se presentan las propiedades de la exponencial compleja. Una propiedad llamativa será que la expresión $e^x > 0$ para todo $x \in \mathbb{R}$ cambia a $e^z \neq 0$ para todo $z \in \mathbb{C}$, pudiendo ahora la exponencial tomar valores negativos e incluso complejos no reales.

Teorema B.3 (Propiedades de la exponencial compleja)**.** Se cumplen las siguientes propiedades:

(a) $e^0 = 1$.

(b) $e^z \neq 0$, para todo $z \in \mathbb{C}$.

(c) $e^{z_1+z_2} = e^{z_1}e^{z_2}$, para todo $z_1, z_2 \in \mathbb{C}$.

(d) $(e^z)^{-1} = e^{-z}$, para todo $z \in \mathbb{C}$.

(e) $e^{z_1-z_2} = \dfrac{e^{z_1}}{e^{z_2}}$, para todo $z_1, z_2 \in \mathbb{C}$.

(f) $e^z = 1 \Longleftrightarrow z = 2k\pi i,\ k \in \mathbb{Z}$.

(g) $e^{z_1} = e^{z_2} \Longleftrightarrow z_1 - z_2 = 2k\pi i,\ k \in \mathbb{Z}$.

Demostración. Se propone como ejercicio. $\square$

B.8. EJERCICIOS

(1) Indicar si las siguientes afirmaciones son verdaderas o falsas:

(a) $\sqrt{-4} \in \mathbb{Z}$.
(b) $\dfrac{\pi}{2} \in \mathbb{Q}$.
(c) $-3 \in \mathbb{N}$.
(d) $\sqrt[3]{8} \in \mathbb{Q}$.
(e) $2\pi \in \mathbb{R}$.
(f) la raíz cuadrada de un número entero es siempre un número real.

(2) Comprobar que:

(a) $(\sqrt{2} - i) - i(1 - \sqrt{2}i) = -2i$.
(b) $\dfrac{1+2i}{3-4i} + \dfrac{2-i}{5i} = -\dfrac{2}{5}$.
(c) $\dfrac{5}{(1-i)(2-i)(3-i)} = \dfrac{1}{2}i$.
(d) $(1-i)^4 = -4$.
(e) $\dfrac{(2+i)^2}{3-4i} = 1$.
(f) $\dfrac{3i^{30} - i^{19}}{2i-1} = 1 + i$.

(3) Hallar el módulo y el argumento de:

(a) $(1+i)^5$.
(b) $(1+4i)^3$.

(4) Calcular el argumento de los siguientes números complejos:

(a) $z = \dfrac{-2}{1+i\sqrt{3}}$.
(b) $z = \dfrac{i}{-2-2i}$.
(c) $z = (\sqrt{3} - i)^6$.

(5) Calcular

$$\left|\frac{(3+4i)(2+i)}{(1+2i)(3-4i)}\right|$$

de dos maneras diferentes.

(6) Obtener la forma binómica de:

(a) $(2i)^{1/2}$.

(b) $(-i)^{1/3}$.

(c) $(-1)^{1/3}$.

(d) $8^{1/6}$.

(e) $(1+i)^{11}$.

(f) $32^{1/5}$.

(7) Comprobar que

$$z = \frac{\sqrt{2}}{2} - \frac{\sqrt{2}}{2}i$$

es una raíz de la ecuación

$$z^{33} = z.$$

Indicar (calculando solo dos de ellas) cómo se pueden hallar las demás raíces de dicha ecuación.

(8) Dibujar en el plano complejo las raíces sextas de la unidad sin calcularlas. Ahora, calcularlas y comprobar que el dibujo anterior es correcto.

(9) Resolver las siguientes ecuaciones en el conjunto de los números complejos:

(a) $(z^4 - 16)(z^3 + 1) = 0$.

(b) $z^6 - 9z^3 + 8 = 0$.

(10) Determinar todos los números complejos tales que:

(a) su cuadrado es igual a su conjugado.

(b) el argumento de $\dfrac{z+1}{z+2}$ sea $\dfrac{\pi}{2}$.

(11) Dado el número complejo

$$z = \frac{3-2ai}{4-3i},$$

hallar $a \in \mathbb{R}$ para que $z \in \mathbb{C}$

(a) sea imaginario puro.

(b) sea un número real.

(c) esté sobre la bisectriz del primer cuadrante.

Repetir los tres apartados anteriores buscando los números $a \in \mathbb{C}$ para que $z \in \mathbb{C}$ cumplan cada una de las condiciones pedidas.

(12) ¿Qué razones pueden darse para invalidar la siguiente "demostración"?

$$1 = \sqrt{1} = \sqrt{(-1)(-1)} = \sqrt{-1}\sqrt{-1} = i \cdot i = i^2 = -1.$$

(13) Calcular y representar gráficamente todos los números $z \in \mathbb{C}$ tales que:

$$\frac{5-3i}{2+2i} + \frac{7}{16}z^3 i^{77} - \left(\frac{1}{2} - i\right)(2+3i) = (-\sqrt{3}+i)^{22} - \frac{3}{2}i - 8^7(1+i\sqrt{3}).$$

(14) Representar gráficamente el conjunto de todos los $z \in \mathbb{C}$ tales que:

$$\frac{\pi}{2} < \operatorname{Arg}(iz) \leq \pi, \qquad \operatorname{Im}(z) > 5\operatorname{Re}(z) \qquad \text{y} \qquad 1 < |z - 5i| \leq \overline{|1 - \sqrt{3}i}$$

simultáneamente.

(15) Hallar todos los valores complejos z, si existen, tales que:

(a) $z + \dfrac{1}{z} = 2z^3$ con

(I) $|z| \neq 1$.

(II) $|z| = 1$.

(b) $(1 + \sqrt{3}i)z^3 - i^{25}z = i^{-13}z - \dfrac{4}{1 - \sqrt{3}i}$.

(16) Hallar el conjunto de los $z \in \mathbb{C}$ que cumplan las condiciones indicadas y dibujarlo en el plano complejo.

(a) $|\operatorname{Re}(z)| < 1 \quad$ y $\quad -1 \leq \operatorname{Im}(z) < 1$.

(b) $\operatorname{Re}(z^2) = 0$.

(c) $|z - 1 + 2i| < 3$.

(d) $4 < |z - i|^2$.

(e) $|z - 1| = |z - 2|$.

(f) $|z + 1| = |\overline{z} - 1|$.

(g) $||z - 5| - |z + 5|| = 8$.

(17) Demostrar que si $z \in \mathbb{C}$ es tal que $|z| = 1$ entonces

$$|1 + z|^2 + |1 - \overline{z}|^2 = 4.$$

(18) Hallar el conjunto de los $z \in \mathbb{C}$ que satisfacen las dos condiciones indicadas en cada apartado y dibujarlo en el plano complejo.

(a) $|z|^2 + 3[\mathrm{Im}(z)]^2 = 20, \quad \mathrm{Re}(z) + 2\mathrm{Im}(z) = 6$.

(b) $|z|^2 + 3[\mathrm{Im}(z)]^2 = 36, \quad |z| = 2\sqrt{3}$.

(c) $|z|^2 - 5[\mathrm{Re}(z)]^2 = 16, \quad \mathrm{Im}(z) - \mathrm{Re}(z) = 4$.

(19) Resolver las ecuaciones siguientes en el conjunto de los números complejos:

(a) $z^6 - 2z^3 + 2 = 0$.

(b) $z^5 + 64z^2 = 0$.

(c) $z^2 - (3 - 4i)z - (1 + 7i) = 0$.

(20) Expresar los siguientes números complejos en forma binómica:

(a) $e^{2i} + e^i$.

(b) $e^{1-i}e^{4+4i}$.

(c) $\dfrac{e^{2+2i}}{e^{1+i}}$.

(d) $ie^{\pi i}$.

(21) Hallar, si existen, los $z \in \mathbb{C}$ tales que:

(a) $e^z = -1$. (Notar que en el conjunto de los números complejos, la imagen de la función exponencial puede tomar valores negativos).

(b) $e^z \in \mathbb{R}$.

(c) $|e^z| = 1$.

(d) $|e^z| = a$, con $a \in \mathbb{R}$.

(e) $e^z e^{1+i} = 0$.

(f) $\dfrac{e^z}{e^{1+i}} = 1$.

(22) Probar que un número complejo z tiene módulo 1 si y sólo si $z = e^{i\theta}$, para algún $\theta \in [0, 2\pi[$.

(23) ¿Cuál es el efecto de multiplicar un número complejo z por $e^{i\pi/2}$?

Anexo C

Definición de polinomio

C.1. Introducción

A lo largo del libro se consideran las expresiones "polinomio" y "función polinomial" y se las manipula tratándolas como sinónimas. De hecho, en el cuerpo de los racionales, reales y complejos es posible considerarlas de este modo. Sin embargo, sobre un cuerpo arbitrario, en general, esto no cierto. Para verlo, en este anexo se presenta la definición algebraica de polinomio en la indeterminada X a coeficientes en el cuerpo $\mathbb{K}$ y, al finalizar, se realiza la correspondencia biunívoca entre los conjuntos que definen los polinomios (en la indeterminada X) y las funciones polinomiales (en la variable x), ambos con coeficientes en $\mathbb{K}$.

Para ello, antes es necesario recordar el concepto de sucesión.

Una sucesión de elementos de un cuerpo $\mathbb{K}$ es una aplicación

$$\begin{array}{rccc} f & : & \{0\} \cup \mathbb{N} & \rightarrow & \mathbb{K} \\ & & n & \mapsto & f(n) = a_n \end{array}$$

Dado que la sucesión f queda determinada una vez que se conocen los elementos $a_n \in \mathbb{K}$ para todos los $n \in \{0\} \cup \mathbb{N}$, es decir, una vez conocidos

$$f(0) = a_0, \quad f(1) = a_1, \quad f(2) = a_2, \quad \dots, \quad f(n) = a_n, \dots,$$

para determinar la sucesión alcanza con escribir los elementos a_i, para $i \in \{0\} \cup \mathbb{N}$, de forma ordenada como

$$a := (a_0, a_1, \dots, a_n, \dots).$$

Los números $a_i \in \mathbb{K}$ con $i \in \{0\} \cup \mathbb{N}$ de la sucesión a se llaman **coeficientes** de la sucesión.

Ejemplo C.1. *La sucesión* $f : \{0\} \cup \mathbb{N} \to \mathbb{R}$ *definida por* $f(n) = 2n$ *puede representarse mediante*

$$a = (0, 2, 4, 6, \dots, 2n, \dots).$$

Definición C.1. Se dice que dos sucesiones con coeficientes en $\mathbb{K}$

$$a = (a_0, a_1, \dots, a_n, \dots) \qquad \text{y} \qquad b = (b_0, b_1, \dots, b_n, \dots)$$

son **iguales** si y sólo si $a_i = b_i$ para todo $i \in \{0\} \cup \mathbb{N}$.

En particular, la sucesión $a = (a_0, a_1, \dots, a_n, \dots)$ es igual a la sucesión $b = (0, 0, \dots, 0, \dots)$ si y sólo si $a_i = 0$, para todo $i \in \{0\} \cup \mathbb{N}$.

C.2. Definición de polinomio

Ahora se considerarán todas las sucesiones cuyos coeficientes son cero desde un subíndice en adelante. Sea

$$\mathbb{K}[X] = \{(a_0, a_1, \dots, a_n, \dots) : a_i \in \mathbb{K} \wedge \exists m \in \{0\} \cup \mathbb{N} : a_i = 0, \forall i > m\}.$$

Como todo cuerpo con más de un elemento tiene los elementos 0 y 1 (con $0 \neq 1$), algunos elementos de $\mathbb{K}[X]$ son:

(a) $(0, 0, 0, \ldots, 0, \ldots)$: $a_i = 0$, $\forall i > 0 = m$ (en realidad, se verifica para cualquier $m \in \{0\} \cup \mathbb{N}$),

(b) $(1, 0, 0, \ldots, 0, \ldots)$: $a_0 = 1$, $a_i = 0$, $\forall i > 0 = m$,

(c) $(0, 1, 0, \ldots, 0, \ldots)$: $a_0 = 0$, $a_1 = 1$, $a_i = 0$, $\forall i > 1 = m$

(d) $(0, 0, 1, \ldots, 0, \ldots)$: $a_0 = a_1 = 0$, $a_2 = 1$, $a_i = 0$, $\forall i > 2 = m$,

etc.

A continuación se algebrizará $\mathbb{K}[X]$ (es decir, se definirán una adición, una multiplicación por escalares y una multiplicación que cumplan ciertas propiedades), y se buscará una representación sencilla[1] de los elementos de $\mathbb{K}[X]$.

Definición C.2. Sean $a = (a_0, a_1, \ldots, a_n, \ldots) \in \mathbb{K}[X]$ y $b = (b_0, b_1, \ldots, b_n, \ldots) \in \mathbb{K}[X]$. Se llama **adición** de a y b a la operación denotada por $+$, dada por la sucesión denominada **suma**, cuyo coeficiente i-ésimo es $(a + b)_i = a_i + b_i$ para todo $i \in \{0\} \cup \mathbb{N}$. Es decir,

$$a + b = (a_0 + b_0, a_1 + b_1, \ldots, a_n + b_n, \ldots).$$

Es fácil probar que la suma de elementos de $\mathbb{K}[X]$ cumple las siguientes propiedades:

(a) Si $a, b \in \mathbb{K}[X]$ entonces $a + b \in \mathbb{K}[X]$.

(b) La suma de $\mathbb{K}[X]$ cumple las siguientes propiedades: asociativa, admite elemento neutro $0 := (0, 0, \ldots, 0, \ldots)$, todo elemento $a = (a_0, a_1, \ldots, a_n, \ldots) \in \mathbb{K}[X]$ admite elemento simétrico $-a := (-a_0, -a_1, \ldots, -a_n, \ldots) \in \mathbb{K}[X]$, conmutativa.

[1] Se trata de la expresión (C.4), que es la estudiada en Secundaria, generalmente como definición de polinomio.

En consecuencia, $(\mathbb{K}[X], +)$ es un grupo abeliano.

Ahora se definirá la multiplicación de un escalar por un elemento de $\mathbb{K}[X]$.

Definición C.3. Sean $a = (a_0, a_1, \ldots, a_n, \ldots) \in \mathbb{K}[X]$ y $\lambda \in \mathbb{K}$. Se llama **multiplicación** de λ por a a la operación denotada por ., definida por la sucesión denominada **producto de λ por** a, cuyo coeficiente i-ésimo es $(\lambda.a)_i = \lambda a_i$ para todo $i \in \{0\} \cup \mathbb{N}$. Es decir,

$$\lambda.a = (\lambda a_0, \lambda a_1, \ldots, \lambda a_n, \ldots).$$

En particular, $0.a = (0, 0, \ldots, 0, \ldots) = 0$.

Es fácil probar que la multiplicación de un escalar por un elemento de $\mathbb{K}[X]$ cumple las siguientes propiedades:

(a) Si $a \in \mathbb{K}[X]$ y $\lambda \in \mathbb{K}$ entonces $\lambda.a \in \mathbb{K}[X]$.

(b) Para todo $\lambda, \mu \in \mathbb{K}$, para todo $a, b \in \mathbb{K}[X]$ se verifica:

   (I) $\lambda.(a + b) = \lambda.a + \lambda.b$.

   (II) $(\lambda + \mu).a = \lambda.a + \mu.a$.

   (III) $(\lambda\mu).a = \lambda.(\mu.a)$.

   (IV) $1_{\mathbb{K}}.a = a$.

En consecuencia, $(\mathbb{K}[X], +, .)$ es un $\mathbb{K}$-espacio vectorial.

A partir de las sucesiones particulares:

$$\begin{aligned}
X_0 &= (1, 0, 0, \ldots, 0, 0, 0, \ldots),\\
X_1 &= (0, 1, 0, \ldots, 0, 0, 0, \ldots),\\
X_2 &= (0, 0, 1, \ldots, 0, 0, 0, \ldots),\\
&\vdots\\
X_i &= (0, 0, 0, \ldots, 0, 1, 0, \ldots), \text{ con } x_i = 1, x_j = 0, \text{ si } j \neq i,\\
&\vdots
\end{aligned}$$

si se considera un elemento arbitrario

$$a = (a_0, a_1, \ldots, a_n, 0, 0, \ldots) \in \mathbb{K}[X] \quad \text{con} \quad a_i = 0 \quad \text{para } i > n \tag{C.2}$$

(para algún $n \in \{0\} \cup \mathbb{N}$) entonces a puede escribirse como

$$\begin{aligned} a &= a_0.(1, 0, \ldots, 0, \ldots) + a_1.(0, 1, \ldots, 0, \ldots) + \\ &\quad + a_2.(0, 0, 1, \ldots, 0, \ldots) + \cdots + a_n.(0, 0, \ldots, \underbrace{1}_{n+1}, \ldots) \\ &= a_0.X_0 + a_1.X_1 + \cdots + a_n.X_n. \end{aligned}$$

Así, todo elemento de $\mathbb{K}[X]$ se puede representar como una *combinación lineal* finita de las sucesiones $X_0, X_1, \ldots, X_i, \ldots$ con coeficientes $a_0, a_1, \ldots, a_i, \ldots$ en $\mathbb{K}$.

También es posible definir una **multiplicación** de elementos de $\mathbb{K}[X]$. Para ello, primero se la define sobre los elementos

$$X_0, X_1, \ldots, X_i, \ldots$$

y se la extiende de modo que se cumpla la propiedad distributiva a todos los elementos de $\mathbb{K}[X]$.

Definición C.4. $X_i \cdot X_j := X_{i+j}$, para todo $i, j \in \{0\} \cup \mathbb{N}$.

El producto de la Definición C.4 cumple las propiedades:

(a) Asociativa y conmutativa.

(b) $X_0 \cdot X_i = X_i$ para todo $i \in \{0\} \cup \mathbb{N}$.

(c) $X_1^2 = X_1 \cdot X_1 = X_2$, $X_1^3 = X_1^2 \cdot X_1 = X_3$, y por inducción, $X_1^n = X_n$, para todo $n \in \mathbb{N}$.

Si se conviene en notar (no se debe confundir la sucesión 1 con el elemento neutro $1_{\mathbb{K}}$):

- $X_1^0 = X_0 =: 1$ (que, por (b), es el neutro de esta multiplicación),

- $X_1^1 = X_1 =: X$,

la expresión de un elemento de $\mathbb{K}[X]$ cualquiera toma la forma

$$a_0.1 + a_1.X + a_2.X^2 + \cdots + a_n.X^n. \tag{C.3}$$

Debido a su papel fundamental, el elemento X tiene un nombre especial y se llama **indeterminada**.

A partir de esta representación para un elemento cualquiera de $\mathbb{K}[X]$, se puede definir la **multiplicación** de dos elementos de $\mathbb{K}[X]$ teniendo en cuenta la siguiente definición y de modo que se cumplan las propiedades distributiva y asociativa.

Definición C.5. $(a_i \cdot X^i) \cdot (b_j \cdot X^j) := a_i b_j.X^{i+j}$ para $i, j \in \{0\} \cup \mathbb{N}$.

De esta forma, el producto de los elementos

$$a = a_0.1 + a_1.X + a_2.X^2 + \cdots + a_n.X^n$$

y

$$b = b_0.1 + b_1.X + b_2.X^2 + \cdots + b_m.X^m$$

de $\mathbb{K}[X]$ se definirá mediante

$$\begin{aligned} a \cdot b = & (a_0.1) \cdot (b_0.1) + \cdots + (a_0.1) \cdot (b_m.X^m) + \\ & +(a_1.X) \cdot (b_0.1) + \cdots + (a_1.X) \cdot (b_m.X^m) + \cdots + \\ & +(a_n.X^n) \cdot (b_0.1) + \cdots + (a_n.X^n) \cdot (b_m.X^m) \\ = & \ (a_0b_0).1 + (a_0b_1 + a_1b_0).X + (a_0b_2 + a_1b_1 + a_2b_0).X^2 + \\ & +(a_1b_0).X + \cdots + (a_1b_m).X^{m+1} + \cdots + \\ & + \sum_{i=0}^{m+n} (a_i b_{m+n-i}).X^{n+m} \\ = & \sum_{k=0}^{m+n} \left(\sum_{i=0}^{k} a_i b_{k-i} \right).X^k. \end{aligned}$$

Definición C.6. Se llama **multiplicación** del elemento $a = a_0 + a_1X + a_2X^2 + \cdots + a_nX^n \in \mathbb{K}[X]$ por el elemento $b = b_0 + b_1X + b_2X^2 + \cdots + b_mX^m \in \mathbb{K}[X]$, y se denota $\cdot$, a la sucesión denominada **producto** dada por

$$a \cdot b = \sum_{k=0}^{m+n} \left(\sum_{i=0}^{k} a_i b_{k-i} \right) .X^k.$$

La multiplicación de la Definición C.6 cumple las propiedades:

(a) Si $a, b \in \mathbb{K}[X]$ entonces $a \cdot b \in \mathbb{K}[X]$.

(b) Para todo $a, b, c \in \mathbb{K}[X]$, para todo $\lambda \in \mathbb{K}$ se verifica:

- (I) $(a \cdot b) \cdot c = a \cdot (b \cdot c)$.
- (II) $a \cdot (b + c) = a \cdot b + a \cdot c$ y $(a + b) \cdot c = a \cdot c + b \cdot c$.
- (III) $\lambda.(a \cdot b) = (\lambda.a) \cdot b = a \cdot (\lambda.b)$.
- (IV) $1_{\mathbb{K}[X]} \cdot a = a \cdot 1_{\mathbb{K}[X]} = a$.
- (V) $a \cdot b = b \cdot a$.

Para $a \in \mathbb{K}$, se identificará $a.1 = (a, 0, \ldots, 0, \ldots) \in \mathbb{K}[X]$ con $a \in \mathbb{K}$ (es decir, se prescindirá de los ceros que siguen a a en la sucesión). Para verlo, se puede definir la aplicación

$$\begin{array}{cccc} \varphi & : & \mathbb{K} & \rightarrow & \mathbb{K}[X] \\ & & a & \mapsto & \varphi(a) = (a, 0, 0, \ldots, 0, \ldots) \end{array}$$

Es claro que φ es inyectiva y respeta las operaciones, es decir, para todo $a, b, \lambda \in \mathbb{K}$,

$$\begin{array}{rcl} \varphi(a + b) & = & \varphi(a) + \varphi(b) \\ \varphi(ab) & = & \varphi(a) \cdot \varphi(b) \end{array}$$

En símbolos,

$$\begin{array}{ccc} a + b & \mapsto & (a + b, 0, 0, \ldots) = (a, 0, 0, \ldots) + (b, 0, 0, \ldots) \\ ab & \mapsto & (ab, 0, 0, \ldots) = (a, 0, 0, \ldots) \cdot (b, 0, 0, \ldots) \end{array}$$

Es decir, puede considerarse que dentro de $\mathbb{K}[X]$ hay una copia de $\mathbb{K}$ y así es posible identificar $a \in \mathbb{K}$ con $(a, 0, 0, \dots, 0, \dots) \in \mathbb{K}[X]$. Esta identificación puede escribirse como $a \cdot 1 = a$.

Puesto que usualmente se escribe $a_i X^i$ en lugar de $a_i \cdot X^i$ (y además X^i en lugar de $1 \cdot X^i$ si fuese $a_i = 1$), con la identificación anterior, la expresión (C.3) de un elemento de $\mathbb{K}[X]$ cualquiera viene finalmente dada mediante la representación

$$a_0 + a_1 X + a_2 X^2 + \cdots + a_n X^n. \tag{C.4}$$

En definitiva, el elemento de $\mathbb{K}[X]$ dado en (C.2) mediante una sucesión (especial) puede ser representado de forma más sencilla por (C.4) en términos de la indeterminada X.

Definición C.7. Los elementos de $\mathbb{K}[X]$ se llaman **polinomios** en la indeterminada X a coeficientes en $\mathbb{K}$.

La notación utilizada para polinomios será $P(X)$, $p(X)$, $f(X)$, etc.

Se llama **polinomio nulo** al que tiene todos sus coeficientes iguales a cero, y se denota $P(X) = 0$.

Definición C.8. Si

$$P(X) = a_0 + a_1 X + a_2 X^2 + \cdots + a_n X^n \in \mathbb{K}[X]$$

es un polinomio no nulo, se llama **grado** del polinomio P al mayor subíndice $k \in \{0\} \cup \mathbb{N}$ tal que $a_k \neq 0$, y se denota $\mathrm{gr}(P(X)) = k$. Al polinomio nulo no se le atribuye **grado**.

Por tanto, un polinomio de grado n se puede representar por

$$P(X) = a_0 + a_1 X + a_2 X^2 + \cdots + a_n X^n \in \mathbb{K}[X], \tag{C.5}$$

con $a_n \neq 0$.

Se utilizan los siguientes nombres:

- Los polinomios de grado 0 junto al polinomio nulo se llaman **polinomios constantes** y son de la forma $P(X) = a_0$, con $a_0 \in \mathbb{K}$.
- Si un polinomio tiene grado n como en (C.5), se llama **término principal** a $a_n X^n$.
- Se llama **coeficiente principal** a a_n en la expresión del polinomio de grado n como en (C.5).
- Cuando $a_n = 1$, se dice que el **polinomio** es **mónico**.

Ejemplo C.2. *En el polinomio*

$$P(X) = -2 + 3X - X^3 \in \mathbb{R}[X],$$

los coeficientes son $a_0 = -2$, $a_1 = 3$, $a_2 = 0$ y el término principal $a_3 = -1$. El polinomio es de grado 3 (por lo tanto, no es un polinomio constante) y no es mónico pues $a_3 = -1 \neq 1$.

Si se consideran solamente las operaciones de adición y multiplicación de polinomios, se ha probado que $(\mathbb{K}[X], +, \cdot)$ es un anillo, llamado el **anillo de polinomios**.

Sin embargo, en este anexo se ha construido un ejemplo de una estructura algebraica más rica denominada álgebra lineal sobre un cuerpo.

Definición C.9. Sea $\mathbb{K}$ un cuerpo. Un **álgebra lineal sobre el cuerpo** $\mathbb{K}$ es un $\mathbb{K}$-espacio vectorial $(V, \oplus, \odot)$ con una operación interna adicional llamada **multiplicación** de vectores:

Operación interna: $\begin{array}{rccc} \cdot & : & V \times V & \rightarrow & V \\ & & (\vec{u}, \vec{v}) & \mapsto & \vec{u} \cdot \vec{v} \end{array}$ sujeta a los axiomas:

(a) $(\vec{u} \cdot \vec{v}) \cdot \vec{w} = \vec{u} \cdot (\vec{v} \cdot \vec{w}), \quad \forall \vec{u}, \vec{v}, \vec{w} \in V.$

(b) $\vec{u}\cdot(\vec{v}\oplus\vec{w}) = \vec{u}\cdot\vec{v}\oplus\vec{u}\cdot\vec{w}$ y $(\vec{u}\oplus\vec{v})\cdot\vec{w} = \vec{u}\cdot\vec{w}\oplus\vec{v}\cdot\vec{w}$, $\forall\vec{u},\vec{v},\vec{w}\in V$.

(c) $\lambda\odot(\vec{u}\cdot\vec{v}) = (\lambda\odot\vec{u})\cdot\vec{v} = \vec{u}\cdot(\lambda\odot\vec{v})$, $\forall\vec{u},\vec{v}\in V$, $\forall\lambda\in\mathbb{K}$.

Si además se cumple que:

$$\exists\vec{1}_V\in V : \vec{1}_V\cdot\vec{u} = \vec{u}\cdot\vec{1}_V = \vec{u}, \quad \forall\vec{u}\in V,$$

se dice que el álgebra lineal admite **elemento neutro** y, si también cumple

$$\vec{u}\cdot\vec{v} = \vec{v}\cdot\vec{u}, \quad \forall\vec{u},\vec{v}\in V.$$

se dice que el álgebra lineal es **conmutativa**.

En consecuencia, se ha probado el siguiente resultado.

Teorema C.1. $(\mathbb{K}[X], +, ., \cdot)$ es un álgebra lineal conmutativa con elemento neutro sobre el cuerpo $\mathbb{K}$.

C.3. Relación entre polinomio y función polinomial

Si se tiene una **función polinomial** $f : \mathbb{R}\to\mathbb{R}$ en la variable x del tipo

$$f(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n, \qquad \text{para todo } x\in\mathbb{R},$$

al ser derivable de todos los órdenes en todo $\mathbb{R}$ (y, por tanto, continua en todo $\mathbb{R}$), es fácil ver que los coeficientes a_k, para $k = 0, 1, 2, \ldots, n$ quedan completamente determinados por f. En efecto, $f(0) = a_0$

y ahora derivando sucesivamente y sustituyendo en 0 se tiene que $a_k = \frac{f^{(k)}(0)}{k!}$ para todo $k = 1, 2, \ldots, n$ con lo que

$$f(x) = f(0) + f'(0)x + \frac{f''(0)}{2!}x^2 + \cdots + \frac{f^{(n)}(0)}{n!}x^n, \qquad \forall x \in \mathbb{R}.$$

El hecho de que los coeficientes de la función polinomial queden completamente (unívocamente) determinados puede enunciarse como sigue.

Propiedad[a][Identificación de polinomios]: Sean $a_0, a_1, a_2, \ldots, a_n, b_0, b_1, b_2, \ldots, b_m \in \mathbb{R}$. Si

$$a_0 + a_1x + a_2x^2 + \cdots + a_nx^n = b_0 + b_1x + b_2x^2 + \cdots + b_mx^m, \forall x \in \mathbb{R}$$

entonces $n = m$ y $a_k = b_k$ para todo $k = 0, 1, 2, \ldots, n$.

[a]Comparar con la Definición C.1.

La propiedad anterior es también válida para funciones polinomiales a coeficientes complejos[2] (y, en particular, para coeficientes racionales pues $\mathbb{Q} \subseteq \mathbb{R} \subseteq \mathbb{C}$).

Dado un **polinomio** $P(X) = a_0 + a_1X + a_2X^2 + \cdots + a_nX^n \in \mathbb{K}[X]$, la aplicación definida por

$$\begin{array}{cccc} \varphi_P & : & \mathbb{K} & \to & \mathbb{K} \\ & & x & \mapsto & \varphi_P(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n \end{array} \tag{C.6}$$

se denomina **función polinomial** asociada a $P(X)$.

Utilizando esta definición, la propiedad anterior se puede expresar de otra manera. Establece que no puede ocurrir que dos polinomios a coeficientes reales sean distintos y sus funciones polinomiales asociadas tomen los mismos valores sobre todos los $x \in \mathbb{R}$.

[2]Su demostración escapa el alcance de este libro pues requiere del conocimiento de *funciones de variable compleja*.

Teorema C.2. Si $\mathbb{K}$ es el cuerpo de los números racionales, reales o complejos entonces existe una correspondencia biunívoca entre polinomios y funciones polinomiales (que, además, respeta las operaciones de adición, multiplicación por escalares y multiplicación).

Demostración. Si

$$\mathbb{K}[x] := \{f : \mathbb{K} \to \mathbb{K} / \ f \text{ función polinomial a coeficientes en } \mathbb{K}\},$$

se define la aplicación

$$\begin{array}{rcccc} \Gamma & : & \mathbb{K}[X] & \to & \mathbb{K}[x] \\ & & P(X) & \mapsto & \Gamma(P(X)) = \varphi_P \end{array}$$

donde $P(X)$ y φ_P se definen como en (C.6). Se cumple:

- $\underline{\Gamma \text{ es inyectiva}}$: Si $P(X) = 0$ ó $Q(X) = 0$, utilizando la definición de Γ y la identificación de polinomios, es fácil probar que $P(X) = 0 = Q(X)$.
  Sean $P, Q \in \mathbb{K}[X]$, no nulos, con $P(X) = a_0 + a_1X + a_2X^2 + \cdots + a_nX^n$ y $Q(X) = b_0 + b_1X + b_2X^2 + \cdots + b_mX^m$ tales que $\varphi_P = \varphi_Q$. De $\varphi_P(x) - \varphi_Q(x) = 0$ para todo $x \in \mathbb{K}$ es claro que $(a_0 + a_1x + a_2x^2 + \cdots + a_nx^n) - (b_0 + b_1x + b_2x^2 + \cdots + b_mx^m) = 0$, $\forall x \in \mathbb{K}$. Se trata de una función polinomial que se anula en todos los elementos de $\mathbb{K} \in \{\mathbb{Q}, \mathbb{R}, \mathbb{C}\}$, que son conjuntos con infinitos elementos, lo que es imposible, a menos que sea la función polinomial idénticamente nula[3]. Luego, $n = m$ y $a_k = b_k$, para todo $k = 1, 2, \ldots, n$. Entonces $Q(X) = a_0 + a_1X + a_2X^2 + \cdots + a_nX^n$ y así, $P(X) = Q(X)$.

[3]Para demostrarlo se realiza un razonamiento similar al del Ejemplo 8.47 mediante el determinante de Vandermonde. La propiedad que comparten los cuerpos $\mathbb{K} \in \{\mathbb{Q}, \mathbb{R}, \mathbb{C}\}$, necesaria para la demostración, es que sean cuerpos infinitos.

- Γ es sobreyectiva: Dada una función polinomial $f(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$, con coeficientes en $\mathbb{K}$ y tal que $a_n \neq 0$ para algún $n \geq 0$, se puede considerar el polinomio $P(X) = a_0 + a_1X + a_2X^2 + \cdots + a_nX^n \in \mathbb{K}[X]$. Se cumple que $\Gamma(P(X)) = \varphi_P(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n = f(x)$, $\forall x \in \mathbb{K}$.
  La función polinomial nula se corresponde con el polinomio nulo.
- Γ respeta las operaciones: Es un cálculo sencillo y se propone como ejercicio.

□

Se recuerda que dos **funciones polinomiales** $f, g : \mathbb{K} \to \mathbb{K}$ dadas por $f(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$ y $g(x) = b_0 + b_1x + b_2x^2 + \cdots + b_mx^m$ son **iguales** si $f(x) = g(x)$ para todo $x \in \mathbb{K}$.

Además, dos **polinomios** $P(X) = a_0 + a_1X + a_2X^2 + \cdots + a_nX^n$ y $Q(X) = b_0 + b_1X + b_2X^2 + \cdots + b_mX^m$ de $\mathbb{K}[X]$ ($n = \text{gr}(P)$ y $m = \text{gr}(Q)$) son **iguales** si $m = n$ y $a_i = b_i$ para todo $i = 0, 1, 2, \ldots, n$.

Con estas dos últimas definiciones presentes, se puede asegurar que el resultado del teorema anterior *no es válido* en general. Basta considerar un cuerpo $\mathbb{K}$ con un número finito de elementos. Por ejemplo, en $\mathbb{K} = \mathbb{Z}_2 = \{0, 1\}$ con las operaciones definidas en el Ejercicio 36 de la página 99. Si se consideran los polinomios $P(X) = X$ y $Q(X) = X^2$, se tiene que:

- $P(X) \neq Q(X)$ (pues sus respectivos coeficientes no coinciden),
- las funciones polinomiales φ_P y φ_Q asociadas a P y a Q, respectivamente, son iguales (ya que $\varphi_P(0) = 0 = \varphi_Q(0)$ y $\varphi_P(1) = 1 = \varphi_Q(1)$ y, por tanto, se tiene que coinciden en todos los elementos su dominio $\mathbb{Z}_2$).

Es claro que las funciones polinomiales se pueden evaluar en todos los valores de su dominio de definición. De la correspondencia encontrada en el Teorema C.2, es natural establecer el concepto de valor de un polinomio en un elemento de $\mathbb{K}$ a partir del valor que toma su función polinomial asociada.

Definición C.10. Si $P(X) = a_0 + a_1 X + a_2 X^2 + \cdots + a_n X^n$ es un polinomio de $\mathbb{K}[X]$ y $k \in \mathbb{K}$, se denomina **valor** de $P(X)$ en k al elemento de $\mathbb{K}$ dado por

$$P(k) = a_0 + a_1 k + a_2 k^2 + \cdots + a_n k^n.$$

Si $P(k) = 0$, se dice que k es un **cero** o una **raíz** de $P(X)$.

Observación C.1. El Teorema C.2 asegura que las funciones polinomiales se comportan del mismo modo que los polinomios en el caso de estar definidos sobre un cuerpo (infinito) como $\mathbb{Q}, \mathbb{R}$ o $\mathbb{C}$ puesto que entre ambos conjuntos hay una correspondencia biunívoca que respeta las operaciones.
En particular, si las raíces de un polinomio con coeficientes reales son números reales, se pueden representar gráficamente en un sistema de coordenadas cartesianas, utilizando su función polinomial asociada.

El lector interesado puede completar la información presentada en este anexo a partir de [10, pág 117].

Anexo D

Teorema de Steinitz

El siguiente resultado permitirá deducir el teorema de equicardinalidad de las bases y relaciona la cantidad de elementos de un sistema de generadores con la de un conjunto linealmente independiente de un espacio vectorial. Este resultado es más general que el Lema 8.2.

Lema D.1. Sea $V \neq \{0\}$ un $\mathbb{K}$-espacio vectorial finitamente generado. Si $\{v_1, v_2, \ldots, v_s\}$ es un conjunto linealmente independiente de V y $\{u_1, u_2, \ldots, u_m\}$ es un sistema generador de V entonces $s \leq m$.

Demostración. Si se supone, por reducción al absurdo, que $s > m$, usando que $\overline{\{u_1, u_2, \ldots, u_m\}} = V$ se deberá llegar a que $\{v_1, v_2, \ldots, v_s\}$ es linealmente dependiente.

En efecto, al ser $\{u_1, u_2, \ldots, u_m\}$ un sistema de generadores, para cada v_i $(i = 1, 2, \ldots, s)$ existen $\alpha_{ij} \in \mathbb{K}$, para $j = 1, 2, \ldots, m$ tales que

$$v_i = \sum_{j=1}^{m} \alpha_{ij} u_j, \qquad i = 1, 2, \ldots, s. \tag{D.1}$$

Es necesario vincular esta información del sistema de generadores con la relación de dependencia lineal de $\{v_1, v_2, \ldots, v_s\}$. Para ello, se plantea una combinación lineal nula de los vectores $v_1, v_2, \ldots, v_s$:

$$0 = x_1 v_1 + x_2 v_2 + \cdots + x_s v_s = \sum_{i=1}^{s} x_i v_i, \tag{D.2}$$

para ciertos escalares $x_1, x_2, \ldots, x_s \in \mathbb{K}$. Se debe probar que es posible encontrar una combinación de estos escalares, no todos nulos, y así concluir que el conjunto $\{v_1, v_2, \ldots, v_s\}$ es linealmente dependiente.

Sustituyendo (D.1) en (D.2) y operando se tiene

$$\begin{aligned} 0 &= \sum_{i=1}^{s} x_i v_i \\ &= \sum_{i=1}^{s} x_i \left(\sum_{j=1}^{m} \alpha_{ij} u_j \right) \\ &= \sum_{i=1}^{s} \sum_{j=1}^{m} x_i \alpha_{ij} u_j \\ &= \sum_{j=1}^{m} \left(\sum_{i=1}^{s} \alpha_{ij} x_i \right) u_j. \end{aligned}$$

La información del paréntesis sugiere considerar[1] el sistema lineal homogéneo de m ecuaciones con s incógnitas cuya matriz de coeficientes está formada por los α_{ij} (fijos), es decir,

$$\left\{ \begin{array}{ccccccccc} \alpha_{11}x_1 & + & \alpha_{21}x_2 & + & \ldots & + & \alpha_{s1}x_s & = & 0 \\ \alpha_{12}x_1 & + & \alpha_{22}x_2 & + & \ldots & + & \alpha_{s2}x_s & = & 0 \\ & & & & & & & \vdots & \\ \alpha_{1m}x_1 & + & \alpha_{2m}x_2 & + & \ldots & + & \alpha_{sm}x_s & = & 0 \end{array} \right. , \tag{D.3}$$

[1] Otra forma de razonar en este paso es que: si se supone que los escalares del paréntesis son todos nulos, esa es una solución posible de escribir al vector nulo como combinación lineal de los vectores $u_1, u_2, \ldots, u_m$. A partir del planteamiento de dicho sistema, es posible encontrar una solución no trivial del mismo, mediante un razonamiento similar al dado a continuación.

que de manera compacta se escribe como $\sum_{i=1}^{s} \alpha_{ij} x_i = 0$, para $j = 1, 2, \ldots, m$.

Al ser $s > m$, la Proposición 4.2 de la página 199 asegura que dicho sistema admite alguna solución no trivial $(c_1, c_2, \ldots, c_s)$, que son los escalares (no todos nulos) buscados. Es decir, escribiendo ahora la combinación lineal (D.2) con los coeficientes $c_1, c_2, \ldots, c_s$ (en lugar de $x_1, x_2, \ldots, x_s$), que determinan una solución no trivial del sistema homogéneo (D.3), se cumple que

$$c_1 v_1 + c_2 v_2 + \cdots + c_s v_s = \sum_{i=1}^{s} c_i v_i = \sum_{j=1}^{m} \left(\sum_{i=1}^{s} \alpha_{ij} c_i \right) u_j = 0.$$

Por lo tanto, $\{v_1, v_2, \ldots, v_s\}$ es linealmente dependiente. Esta contradicción proviene de suponer que $s > m$. Luego, se ha probado que $s \leq m$. $\square$

Se puede profundizar aún más. Bajo las hipótesis del Lema D.1 es posible proporcionar un nuevo sistema generador con la misma cantidad de vectores que el de partida y que sustituya algunos de sus vectores por los vectores linealmente independientes dados. Es decir, dados un conjunto linealmente independiente L (no vacío) y un sistema generador G con $n \geq 1$ vectores de un $\mathbb{K}$-espacio vectorial V, la idea es encontrar otro sistema generador de V con n vectores donde algunos vectores de G (eventualmente todos) son intercambiados por todos los vectores de L.

Antes de establecer el teorema se presenta un lema técnico.

Lema D.2. Sea $V \neq \{0\}$ un $\mathbb{K}$-espacio vectorial. Si $X = \{u_1, u_2, \ldots, u_m\} \subseteq V$ y

$$v = \alpha_1 u_1 + \cdots + \alpha_{j-1} u_{j-1} + \alpha_j u_j + \alpha_{j+1} u_{j+1} + \cdots + \alpha_m u_m,$$

con algún $j \in \{1, 2, \ldots, m\}$ tal que $\alpha_j \neq 0$, entonces

$$\overline{\{u_1, \ldots, u_{j-1}, v, u_{j+1}, \ldots, u_m\}} = \overline{X}.$$

Demostración. Sigue inmediatamente de aplicar el Ejercicio (41) de la página 451 de modo que las operaciones elementales permitan convertir v en u_j. □

Ahora se expondrá el resultado importante de este anexo, que es debido al matemático alemán Ernst Steinitz (1891-1928). Se conoce como **Teorema del intercambio, de sustitución o del canje de Steinitz**. Si bien es válido para espacios vectoriales arbitrarios, se probará sólo para espacios finitamente generados. Se observará que la primera parte del enunciado del Teorema de Steinitz coincide con el resultado probado en el Lema D.1. No sólo se probará de una forma completamente distinta, sino que, además, la demostración proporcionará un método para hallar un nuevo sistema generador de un espacio vectorial añadiendo a los elementos de un conjunto linealmente independiente (dado) algunos elementos de otro sistema generador (también conocido).

Teorema D.1 (Teorema de Steinitz)**.** Sea $V \neq \{0\}$ un $\mathbb{K}$-espacio vectorial finitamente generado. Si $G = \{u_1, u_2, \ldots, u_n\}$ es un sistema generador de V y $L = \{v_1, v_2, \ldots, v_m\}$ es un conjunto linealmente independiente de V entonces $m \leq n$. Además, existe un subconjunto G_L de G con $n - m$ elementos tal que $G_L \cup L$ es un sistema generador de V con n elementos.

Demostración. Como $\overline{G} = \overline{\{u_1, u_2, \dots, u_n\}} = V$ y $v_1 \in V$, es claro que

$$v_1 = a_1 u_1 + a_2 u_2 + a_3 u_3 + \cdots + a_n u_n,$$

para ciertos escalares $a_1, a_2, a_3, \dots, a_n \in \mathbb{K}$. Al ser L un conjunto linealmente independiente, se tiene que $v_1 \neq 0$. Luego, alguno de los escalares a_i es no nulo. Sin pérdida de generalidad es posible suponer que $a_1 \neq 0$ (en caso contrario, se deberían renumerar los subíndices para que así sea). Por el Lema D.2,

$$\overline{\{v_1, u_2, \dots, u_n\}} = \overline{\{u_1, u_2, \dots, u_n\}} = V.$$

Hasta ahora, en el sistema generador G se ha cambiado u_1 por el vector v_1 del conjunto L.

Ahora se procede de forma similar con v_2. En efecto, como $v_2 \in V = \overline{\{v_1, u_2, \dots, u_n\}}$, es claro que

$$v_2 = b_1 v_1 + b_2 u_2 + b_3 u_3 + \cdots + b_n u_n,$$

para ciertos escalares $b_1, b_2, b_3, \dots, b_n \in \mathbb{K}$. Se debe cumplir que alguno de los b_i, $i \in \{2, 3, \dots, n\}$ sea no nulo, pues si fuesen nulos se tendría $v_2 = b_1 v_1$, lo que contradice que L es un conjunto linealmente independiente. Sin pérdida de generalidad es posible suponer que $b_2 \neq 0$ (en caso contrario, de nuevo se deberían renumerar los subíndices para que así sea). Por el Lema D.2,

$$\overline{\{v_1, u_2, u_3, \dots, u_n\}} = \overline{\{v_1, v_2, u_3, \dots, u_n\}} = V.$$

Hasta ahora, en el sistema generador G se han cambiado u_1 y u_2 por los vectores v_1 y v_2 del conjunto L.

Mientras que sea $m \leq n$, se puede continuar mediante un proceso similar hasta obtener

$$\overline{\{v_1, \dots, v_m, u_{m+1}, \dots, u_n\}} = \overline{\{u_1, \dots, u_m, u_{m+1}, \dots, u_n\}} = V.$$

Definiendo $G_L := \{u_{m+1}, \ldots, u_n\}$, es claro que $G_L \subseteq G$ tiene $n - m$ elementos y $G_L \cup L$ es un sistema generador de V con n vectores.

Sin embargo, si fuese $m > n$, en el proceso de intercambio, los n primeros vectores de L (es decir, $v_1, v_2, v_3, \ldots, v_n$) se habrían intercambiado por todos los vectores de G con lo cual se habría llegado a

$$\overline{\{v_1, v_2, v_3, \ldots, v_n\}} = \overline{\{u_1, u_2, u_3, \ldots, u_n\}} = V,$$

quedando aún, al menos un vector v_{n+1} sin intercambiar. En este caso, es claro que $v_{n+1} \in V = \overline{\{v_1, v_2, v_3, \ldots, v_n\}}$ se llega a una contradicción con la hipótesis sobre la indepenencia lineal de L. Es decir, esta situación no puede ocurrir. $\square$

Ejemplo D.1. *Encontrar un sistema generador con 5 vectores del espacio vectorial $\mathbb{R}^3$ que esté formado por el conjunto linealmente independiente*

$$L = \{v_1 = (0, 1, 0), v_2 = (1, -2, 4)\}$$

al que se añadan vectores del sistema generador

$$\begin{aligned} G &= \{u_1 = (3, -6, 9), u_2 = (1, 0, 0), u_3 = (0, -3, 0), \\ &\quad u_4 = (0, 0, 2), u_5 = (1, 1, 1)\}. \end{aligned}$$

$\triangleleft$ Se observa que se piden tantos vectores como tiene el sistema generador G, con lo cual se puede aplicar el método utilizado en la demostración[2]. Dado que, por ejemplo,

$$v_1 = 0u_1 + 0u_2 + \left(-\frac{1}{3}\right) u_3 + 0u_4 + 0u_5,$$

[2]En este caso los cálculos son sencillos y por simple observación se podría responder. El método aplicado sería necesario cuando las combinacionoes lineales no se aprecian tan fácilmente.

es posible escribir

$$u_3 = 0u_1 + 0u_2 + (-3)v_1 + 0u_4 + 0u_5,$$

lo que permite concluir, por el Lema D.2, que

$$\overline{G} = \overline{\{u_1, u_2, v_1, u_4, u_5\}},$$

con lo cual, hasta ahora, L le ha cedido un vector al sistema generador inicial G. Falta intercambiar el otro vector. En efecto, de

$$v_2 \in \mathbb{R}^3 = \overline{\{u_1, u_2, v_1, u_4, u_5\}},$$

se puede establecer que, por ejemplo,

$$v_2 = \frac{4}{9}u_1 + \left(-\frac{1}{3}\right)u_2 + \frac{2}{3}v_1 + 0u_4 + 0u_5,$$

de donde

$$u_1 = \frac{9}{4}v_2 + \left(\frac{3}{4}\right)u_2 - \frac{3}{2}v_1 + 0u_4 + 0u_5,$$

para finalmente concluir, por el Lema D.2, que

$$\overline{G} = \overline{\{v_2, u_2, v_1, u_4, u_5\}}.$$

Es claro que en este caso se ha elegido $G_L = \{u_2, u_4, u_5\}$ y es evidente que hay otras formas de realizar el intercambio. $\triangleright$

Anexo E

La dimensión de $\mathbb{R}$ sobre $\mathbb{Q}$ es infinita

En el Ejemplo 8.45 de la página 390, se ha probado que la dimensión del espacio vectorial $\mathbb{R}$ sobre el cuerpo $\mathbb{Q}$ es infinita a partir de la trascendencia del número π.

En este anexo se probará que $\mathbb{R}$ es de dimensión infinita sobre $\mathbb{Q}$ a partir de los siguientes resultados.

Se recuerda que un **número** entero p se llama **primo** si es distinto de 0, 1 y -1 y es divisible por ± 1 y $\pm p$, es decir, no tiene divisores propios.

Teorema E.1 (Primer Teorema de Euclides)**.** Existen infinitos números primos.

Demostración. Alcanza con probar que en $\mathbb{Z}$ existen infinitos números primos positivos.

Si se supone, por reducción al absurdo, que existe sólo un número finito de números primos positivos, a todos ellos es posible enumerarlos como

$$p_1, p_2, \ldots, p_n.$$

Ahora se considera el número

$$N := p_1 p_2 \dots p_n + 1. \tag{E.1}$$

Es evidente que $N > 1$ y además, N es estrictamente mayor que todos los números de la lista anterior.

Es claro que el número natural N puede ser primo o compuesto. Se analiza cada caso.

- Si N fuese primo, se llega a una contradicción, por ser N un primo que no está en la lista anterior de todos los números primos.
- Si N fuese compuesto, existiría algún factor primo p que divide a N. Al ser p un primo, debe ser uno de los de la lista anterior, es decir, $p = p_i$ para algún $i \in \{1, 2, \dots, n\}$. Luego, $N = p_i q$, para algún número natural q. Sustituyendo en (E.1),

  $$p_i q = N = p_1 p_2 \dots p_n + 1,$$

  de donde, trasponiendo términos y sacando factor común el primo p_i común a ambas factorizaciones, se tiene

  $$p_i(q - p_1 p_2 \dots p_{i-1} p_{i+1} \dots p_n) = p_i q - p_1 p_2 \dots p_n = 1,$$

  Es claro que el factor $q - p_1 p_2 \dots p_{i-1} p_{i+1} \dots p_n \notin \{0, 1, -1\}$ (pues si fuese 0, se tendría que $0 = 1$, si fuese 1, sería $p_i = 1$ y si fuese -1, sería $p_i = -1$ y ninguna de las situaciones es posible), luego $|q - p_1 p_2 \dots p_{i-1} p_{i+1} \dots p_n| > 1$, con lo cual el número 1 es un número compuesto (por ser producto de dos números enteros distintos de 0, 1 y -1), que es una contradicción.

$\square$

Teorema E.2 (Teorema fundamental de la Aritmética)**.** Todo número entero $n \in \mathbb{Z}$, distinto de $0, 1$ y -1, se puede factorizar como producto de una cantidad finita de números primos y esta factorización es única salvo el orden de los factores.

Demostración. Una demostración se puede encontrar en [2, pág. 14]. □

Teorema E.3. El espacio vectorial $\mathbb{R}$ es de dimensión infinita sobre $\mathbb{Q}$, esto es

$$\dim_{\mathbb{Q}}(\mathbb{R}) = \infty.$$

Demostración. Se considera el siguiente conjunto

$$S = \{\log(p) : p \text{ es un primo positivo}\} \subset \mathbb{R},$$

donde log denota la función logaritmo natural.

Por el Primer Teorema de Euclides y recordando que la función logaritmo es inyectiva, se tiene que el conjunto S es infinito.

Para probar que S es linealmente independiente sobre $\mathbb{Q}$, se consideran los números primos positivos $p_1, p_2, \ldots, p_n$ (distintos dos a dos) y se supone que

$$0 = \alpha_1 \log(p_1) + \alpha_2 \log(p_2) + \cdots + \alpha_n \log(p_n)$$

con $\alpha_i \in \mathbb{Q}$ para $i = 1, 2, \ldots, n$.

Sin pérdida de generalidad, es posible suponer que los coeficientes son enteros, es decir, $\alpha_i \in \mathbb{Z}$ para $i = 1, 2, \ldots, n$. Si no lo fuesen, alcanzaría con multiplicar por su mínimo común denominador para conseguirlo.

De acuerdo a las propiedades de los logaritmos, se tiene que

$$0 = \log(p_1^{\alpha_1} p_2^{\alpha_2} \ldots p_n^{\alpha_n}),$$

y por definición de logaritmo se obtiene

$$1 = p_1^{\alpha_1} p_2^{\alpha_2} \ldots p_n^{\alpha_n}.$$

Puesto que los coeficientes enteros α_i pueden ser positivos o negativos (si fuera nulo se quitaría de la expresión pues $p_i^0 = 1$), escribiendo en un miembro el producto de las potencias de p_i con exponente positivo y en el otro miembro el producto de las potencias de p_i que en principio tenían exponente negativo, se consigue una igualdad de dos factorizaciones de un número entero como producto de primos. Al ser todos los primos diferentes entre sí, por la unicidad que asegura el Teorema Fundamental de la Aritmética, los exponentes deben ser todos nulos. Esto es, $\alpha_1 = \alpha_2 = \cdots = \alpha_n = 0$.

Luego, el conjunto S es linealmente independiente en $\mathbb{R}$. $\square$

“Tal es el destino del filósofo. Respecto a sí mismo, elevará su alma hasta el más alto grado del conocimiento inteligible, para fijar así las miradas de su espíritu en ese foco, de donde irradia toda luz, y de donde nace toda realidad visible e inteligible: ‘En los confines del mundo intelectual está la idea del bien, que se percibe con dificultad, pero que no es posible percibir, sin deducir que ella es la causa de todo cuanto existe de bello y de bueno; que en el mundo visible produce la luz y el astro de que esta procede; que en el mundo invisible produce directamente la verdad y el conocimiento; y, en fin, que es preciso fijar bien las miradas en esta idea, para conducirse con sabiduría en la vida pública y en la privada.’ Esta idea es Dios mismo, principio eterno e inmutable del orden moral y del orden político. Ahora se comprenderá por qué el fin de la educación filosófica, destinada a formar los jefes futuros del Estado, debe ser el de dirigir la inteligencia de estos hacia la idea del Bien. Esto es lo mismo que presentar el orden divino como modelo de gobierno.

¿Cómo se elevará el alma progresivamente de las primeras tinieblas a esta pura luz? Esto será objeto de ciertas ciencias, que los magistrados futuros cultivarán con preferencia, como una especie de aprendizaje intelectual.

En primera línea entra la aritmética, en lo que tiene de más elevada, ‘no para hacerla servir, como sucede entre los mercaderes y comerciantes, para las compras y las ventas, sino para elevarse por medio de la pura inteligencia a la contemplación de la esencia de los números.’

Después viene la geometría, muy propia para formar en el alma ‘ese espíritu filosófico, que eleva nuestras miradas hacia las cosas de lo alto, en lugar de abatirlas sobre las cosas de este mundo’, con tal que procuremos fijarnos, no en las figuras, sino en la ideas que representan.

En tercer lugar, será preciso crear una ciencia, aún no inventada, pero necesaria para completar la precedente, una geometría de los sólidos de tres dimensiones.

Y en cuarto lugar, la astronomía, estudiada con el mismo espíritu que las tres primeras ciencias.

Pero todas éstas no serán más que preludios de la verdadera ciencia filosófica, la que pone al hombre en situación de dar y entender la razón de todas las cosas. ¿Cuál es? La dialéctica, ciencia y método a la vez, que da al alma la facultad de elevarse desde los objetos más humildes hasta la idea del bien, y de descender luego de la idea del bien hasta los más humildes objetos, recorriendo así en su marcha todos los grados del ser. Esta es la ciencia última, ‘la cima y el coronamiento de las demás ciencias’.”

Platón, Obras completas, edición de Patricio de Azcárate, tomo 7, Madrid 1872.

Bibliografía

[1] C. Badesa, I. Jané, R. Jansana, *Elementos de lógica formal*, Ariel Filosofía, Barcelona, 1998.

[2] M. Castellet, I. Llerena, *Álgebra Lineal y Geometría*, Editorial Reverté, Barcelona, 2000.

[3] J. de Burgos, *Álgebra Lineal y Geometría Cartesiana*, Tercera Edición, McGrawHill, Madrid, 2006.

[4] S.H. Friedberg, A.J. Insel, L.E. Spence, *Elementary Linear Algebra*, Pearson Education, Londres, 2008.

[5] E. Gentile, *Anillo de polinomios*, Editorial Docencia, Buenos Aires, 1980.

[6] S.I. Grossman, J.J. Flores Godoy, *Álgebra Lineal*, Séptima edición, McGrawHill, Madrid, 2012.

[7] P.R. Halmos, *Teoría intuitiva de los conjuntos*, Octava edición, Editorial Continental, México, 1973.

[8] E. Hernández Rodríguez, M.J. Vázquez Gallo, M.A. Zurro Moro, *Álgebra Lineal y Geometría*, Tercera edición, Pearson Education, Madrid, 2012.

[9] I. Herstein, D. Winter, *Matrix Theory and Linear Algebra*, Macmillan Publishing Company, Nueva York, 1988.

[10] K. Hoffman, R. Kunze, *Linear Algebra*, Segunda Edición, Prentice-Hall, Londres, 1971.

[11] N. Jacobson, *Lectures in Abstract Algebra II*, Springer, Berlín, 1953.

[12] G. Jerónimo, J. Sabia, S. Tesauri, *Álgebra Lineal*, Fascículo 2, Cursos de grado, Universidad de Buenos Aires, Argentina, 2008.

[13] C. Kuratowski, *Sur la notion de l'ordre dans la théorie des ensembles*, Fundamenta Mathematicae 2, 161–171, 1921.

[14] S. Lang, *Linear Algebra*, Segunda edición, Addison-Wesley Publishing Company, Londres, 1972.

[15] D.C. Lay, S.R. Lay, J.J. McDonald, *Linear Algebra and its Applications*, Quinta edición, Pearson, Madrid, 2016.

[16] K.T. Leung, *Linear Algebra and Geometry*, Hong Kong University Press, Hong Kong, 1974.

[17] L. Merino, E. Santos, *Álgebra Lineal con métodos elementales*, Thomson, España, 2006.

[18] C.D. Meyer, Matrix Analysis and Applied Linear Algebra, SIAM, Nueva York, 2010.

[19] D. Poole, *Linear Algebra: A Modern Introduction*, Segunda edición, Thomson, 2006.

[20] F. Puerta, Red ALAMA, 2011, http://www.red-alama.es/wp-content/uploads/2011/04/isomorfismos-canónicos.pdf.

[21] A. Raya, A. Rider, R. Rubio, *Álgebra y Geometría lineal*, Reverté, Barcelona, 2018.

[22] J. Sancho San Román. Álgebra Lineal y Geometría, Octavio y Félez, Zaragoza, 1974.

[23] Gilbert Strang, Introduction to Linear Algebra, Cuarta edición, Wellesley -Cambridge Press, 2009.

[24] N. Thome, Álgebra Lineal y Geometría I, Problemas resueltos, Editorial de la Sociedad Matemática Mexicana, 2023.

[25] N. Thome, Álgebra Lineal y Geometría I, Prácticas Informáticas con MATLAB, Editorial de la Sociedad Matemática Mexicana, 2023.

[26] T. Yuster, The Reduced Row Echelon Form of a Matrix is Unique: A Simple Proof, Mathematics Magazine, Mathematical Association of America, 57, 2, 93–94, 1984.

Aspectos históricos y divulgativos

[27] *El libro de las Matemáticas*. Editorial Akal, Madrid, 2020.

[28] A. Doxiadis, *El tío Petris y la conjetura de Goldbach*, Zeta, 2009.

[29] C. B. Boyer, *History of Analytic Geometry*, Dover, Nueva York, 2004.

[30] M. Kline, *Mathematical Thought From Ancient to Modern Times*, Oxford University Press, 1972.

[31] D. Luzardo, A.J. Peña. *Historia del Álgebra Lineal hasta los albores del Siglo XX*, Divulgaciones Matemáticas, 14, 2, 153–170, (2006).

[32] K. Ríbnikov. *Historia de las Matemáticas*, Editorial MIR, Moscú, 1974.

[33] I. Stewart, *Historia de las Matemáticas: En los último 10000 años*, Segunda edición, Crítica, Barcelona, 2009.

[34] F. Vera, *Veinte matemáticos célebres*, Compañía General Fabril Editora, Buenos Aires, 1964.