



ÁLGEBRA LINEAL

Fundamentos y principios del álgebra lineal

Antonio M. Mendoza • Luis A. Godínez

ÁLGEBRA LINEAL

Luis Alfonso
Godínez Contreras

Marco Antonio
Molina Mendoza



Instituto Politécnico Nacional

Quiero expresar mi agradecimiento al Lic. Luis Alfonso Godínez Contreras, quien aparte de haber sido mi profesor en algunos cursos, me permitió recopilar el presente trabajo.

PREFACIO

En el ámbito de las ciencias matemáticas, el álgebra lineal emerge como una disciplina fundamental cuya influencia se extiende a diversas áreas del conocimiento. Desde su concepción en los albores del desarrollo matemático hasta su aplicación en campos tan variados como la física teórica, la ingeniería, la economía y la inteligencia artificial, el álgebra lineal ha probado ser un pilar inquebrantable en la comprensión y modelado de fenómenos complejos.

Al abordar la redacción de este texto, he tenido en cuenta la necesidad de equilibrar la claridad expositiva con la profundidad conceptual. Cada capítulo se estructura de manera que los fundamentos se presenten de forma gradual y sistemática, permitiendo al lector asimilar los conceptos básicos antes de adentrarse en temas más complejos. He incorporado numerosos ejemplos, ilustraciones y problemas resueltos para enriquecer la comprensión y facilitar la asimilación de los contenidos.

Este libro también pretende fomentar una apreciación más profunda del álgebra lineal como una disciplina en constante evolución, cuyas aplicaciones van más allá del ámbito académico y se extienden a la resolución de problemas prácticos en la vida cotidiana y en la investigación científica de vanguardia. Se alienta a los lectores a explorar conexiones interdisciplinarias y a aplicar los conceptos aprendidos en contextos diversos, cultivando así una comprensión holística de la materia. El propósito de este libro es servir como recurso educativo y de referencia, cada página de este volumen ha sido meticulosamente diseñada para guiar al lector a través de un viaje intelectual que abarca desde los conceptos más elementales hasta las estructuras más abstractas y sofisticadas.

Esta obra, aunque extensa, está dirigida a estudiantes de Ingeniería, que buscan ampliar su comprensión de esta área de las matemáticas y obtener herramientas prácticas para resolver problemas complejos de la vida real. En este libro se presentan los principios fundamentales de la matemática a través de un enfoque detallado en los apéndices, que abordan temas tan diversos como la inducción matemática, los números complejos, el teorema fundamental del álgebra y la resolución de ecuaciones bajo radicales. Aunque pueden parecer temas adicionales, estos apéndices son esenciales para comprender completamente las teorías matemáticas más avanzadas.

Recalco que esta obra son notas del curso de Álgebra Lineal del periodo 2024/1 y 2024/2, por lo que recomiendo consultar la bibliografía proporcionada en este trabajo. La presente obra consta de todo el contenido que adjudicó el Lic. Luis Alfonso Godínez Contreras, quien imparte clases en la Escuela Superior de Física y Matemáticas (ESFM) en el Instituto Politécnico Nacional (IPN). Hice el mayor esfuerzo en estructurar el libro a base de proposiciones, ejercicios, notas, etc.; y agregué y modifiqué algunas definiciones, a fin de lograr una mayor comprensión sin llegar a tener alguna ambigüedad.

Cabe mencionar que este libro fue mecanografiado por Marco Antonio Molina Mendoza, estudiante de la Escuela Superior de Física y Matemáticas. Sigo trabajando para corregir errores y/o mejorar este trabajo, por lo que está en constante cambio. Recomiendo visitar *Biblioteca ESFM*, un proyecto personal que compila apuntes actualizados. Allí podrán acceder a la última edición del documento.



Algunos cambios han sido resultado de los comentarios de mis amigos y estudiantes de la Escuela Superior de Física y Matemáticas. Estas son algunas de las muchas mejoras que he incorporado.

- Aspecto estético mejorado: He trabajado arduamente en mejorar el aspecto visual del libro. Incorpore nuevas fuentes y estilos para que sea más fácil de leer y asimilar los conceptos.
- Errores ortográficos y matemáticos corregidos: He realizado una revisión minuciosa para garantizar la precisión en todos los aspectos del contenido. Los errores ortográficos y matemáticos se han corregido para ofrecer una experiencia de aprendizaje fluida y confiable.
- Imágenes actualizadas: He renovado todas las imágenes y gráficos del libro para asegurarme de que sean claros, informativos y visualmente atractivos. Las nuevas ilustraciones ayudarán a comprender mejor los conceptos y aplicaciones del álgebra lineal en el mundo real.
- Ejemplos adicionales: He añadido más ejemplos paso a paso y problemas resueltos para reforzar el entendimiento de los temas discutidos.

Nota

Las imágenes que se presentan en este libro están hechas con TikZ, una herramienta para la producción de gráficos vectoriales. Dichas imágenes son de elaboración propia y son de dominio público.

En la elaboración de este libro, he optado por una presentación que refleje la esencia misma de la disciplina: precisión, claridad y elegancia matemática. Con este propósito en mente, he desarrollado una plantilla en LaTeX que resalta la pureza del contenido, utilizando únicamente el contraste entre el blanco y el negro para enfocar la atención en los conceptos fundamentales que exploraremos a lo largo de estas páginas.

La elección de un diseño en blanco y negro no es solo estética, sino una decisión que busca facilitar la comprensión del material para los lectores. Al eliminar cualquier distracción visual que pueda surgir del uso de colores llamativos, me he enfocado en ofrecer una experiencia de lectura que permita una inmersión total en el mundo del álgebra lineal.

Las ecuaciones, teoremas y demostraciones, que constituyen el núcleo de este texto, se presentan de manera clara y legible, sin interferencias visuales. La tipografía se ha seleccionado con cuidado para garantizar una lectura fluida y agradable, mientras que las figuras y gráficos han sido diseñados con precisión para complementar y reforzar los conceptos expuestos en el texto.

Entiendo la importancia de la legibilidad tanto en la versión impresa como en la digital, por lo que he dedicado especial atención a la adaptabilidad de mi plantilla. Ya sea en un libro físico o en un documento electrónico, los lectores encontrarán una presentación coherente y accesible que facilitará su inmersión en los temas tratados.

Con una presentación clara y concisa de los temas, busco fomentar la apreciación de la belleza y utilidad de esta rama de las matemáticas. Se aspira a que los lectores adquieran las habilidades y conocimientos necesarios para resolver problemas complejos y se sientan motivados a explorar aplicaciones más avanzadas del álgebra lineal en su futuro académico y profesional.

Con gratitud y humildad, me complace presentar este volumen a la comunidad académica y a todos aquellos que tienen interés en explorar esta área de las matemáticas. Espero sinceramente que su lectura les resulte tan enriquecedora y reveladora como ha sido para mí su elaboración.

Marco Antonio Molina Mendoza
México, 22 de octubre de 2025

CONTENIDO

CAPÍTULO 1	ESPACIOS VECTORIALES REALES	11
1.1	Definición y ejemplos	11
1.2	Subespacios vectoriales	26
1.3	Conjuntos generadores	30
1.4	Conjuntos linealmente independientes	33
1.5	Base y dimensión	37
1.6	Ejercicios del Capítulo 1	47
CAPÍTULO 2	MATRICES	53
2.1	El espacio vectorial de matrices	54
2.2	Multiplicación de matrices y propiedades	56
2.3	Sistemas de ecuaciones	66
2.4	Núcleo e imagen de una matriz	70
2.5	Operaciones elementales en los renglones de una matriz	77
2.6	Matrices no singulares y matrices elementales	83
2.7	Ejercicios del Capítulo 2	91
CAPÍTULO 3	DETERMINANTES	99
3.1	Determinantes por expansión de cofactores	99
3.2	Evaluación de determinantes por reducción de filas	105
3.3	Propiedades de los determinantes	108
3.4	Otros métodos para calcular determinantes	114
3.5	Algunas aplicaciones del determinante	121
CAPÍTULO 4	TRANSFORMACIONES LINEALES	131
4.1	Definición y ejemplos	132
CAPÍTULO 5	ESPACIOS VECTORIALES CON PRODUCTO INTERNO	135
5.1	Producto punto, norma y proyecciones	136
5.2	Proceso de ortonormalización de Gram-Schmidt	145
CAPÍTULO 6	DIAGONALIZACIÓN Y FORMAS CUADRÁTICAS	147

CAPÍTULO 7 ESPACIOS VECTORIALES COMPLEJOS 149

CAPÍTULO 8 APLICACIONES DEL ÁLGEBRA LINEAL 151

8.1	Construcción de curvas y superficies a través de puntos específicos	151
8.2	Interpolación mediante splines cúbicos	156
8.3	Cadenas de Markov	164
8.4	Teoría de grafos	173
8.5	Juegos de estrategia	180
8.6	Algunos modelos económicos de Leontief	187
8.7	Criptografía	194
8.8	Genética	203
8.9	Ejercicios del Capítulo 8	213

APÉNDICE A PRINCIPIOS DE INDUCCIÓN Y DE BUEN ORDEN 221

A.1	Inducción errónea en las matemáticas	224
A.2	El principio de la inducción matemática	225
A.3	El principio de buen orden	228
A.4	Más ejemplos de inducción	230
A.5	Ejercicios del Apéndice A	239

APÉNDICE B NÚMEROS COMPLEJOS 241

B.1	Operaciones con números complejos	242
B.2	Forma polar de un número complejo	250
B.3	Fórmula de De Moivre y de Euler	254
B.4	Ejercicios del Apéndice B	255

APÉNDICE C EL TEOREMA FUNDAMENTAL DEL ÁLGEBRA 257

C.1	El teorema fundamental del álgebra	257
C.2	Consecuencias del teorema fundamental	260

APÉNDICE D RESOLUCIÓN DE ECUACIONES BAJO RADICALES 265

D.1	La definición del discriminante	266
D.2	Ecuación de segundo grado	266
D.3	El discriminante de la ecuación general de segundo grado	267
D.4	La ecuación de tercer grado	267
D.5	El discriminante de la ecuación general de tercer grado	270
D.6	La ecuación de cuarto grado	273
D.7	La ecuación de quinto grado	277

1

Capítulo

ESPACIOS VECTORIALES REALES

En las páginas siguientes, exploraremos los espacios vectoriales, estructuras fundamentales en matemáticas que sustentan numerosos campos científicos y tecnológicos. Un espacio vectorial real se define mediante diez axiomas básicos, algo que debe comprenderse como un conjunto de principios que no se *demuestran*; más bien, son suposiciones que sirven como punto de partida para desarrollar teoremas.

1.1 Definición y ejemplos

Definición 1.1.1:

Sea V un conjunto no vacío y sea K un campo. Definimos la suma de elementos en V y una multiplicación por un escalar respectivamente, como

$$\begin{aligned} +: V \times V &\longrightarrow V \\ (\mathfrak{u}, \mathfrak{v}) &\longmapsto \mathfrak{u} + \mathfrak{v} \end{aligned}$$

$$\begin{aligned} \cdot: K \times V &\longrightarrow V \\ (\alpha, \mathfrak{u}) &\longmapsto \alpha \cdot \mathfrak{u} \end{aligned}$$

Decimos que V es un *espacio vectorial sobre K* , si cumple con los siguientes axiomas: Para toda $\mathfrak{x}, \mathfrak{y}, \mathfrak{z} \in V$ y $\alpha, \beta \in K$

- i) Cerradura: $\mathfrak{x} + \mathfrak{y} \in V$.
- ii) Asociatividad: $\mathfrak{x} + (\mathfrak{y} + \mathfrak{z}) = (\mathfrak{x} + \mathfrak{y}) + \mathfrak{z}$.

- iii) Conmutatividad: $x + y = y + x$.
- iv) Neutro aditivo: Existe un elemento $0 \in V$ que llamaremos cero tal que $x + 0 = 0 + x = x$.
- v) Inverso aditivo: Para cada $x \in V$, existe un elemento $-x \in V$ tal que $x + (-x) = 0$. A $-x$ se le llama inverso de x .
- vi) Cerradura: $\alpha \cdot x \in V$.
- vii) Asociatividad: $\alpha \cdot (\beta \cdot x) = (\alpha \cdot \beta) \cdot x$.
- viii) Distributividad (con un escalar): $\alpha \cdot (x + y) = \alpha x + \alpha y$.
- ix) Distributividad (con dos escalares): $(\alpha + \beta) \cdot x = \alpha x + \beta x$.
- x) Identidad multiplicativa: $1 \cdot x = x$.

Si V es un espacio vectorial sobre un campo K y $u \in V$, entonces diremos que u es un *vector*. A menos que se indique lo contrario, asumiremos que el campo K es \mathbb{R} , es decir, que trabajamos con espacios vectoriales reales.

Teorema 1.1.1: unicidad del neutro y el inverso aditivo: En un espacio vectorial V sobre K , el inverso y el neutro aditivo son únicos.

Demostración: Veamos primero que el neutro aditivo es único. Del axioma iv), existe $0 \in V$ tal que para toda $x \in V$

$$x + 0 = 0 + x = x \quad (1.1.1)$$

Supongamos que existe $\tilde{0} \in V$ tal que

$$x + \tilde{0} = \tilde{0} + x = x \quad (1.1.2)$$

de (1.1.1), como se cumple para toda $x \in V$, en particular se cumple para $x = \tilde{0}$, es decir

$$\tilde{0} + 0 = 0 + \tilde{0} = \tilde{0} \quad (1.1.3)$$

de igual forma, de la expresión (1.1.2) se cumple para toda $x \in V$, en particular, para $x = 0$. Así

$$0 + \tilde{0} = \tilde{0} + 0 = 0 \quad (1.1.4)$$

de (1.1.4) y (1.1.3),

$$\tilde{0} \stackrel{(1.1.3)}{=} 0 + \tilde{0} \stackrel{(1.1.4)}{=} 0$$

Por tanto, $\tilde{0} = 0$. Ahora, demostremos que el inverso es único. Del axioma v), tenemos que para cada $x \in V$ existe $-x$ tal que

$$x + (-x) = 0 \quad (1.1.5)$$

Supongamos que existe $y \in V$ tal que

$$x + y = 0 \quad (1.1.6)$$

Demostraremos que $y = -x$

$y = y + 0$	por axioma iv)
$= y + (x + (-x))$	por axioma v)
$= (y + x) + (-x)$	por axioma ii)
$= (x + y) + (-x)$	por axioma iii)
$= 0 + (-x)$	por la expresión (1.1.6)
$= -x$	por axioma iv)

Por lo tanto, $y = -x$.

Teorema 1.1.2: Sea V un espacio vectorial sobre K , entonces

- i. $\alpha \cdot 0 = 0, \forall \alpha \in K$.
- ii. $0 \cdot x = 0, \forall x \in V$.
- iii. Si $\alpha \cdot x = 0$, entonces $\alpha = 0$ o $x = 0$.
- iv. $(-1) \cdot x = -x$.

Demostración:

i. Sea $\alpha \in K$, tenemos que

$$0 = 0 + 0 \quad \text{por axioma iv)} \quad (1.1.7)$$

$$= \alpha \cdot 0 + (-\alpha \cdot 0) \quad \text{por axioma v)} \quad (1.1.8)$$

de (1.1.7),

$$\alpha \cdot 0 = \alpha \cdot (0 + 0) \quad \text{por def. de producto escalar} \quad (1.1.9)$$

$$= \alpha \cdot 0 + \alpha \cdot 0 \quad \text{por axioma viii)} \quad (1.1.10)$$

sustituyendo (1.1.10) en la ecuación (1.1.8), obtenemos

$$\begin{aligned} 0 &= (\alpha \cdot 0 + \alpha \cdot 0) + (-\alpha \cdot 0) \\ &= \alpha \cdot 0 + (\alpha \cdot 0 + (-\alpha \cdot 0)) && \text{por axioma ii)} \\ &= \alpha \cdot 0 + 0 && \text{por axioma v)} \\ &= \alpha \cdot 0 && \text{por axioma iv)} \end{aligned}$$

Por tanto, $\alpha \cdot 0 = 0$.

ii. Sea $x \in V$. Sabemos que en $K = \mathbb{R}$, se cumple que $0 + 0 = 0$, ahora

$$0 \cdot x = (0 + 0) \cdot x \quad \text{por def. de producto escalar} \quad (1.1.11)$$

$$= 0 \cdot x + 0 \cdot x \quad \text{por ix)} \quad (1.1.12)$$

Por otra parte,

$$0 = 0 \cdot x + (-0 \cdot x) \quad \text{por axioma v)} \quad (1.1.13)$$

Sustituyendo (1.1.12) en (1.1.13),

$$\begin{aligned} 0 &= (0 \cdot 0 + 0 \cdot 0) + (-0 \cdot x) \\ &= 0 \cdot 0 + (0 \cdot x + (-0 \cdot x)) && \text{por axioma ii)} \\ &= 0 \cdot x + 0 && \text{por axioma v)} \\ &= 0 \cdot x && \text{por axioma iv)} \end{aligned}$$

Por tanto, $0 \cdot x = 0$.

iii. Supongamos que $\alpha \neq 0$, entonces hay que demostrar que $x = 0$. Como $\alpha \in K$ y $\alpha \neq 0$, entonces existe $\alpha^{-1} \in K$ tal que

$$\alpha^{-1} \cdot \alpha = 1 \quad (1.1.14)$$

De la hipótesis,

$$\begin{aligned} 0 &= \alpha^{-1} \cdot 0 && \text{por (i) de este teorema} \\ &= \alpha^{-1} \cdot (\alpha \cdot x) && \text{por hipótesis} \\ &= (\alpha^{-1} \cdot \alpha) \cdot x && \text{por axioma vii)} \\ &= 1 \cdot x && \text{por ecuación (1.1.14)} \\ &= x && \text{por axioma x)} \end{aligned}$$

Por tanto, $x = 0$. Ahora, supongamos que $x \neq 0$. Procedamos por reducción al absurdo, supongamos que $\alpha \neq 0$, entonces existe $\alpha^{-1} \in K$ tal que

$$\alpha^{-1} \cdot \alpha = 1 \quad (1.1.15)$$

Entonces

$$\begin{aligned} x &= 1 \cdot x && \text{por axioma x)} \\ &= (\alpha^{-1} \cdot \alpha) \cdot x && \text{por la ecuación (1.1.15)} \\ &= \alpha^{-1} \cdot (\alpha \cdot x) && \text{por axioma vii)} \\ &= \alpha^{-1} \cdot 0 && \text{por hipótesis} \\ &= 0 \end{aligned}$$

Por lo tanto, $x = 0$, como se quería demostrar. Ahora, si $\alpha = 0$ y $x \neq 0$, la propiedad (iii) queda demostrada por los incisos (i) y (ii) de este teorema.

iv. Como

$$0 = 1 + (-1) \in K \quad (1.1.16)$$

por (i) de este teorema,

$$\begin{aligned} 0 &= 0 \cdot x \\ &= (1 + (-1)) \cdot x && \text{por la expresión (1.1.16)} \\ &= 1 \cdot x + (-1) \cdot x && \text{por axioma ix)} \end{aligned}$$

esto es

$$0 = 1 \cdot x + (-1) \cdot x \quad (1.1.17)$$

Ahora,

$$\begin{aligned} -x &= -x + 0 \\ &= -x + (1 \cdot x + (-1) \cdot x) && \text{por la expresión (1.1.17)} \\ &= (-x + 1 \cdot x) + (-1) \cdot x && \text{por axioma ii)} \\ &= (-x + x) + (-1) \cdot x && \text{por axioma x)} \\ &= 0 + (-1) \cdot x && \text{por axioma v)} \\ &= (-1) \cdot x && \text{por axioma iv)} \end{aligned}$$

Por tanto, $-x = (-1) \cdot x$.

En el siguiente ejemplo, demostraremos que \mathbb{R}^2 es un espacio vectorial sobre \mathbb{R} , verificando que cumple con todas las propiedades que definen un espacio vectorial.

Ejemplo 1.1.1:

Recordemos por el curso de Geometría Analítica que

$$\mathbb{R}^2 = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} \mid x \in \mathbb{R}, y \in \mathbb{R} \right\}$$

en el cual se definen dos operaciones dados $x, y \in \mathbb{R}^2$ con $x = \begin{pmatrix} a \\ b \end{pmatrix}$ e $y = \begin{pmatrix} c \\ d \end{pmatrix}$

$$\begin{aligned} +: \mathbb{R}^2 \times \mathbb{R}^2 &\longrightarrow \mathbb{R}^2 \\ (x, y) &\longmapsto \begin{pmatrix} a+c \\ b+d \end{pmatrix} \end{aligned}$$

$$\begin{aligned} \cdot: \mathbb{R} \times \mathbb{R}^2 &\longrightarrow \mathbb{R}^2 \\ (\alpha, x) &\longmapsto \begin{pmatrix} \alpha a \\ \alpha b \end{pmatrix} \end{aligned}$$

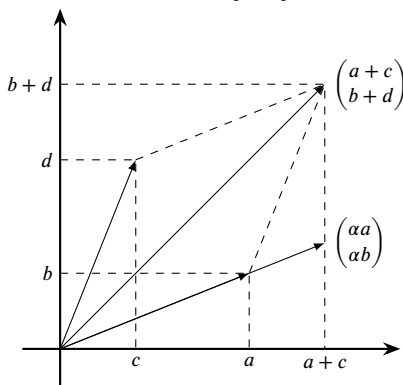


Figura 1.1

Notemos que las operaciones antes mencionadas admiten una interpretación geométrica (vea la figura 1.1). El conjunto \mathbb{R}^2 representa el conjunto de todos los pares ordenados de números reales. \mathbb{R}^2 es bidimensional, lo que significa que tiene dos dimensiones: la dimensión horizontal (eje x por convención) y la dimensión vertical (eje y por convención). En geometría analítica, se utiliza para representar figuras geométricas, ecuaciones de rectas, círculos y cualquier otra forma bidimensional. Probemos que $V = \mathbb{R}^2$, con las operaciones “+” y “·” definidas previamente es un espacio vectorial sobre \mathbb{R} . Verifiquemos los diez axiomas de un espacio vectorial (definición 1.1.1) como sigue:

i) Sea $\mathbf{x}, \mathbf{y} \in \mathbb{R}^2$ con $\mathbf{x} = \begin{pmatrix} a \\ b \end{pmatrix}$, $\mathbf{y} = \begin{pmatrix} c \\ d \end{pmatrix}$, entonces

$$\begin{aligned} \mathbf{x} + \mathbf{y} &= \begin{pmatrix} a \\ b \end{pmatrix} + \begin{pmatrix} c \\ d \end{pmatrix} \\ &= \begin{pmatrix} a + c \\ b + d \end{pmatrix} && \text{por def. de suma} \\ &= \begin{pmatrix} a + c \\ b + d \end{pmatrix} \in \mathbb{R}^2 \end{aligned}$$

Ya que $a, b, c, d \in \mathbb{R}$, entonces $a + c, b + d \in \mathbb{R}$. Por tanto, se cumple la propiedad de cerradura.

ii) Sea $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^2$ con $\mathbf{x} = \begin{pmatrix} a \\ b \end{pmatrix}$, $\mathbf{y} = \begin{pmatrix} c \\ d \end{pmatrix}$, $\mathbf{z} = \begin{pmatrix} m \\ l \end{pmatrix}$, entonces

$$\begin{aligned} \mathbf{x} + (\mathbf{y} + \mathbf{z}) &= \begin{pmatrix} a \\ b \end{pmatrix} + \left[\begin{pmatrix} c \\ d \end{pmatrix} + \begin{pmatrix} m \\ l \end{pmatrix} \right] \\ &= \begin{pmatrix} a \\ b \end{pmatrix} + \begin{pmatrix} c + m \\ d + l \end{pmatrix} && \text{por def. de suma} \\ &= \begin{pmatrix} a + (c + m) \\ b + (d + l) \end{pmatrix} && \text{por def. de suma} \\ &= \begin{pmatrix} (a + c) + m \\ (b + d) + l \end{pmatrix} && \text{por asociatividad en } \mathbb{R} \\ &= \begin{pmatrix} a + c \\ b + d \end{pmatrix} + \begin{pmatrix} m \\ l \end{pmatrix} && \text{por def. de suma} \\ &= \left[\begin{pmatrix} a \\ b \end{pmatrix} + \begin{pmatrix} c \\ d \end{pmatrix} \right] + \begin{pmatrix} m \\ l \end{pmatrix} && \text{por def. de suma} \\ &= (\mathbf{x} + \mathbf{y}) + \mathbf{z} \end{aligned}$$

Por tanto, se cumple la asociatividad.

iii) Sea $\mathbf{x}, \mathbf{y} \in \mathbb{R}^2$ con $\mathbf{x} = \begin{pmatrix} a \\ b \end{pmatrix}$, $\mathbf{y} = \begin{pmatrix} c \\ d \end{pmatrix}$, entonces

$$\begin{aligned} \mathbf{x} + \mathbf{y} &= \begin{pmatrix} a \\ b \end{pmatrix} + \begin{pmatrix} c \\ d \end{pmatrix} \\ &= \begin{pmatrix} a + c \\ b + d \end{pmatrix} && \text{por def. de suma} \\ &= \begin{pmatrix} c + a \\ d + b \end{pmatrix} && \text{por conmutatividad en } \mathbb{R} \\ &= \begin{pmatrix} c \\ d \end{pmatrix} + \begin{pmatrix} a \\ b \end{pmatrix} && \text{por def. de suma} \\ &= \mathbf{y} + \mathbf{x} \end{aligned}$$

Por tanto, se cumple la conmutatividad.

iv) Sea $\mathbf{0} \in \mathbb{R}^2$ con $\mathbf{0} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, entonces

$$\begin{aligned} \mathbf{0} + \mathbf{x} &= \begin{pmatrix} 0 \\ 0 \end{pmatrix} + \begin{pmatrix} a \\ b \end{pmatrix} \\ &= \begin{pmatrix} 0 + a \\ 0 + b \end{pmatrix} && \text{por def. de suma} \\ &= \begin{pmatrix} a \\ b \end{pmatrix} && \text{por neutro aditivo en } \mathbb{R} \\ &= \mathbf{x} \end{aligned}$$

Por tanto, se cumple la propiedad del neutro aditivo.

v) Dado $\mathbf{x} = \begin{pmatrix} a \\ b \end{pmatrix}$, existe $-\mathbf{x} = \begin{pmatrix} -a \\ -b \end{pmatrix} \in \mathbb{R}^2$ tal que

$$\begin{aligned} \mathbf{x} + (-\mathbf{x}) &= \begin{pmatrix} a \\ b \end{pmatrix} + \begin{pmatrix} -a \\ -b \end{pmatrix} \\ &= \begin{pmatrix} a + (-a) \\ b + (-b) \end{pmatrix} && \text{por def. de suma} \\ &= \begin{pmatrix} a - a \\ b - b \end{pmatrix} && \text{por definición en } \mathbb{R} \\ &= \begin{pmatrix} 0 \\ 0 \end{pmatrix} && \text{por inv. aditivo en } \mathbb{R} \\ &= \mathbf{0} \end{aligned}$$

Por tanto, se cumple la propiedad del inverso aditivo.

vi) Sea $\alpha \in \mathbb{R}$ y sea $\mathbf{x} \in \mathbb{R}^2$ con $\mathbf{x} = \begin{pmatrix} a \\ b \end{pmatrix}$, entonces

$$\begin{aligned} \alpha \cdot \mathbf{x} &= \alpha \cdot \begin{pmatrix} a \\ b \end{pmatrix} \\ &= \begin{pmatrix} \alpha a \\ \alpha b \end{pmatrix} \in \mathbb{R}^2 && \text{por def. de producto} \end{aligned}$$

Ya que $\alpha a, \alpha b \in \mathbb{R}$, se cumple la propiedad de cerradura.

vii) Sea $\alpha, \beta \in \mathbb{R}$ y sea $\mathbf{x} \in \mathbb{R}^2$ con $\mathbf{x} = \begin{pmatrix} a \\ b \end{pmatrix}$, entonces

$$\begin{aligned} \alpha \cdot (\beta \cdot \mathbf{x}) &= \alpha \cdot \left[\beta \cdot \begin{pmatrix} a \\ b \end{pmatrix} \right] \\ &= \alpha \cdot \begin{pmatrix} \beta a \\ \beta b \end{pmatrix} && \text{por def. de producto} \\ &= \begin{pmatrix} \alpha(\beta a) \\ \alpha(\beta b) \end{pmatrix} && \text{por def. de producto} \\ &= \begin{pmatrix} (\alpha\beta)a \\ (\alpha\beta)b \end{pmatrix} && \text{por asociatividad en } \mathbb{R} \\ &= (\alpha\beta) \cdot \begin{pmatrix} a \\ b \end{pmatrix} && \text{por def. de producto} \\ &= (\alpha\beta) \cdot \mathbf{x} \end{aligned}$$

Por tanto, se cumple la asociatividad.

viii) Sea $\alpha \in \mathbb{R}$ y sean $\mathbf{x}, \mathbf{y} \in \mathbb{R}^2$ con $\mathbf{x} = \begin{pmatrix} a \\ b \end{pmatrix}$ e $\mathbf{y} = \begin{pmatrix} c \\ d \end{pmatrix}$, entonces

$$\begin{aligned}
 \alpha \cdot (\mathbf{x} + \mathbf{y}) &= \alpha \cdot \left[\begin{pmatrix} a \\ b \end{pmatrix} + \begin{pmatrix} c \\ d \end{pmatrix} \right] \\
 &= \alpha \cdot \begin{pmatrix} a+c \\ b+d \end{pmatrix} && \text{por def. de suma} \\
 &= \begin{pmatrix} \alpha(a+c) \\ \alpha(b+d) \end{pmatrix} && \text{por def. de producto} \\
 &= \begin{pmatrix} \alpha a + \alpha c \\ \alpha b + \alpha d \end{pmatrix} && \text{por distributividad en } \mathbb{R} \\
 &= \begin{pmatrix} \alpha a \\ \alpha c \end{pmatrix} + \begin{pmatrix} \alpha b \\ \alpha d \end{pmatrix} && \text{por def. de suma} \\
 &= \alpha \cdot \begin{pmatrix} a \\ b \end{pmatrix} + \alpha \cdot \begin{pmatrix} c \\ d \end{pmatrix} && \text{por def. de producto} \\
 &= \alpha \cdot \mathbf{x} + \alpha \cdot \mathbf{y}
 \end{aligned}$$

Por tanto, se cumple la distributividad con dos vectores y un escalar.

ix) Sean $\alpha, \beta \in \mathbb{R}$ y sea $\mathbf{x} \in \mathbb{R}^2$ con $\mathbf{x} = \begin{pmatrix} a \\ b \end{pmatrix}$, entonces

$$\begin{aligned}
 (\alpha + \beta) \cdot \mathbf{x} &= (\alpha + \beta) \cdot \begin{pmatrix} a \\ b \end{pmatrix} \\
 &= \begin{pmatrix} (\alpha + \beta)a \\ (\alpha + \beta)b \end{pmatrix} && \text{por def. de producto} \\
 &= \begin{pmatrix} \alpha a + \beta a \\ \alpha b + \beta b \end{pmatrix} && \text{por distributividad en } \mathbb{R} \\
 &= \begin{pmatrix} \alpha a \\ \alpha b \end{pmatrix} + \begin{pmatrix} \beta a \\ \beta b \end{pmatrix} && \text{por def. de suma} \\
 &= \alpha \cdot \begin{pmatrix} a \\ b \end{pmatrix} + \beta \cdot \begin{pmatrix} a \\ b \end{pmatrix} && \text{por def. de producto} \\
 &= \alpha \cdot \mathbf{x} + \beta \cdot \mathbf{x}
 \end{aligned}$$

Por tanto, se cumple la distributividad con dos escalares y un vector.

x) Sea $\mathbf{x} \in \mathbb{R}^2$ con $\mathbf{x} = \begin{pmatrix} a \\ b \end{pmatrix}$, entonces

$$\begin{aligned}
 1 \cdot \mathbf{x} &= 1 \cdot \begin{pmatrix} a \\ b \end{pmatrix} \\
 &= \begin{pmatrix} 1 \cdot a \\ 1 \cdot b \end{pmatrix} && \text{por def. de producto} \\
 &= \begin{pmatrix} a \\ b \end{pmatrix} && \text{por identidad multiplicativa en } \mathbb{R} \\
 &= \mathbf{x}
 \end{aligned}$$

Por tanto, se cumple la propiedad de identidad multiplicativa.

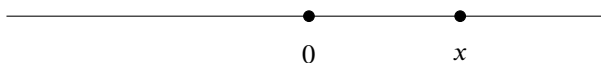
Por lo tanto, \mathbb{R}^2 es un espacio vectorial sobre \mathbb{R} .

La idea de usar pares ordenados y ternas de números reales para representar puntos en el espacio bidimensional y tridimensional era bien conocida en los siglos XVIII y XIX. Al inicio del siglo XX, matemáticos y físicos comenzaron a explorar el uso de espacios “de dimensiones superiores” en matemáticas y física. Hoy en día, incluso el estudiante común está familiarizado con la noción del tiempo como una cuarta dimensión, una idea utilizada por Albert Einstein al desarrollar la teoría

general de la relatividad. Actualmente, los físicos que trabajan en el campo de la “teoría de cuerdas” suelen utilizar el espacio de 11 dimensiones en su búsqueda de una teoría unificada que explique cómo funcionan las fuerzas fundamentales de la naturaleza. Gran parte del trabajo restante en esta sección se ocupa de extender la noción del espacio a n dimensiones.

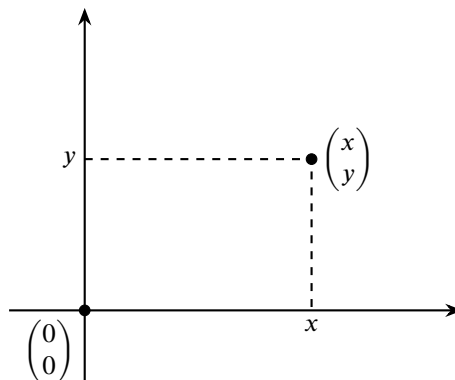
Para explorar estas ideas más a fondo, comenzamos con algo de terminología y notación. El conjunto de todos los números reales puede ser visto geoméricamente como una línea. Se le llama la recta real y se denota por \mathbb{R} o \mathbb{R}^1 . El superíndice refuerza la idea intuitiva de que una línea es unidimensional. El conjunto de todos los pares ordenados de números reales (llamados 2-tuplas) y el conjunto de todas las ternas ordenadas de números reales (llamados 3-tuplas) se denotan como \mathbb{R}^2 y \mathbb{R}^3 , respectivamente. El superíndice refuerza la idea de que los pares ordenados corresponden a puntos en el plano (bidimensional) y las ternas ordenadas a puntos en el espacio (tridimensional). Geométricamente, un número real x puede ser representado por un punto sobre una línea.

Figura 1.2 Representación de un número real x



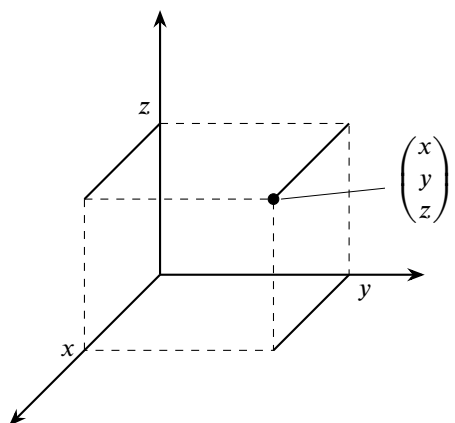
Dados $x, y \in \mathbb{R}$, la pareja ordenada $\begin{pmatrix} x \\ y \end{pmatrix}$ puede ser representado por un punto en el plano. Es decir, \mathbb{R}^2 representa el conjunto de todos los pares ordenados de números reales, lo que esencialmente constituye un plano bidimensional.

Figura 1.3 Representación de una pareja ordenada $\begin{pmatrix} x \\ y \end{pmatrix}$



Dados $x, y, z \in \mathbb{R}$, la terna ordenada $\begin{pmatrix} x \\ y \\ z \end{pmatrix}$ puede ser representada por un punto en el espacio. Es decir, \mathbb{R}^3 es el espacio tridimensional que experimentamos en la vida cotidiana, donde cada punto puede describirse con tres coordenadas reales.

Figura 1.4 Representación de una terna ordenada $\begin{pmatrix} x \\ y \\ z \end{pmatrix}$



Definición 1.1.2:

Sea $n \in \mathbb{N}$, definimos

$$\mathbb{R}^n = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \mid x_i \in \mathbb{R}, \text{ con } i = 1, 2, \dots, n \right\}$$

en donde existen dos operaciones: Dados $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ con $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ e $\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$,

$$+: \mathbb{R}^n \times \mathbb{R}^n \longrightarrow \mathbb{R}^n$$

$$(\mathbf{x}, \mathbf{y}) \longmapsto \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{pmatrix}$$

$$\cdot: \mathbb{R} \times \mathbb{R}^n \longrightarrow \mathbb{R}^n$$

$$(\alpha, \mathbf{x}) \longmapsto \begin{pmatrix} \alpha x_1 \\ \alpha x_2 \\ \vdots \\ \alpha x_n \end{pmatrix}$$

A los números x_1, x_2, \dots, x_n les llamaremos entradas o componentes de \mathbf{x} .

Definición 1.1.3:

Sean $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. Diremos que \mathbf{x} e \mathbf{y} son *iguales* (también llamado *equivalentes*), denotado por $\mathbf{x} = \mathbf{y}$, si y solo si $x_i = y_i$ para toda $i = 1, 2, \dots, n$. Es decir, si todas las entradas de \mathbf{x} e \mathbf{y} son iguales, entonces $\mathbf{x} = \mathbf{y}$.

Ejemplo 1.1.2:

Sea

$$V = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2 \mid y = mx, \text{ con } m \in \mathbb{R} \text{ fijo} \right\}$$

con la suma y multiplicación por un escalar definidas en \mathbb{R}^2 . Notemos que la expresión $y = mx$ representa una recta en el plano cartesiano con pendiente m que pasa por el origen, donde x y y son las coordenadas de un punto y m es una constante. Este conjunto admite una interpretación geométrica (vea la figura 1.5). Observemos que

$$V = \left\{ \begin{pmatrix} x \\ mx \end{pmatrix} \in \mathbb{R}^2 \mid m \in \mathbb{R} \text{ fijo} \right\}.$$

Verifiquemos entonces, los diez axiomas de un espacio vectorial. Así:

i) Sea $\mathbf{x}, \mathbf{y} \in V$ con $\mathbf{x} = \begin{pmatrix} x \\ mx \end{pmatrix}$, $\mathbf{y} = \begin{pmatrix} y \\ my \end{pmatrix}$, entonces

$$\begin{aligned} \mathbf{x} + \mathbf{y} &= \begin{pmatrix} x \\ mx \end{pmatrix} + \begin{pmatrix} y \\ my \end{pmatrix} \\ &= \begin{pmatrix} x + y \\ mx + my \end{pmatrix} && \text{por def. de suma} \\ &= \begin{pmatrix} x + y \\ m(x + y) \end{pmatrix} && \text{por distributividad en } \mathbb{R} \\ &= \begin{pmatrix} \chi \\ m\chi \end{pmatrix} \in V && \text{siendo } \chi = x + y \end{aligned}$$

Por tanto, se cumple la propiedad de cerradura.

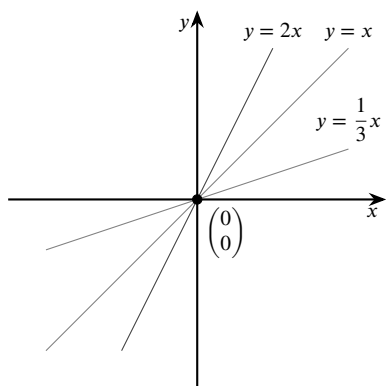


Figura 1.5 Representación geométrica de algunos ejemplos del conjunto V

ii) Sea $x, y, z \in V$ con $x = \begin{pmatrix} x \\ mx \end{pmatrix}$, $y = \begin{pmatrix} y \\ my \end{pmatrix}$, $z = \begin{pmatrix} z \\ mz \end{pmatrix}$, entonces

$$\begin{aligned}
 x + (y + z) &= \begin{pmatrix} x \\ mx \end{pmatrix} + \left[\begin{pmatrix} y \\ my \end{pmatrix} + \begin{pmatrix} z \\ mz \end{pmatrix} \right] \\
 &= \begin{pmatrix} x \\ mx \end{pmatrix} + \begin{pmatrix} y + z \\ my + mz \end{pmatrix} && \text{por def. de suma} \\
 &= \begin{pmatrix} x + (y + z) \\ mx + (my + mz) \end{pmatrix} && \text{por def. de suma} \\
 &= \begin{pmatrix} (x + y) + z \\ (mx + my) + mz \end{pmatrix} && \text{por asociatividad en } \mathbb{R} \\
 &= \begin{pmatrix} x + y \\ mx + my \end{pmatrix} + \begin{pmatrix} z \\ mz \end{pmatrix} && \text{por def. de suma} \\
 &= \left[\begin{pmatrix} x \\ mx \end{pmatrix} + \begin{pmatrix} y \\ my \end{pmatrix} \right] + \begin{pmatrix} z \\ mz \end{pmatrix} && \text{por def. de suma} \\
 &= (x + y) + z
 \end{aligned}$$

Por tanto, se cumple la asociatividad.

iii) Sea $x, y \in V$ con $x = \begin{pmatrix} x \\ mx \end{pmatrix}$, $y = \begin{pmatrix} y \\ my \end{pmatrix}$, entonces

$$\begin{aligned}
 x + y &= \begin{pmatrix} x \\ mx \end{pmatrix} + \begin{pmatrix} y \\ my \end{pmatrix} \\
 &= \begin{pmatrix} x + y \\ mx + my \end{pmatrix} && \text{por def. de suma} \\
 &= \begin{pmatrix} y + x \\ my + mx \end{pmatrix} && \text{por conmutatividad en } \mathbb{R} \\
 &= \begin{pmatrix} y \\ my \end{pmatrix} + \begin{pmatrix} x \\ mx \end{pmatrix} && \text{por def. de suma} \\
 &= y + x
 \end{aligned}$$

Por tanto, se cumple la propiedad de conmutatividad.

iv) Existe $0 = \begin{pmatrix} 0 \\ m \cdot 0 \end{pmatrix} \in V$ tal que

$$x + 0 = \begin{pmatrix} x \\ mx \end{pmatrix} + \begin{pmatrix} 0 \\ m \cdot 0 \end{pmatrix} = x.$$

El elemento 0 pertenece al conjunto V porque cumple con la condición que define a los elementos de V , es decir, que la segundo componente sea igual a m multiplicado por el primero. Además, 0 actúa como el neutro aditivo porque,

$$\begin{aligned}
 x + 0 &= \begin{pmatrix} x \\ mx \end{pmatrix} + \begin{pmatrix} 0 \\ m \cdot 0 \end{pmatrix} \\
 &= \begin{pmatrix} x + 0 \\ mx + m \cdot 0 \end{pmatrix} && \text{por def. de suma} \\
 &= \begin{pmatrix} x \\ mx + 0 \end{pmatrix} && \text{por propiedad en } \mathbb{R} \\
 &= \begin{pmatrix} x \\ mx \end{pmatrix} && \text{por neutro aditivo en } \mathbb{R} \\
 &= x
 \end{aligned}$$

Por lo tanto, 0 existe en V y cumple con la propiedad requerida de ser el elemento neutro para la suma en este espacio.

v) Para cada $\mathbf{x} = \begin{pmatrix} x \\ mx \end{pmatrix} \in V$, existe $-\mathbf{x} = \begin{pmatrix} -x \\ m(-x) \end{pmatrix} \in V$ tal que

$$\mathbf{x} + (-\mathbf{x}) = \begin{pmatrix} x \\ mx \end{pmatrix} + \begin{pmatrix} -x \\ m(-x) \end{pmatrix} = \mathbf{0}.$$

Es claro que el elemento $-\mathbf{x}$ pertenece al conjunto V porque cumple con la condición que define a los elementos de V , es decir, que el segundo componente sea igual a m multiplicado por el primero. En este caso, para $-\mathbf{x}$, el segundo componente es $m(-x)$. Además, $-\mathbf{x}$ actúa como el opuesto aditivo de \mathbf{x} , ya que,

$$\begin{aligned} \mathbf{x} + (-\mathbf{x}) &= \begin{pmatrix} x \\ mx \end{pmatrix} + \begin{pmatrix} -x \\ m(-x) \end{pmatrix} \\ &= \begin{pmatrix} x + (-x) \\ mx + m(-x) \end{pmatrix} && \text{por def. de suma} \\ &= \begin{pmatrix} x - x \\ m(x + (-x)) \end{pmatrix} && \text{por distributividad en } \mathbb{R} \\ &= \begin{pmatrix} 0 \\ m \cdot 0 \end{pmatrix} && \text{por inv. aditivo en } \mathbb{R} \\ &= \mathbf{0} \end{aligned}$$

Por tanto, se cumple la propiedad del inverso aditivo.

vi) Sea $\alpha \in \mathbb{R}$ y sea $\mathbf{x} \in V$ con $\mathbf{x} = \begin{pmatrix} x \\ mx \end{pmatrix}$, entonces

$$\begin{aligned} \alpha \cdot \mathbf{x} &= \alpha \cdot \begin{pmatrix} x \\ mx \end{pmatrix} \\ &= \begin{pmatrix} \alpha x \\ \alpha mx \end{pmatrix} && \text{por def. de producto} \\ &= \begin{pmatrix} (\alpha x) \\ m(\alpha x) \end{pmatrix} && \text{por asociatividad en } \mathbb{R} \\ &= \begin{pmatrix} \xi \\ m\xi \end{pmatrix} \in V && \text{siendo } \xi = \alpha x \end{aligned}$$

Por tanto, se cumple la propiedad de cerradura.

vii) Se deja como ejercicio al lector.

viii) Sea $\alpha \in \mathbb{R}$ y sean $\mathbf{x}, \mathbf{y} \in V$ con $\mathbf{x} = \begin{pmatrix} x \\ mx \end{pmatrix}$ e $\mathbf{y} = \begin{pmatrix} y \\ my \end{pmatrix}$, entonces

$$\begin{aligned} \alpha \cdot (\mathbf{x} + \mathbf{y}) &= \alpha \cdot \left[\begin{pmatrix} x \\ mx \end{pmatrix} + \begin{pmatrix} y \\ my \end{pmatrix} \right] \\ &= \alpha \cdot \begin{pmatrix} x + y \\ mx + my \end{pmatrix} && \text{por def. de suma} \\ &= \begin{pmatrix} \alpha(x + y) \\ \alpha(mx + my) \end{pmatrix} && \text{por def. de producto} \\ &= \begin{pmatrix} \alpha x + \alpha y \\ \alpha mx + \alpha my \end{pmatrix} && \text{por distributividad en } \mathbb{R} \\ &= \begin{pmatrix} \alpha x \\ \alpha mx \end{pmatrix} + \begin{pmatrix} \alpha y \\ \alpha my \end{pmatrix} && \text{por def. de suma} \\ &= \alpha \cdot \begin{pmatrix} x \\ mx \end{pmatrix} + \alpha \cdot \begin{pmatrix} y \\ my \end{pmatrix} && \text{por def. de producto} \\ &= \alpha \cdot \mathbf{x} + \alpha \cdot \mathbf{y} \end{aligned}$$

Por tanto, se cumple la distributividad con dos vectores y un escalar.

ix) Sea $\alpha, \beta \in \mathbb{R}$ y sea $\mathfrak{x} \in V$ con $\mathfrak{x} = \begin{pmatrix} x \\ mx \end{pmatrix}$, entonces

$$\begin{aligned}
 (\alpha + \beta) \cdot \mathfrak{x} &= (\alpha + \beta) \cdot \begin{pmatrix} x \\ mx \end{pmatrix} \\
 &= \begin{pmatrix} (\alpha + \beta) \cdot x \\ (\alpha + \beta) \cdot mx \end{pmatrix} && \text{por def. de producto} \\
 &= \begin{pmatrix} \alpha x + \beta x \\ \alpha mx + \beta mx \end{pmatrix} && \text{por distributividad en } \mathbb{R} \\
 &= \begin{pmatrix} \alpha x \\ \alpha mx \end{pmatrix} + \begin{pmatrix} \beta x \\ \beta mx \end{pmatrix} && \text{por def. de suma} \\
 &= \alpha \cdot \mathfrak{x} + \beta \cdot \mathfrak{x} && \text{por def. de producto}
 \end{aligned}$$

Por tanto, se cumple la distributividad con dos escalares y un vector.

x) Sea $\mathfrak{x} \in V$ con $\mathfrak{x} = \begin{pmatrix} x \\ mx \end{pmatrix}$, entonces

$$\begin{aligned}
 1 \cdot \mathfrak{x} &= 1 \cdot \begin{pmatrix} x \\ mx \end{pmatrix} \\
 &= \begin{pmatrix} 1 \cdot x \\ 1 \cdot mx \end{pmatrix} && \text{por def. de producto} \\
 &= \begin{pmatrix} x \\ mx \end{pmatrix} && \text{por identidad multiplicativa en } \mathbb{R} \\
 &= \mathfrak{x}
 \end{aligned}$$

Por tanto, se cumple la propiedad de identidad multiplicativa.

Por tanto, V es un espacio vectorial sobre \mathbb{R} .

Ejemplo 1.1.3:

Sea

$$U = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2 \mid y = mx + 1, \text{ con } m \in \mathbb{R} \text{ fijo} \right\}$$

con las operaciones de suma y multiplicación por un escalar definidas en \mathbb{R}^2 . Demostremos que U no es un espacio vectorial sobre \mathbb{R} . Observemos que

$$U = \left\{ \begin{pmatrix} x \\ mx + 1 \end{pmatrix} \in \mathbb{R}^2 \mid m \in \mathbb{R} \text{ fijo} \right\}$$

Así, se sigue que:

i) Sea $\mathfrak{x}, \mathfrak{y} \in V$ con $\mathfrak{x} = \begin{pmatrix} x \\ mx + 1 \end{pmatrix}$, $\mathfrak{y} = \begin{pmatrix} y \\ my + 1 \end{pmatrix}$, entonces

$$\begin{aligned}
 \mathfrak{x} + \mathfrak{y} &= \begin{pmatrix} x \\ mx + 1 \end{pmatrix} + \begin{pmatrix} y \\ my + 1 \end{pmatrix} \\
 &= \begin{pmatrix} x + y \\ mx + 1 + my + 1 \end{pmatrix} && \text{por def. de suma} \\
 &= \begin{pmatrix} x + y \\ m(x + y) + 2 \end{pmatrix} && \text{por distributividad en } \mathbb{R} \\
 &= \begin{pmatrix} \chi \\ m\chi + 2 \end{pmatrix} \notin U && \text{siendo } \chi = x + y
 \end{aligned}$$

Por tanto, no se cumple la propiedad de cerradura.

Por tanto, U no es un espacio vectorial sobre \mathbb{R} . De igual forma al anterior ejemplo, U admite una interpretación geométrica sencilla, vea la figura 1.6.

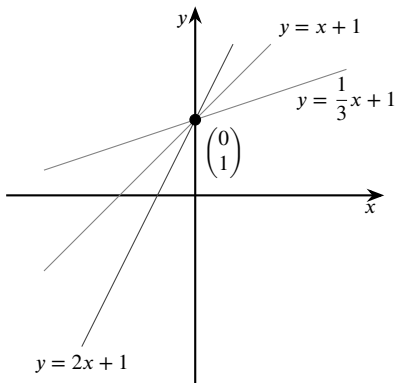


Figura 1.6 Representación geométrica de algunos ejemplos del conjunto U

El siguiente ejemplo es el espacio vectorial de funciones reales en $(-\infty, \infty)$.

Ejemplo 1.1.4:

Sea V el conjunto de funciones con valores reales definidas en cada x en el intervalo $(-\infty, \infty)$. Sean $f = f(x)$ y $g = g(x)$ dos funciones en V y α un escalar, definimos

$$(f + g)(x) = f(x) + g(x) \quad (1.1.18)$$

y para α ,

$$(\alpha \cdot f)(x) = \alpha f(x). \quad (1.1.19)$$

Una forma de interpretar estas operaciones es ver los valores $f(x)$ y $g(x)$ como las “componentes” de f y g en el punto x . En este caso, las ecuaciones (1.1.18) y (1.1.19) establecen que dos funciones se suman agregando sus componentes correspondientes, y que una función se multiplica por un escalar multiplicando cada componente por dicho escalar, exactamente como en \mathbb{R}^n . Esta idea se ilustra en las partes (a) y (b) de la figura 1.7. El conjunto V con estas operaciones se denota por el símbolo $F(-\infty, \infty)$. Podemos demostrar que este es un espacio vectorial de la siguiente manera:

- i) El axioma de cerradura para la suma requiere que si sumamos dos funciones definidas en cada x en el intervalo $(-\infty, \infty)$, entonces la suma de esas funciones también deben estar definidas en cada x en el intervalo $(-\infty, \infty)$. Esto se sigue de la fórmula (1.1.18).

- ii) Sea $f, g, h \in V$ con $f = f(x)$, $g = g(x)$, $h = h(x)$,

$$\begin{aligned} [f + (g + h)](x) &= f(x) + (g(x) + h(x)) \\ &= (f(x) + g(x)) + h(x) \\ &= [(f + g) + h](x) \end{aligned}$$

Por tanto, se cumple la asociatividad sobre la suma de funciones.

- iii) Sea $f, g \in V$ con $f = f(x)$, $g = g(x)$,

$$\begin{aligned} (f + g)(x) &= f(x) + g(x) \\ &= g(x) + f(x) \\ &= (g + f)(x) \end{aligned}$$

Por tanto, se cumple la propiedad de conmutatividad.

- iv) Existe una función 0 en $F(-\infty, \infty)$, la cual, al sumarse con cualquier otra función f en $F(-\infty, \infty)$, devuelve f como resultado. La función cuyo valor en cada punto x del intervalo $(-\infty, \infty)$ es cero cumple con esta propiedad. Geométricamente, el gráfico de la función 0 es la línea que coincide con el eje x . De esta forma, si tenemos $f, 0 \in V$ con $f = f(x)$, $0 = 0$,

$$\begin{aligned} (f + 0)(x) &= f(x) + 0 \\ &= f(x) \\ &= f \end{aligned}$$

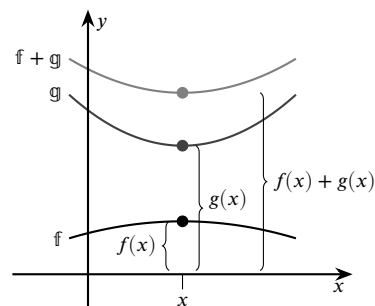
Por tanto, se cumple la propiedad de neutro aditivo.

- v) Para cada función f en $F(-\infty, \infty)$ existe una función $-f$ en $F(-\infty, \infty)$, la cual, al sumarse con f , produce la función 0 . La función definida por $-f(x) = -f(x)$ cumple con esta propiedad. Geométricamente, el gráfico de $-f$ se obtiene reflejando el de f con respecto al eje x (figura 1.7c). Así,

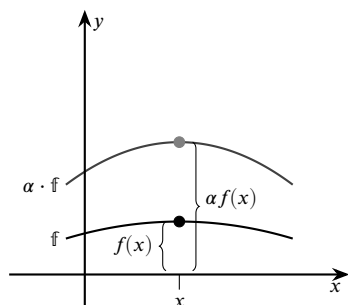
$$\begin{aligned} [f + (-f)](x) &= f(x) + (-f(x)) \\ &= f(x) - f(x) \\ &= 0 \\ &= 0 \end{aligned}$$

Por tanto, se cumple la propiedad de inverso aditivo.

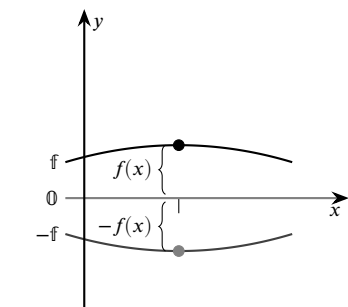
- vi) Es análogo al inciso (i) y esto se sigue de la fórmula (1.1.19).



(a)



(b)



(c)

Figura 1.7

vii) Sea $f, g \in V$ con $f = f(x)$, $g = g(x)$ y sea $\alpha \in \mathbb{R}$,

$$\begin{aligned} [\alpha \cdot (f + g)](x) &= \alpha(f(x) + g(x)) \\ &= \alpha f(x) + \alpha g(x) \\ &= (\alpha \cdot f)(x) + (\alpha \cdot g)(x) \\ &= (\alpha \cdot f + \alpha \cdot g)(x) \end{aligned}$$

Por tanto, se cumple la distributividad sobre la suma de vectores.

viii) Sea $f \in V$ con $f = f(x)$ y sea $\alpha, \beta \in \mathbb{R}$,

$$\begin{aligned} [(\alpha + \beta) \cdot f](x) &= (\alpha + \beta)f(x) \\ &= \alpha f(x) + \beta f(x) \\ &= (\alpha \cdot f)(x) + (\beta \cdot f)(x) \\ &= (\alpha \cdot f + \beta \cdot f)(x) \end{aligned}$$

Por tanto, se cumple la distributividad sobre la suma de escalares.

ix) Se deja como ejercicio al lector.

x) Sea $f \in V$ con $f = f(x)$ y sea $1 \in \mathbb{R}$,

$$\begin{aligned} (1 \cdot f)(x) &= 1f(x) \\ &= f(x) \\ &= f \end{aligned}$$

Por tanto, se cumple la propiedad de identidad multiplicativa.

Dado que $F(-\infty, \infty)$ satisface todas las propiedades requeridas para ser un espacio vectorial, concluimos que $F(-\infty, \infty)$ es un espacio vectorial sobre \mathbb{R} .

Es crucial entender que no se pueden imponer arbitrariamente dos operaciones en un conjunto V y esperar que se cumplan los axiomas de espacio vectorial. Por ejemplo, si V es el conjunto de vectores de \mathbb{R}^n con componentes positivas y se usan las operaciones estándar de \mathbb{R}^n , entonces V no es cerrado bajo la multiplicación por escalares, ya que $(-1) \cdot u$ es un vector con al menos un componente negativo, que no pertenecería a V .

El siguiente es un ejemplo menos evidente en el que solo uno de los diez axiomas de espacio vectorial deja de cumplirse.

Ejemplo 1.1.5:

Sea $V = \mathbb{R}^2$ y definamos las operaciones de suma y producto escalar en V de la siguiente manera: Si $u = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$, $v = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$ y si α es un número real cualquiera,

$$u + v = \begin{pmatrix} u_1 + v_1 \\ u_2 + v_2 \end{pmatrix} \quad \text{y} \quad \alpha \cdot u = \begin{pmatrix} \alpha u_1 \\ 0 \end{pmatrix}.$$

Por ejemplo, si $u = \begin{pmatrix} 2 \\ 4 \end{pmatrix}$, $v = \begin{pmatrix} -3 \\ 5 \end{pmatrix}$ y $\alpha = 7$, entonces

$$u + v = \begin{pmatrix} 2 + (-3) \\ 4 + 5 \end{pmatrix} = \begin{pmatrix} -1 \\ 9 \end{pmatrix} \quad \text{y} \quad \alpha \cdot u = 7 \cdot u = \begin{pmatrix} 7 \cdot 2 \\ 0 \end{pmatrix} = \begin{pmatrix} 14 \\ 0 \end{pmatrix}.$$

La operación de suma es la estándar en \mathbb{R}^2 , pero el producto escalar no lo es. Los primeros nueve axiomas se cumplen, sin embargo, el axioma (x) no se cumple para ciertos vectores. Por ejemplo, si u es tal que $u_2 \neq 0$, entonces

$$1 \cdot u = 1 \cdot \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} 1 \cdot u_1 \\ 0 \end{pmatrix} = \begin{pmatrix} u_1 \\ 0 \end{pmatrix} \neq u.$$

Por lo tanto, V no es un espacio vectorial con las operaciones establecidas.

Nuestro siguiente ejemplo será un espacio vectorial inusual que hemos incluido para ilustrar la variedad de espacios vectoriales.

Ejemplo 1.1.6:

Sea V el conjunto de los números reales positivos, y sean $u = u$ y $v = v$ cualesquiera vectores (es decir, números reales positivos) en V . Sea α un escalar cualquiera. Definimos las operaciones en V como

$$u + v = uv \quad \text{y} \quad \alpha \cdot u = u^\alpha.$$

Por ejemplo, $1 + 1 = (1)(1) = 1$ y $(2)(1) = 1^2 = 1$, lo cual resulta extraño, pero, no obstante, el conjunto V con estas operaciones satisface los diez axiomas de los espacios vectoriales y, por lo tanto, es un espacio vectorial. Verificaremos algunos axiomas, y dejaremos los demás como ejercicio al lector.

i) Dado que $u = u$ y $v = v$ son números reales positivos, el producto uv es también un número real positivo, ya que el producto de dos números reales positivos siempre da como resultado otro número real positivo.

ii) Sea $u, v \in V$ con $u = u$, $v = v$,

$$\begin{aligned} u + v &= uv \\ &= vu \\ &= v + u \end{aligned}$$

Por tanto, se cumple la propiedad de conmutatividad.

iii) Se deja como ejercicio al lector.

iv) Sea $u, 0 \in V$ con $u = u$, $0 = 1$,

$$\begin{aligned} u + 0 &= u^1 \\ &= u \\ &= u \end{aligned}$$

Por tanto, se cumple la propiedad de neutro aditivo.

v) Dado $u = u$, existe $-u = \frac{1}{u}$ que satisface

$$\begin{aligned} u + (-u) &= u \left(\frac{1}{u} \right) \\ &= 1 \\ &= 0 \end{aligned}$$

Por tanto, se cumple la propiedad de inverso aditivo.

vi) Dado que $u = u$ es un número real positivo, y α es un número real cualquiera, la operación u^α es también un número real positivo.

vii) Sea $u, v \in V$ con $u = u$, $v = v$ y sea $\alpha \in \mathbb{R}$,

$$\begin{aligned} \alpha \cdot (u + v) &= (uv)^\alpha \\ &= u^\alpha v^\alpha \\ &= \alpha \cdot u + \alpha \cdot v \end{aligned}$$

Por tanto, se cumple la distributividad sobre la suma de vectores.

viii) Sea $u \in V$ con $u = u$ y sea $\alpha, \beta \in \mathbb{R}$,

$$\begin{aligned} (\alpha + \beta) \cdot u &= u^{\alpha+\beta} \\ &= u^\alpha u^\beta \\ &= \alpha \cdot u + \beta \cdot u \end{aligned}$$

Por tanto, se cumple la distributividad sobre la suma de escalares.

ix) Se deja como ejercicio al lector.

x) Se deja como ejercicio al lector.

En conclusión, aunque estas operaciones resulten inusuales, la verificación de los diez axiomas, confirma que esta estructura cumple con todos los requisitos para ser considerada un espacio vectorial sobre \mathbb{R} .

El siguiente ejemplo es el espacio vectorial de los polinomios con coeficientes reales, que cumple con los axiomas correspondientes.

Ejemplo 1.1.7: El conjunto de polinomios reales de grado menor o igual a n se define como

$$P_n = \{a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0 \mid a_i \in \mathbb{R}, \text{ para } i = 0, 1, 2, \dots, n\}$$

y se expresan los elementos como

$$p(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$$

En P_n , definimos dos operaciones: Dado $p(x) \in P_n$ y $\alpha \in \mathbb{R}$,

$$\begin{aligned} +: \quad P_n \times P_n &\longrightarrow P_n \\ (p(x), q(x)) &\longmapsto p(x) + q(x) \end{aligned}$$

$$\begin{aligned} \cdot: \quad \mathbb{R} \times P_n &\longrightarrow P_n \\ (\alpha, p(x)) &\longmapsto \alpha \cdot p(x) \end{aligned}$$

donde

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

y

$$\alpha \cdot p(x) = \alpha a_n x^n + \alpha a_{n-1} x^{n-1} + \cdots + \alpha a_1 x + \alpha a_0.$$

Entonces P_n es un espacio vectorial sobre \mathbb{R} . Es obvio que la suma de dos polinomios de grado menor o igual a n es otro polinomio de grado menor o igual a n , por lo que se cumple el axioma (i). Las propiedades (ii) y (v) a (x) son claras. Si se define el polinomio

$$\mathbb{0} = 0x^n + 0x^{n-1} + \cdots + 0x + 0,$$

entonces $\mathbb{0} \in P_n$ y el axioma (iii) se cumple. Por último, sea

$$-p(x) = -a_n x^n - a_{n-1} x^{n-1} - \cdots - a_1 x - a_0;$$

se ve que el axioma (iv) se cumple, con lo que P_n es un espacio vectorial real.

Nuestro último ejemplo es el más sencillo, ya que corresponde al espacio vectorial cero. A diferencia de los ejemplos anteriores, más complejos, este caso se verifica de manera trivial, ya que solo contiene el vector nulo.

Ejemplo 1.1.8: Sea W un conjunto que consta de un único elemento, $\mathbb{0}$, y definimos

$$\mathbb{0} + \mathbb{0} = \mathbb{0} \quad \text{y} \quad \alpha \cdot \mathbb{0} = \mathbb{0}$$

para todo escalar α . A esto lo llamamos el *espacio vectorial cero*.

1.2 Subespacios vectoriales

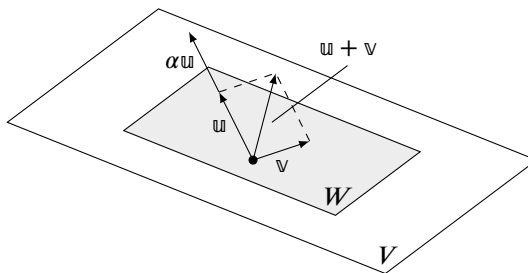
A menudo ocurre que algún espacio vectorial de interés está contenido dentro de un espacio vectorial más grande cuyas propiedades son conocidas. En esta sección mostraremos cómo reconocer cuándo ocurre esto, explicaremos cómo las propiedades del espacio vectorial más grande pueden usarse para obtener propiedades del espacio vectorial más pequeño, y proporcionaremos una variedad de ejemplos.

Definición 1.2.1:

Se dice que W es un *subespacio vectorial* de V si W es un subconjunto no vacío de V , y W es un espacio vectorial, junto con las operaciones de suma entre vectores y multiplicación por un escalar definidas para V .

En general, para demostrar que un conjunto no vacío W con dos operaciones es un espacio vectorial, se deben verificar los diez axiomas de espacio vectorial. Sin embargo, si W es un subespacio de un espacio vectorial conocido V , entonces ciertos axiomas no necesitan ser verificados porque son “heredados” de V . Por ejemplo, no es necesario verificar que $u + v = v + u$ se cumple en W porque se cumple para todos los vectores en V , incluidos los de W . Por otro lado, es necesario verificar que W está cerrado bajo la suma y la multiplicación escalar, ya que es posible que sumar dos vectores en W o multiplicar un vector en W por un escalar produzca un vector en V que no esté en W (figura 1.8).

Figura 1.8 Los vectores u y v están en W , pero los vectores $u + v$ y αu no lo están



Los axiomas que no son heredados por W son:

- **Axioma 1:** Cerradura de W bajo la suma.
- **Axioma 4:** Existencia de un vector cero en W .
- **Axioma 5:** Existencia de un negativo en W para cada vector en W .
- **Axioma 6:** Cerradura de W bajo la multiplicación escalar.

Por lo tanto, estos deben ser verificados para probar que W es un subespacio de V . Sin embargo, el siguiente teorema muestra que si los axiomas 1 y 6 se cumplen en W , entonces los axiomas 4 y 5 se cumplen en W como consecuencia.

Teorema 1.2.1:

Si W es un conjunto de uno o más vectores en un espacio vectorial V , entonces W es un subespacio de V si y solo si se satisfacen las siguientes condiciones:

- a) Si u y v son vectores en W , entonces $u + v$ está en W .
- b) Si α es un escalar y u es un vector en W , entonces $\alpha \cdot u$ está en W .

Demostración: Por definición, un subespacio es un espacio vectorial cuyos elementos pertenecen a V y cuyas operaciones son las mismas que en V . Como V es un espacio vectorial, se tiene que W está cerrado bajo la suma y el producto escalar, que corresponden exactamente a las condiciones (a) y (b). Recíprocamente, supongamos que se satisfacen las condiciones (a) y (b). Dado que los axiomas (i), (ii), (iii), (vii), (viii), (ix) y (x) son heredados de V , solo necesitamos demostrar que los axiomas (iv) y (v) se cumplen en W . Para ello, sea u un vector cualquiera en W . De la condición (b) se sigue que $\alpha \cdot u$ es un vector en W para cualquier escalar α . En particular, $0 \cdot u = 0$ y $(-1) \cdot u = -u$ están en W , lo que muestra que los axiomas (iv) y (v) se cumplen en W .

Ejemplo 1.2.1:

Si V es un espacio vectorial y $W = \{0\}$ es el subconjunto de V que consiste únicamente en el vector cero, entonces W está cerrado bajo la suma y la multiplicación por un escalar, ya que

$$0 + 0 = 0 \quad \text{y} \quad \alpha \cdot 0 = 0$$

para cualquier escalar α . A W lo llamamos el *subespacio cero* de V .

Ejemplo 1.2.2: Si W es una recta que pasa por el origen en \mathbb{R}^2 o \mathbb{R}^3 , entonces la suma de dos vectores en la recta o la multiplicación de un vector en la recta por un escalar produce otro vector en la recta. Por lo tanto, W está cerrado bajo la suma y el producto escalar (véase la figura 1.9 para una ilustración en \mathbb{R}^3).

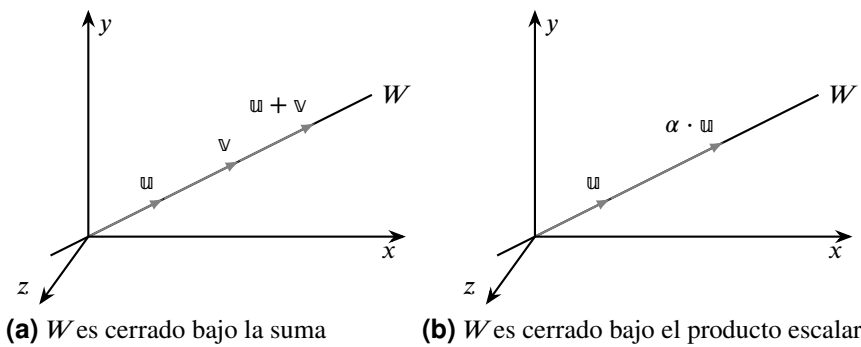


Figura 1.9 Los vectores $u + v$ y $\alpha \cdot u$ están en la misma recta que u y v

Ejemplo 1.2.3: Si u y v son vectores en un plano W que pasa por el origen en \mathbb{R}^3 , entonces es evidente geoméricamente que $u + v$ y $\alpha \cdot u$ también pertenecen al mismo plano W para cualquier escalar α (figura 1.10). Por lo tanto, W está cerrado bajo la suma y el producto escalar.

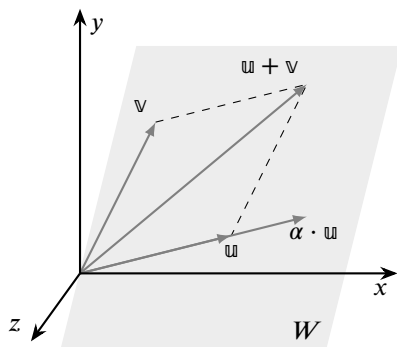


Figura 1.10 Los vectores $u + v$ y $\alpha \cdot u$ están en el mismo plano que u y v

La tabla 1.1 a continuación presenta una lista de los subespacios de \mathbb{R}^2 y \mathbb{R}^3 que hemos encontrado hasta ahora.

Tabla 1.1 Subespacios vectoriales de \mathbb{R}^2 y \mathbb{R}^3

Subespacios de \mathbb{R}^2	Subespacios de \mathbb{R}^3
<ul style="list-style-type: none">• $\{0\}$• Rectas que pasan por el origen• \mathbb{R}^2	<ul style="list-style-type: none">• $\{0\}$• Rectas que pasan por el origen• Planos que pasan por el origen• \mathbb{R}^3

En los ejemplos anteriores, hemos visto subespacios vectoriales en los espacios \mathbb{R}^2 y \mathbb{R}^3 . Estos subespacios son fundamentales para comprender las propiedades estructurales de los espacios vectoriales. A continuación, daremos un ejemplo que ilustra cómo el conjunto de funciones continuas, bajo ciertas operaciones, también forma un subespacio.

Ejemplo 1.2.4: Existe un teorema en cálculo que establece que la suma de funciones continuas es continua y que una constante multiplicada por una función continua también es continua. En lenguaje vectorial, el conjunto de funciones continuas en $(-\infty, \infty)$ es un subespacio de $F(-\infty, \infty)$. Denotaremos este subespacio como $C(-\infty, \infty)$.

Ejemplo 1.2.5: Una función con una derivada continua se dice que es *continuamente diferenciable*. Existe un teorema en cálculo que establece que la suma de dos funciones continuamente diferenciables es continuamente diferenciable y que una constante multiplicada por una función continuamente diferenciable también es continuamente diferenciable. Por lo tanto, las funciones que son continuamente diferenciables en $(-\infty, \infty)$ forman un subespacio de $F(-\infty, \infty)$. Denotaremos este subespacio por $C^1(-\infty, \infty)$, donde el superíndice enfatiza que las primeras derivadas son continuas. Para llevar esto un paso más allá, el conjunto de funciones con m derivadas continuas en $(-\infty, \infty)$ es un subespacio de $F(-\infty, \infty)$, al igual que el conjunto de funciones con derivadas de todos los órdenes en $(-\infty, \infty)$. Denotaremos estos subespacios por $C^m(-\infty, \infty)$ y $C^\infty(-\infty, \infty)$, respectivamente.

Ejemplo 1.2.6: Recordemos que un *polinomio* es una función que puede expresarse en la forma

$$p(x) = a_0 + a_1x + \cdots + a_nx^n$$

donde a_0, a_1, \dots, a_n son constantes. Es evidente que la suma de dos polinomios es un polinomio y que una constante multiplicada por un polinomio también es un polinomio. Por lo tanto, el conjunto W de todos los polinomios es un subespacio de $F(-\infty, \infty)$. Este espacio se denota por P_∞ .

Ejemplo 1.2.7: Recordemos que el grado de un polinomio es la mayor potencia de la variable que aparece con un coeficiente distinto de cero. No es cierto que el conjunto W de polinomios de grado exacto n sea un subespacio de $F(-\infty, \infty)$ porque ese conjunto no está cerrado bajo la suma. Por ejemplo, los polinomios

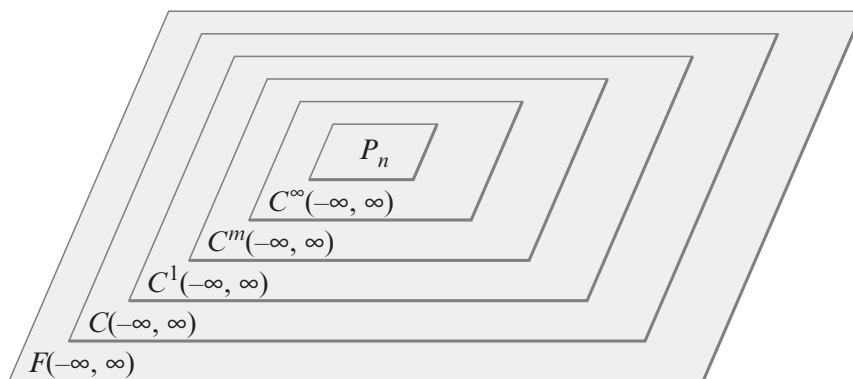
$$1 + 2x + 3x^2 \quad \text{y} \quad 5 + 7x - 3x^2$$

tienen grado 2, pero su suma tiene grado 1. Sin embargo, para cada entero no negativo n , el conjunto de polinomios de grado menor o igual a n forma un subespacio de $F(-\infty, \infty)$, ya que está cerrado bajo la suma y multiplicación por un escalar.

En nuestros ejemplos anteriores consideramos funciones que estaban definidas en todos los puntos del intervalo $(-\infty, \infty)$. A veces queremos considerar funciones que están definidas únicamente en algún subintervalo de $(-\infty, \infty)$, como el intervalo cerrado $[a, b]$ o el intervalo abierto (a, b) . En tales casos, realizaremos un cambio de notación apropiado. Por ejemplo, $C[a, b]$ es el espacio de funciones continuas en $[a, b]$, y $C(a, b)$ es el espacio de funciones continuas en (a, b) .

En cálculo se demuestra que los polinomios son funciones continuas y tienen derivadas continuas de todos los órdenes en $(-\infty, \infty)$. Por lo tanto, se deduce que P_∞ no solo es un subespacio de $F(-\infty, \infty)$, como se observó anteriormente, sino que también es un subespacio de $C^\infty(-\infty, \infty)$. Dejamos al lector la tarea de convencerse de que los espacios vectoriales discutidos en los ejemplos 1.2.4 a 1.2.7 están “anidados” unos dentro de otros, como se ilustra en la figura 1.11.

Figura 1.11 Jerarquía de espacios funcionales por diferenciabilidad



Teorema 1.2.2: Si W_1, W_2, \dots, W_r son subespacios de un espacio vectorial V , entonces la intersección de estos subespacios también es un subespacio de V .

Demostración: Sea $W = W_1 \cap W_2 \cap \dots \cap W_r$. Este conjunto no es vacío porque cada uno de estos subespacios contiene el vector cero de V , y por lo tanto, también lo hace su intersección. Así, queda por demostrar que W está cerrado bajo la suma y la multiplicación por escalares. Para demostrar la cerradura bajo la suma, sean u y v vectores en W . Dado que W es la intersección de W_1, W_2, \dots, W_r , se deduce que u y v también pertenecen a cada uno de estos subespacios. Además, como estos subespacios están cerrados bajo la suma y la multiplicación por escalares, también contienen los vectores $u + v$ y $\alpha \cdot u$ para todo escalar α , y por lo tanto, su intersección W también los contiene. Esto prueba que W está cerrado bajo la suma y la multiplicación por escalares.

1.3 Conjuntos generadores

Definición 1.3.1:

Sea V un espacio vectorial sobre K . Si u es un vector en un espacio vectorial V , entonces se dice que u es una *combinación lineal* de los vectores v_1, v_2, \dots, v_n en V si u puede expresarse en la forma

$$u = k_1 v_1 + k_2 v_2 + \dots + k_n v_n,$$

donde k_1, k_2, \dots, k_n son escalares. Estos escalares se denominan los *coeficientes* de la combinación lineal.

Si $n = 1$, entonces la expresión de la anterior definición tiene la forma

$$u = k_1 v_1,$$

en cuyo caso la combinación lineal es simplemente un múltiplo escalar de v_1 .

Teorema 1.3.1: Sea V un espacio vectorial sobre K . Si $S = \{v_1, v_2, \dots, v_n\}$ es un conjunto no vacío de vectores en V , entonces:

- El conjunto W de todas las combinaciones lineales posibles de los vectores en S es un subespacio de V .
- El conjunto W en el inciso anterior, es el subespacio más pequeño de V que contiene a todos los vectores en S , en el sentido de que cualquier otro subespacio que contenga esos vectores también contiene a W .

Demostración:

- Sea W el conjunto de todas las combinaciones lineales posibles de los vectores en S . Debemos demostrar que W está cerrado bajo la suma y la multiplicación por escalares. Para demostrar la cerradura bajo la suma, sean

$$x_1 = a_1 v_1 + a_2 v_2 + \dots + a_n v_n \quad \text{y} \quad x_2 = b_1 v_1 + b_2 v_2 + \dots + b_n v_n$$

dos vectores en W . Se deduce que su suma puede escribirse como

$$x_1 + x_2 = (a_1 + b_1)v_1 + (a_2 + b_2)v_2 + \dots + (a_n + b_n)v_n,$$

que es una combinación lineal de los vectores en S . Por lo tanto, W está cerrado bajo la suma. Se deja como ejercicio al lector la demostración de que W también está cerrado bajo la multiplicación por escalares y, por ende, es un subespacio de V .

- Sea W' cualquier subespacio de V que contiene a todos los vectores en S . Dado que W' está cerrado bajo la suma y la multiplicación por escalares, contiene todas las combinaciones lineales de los vectores en S y, por ende, contiene a W .

Definición 1.3.2:

Se dice que los vectores v_1, v_2, \dots, v_n de un espacio vectorial V *generan* a V si todo vector en V se puede escribir como una combinación lineal de los mismos. Es decir, para todo $v \in V$ existen escalares a_1, a_2, \dots, a_n tales que

$$v = a_1 v_1 + a_2 v_2 + \dots + a_n v_n.$$

Ejemplo 1.3.1:

Los vectores unitarios estándar en \mathbb{R}^n son

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \dots, e_n = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}.$$

Estos vectores generan \mathbb{R}^n ya que cualquier vector en \mathbb{R}^n se puede expresar como

$$v = v_1 e_1 + v_2 e_2 + \dots + v_n e_n$$

lo cual es una combinación lineal de e_1, e_2, \dots, e_n . Así, por ejemplo, los vectores

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, e_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

generan \mathbb{R}^3 ya que cualquier vector se puede expresar como

$$v = \begin{pmatrix} a \\ b \\ c \end{pmatrix} = a \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + b \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + c \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = a e_1 + b e_2 + c e_3.$$

Definición 1.3.3:

El *espacio generado* por $\{v_1, v_2, \dots, v_k\}$ es el conjunto de combinaciones lineales v_1, v_2, \dots, v_k . Es decir,

$$\text{gen}(\{v_1, v_2, \dots, v_k\}) = \{v \mid v = a_1 v_1 + a_2 v_2 + \dots + a_k v_k\}$$

donde a_1, a_2, \dots, a_k son escalares arbitrarios.

En las definiciones 1.3.2 y 1.3.3 se utilizaron dos términos diferentes: “genera” y “espacio generado”. Se hace hincapié en que un conjunto de vectores v_1, v_2, \dots, v_n *genera* a V (que es un verbo) si todo vector en V se puede escribir como una combinación lineal de v_1, v_2, \dots, v_n . Por otro lado, el *espacio generado* (que es un sustantivo) por los n vectores v_1, v_2, \dots, v_k es el conjunto de combinaciones lineales de estos vectores. Estos dos conceptos son diferentes, aún cuando los términos se parezcan. Por ejemplo, si consideramos $V = \mathbb{R}^3$,

- Definición 1.3.2: $\{e_1, e_2, e_3\}$ es un conjunto generador de V porque cualquier vector en \mathbb{R}^3 se puede escribir como combinación lineal de estos vectores.
- Definición 1.3.3: Si tomamos $\{e_1, e_2\}$, el espacio generado es el subespacio $\text{gen}(\{e_1, e_2\})$, que es un plano en \mathbb{R}^3 .

La diferencia está en que la definición 1.3.2 cubre el espacio completo, mientras que la 1.3.3 describe subespacios.

Ejemplo 1.3.2:

Los polinomios $1, x, x^2, \dots, x^n$ generan el espacio vectorial P_n definido en el ejemplo 1.1.7, ya que cada polinomio p en P_n puede escribirse como

$$p = a_0 + a_1 x + \dots + a_n x^n,$$

que es una combinación lineal de $1, x, x^2, \dots, x^n$. Podemos denotar esto como

$$P_n = \text{gen}(\{1, x, x^2, \dots, x^n\}).$$

Ejemplo 1.3.3: Sea $V = \mathbb{R}^3$ un espacio vectorial sobre \mathbb{R} . Por el teorema 1.3.1, se tiene que

$$W = \text{gen} \left(\left\{ \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \right\} \right) \quad (1.3.1)$$

es un subespacio de \mathbb{R}^3 . Esto implica que W está generado por las combinaciones lineales de los vectores dados. De la ecuación (1.3.1), podemos expresar W como

$$\begin{aligned} W &= \left\{ \alpha \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + \beta \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \mid \alpha, \beta \in \mathbb{R} \right\} \\ &= \left\{ \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} + \alpha \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + \beta \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \mid \alpha, \beta \in \mathbb{R} \right\} \\ &= \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3 \mid \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} \alpha \\ \alpha + \beta \\ \beta \end{pmatrix} \text{ donde } \alpha, \beta \in \mathbb{R} \right\} \end{aligned}$$

De esta forma, las coordenadas de un vector genérico en W se pueden describir en términos de los parámetros α y β como

$$\begin{aligned} x &= \alpha \\ y &= \alpha + \beta \quad \text{con } \alpha, \beta \in \mathbb{R} \\ z &= \beta \end{aligned}$$

Esto representa la forma paramétrica de un plano en \mathbb{R}^3 . Ahora, eliminemos los parámetros α y β para obtener la ecuación implícita del plano. De las expresiones para y y z , tenemos

$$y = x + z \implies x - y + z = 0.$$

Si tomamos los vectores v_1 y v_2 respectivamente que generan al subespacio W , geoméricamente tenemos que

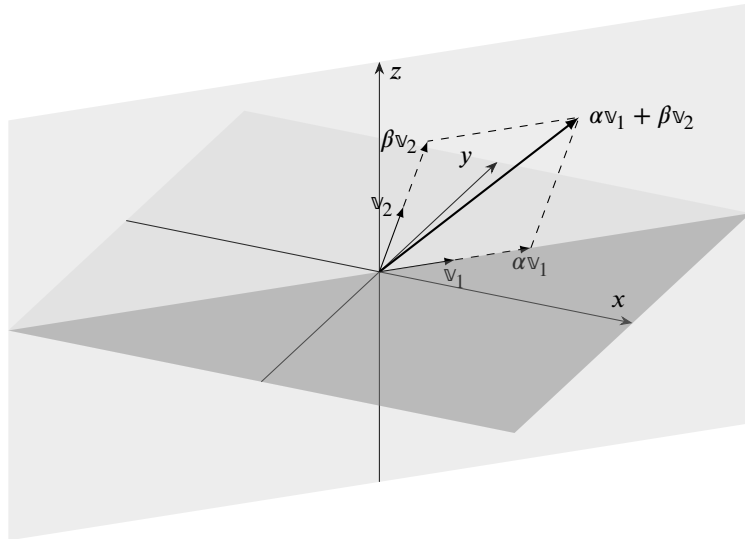


Figura 1.12 Representación del plano $x - y + z = 0$

Por lo tanto, el subespacio W es el plano en \mathbb{R}^3 definido por la ecuación implícita

$$x - y + z = 0.$$

Esto demuestra que W es efectivamente un subespacio vectorial y, además, está contenido en el plano descrito.

1.4 Conjuntos linealmente independientes

Definición 1.4.1:

Sea V un espacio vectorial sobre K . Un subconjunto S del espacio vectorial V se llama *linealmente dependiente* si existen vectores distintos v_1, v_2, \dots, v_n en S y a_1, a_2, \dots, a_n en K no todos 0 tales que

$$a_1 v_1 + a_2 v_2 + \dots + a_n v_n = 0.$$

En este caso también decimos que los vectores de S son linealmente dependientes. En caso contrario, decimos que los vectores son *linealmente independientes* y por tanto, S también es linealmente independiente.

Teorema 1.4.1:

Los vectores no nulos v_1, v_2, \dots, v_k en un espacio vectorial V , son linealmente dependientes si, y solo si uno de los vectores v_j , con $j \geq 2$, es una combinación lineal de los vectores que lo preceden v_1, v_2, \dots, v_{j-1} .

Demostración: Si v_j es una combinación lineal de v_1, v_2, \dots, v_{j-1} ,

$$v_j = c_1 v_1 + c_2 v_2 + \dots + c_{j-1} v_{j-1},$$

entonces

$$c_1 v_1 + c_2 v_2 + \dots + c_{j-1} v_{j-1} + (-1)v_j + 0v_{j+1} + \dots + 0v_n = 0.$$

Como por lo menos uno de los coeficientes es diferente de cero, v_1, v_2, \dots, v_n son linealmente dependientes. Supongamos ahora que v_1, v_2, \dots, v_n son linealmente dependientes. Entonces existen escalares c_1, c_2, \dots, c_n , no todos cero, tales que

$$c_1 v_1 + c_2 v_2 + \dots + c_n v_n = 0.$$

Sea j el mayor subíndice para el cual $c_j \neq 0$. Si $j > 1$, entonces

$$v_j = -\left(\frac{c_1}{c_j}\right)v_1 - \left(\frac{c_2}{c_j}\right)v_2 - \dots - \left(\frac{c_{j-1}}{c_j}\right)v_{j-1}.$$

Si $j = 1$, entonces $c_1 v_1 = 0$, lo que implica que $v_1 = 0$, contradiciendo la hipótesis de que ninguno de los vectores es el vector cero. Por lo tanto, uno de los vectores v_j es una combinación lineal de los vectores que lo preceden v_1, v_2, \dots, v_{j-1} .

Corolario 1.4.1:

Dos vectores en un espacio vectorial son linealmente dependientes si y solo si uno de ellos es un múltiplo escalar del otro.

Ejemplo 1.4.1:

Consideremos el espacio vectorial \mathbb{R}^3 y los vectores

$$v_1 = \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}, v_2 = \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}, v_3 = \begin{pmatrix} -3 \\ 2 \\ -1 \end{pmatrix}, v_4 = \begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix}$$

Notemos que

$$v_1 + v_2 + 0v_3 - v_4 = 0,$$

es decir, v_1, v_2, v_3 y v_4 son linealmente dependientes. En este caso,

$$v_4 = v_1 + v_2.$$

i Nota

El teorema 1.4.1 no dice que en un conjunto de vectores linealmente dependientes todo vector v del conjunto es una combinación lineal de los vectores que le preceden. En el ejemplo 1.4.1, también se cumple que $v_1 + 2v_2 + v_3 + 0v_4 = 0$. Sin embargo, no podemos despejar a v_4 para expresarlo como combinación lineal de v_1, v_2 y v_3 , puesto que su coeficiente es cero.

En otras palabras, si S es un conjunto de dos o más vectores en V , se dice que es *linealmente independiente* si ningún vector en S puede expresarse como una combinación lineal de los demás. En caso contrario, donde al menos uno de los vectores puede escribirse como combinación lineal de los demás, el conjunto S se considera *linealmente dependiente*, según el precedente teorema y corolario.

La independencia lineal tiene interpretaciones geométricas útiles en \mathbb{R}^2 y \mathbb{R}^3 . Dos vectores en \mathbb{R}^2 o \mathbb{R}^3 son linealmente independientes si y solo si no están sobre la misma recta cuando tienen sus puntos iniciales en el origen. De lo contrario, uno sería un múltiplo escalar del otro (figura 1.13).

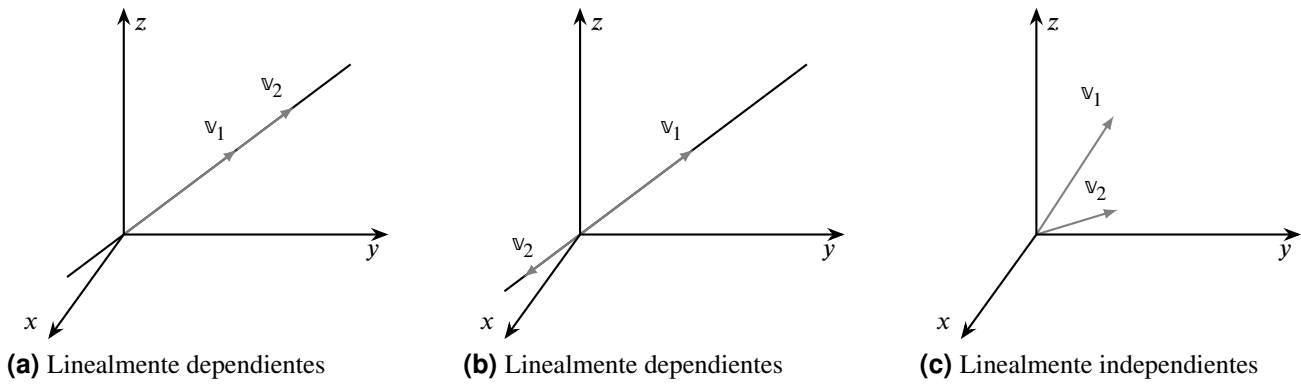


Figura 1.13

Tres vectores en \mathbb{R}^3 son linealmente independientes si y solo si no están sobre el mismo plano cuando tienen sus puntos iniciales en el origen. De lo contrario, al menos uno sería una combinación lineal de los otros dos (figura 1.14).

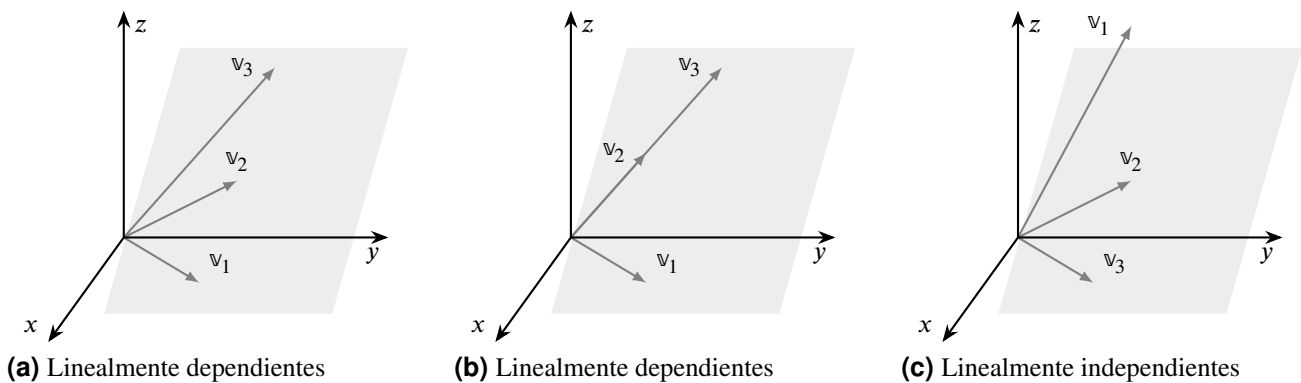


Figura 1.14

Observemos que un tercer eje de coordenadas en \mathbb{R}^2 es superfluo, ya que un vector unitario sobre tal eje debe ser expresable como una combinación lineal de los vectores unitarios sobre los ejes x y y positivos. En un sistema rectangular de coordenadas xy , cada vector en el plano puede expresarse de una única manera como una combinación lineal de los vectores unitarios estándar. Por ejemplo,

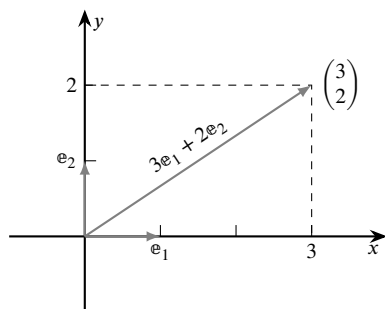


Figura 1.15

$$v = \begin{pmatrix} 3 \\ 2 \end{pmatrix} = 3 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + 2 \begin{pmatrix} 0 \\ 1 \end{pmatrix} = 3e_1 + 2e_2 \quad (1.4.1)$$

según la figura 1.15. Supongamos, sin embargo, que introducimos un tercer eje de coordenadas que forma un ángulo de 45° con el eje x . Llamémoslo el eje w . Como se ilustra en la figura 1.16, el vector unitario a lo largo del eje w es

$$w = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}$$

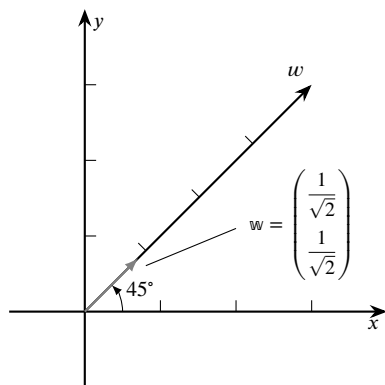


Figura 1.16

Mientras que la fórmula (1.4.1) muestra la única forma de expresar el vector v como una combinación lineal de e_1 y e_2 , existen infinitas formas de expresar este vector como una combinación lineal de e_1 , e_2 y w . Tres de esas posibilidades son:

$$\begin{pmatrix} 3 \\ 2 \end{pmatrix} = 3 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + 2 \begin{pmatrix} 0 \\ 1 \end{pmatrix} + 0 \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} = 3e_1 + 2e_2 + 0w$$

$$\begin{pmatrix} 3 \\ 2 \end{pmatrix} = 2 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} + \sqrt{2} \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} = 2e_1 + 1e_2 + \sqrt{2}w$$

$$\begin{pmatrix} 3 \\ 2 \end{pmatrix} = 4 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + 3 \begin{pmatrix} 0 \\ 1 \end{pmatrix} - \sqrt{2} \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} = 4e_1 + 3e_2 - \sqrt{2}w$$

En resumen, al introducir un eje superfluo, creamos la complicación de tener múltiples formas de asignar coordenadas a los puntos en el plano. Lo que hace que el vector w sea superfluo es el hecho de que puede expresarse como una combinación lineal de los vectores e_1 y e_2 , es decir,

$$w = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} = \frac{1}{\sqrt{2}}e_1 + \frac{1}{\sqrt{2}}e_2$$

Este resultado es una consecuencia del siguiente teorema.

Teorema 1.4.2: Un conjunto de m vectores en \mathbb{R}^n es siempre linealmente dependiente si $m > n$.

Demostración: Se deja como ejercicio al lector.

El siguiente teorema trata sobre la independencia y dependencia lineal de conjuntos que contienen el vector cero.

Teorema 1.4.3:

- i) Un conjunto finito que contiene el vector 0 es linealmente dependiente.
- ii) Un conjunto con exactamente un vector, es linealmente independiente si y solo si ese vector es distinto de 0 .

Demostración:

i) Para cualesquiera vectores v_1, v_2, \dots, v_n en un espacio vectorial, el conjunto $S = \{v_1, v_2, \dots, v_n, 0\}$ es linealmente dependiente, ya que la ecuación

$$0v_1 + 0v_2 + \dots + 0v_n + 1(0) = 0$$

expresa a 0 como una combinación lineal de los vectores en S con coeficientes que no son todos iguales a 0 .

ii) Se deja como ejercicio al lector.

Ejemplo 1.4.2:

Las funciones $f_1 = x$ y $f_2 = \sin(x)$ son vectores linealmente independientes en el espacio $F(-\infty, \infty)$, ya que ninguna de las funciones es múltiplo escalar de la otra. Por otro lado, las dos funciones $g_1 = \sin(2x)$ y $g_2 = \sin(x) \cos(x)$ son linealmente dependientes porque la identidad $\sin(2x) = 2 \sin(x) \cos(x)$ revela que g_1 y g_2 son múltiplos escalares entre sí.

Ejemplo 1.4.3:

El conjunto más básico de vectores linealmente independientes en \mathbb{R}^n es el conjunto de los vectores unitarios estándar

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \dots, \mathbf{e}_n = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}.$$

Para ilustrar esto en \mathbb{R}^3 , consideremos los vectores unitarios estándar

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \mathbf{e}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

Para probar la independencia lineal, debemos mostrar que los únicos coeficientes que satisfacen la ecuación vectorial

$$a_1 \mathbf{e}_1 + a_2 \mathbf{e}_2 + a_3 \mathbf{e}_3 = \mathbf{0}$$

son $a_1 = 0$, $a_2 = 0$, $a_3 = 0$. Esto se hace evidente al escribir esta ecuación en su forma de componentes

$$\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

No debería haber problemas para adaptar este argumento y establecer la independencia lineal de los vectores unitarios estándar en \mathbb{R}^n .

Aplicación de combinaciones lineales a modelos de color:

Los colores en los monitores de computadora se basan comúnmente en lo que se llama el modelo de color RGB. Los colores en este sistema se crean sumando porcentajes de los colores primarios rojo (R, red), verde (G, green) y azul (B, blue). Una forma de hacerlo es identificar los colores primarios con los vectores.

$$\begin{aligned} \mathbf{r} &= (1, 0, 0) && \text{(rojo puro),} \\ \mathbf{g} &= (0, 1, 0) && \text{(verde puro),} \\ \mathbf{b} &= (0, 0, 1) && \text{(azul puro),} \end{aligned}$$

en \mathbb{R}^3 , y crear todos los demás colores formando combinaciones lineales de \mathbf{r} , \mathbf{g} y \mathbf{b} usando coeficientes entre 0 y 1, inclusive; estos coeficientes representan el porcentaje de cada

color puro en la mezcla. El conjunto de todos estos vectores de color se llama el espacio RGB o el cubo RGB (figura 1.17). Así, cada vector de color \mathbf{c} en este cubo se puede expresar como una combinación lineal de la forma

$$\begin{aligned} \mathbf{c} &= k_1 \mathbf{r} + k_2 \mathbf{g} + k_3 \mathbf{b} \\ &= k_1(1, 0, 0) + k_2(0, 1, 0) + k_3(0, 0, 1) \\ &= (k_1, k_2, k_3) \end{aligned}$$

donde $0 \leq k_i \leq 1$. Como se indica en la figura, las esquinas del cubo representan los colores primarios puros junto con los colores blanco, magenta, cian y amarillo. Los vectores a lo largo de la diagonal, que van de negro a blanco, corresponden a los tonos de gris.

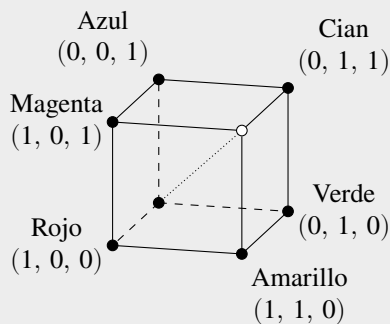


Figura 1.17 Representación del espacio de color RGB como un cubo en \mathbb{R}^3

Ejemplo 1.4.4: Demostrar que los polinomios

$$1, x, x^2, \dots, x^n$$

forman un conjunto linealmente independiente en P_n .

Solución: Por conveniencia, denotemos los polinomios como

$$p_0 = 1, p_1 = x, p_2 = x^2, \dots, p_n = x^n.$$

Debemos demostrar que los únicos coeficientes que satisfacen la ecuación vectorial

$$a_0 p_0 + a_1 p_1 + a_2 p_2 + \dots + a_n p_n = 0 \quad (1.4.2)$$

son

$$a_0 = a_1 = a_2 = \dots = a_n = 0.$$

La ecuación (1.4.2) es equivalente a la afirmación de que

$$a_0 + a_1 x + a_2 x^2 + \dots + a_n x^n = 0 \quad (1.4.3)$$

para todo $x \in (-\infty, \infty)$. Por lo tanto, debemos demostrar que esto es cierto si y solo si cada coeficiente en (1.4.3) es igual a cero. Para ver que esto es cierto, recordemos del álgebra que un polinomio no nulo de grado n tiene como máximo n raíces distintas. Así, si alguno de los coeficientes en (1.4.3) no es cero, el lado izquierdo sería un polinomio no nulo con un número infinito de raíces, lo cual es una contradicción. Por lo tanto, la ecuación (1.4.2) solo tiene la solución trivial.

1.5 Base y dimensión

Definición 1.5.1:

Decimos que un conjunto finito de vectores $\{v_1, v_2, \dots, v_n\}$ es una *base* para un espacio vectorial V si:

- i) $\{v_1, v_2, \dots, v_n\}$ es linealmente independiente,
- ii) $\{v_1, v_2, \dots, v_n\}$ genera a V .

Si se considera una base como la descripción de un sistema de coordenadas para un espacio vectorial V de dimensión finita, entonces la parte (i) de esta definición garantiza que no existe ninguna interrelación entre los vectores base, la parte (ii) garantiza que hay suficientes vectores base para proporcionar coordenadas a todos los vectores en V . A continuación, se presentan algunos ejemplos.

Ejemplo 1.5.1: Recordemos del ejemplo 1.3.1 que los vectores unitarios estándar en

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \dots, e_n = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}.$$

generan \mathbb{R}^n y del ejemplo 1.4.3, que son linealmente independientes. Por lo tanto, forman una base para \mathbb{R}^n que llamamos *base natural*, *base estándar* o *base canónica* para \mathbb{R}^n . En particular,

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, e_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

es la base estándar para \mathbb{R}^3 .

Ejemplo 1.5.2: Demostrar que $\mathcal{B} = \{1, x, x^2, \dots, x^n\}$ es una base para el espacio vectorial P_n de polinomios de grado n o menor, que llamamos la *base estándar* de P_n .

Solución: Debemos demostrar que los polinomios en \mathcal{B} son linealmente independientes y generan P_n . Denotemos estos polinomios como

$$p_0 = 1, p_1 = x, p_2 = x^2, \dots, p_n = x^n.$$

Mostramos en el ejemplo 1.3.2 que estos vectores generan P_n y en el ejemplo 1.4.4 que son linealmente independientes. Por lo tanto, forman una base para P_n .

Ejemplo 1.5.3: Consideremos el espacio vectorial \mathbb{R}^3 y el conjunto

$$\mathcal{B} = \left\{ \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \right\}.$$

Queremos demostrar que este conjunto es una base de \mathbb{R}^3 . Para verificar que \mathcal{B} es linealmente independiente, formamos la ecuación

$$c_1 \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + c_2 \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} + c_3 \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

y resolvemos para c_1, c_2, c_3 . De lo anterior, obtenemos el siguiente sistema lineal

$$\begin{aligned} c_1 + c_3 &= 0 \\ c_1 + c_2 &= 0 \\ c_2 + c_3 &= 0 \end{aligned}$$

que tiene como única solución $c_1 = c_2 = c_3 = 0$, lo cual muestra que \mathcal{B} es linealmente independiente. Para mostrar que \mathcal{B} genera a \mathbb{R}^3 , debemos probar que cualquier vector en \mathbb{R}^3 se puede expresar como combinación lineal de \mathcal{B} , es decir, debemos encontrar constantes c_1, c_2, c_3 tales que

$$c_1 \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + c_2 \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} + c_3 \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \quad \text{para cualquier } \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3$$

De la expresión anterior, obtenemos el siguiente sistema lineal

$$\begin{aligned} c_1 + c_3 &= x \\ c_1 + c_2 &= y \\ c_2 + c_3 &= z \end{aligned}$$

que tiene como única solución

$$c_1 = \frac{x + y - z}{2}, \quad c_2 = -\frac{x - y - z}{2}, \quad c_3 = \frac{x - y + z}{2}.$$

Por lo tanto, \mathcal{B} genera a \mathbb{R}^3 . Se concluye así que \mathcal{B} es una base para \mathbb{R}^3 .

Ejemplo 1.5.4: Determine una base para el subespacio de P_2 , formado por los vectores de la forma $at^2 + bt + c$ donde $c = a - b$.

Solución: Todo vector en V es de la forma

$$at^2 + bt + a - b,$$

que puede escribirse como

$$a(t^2 + 1) + b(t - 1),$$

lo cual muestra que los vectores $t^2 + 1$ y $t - 1$ generan a V . Además, estos vectores son linealmente independientes ya que ninguno es múltiplo del otro. También podría haberse obtenido (con mayor trabajo) esta conclusión sobre independencia lineal, escribiendo la ecuación

$$a_1(t^2 + 1) + a_2(t - 1) = 0$$

o

$$t^2 a_1 + t a_2 + (a_1 - a_2) = 0.$$

Como esta ecuación se cumple para todos los valores de t , debemos tener $a_1 = 0$ y $a_2 = 0$, que es la condición de independencia lineal.

Definición 1.5.2:

Sea V un espacio vectorial sobre K . Se llama *dimensión* del espacio vectorial V , a la cardinalidad de la base de V , y se denotará por $\dim V$.

Un espacio vectorial es de *dimensión finita* si existe un subconjunto finito de V que es una base para V ; en caso contrario, es decir, si no existe tal subconjunto finito, el espacio es de *dimensión infinita*.

Casi todos los espacios vectoriales considerados en este libro son de dimensión finita. Sin embargo, es conveniente anotar que muchos espacios vectoriales de importancia en matemáticas y en física son de dimensión infinita; su estudio excede el alcance de este libro. El espacio vectorial P , de todos los polinomios, y el espacio vectorial $C(-\infty, \infty)$ de las funciones continuas $f: \mathbb{R} \rightarrow \mathbb{R}$, no son de dimensión finita.

Ahora estableceremos algunos resultados relativos a espacios vectoriales de dimensión finita, que hacen referencia al número de vectores en una base, comparan el número de vectores de dos bases diferentes y establecen propiedades de las bases.

Teorema 1.5.1:

Sea V un espacio vectorial y sea $\mathcal{B} = \{v_1, v_2, \dots, v_n\}$ una base de V . Entonces todo elemento de V se expresa de manera única a partir de los elementos de la base. Es decir, dado $u \in V$, existen escalares a_1, a_2, \dots, a_n únicos tales que

$$u = a_1 v_1 + a_2 v_2 + \dots + a_n v_n \quad (1.5.1)$$

Demostración: Supongamos que existen escalares b_1, b_2, \dots, b_n tales que

$$u = b_1 v_1 + b_2 v_2 + \dots + b_n v_n \quad (1.5.2)$$

Se demostrará que $a_i = b_i$, para $i = 1, 2, \dots, n$. Sea

$$\begin{aligned} -u &= (-1) \cdot u \\ &= (-1) \cdot (b_1 v_1 + b_2 v_2 + \dots + b_n v_n) \\ &= (-1)b_1 v_1 + (-1)b_2 v_2 + \dots + (-1)b_n v_n \\ &= (-b_1)v_1 + (-b_2)v_2 + \dots + (-b_n)v_n \end{aligned}$$

Ahora, de las expresiones (1.5.1) y (1.5.2) tenemos

$$\begin{aligned} 0 &= u + (-u) \\ &= a_1 v_1 + a_2 v_2 + \dots + a_n v_n + (-b_1)v_1 + (-b_2)v_2 + \dots + (-b_n)v_n \\ &= (a_1 + (-b_1))v_1 + (a_2 + (-b_2))v_2 + \dots + (a_n + (-b_n))v_n \end{aligned}$$

Como \mathcal{B} es linealmente independiente,

$$(a_1 + (-b_1)) = 0, (a_2 + (-b_2)) = 0, \dots, (a_n + (-b_n)) = 0,$$

de modo que $b_1 = a_1, b_2 = a_2, \dots, b_n = a_n$. Por lo tanto, solo hay una forma de expresar u como combinación lineal de los vectores en \mathcal{B} .

Supongamos que $\mathcal{B} = \{v_1, v_2, \dots, v_n\}$ genera un espacio vectorial V y que v_j es una combinación lineal de los vectores que le preceden en \mathcal{B} . Entonces el conjunto

$$\mathcal{A} = \{v_1, v_2, \dots, v_{j-1}, v_{j+1}, \dots, v_n\}$$

que consta de los vectores en \mathcal{B} , con excepción de v_j , también genera a V . Para demostrar este resultado, observemos que si v es cualquier vector en V , entonces, como \mathcal{B} genera a V , podemos encontrar escalares a_1, a_2, \dots, a_n tales que

$$v = a_1 v_1 + a_2 v_2 + \dots + a_{j-1} v_{j-1} + a_j v_j + a_{j+1} v_{j+1} + \dots + a_n v_n.$$

Ahora, si

$$v_j = b_1 v_1 + b_2 v_2 + \dots + b_{j-1} v_{j-1},$$

entonces

$$\begin{aligned} v &= a_1 v_1 + a_2 v_2 + \dots + a_{j-1} v_{j-1} + a_j (b_1 v_1 + b_2 v_2 + \dots + b_{j-1} v_{j-1}) \\ &\quad + a_{j+1} v_{j+1} + \dots + a_n v_n \\ &= c_1 v_1 + c_2 v_2 + \dots + c_{j-1} v_{j-1} + c_{j+1} v_{j+1} + \dots + c_n v_n \end{aligned}$$

lo cual significa que $\text{gen } \mathcal{A} = V$.

Teorema 1.5.2: Sea $\mathcal{B} = \{v_1, v_2, \dots, v_n\}$ un conjunto de vectores no nulos en un espacio vectorial V y sea $W = \text{gen } \mathcal{B}$. Entonces, algún subconjunto de \mathcal{B} es una base para W .

Demostración: Procedamos por casos.

- i) Si \mathcal{B} es linealmente independiente, entonces \mathcal{B} es una base para W porque \mathcal{B} es generador de W , de acuerdo con la hipótesis.
- ii) Si \mathcal{B} es linealmente dependiente, entonces existen c_1, c_2, \dots, c_n , no todos iguales a cero, tales que

$$c_1 v_1 + c_2 v_2 + \dots + c_n v_n = 0.$$

Por lo tanto, algún v_j es una combinación lineal de los vectores anteriores en \mathcal{B} (teorema 1.4.1). Ahora, eliminamos v_j de \mathcal{B} , para obtener un subconjunto \mathcal{B}_1 de \mathcal{B} . Entonces, por la observación anterior a este teorema, concluimos que $\mathcal{B}_1 = \{v_1, v_2, \dots, v_{j-1}, v_{j+1}, \dots, v_n\}$ genera a W . Si \mathcal{B}_1 es linealmente independiente, entonces \mathcal{B}_1 es una base. Si \mathcal{B}_1 es linealmente dependiente, eliminamos un vector de \mathcal{B}_1 que sea combinación lineal de los vectores que le preceden en \mathcal{B}_1 y obtenemos un nuevo conjunto \mathcal{B}_2 que también genera a W . Continuando de esta forma, y dado que \mathcal{B} es un conjunto finito, en algún momento encontraremos un subconjunto \mathcal{A} de \mathcal{B} que es linealmente independiente y genera a W . Tal conjunto \mathcal{A} es una base para W .

Ejemplo 1.5.5: La dimensión de \mathbb{R}^2 es 2; la dimensión de \mathbb{R}^3 es 3; y en general, la dimensión de \mathbb{R}^n es n . Esto se expresa respectivamente como $\dim \mathbb{R}^2 = 2$, $\dim \mathbb{R}^3 = 3$, $\dim \mathbb{R}^n = n$.

Ejemplo 1.5.6: La dimensión de P_2 es 3; la dimensión de P_3 es 4; y en general, la dimensión de P_n es $n + 1$. Esto se expresa respectivamente como $\dim P_2 = 3$, $\dim P_3 = 4$, $\dim P_n = n + 1$.

Ejemplo 1.5.7: Si $\mathcal{B} = \{v_1, v_2, \dots, v_r\}$, entonces cada vector en $\text{gen } \mathcal{B}$ se puede expresar como una combinación lineal de los vectores en \mathcal{B} . Así, si los vectores en \mathcal{B} son linealmente independientes, automáticamente forman una base para $\text{gen } \mathcal{B}$, de lo cual podemos concluir que

$$\dim \text{gen}(\{v_1, v_2, \dots, v_r\}) = r.$$

En otras palabras, la dimensión del espacio generado por un conjunto linealmente independiente de vectores es igual al número de vectores en ese conjunto.

Observemos que si $\{v_1, v_2, \dots, v_n\}$ es una base para un espacio vectorial V , entonces $\{cv_1, cv_2, \dots, cv_n\}$ también es una base, si $c \neq 0$. Esta observación muestra que un espacio vectorial real diferente de $\{0\}$ tiene siempre infinitas bases.

Teorema 1.5.3: Sea V un espacio vectorial de dimensión finita. Si $\mathcal{B} = \{v_1, v_2, \dots, v_n\}$ es una base para V y $\mathcal{A} = \{w_1, w_2, \dots, w_r\}$ es un conjunto linealmente independiente de vectores en V , entonces $r \leq n$.

Demostración: Sea $\mathcal{A}_1 = \{w_1, v_1, v_2, \dots, v_n\}$. Como \mathcal{B} genera a V , \mathcal{A}_1 también lo genera. Y, puesto que w_1 es una combinación lineal de los vectores de \mathcal{B} , \mathcal{A}_1 es linealmente dependiente. Entonces, de acuerdo con el teorema 1.4.1, algún v_j es una combinación lineal de los vectores que le preceden en \mathcal{A}_1 . Eliminemos ese vector particular v_j . Sea $\mathcal{B}_1 = \{w_1, v_1, \dots, v_{j-1}, v_{j+1}, \dots, v_n\}$. Observemos que \mathcal{B}_1 genera a V . A continuación, sea $\mathcal{A}_2 = \{w_2, w_1, v_1, \dots, v_{j-1}, v_{j+1}, \dots, v_n\}$. Entonces \mathcal{A}_2 es linealmente dependiente, y algún vector en \mathcal{A}_2 es una combinación lineal de los vectores precedentes en \mathcal{A}_2 . Como \mathcal{A} es linealmente independiente, dicho vector no puede ser w_1 , así que es v_i , con $i \neq j$. Repetimos este proceso una y otra vez. Si se eliminan todos los vectores v antes de que se puedan incluir todos los vectores w , entonces el conjunto resultante de vectores w , un subconjunto de \mathcal{A} , es linealmente dependiente, lo cual implica que también \mathcal{A} es linealmente dependiente. Esta contradicción permite concluir que el número r de vectores w no puede ser mayor que el número n de vectores v . Esto es, $r \leq n$.

Corolario 1.5.1: Si $\mathcal{B} = \{v_1, v_2, \dots, v_n\}$ y $\mathcal{A} = \{w_1, w_2, \dots, w_m\}$ son dos bases para un espacio vectorial, entonces $n = m$.

Demostración: Como \mathcal{A} es un conjunto linealmente independiente de vectores, el teorema 1.5.3 implica que $m \leq n$. Igualmente, tenemos que $n \leq m$, puesto que \mathcal{B} es linealmente independiente. Entonces, $n = m$.

Se puede probar que todos los espacios de dimensión finita que tienen igual dimensión difieren solo en la naturaleza de sus elementos; sus propiedades algebraicas son idénticas.

A continuación, enunciaremos una consecuencia directa del teorema 1.5.3.

Teorema 1.5.4: Sea V un espacio vectorial de dimensión finita n , y sea W un subespacio no nulo de V , entonces $\dim W \leq \dim V$.

Demostración: Sea $\{v_1, v_2, \dots, v_k\}$ una base de W . Al ser una base, los vectores v_1, v_2, \dots, v_k son linealmente independientes. Por el teorema 1.5.3, sabemos que $k \leq n$, de donde se sigue que

$$\dim W = k \leq n = \dim V.$$

Por lo tanto, $\dim W \leq \dim V$.

En la observación anterior al teorema 1.5.3 mencionamos que un espacio vectorial tiene muchas bases; pero acabamos de demostrar que, para un espacio vectorial particular V , todas las bases tienen el mismo número de vectores. Este hecho conduce a la propia definición de dimensión.

Se puede probar que si un espacio vectorial V tiene dimensión finita n , entonces cualquier conjunto de $n + 1$ vectores en V es linealmente dependiente. En particular, cualquier conjunto con más de n vectores en \mathbb{R}^n es linealmente dependiente (teorema 1.4.2). A continuación, demostraremos que si $\dim V = n$, entonces ningún conjunto de $n - 1$ vectores en V puede generar a V . Por ejemplo, sabemos que la dimensión de P_2 es 3. Entonces el conjunto $\{1 - x^2, x\}$ no puede ser una base de P_2 , pues tendríamos a lo más dos vectores linealmente independientes y $2 < 3 = \dim P_2$.

Teorema 1.5.5: Si V es un espacio vectorial de dimensión finita n , entonces ningún conjunto con $n - 1$ vectores de V puede generar a V .

Demostración: Sea V un espacio vectorial de dimensión finita n . Supongamos que $\mathcal{B} = \{v_1, v_2, \dots, v_{n-1}\}$ es subconjunto de V y que además, \mathcal{B} genera V . Entonces, por el teorema 1.5.2 existe un subconjunto \mathcal{B}_1 de \mathcal{B} tal que \mathcal{B}_1 es una base de V . Dado que \mathcal{B}_1 es una base de V , tenemos que

$$\dim V = |\mathcal{B}_1|. \quad (1.5.3)$$

Como \mathcal{B}_1 es subconjunto de \mathcal{B} , se cumple que

$$|\mathcal{B}_1| \leq |\mathcal{B}| = n - 1. \quad (1.5.4)$$

Las ecuaciones (1.5.3) y (1.5.4) juntas implican que

$$\dim V \leq n - 1$$

lo cual es una contradicción con la hipótesis de que $\dim V = n$. Por lo tanto, \mathcal{B} no puede generar V .

De acuerdo con la definición, un conjunto de vectores en un espacio vectorial V es una base para V si genera a V y es linealmente independiente. Sin embargo, si sabemos que la dimensión de V es n , y el conjunto tiene n vectores, solo necesitamos verificar una de las dos condiciones, de acuerdo con el siguiente teorema.

Teorema 1.5.6: Sea V un espacio vectorial de dimensión n y sea $\mathcal{B} = \{v_1, v_2, \dots, v_n\}$ un conjunto de n vectores en V .

- i) Si \mathcal{B} es linealmente independiente, entonces es una base para V .
- ii) Si \mathcal{B} genera a V , entonces es una base para V .

Demostración:

- i) Dado que \mathcal{B} es linealmente independiente y $\dim V = n$, entonces existe una base \mathcal{A} para V , que contiene a \mathcal{B} . Pero, dado que $\dim V = n$, \mathcal{A} no puede contener más elementos que \mathcal{B} . Esto implica que $\mathcal{B} = \mathcal{A}$. Por lo tanto, \mathcal{B} es una base de V .
- ii) Dado que \mathcal{B} genera V , podemos decir que algún subconjunto de \mathcal{B} , digamos \mathcal{A} , es una base de V . Como $\dim V = n$, el conjunto \mathcal{A} contiene exactamente n elementos. Esto implica que $\mathcal{B} = \mathcal{A}$. Por lo tanto, \mathcal{B} es una base de V .

A continuación enunciaremos un teorema para construir una base que contenga un conjunto dado de vectores linealmente independientes.

Teorema 1.5.7: Si \mathcal{A} es un conjunto de vectores linealmente independientes en un espacio vectorial de dimensión finita V , existe una base \mathcal{B} para V , que contiene a \mathcal{A} .

Demostración: Sea $\mathcal{A} = \{v_1, v_2, \dots, v_k\}$ un conjunto de vectores linealmente independientes en V , donde $k \leq n$. Como V es de dimensión finita n , sabemos que existe una base $\mathcal{B} = \{v_1, v_2, \dots, v_k, w_1, w_2, \dots, w_{n-k}\}$ de V , donde los w_i son vectores adicionales que completan la base. Entonces, el conjunto \mathcal{B} genera a V . Ahora bien, queremos demostrar que $\mathcal{A} \cup \{w_1, w_2, \dots, w_{n-k}\}$ es un conjunto de vectores linealmente independientes. Es decir, debemos demostrar que

$$c_1 v_1 + c_2 v_2 + \dots + c_k v_k + d_1 w_1 + d_2 w_2 + \dots + d_{n-k} w_{n-k} = 0.$$

tiene como solución la trivial. Pero, \mathcal{A} es linealmente independiente, por lo que

$$c_1 = c_2 = \dots = c_k = 0.$$

Por lo tanto, la ecuación se reduce a

$$d_1 w_1 + d_2 w_2 + \cdots + d_{n-k} w_{n-k} = 0.$$

Como $\{w_1, w_2, \dots, w_{n-k}\}$ son parte de la base de V , también son linealmente independientes, lo que implica que

$$d_1 = d_2 = \cdots = d_{n-k} = 0.$$

Esto demuestra que el conjunto $\{v_1, v_2, \dots, v_k, w_1, w_2, \dots, w_{n-k}\}$ es linealmente independiente. Como el conjunto $\mathcal{A} \cup \{w_1, w_2, \dots, w_{n-k}\}$ es linealmente independiente y tiene n vectores, entonces este conjunto es una base de V .

El teorema anterior establece que un conjunto linealmente independiente de vectores en un espacio vectorial V puede extenderse a una base para V .

Ejemplo 1.5.8: Sea $V = \mathbb{R}^3$ y sea el conjunto de vectores $\mathcal{A} = \{v_1, v_2\}$, donde

$$v_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \quad \text{y} \quad v_2 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}.$$

Queremos demostrar que existe una base \mathcal{B} para V que contenga a \mathcal{A} . Notemos que \mathcal{A} es linealmente independiente, ya que no existe un escalar c tal que $cv_1 = v_2$. Aunque \mathcal{A} es linealmente independiente en \mathbb{R}^3 , no contiene tres vectores. Como \mathbb{R}^3 tiene dimensión 3, necesitamos agregar un tercer vector a \mathcal{A} para completar la base de V . Podemos elegir un tercer vector, por ejemplo,

$$v_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

que no es combinación lineal de los vectores en \mathcal{A} . Ahora, debemos verificar que

$$\mathcal{B} = \{v_1, v_2, v_3\}$$

es linealmente independiente, pero este es linealmente independiente ya que ninguno de estos vectores puede escribirse como una combinación lineal de los otros. Además, \mathcal{B} genera \mathbb{R}^3 , porque cualquier vector en \mathbb{R}^3 puede escribirse como una combinación lineal de estos tres vectores. Es decir,

$$c_1 \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + c_2 \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} + c_3 \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \quad \text{para cualquier} \quad \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3$$

cuya solución es

$$c_1 = y, \quad c_2 = x - y, \quad c_3 = -x + y + z.$$

Por lo tanto, \mathcal{B} es una base para \mathbb{R}^3 y contiene al conjunto \mathcal{A} .

Recordemos que la dimensión de un espacio vectorial se define como la cantidad de vectores linealmente independientes necesarios para generar todo el espacio. En el caso del vector cero, su dimensión es 0 porque no puede generar ningún otro vector aparte de sí mismo mediante combinaciones lineales. Para entenderlo mejor, consideremos que cualquier vector v en un espacio vectorial V puede ser expresado como

$$v = c_1 v_1 + c_2 v_2 + \cdots + c_n v_n$$

donde v_1, v_2, \dots, v_n son vectores linealmente independientes en V , y c_1, c_2, \dots, c_n son escalares. Ahora, si consideramos el vector cero, no importa cuánto intentemos expresarlo como una combinación lineal de otros vectores v_1, v_2, \dots, v_n , siempre obtendremos $\mathbf{0}$ como resultado, ya que cualquier escalar multiplicado por $\mathbf{0}$ sigue siendo $\mathbf{0}$.

Teorema 1.5.8: Sea V un espacio vectorial de dimensión finita n . Si S y T son subespacios de V , entonces

$$\dim(S + T) = \dim(S) + \dim(T) - \dim(S \cap T).$$

Demostración: Como S y T son subespacios de V , entonces $S + T$ y $S \cap T$ son subespacios de V (por el ejercicio 68, pág. 49 y teorema 1.2.2). Además, como V es de dimensión finita, entonces S , T , $S + T$ y $S \cap T$ son de dimensión finita. Sean

$$\dim S = r, \quad \dim T = m, \quad \dim(S \cap T) = k.$$

Sea

$$\{\mathfrak{u}_1, \mathfrak{u}_2, \dots, \mathfrak{u}_k\}$$

una base de $S \cap T$. Claro que $\{\mathfrak{u}_1, \mathfrak{u}_2, \dots, \mathfrak{u}_k\} \subset S$ es linealmente independiente y podemos completarlo con $\mathfrak{v}_1, \mathfrak{v}_2, \dots, \mathfrak{v}_{r-k} \in S$ a una base de S , y por tanto

$$\{\mathfrak{u}_1, \mathfrak{u}_2, \dots, \mathfrak{u}_k, \mathfrak{v}_1, \mathfrak{v}_2, \dots, \mathfrak{v}_{r-k}\}$$

es una base de S . Análogamente, $\{\mathfrak{u}_1, \mathfrak{u}_2, \dots, \mathfrak{u}_k\} \subset T$ es linealmente independiente y podemos completarlo con $\mathfrak{w}_1, \mathfrak{w}_2, \dots, \mathfrak{w}_{m-k} \in T$ a una base de T , y por tanto

$$\{\mathfrak{u}_1, \mathfrak{u}_2, \dots, \mathfrak{u}_k, \mathfrak{w}_1, \mathfrak{w}_2, \dots, \mathfrak{w}_{m-k}\}$$

es una base de T . Probaremos enseguida que

$$\mathcal{B} = \{\mathfrak{v}_1, \mathfrak{v}_2, \dots, \mathfrak{v}_{r-k}, \mathfrak{u}_1, \mathfrak{u}_2, \dots, \mathfrak{u}_k, \mathfrak{w}_1, \mathfrak{w}_2, \dots, \mathfrak{w}_{m-k}\}$$

es una base de $S + T$. Primero, demostremos que, en efecto, \mathcal{B} genera a $S + T$. Sea $\mathfrak{s} \in S + T$, entonces $\mathfrak{s} = \mathfrak{u} + \mathfrak{v}$. Es decir, existen $s_1, \dots, s_k, b_1, \dots, b_{r-k}$ y $t_1, \dots, t_k, a_1, \dots, a_{m-k}$ escalares tales que

$$\mathfrak{u} = s_1 \mathfrak{u}_1 + \dots + s_k \mathfrak{u}_k + b_1 \mathfrak{v}_1 + \dots + b_{r-k} \mathfrak{v}_{r-k}$$

y

$$\mathfrak{v} = t_1 \mathfrak{u}_1 + \dots + t_k \mathfrak{u}_k + a_1 \mathfrak{w}_1 + \dots + a_{m-k} \mathfrak{w}_{m-k}.$$

Lo que nos lleva a que

$$\mathfrak{s} = b_1 \mathfrak{v}_1 + \dots + b_{r-k} \mathfrak{v}_{r-k} + (s_1 + t_1) \mathfrak{u}_1 + \dots + (s_k + t_k) \mathfrak{u}_k + a_1 \mathfrak{w}_1 + \dots + a_{m-k} \mathfrak{w}_{m-k}.$$

Por lo tanto, $\mathfrak{s} \in \text{gen}(\{\mathfrak{v}_1, \dots, \mathfrak{v}_{r-k}, \mathfrak{u}_1, \dots, \mathfrak{u}_k, \mathfrak{w}_1, \dots, \mathfrak{w}_{m-k}\})$, lo que implica que \mathcal{B} genera a $S + T$. Ahora, demostremos que, en efecto, \mathcal{B} es linealmente independiente. Sean $x_1, \dots, x_{r-k}, y_1, \dots, y_k, z_1, \dots, z_{m-k}$ escalares tales que

$$0 = x_1 \mathfrak{v}_1 + \dots + x_{r-k} \mathfrak{v}_{r-k} + y_1 \mathfrak{u}_1 + \dots + y_k \mathfrak{u}_k + z_1 \mathfrak{w}_1 + \dots + z_{m-k} \mathfrak{w}_{m-k}.$$

Entonces

$$x_1 \mathfrak{v}_1 + \dots + x_{r-k} \mathfrak{v}_{r-k} = -(y_1 \mathfrak{u}_1 + \dots + y_k \mathfrak{u}_k + z_1 \mathfrak{w}_1 + \dots + z_{m-k} \mathfrak{w}_{m-k}),$$

por tanto

$$x_1 \mathfrak{v}_1 + \dots + x_{r-k} \mathfrak{v}_{r-k} \in T.$$

Como

$$x_1 \mathfrak{v}_1 + \dots + x_{r-k} \mathfrak{v}_{r-k} \in S,$$

entonces

$$x_1 \mathfrak{v}_1 + \dots + x_{r-k} \mathfrak{v}_{r-k} \in S \cap T.$$

Por tanto existen c_1, \dots, c_k escalares tales que

$$x_1 \mathfrak{v}_1 + \dots + x_{r-k} \mathfrak{v}_{r-k} = c_1 \mathfrak{u}_1 + \dots + c_k \mathfrak{u}_k.$$

Consecuentemente

$$y_1 \mathfrak{u}_1 + \dots + y_k \mathfrak{u}_k + z_1 \mathfrak{w}_1 + \dots + z_{m-k} \mathfrak{w}_{m-k} = 0,$$

y como

$$\{\mathbb{U}_1, \mathbb{U}_2, \dots, \mathbb{U}_k, \mathbb{W}_1, \mathbb{W}_2, \dots, \mathbb{W}_{m-k}\}$$

es base de T , entonces

$$0 = y_1 = \dots = y_k = z_1 = \dots = z_{m-k}.$$

Por tanto

$$\{\mathbb{V}_1, \mathbb{V}_2, \dots, \mathbb{V}_{r-k}, \mathbb{U}_1, \mathbb{U}_2, \dots, \mathbb{U}_k, \mathbb{W}_1, \mathbb{W}_2, \dots, \mathbb{W}_{m-k}\}$$

es linealmente independiente. Por todo lo anterior,

$$\{\mathbb{V}_1, \mathbb{V}_2, \dots, \mathbb{V}_{r-k}, \mathbb{U}_1, \mathbb{U}_2, \dots, \mathbb{U}_k, \mathbb{W}_1, \mathbb{W}_2, \dots, \mathbb{W}_{m-k}\}$$

es base de $S + T$, entonces

$$\begin{aligned} \dim(S + T) &= (r - k) + k + (m - k) \\ &= r + m - k \\ &= \dim S + \dim T - \dim(S \cap T) \end{aligned}$$

Ejemplo 1.5.9: En P_3 , sean

$$W_1 = \{p(x) \in P_3 \text{ tal que } x^2 - 1 \mid p(x)\}$$

y

$$W_2 = \{p(x) \in P_3 \text{ tal que } x^2 + x - 2 \mid p(x)\}.$$

Obtenga una base y la dimensión de los siguientes subespacios:

- a) W_1 , b) W_2 , c) $W_1 \cap W_2$, d) $W_1 + W_2$.

Demostración: Antes de empezar, recordemos la definición de divisibilidad de polinomios: Sean $f(x), g(x) \in P_n$. Decimos que $g(x)$ divide a $f(x)$, o que $g(x)$ es un factor de $f(x)$, si existe $q(x) \in P_n$ tal que $f(x) = g(x)q(x)$. Comúnmente, para decir que $g(x)$ divide a $f(x)$ se escribe $g(x) \mid f(x)$.

- a) Sea $p(x) \in P_3$, con $p(x) = a_0 + a_1x + a_2x^2 + a_3x^3$, entonces

$$\begin{aligned} p(x) \in W_1 &\iff x^2 - 1 \mid p(x) \\ &\iff \exists q(x) \in P_3 \text{ tal que } p(x) = (x^2 - 1)q(x) \\ &\iff p(x) = (x^2 - 1)(ax + b) \\ &\iff p(x) = a(x^3 - x) + b(x^2 - 1) \end{aligned}$$

Por lo tanto, $p(x) \in \text{gen}(\{x^3 - x, x^2 - 1\})$, y como no existe un escalar c tal que $x^2 - 1 = c(x^3 - x)$, entonces el conjunto $\{x^3 - x, x^2 - 1\}$ es linealmente independiente. Además, es claro que dicho conjunto genera a W_1 , pues para cualquier escalar a y b , $p(x) \in W_1$. Por lo tanto, demostramos que $\mathcal{A} = \{x^3 - x, x^2 - 1\}$ es una base para W_1 y por consiguiente, tenemos que $\dim W_1 = 2$.

- b) Sea $p(x) \in P_3$, con $p(x) = a_0 + a_1x + a_2x^2 + a_3x^3$, entonces

$$\begin{aligned} p(x) \in W_2 &\iff x^2 + x - 2 \mid p(x) \\ &\iff \exists q(x) \in P_3 \text{ tal que } p(x) = (x^2 + x - 2)q(x) \\ &\iff p(x) = (x^2 + x - 2)(ax + b) \\ &\iff p(x) = a(x^3 + x^2 - 2x) + b(x^2 + x - 2) \end{aligned}$$

Por lo tanto, $p(x) \in \text{gen}(\{x^3 + x^2 - 2x, x^2 + x - 2\})$, y como no existe un escalar c tal que $x^3 + x^2 - 2x = c(x^2 + x - 2)$, entonces el conjunto

$\{x^3 + x^2 - 2x, x^2 + x - 2\}$ es linealmente independiente. Además, es claro que dicho conjunto genera a W_2 , pues para cualquier a y b , $p(x) \in W_2$. Por lo tanto, demostramos que $\mathcal{B} = \{x^3 + x^2 - 2x, x^2 + x - 2\}$ es una base para W_2 y por consiguiente, tenemos que $\dim W_2 = 2$.

c) Sea $p(x) \in P_3$, con $p(x) = a_0 + a_1x + a_2x^2 + a_3x^3$, entonces

$$\begin{aligned} p(x) \in W_1 \cap W_2 &\iff x^2 - 1 \mid p(x) \text{ y } x^2 + x - 2 \mid p(x) \\ &\iff (x-1)(x+1) \mid p(x) \text{ y } (x-1)(x+2) \mid p(x) \\ &\iff (x-1)(x+1)(x+2) \mid p(x) \\ &\iff c(x^3 + 2x^2 - x - 2) \end{aligned}$$

Por lo tanto, $p(x) \in \text{gen}(\{x^3 + 2x^2 - x - 2\})$, y es linealmente independiente, pues es un conjunto de un solo elemento. Además, es claro que dicho conjunto genera a $W_1 \cap W_2$, pues para cualquier escalar c , $p(x) \in W_1 \cap W_2$. Por lo tanto, demostramos que $\mathcal{C} = \{x^3 + 2x^2 - x - 2\}$ es una base de $W_1 \cap W_2$ y por consiguiente, $\dim(W_1 \cap W_2) = 1$.

d) Por el ejercicio 68 en la página 49, tenemos que

$$\begin{aligned} W_1 + W_2 &= \text{gen}(\mathcal{A} \cup \mathcal{B}) \\ &= \text{gen}(\{x^3 - x, x^2 - 1, x^3 + x^2 - 2x, x^2 + x - 2\}) \end{aligned}$$

Observemos que

$$x^2 + x - 2 = (x^3 - x) + 2(x^2 - 1) - (x^3 + x^2 - 2x).$$

Por lo tanto, $x^2 + x - 2$ es linealmente dependiente de los otros polinomios y se elimina. Los polinomios restantes forman el conjunto

$$\mathcal{D} = \{x^3 - x, x^2 - 1, x^3 + x^2 - 2x\}$$

que es linealmente independiente, pues si formamos la ecuación

$$a(x^3 - x) + b(x^2 - 1) + c(x^3 + x^2 - 2x) = 0,$$

que es equivalente a

$$(a+c)x^3 + (b+c)x^2 + (-a-2c)x + (-b) = 0,$$

obtenemos el siguiente sistema lineal

$$\begin{aligned} a + c &= 0 \\ b + c &= 0 \\ -a - 2c &= 0 \\ -b &= 0 \end{aligned}$$

que tiene como única solución

$$a = b = c = 0.$$

Por el teorema 1.5.8 y por los incisos anteriores de este ejemplo,

$$\begin{aligned} \dim(W_1 + W_2) &= \dim(W_1) + \dim(W_2) - \dim(W_1 \cap W_2) \\ &= 2 + 2 - 1 = 3 \end{aligned}$$

Ahora, sabemos que \mathcal{D} es linealmente independiente y que además, coincide con la dimensión de $W_1 + W_2$. Así que por el teorema 1.5.6, una base para $W_1 + W_2$ es simplemente $\mathcal{D} = \{x^3 - x, x^2 - 1, x^3 + x^2 - 2x\}$.

1.6 Ejercicios del Capítulo 1

De los problemas 1 al 15 determine si el conjunto dado es un espacio vectorial. De no ser así proporcione una lista de los axiomas que no se cumplen.

1. El conjunto de números naturales \mathbb{N} como vectores, el conjunto de números naturales \mathbb{N} como escalares y la operación de multiplicación para números naturales.
2. El conjunto de números naturales \mathbb{N} como vectores, el conjunto de números naturales \mathbb{N} como escalares, la operación de suma para números naturales y la multiplicación entre números naturales para la operación de multiplicación de escalar y vector.
3. El conjunto de números enteros \mathbb{Z} como vectores, el conjunto de números naturales \mathbb{N} como escalares, la operación de suma para números enteros y la multiplicación entre números enteros para la operación de multiplicación de escalar y vector.
4. $\left\{ \begin{pmatrix} x \\ y \end{pmatrix} \mid y \leq 0 \text{ donde } x, y \in \mathbb{R} \right\}$ con la suma de vectores y multiplicación por un escalar usuales.
5. Los vectores en el plano que está en el primer cuadrante.
6. El conjunto de vectores en \mathbb{R}^2 de la forma $\begin{pmatrix} x \\ x \end{pmatrix}$.
7. El conjunto de vectores los números racionales \mathbb{Q} con la operación de suma, el conjunto de escalares los números enteros \mathbb{Z} y la operación de multiplicación de escalar y vector la multiplicación usual.
8. El conjunto de polinomios de grado menor o igual a n con término constante cero.
9. El conjunto de polinomios de grado menor o igual a n con término constante a_0 positivo.
10. El conjunto de polinomios de grado menor o igual a n con término constante a_0 negativo.
11. El conjunto de funciones continuas de valores reales definidas en $[0, 1]$ con $f(0) = 0$ y $f(1) = 0$ bajo las operaciones del ejemplo 1.1.4.
12. El conjunto de puntos en \mathbb{R}^3 que se encuentran sobre una recta que pasa por el origen.
13. El conjunto de puntos en \mathbb{R}^3 que se encuentran sobre la recta $x = t + 1$, $y = 2t$, $z = t - 1$.
14. El conjunto de funciones diferenciables definidas en $[0, 1]$ con las operaciones del ejemplo 1.1.4.
15. El conjunto de números reales de la forma $a + b\sqrt{2}$, donde a y b son números racionales, bajo la suma de números reales usual y la multiplicación por un escalar definida solo para escalares racionales.

De los problemas 16 al 34 determine si el subconjunto dado W del espacio vectorial V es un subespacio de V .

16. $V = \mathbb{R}^2$; $W = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} \mid x = 3, y \in \mathbb{R} \right\}$
17. $V = \mathbb{R}^2$; $W = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} \mid y \geq 0 \right\}$

$$18. V = \mathbb{R}^2; W = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} \mid x = y \right\}$$

$$19. V = \mathbb{R}^2; W = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} \mid y = 2x \right\}$$

$$20. V = \mathbb{R}^3; W = \text{el plano } xy$$

$$21. V = \mathbb{R}^2; W = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} \mid x^2 + y^2 \leq 1 \right\}$$

$$22. V = \mathbb{R}^2; W = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} \mid x^2 + y^3 < 1 \right\}$$

$$23. V = \mathbb{R}; W = \mathbb{Q}$$

$$24. V = P_n; W = \{p \in P_n \mid p(0) = 0 \text{ y } p'(0) = 0\}$$

$$25. V = P_4; W = \{p \in P_4 \mid p(0) = 0\}$$

$$26. V = P_n; W = \{p \in P_n \mid p(0) = 0\}$$

$$27. V = P_n; W = \{p \in P_n \mid p(0) = 1\}$$

$$28. V = C[0, 1]; W = \{f \in C[0, 1] \mid f(0) = f(1) = 0\}$$

$$29. V = C[0, 1]; W = \{f \in C[0, 1] \mid f(0) = 2\}$$

$$30. V = C^1[0, 1]; W = \{f \in C^1[0, 1] \mid f'(0) = 0\}$$

$$31. V = C[a, b]; \text{ con } a, b \in \mathbb{R} \text{ y } a < b; W = \left\{ f \in C[a, b] \mid \int_a^b f(x)dx = 0 \right\}$$

$$32. V = C[a, b]; W = \left\{ f \in C[a, b] \mid \int_a^b f(x)dx = 1 \right\}$$

$$33. V = C[a, b]; W = \left\{ f \in C[a, b] \mid \int_a^b f^2(x)dx = 0 \right\}$$

$$34. \text{ Sea } W = \left\{ \begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} \mid ax + by + cz + dw = 0 \right\}, \text{ donde } a, b, c \text{ y } d \text{ son números reales, no todos cero. Demuestre que } W \text{ es un subespacio propio de } \mathbb{R}^4.$$

W se llama un hiperplano en \mathbb{R}^4 que pasa por el origen.

De los problemas 35 al 52 determine si el conjunto dado de vectores genera el espacio vectorial dado.

$$35. \text{ En } \mathbb{R}^2: \begin{pmatrix} 2 \\ 10 \end{pmatrix}, \begin{pmatrix} 10 \\ 8 \end{pmatrix}$$

$$36. \text{ En } \mathbb{R}^2: \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 3 \\ 4 \end{pmatrix}$$

$$37. \text{ En } \mathbb{R}^2: \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 2 \end{pmatrix}$$

$$38. \text{ En } \mathbb{R}^2: \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 3 \\ 4 \end{pmatrix}, \begin{pmatrix} -1 \\ -2 \end{pmatrix}$$

$$39. \text{ En } \mathbb{R}^2: \begin{pmatrix} -12 \\ 5 \end{pmatrix}, \begin{pmatrix} -3 \\ 0 \end{pmatrix}, \begin{pmatrix} 4 \\ -8 \end{pmatrix}$$

$$40. \text{ En } \mathbb{R}^2: \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 2 \end{pmatrix}, \begin{pmatrix} 5 \\ 5 \end{pmatrix}$$

$$41. \text{ En } \mathbb{R}^2: \begin{pmatrix} -6 \\ 5 \end{pmatrix}, \begin{pmatrix} 7 \\ 9 \end{pmatrix}, \begin{pmatrix} 7 \\ -12 \end{pmatrix}, \begin{pmatrix} -10 \\ 6 \end{pmatrix}$$

$$42. \text{ En } \mathbb{R}^3: \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} -1 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 5 \\ 2 \\ 3 \end{pmatrix}$$

$$43. \text{ En } \mathbb{R}^3: \begin{pmatrix} 0 \\ 5 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \\ 3 \end{pmatrix}, \begin{pmatrix} -1 \\ -1 \\ 5 \end{pmatrix}$$

$$44. \text{ En } \mathbb{R}^3: \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \quad 45. \text{ En } \mathbb{R}^3: \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 3 \\ 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 7 \\ 3 \\ 5 \end{pmatrix}$$

$$46. \text{ En } \mathbb{R}^3: \begin{pmatrix} 7 \\ 6 \\ 9 \end{pmatrix}, \begin{pmatrix} 7 \\ 3 \\ 9 \end{pmatrix}, \begin{pmatrix} 7 \\ 0 \\ 3 \end{pmatrix}, \begin{pmatrix} 9 \\ 6 \\ 7 \end{pmatrix} \quad 47. \text{ En } \mathbb{R}^3: \begin{pmatrix} 4 \\ 4 \\ -6 \end{pmatrix}, \begin{pmatrix} -8 \\ 4 \\ -24 \end{pmatrix}, \begin{pmatrix} -4 \\ 0 \\ -6 \end{pmatrix}$$

$$48. \text{ En } P_2: 1 - x, 3 - x^2$$

$$49. \text{ En } P_2: 1 - x, 3 - x^2, x$$

$$50. \text{ En } P_2: x^2 + 1, x^2 - 1, x + 6$$

$$51. \text{ En } P_2: -12x + 5x^2, -9 - 27x + 8x^2, -3 - 5x + x^2$$

$$52. \text{ En } P_2: -10 + 3x + 11x^2, 10 + 9x - 4x^2, 5 + x + 4x^2$$

De los problemas 53 al 60 describa el espacio generado por los vectores.

$$53. \begin{pmatrix} -6 \\ 3 \end{pmatrix}, \begin{pmatrix} -11 \\ 5 \end{pmatrix}$$

$$54. \begin{pmatrix} -5 \\ -8 \end{pmatrix}, \begin{pmatrix} -4 \\ -8 \end{pmatrix}, \begin{pmatrix} 10 \\ -5 \end{pmatrix}$$

$$55. \begin{pmatrix} -12 \\ -16 \end{pmatrix}, \begin{pmatrix} 6 \\ 8 \end{pmatrix}, \begin{pmatrix} 18 \\ 24 \end{pmatrix}$$

$$56. \begin{pmatrix} 20 \\ -23 \\ -8 \end{pmatrix}, \begin{pmatrix} 2 \\ 7 \\ -2 \end{pmatrix}, \begin{pmatrix} 8 \\ -3 \\ -4 \end{pmatrix}, \begin{pmatrix} -2 \\ 24 \\ -2 \end{pmatrix}$$

$$57. \begin{pmatrix} -3 \\ -3 \\ -2 \end{pmatrix}, \begin{pmatrix} -2 \\ 4 \\ -8 \end{pmatrix}, \begin{pmatrix} 6 \\ -6 \\ 12 \end{pmatrix}$$

$$58. \begin{pmatrix} -9 \\ 8 \\ -4 \end{pmatrix}, \begin{pmatrix} 39 \\ 20 \\ 38 \end{pmatrix}, \begin{pmatrix} -34 \\ 12 \\ -22 \end{pmatrix}, \begin{pmatrix} 7 \\ 12 \\ 10 \end{pmatrix}$$

$$59. \begin{pmatrix} 2 \\ -1 \\ -1 \end{pmatrix}, \begin{pmatrix} -4 \\ 2 \\ 2 \end{pmatrix}, \begin{pmatrix} 6 \\ -3 \\ -3 \end{pmatrix}$$

$$60. \begin{pmatrix} -2 \\ 1 \\ 3 \\ -4 \end{pmatrix}, \begin{pmatrix} 9 \\ 12 \\ -18 \\ 6 \end{pmatrix}, \begin{pmatrix} -23 \\ 25 \\ 25 \\ -56 \end{pmatrix}, \begin{pmatrix} -1 \\ 6 \\ 0 \\ -6 \end{pmatrix}$$

61. Demuestre que dos polinomios de grado menor o igual a dos, no pueden generar P_2 .

62. Si p_1, p_2, \dots, p_m genera P_m , demuestre que $m \geq n + 1$.

63. Demuestre que si u y v están en $\text{gen}(\{v_1, v_2, \dots, v_k\})$, entonces $u + v$ y αu están en $\text{gen}(\{v_1, v_2, \dots, v_k\})$.

64. Demuestre que el conjunto infinito $\{1, x, x^2, x^3, \dots\}$ genera P , el espacio vectorial de polinomios.

65. Si $S = \{v_1, v_2, \dots, v_r\}$ y $S' = \{u_1, u_2, \dots, u_k\}$ son conjuntos no vacíos de vectores en un espacio vectorial V , entonces

$$\text{gen}(\{v_1, v_2, \dots, v_r\}) = \text{gen}(\{u_1, u_2, \dots, u_k\})$$

si y solo si cada vector en S es una combinación lineal de los vectores en S' , y cada vector en S' es una combinación lineal de los vectores en S .

66. Sean v_1 y v_2 en \mathbb{R}^3 . Demuestre que si $v_2 = cv_1$, entonces $\text{gen}(\{v_1, v_2\})$ es una recta que pasa por el origen.

67. En el problema anterior suponga que v_1 y v_2 no son paralelos. Demuestre que $W = \text{gen}(\{v_1, v_2\})$ es un plano que pasa por el origen. ¿Cuál es la ecuación del plano?

68. Sea V un espacio vectorial y sean S y T subespacios de V . Demuestre que

i) $S \cup T$ es subespacio de V si y solo si $S \subseteq T$ o $T \subseteq S$.

ii) $S + T = \{u + v \mid u \in S \text{ y } v \in T\}$ es subespacio de V .

iii) $S + T = \text{gen}(S \cup T)$.

iv) Sean A, B subconjuntos de V . Si A genera a S y B genera a T , entonces $S + T = \text{gen}(A \cup B)$.

De los problemas 69 al 90 determine si el conjunto de vectores dado es linealmente dependiente o independiente.

69. $\begin{pmatrix} 9 \\ -8 \end{pmatrix}, \begin{pmatrix} -11 \\ -3 \end{pmatrix}$

70. $\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} -1 \\ -3 \end{pmatrix}$

71. $\begin{pmatrix} 2 \\ -1 \\ 4 \end{pmatrix}, \begin{pmatrix} 4 \\ -2 \\ 7 \end{pmatrix}$

72. $\begin{pmatrix} -6 \\ 1 \end{pmatrix}, \begin{pmatrix} 12 \\ -2 \end{pmatrix}$

73. $\begin{pmatrix} 2 \\ -1 \\ 4 \end{pmatrix}, \begin{pmatrix} 4 \\ -2 \\ 8 \end{pmatrix}$

74. $\begin{pmatrix} -2 \\ 3 \end{pmatrix}, \begin{pmatrix} 4 \\ 7 \end{pmatrix}$

75. $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$

76. $\begin{pmatrix} -10 \\ -6 \end{pmatrix}, \begin{pmatrix} 10 \\ -6 \end{pmatrix}, \begin{pmatrix} 5 \\ 9 \end{pmatrix}$

77. $\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$

78. $\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$

79. $\begin{pmatrix} 8 \\ 7 \\ -8 \end{pmatrix}, \begin{pmatrix} -11 \\ -12 \\ -7 \end{pmatrix}, \begin{pmatrix} 12 \\ -3 \\ 7 \end{pmatrix}$

80. $\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 4 \\ -1 \\ 1 \end{pmatrix}$

81. $\begin{pmatrix} -3 \\ 4 \\ 2 \end{pmatrix}, \begin{pmatrix} 7 \\ -1 \\ 3 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 8 \end{pmatrix}$

82. $\begin{pmatrix} -1 \\ 0 \\ 11 \end{pmatrix}, \begin{pmatrix} 7 \\ -20 \\ -29 \end{pmatrix}, \begin{pmatrix} 1 \\ -5 \\ 1 \end{pmatrix}$

83. En P_2 : $1 - x, x$

84. En P_2 : $-x, x^2 - 2x, 3x + 5x^2$

85. En P_2 : $-3 - 2x - 11x^2, -39 - 6x - 3x^2, -12 - 9x^2, 20 - 4x + 5x^2$

86. En P_4 : $x - 1, (x - 1)(x - 2), (x - 1)(x - 2)(x - 3), x^4$

87. En P_2 : $x, x^2 - x, x^3 - x$

88. En $C[0, 1]$: e^x, e^{-x}

89. En $C[0, 1]$: $\sin x, \cos x$

90. En $C[0, 1]$: $x, \sqrt{x}, \sqrt[3]{x}$

91. Determine una condición sobre los números a, b, c y d tal que los vectores $\begin{pmatrix} a \\ b \end{pmatrix}$ y $\begin{pmatrix} c \\ d \end{pmatrix}$ sean linealmente dependientes.

92. Encuentre una condición sobre los números a_{ij} tal que los vectores

$$\begin{pmatrix} a_{11} \\ a_{21} \\ a_{31} \end{pmatrix}, \begin{pmatrix} a_{12} \\ a_{22} \\ a_{32} \end{pmatrix}, \begin{pmatrix} a_{13} \\ a_{23} \\ a_{33} \end{pmatrix}$$

sean linealmente independientes.

93. ¿Para qué valor o valores de α serán linealmente dependientes los vectores

$$\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 2 \\ -1 \\ 4 \end{pmatrix}, \begin{pmatrix} 3 \\ \alpha \\ 4 \end{pmatrix}?$$

94. ¿Para qué valor o valores de α serán linealmente dependientes los vectores $\begin{pmatrix} 2 \\ -3 \\ 1 \end{pmatrix}$, $\begin{pmatrix} -4 \\ 6 \\ -2 \end{pmatrix}$, $\begin{pmatrix} \alpha \\ 1 \\ 2 \end{pmatrix}$?

95. ¿Para qué valor o valores de α serán linealmente dependientes los vectores $\begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix}$, $\begin{pmatrix} -2 \\ -1 \\ -1 \end{pmatrix}$, $\begin{pmatrix} \alpha \\ 5 \\ 2 \end{pmatrix}$?

96. ¿Para qué valor o valores de α y β serán linealmente independientes los vectores $\begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix}$, $\begin{pmatrix} -2 \\ -1 \\ \beta \end{pmatrix}$, $\begin{pmatrix} \alpha \\ 5 \\ 2 \end{pmatrix}$?

97. Demuestre que si los vectores v_1, v_2, \dots, v_n son linealmente dependientes en \mathbb{R}^m , con $m < n$, y si v_{n+1} es cualquier otro vector en \mathbb{R}^m , entonces el conjunto $v_1, v_2, \dots, v_n, v_{n+1}$ es linealmente dependiente.

98. Demuestre que si v_1, v_2, \dots, v_n ($n \geq 2$) son linealmente independientes, entonces también lo son v_1, v_2, \dots, v_k , donde $k < n$.

99. Demuestre que cualesquiera cuatro polinomios en P_2 son linealmente dependientes.

100. Demuestre que dos polinomios no pueden generar a P_2 .

101. Demuestre que cualesquiera $n + 2$ polinomios en P_n son linealmente dependientes.

102. Demuestre que cualquier subconjunto de un conjunto de vectores linealmente independientes es linealmente independiente.

103. Sean S_1 y S_2 dos conjuntos finitos linealmente independientes en un espacio vectorial V . Demuestre que $S_1 \cap S_2$ es un conjunto linealmente independiente.

104. Sea $\{v_1, v_2, \dots, v_n\}$ un conjunto linealmente independiente. Demuestre que los vectores $v_1, v_1 + v_2, v_1 + v_2 + v_3, \dots, v_1 + v_2 + \dots + v_n$ son linealmente independientes.

105. Sea $\{v_1, v_2, \dots, v_n\}$ un conjunto de vectores que tiene la propiedad de que el conjunto $\{v_i, v_j\}$ es linealmente dependiente cuando $i \neq j$. Demuestre que cada vector del conjunto es un múltiplo de un solo vector de ese conjunto.

106. Suponga que u, v y w , son linealmente independientes. Pruebe o desapruebe: $u + v, u + w$ y $u + w$ son linealmente independientes.

107. Sea $\{v_1, v_2, \dots, v_n\}$ un conjunto linealmente independiente y suponga que $v \notin \text{gen}(\{v_1, v_2, \dots, v_n\})$. Demuestre que $\{v_1, v_2, \dots, v_n\}$ es un conjunto linealmente independiente.

108. Encuentre un conjunto linealmente independiente de vectores en P_2 que contenga a los polinomios $1 - x^2$ y $1 + x^2$

109. Encuentre un conjunto linealmente independiente de vectores en P_2 que contenga a los polinomios $x + x^2$ y $1 + x$.

De los problemas 110 al 117 determine si el conjunto dado es una base para el espacio vectorial a que se refiere.

110. En P_2 : $-2 - 11x + 7x^2, -5 - x - 5x^2$

111. En P_2 : $1 - x^2, x$

112. En P_2 : $-3x$, $1 + x^2$, $x^2 - 5$
113. En P_2 : $1 + 3x + 7x^2$, $5 + 12x + 35x^2$, $8 + 5x - 12x^2$
114. En P_2 : $x^2 - 1$, $x^2 - 2$, $x^2 - 3$
115. En P_3 : 1 , $1 + x$, $1 + x^2$, $1 + x^3$
116. En P_2 : $10 - x - 10x^2$, $-23 + 14x + 53x^2$, $-1 + 4x + 11x^2$
117. En P_3 : 3 , $x^3 - 4x + 6$, x^2
118. Encuentre una base en \mathbb{R}^3 para el conjunto de vectores en el plano dado por $3x - 2y + 5z = 0$.
119. Encuentre una base en \mathbb{R}^3 para el conjunto de vectores en el plano dado por $3x - 2y + z = 0$.
120. Encuentre una base en \mathbb{R}^3 para el conjunto de vectores en la recta $x = 2$, $y = -2t$, $z = 3t$.
121. Encuentre una base en \mathbb{R}^3 para el conjunto de vectores en la recta $x = 3t$, $y = -2t$, $z = t$.
122. Demuestre que los únicos subespacios propios en \mathbb{R}^2 son rectas que pasan por el origen.
123. En \mathbb{R}^n un hiperplano que contiene a $\mathbf{0}$ es un subespacio de dimensión $n - 1$. Si W es un hiperplano en \mathbb{R}^n que contiene a $\mathbf{0}$, demuestre que

$$W = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \mid a_1x_1 + a_2x_2 + \cdots + a_nx_n = 0 \right\}$$

donde a_1, a_2, \dots, a_n son números reales fijos, no todos cero.

124. Demuestre que dos vectores \mathbf{v}_1 y \mathbf{v}_2 en \mathbb{R}^2 con puntos terminales en el origen son colineales si y solo si $\dim \text{gen}(\{\mathbf{v}_1, \mathbf{v}_2\}) = 1$.
125. Demuestre que los tres vectores $\mathbf{v}_1, \mathbf{v}_2$ y \mathbf{v}_3 en \mathbb{R}^3 con puntos terminales en el origen son coplanares si y solo si $\dim \text{gen}(\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}) \leq 2$.
126. Demuestre que cualesquiera n vectores que generan un espacio V de dimensión n forman una base para V .
127. Demuestre que todo subespacio de un espacio vectorial de dimensión finita tiene una base.
128. Determine la dimensión del subespacio de P_2 formado por los vectores de la forma $at^2 + bt + c$, donde $c = b - 2a$.
129. Determine la dimensión del subespacio de P_3 formado por los vectores de la forma $at^3 + bt^2 + ct + d$, donde $b = 3a - 5d$ y $c = d + 4a$.
130. En el espacio P_2 , sea

$$W_1 = \{p(x) \in P_2 \mid p(0) = 0\} \quad \text{y} \quad W_2 = \{p(x) \in P_2 \mid p(1) = 0\}.$$

Obtenga una base y la dimensión para los siguientes subespacios: W_1 , W_2 , $W_1 \cap W_2$ y $W_1 + W_2$.

2

Capítulo

MATRICES

Las matrices son herramientas fundamentales en el álgebra lineal, con aplicaciones que abarcan desde la solución de sistemas de ecuaciones lineales hasta la representación de transformaciones lineales y la modelación de fenómenos en disciplinas como la física, la ingeniería y la economía. Para simplificar, en gran parte de este libro restringiremos nuestra atención al análisis de las matrices cuyas entradas son números reales. Sin embargo, también se estudiarán las matrices con entradas complejas, mismas que tienen gran importancia en muchas aplicaciones.

Este capítulo está dedicado al estudio de las matrices reales, sus propiedades algebraicas y las operaciones que pueden realizarse sobre ellas. Iniciaremos estableciendo el concepto de matriz real como una estructura rectangular formada por elementos del conjunto de los números reales. Posteriormente, introduciremos las operaciones básicas que se pueden realizar entre matrices: la suma, el producto por un escalar y el producto matricial, analizando las condiciones necesarias para que estas operaciones estén bien definidas y sus propiedades fundamentales, como la asociatividad, distributividad y conmutatividad.

Las matrices reales son más que simples arreglos de números; representan sistemas lineales y sirven como base para conceptos avanzados, como determinantes, transformaciones lineales y espacios vectoriales. En este capítulo, también exploraremos tipos específicos de matrices reales, como las matrices cuadradas, diagonales, simétricas y triangulares, cada una con propiedades particulares que las hacen esenciales en aplicaciones matemáticas y computacionales.

El estudio de las matrices reales no solo proporciona herramientas algebraicas para manipular datos y resolver problemas matemáticos, sino que también abre la puerta a la comprensión de conceptos más abstractos, como valores y vectores propios, y su conexión con la diagonalización. Este enfoque nos permitirá sentar las bases para abordar temas más avanzados en capítulos posteriores, donde se extiende el análisis a matrices con entradas complejas y espacios vectoriales complejos.

2.1 El espacio vectorial de matrices

Definición 2.1.1:

Una *matriz* de $m \times n$ es un arreglo rectangular de mn números reales ordenados en m filas (renglones) horizontales y n columnas verticales:

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1j} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2j} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{i1} & a_{i2} & \dots & a_{ij} & \dots & a_{in} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mj} & \dots & a_{mn} \end{bmatrix} \quad \begin{array}{l} \longleftarrow i\text{-ésimo renglón} \\ \uparrow j\text{-ésima columna} \end{array}$$

Nos referimos al número a_{ij} , que está en la i -ésima fila (renglón) y la j -ésima columna donde $i \in \{1, 2, \dots, m\}$ y $j \in \{1, 2, \dots, n\}$, como el ij -ésimo *elemento* de dicha matriz, o la *entrada* (i, j) de dicha matriz.

Vamos a denotar a las matrices con letras mayúsculas: A, B, C, D, \dots , y al arreglo de la matriz A , lo denotaremos por $[a_{ij}]$. Diremos que A es m por n (que se escribe $m \times n$). Al conjunto de las matrices de tamaño $m \times n$ con entradas en \mathbb{R} lo denotaremos por $\mathcal{M}_{m \times n}(\mathbb{R})$, donde

$$\mathcal{M}_{m \times n}(\mathbb{R}) = \{[a_{ij}] \mid a_{ij} \in \mathbb{R}, \text{ con } i \in \{1, 2, \dots, m\} \text{ y } j \in \{1, 2, \dots, n\}\}.$$

Si $m = n$, decimos que A es una *matriz cuadrada de orden n* , y que los números $a_{11}, a_{22}, \dots, a_{nn}$ forman la *diagonal principal* de A .

Ejemplo 2.1.1:

Sean

$$A = \begin{bmatrix} 2 & 0 & 3 \\ -2 & 1 & 8 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 4 \\ 2 & -3 \end{bmatrix}, \quad C = \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}.$$

Entonces A es una matriz de tamaño 2×3 con $a_{12} = 0, a_{13} = 3, a_{22} = 1, a_{23} = 8$; B es una matriz de tamaño 2×2 con $b_{11} = 1, b_{22} = -3$; C es una matriz de tamaño 3×1 con $c_{11} = 1, c_{21} = -1, c_{31} = 2$.

Definición 2.1.2:

Dadas $A, B \in \mathcal{M}_{m \times n}(\mathbb{R})$ con $A = [a_{ij}]$, $B = [b_{ij}]$, y además, $\alpha \in \mathbb{R}$, definimos la matriz suma de A y B (denotada por $A + B$) y la matriz producto por escalar de α y A (denotada por $\alpha \cdot A$), respectivamente como

$$\begin{aligned} +: \quad \mathcal{M}_{m \times n}(\mathbb{R}) \times \mathcal{M}_{m \times n}(\mathbb{R}) &\longrightarrow \mathcal{M}_{m \times n}(\mathbb{R}) \\ (A, B) &\longmapsto A + B \end{aligned}$$

$$\begin{aligned} \cdot: \quad \mathbb{R} \times \mathcal{M}_{m \times n}(\mathbb{R}) &\longrightarrow \mathcal{M}_{m \times n}(\mathbb{R}) \\ (\alpha, A) &\longmapsto \alpha \cdot A \end{aligned}$$

Es decir, para toda $i \in \{1, 2, \dots, m\}$ y $j \in \{1, 2, \dots, n\}$ tenemos que

$$[a_{ij}] + [b_{ij}] = [a_{ij} + b_{ij}], \quad \alpha \cdot [a_{ij}] = [\alpha \cdot a_{ij}].$$

Ejemplo 2.1.2:

Sean

$$A = \begin{bmatrix} 1 & -2 & 4 \\ 2 & -1 & 3 \end{bmatrix} \quad \text{y} \quad B = \begin{bmatrix} 0 & 2 & -4 \\ 1 & 3 & -1 \end{bmatrix}.$$

Entonces

$$A + B = \begin{bmatrix} 1+0 & -2+2 & 4+(-4) \\ 2+1 & -1+3 & 3+(-1) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 3 & 2 & 2 \end{bmatrix}.$$

Definición 2.1.3:

Decimos que dos matrices A y B son iguales si:

- i) Tienen el mismo tamaño.
- ii) Tienen las mismas entradas.

Teorema 2.1.1:

El conjunto de las matrices de tamaño $m \times n$ con entradas en \mathbb{R}

$$\mathcal{M}_{m \times n}(\mathbb{R}) = \{[a_{ij}] \mid a_{ij} \in \mathbb{R}, \text{ con } i \in \{1, 2, \dots, m\} \text{ y } j \in \{1, 2, \dots, n\}\}$$

con las operaciones de la definición 2.1.2, es un espacio vectorial sobre \mathbb{R} .

Demostración: Probemos que el conjunto $\mathcal{M}_{m \times n}(\mathbb{R})$ cumple las diez propiedades de la definición 1.1.1 como sigue:

- i) Es claro que se cumple la primer propiedad de suma, pues dados $[a_{ij}]$, $[b_{ij}] \in \mathcal{M}_{m \times n}(\mathbb{R})$

$$[a_{ij}] + [b_{ij}] = [a_{ij} + b_{ij}] \in \mathcal{M}_{m \times n}(\mathbb{R}) \quad \text{por def. de suma de matrices}$$

- ii) Se deja como ejercicio al lector.

- iii) Dados $[a_{ij}]$, $[b_{ij}] \in \mathcal{M}_{m \times n}(\mathbb{R})$

$$\begin{aligned} [a_{ij}] + [b_{ij}] &= [a_{ij} + b_{ij}] && \text{por def. de suma de matrices} \\ &= [b_{ij} + a_{ij}] && \text{por conmutatividad en } \mathbb{R} \\ &= [b_{ij}] + [a_{ij}] \end{aligned}$$

Por tanto, se cumple la conmutatividad.

- iv) Existe $[0_{ij}] \in \mathcal{M}_{m \times n}(\mathbb{R})$, siendo

$$\begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & & \ddots & \\ 0 & 0 & \cdots & 0 \end{bmatrix}$$

tal que

$$\begin{aligned} [a_{ij}] + [0_{ij}] &= [a_{ij} + 0] && \text{por def. de suma de matrices} \\ &= [a_{ij}] && \text{por neutro aditivo en } \mathbb{R} \end{aligned}$$

Por tanto, se cumple la propiedad del neutro aditivo.

- v) Dada $[a_{ij}] \in \mathcal{M}_{m \times n}(\mathbb{R})$, existe $[-a_{ij}] \in \mathcal{M}_{m \times n}(\mathbb{R})$ tal que

$$\begin{aligned} [a_{ij}] + [-a_{ij}] &= [a_{ij} - a_{ij}] && \text{por def. de suma de matrices} \\ &= [0_{ij}] && \text{por inverso aditivo en } \mathbb{R} \end{aligned}$$

A la matriz $[-a_{ij}]$ se le llama matriz inversa de $[a_{ij}]$ para la suma. Por tanto, se cumple la propiedad del inverso aditivo.

- vi) Es claro que se cumple la primer propiedad de multiplicación escalar, pues dados $\alpha \in \mathbb{R}$ y $[a_{ij}] \in \mathcal{M}_{m \times n}(\mathbb{R})$

$$\alpha \cdot [a_{ij}] = [\alpha \cdot a_{ij}] \in \mathcal{M}_{m \times n}(\mathbb{R}) \quad \text{por def. de multiplicación escalar}$$

- vii) Dados $\alpha, \beta \in \mathbb{R}$ y $[a_{ij}] \in \mathcal{M}_{m \times n}(\mathbb{R})$

$$\begin{aligned} \alpha \cdot (\beta \cdot [a_{ij}]) &= \alpha \cdot [\beta \cdot a_{ij}] && \text{por def. de multiplicación escalar} \\ &= [\alpha \cdot (\beta \cdot a_{ij})] && \text{por def. de multiplicación escalar} \\ &= [(\alpha \cdot \beta) \cdot a_{ij}] && \text{por conmutatividad en } \mathbb{R} \\ &= (\alpha \cdot \beta) \cdot [a_{ij}] \end{aligned}$$

Por tanto, se cumple la asociatividad.

viii) Se deja como ejercicio al lector.

ix) Se deja como ejercicio al lector.

x) Se deja como ejercicio al lector.

Por tanto, $\mathcal{M}_{m \times n}(\mathbb{R})$ es un espacio vectorial sobre \mathbb{R} .

2.2 Multiplicación de matrices y propiedades

Definición 2.2.1:

Sean $\mathfrak{u}, \mathfrak{v} \in \mathbb{R}^n$ con

$$\mathfrak{u} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} \quad \text{y} \quad \mathfrak{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}.$$

Se define el *producto punto* o *producto escalar* de los vectores \mathfrak{u} y \mathfrak{v} , denotado por $\mathfrak{u} \bullet \mathfrak{v}$, como sigue:

$$\mathfrak{u} \bullet \mathfrak{v} = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n = \sum_{i=1}^n u_i v_i.$$

Teorema 2.2.1: El producto punto cumple lo siguiente:

- i) $\mathfrak{u} \bullet \mathbf{0} = 0$, siendo $\mathfrak{u}, \mathbf{0} \in \mathbb{R}^n$ y $0 \in \mathbb{R}$.
- ii) $\mathfrak{u} \bullet \mathfrak{v} = \mathfrak{v} \bullet \mathfrak{u}$, siendo $\mathfrak{u}, \mathfrak{v} \in \mathbb{R}^n$.
- iii) $\mathfrak{u} \bullet (\mathfrak{v} + \mathfrak{w}) = \mathfrak{u} \bullet \mathfrak{v} + \mathfrak{u} \bullet \mathfrak{w}$, siendo $\mathfrak{u}, \mathfrak{v}, \mathfrak{w} \in \mathbb{R}^n$.
- iv) $(\alpha \cdot \mathfrak{u}) \bullet \mathfrak{v} = \mathfrak{u} \bullet (\alpha \cdot \mathfrak{v}) = \alpha \cdot (\mathfrak{u} \bullet \mathfrak{v})$, siendo $\mathfrak{u}, \mathfrak{v} \in \mathbb{R}^n$ y $\alpha \in \mathbb{R}$.

Demostración:

i) Sea $\mathfrak{u}, \mathbf{0} \in \mathbb{R}^n$,

$$\begin{aligned} \mathfrak{u} \bullet \mathbf{0} &= \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} \bullet \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \\ &= u_1 \cdot 0 + u_2 \cdot 0 + \cdots + u_n \cdot 0 \\ &= 0 + 0 + \cdots + 0 \\ &= 0 \end{aligned}$$

ii) Sea $\mathfrak{u}, \mathfrak{v} \in \mathbb{R}^n$,

$$\begin{aligned} \mathfrak{u} \bullet \mathfrak{v} &= \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} \bullet \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} \\ &= u_1 v_1 + u_2 v_2 + \cdots + u_n v_n \\ &= v_1 u_1 + v_2 u_2 + \cdots + v_n u_n \\ &= \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} \bullet \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} \\ &= \mathfrak{v} \bullet \mathfrak{u} \end{aligned}$$

iii) Sea $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^n$,

$$\begin{aligned}
 \mathbf{u} \bullet (\mathbf{v} + \mathbf{w}) &= \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} \bullet \begin{pmatrix} v_1 + w_1 \\ v_2 + w_2 \\ \vdots \\ v_n + w_n \end{pmatrix} \\
 &= u_1(v_1 + w_1) + u_2(v_2 + w_2) + \cdots + u_n(v_n + w_n) \\
 &= u_1v_1 + u_1w_1 + u_2v_2 + u_2w_2 + \cdots + u_nv_n + u_nw_n \\
 &= u_1v_1 + u_2v_2 + \cdots + u_nv_n + u_1w_1 + u_2w_2 + \cdots + u_nw_n \\
 &= \mathbf{u} \bullet \mathbf{v} + \mathbf{u} \bullet \mathbf{w}
 \end{aligned}$$

iv) Sea $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n, \alpha \in \mathbb{R}$,

$$\begin{aligned}
 (\alpha \cdot \mathbf{u}) \bullet \mathbf{v} &= \begin{pmatrix} \alpha u_1 \\ \alpha u_2 \\ \vdots \\ \alpha u_n \end{pmatrix} \bullet \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} \\
 &= (\alpha u_1)v_1 + (\alpha u_2)v_2 + \cdots + (\alpha u_n)v_n \\
 &= \alpha u_1v_1 + \alpha u_2v_2 + \cdots + \alpha u_nv_n \\
 &= u_1\alpha v_1 + u_2\alpha v_2 + \cdots + u_n\alpha v_n \\
 &= u_1(\alpha v_1) + u_2(\alpha v_2) + \cdots + u_n(\alpha v_n) \\
 &= \mathbf{u} \bullet (\alpha \cdot \mathbf{v}) \\
 &= u_1(\alpha v_1) + u_2(\alpha v_2) + \cdots + u_n(\alpha v_n) \\
 &= u_1\alpha v_1 + u_2\alpha v_2 + \cdots + u_n\alpha v_n \\
 &= \alpha u_1v_1 + \alpha u_2v_2 + \cdots + \alpha u_nv_n \\
 &= \alpha(u_1v_1 + u_2v_2 + \cdots + u_nv_n) \\
 &= \alpha(\mathbf{u} \bullet \mathbf{v})
 \end{aligned}$$

Ejemplo 2.2.1:

Consideremos dos vectores en \mathbb{R}^3

$$\mathbf{u} = \begin{pmatrix} 2 \\ -3 \\ 1 \end{pmatrix} \quad \text{y} \quad \mathbf{v} = \begin{pmatrix} 4 \\ 5 \\ -2 \end{pmatrix}.$$

El producto punto de estos dos vectores se calcula como:

$$\mathbf{u} \bullet \mathbf{v} = (2)(4) + (-3)(5) + (1)(-2) = 8 - 15 - 2 = -9$$

Por lo tanto, el producto punto de \mathbf{u} y \mathbf{v} es -9 .

Las matrices son una forma elegante y eficiente de organizar y manipular conjuntos de números. Son especialmente útiles en el álgebra lineal, ya que nos permiten representar sistemas de ecuaciones lineales, realizar transformaciones lineales, modelar datos y mucho más.

El producto de matrices es una operación crucial que nos permite combinar dos matrices para obtener una nueva matriz. Pero a diferencia de la suma o resta de matrices, el producto no es simplemente una operación elemento por elemento. En cambio, implica un procedimiento más complejo que puede parecer un poco misterioso al principio. Supongamos que tenemos dos matrices, A y B . Para multiplicar A por B (denotado como AB), tomamos cada fila de A y la “emparejamos” con cada columna de B , realizando una serie de multiplicaciones y sumas. Es importante tener en cuenta que no siempre se pueden multiplicar matrices.

Para que sea más claro de ver, observe la figura 2.1. Proporcionaremos ejemplos prácticos y una variedad de ejercicios para que los lectores puedan poner en práctica sus conocimientos sobre el producto de matrices.

Figura 2.1 La representación del producto de matrices

$$B = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{bmatrix}$$

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

$a_{11}b_{11} + a_{12}b_{21}$
 $a_{21}b_{13} + a_{22}b_{23}$

Ejemplo 2.2.2: Sean las matrices $A \in \mathcal{M}_{2 \times 3}(\mathbb{R})$ y $B \in \mathcal{M}_{3 \times 4}(\mathbb{R})$ dadas por

$$A = \begin{bmatrix} 2 & 0 & -3 \\ 4 & 1 & 5 \end{bmatrix} \quad \text{y} \quad B = \begin{bmatrix} 7 & -1 & 4 & 7 \\ 2 & 5 & 0 & -4 \\ -3 & 1 & 2 & 3 \end{bmatrix}.$$

Al multiplicar la matriz A por la matriz B como la figura 2.1, obtenemos:

$$\begin{aligned}
 AB &= \begin{bmatrix} 2 & 0 & -3 \\ 4 & 1 & 5 \end{bmatrix} \begin{bmatrix} 7 & -1 & 4 & 7 \\ 2 & 5 & 0 & -4 \\ -3 & 1 & 2 & 3 \end{bmatrix} \\
 &= \begin{bmatrix} 23 & -5 & 2 & 5 \\ 15 & 6 & 26 & 39 \end{bmatrix}
 \end{aligned}$$

Definición 2.2.2:

Sean $A = [a_{ij}] \in \mathcal{M}_{m \times p}(\mathbb{R})$ y $B = [b_{jk}] \in \mathcal{M}_{p \times n}(\mathbb{R})$. Se define el producto de las matrices A con B , como sigue

$$\begin{aligned}
 AB &= [a_{ij}][b_{jk}] \\
 &= \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & & \ddots & \\ a_{m1} & a_{m2} & \cdots & a_{mp} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & & \ddots & \\ b_{p1} & b_{p2} & \cdots & b_{pn} \end{bmatrix} \\
 &= \begin{bmatrix} \sum_{q=1}^p a_{1q}b_{q1} & \sum_{q=1}^p a_{1q}b_{q2} & \cdots & \sum_{q=1}^p a_{1q}b_{qn} \\ \sum_{q=1}^p a_{2q}b_{q1} & \sum_{q=1}^p a_{2q}b_{q2} & \cdots & \sum_{q=1}^p a_{2q}b_{qn} \\ \vdots & & \ddots & \\ \sum_{q=1}^p a_{mq}b_{q1} & \sum_{q=1}^p a_{mq}b_{q2} & \cdots & \sum_{q=1}^p a_{mq}b_{qn} \end{bmatrix} \\
 &= [c_{ik}]
 \end{aligned}$$

donde

$$c_{ik} = \sum_{q=1}^p a_{iq}b_{qk} = a_{i1}b_{1k} + a_{i2}b_{2k} + \cdots + a_{ip}b_{pk} \quad (2.2.1)$$

Nótese que no siempre se puede multiplicar matrices. Dos matrices se pueden multiplicar únicamente si el número de columnas de la primera matriz es igual al número de renglones de la segunda. De otro modo, los vectores que forman el renglón i en A y la columna j de B no tendrán el mismo número de componentes y el producto punto en la ecuación (2.2.1) no estará definido. Dicho de otro modo, las matrices A y B serán *incompatibles* bajo la multiplicación.

Ejemplo 2.2.3:

Grupo	Elementos
X	a_1, a_2, a_3, a_4
Y	$b_1, b_2, b_3, b_4, b_5, b_6$
Z	c_1, c_2, c_3, c_4, c_5

Tabla 2.1

En este ejemplo se muestra la forma en la cual se puede usar la multiplicación de matrices para modelar la manera en que se extiende una enfermedad contagiosa. Las enfermedades contagiosas son aquellas que se propagan de una persona a otra a través del contacto directo. Los ejemplos incluyen la gripe, el resfriado común y la COVID-19. Considere tres grupos de personas X , Y y Z con 4, 6 y 5 elementos respectivamente como se muestra en la tabla 2.1. Suponga que cuatro individuos han contraído esta enfermedad. Este grupo entra en contacto con seis personas de un segundo grupo. Estos contactos, llamados *contactos directos*, se pueden representar por una matriz de 4×6 . Consideremos la siguiente matriz para los grupos X y Y :

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

En este caso se hace $a_{ij} = 1$ si la i -ésima persona del primer grupo entra en contacto con la j -ésima persona del segundo grupo. Por ejemplo, el 1 en la posición $(2, 4)$ significa que la segunda persona del primer grupo (infectada) entró en contacto con la cuarta persona del segundo grupo. Ahora suponga que un tercer grupo de cinco personas tiene varios contactos directos con individuos del segundo grupo. Esto también se puede representar mediante una matriz.

$$B = \begin{bmatrix} 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}.$$

Observe que $b_{64} = 0$, lo que quiere decir que la sexta persona del segundo grupo no tiene contacto con la cuarta persona del tercer grupo. Los contactos *indirectos* o de *segundo orden* entre individuos del primero y tercer grupos se representan mediante la matriz de 4×5 , $C = AB$. Para ver esto, observe que una persona del grupo 3 puede quedar contagiada por alguien del grupo 2, quien a su vez fue contagiada por alguien del grupo 1. Por ejemplo, como $a_{24} = 1$ y $b_{45} = 1$ se ve que, indirectamente, la quinta persona del grupo 3 tuvo contacto (a través de la cuarta persona del grupo 2) con la segunda persona del grupo 1. El número total de contactos indirectos entre la segunda persona del grupo 1 y la quinta persona del grupo 3 está dado por

$$\begin{aligned} c_{25} &= a_{21}b_{15} + a_{22}b_{25} + a_{23}b_{35} + a_{24}b_{45} + a_{25}b_{55} + a_{26}b_{65} \\ &= 1 \cdot 1 + 0 \cdot 0 + 0 \cdot 0 + 1 \cdot 1 + 0 \cdot 0 + 1 \cdot 0 = 2 \end{aligned}$$

Así pues,

$$C = AB = \begin{bmatrix} 0 & 0 & 0 & 2 & 0 \\ 1 & 0 & 2 & 0 & 2 \\ 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 2 & 0 & 1 \end{bmatrix}.$$

Observe que únicamente la segunda persona del grupo 3 no tiene contactos indirectos con la enfermedad. La quinta persona de este grupo tiene $2 + 1 + 1 = 4$ contactos de segundo orden.

Ejemplo 2.2.4:

Un torneo de Tenis se puede organizar de la siguiente manera. Cada uno de los n tenistas juega contra todos los demás y se registran los resultados en una matriz $R = [r_{ij}] \in \mathcal{M}_{n \times n}(\mathbb{R})$ como sigue:

$$r_{ij} = \begin{cases} 1, & \text{si el jugador } i \text{ le gana al } j \\ 0, & \text{si el jugador } i \text{ pierde contra el } j \\ 0, & \text{si } i = j \end{cases}$$

Se asigna al tenista i la calificación siguiente:

$$S_i = \sum_{j=1}^n r_{ij} + \frac{1}{2} \sum_{j=1}^n (R^2)_{ij} \quad (2.2.2)$$

Para un torneo de $n = 4$ tenistas, tenemos

$$R = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \end{bmatrix} = [r_{ij}].$$

Para clasificar a los jugadores, primero calculemos R^2 como sigue:

$$R^2 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 2 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{bmatrix}.$$

Así, utilizando la ecuación (2.2.2), obtenemos:

$$\begin{aligned} S_1 &= 1 + \frac{1}{2}(2) & S_2 &= 2 + \frac{1}{2}(3) & S_3 &= 1 + \frac{1}{2}(1) & S_4 &= 2 + \frac{1}{2}(2) \\ &= 2 & &= \frac{7}{2} & &= \frac{3}{2} & &= 3 \end{aligned}$$

Por tanto, $S_1 = 2$, $S_2 = 3.5$, $S_3 = 1.5$ y $S_4 = 3$, y además

$$S_3 < S_1 < S_4 < S_2.$$

El puntaje es el número de juegos ganados por el jugador i más la mitad del número de juegos ganados por los jugadores a quienes el jugador i venció.

Ejemplo 2.2.5:

En inteligencia artificial, el producto de matrices es esencial para calcular la salida de una capa en una *red neuronal artificial*. Supongamos que tenemos una *capa neuronal* con 3 neuronas y 4 entradas, donde: X contiene todas las características del dato, W es la *matriz de pesos* que conecta las entradas con las neuronas, B contiene todos los *sesgos* de los datos, Y es la *salida de la capa neuronal*. Así,

$$X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}, \quad W = \begin{bmatrix} w_{11} & w_{12} & w_{13} & w_{14} \\ w_{21} & w_{22} & w_{23} & w_{24} \\ w_{31} & w_{32} & w_{33} & w_{34} \end{bmatrix}, \quad B = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}.$$

La salida de la capa se obtiene mediante el producto de matrices:

$$Y = WX + B$$

Expandiendo este cálculo, el cual se realiza en *todas las capas de una red neuronal*:

$$Y = \begin{bmatrix} w_{11}x_1 + w_{12}x_2 + w_{13}x_3 + w_{14}x_4 + b_1 \\ w_{21}x_1 + w_{22}x_2 + w_{23}x_3 + w_{24}x_4 + b_2 \\ w_{31}x_1 + w_{32}x_2 + w_{33}x_3 + w_{34}x_4 + b_3 \end{bmatrix}.$$

Teorema 2.2.2: El producto de matrices es asociativo.

El producto de matrices, a diferencia de la multiplicación numérica, no es conmutativo en general. Esto significa que el orden de los factores altera significativamente el resultado. Para ilustrar este fenómeno, consideremos el siguiente contraejemplo con matrices cuadradas de orden 2:

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \quad \text{y} \quad B = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$$

Calculamos primero AB ,

$$\begin{aligned} AB &= \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 3 & 2 \\ 1 & 1 \end{bmatrix} \end{aligned}$$

Ahora, invirtiendo el orden

$$\begin{aligned} BA &= \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 2 & 3 \\ 1 & 2 \end{bmatrix} \end{aligned}$$

Así que, en general $AB \neq BA$.

De aquí en adelante se escribirá el producto de tres matrices simplemente como ABC . Se puede hacer esto porque $(AB)C = A(BC)$; entonces se obtiene la misma respuesta independientemente de cómo se lleve a cabo la multiplicación (siempre y cuando no se conmute ninguna de las matrices). La ley asociativa se puede extender a productos de más matrices. Por ejemplo, suponga que AB , BC y CD están definidas. Entonces

$$ABCD = A(B(CD)) = ((AB)C)D = A(BC)D = (AB)(CD)$$

Existe una ley distributiva para la multiplicación de matrices.

Teorema 2.2.3: Si todas las sumas y todos los productos siguientes están definidos, entonces

$$A(B + C) = AB + AC \quad \text{y} \quad (A + B)C = AC + BC$$

Demostración: Demostraremos la primer igualdad, la segunda se deja como ejercicio al lector. Sean $A \in \mathcal{M}_{m \times p}(\mathbb{R})$ con $A = [a_{ij}]$, y $B, C \in \mathcal{M}_{p \times n}(\mathbb{R})$ con $B = [b_{kj}]$ y $C = [c_{kj}]$. Por las dimensiones de las matrices, la suma $B + C$ y los productos AB , AC están bien definidos. Primero calculamos $B + C$, operación que se realiza entrada por entrada

$$\begin{aligned} B + C &= \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & & \ddots & \\ b_{p1} & b_{p2} & \cdots & b_{pn} \end{bmatrix} + \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & & \ddots & \\ c_{p1} & c_{p2} & \cdots & c_{pn} \end{bmatrix} \\ &= \begin{bmatrix} b_{11} + c_{11} & b_{12} + c_{12} & \cdots & b_{1n} + c_{1n} \\ b_{21} + c_{21} & b_{22} + c_{22} & \cdots & b_{2n} + c_{2n} \\ \vdots & & \ddots & \\ b_{p1} + c_{p1} & b_{p2} + c_{p2} & \cdots & b_{pn} + c_{pn} \end{bmatrix} \end{aligned}$$

Aplicando la definición de multiplicación matricial, cada entrada (i, j) del producto es el producto punto de la fila i de A con la columna j de $B + C$,

$$\begin{aligned}
A(B+C) &= \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & & \ddots & \\ a_{m1} & a_{m2} & \cdots & a_{mp} \end{bmatrix} \begin{bmatrix} b_{11} + c_{11} & b_{12} + c_{12} & \cdots & b_{1n} + c_{1n} \\ b_{21} + c_{21} & b_{22} + c_{22} & \cdots & b_{2n} + c_{2n} \\ \vdots & & \ddots & \\ b_{p1} + c_{p1} & b_{p2} + c_{p2} & \cdots & b_{pn} + c_{pn} \end{bmatrix} \\
&= \begin{bmatrix} \sum_{q=1}^p a_{1q}(b_{q1} + c_{q1}) & \sum_{q=1}^p a_{1q}(b_{q2} + c_{q2}) & \cdots & \sum_{q=1}^p a_{1q}(b_{qn} + c_{qn}) \\ \sum_{q=1}^p a_{2q}(b_{q1} + c_{q1}) & \sum_{q=1}^p a_{2q}(b_{q2} + c_{q2}) & \cdots & \sum_{q=1}^p a_{2q}(b_{qn} + c_{qn}) \\ \vdots & & \ddots & \\ \sum_{q=1}^p a_{mq}(b_{q1} + c_{q1}) & \sum_{q=1}^p a_{mq}(b_{q2} + c_{q2}) & \cdots & \sum_{q=1}^p a_{mq}(b_{qn} + c_{qn}) \end{bmatrix}
\end{aligned}$$

En cada entrada, la distributividad en \mathbb{R} permite separar las sumas

$$\sum_{q=1}^p a_{iq}(b_{qj} + c_{qj}) = \sum_{q=1}^p a_{iq}b_{qj} + \sum_{q=1}^p a_{iq}c_{qj}$$

Esto permite expresar la suma de dos productos matriciales, es decir,

$$\begin{aligned}
A(B+C) &= \begin{bmatrix} \sum_{q=1}^p a_{1q}b_{q1} & \sum_{q=1}^p a_{1q}b_{q2} & \cdots & \sum_{q=1}^p a_{1q}b_{qn} \\ \sum_{q=1}^p a_{2q}b_{q1} & \sum_{q=1}^p a_{2q}b_{q2} & \cdots & \sum_{q=1}^p a_{2q}b_{qn} \\ \vdots & & \ddots & \\ \sum_{q=1}^p a_{mq}b_{q1} & \sum_{q=1}^p a_{mq}b_{q2} & \cdots & \sum_{q=1}^p a_{mq}b_{qn} \end{bmatrix} \\
&\quad + \begin{bmatrix} \sum_{q=1}^p a_{1q}c_{q1} & \sum_{q=1}^p a_{1q}c_{q2} & \cdots & \sum_{q=1}^p a_{1q}c_{qn} \\ \sum_{q=1}^p a_{2q}c_{q1} & \sum_{q=1}^p a_{2q}c_{q2} & \cdots & \sum_{q=1}^p a_{2q}c_{qn} \\ \vdots & & \ddots & \\ \sum_{q=1}^p a_{mq}c_{q1} & \sum_{q=1}^p a_{mq}c_{q2} & \cdots & \sum_{q=1}^p a_{mq}c_{qn} \end{bmatrix} \\
&= AB + AC
\end{aligned}$$

Por lo tanto, queda demostrada la igualdad $A(B+C) = AB + AC$.

Cuando trabajamos con una misma matriz cuadrada, es posible definir potencias enteras positivas mediante la multiplicación iterada. Es decir, si $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ y $n \in \mathbb{N}$, definimos A^n como

$$A^n = \underbrace{AA \cdots A}_{n-\text{veces}}$$

A continuación, presentaremos una variedad de matrices con diferentes propiedades y características. Cada tipo de matriz posee atributos específicos que permiten clasificarla y estudiarla en el álgebra lineal.

Definición 2.2.3:

Sea $A \in \mathcal{M}_{n \times n}(\mathbb{R})$, decimos que A es una matriz idempotente si $A^2 = A$.

Ejemplo 2.2.6:

Sea $A = \begin{bmatrix} -1 & 3 & 5 \\ 1 & -3 & -5 \\ -1 & 3 & 5 \end{bmatrix}$, entonces

$$A^2 = \begin{bmatrix} -1 & 3 & 5 \\ 1 & -3 & -5 \\ -1 & 3 & 5 \end{bmatrix} \begin{bmatrix} -1 & 3 & 5 \\ 1 & -3 & -5 \\ -1 & 3 & 5 \end{bmatrix} = \begin{bmatrix} -1 & 3 & 5 \\ 1 & -3 & -5 \\ -1 & 3 & 5 \end{bmatrix} = A.$$

Por tanto, A es idempotente.

Definición 2.2.4:

Sea $A \in \mathcal{M}_{n \times n}(\mathbb{R})$, decimos que A es una matriz nilpotente si existe un entero positivo p tal que $A^p = \mathbf{0}$. El grado o índice de nilpotencia es el menor entero positivo p tal que $A^p = \mathbf{0}$.

Ejemplo 2.2.7:

Sea $A = \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix}$, entonces

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

Por tanto, A es nilpotente de índice de nilpotencia 2.

Una matriz que desempeña un papel central en el álgebra lineal es la llamada matriz identidad. Esta matriz actúa como el elemento neutro de la multiplicación matricial, análoga al número uno en la multiplicación de escalares.

Definición 2.2.5:

La *matriz identidad* I_n es una matriz de $n \times n$ cuyos elementos de la diagonal principal son iguales a 1 y todos los demás son 0. Esto es,

$$I_n = [\delta_{ij}] = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{si } i \neq j \end{cases}$$

Si no es importante enfatizar el tamaño, escribiremos solamente I para denotar a la matriz identidad. Para explicar el papel de las matrices identidad en la aritmética de matrices, consideremos el efecto de multiplicar una matriz general A de tamaño 2×3 en cada lado por una matriz identidad. Multiplicar por la derecha con la matriz identidad 3×3 da

$$AI_3 = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} = A$$

y multiplicar por la izquierda con la matriz identidad 2×2 da

$$I_2A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} = A.$$

El mismo resultado se cumple en general; es decir, si A es cualquier matriz $m \times n$,

$$AI_n = A \quad \text{y} \quad I_m A = A.$$

Dado que las matrices pueden elevarse a potencias y combinarse mediante operaciones algebraicas, es natural extender a ellas la noción de aplicarles un polinomio.

Definición 2.2.6:

Sea $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ y sea $f(x) \in P_m$, con

$$f(x) = a_m x^m + \cdots + a_1 x + a_0.$$

Definimos a $f(A)$ como:

$$f(A) = a_m A^m + \cdots + a_1 A + a_0 I_n.$$

Ejemplo 2.2.8:

Si $f(x) = x^2 - 5x + 2$ y $A = \begin{bmatrix} 2 & 0 \\ 4 & 5 \end{bmatrix}$, entonces:

$$f(A) = \begin{bmatrix} 2 & 0 \\ 4 & 5 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 4 & 5 \end{bmatrix} - 5 \begin{bmatrix} 2 & 0 \\ 4 & 5 \end{bmatrix} + 2 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} -4 & 0 \\ 8 & 2 \end{bmatrix}.$$

Definición 2.2.7:

Si A es una matriz cuadrada, entonces la *traza de A* , denotada por $\text{Tr}(A)$, se define como la suma de las entradas en la diagonal principal de A . La traza de A no está definida si A no es una matriz cuadrada.

Ejemplo 2.2.9:

Si consideramos la matriz del ejemplo 2.2.6, entonces la traza de A es

$$\text{Tr}(A) = (-1) + (-3) + 5 = 1.$$

Definición 2.2.8:

Sea $A \in \mathcal{M}_{m \times n}(\mathbb{R})$. La *matriz transpuesta* de una matriz, denotada por A^T , es la matriz que se obtiene al intercambiar las filas y las columnas de A . Es decir, si $A = [a_{ij}]$, entonces su transpuesta A^T es una matriz $n \times m$ dada por $A^T = [a_{ji}]$.

Ejemplo 2.2.10:

Consideremos

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}.$$

La *matriz transpuesta A^T* se obtiene intercambiando filas por columnas

$$A^T = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}.$$

En el caso especial en que A es una matriz cuadrada, la transpuesta de A puede obtenerse intercambiando las entradas que están posicionadas simétricamente con respecto a la diagonal principal. En la definición 2.2.8, observemos que A^T también puede obtenerse *reflejando* A respecto a su diagonal principal.

$$A = \begin{bmatrix} 1 & -2 & 4 \\ 3 & 7 & 0 \\ -5 & 8 & 6 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & -2 & 4 \\ 3 & 7 & 0 \\ -5 & 8 & 6 \end{bmatrix} \rightarrow A^T = \begin{bmatrix} 1 & 3 & -5 \\ -2 & 7 & 8 \\ 4 & 0 & 6 \end{bmatrix}$$

↑
Intercambia las entradas que están
posicionadas simétricamente
respecto a la diagonal principal.

Definición 2.2.9:

Sea A una matriz cuadrada.

- i) Si $A^T = A$, decimos que la matriz A es *simétrica*.
- ii) Si $A^T = -A$, decimos que la matriz A es *antisimétrica*.

Ejemplo 2.2.11:

Consideremos

$$A = \begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix}.$$

Notemos que $A^T = A$, por lo que A es simétrica. Sin embargo, si consideramos

$$B = \begin{bmatrix} 0 & 2 \\ -2 & 0 \end{bmatrix}.$$

Notemos que $B^T = -B$, por lo que B es antisimétrica.

A continuación, se presentan algunas propiedades de la transpuesta que facilitan su manipulación y que se utilizarán en diversos contextos posteriores.

Teorema 2.2.4: Sean A y B matrices, cuyos tamaños son tales que pueden realizarse las operaciones indicadas, y sea α un escalar. Entonces

- i) $(A^T)^T = A$.
- ii) $(A + B)^T = A^T + B^T$.
- iii) $(\alpha A)^T = \alpha A^T$.
- iv) $(AB)^T = B^T A^T$.

Demostración: Se dejan como ejercicio al lector.

La segunda y cuarta propiedad del teorema 2.2.4 pueden generalizarse a sumas y productos de un número finito de matrices. Es decir,

$$(A_1 + A_2 + \cdots + A_k)^T = A_1^T + A_2^T + \cdots + A_k^T \quad \text{y} \quad (A_1 A_2 \cdots A_k)^T = A_k^T \cdots A_2^T A_1^T$$

si se supone que los tamaños de las matrices son tales que todas las operaciones pueden realizarse.

Teorema 2.2.5: a) Si A es una matriz cuadrada, entonces $A + A^T$ es una matriz simétrica.
b) Para cualquier matriz A , AA^T y $A^T A$ son matrices simétricas.

Definición 2.2.10:

Sea $A \in \mathcal{M}_{n \times n}(\mathbb{R})$, decimos que A es una matriz diagonal si todas sus entradas fuera de la diagonal principal son cero, es decir, $[a_{ij}] = 0$, para toda $i \neq j$. Si $[a_{ij}] = a_{ij}$, entonces escribiremos

$$A = \text{diag} \{a_{11}, a_{22}, \dots, a_{nn}\}$$

Ejemplo 2.2.12:

Sea

$$A = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -6 & 0 \\ 0 & 0 & -9 \end{bmatrix},$$

entonces

$$A = \text{diag}\{-1, -6, -9\}.$$

Las matrices diagonales, constituyen un caso particularmente simple de las matrices cuadradas. Gracias a su estructura, son fáciles de manipular algebraicamente y presentan un comportamiento estable bajo operaciones comunes. Las siguientes propiedades muestran que el conjunto de matrices diagonales es cerrado bajo suma, multiplicación escalar, multiplicación entre matrices y transposición:

Teorema 2.2.6: Sean A, B matrices diagonales y sea $\alpha \in \mathbb{R}$, entonces

- i) $A + B$ es diagonal.
- ii) AB es diagonal.
- iii) αA es diagonal.
- iv) $A^T = A$.

Demostración: Se dejan como ejercicio al lector.

Definición 2.2.11:

Sea A una matriz cuadrada.

- i) Decimos que A es una matriz triangular superior si todas las entradas bajo la diagonal son cero, es decir, $[a_{ij}] = 0$, para toda $i > j$.
- ii) decimos que A es una matriz triangular inferior si todas las entradas sobre la diagonal son cero, es decir, $[a_{ij}] = 0$, para toda $i < j$.

Ejemplo 2.2.13: Un ejemplo de una matriz triangular superior es:

$$\begin{bmatrix} 1 & 6 & 1 \\ 0 & 2 & 7 \\ 0 & 0 & 9 \end{bmatrix}$$

Un ejemplo de una matriz triangular inferior es:

$$\begin{bmatrix} 6 & 0 & 0 \\ 4 & 3 & 0 \\ 8 & 5 & 1 \end{bmatrix}$$

Las matrices triangulares, ya sean superiores o inferiores, poseen estructuras algebraicamente convenientes que se conservan bajo ciertas operaciones. En el caso de las matrices triangulares superiores, se cumple lo siguiente:

Teorema 2.2.7: Sean A, B matrices triangulares superiores y sea $\alpha \in \mathbb{R}$, entonces

- i) $A + B$ es triangular superior.
- ii) AB es triangular superior.
- iii) αA es triangular superior.

Demostración: Se dejan como ejercicio al lector.

De manera análoga, las matrices triangulares inferiores también son cerradas bajo suma, multiplicación y multiplicación escalar, como se muestra a continuación:

Teorema 2.2.8: Sean A, B matrices triangulares inferiores y sea $\alpha \in \mathbb{R}$, entonces

- i) $A + B$ es triangular inferior.
- ii) AB es triangular inferior.
- iii) αA es triangular inferior.

Demostración: Se dejan como ejercicio al lector.

2.3 Sistemas de ecuaciones

En el estudio de diversos fenómenos, desde problemas físicos y económicos hasta modelos computacionales y redes eléctricas, es común encontrar relaciones entre cantidades desconocidas que se expresan mediante ecuaciones. En muchos casos, estas relaciones tienen una estructura algebraicamente simple: las incógnitas aparecen sin exponentes ni productos entre ellas. A este tipo de expresiones se les conoce como ecuaciones lineales.

Definición 2.3.1: Sean a_1, a_2, \dots, a_n, b escalares. Una ecuación lineal con n incógnitas es una expresión de la forma

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

donde a_1, a_2, \dots, a_n, b se llaman *coeficientes de la ecuación* y x_1, x_2, \dots, x_n se llaman *incógnitas*.

Una *solución* de la ecuación lineal anterior es un conjunto de valores de las incógnitas, digamos $x_1 = k_1, x_2 = k_2, \dots, x_n = k_n$, con la propiedad de que es cierta la siguiente expresión (obtenida sustituyendo cada x_i por k_i en la ecuación):

$$a_1k_1 + a_2k_2 + \dots + a_nk_n = b.$$

Se dice entonces que este conjunto de valores *satisface* la ecuación. El conjunto de todas las soluciones se llama *conjunto solución*, *solución general* o, simplemente, la *solución* de la ecuación.

Las nociones anteriores suponen que las incógnitas están ordenadas. Cuando se eviten los subíndices, normalmente utilizaremos variables x, y, z para denotar tres incógnitas, x, y, z, t para denotar cuatro incógnitas y así sucesivamente.

Ejemplo 2.3.1:

a) La ecuación $2x - 5y + 3xz = 4$ no es lineal, porque el producto de dos incógnitas es de segundo grado.

b) La ecuación $x + 2y - 4z + t = 3$ es lineal en las cuatro incógnitas x, y, z, t . Una solución a esta ecuación es: $x = 3, y = 2, z = 1, t = 0$. Esto ya que

$$3 + 2(2) - 4(1) + 0 = 3 \implies 3 = 3$$

es una expresión cierta. Sin embargo: $x = 1, y = 2, z = 4, t = 5$ no es una solución de la ecuación porque

$$1 + 2(2) - 4(4) + 5 = 3 \implies -6 = 3$$

no es cierto.

Antes de estudiar sistemas más generales, es útil analizar el caso más simple: una ecuación lineal con una sola incógnita.

Teorema 2.3.1: Consideremos la ecuación lineal $ax = b$.

- i) Si $a \neq 0$, $x = a^{-1}b$ es la solución única de $ax = b$.
- ii) Si $a = 0$, pero $b \neq 0$, $ax = b$ no tiene solución.
- iii) Si $a = 0$ y $b = 0$, todo escalar k es solución de $ax = b$.

Demostración:

- i) Supongamos $a \neq 0$. Entonces existe el escalar $a^{-1}b$. Sustituyendo $a^{-1}b$ en $ax = b$ se obtiene $a(a^{-1}b) = b$, o bien $b = b$; por consiguiente, $a^{-1}b$ es una solución. Por otra parte, supongamos que x_0 es solución de $ax = b$, de forma que $ax_0 = b$. Multiplicando ambos miembros por a^{-1} se obtiene $x_0 = a^{-1}b$. De aquí $a^{-1}b$ es la única solución de $ax = b$.
- ii) Sea ahora, $a = 0$. Entonces, para todo escalar k , tenemos $ak = 0k = 0$. Si $b \neq 0$, $ax \neq b$. De acuerdo con esto, k no es solución de $ax = b$.
- iii) Si $b = 0$, $ak = b$. Esto es, cualquier escalar k es una solución de $ax = b$.

Una ecuación lineal se dice *degenerada* si tiene la forma

$$0x_1 + 0x_2 + \cdots + 0x_n = b$$

esto es, si cada coeficiente es igual a cero. En este caso, el comportamiento de la ecuación depende únicamente del valor del término independiente b . Si $b \neq 0$, la ecuación se reduce a una contradicción del tipo $0 = b$, por lo que no tiene solución. En cambio, si $b = 0$, la igualdad se cumple para cualquier vector $\mathbf{x} \in \mathbb{R}^n$, y por tanto, todo vector es solución.

Definición 2.3.2:

Consideremos la ecuación lineal no degenerada

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

Por *primera incógnita* en dicha ecuación, entendemos la primera con coeficiente no nulo. Su posición p en la ecuación es entonces el menor entero de j para el cual $a_j \neq 0$. Es decir, x_p es la primer incógnita si $a_j = 0$ para $j < p$, pero $a_p \neq 0$.

Ejemplo 2.3.2: Consideremos la ecuación lineal $5y - 2z = 3$. Aquí y es la primera incógnita. Si las incógnitas son x, y y z , entonces $p = 2$ es su posición, pero si y y z son las únicas incógnitas, es $p = 1$.

Teorema 2.3.2: Consideremos una ecuación lineal no degenerada

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

con primer incógnita x_p .

- i) Cualquier conjunto de valores de las incógnitas x_j con $j \neq p$ dará una única solución de la ecuación. Las incógnitas x_j se llaman *variables libres* porque se les puede asignar cualquier valor.
- ii) Toda solución de la ecuación se obtiene en el inciso anterior. El conjunto de todas las soluciones se llama *solución general* de la ecuación.

Ejemplo 2.3.3: Encuentre dos soluciones particulares y la solución general de la ecuación lineal:

$$2x - 4y + z = 8.$$

Solución: Aquí x es la primera incógnita. De acuerdo con ello, asignamos valores cualesquiera a las variables libres y y z , entonces despejamos x para obtener una solución. Por ejemplo, si tomamos $y = 1$ y $z = 0$, la sustitución en la ecuación proporciona

$$2x - 4(1) + 0 = 8 \implies x = 6.$$

Por lo tanto, una solución es:

$$x = 6, \quad y = 1, \quad z = 0.$$

Ahora bien, si tomamos $y = 3$ y $z = 4$, la sustitución en la ecuación proporciona

$$2x - 4(3) + 4 = 8 \implies x = 8.$$

Por lo tanto, otra solución es:

$$x = 8, \quad y = 3, \quad z = 4.$$

La solución general de la ecuación principal, se obtiene como sigue: En primer lugar, asignamos valores arbitrarios (llamados *parámetros*) a las variables libres, digamos $y = a$ y $z = b$. A continuación sustituimos en la ecuación para obtener

$$2x - 4a + b = 8,$$

o equivalentemente

$$x = 4 + 2a - \frac{1}{2}b.$$

Entonces

$$x = 4 + 2a - \frac{1}{2}b, \quad y = a, \quad z = b$$

es la solución general.

Definición 2.3.3:

Sea $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ y $\mathbf{x} \in \mathbb{R}^n$, se define el producto de la matriz A con el vector \mathbf{x} como:

$$A\mathbf{x} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & \ddots & \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{bmatrix}.$$

Definición 2.3.4:

Si $A\mathbf{x} = \mathbf{b}$, obtenemos

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned} \quad (2.3.1)$$

A la expresión (2.3.1) se le llamará *sistema de ecuaciones lineales* (SEL) de m ecuaciones con n incógnitas. Además, si $\mathbf{b} = \mathbf{0}$, obtenemos

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= 0 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= 0 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= 0 \end{aligned}$$

A esta expresión se le llamará *sistema homogéneo*.

Observemos que en la definición 2.3.3 se puso a \mathbf{x} como una matriz de tamaño $n \times 1$, cuando al principio de dicha definición se dijo que pertenecía a \mathbb{R}^n . Aunque pueda parecer un error, no es así, pues \mathbf{x} se puede representar de ambas maneras, sea como vector o como matriz de tamaño $n \times 1$. Así que si $\mathbf{x} \in \mathbb{R}^n$, entonces $\mathbf{x} \in \mathcal{M}_{n \times 1}(\mathbb{R})$. Cuando consideremos a $\mathbf{x} \in \mathcal{M}_{n \times 1}(\mathbb{R})$, escribiremos a \mathbf{x} como

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \text{ en vez de } \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}.$$

Esto nos será de gran ayuda para entender de mejor manera los sistemas de ecuaciones lineales y no confundirnos con la notación.

Consideremos inicialmente un sistema de dos ecuaciones lineales con dos incógnitas x_1 y x_2 :

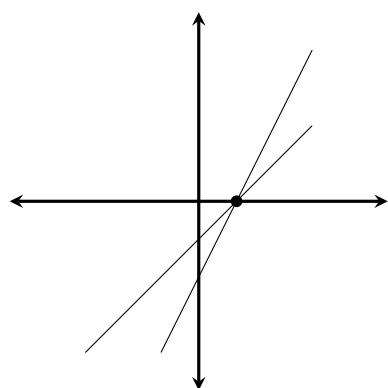
$$\begin{aligned} a_{11}x_1 + a_{12}x_2 &= b_1 \\ a_{21}x_1 + a_{22}x_2 &= b_2 \end{aligned}$$

donde a_{11} , a_{12} , a_{21} , a_{22} , b_1 , y b_2 son coeficientes reales. Geométricamente, cada ecuación representa una recta en el plano cartesiano, y la solución del sistema corresponde a sus puntos de intersección. Este sistema admite tres escenarios posibles: solución única (rectas que se cortan en un punto), infinitas soluciones (rectas coincidentes), o ninguna solución (rectas paralelas no coincidentes), como ilustra la figura 2.2.

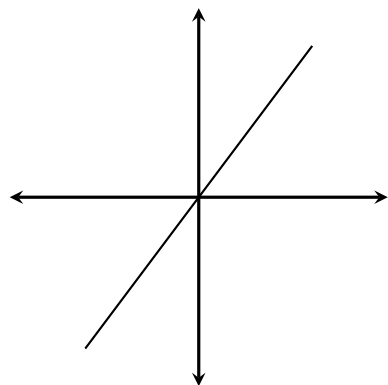
Al extender este concepto al espacio tridimensional, obtenemos un sistema de tres ecuaciones con tres incógnitas x_1 , x_2 , y x_3 :

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 &= b_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 &= b_2 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 &= b_3 \end{aligned}$$

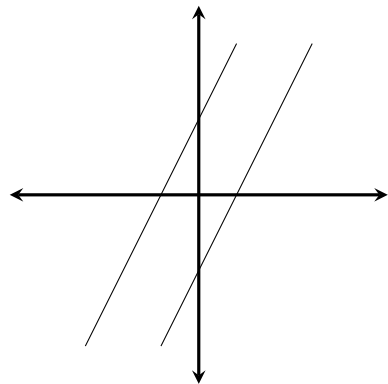
donde los coeficientes a_{ij} y términos independientes b_k son números reales. En este caso, cada ecuación describe un plano en el espacio tridimensional, y la solución representa la intersección de estos planos. Análogamente al caso bidimensional, el sistema puede tener: solución única (tres planos que se intersectan en un punto), infinitas soluciones (planos coincidentes o que se cortan formando una recta), o ninguna solución (planos paralelos o sin intersección común). Aunque la visualización geométrica tridimensional es más compleja, la figura 2.3 muestra una representación simplificada de estos posibles casos.



i) Sistemas con única solución



ii) Sistemas con infinitas soluciones



iii) Sistemas sin solución

Figura 2.2 Representación de los posibles casos del conjunto solución de un SEL 2×2

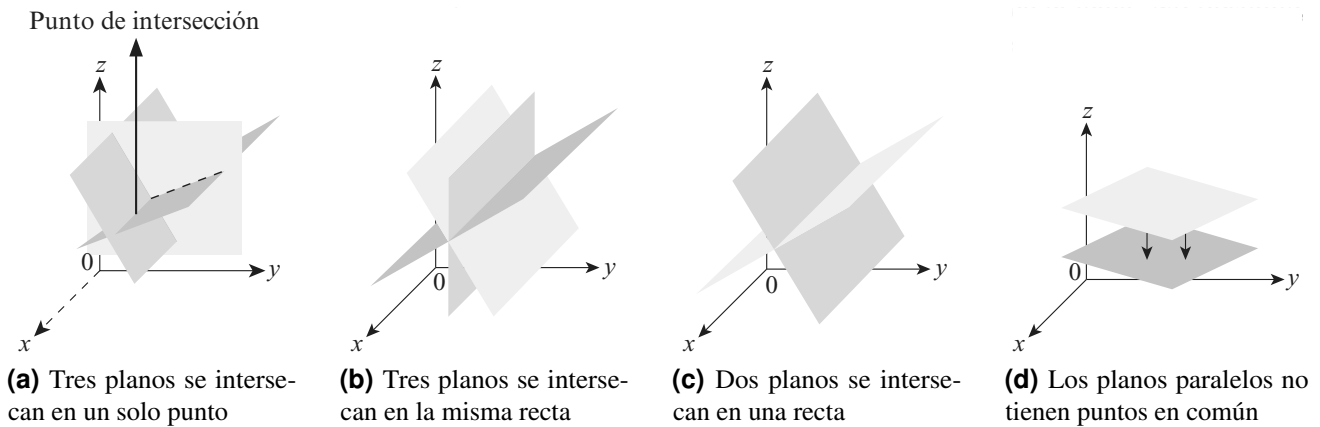


Figura 2.3 Representación de los posibles casos del conjunto solución de un SEL 3×3

La figura 2.3 ilustra cómo interactúan los planos en un sistema de tres ecuaciones lineales con tres incógnitas. Cada caso refleja una relación distinta entre las ecuaciones, lo que determina si el sistema tiene una solución única, infinitas soluciones o ninguna.

Por último, este análisis geométrico sirve como base para introducir nociones más abstractas, como el núcleo y la imagen de una matriz, que exploraremos en la siguiente sección. Estas ideas no solo son esenciales para resolver sistemas, sino también para comprender mejor las estructuras que los sustentan.

2.4 Núcleo e imagen de una matriz

Definición 2.4.1:

Sea $A \in \mathcal{M}_{m \times n}(\mathbb{R})$, se define el *espacio nulo* de A , denotado por $\mathcal{N}(A)$, como:

$$\mathcal{N}(A) = \{\mathbf{x} \in \mathbb{R}^n \mid A\mathbf{x} = \mathbf{0} \in \mathbb{R}^m\}$$

Al espacio nulo de A también se le llama *kernel* de A , o bien, *núcleo* de A .

Teorema 2.4.1: Sea $A \in \mathcal{M}_{m \times n}(\mathbb{R})$, entonces $\mathcal{N}(A)$ es un subespacio de \mathbb{R}^n .

Demostración:

i) Sean $\mathbf{x}, \mathbf{y} \in \mathcal{N}(A)$, entonces

$$A\mathbf{x} = \mathbf{0} \quad \text{y} \quad A\mathbf{y} = \mathbf{0}$$

Ahora

$$\begin{aligned} A(\mathbf{x} + \mathbf{y}) &= A\mathbf{x} + A\mathbf{y} \\ &= \mathbf{0} + \mathbf{0} \\ &= \mathbf{0} \end{aligned}$$

Entonces $\mathbf{x} + \mathbf{y} \in \mathcal{N}(A)$. Por tanto, se cumple la propiedad de cerradura para la suma.

ii) Se deja como ejercicio al lector.

De (i) y (ii), se sigue que $\mathcal{N}(A)$ es subespacio de \mathbb{R}^n .

Definición 2.4.2:

Llamaremos *nulidad* de A , denotada por $\nu(A)$, a la dimensión del espacio nulo, es decir,

$$\nu(A) = \dim(\mathcal{N}(A)).$$

Ejemplo 2.4.1: Determine $\mathcal{N}(A)$ y $\nu(A)$ de la siguiente matriz

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

Solución: Por definición,

$$\begin{aligned} \mathcal{N}(A) &= \{ \mathbf{x} \in \mathbb{R}^3 \mid A\mathbf{x} = \mathbf{0} \} \\ &= \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \in \mathbb{R}^3 \mid \begin{bmatrix} 1 & 2 & -1 \\ 2 & -1 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right\} \\ &= \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \in \mathbb{R}^3 \mid \begin{bmatrix} x_1 + 2x_2 - x_3 \\ 2x_1 - x_2 + 3x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right\} \end{aligned}$$

Entonces

$$\begin{aligned} x_1 + 2x_2 - x_3 &= 0 \\ 2x_1 - x_2 + 3x_3 &= 0 \end{aligned}$$

Multiplicando la segunda ecuación por 2, obtenemos

$$\begin{aligned} x_1 + 2x_2 - x_3 &= 0 \\ 4x_1 - 2x_2 + 6x_3 &= 0 \end{aligned}$$

Sumando la primer ecuación con la segunda, obtenemos

$$5x_1 + 5x_3 = 0$$

Por tanto, $x_3 = -x_1$. Sustituyendo en la primer ecuación, obtenemos

$$2x_1 + 2x_2 = 0$$

Por tanto, $x_2 = -x_1$. Se sigue entonces

$$\begin{aligned} \mathcal{N}(A) &= \left\{ \begin{pmatrix} x_1 \\ -x_1 \\ -x_1 \end{pmatrix} \mid x_1 \in \mathbb{R} \right\} \\ &= \left\{ x_1 \cdot \begin{pmatrix} 1 \\ -1 \\ -1 \end{pmatrix} \mid x_1 \in \mathbb{R} \right\} \\ &= \text{gen} \left(\left\{ \begin{pmatrix} 1 \\ -1 \\ -1 \end{pmatrix} \right\} \right) \end{aligned}$$

Por tanto, $\mathcal{N}(A) = \text{gen} \left(\left\{ \begin{pmatrix} 1 \\ -1 \\ -1 \end{pmatrix} \right\} \right)$. Además, $\nu(A) = 1$.

Si A es la matriz cero de tamaño $m \times n$, entonces el sistema se convierte en

$$\mathbf{0}_{m \times n} \mathbf{x} = \mathbf{0}_{n \times 1}.$$

Dado que la matriz cero anula cualquier vector $\mathbf{x} \in \mathbb{R}^n$, el sistema siempre tiene todas las soluciones posibles en \mathbb{R}^n . Esto significa que el espacio nulo de $\mathbf{0}_{m \times n}$ es todo el espacio \mathbb{R}^n

$$\mathcal{N}(\mathbf{0}) = \mathbb{R}^n \quad \text{y} \quad \nu(\mathbf{0}) = n.$$

Ejemplo 2.4.2: Determine $\mathcal{N}(A)$ y $\nu(A)$ de la siguiente matriz

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

Solución: Por definición,

$$\begin{aligned} \mathcal{N}(A) &= \{ \mathbf{x} \in \mathbb{R}^2 \mid A\mathbf{x} = \mathbf{0} \in \mathbb{R}^2 \} \\ &= \left\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2 \mid \begin{bmatrix} 1 & 2 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right\} \\ &= \left\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2 \mid \begin{bmatrix} x_1 + 2x_2 \\ x_1 - x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right\} \end{aligned}$$

Entonces

$$\begin{aligned} x_1 + 2x_2 &= 0 \\ x_1 - x_2 &= 0 \end{aligned}$$

Restando la primera ecuación de la segunda, obtenemos que

$$3x_2 = 0 \implies x_2 = 0$$

y sustituyendo en la segunda ecuación se obtiene que

$$x_1 = 0$$

Se sigue entonces que

$$x_1 = 0 \quad \text{y} \quad x_2 = 0$$

por lo que

$$\begin{aligned} \mathcal{N}(A) &= \left(\left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix} \right\} \right) \\ &= \text{gen} \left(\left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix} \right\} \right) \end{aligned}$$

Por tanto, $\mathcal{N}(A) = \text{gen} \left(\left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix} \right\} \right)$. Además, $\nu(A) = 0$.

Definición 2.4.3:

Sea $A \in \mathcal{M}_{m \times n}(\mathbb{R})$, se define la *imagen* de A , denotado por $\mathcal{R}(A)$, como

$$\mathcal{R}(A) = \{ \mathbf{y} \in \mathbb{R}^m \mid A\mathbf{x} = \mathbf{y}, \text{ para algún } \mathbf{x} \in \mathbb{R}^n \}$$

Teorema 2.4.2: Sea $A \in \mathcal{M}_{m \times n}(\mathbb{R})$, entonces $\mathcal{R}(A)$ es un subespacio de \mathbb{R}^m

Demostración:

i) Sean $\mathbf{y}_1, \mathbf{y}_2 \in \mathcal{R}(A)$, entonces

$$A\mathbf{x}_1 = \mathbf{y}_1, \text{ para algún } \mathbf{x}_1 \in \mathbb{R}^n \quad (2.4.1)$$

y

$$A\mathbf{x}_2 = \mathbf{y}_2, \text{ para algún } \mathbf{x}_2 \in \mathbb{R}^n \quad (2.4.2)$$

Así, de las expresiones (2.4.1) y (2.4.2),

$$\begin{aligned} \mathbf{y}_1 + \mathbf{y}_2 &= A\mathbf{x}_1 + A\mathbf{x}_2 \\ &= A(\mathbf{x}_1 + \mathbf{x}_2) \\ &= A\mathbf{w} \end{aligned}$$

siendo $w = x_1 + x_2$. Por tanto, $y_1 + y_2 \in \mathcal{R}(A)$, es decir, se cumple la cerradura bajo la suma.

ii) Se deja como ejercicio al lector.

De (i) y (ii), se sigue que $\mathcal{R}(A)$ es un subespacio de \mathbb{R}^m .

Definición 2.4.4:

Llamaremos *rango* de A , denotado por $\rho(A)$, a la dimensión de la imagen, es decir,

$$\rho(A) = \dim(\mathcal{R}(A))$$

Ejemplo 2.4.3:

Determine $\mathcal{R}(A)$ y $\rho(A)$ de la siguiente matriz

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

Solución: Por definición,

$$\begin{aligned} \mathcal{R}(A) &= \{y \in \mathbb{R}^2 \mid Ax = y, \text{ para algún } x \in \mathbb{R}^2\} \\ &= \left\{ \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \in \mathbb{R}^2 \mid \begin{bmatrix} 1 & 2 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}, \text{ para algún } x \in \mathbb{R}^2 \right\} \\ &= \left\{ \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \in \mathbb{R}^2 \mid \begin{bmatrix} x_1 + 2x_2 \\ x_1 - x_2 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}, \text{ para algún } x \in \mathbb{R}^2 \right\} \\ &= \left\{ \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = x_1 \begin{pmatrix} 1 \\ 1 \end{pmatrix} + x_2 \begin{pmatrix} 2 \\ -1 \end{pmatrix} \mid x_1, x_2 \in \mathbb{R} \right\} \\ &= \text{gen} \left(\left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ -1 \end{pmatrix} \right\} \right) \end{aligned}$$

Observemos que el conjunto de vectores anterior es linealmente independiente.

Por tanto, $\mathcal{R}(A) = \text{gen} \left(\left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ -1 \end{pmatrix} \right\} \right)$ y $\rho(A) = 2$.

Ejemplo 2.4.4:

Determine $\mathcal{R}(A)$ y $\rho(A)$ de la siguiente matriz

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

Solución: Por definición,

$$\begin{aligned} \mathcal{R}(A) &= \{y \in \mathbb{R}^2 \mid Ax = y, \text{ para algún } x \in \mathbb{R}^3\} \\ &= \left\{ \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \in \mathbb{R}^2 \mid \begin{bmatrix} 1 & 2 & -1 \\ 2 & -1 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}, \text{ para algún } x \in \mathbb{R}^3 \right\} \\ &= \left\{ \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \in \mathbb{R}^2 \mid \begin{bmatrix} x_1 + 2x_2 - x_3 \\ 2x_1 - x_2 + 3x_3 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}, \text{ para algún } x \in \mathbb{R}^3 \right\} \\ &= \left\{ \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = x_1 \begin{pmatrix} 1 \\ 2 \end{pmatrix} + x_2 \begin{pmatrix} 2 \\ -1 \end{pmatrix} + x_3 \begin{pmatrix} -1 \\ 3 \end{pmatrix} \mid x_1, x_2, x_3 \in \mathbb{R} \right\} \\ &= \text{gen} \left(\left\{ \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ -1 \end{pmatrix}, \begin{pmatrix} -1 \\ 3 \end{pmatrix} \right\} \right) \end{aligned}$$

Observemos que el conjunto anterior de vectores no es linealmente independiente, pues el primer vector puede expresarse como combinación lineal de los otros dos.

En particular, si denotamos los vectores respectivamente como v_1, v_2, v_3 , entonces existen escalares $\alpha = 1$ y $\beta = 1$ tales que $v_1 = \alpha v_2 + \beta v_3$. Por tanto, esto implica

que $\mathcal{R}(A) = \text{gen} \left(\left\{ \begin{pmatrix} 2 \\ -1 \end{pmatrix}, \begin{pmatrix} -1 \\ 3 \end{pmatrix} \right\} \right)$ y $\rho(A) = 2$.

Definición 2.4.5:

Si $A \in \mathcal{M}_{m \times n}(\mathbb{R})$, sean $\{\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_m\}$ los renglones de A y $\{\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_n\}$ las columnas de A . Es decir,

$$A = \begin{matrix} & \mathbf{c}_1 & \mathbf{c}_2 & \cdots & \mathbf{c}_j & \cdots & \mathbf{c}_n \\ \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1j} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2j} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{ij} & \cdots & a_{in} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mj} & \cdots & a_{mn} \end{bmatrix} & \mathbf{r}_1 \\ & & & & & & \mathbf{r}_2 \\ & & & & & & \vdots \\ & & & & & & \mathbf{r}_i \\ & & & & & & \vdots \\ & & & & & & \mathbf{r}_m \end{matrix}$$

Definimos el espacio renglón de A , denotado por R_A , como

$$R_A = \text{gen}(\{\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_m\}) \subseteq \mathbb{R}^n$$

y el espacio columna, denotado por C_A , como

$$C_A = \text{gen}(\{\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_n\}) \subseteq \mathbb{R}^m$$

Teorema 2.4.3: La imagen de una matriz es igual al espacio columna.

Demostración: Para demostrar que $C_A = \mathcal{R}(A)$, tenemos que demostrar que $\mathcal{R}(A) \subseteq C_A$ e $C_A \subseteq \mathcal{R}(A)$.

- i) Queremos probar que $\mathcal{R}(A) \subseteq C_A$. Supongamos que $\mathbf{y} \in \mathcal{R}(A)$. Entonces existe un vector \mathbf{x} tal que $\mathbf{y} = A\mathbf{x}$, pero podemos expresar a $A\mathbf{x}$ como una combinación lineal de las columnas de A . Para ver esto, consideremos

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & \ddots & \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \quad \text{y} \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

Entonces

$$A\mathbf{x} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{bmatrix}$$

y por consiguiente,

$$A\mathbf{x} = x_1 \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix} + x_2 \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix} + \cdots + x_n \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix}.$$

Por lo tanto, $\mathbf{y} \in C_A$, de manera que $\mathcal{R}(A) \subseteq C_A$.

- ii) Se deja como ejercicio al lector.

Si A es la matriz cero de tamaño $m \times n$, entonces el sistema se convierte en

$$\mathbf{0}_{m \times n} \mathbf{x} = \mathbf{y}.$$

Es decir, la única salida posible es el vector cero en \mathbb{R}^m . Esto implica que

$$\mathcal{R}(\mathbf{0}) = \{\mathbf{0}_m\}.$$

Esto significa que la imagen de la matriz cero es el subespacio trivial $\{\mathbf{0}_m\}$ en \mathbb{R}^m . Dado que $\mathcal{R}(\mathbf{0}) = \{\mathbf{0}_m\}$, su dimensión es 0. Es decir,

$$\rho(\mathbf{0}) = 0.$$

Teorema 2.4.4: Si A es una matriz de $m \times n$, entonces

$$\dim R_A = \dim C_A = \dim (\mathcal{R}(A)) = \rho(A)$$

Demostración: Denotaremos por α_{ij} la componente ij de A . Debemos demostrar que $\dim R_A = \dim C_A$. Los renglones de A se denotan por $\mathbb{r}_1, \mathbb{r}_2, \dots, \mathbb{r}_m$ y sea $k = \dim R_A$. Sea $S = \{\mathbb{s}_1, \mathbb{s}_2, \dots, \mathbb{s}_k\}$ una base para R_A . Entonces cada renglón de A se puede expresar como una combinación lineal de los vectores en S , y se tiene, para algunas constantes α_{ij} ,

$$\begin{aligned} \mathbb{r}_1 &= \alpha_{11}\mathbb{s}_1 + \alpha_{12}\mathbb{s}_2 + \dots + \alpha_{1k}\mathbb{s}_k \\ \mathbb{r}_2 &= \alpha_{21}\mathbb{s}_1 + \alpha_{22}\mathbb{s}_2 + \dots + \alpha_{2k}\mathbb{s}_k \\ &\vdots \\ \mathbb{r}_m &= \alpha_{m1}\mathbb{s}_1 + \alpha_{m2}\mathbb{s}_2 + \dots + \alpha_{mk}\mathbb{s}_k \end{aligned} \quad (2.4.3)$$

Ahora la componente j de \mathbb{r}_i es a_{ij} . Entonces si se igualan las componentes j de ambos lados de (2.4.3), se obtiene

$$\begin{aligned} a_{1j} &= \alpha_{11}s_{1j} + \alpha_{12}s_{2j} + \dots + \alpha_{1k}s_{kj} \\ a_{2j} &= \alpha_{21}s_{1j} + \alpha_{22}s_{2j} + \dots + \alpha_{2k}s_{kj} \\ &\vdots \\ a_{mj} &= \alpha_{m1}s_{1j} + \alpha_{m2}s_{2j} + \dots + \alpha_{mk}s_{kj} \end{aligned} \quad \text{donde } \mathbb{s}_i = \begin{pmatrix} s_{i1} \\ s_{i2} \\ \vdots \\ s_{in} \end{pmatrix}$$

Es decir,

$$\begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{bmatrix} = s_{1j} \underbrace{\begin{bmatrix} \alpha_{11} \\ \alpha_{21} \\ \vdots \\ \alpha_{m1} \end{bmatrix}}_{\mathbb{w}_1} + s_{2j} \underbrace{\begin{bmatrix} \alpha_{12} \\ \alpha_{22} \\ \vdots \\ \alpha_{m2} \end{bmatrix}}_{\mathbb{w}_2} + \dots + s_{kj} \underbrace{\begin{bmatrix} \alpha_{1k} \\ \alpha_{2k} \\ \vdots \\ \alpha_{mk} \end{bmatrix}}_{\mathbb{w}_k} \quad (2.4.4)$$

Entonces como el lado izquierdo de (2.4.4) es la columna j de A , se observa que cada columna de A se puede escribir como una combinación lineal de $\mathbb{w}_1, \mathbb{w}_2, \dots, \mathbb{w}_k$, lo que significa que los vectores $\mathbb{w}_1, \mathbb{w}_2, \dots, \mathbb{w}_k$ generan a C_A y

$$\dim C_A \leq k = \dim R_A. \quad (2.4.5)$$

Pero la ecuación (2.4.5) se cumple para cualquier matriz A , y en particular para A^T . Pero $C_{A^T} = R_A$ y $R_{A^T} = C_A$. Como de (2.4.5) $\dim C_{A^T} \leq \dim R_{A^T}$, se tiene

$$\dim R_A \leq \dim C_A. \quad (2.4.6)$$

Combinando (2.4.5) y (2.4.6) la demostración queda completa.

Ejemplo 2.4.5: Tomando la matriz del ejemplo 2.4.4, tenemos que

$$R_A = \text{gen} \left(\left\{ \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}, \begin{pmatrix} 2 \\ -1 \\ 3 \end{pmatrix} \right\} \right), \quad C_A = \mathcal{R}(A) = \text{gen} \left(\left\{ \begin{pmatrix} 2 \\ -1 \end{pmatrix}, \begin{pmatrix} -1 \\ 3 \end{pmatrix} \right\} \right).$$

De acuerdo con el teorema demostrado previamente, tenemos que

$$\dim R_A = \dim C_A = \dim (\mathcal{R}(A)) = \rho(A)$$

En este caso particular, se verifica que $\dim R_A = \dim C_A = \dim (\mathcal{R}(A)) = 2$.

El siguiente teorema establece una relación fundamental entre el rango y la nulidad de una matriz.

Teorema 2.4.5: Si A es una matriz con n columnas, entonces

$$\rho(A) + \nu(A) = n. \quad (2.4.7)$$

De lo visto anteriormente, hay seis espacios vectoriales importantes asociados con una matriz A y A^T :

espacio renglón de A	espacio renglón de A^T
espacio columna de A	espacio columna de A^T
espacio nulo de A	espacio nulo de A^T

Sin embargo, transponer una matriz convierte los vectores renglón en vectores columna y viceversa, por lo tanto, salvo por una diferencia de notación, el espacio renglón de A^T es el mismo que el espacio columna de A , y el espacio columna de A^T es el mismo que el espacio renglón de A . Así, de los seis espacios mencionados anteriormente, solo los siguientes cuatro son distintos:

espacio renglón de A	espacio columna de A
espacio nulo de A	espacio nulo de A^T

A estos se les llama los *espacios fundamentales* de una matriz A . Ahora consideraremos cómo se relacionan estos cuatro subespacios.

Pongamos atención por un momento en la matriz A^T . Como el espacio renglón y el espacio columna de una matriz tienen la misma dimensión, y como al transponer una matriz se intercambian sus columnas por filas y sus filas por columnas, el siguiente resultado no debería ser sorprendente.

Teorema 2.4.6: Si A es cualquier matriz, entonces

$$\rho(A) = \rho(A^T).$$

Demostración: Es evidente, pues $\rho(A) = \dim(R_A) = \dim(C_{A^T}) = \rho(A^T)$.

Este resultado tiene algunas implicaciones importantes. Por ejemplo, si A es una matriz $m \times n$, entonces al aplicar la fórmula (2.4.7) a la matriz A^T y usando el hecho de que esta matriz tiene m columnas, se obtiene

$$\rho(A^T) + \nu(A^T) = m$$

lo cual, en virtud del teorema 2.4.6, se puede reescribir como

$$\rho(A) + \nu(A^T) = m.$$

Esta forma alternativa de la fórmula (2.4.7) permite expresar las dimensiones de los cuatro espacios fundamentales en términos del tamaño y el rango de A . Específicamente, si $\rho(A) = r$, entonces

$$\begin{aligned} \dim(R_A) &= r & \dim(C_A) &= r \\ \dim(\mathcal{N}(A)) &= n - r & \dim(\mathcal{N}(A^T)) &= m - r \end{aligned} \quad (2.4.8)$$

Las cuatro fórmulas en (2.4.8) proporcionan una relación *algebraica* entre el tamaño de una matriz y las dimensiones de sus espacios fundamentales.

Ejemplo 2.4.6: ¿Cuál es el rango máximo posible de una matriz $m \times n$ que no es cuadrada?

Solución: Dado que los vectores renglón de A pertenecen a \mathbb{R}^n y los vectores columna a \mathbb{R}^m , el espacio renglón de A tiene como máximo dimensión n y el espacio columna como máximo dimensión m . Como el rango de A es la dimensión común de su espacio renglón y su espacio columna, se deduce que el rango es como máximo el menor entre m y n . Esto se expresa como: $\rho(A) \leq \min(m, n)$.

2.5 Operaciones elementales en los renglones de una matriz

Las operaciones elementales por renglón son un conjunto de tres operaciones básicas que se pueden aplicar a los renglones de una matriz sin alterar el espacio renglón de la misma. Estas operaciones son fundamentales en el álgebra lineal para simplificar matrices y resolver sistemas de ecuaciones lineales. La idea principal es manipular la matriz de manera sistemática para alcanzar una forma más sencilla, como la forma escalonada o la forma escalonada reducida, lo que facilita el análisis de sus propiedades. Así, dada una matriz, las operaciones elementales son:

- **Intercambiar dos renglones:** Si intercambiamos los renglones i y j , escribiremos $r_i \rightleftharpoons r_j$.
- **Multiplicar un renglón por una constante no cero:** Si multiplicamos el renglón i por la constante no cero α , escribiremos $r_i \leftarrow \alpha r_i$.
- **Sumarle a un renglón un múltiplo de otro renglón:** Si le sumamos al renglón i , α -veces el renglón j , escribiremos $r_i \leftarrow r_i + \alpha r_j$.

Para denotar una operación elemental, se utiliza una flecha orientada hacia la derecha que indica el paso de una matriz a otra. Encima de esta flecha se escribe la operación elemental correspondiente, especificando claramente qué tipo de operación se está realizando. Por ejemplo:

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \xrightarrow{r_1 \rightleftharpoons r_2} \begin{bmatrix} 3 & 4 \\ 1 & 2 \end{bmatrix}, \quad \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \xrightarrow{r_1 \leftarrow 3r_1} \begin{bmatrix} 3 & 6 \\ 3 & 4 \end{bmatrix}, \quad \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \xrightarrow{r_2 \leftarrow r_2 + 2r_1} \begin{bmatrix} 1 & 2 \\ 5 & 8 \end{bmatrix}$$

Definición 2.5.1:

Sean $A, B \in \mathcal{M}_{m \times n}(\mathbb{R})$. Decimos que A es equivalente por renglones a B , si existe un número finito de operaciones elementales tales que al aplicárselas a la matriz A se obtiene la matriz B .

Definición 2.5.2:

Decimos que una matriz $A \in \mathcal{M}_{m \times n}(\mathbb{R})$ está en forma escalonada si:

- Los renglones cero están en la parte inferior de la matriz.
- Los renglones no cero satisfacen que su primera entrada no cero es 1, llamado uno principal.
- En dos renglones no cero consecutivos, el uno principal del renglón inferior aparece más a la derecha que el uno principal del renglón superior.

Ejemplo 2.5.1:

Un ejemplo muy claro es

$$A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & 5 & 6 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Esta matriz está en forma escalonada, ya que:

- Esta matriz no tiene renglones completamente cero, por lo que la primera condición se cumple trivialmente.
- En el primer renglón, la primera entrada no cero es 1 (columna 1). En el segundo renglón, la primera entrada no cero es 1 (columna 2, después del cero inicial). En el tercer renglón, la primera entrada no cero es 1 (columna 4). Todos los renglones no cero cumplen con la segunda condición.
- Del primer al segundo renglón, el uno principal del segundo renglón (columna 2) está a la derecha del uno principal del primer renglón (columna 1). Del segundo al tercer renglón, el uno principal del tercer renglón (columna 4) está a la derecha del uno principal del segundo renglón (columna 2). Esto satisface la tercera condición.

Definición 2.5.3:

Decimos que una matriz $A \in \mathcal{M}_{m \times n}(\mathbb{R})$ está en forma escalonada reducida si la matriz está en forma escalonada y cada columna que contenga un uno principal tiene ceros en todas sus demás entradas.

Ejemplo 2.5.2:

Un ejemplo muy claro es

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

Esta matriz está en forma escalonada reducida, ya que: está en forma escalonada y cada columna con un uno principal (columnas 1, 2 y 4) tiene ceros en todas las demás entradas. La columna 3 no tiene un uno principal, por lo que no hay restricción.

Teorema 2.5.1:

Toda matriz tamaño $m \times n$ no cero es equivalente por renglones a una matriz en forma escalonada.

Teorema 2.5.2:

Toda matriz tamaño $m \times n$ no cero es equivalente por renglones a una única matriz en forma escalonada reducida.

Después de haber definido ya las tres operaciones elementales sobre renglones, estamos en condiciones de usarlas para resolver sistemas de ecuaciones lineales de manera sistemática.

A medida que aumenta el número de ecuaciones e incógnitas en un sistema lineal, también lo hace la complejidad del álgebra involucrado en encontrar soluciones. Los cálculos requeridos pueden hacerse más manejables simplificando la notación y estandarizando los procedimientos. Del sistema visto en (2.3.1), podemos abreviar el sistema escribiendo únicamente el arreglo rectangular de números como

$$\left[\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & & \ddots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{array} \right].$$

Esto se llama la *matriz aumentada* del sistema. Por ejemplo, la matriz aumentada para el sistema de ecuaciones

$$\begin{array}{l} 7x_1 + 2x_2 + 4x_3 = 9 \\ 2x_1 + 4x_2 - 3x_3 = 1 \\ 3x_1 - 8x_2 - 5x_3 = 0 \end{array} \quad \text{es} \quad \left[\begin{array}{ccc|c} 7 & 2 & 4 & 9 \\ 2 & 4 & -3 & 1 \\ 3 & -8 & -5 & 0 \end{array} \right].$$

Método de eliminación Gaussiana:

La eliminación Gaussiana es un método algebraico utilizado para resolver sistemas de ecuaciones lineales. Consiste en transformar una matriz aumentada del sistema en una forma escalonada mediante operaciones elementales por renglones. El procedimiento es el siguiente:

1. Considerar la matriz aumentada del sistema, llamémosle A .
2. Aplicar a la matriz A operaciones elementales hasta obtener una matriz en forma escalonada, llamémosle R .
3. Escribir el sistema de ecuaciones que tiene por matriz aumentada a la matriz R .
4. Se despeja el valor de la última incógnita que esté asociada a un uno principal.
5. Se usa la sustitución hacia atrás para determinar las demás incógnitas que estén asociadas a un uno principal.
6. Escribir el conjunto solución, llamémosle S .

Ejemplo 2.5.3: Resuelva el siguiente sistema por el método de eliminación Gaussiana:

$$-x - y + 3z = -1$$

$$5x - y - 3z = -7$$

$$4x + 3y + z = 2$$

Solución: Apliquemos el método de eliminación Gaussiana. Primero formamos la matriz aumentada del sistema, la cual es

$$\left[\begin{array}{ccc|c} -1 & -1 & 3 & -1 \\ 5 & -1 & -3 & -7 \\ 4 & 3 & 1 & 2 \end{array} \right].$$

Ahora, aplicamos operaciones elementales hasta llevarla a su forma escalonada. Sumando 5 veces el primer renglón al segundo renglón, obtenemos

$$\left[\begin{array}{ccc|c} -1 & -1 & 3 & -1 \\ 5 & -1 & -3 & -7 \\ 4 & 3 & 1 & 2 \end{array} \right] \xrightarrow{r_2 \leftarrow r_2 + 5r_1} \left[\begin{array}{ccc|c} -1 & -1 & 3 & -1 \\ 0 & -6 & 12 & -12 \\ 4 & 3 & 1 & 2 \end{array} \right]$$

Sumando 4 veces el primer renglón al tercer renglón, obtenemos

$$\left[\begin{array}{ccc|c} -1 & -1 & 3 & -1 \\ 0 & -6 & 12 & -12 \\ 4 & 3 & 1 & 2 \end{array} \right] \xrightarrow{r_3 \leftarrow r_3 + 4r_1} \left[\begin{array}{ccc|c} -1 & -1 & 3 & -1 \\ 0 & -6 & 12 & -12 \\ 0 & -1 & 13 & -2 \end{array} \right]$$

Multiplicando por -1 el primer renglón, obtenemos

$$\left[\begin{array}{ccc|c} -1 & -1 & 3 & -1 \\ 0 & -6 & 12 & -12 \\ 0 & -1 & 13 & -2 \end{array} \right] \xrightarrow{r_1 \leftarrow (-1)r_1} \left[\begin{array}{ccc|c} 1 & 1 & -3 & 1 \\ 0 & -6 & 12 & -12 \\ 0 & -1 & 13 & -2 \end{array} \right]$$

Multiplicando por $-\frac{1}{6}$ el segundo renglón, obtenemos

$$\left[\begin{array}{ccc|c} 1 & 1 & -3 & 1 \\ 0 & -6 & 12 & -12 \\ 0 & -1 & 13 & -2 \end{array} \right] \xrightarrow{r_2 \leftarrow (-\frac{1}{6})r_2} \left[\begin{array}{ccc|c} 1 & 1 & -3 & 1 \\ 0 & 1 & -2 & 2 \\ 0 & -1 & 13 & -2 \end{array} \right]$$

Sumando 1 vez el segundo renglón al tercer renglón, obtenemos

$$\left[\begin{array}{ccc|c} 1 & 1 & -3 & 1 \\ 0 & 1 & -2 & 2 \\ 0 & -1 & 13 & -2 \end{array} \right] \xrightarrow{r_3 \leftarrow r_3 + r_2} \left[\begin{array}{ccc|c} 1 & 1 & -3 & 1 \\ 0 & 1 & -2 & 2 \\ 0 & 0 & 11 & 0 \end{array} \right]$$

Multiplicando por $\frac{1}{11}$ el tercer renglón, obtenemos

$$\left[\begin{array}{ccc|c} 1 & 1 & -3 & 1 \\ 0 & 1 & -2 & 2 \\ 0 & 0 & 11 & 0 \end{array} \right] \xrightarrow{r_3 \leftarrow (\frac{1}{11})r_3} \left[\begin{array}{ccc|c} 1 & 1 & -3 & 1 \\ 0 & 1 & -2 & 2 \\ 0 & 0 & 1 & 0 \end{array} \right] = R$$

Ahora bien, escribimos el sistema de ecuaciones que tiene por matriz aumentada a la matriz R :

$$x + y - 3z = 1$$

$$y - 2z = 2$$

$$z = 0$$

Del sistema, ya sabemos que $z = 0$, sustituyendo en la segunda ecuación obtenemos que $y = 2$. Luego, sustituyendo estos dos valores en la primera ecuación obtenemos que $x = -1$. Por lo tanto, el conjunto solución es

$$S = \left\{ \begin{pmatrix} -1 \\ 2 \\ 0 \end{pmatrix} \right\}.$$

Es importante destacar que no existe un único camino para resolver un sistema de ecuaciones lineales. Aunque todos los caminos correctos conducen, en principio, a la misma solución (si existe una), las operaciones que se aplican pueden variar dependiendo de la estrategia o del orden en que se elijan. Esta flexibilidad también implica que es fundamental verificar cuidadosamente cada paso, ya que una mala elección o un error de cálculo puede llevar a una matriz incorrecta.

Método de eliminación Gauss-Jordan:

La eliminación Gauss-Jordan es también un método algebraico utilizado para resolver sistemas de ecuaciones lineales. Este método es una extensión natural de la eliminación gaussiana, y tiene como objetivo llevar la matriz no sólo a una forma más sencilla, sino a su forma más reducida mediante operaciones elementales por renglones. El procedimiento es el siguiente:

1. Considerar la matriz aumentada del sistema, llamémosle A .
2. Aplicar a la matriz A operaciones elementales hasta obtener una matriz en forma escalonada reducida, llamémosle R .
3. Escribir el sistema de ecuaciones que tiene por matriz aumentada a la matriz R .
4. Despejar las incógnitas que estén asociadas a un uno principal.
5. Escribir el conjunto solución, llamémosle S .

Ejemplo 2.5.4: Resuelva el siguiente sistema por el método de eliminación Gauss-Jordan:

$$\begin{aligned}x + 3y - 2z &= 1 \\2x - 9y + 2z &= 2 \\9x + 4y + 2z &= 9\end{aligned}$$

Solución: Apliquemos el método de eliminación Gauss-Jordan. Primero formamos la matriz aumentada del sistema, la cual es

$$\left[\begin{array}{ccc|c} 1 & 3 & -2 & 1 \\ 2 & -9 & 2 & 2 \\ 9 & 4 & 2 & 9 \end{array} \right].$$

Ahora, aplicamos operaciones elementales hasta llevarla a su forma escalonada reducida. Sumando -2 veces el primer renglón al segundo renglón, obtenemos

$$\left[\begin{array}{ccc|c} 1 & 3 & -2 & 1 \\ 2 & -9 & 2 & 2 \\ 9 & 4 & 2 & 9 \end{array} \right] \xrightarrow{r_2 \leftarrow r_2 + (-2)r_1} \left[\begin{array}{ccc|c} 1 & 3 & -2 & 1 \\ 0 & -15 & 6 & 0 \\ 9 & 4 & 2 & 9 \end{array} \right]$$

Sumando -9 veces el primer renglón al tercer renglón, obtenemos

$$\left[\begin{array}{ccc|c} 1 & 3 & -2 & 1 \\ 0 & -15 & 6 & 0 \\ 9 & 4 & 2 & 9 \end{array} \right] \xrightarrow{r_3 \leftarrow r_3 + (-9)r_1} \left[\begin{array}{ccc|c} 1 & 3 & -2 & 1 \\ 0 & -15 & 6 & 0 \\ 0 & -23 & 20 & 0 \end{array} \right]$$

Multiplicando por $\frac{1}{3}$ el segundo renglón, obtenemos

$$\left[\begin{array}{ccc|c} 1 & 3 & -2 & 1 \\ 0 & -15 & 6 & 0 \\ 0 & -23 & 20 & 0 \end{array} \right] \xrightarrow{r_2 \leftarrow \left(\frac{1}{3}\right)r_2} \left[\begin{array}{ccc|c} 1 & 3 & -2 & 1 \\ 0 & -5 & 2 & 0 \\ 0 & -23 & 20 & 0 \end{array} \right]$$

Sumando 1 vez el primer renglón al segundo renglón, obtenemos

$$\left[\begin{array}{ccc|c} 1 & 3 & -2 & 1 \\ 0 & -5 & 2 & 0 \\ 0 & -23 & 20 & 0 \end{array} \right] \xrightarrow{r_1 \leftarrow r_1 + r_2} \left[\begin{array}{ccc|c} 1 & -2 & 0 & 1 \\ 0 & -5 & 2 & 0 \\ 0 & -23 & 20 & 0 \end{array} \right]$$

Sumando -10 veces el segundo renglón al tercer renglón, obtenemos

$$\left[\begin{array}{ccc|c} 1 & -2 & 0 & 1 \\ 0 & -5 & 2 & 0 \\ 0 & -23 & 20 & 0 \end{array} \right] \xrightarrow{r_3 \leftarrow r_3 + (-10)r_2} \left[\begin{array}{ccc|c} 1 & -2 & 0 & 1 \\ 0 & -5 & 2 & 0 \\ 0 & 27 & 0 & 0 \end{array} \right]$$

Multiplicando por $\frac{1}{27}$ el tercer renglón, obtenemos

$$\left[\begin{array}{ccc|c} 1 & -2 & 0 & 1 \\ 0 & -5 & 2 & 0 \\ 0 & 27 & 0 & 0 \end{array} \right] \xrightarrow{r_3 \leftarrow \left(\frac{1}{27}\right)r_3} \left[\begin{array}{ccc|c} 1 & -2 & 0 & 1 \\ 0 & -5 & 2 & 0 \\ 0 & 1 & 0 & 0 \end{array} \right]$$

Sumando 5 veces el tercer renglón al segundo renglón, obtenemos

$$\left[\begin{array}{ccc|c} 1 & -2 & 0 & 1 \\ 0 & -5 & 2 & 0 \\ 0 & 1 & 0 & 0 \end{array} \right] \xrightarrow{r_2 \leftarrow r_2 + 5r_3} \left[\begin{array}{ccc|c} 1 & -2 & 0 & 1 \\ 0 & 0 & 2 & 0 \\ 0 & 1 & 0 & 0 \end{array} \right]$$

Sumando 2 veces el tercer renglón al primer renglón, obtenemos

$$\left[\begin{array}{ccc|c} 1 & -2 & 0 & 1 \\ 0 & 0 & 2 & 0 \\ 0 & 1 & 0 & 0 \end{array} \right] \xrightarrow{r_1 \leftarrow r_1 + 2r_3} \left[\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 0 & 2 & 0 \\ 0 & 1 & 0 & 0 \end{array} \right]$$

Intercambiando el segundo renglón y el tercer renglón, obtenemos

$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 0 & 2 & 0 \\ 0 & 1 & 0 & 0 \end{array} \right] \xrightarrow{r_2 \leftrightarrow r_3} \left[\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \end{array} \right]$$

Multiplicando por $\frac{1}{2}$ el tercer renglón, obtenemos

$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \end{array} \right] \xrightarrow{r_3 \leftarrow \left(\frac{1}{2}\right)r_3} \left[\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array} \right] = R$$

Ahora bien, escribimos el sistema de ecuaciones que tiene por matriz aumentada a la matriz R :

$$\begin{aligned} x &= 1 \\ y &= 0 \\ z &= 0 \end{aligned}$$

Pero las incógnitas ya están despejadas. Por lo tanto, el conjunto solución es

$$S = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \right\}.$$

Ejemplo 2.5.5: Encuentra la nulidad y el rango de la matriz

$$A = \begin{bmatrix} -1 & 2 & 0 & 4 & 5 & -3 \\ 3 & -7 & 2 & 0 & 1 & 4 \\ 2 & -5 & 2 & 4 & 6 & 1 \\ 4 & -9 & 2 & -4 & -4 & 7 \end{bmatrix}.$$

Solución: Para poder encontrar la nulidad de la matriz A , primero debemos encontrar el espacio nulo de A . Sin embargo, por definición es equivalente a encontrar la solución de $A\mathbf{x} = \mathbf{0}$. También podemos aplicar el método de eliminación

de Gauss-Jordan. Formamos la matriz aumentada,

$$\left[\begin{array}{cccccc|c} -1 & 2 & 0 & 4 & 5 & -3 & 0 \\ 3 & -7 & 2 & 0 & 1 & 4 & 0 \\ 2 & -5 & 2 & 4 & 6 & 1 & 0 \\ 4 & -9 & 2 & -4 & -4 & 7 & 0 \end{array} \right].$$

Aplicando operaciones elementales a dicha matriz, obtenemos que su forma escalonada reducida es (verificar):

$$\left[\begin{array}{cccccc|c} 1 & 0 & -4 & -28 & -37 & 13 & 0 \\ 0 & 1 & -2 & -12 & -16 & 5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right]. \quad (2.5.1)$$

Por tanto, el sistema correspondiente de ecuaciones es

$$x_1 - 4x_3 - 28x_4 - 37x_5 + 13x_6 = 0$$

$$x_2 - 2x_3 - 12x_4 - 16x_5 + 5x_6 = 0$$

Despejando las incógnitas que estén asociadas a un uno principal, tenemos que

$$x_1 = 4x_3 + 28x_4 + 37x_5 - 13x_6$$

$$x_2 = 2x_3 + 12x_4 + 16x_5 - 5x_6$$

Resolviendo estas ecuaciones para las variables principales se obtiene

$$x_1 = 4r + 28s + 37t - 13u$$

$$x_2 = 2r + 12s + 16t - 5u$$

$$x_3 = r$$

$$x_4 = s$$

$$x_5 = t$$

$$x_6 = u$$

Por lo tanto, el conjunto solución del sistema homogéneo es

$$S = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{pmatrix} = r \begin{pmatrix} 4 \\ 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + s \begin{pmatrix} 28 \\ 12 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} + t \begin{pmatrix} 37 \\ 16 \\ 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} + u \begin{pmatrix} -13 \\ -5 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \mid r, s, t, u \in \mathbb{R} \right\}.$$

Es decir,

$$\mathcal{N}(A) = \text{gen} \left(\left\{ \begin{pmatrix} 4 \\ 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 28 \\ 12 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 37 \\ 16 \\ 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -13 \\ -5 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \right\} \right),$$

de donde se sigue que $\nu(A) = 4$. Para hallar el rango de A , debemos encontrar la dimensión del espacio solución del sistema lineal $A\mathbf{x} = \mathbf{y}$. Este sistema puede resolverse reduciendo la matriz aumentada a su forma escalonada reducida. La matriz resultante será idéntica a la de (2.5.1), excepto que no tendrá una columna adicional de ceros al final, y dado que esta matriz tiene dos unos principales, su espacio columna es bidimensional y $\rho(A) = 2$.

2.6 Matrices no singulares y matrices elementales

En los números reales, todo número no nulo a tiene un recíproco a^{-1} con la propiedad

$$a \cdot a^{-1} = a^{-1} \cdot a = 1.$$

Al número a^{-1} se le llama comúnmente como el *inverso multiplicativo* de a . Nuestro siguiente objetivo es desarrollar un análogo de este resultado para la aritmética de matrices. Para este propósito, hacemos la siguiente definición.

Definición 2.6.1:

Si A es una matriz cuadrada y si se puede encontrar una matriz B del mismo tamaño tal que $AB = BA = I$, entonces se dice que A es *invertible* o *no singular* y que B es una *inversa* de A . Si no se puede encontrar tal matriz B , entonces se dice que A es *no invertible* o *singular*.

La relación $AB = BA = I$ no cambia al intercambiar A y B , así que si A es invertible y B es una inversa de A , entonces también es cierto que B es invertible y A es una inversa de B . Así, cuando

$$AB = BA = I$$

decimos que A y B son *inversas una de la otra*.

Ejemplo 2.6.1:

Sean

$$A = \begin{bmatrix} 2 & -5 \\ -1 & 3 \end{bmatrix} \quad \text{y} \quad B = \begin{bmatrix} 3 & 5 \\ 1 & 2 \end{bmatrix}.$$

Entonces

$$\begin{aligned} AB &= \begin{bmatrix} 2 & -5 \\ -1 & 3 \end{bmatrix} \begin{bmatrix} 3 & 5 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I \\ BA &= \begin{bmatrix} 3 & 5 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 2 & -5 \\ -1 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I \end{aligned}$$

Por lo tanto, A y B son no singulares y cada una es inversa de la otra.

Es razonable preguntarse si una matriz invertible puede tener más de una inversa. El siguiente teorema muestra que la respuesta es no: una matriz invertible tiene exactamente una inversa.

Teorema 2.6.1:

Si B y C son ambas inversas de la matriz A , entonces $B = C$.

Demostración: Como B es una inversa de A , tenemos que $BA = I$. Multiplicando ambos lados por la derecha con C se obtiene

$$(BA)C = IC = C.$$

Pero también es cierto que

$$(BA)C = B(AC) = BI = B,$$

por lo tanto, $C = B$.

Como consecuencia de este resultado importante, ahora podemos hablar de “la” inversa de una matriz invertible. Si A es invertible, entonces su inversa se denotará con el símbolo A^{-1} . Así,

$$AA^{-1} = I \quad \text{y} \quad A^{-1}A = I.$$

La inversa de A cumple un papel muy similar en la aritmética de matrices al que cumple el recíproco a^{-1} en las relaciones numéricas $aa^{-1} = 1$ y $a^{-1}a = 1$.

i Nota

El símbolo A^{-1} no debe interpretarse como $1/A$. La división entre matrices no será una operación definida en este texto.

El siguiente teorema se refiere a las inversas de productos de matrices.

Teorema 2.6.2: Si A y B son matrices invertibles del mismo tamaño, entonces AB es invertible y

$$(AB)^{-1} = B^{-1}A^{-1}.$$

Demostración: Podemos establecer que AB es invertible y obtener la fórmula dada al mismo tiempo, mostrando que

$$(AB)(B^{-1}A^{-1}) = (B^{-1}A^{-1})(AB) = I.$$

Pero

$$(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AIA^{-1} = AA^{-1} = I$$

y de manera similar,

$$(B^{-1}A^{-1})(AB) = I.$$

Aunque no lo demostraremos, este resultado puede extenderse a tres o más factores: El producto de cualquier cantidad de matrices invertibles es invertible, y la inversa del producto es el producto de las inversas en orden inverso.

Ejemplo 2.6.2: Consideremos las matrices

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 5 \end{bmatrix}, \quad B = \begin{bmatrix} 2 & 3 \\ 1 & 2 \end{bmatrix}.$$

Tenemos que

$$AB = \begin{bmatrix} 4 & 7 \\ 11 & 19 \end{bmatrix}, \quad (AB)^{-1} = \begin{bmatrix} -19 & 7 \\ 11 & -4 \end{bmatrix}$$

y también que

$$A^{-1} = \begin{bmatrix} -5 & 2 \\ 3 & -1 \end{bmatrix}, \quad B^{-1} = \begin{bmatrix} 2 & -3 \\ -1 & 2 \end{bmatrix}, \quad B^{-1}A^{-1} = \begin{bmatrix} -19 & 7 \\ 11 & -4 \end{bmatrix}.$$

Así, $(AB)^{-1} = B^{-1}A^{-1}$ como lo garantiza el teorema anterior.

Si A es una matriz cuadrada invertible, entonces definimos las potencias enteras negativas de A como

$$A^{-n} = (A^{-1})^n = \underbrace{A^{-1}A^{-1}\dots A^{-1}}_{n\text{-veces}}$$

Dado que estas definiciones son análogas a las de los números reales, se cumplen las leyes usuales de exponentes no negativos. Por ejemplo,

$$A^r A^s = A^{r+s} \quad \text{y} \quad (A^r)^s = A^{rs}.$$

Además, tenemos las siguientes propiedades para exponentes negativos:

Teorema 2.6.3: Si A es invertible y n es un entero no negativo, entonces:

- a) A^{-1} es invertible y $(A^{-1})^{-1} = A$.
- b) A^n es invertible y $(A^n)^{-1} = A^{-n} = (A^{-1})^n$.
- c) kA es invertible para todo escalar $k \neq 0$, y $(kA)^{-1} = k^{-1}A^{-1}$.

El siguiente teorema establece una relación entre la inversa de una matriz y la inversa de su transpuesta.

Teorema 2.6.4: Si A es una matriz invertible, entonces A^T también es invertible y

$$(A^T)^{-1} = (A^{-1})^T.$$

Ejemplo 2.6.3: Consideremos las matrices

$$A = \begin{bmatrix} 1 & 2 \\ 1 & 3 \end{bmatrix} \quad \text{y} \quad A^{-1} = \begin{bmatrix} 3 & -2 \\ -1 & 1 \end{bmatrix}.$$

Entonces,

$$A^{-3} = (A^{-1})^3 = \begin{bmatrix} 3 & -2 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 3 & -2 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 3 & -2 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 41 & -30 \\ -15 & 11 \end{bmatrix}.$$

También,

$$A^3 = \begin{bmatrix} 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 11 & 30 \\ 15 & 41 \end{bmatrix}.$$

Por lo tanto, como se esperaba según el teorema 2.6.3b),

$$(A^3)^{-1} = \begin{bmatrix} 41 & -30 \\ -15 & 11 \end{bmatrix} = (A^{-1})^3.$$

Ejemplo 2.6.4: En la aritmética real, donde existe una ley conmutativa para la multiplicación, podemos escribir

$$(a + b)^2 = a^2 + ab + ba + b^2 = a^2 + ab + ab + b^2 = a^2 + 2ab + b^2.$$

Sin embargo, en la aritmética de matrices, donde no existe una ley conmutativa para la multiplicación, lo mejor que podemos hacer es escribir

$$(A + B)^2 = A^2 + AB + BA + B^2.$$

Solo en el caso especial en que A y B conmutan (es decir, $AB = BA$) podemos ir un paso más allá y escribir

$$(A + B)^2 = A^2 + 2AB + B^2.$$

Ejemplo 2.6.5: Una matriz cuadrada con un renglón o una columna de ceros es singular. Para ayudar a entender por qué, consideremos la matriz

$$A = \begin{bmatrix} 1 & 1 & 6 & 0 \\ 9 & 7 & 2 & 0 \\ 2 & 9 & 4 & 0 \\ 6 & 3 & 5 & 0 \end{bmatrix}.$$

Para probar que A es singular, debemos mostrar que no existe una matriz B de tamaño 4×4 tal que $AB = BA = I$. Con este propósito, sean

$$c_1 = \begin{pmatrix} 1 \\ 9 \\ 2 \\ 6 \end{pmatrix}, \quad c_2 = \begin{pmatrix} 1 \\ 7 \\ 9 \\ 3 \end{pmatrix}, \quad c_3 = \begin{pmatrix} 6 \\ 2 \\ 4 \\ 5 \end{pmatrix}, \quad 0 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

las columnas de A . Entonces, para cualquier matriz B de tamaño 4×4 , podemos expresar el producto BA como

$$BA = B[c_1 \quad c_2 \quad c_3 \quad 0] = [Bc_1 \quad Bc_2 \quad Bc_3 \quad B0].$$

La columna de ceros muestra que $BA \neq I$ y, por lo tanto, A es singular.

Nuestro próximo objetivo es mostrar cómo se puede utilizar la multiplicación de matrices para realizar una operación elemental por renglón.

Definición 2.6.2:

Una matriz E se denomina *matriz elemental* si dicha matriz se obtiene de la matriz identidad, aplicando solo una operación elemental.

Ejemplo 2.6.6:

Las siguientes tres matrices elementales son obtenidas de I_3 aplicando las operaciones elementales en sus renglones vistas en la sección anterior.

- Intercambiando el primer renglón y el tercer renglón, se obtiene

$$\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}.$$

- Multiplicando por -3 el segundo renglón, se obtiene

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

- Sumando 3 veces el tercer renglón al primer renglón, se obtiene

$$\begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

El siguiente teorema, cuya demostración se deja como ejercicio al lector, muestra que cuando una matriz A se multiplica por la izquierda por una matriz elemental E , el efecto es realizar una operación elemental por renglones sobre A .

Teorema 2.6.5:

Si la matriz elemental E resulta de realizar una cierta operación elemental sobre I_m , y si A es una matriz $m \times n$, entonces el producto EA es la matriz que resulta al realizar esa misma operación elemental sobre A .

Corolario 2.6.1:

Sean A y B matrices de tamaño $m \times n$, entonces A es equivalente por renglones si y solo si existen E_1, E_2, \dots, E_r matrices elementales obtenidas de realizar una cierta operación elemental sobre I_m , tales que $E_r \cdots E_2 E_1 A = B$.

Ejemplo 2.6.7:

Consideremos la matriz

$$A = \begin{bmatrix} 1 & 0 & 2 & 3 \\ 2 & -1 & 3 & 6 \\ 1 & 4 & 4 & 0 \end{bmatrix}$$

y consideremos la matriz elemental

$$E = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 3 & 0 & 1 \end{bmatrix}$$

que es obtenida de sumar 3 veces el primer renglón de I_3 al tercer renglón. El producto EA es

$$EA = \begin{bmatrix} 1 & 0 & 2 & 3 \\ 2 & -1 & 3 & 6 \\ 4 & 4 & 10 & 9 \end{bmatrix}$$

que es precisamente la matriz que resulta de sumar 3 veces el primer renglón de A al tercer renglón.

Sabemos que si E es una matriz elemental que resulta de realizar una operación elemental sobre la matriz identidad I , entonces existe una segunda operación elemental que, al aplicarse a E , produce nuevamente la identidad I . La tabla 2.2 enumera estas operaciones. Las operaciones del lado derecho de la tabla se denominan *operaciones inversas* de las correspondientes operaciones del lado izquierdo.

Esto implica que toda matriz elemental es invertible, y su inversa también es una matriz elemental. Más aún, al aplicar una operación y su inversa de manera sucesiva, se recupera exactamente el estado original de la matriz.

Tabla 2.2 Operaciones elementales y sus inversas

Operación elemental sobre I que produce E	Operación elemental sobre E que reproduce I
Intercambiar los renglones i y j	Intercambiar los renglones i y j
Multiplicar el i -ésimo renglón por una constante $c \neq 0$	Multiplicar el i -ésimo renglón por una constante c^{-1}
Sumar c veces el renglón i al renglón j	Sumar $-c$ veces el renglón i al renglón j

El siguiente teorema es un resultado clave sobre la invertibilidad de matrices elementales. Servirá como base para muchos resultados posteriores.

Teorema 2.6.6: Toda matriz elemental es invertible, y su inversa también es una matriz elemental.

Demostración: Si E es una matriz elemental, entonces E resulta de realizar una cierta operación elemental sobre I . Sea E_0 la matriz que resulta al aplicar la operación elemental inversa sobre I . Aplicando el teorema 2.6.5 y usando el hecho de que las operaciones elementales inversas cancelan el efecto entre sí, se concluye que

$$E_0 E = I \quad \text{y} \quad E E_0 = I.$$

Por lo tanto, la matriz elemental E_0 es la inversa de E .

Uno de nuestros objetivos al avanzar en este texto es mostrar cómo ideas aparentemente diversas en álgebra lineal están relacionadas. El siguiente teorema, relaciona resultados que hemos obtenido anteriormente.

Teorema 2.6.7: Si A es una matriz $n \times n$, entonces los siguientes enunciados son equivalentes, es decir, todos son verdaderos o todos falsos.

- a) A es no singular.
- b) $A\mathbf{x} = \mathbf{0}$ tiene como única solución la trivial.
- c) La forma escalonada reducida de A es I_n .
- d) A se puede expresar como producto de matrices elementales.

Demostración: Demostraremos la equivalencia estableciendo la cadena de implicaciones: a) \Rightarrow b) \Rightarrow c) \Rightarrow d) \Rightarrow a).

- a) \Rightarrow b) Supongamos que A es no singular y sea \mathbf{x}_0 una solución del sistema homogéneo $A\mathbf{x} = \mathbf{0}$. Multiplicando ambos lados de esta ecuación por la matriz A^{-1} , obtenemos $A^{-1}(A\mathbf{x}_0) = A^{-1}\mathbf{0}$, o bien, $(A^{-1}A)\mathbf{x}_0 = \mathbf{0}$, y por consiguiente, $I\mathbf{x}_0 = \mathbf{0}$, lo que implica que $\mathbf{x}_0 = \mathbf{0}$. Por lo tanto, $A\mathbf{x} = \mathbf{0}$ tiene sólo la solución trivial.
- b) \Rightarrow c) Se deja como ejercicio al lector.
- c) \Rightarrow d) Supongamos que la forma escalonada reducida por filas de A es I_n , de modo que A puede reducirse a I_n mediante una secuencia finita de operaciones fila elementales. Por el teorema 2.6.5, cada una de estas operaciones puede lograrse multiplicando a la izquierda por una matriz elemental apropiada. Así, podemos encontrar matrices elementales E_1, E_2, \dots, E_k tales que

$$E_k \cdots E_2 E_1 A = I_n \quad (2.6.1)$$

Por el teorema 2.6.6, E_1, E_2, \dots, E_k son invertibles. Multiplicando ambos lados de la ecuación (2.6.1) sucesivamente por la izquierda con $E_k^{-1}, \dots, E_2^{-1}, E_1^{-1}$ obtenemos

$$A = E_1^{-1} E_2^{-1} \cdots E_k^{-1} I_n = E_1^{-1} E_2^{-1} \cdots E_k^{-1}.$$

Por el teorema 2.6.6, esta ecuación expresa a A como producto de matrices elementales.

d) \Rightarrow a) Si A es un producto de matrices elementales, entonces, por los teoremas 2.6.3 y 2.6.6, la matriz A es un producto de matrices invertibles, y por lo tanto, es invertible.

Una primera aplicación del teorema anterior, es desarrollar un procedimiento (o algoritmo) que se puede usar para determinar si una matriz dada es no singular, y si lo es, calcular su inversa. Para derivar este algoritmo, supongamos por el momento que A es una matriz no singular de tamaño $n \times n$. En la ecuación (2.6.1), las matrices elementales ejecutan una secuencia de operaciones elementales que reducen A a I_n . Si multiplicamos ambos lados de esta ecuación por la derecha con A^{-1} y simplificamos, obtenemos

$$A^{-1} = E_k \cdots E_2 E_1 I_n.$$

Pero esta ecuación nos dice que *la misma secuencia de operaciones elementales que reduce A a I_n transformará I_n en A^{-1} .*

Método de Gauss-Jordan para hallar la inversa de una matriz:

El método de Gauss-Jordan sirve para saber si una matriz cuadrada es no singular y permite calcular la inversa de dicha matriz. Este método se basa en aplicar operaciones elementales por renglones para transformar una matriz aumentada en una forma específica. En este caso, no se parte de un sistema de ecuaciones, sino de la propia matriz cuya inversa se desea encontrar. El procedimiento es el siguiente:

1. Se escribe la matriz A junto con la matriz identidad del mismo tamaño, formando la matriz aumentada $[A \mid I]$.
2. Se realizan operaciones elementales con el objetivo de transformar la parte izquierda de la matriz aumentada en la identidad. Estas operaciones deben aplicarse también, en paralelo, a la parte derecha.
3. Si al finalizar el proceso, la parte izquierda de la matriz aumentada se convierte exactamente en la matriz identidad, entonces la matriz original es no singular, y la parte derecha de la matriz aumentada será su inversa. En cambio, si al terminar las operaciones la parte izquierda no es la matriz identidad, entonces la matriz A no tiene inversa; es decir, es una matriz singular.

Ejemplo 2.6.8: Verifique que la siguiente matriz es no singular y encuentre su inversa:

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

Solución: Apliquemos el método de Gauss-Jordan para comprobar que la matriz A es no singular y encontrar su inversa. Primero formamos la matriz aumentada con la matriz identidad, la cual es

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

Ahora, aplicamos operaciones elementales a ambos lados de la matriz aumentada. Sumando -2 veces el segundo renglón al primer renglón, obtenemos

$$\left[\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 1 & 1 & 2 & 0 & 1 & 0 \\ 0 & 1 & 2 & 0 & 0 & 1 \end{array} \right] \xrightarrow{r_1 \leftarrow r_1 + (-2)r_2} \left[\begin{array}{ccc|ccc} 1 & 0 & -1 & 1 & 0 & -2 \\ 1 & 1 & 2 & 0 & 1 & 0 \\ 0 & 1 & 2 & 0 & 0 & 1 \end{array} \right]$$

Sumando -1 veces el tercer renglón al segundo renglón, obtenemos

$$\left[\begin{array}{ccc|ccc} 1 & 0 & -1 & 1 & 0 & -2 \\ 1 & 1 & 2 & 0 & 1 & 0 \\ 0 & 1 & 2 & 0 & 0 & 1 \end{array} \right] \xrightarrow{r_2 \leftarrow r_2 + (-1)r_3} \left[\begin{array}{ccc|ccc} 1 & 0 & -1 & 1 & 0 & -2 \\ 1 & 0 & 0 & 0 & 1 & -1 \\ 0 & 1 & 2 & 0 & 0 & 1 \end{array} \right]$$

Sumando -1 veces el segundo renglón al primer renglón, obtenemos

$$\left[\begin{array}{ccc|ccc} 1 & 0 & -1 & 1 & 0 & -2 \\ 1 & 0 & 0 & 0 & 1 & -1 \\ 0 & 1 & 2 & 0 & 0 & 1 \end{array} \right] \xrightarrow{r_1 \leftarrow r_1 + (-1)r_2} \left[\begin{array}{ccc|ccc} 0 & 0 & -1 & 1 & -1 & -1 \\ 1 & 0 & 0 & 0 & 1 & -1 \\ 0 & 1 & 2 & 0 & 0 & 1 \end{array} \right]$$

Multiplicando por -1 el primer renglón, obtenemos

$$\left[\begin{array}{ccc|ccc} 0 & 0 & -1 & 1 & -1 & -1 \\ 1 & 0 & 0 & 0 & 1 & -1 \\ 0 & 1 & 2 & 0 & 0 & 1 \end{array} \right] \xrightarrow{r_1 \leftarrow (-1)r_1} \left[\begin{array}{ccc|ccc} 0 & 0 & 1 & -1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 & -1 \\ 0 & 1 & 2 & 0 & 0 & 1 \end{array} \right]$$

Sumando -2 veces el primer renglón al tercer renglón, obtenemos

$$\left[\begin{array}{ccc|ccc} 0 & 0 & 1 & -1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 & -1 \\ 0 & 1 & 2 & 0 & 0 & 1 \end{array} \right] \xrightarrow{r_3 \leftarrow r_3 + (-2)r_1} \left[\begin{array}{ccc|ccc} 0 & 0 & 1 & -1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 & -1 \\ 0 & 1 & 0 & 2 & -2 & -1 \end{array} \right]$$

Intercambiando el primer renglón y el segundo renglón, obtenemos

$$\left[\begin{array}{ccc|ccc} 0 & 0 & 1 & -1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 & -1 \\ 0 & 1 & 0 & 2 & -2 & -1 \end{array} \right] \xrightarrow{r_1 \leftrightarrow r_2} \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 1 & -1 & 1 & 1 \\ 0 & 1 & 0 & 2 & -2 & -1 \end{array} \right]$$

Intercambiando el segundo renglón y el tercer renglón, obtenemos

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 1 & -1 & 1 & 1 \\ 0 & 1 & 0 & 2 & -2 & -1 \end{array} \right] \xrightarrow{r_2 \leftrightarrow r_3} \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 0 & 1 & -1 \\ 0 & 1 & 0 & 2 & -2 & -1 \\ 0 & 0 & 1 & -1 & 1 & 1 \end{array} \right]$$

Como la parte izquierda de la matriz aumentada es exactamente la matriz identidad, entonces la matriz A es no singular y además,

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

Se deja al lector comprobar, que en efecto, esta matriz es la inversa de A .

Hasta ahora hemos visto dos *procedimientos* para resolver sistemas lineales: la eliminación Gaussiana y la eliminación de Gauss-Jordan. El siguiente teorema se basa en el concepto de la matriz inversa y proporciona una fórmula explícita para la solución de un sistema lineal de n ecuaciones con n incógnitas en el caso en que la matriz de coeficientes es no singular.

Teorema 2.6.8: Si A es una matriz $n \times n$ no singular, entonces para cada vector \mathbf{b} de $n \times 1$, el sistema de ecuaciones $A\mathbf{x} = \mathbf{b}$ tiene exactamente una solución, a saber: $\mathbf{x} = A^{-1}\mathbf{b}$.

Demostración: Dado que $A(A^{-1}\mathbf{b}) = \mathbf{b}$, se deduce fácilmente que $\mathbf{x} = A^{-1}\mathbf{b}$ es una solución de $A\mathbf{x} = \mathbf{b}$. Para mostrar que esta es la única solución, supongamos que \mathbf{x}_0 es una solución arbitraria y demostraremos que debe ser igual a $A^{-1}\mathbf{b}$. Si \mathbf{x}_0 es cualquier solución de $A\mathbf{x} = \mathbf{b}$, entonces $A\mathbf{x}_0 = \mathbf{b}$. Multiplicando ambos lados de esta ecuación por A^{-1} , obtenemos $\mathbf{x}_0 = A^{-1}\mathbf{b}$.

Ejemplo 2.6.9: Considere el sistema de ecuaciones lineales

$$\begin{aligned}x_1 + 2x_2 + 3x_3 &= 5 \\x_1 + x_2 + 2x_3 &= 3 \\x_2 + 2x_3 &= 17\end{aligned}$$

En forma de matricial, este sistema se puede escribir como $A\mathbf{x} = \mathbf{b}$, donde

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 1 & 2 \\ 0 & 1 & 2 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 5 \\ 3 \\ 17 \end{bmatrix}.$$

En el ejemplo 2.6.8, demostramos que A es invertible y

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

Por el teorema anterior, la solución del sistema es

$$\mathbf{x} = A^{-1}\mathbf{b} = \begin{bmatrix} 0 & 1 & -1 \\ 2 & -2 & -1 \\ -1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 5 \\ 3 \\ 17 \end{bmatrix} = \begin{bmatrix} -14 \\ -13 \\ 15 \end{bmatrix}.$$

Frecuentemente, se tiene interés en resolver una secuencia de sistemas

$$A\mathbf{x} = \mathbf{b}_1, \quad A\mathbf{x} = \mathbf{b}_2, \quad A\mathbf{x} = \mathbf{b}_3, \quad \dots, \quad A\mathbf{x} = \mathbf{b}_k$$

cada uno de los cuales tiene la misma matriz cuadrada de coeficientes A . Si A es invertible, entonces las soluciones

$$\mathbf{x}_1 = A^{-1}\mathbf{b}_1, \quad \mathbf{x}_2 = A^{-1}\mathbf{b}_2, \quad \mathbf{x}_3 = A^{-1}\mathbf{b}_3, \quad \dots, \quad \mathbf{x}_k = A^{-1}\mathbf{b}_k$$

pueden obtenerse con una única inversión de matriz y k multiplicaciones matriciales. Una manera eficiente de hacer esto es formar la matriz particionada

$$[A \mid \mathbf{b}_1 \mid \mathbf{b}_2 \mid \dots \mid \mathbf{b}_k] \quad (2.6.2)$$

donde la matriz de coeficientes A está “aumentada” con las k matrices $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_k$, y luego reducir (2.6.2) a su forma escalonada reducida usando la eliminación de Gauss–Jordan. De esta forma, se pueden resolver los k sistemas simultáneamente. Este método tiene la ventaja adicional de que también se aplica cuando A no es invertible.

Ejemplo 2.6.10: Resuelve los sistemas

$$\begin{array}{ll} \text{a)} & \begin{aligned} x_1 + 2x_2 + 3x_3 &= 4 \\ 2x_1 + 5x_2 + 3x_3 &= 5 \\ x_1 + 8x_3 &= 9 \end{aligned} & \text{b)} & \begin{aligned} x_1 + 2x_2 + 3x_3 &= 1 \\ 2x_1 + 5x_2 + 3x_3 &= 6 \\ x_1 + 8x_3 &= -6 \end{aligned} \end{array}$$

Solución: Los dos sistemas tienen la misma matriz de coeficientes. Si aumentamos esta matriz de coeficientes con las columnas de constantes del lado derecho de los sistemas y la llevamos a su forma escalonada reducida, obtenemos

$$\left[\begin{array}{ccc|c|c} 1 & 2 & 3 & 4 & 1 \\ 2 & 5 & 3 & 5 & 6 \\ 1 & 0 & 8 & 9 & -6 \end{array} \right] \longrightarrow \dots \longrightarrow \left[\begin{array}{ccc|c|c} 1 & 0 & 0 & 1 & 2 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & -1 \end{array} \right]$$

De las dos últimas columnas de la matriz escalonada reducida, se deduce que la solución del primer sistema es $x_1 = 1, x_2 = 0, x_3 = 1$ y la solución del segundo sistema es $x_1 = 2, x_2 = 1, x_3 = -1$.

2.7 Ejercicios del Capítulo 2

En los problemas 1 a 8 calcule el producto punto de los dos vectores.

$$1. \begin{pmatrix} 1 \\ 2 \\ -1 \\ 0 \end{pmatrix}; \begin{pmatrix} 3 \\ -7 \\ 4 \\ -2 \end{pmatrix} \quad 2. \begin{pmatrix} 4 \\ -3 \\ 2 \end{pmatrix}; \begin{pmatrix} 1 \\ 6 \\ 6 \end{pmatrix} \quad 3. \begin{pmatrix} 5 \\ 7 \end{pmatrix}; \begin{pmatrix} 3 \\ -2 \end{pmatrix}$$

$$4. \begin{pmatrix} 7 \\ -4 \end{pmatrix}; \begin{pmatrix} 1 \\ -4 \end{pmatrix} \quad 5. \begin{pmatrix} a \\ b \end{pmatrix}; \begin{pmatrix} c \\ d \end{pmatrix} \quad 6. \begin{pmatrix} \sqrt{2} \\ -\sqrt{2} \\ 2 \end{pmatrix}; \begin{pmatrix} \sqrt{18} \\ \sqrt{32} \\ 1 \end{pmatrix}$$

$$7. \begin{pmatrix} \pi \\ 3\pi \\ 3 \end{pmatrix}; \begin{pmatrix} \pi^2 \\ -9\pi \\ \pi^3 \end{pmatrix} \quad 8. \begin{pmatrix} x \\ y \\ z \end{pmatrix}; \begin{pmatrix} y \\ z \\ x \end{pmatrix}$$

9. Sea \mathbf{u} un vector de dimensión n . Pruebe que $\mathbf{u} \bullet \mathbf{u} \geq 0$.

10. Encuentre las condiciones sobre un vector \mathbf{a} tales que $\mathbf{u} \bullet \mathbf{u} = 0$.

En los problemas 11 a 18 considere los vectores

$$\mathbf{u} = \begin{pmatrix} 4 \\ -1 \\ 3 \end{pmatrix}, \quad \mathbf{v} = \begin{pmatrix} 2 \\ 5 \\ -7 \end{pmatrix}, \quad \mathbf{w} = \begin{pmatrix} -6 \\ 8 \\ 0 \end{pmatrix}.$$

Realice las operaciones indicadas con los vectores anteriores.

$$\begin{array}{lll} 11. (2\mathbf{u}) \bullet (3\mathbf{v}) & 12. (\mathbf{u} + \mathbf{v}) \bullet \mathbf{w} & 13. \mathbf{u} \bullet (\mathbf{v} + \mathbf{w}) \\ 14. \mathbf{w} \bullet (\mathbf{u} - \mathbf{v}) & 15. (2\mathbf{v}) \bullet (3\mathbf{w} - 5\mathbf{u}) & 16. (\mathbf{u} - \mathbf{w}) \bullet (3\mathbf{v} - 4\mathbf{u}) \\ 17. \frac{1}{\mathbf{u} \bullet (4\mathbf{w})} \mathbf{v} - 4\mathbf{w} & 18. \frac{\mathbf{u} \bullet \mathbf{w}}{\mathbf{u} \bullet \mathbf{u}} \mathbf{u} \end{array}$$

En los problemas 19 a 34 realice los cálculos indicados.

$$\begin{array}{ll} 19. \begin{bmatrix} 3 & -2 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} -5 & 6 \\ 1 & 3 \end{bmatrix} & 20. \begin{bmatrix} 2 & 3 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 4 & 1 \\ 0 & 6 \end{bmatrix} \\ 21. \begin{bmatrix} -5 & 6 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 3 & -2 \\ 1 & 4 \end{bmatrix} & 22. \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 2 & 3 \end{bmatrix} \\ 23. \begin{bmatrix} 7 & 1 & 4 \\ 2 & -3 & 5 \end{bmatrix} \begin{bmatrix} 1 & 6 \\ 0 & 4 \\ -2 & 3 \end{bmatrix} & 24. \begin{bmatrix} -4 & 5 & 1 \\ 0 & 4 & 2 \end{bmatrix} \begin{bmatrix} 3 & -1 & 1 \\ 5 & 6 & 4 \\ 0 & 1 & 2 \end{bmatrix} \\ 25. \begin{bmatrix} 1 & 4 & -2 \\ 3 & 0 & 4 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 2 & 3 \end{bmatrix} & 26. \begin{bmatrix} 1 & 6 \\ 0 & 4 \\ -2 & 3 \end{bmatrix} \begin{bmatrix} 7 & 1 & 4 \\ 2 & -3 & 5 \end{bmatrix} \\ 27. \begin{bmatrix} 1 & 4 & 6 \\ -2 & 3 & 5 \\ 1 & 0 & 4 \end{bmatrix} \begin{bmatrix} 2 & -3 & 5 \\ 1 & 0 & 6 \\ 2 & 3 & 1 \end{bmatrix} & 28. \begin{bmatrix} 3 & -4 & 6 \\ 1 & 2 & 5 \end{bmatrix} \begin{bmatrix} 1 \\ -2 \end{bmatrix} \\ 29. \begin{bmatrix} 1 & 4 & 0 & 2 \end{bmatrix} \begin{bmatrix} 3 & -6 \\ 2 & 4 \\ 1 & 0 \\ -2 & 3 \end{bmatrix} & 30. \begin{bmatrix} 2 & -3 & 5 \\ 1 & 0 & 6 \\ 2 & 3 & 1 \end{bmatrix} \begin{bmatrix} 1 & 4 & 6 \\ -2 & 3 & 5 \\ 1 & 0 & 4 \end{bmatrix} \end{array}$$

$$31. \begin{bmatrix} 3 & -2 & 1 \\ 4 & 0 & 6 \\ 5 & 1 & 9 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad 32. \begin{bmatrix} 3 & 2 & 1 & -2 \\ -6 & 4 & 0 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 4 \\ 0 \\ 2 \end{bmatrix}$$

$$33. \begin{bmatrix} 5 & -1 & -2 \\ -1 & 3 & 2 \\ 1 & 1 & -5 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad 34. \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 3 & -2 & 1 \\ 4 & 0 & 6 \\ 5 & 1 & 9 \end{bmatrix}$$

35. Considere dos matrices A y B tales que el producto AB existe. Demuestre que el producto AB puede ser calculado columna por columna

$$AB = A [\mathbf{b}_1 \quad \mathbf{b}_2 \quad \cdots \quad \mathbf{b}_n] = [A\mathbf{b}_1 \quad A\mathbf{b}_2 \quad \cdots \quad A\mathbf{b}_n]$$

y renglón por renglón

$$AB = \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_m \end{bmatrix} B = \begin{bmatrix} \mathbf{a}_1 B \\ \mathbf{a}_2 B \\ \vdots \\ \mathbf{a}_m B \end{bmatrix}.$$

36. Sea $A = \begin{bmatrix} 2 & 6 \\ 8 & -6 \end{bmatrix}$, encuentre un vector no nulo $\mathbf{b} = \begin{pmatrix} x \\ y \end{pmatrix}$ tal que $A\mathbf{b} = 6\mathbf{b}$.

37. Encuentre una matriz $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ tal que $A \begin{bmatrix} 2 & 3 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.

38. Sea $A = \begin{bmatrix} 5 & 0 \\ 2 & \alpha \end{bmatrix}$. Determine el valor de α para el cual A es una raíz del polinomio $f(x) = x^2 - 25$.

39. Encuentre B tal que $AB = C$. Si $A = \begin{bmatrix} 5 & 0 & 3 & 4 \\ -1 & 2 & 0 & 1 \end{bmatrix}$ y $C = \begin{bmatrix} 6 & 5 \\ 3 & 5 \end{bmatrix}$.

40. Sea $A = \begin{bmatrix} 2 & 2 \\ 8 & -2 \end{bmatrix}$ y $B = \begin{bmatrix} 2 & -2 \\ 4 & -2 \end{bmatrix}$, pruebe que $A^2 + B^2 = (A + B)^2$.

41. Si $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ y $B = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, encuentre las condiciones para a , b , c y d tal que $AB = BA$.

42. Una matriz A de $n \times n$ tal que $A^2 = I_n$ se llama involutiva. Pruebe que la siguiente matriz es involutiva:

$$A = \begin{bmatrix} 0 & 1 & -1 \\ 4 & -3 & 4 \\ 3 & -3 & 4 \end{bmatrix}.$$

43. Demuestre que $\begin{bmatrix} \alpha & 1 \\ 0 & \alpha \end{bmatrix}^n = \begin{bmatrix} \alpha^n & n\alpha^{n-1} \\ 0 & \alpha^n \end{bmatrix}$ con $n \in \mathbb{Z}^+$.

44. Sean a_{11} , a_{12} , a_{21} y a_{22} números reales dados tales que

$$a_{11}a_{22} - a_{12}a_{21} \neq 0$$

Encuentre los números b_{11} , b_{12} , b_{21} y b_{22} tales que

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

45. Dada la siguiente matriz pruebe que $A^2 = A$:

$$A = \begin{bmatrix} -1 & 3 & 5 \\ 1 & -3 & -5 \\ -1 & 3 & 5 \end{bmatrix}.$$

46. Verifique la ley asociativa para la multiplicación de las matrices, siendo

$$A = \begin{bmatrix} 2 & -1 & 4 \\ 1 & 0 & 6 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 0 & 1 \\ 2 & -1 & 2 \\ 3 & -2 & 0 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 6 \\ -2 & 4 \\ 0 & 5 \end{bmatrix}.$$

47. De la misma forma que en el ejemplo 2.2.3, suponga que un grupo de personas ha contraído una enfermedad contagiosa. Estas personas tienen contacto con un segundo grupo que, a su vez, tiene contacto con un tercer grupo. Si A representa los contactos entre el grupo contagioso y los miembros del grupo 2, y B representa los contactos entre los grupos 2 y 3, donde

$$A = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} \quad \text{y} \quad B = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 \end{bmatrix}.$$

- a) ¿Cuántas personas hay en cada grupo?
b) Encuentre la matriz de contactos indirectos entre los grupos 1 y 3.

48. Conteste la pregunta del problema anterior para

$$A = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 \end{bmatrix} \quad \text{y} \quad B = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}.$$

Se dice que dos vectores \mathbf{u} y \mathbf{v} son ortogonales si $\mathbf{u} \bullet \mathbf{v} = 0$. En los problemas 49 a 54 determine cuáles pares de vectores son ortogonales.

Nota

El tema de ortogonalidad se verá más adelante, en la sección 5.1.

49. $\begin{pmatrix} 2 \\ -3 \end{pmatrix}; \begin{pmatrix} 3 \\ 2 \end{pmatrix}$ 50. $\begin{pmatrix} 7 \\ -5 \\ 4 \\ 1 \end{pmatrix}; \begin{pmatrix} 2 \\ 4 \\ -3 \\ 2 \end{pmatrix}$ 51. $\begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix}; \begin{pmatrix} 0 \\ 1 \\ 0 \\ 1 \end{pmatrix}$

52. $\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}; \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}$ 53. $\begin{pmatrix} 1 \\ 4 \\ -7 \end{pmatrix}; \begin{pmatrix} 2 \\ 3 \\ 2 \end{pmatrix}$ 54. $\begin{pmatrix} a \\ 0 \\ b \\ 0 \\ c \end{pmatrix}; \begin{pmatrix} 0 \\ d \\ 0 \\ e \\ 0 \end{pmatrix}$

55. Determine el número α tal que $\begin{pmatrix} 1 \\ -2 \\ 3 \\ 5 \end{pmatrix}$ es ortogonal a $\begin{pmatrix} -4 \\ \alpha \\ 6 \\ -1 \end{pmatrix}$.

56. Determine todos los números α y β tales que los vectores $\begin{pmatrix} 1 \\ -\alpha \\ 2 \\ 3 \end{pmatrix}$ y $\begin{pmatrix} 4 \\ 5 \\ -2\beta \\ 3 \end{pmatrix}$ son ortogonales.

De los problemas 57 al 76 encuentre el rango y la nulidad de la matriz dada.

$$57. \begin{bmatrix} 4 & 3 \\ 2 & -2 \end{bmatrix}$$

$$58. \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$$

$$59. \begin{bmatrix} 1 & -1 & 2 \\ 3 & 1 & 0 \end{bmatrix}$$

$$60. \begin{bmatrix} -3 & 1 \\ 3 & -2 \\ -1 & 1 \end{bmatrix}$$

$$61. \begin{bmatrix} -1 & 3 & 2 \\ 2 & -6 & -4 \end{bmatrix}$$

$$62. \begin{bmatrix} 1 & -1 & 2 \\ 3 & 1 & 4 \\ -1 & 0 & 4 \end{bmatrix}$$

$$63. \begin{bmatrix} 0 & 3 & -1 \\ 2 & 1 & -1 \end{bmatrix}$$

$$64. \begin{bmatrix} 1 & -1 & 2 \\ 3 & 1 & 4 \\ 5 & -1 & 8 \end{bmatrix}$$

$$65. \begin{bmatrix} -1 & 2 & 1 \\ 2 & -4 & -2 \\ -3 & 6 & 3 \end{bmatrix}$$

$$66. \begin{bmatrix} 2 & 2 & -3 \\ 0 & 0 & -1 \\ -3 & -2 & -3 \end{bmatrix}$$

$$67. \begin{bmatrix} 1 & -1 & 2 & 3 \\ 0 & 1 & 4 & 3 \\ 1 & 0 & 6 & 5 \end{bmatrix}$$

$$68. \begin{bmatrix} 0 & 4 & 2 & 0 \\ 0 & 0 & 1 & 6 \\ 1 & 0 & -1 & 2 \end{bmatrix}$$

$$69. \begin{bmatrix} 3 & -2 & 3 \\ -3 & 1 & -1 \\ 0 & 0 & -2 \\ -1 & -3 & 2 \end{bmatrix}$$

$$70. \begin{bmatrix} 1 & -1 & 2 & 3 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$71. \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -2 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

$$72. \begin{bmatrix} -3 & 0 & -1 & -1 \\ -1 & 4 & 4 & -1 \\ 0 & 2 & 3 & -2 \end{bmatrix}$$

$$73. \begin{bmatrix} 1 & -1 & 2 & 3 \\ -2 & 2 & -4 & -6 \\ 2 & -2 & 4 & 6 \\ 3 & -3 & 6 & 9 \end{bmatrix}$$

$$74. \begin{bmatrix} -1 & -1 & 0 & 0 \\ 0 & 0 & 2 & 3 \\ 4 & 0 & -2 & 1 \\ 3 & -1 & 0 & 4 \end{bmatrix}$$

$$75. \begin{bmatrix} 0 & 4 & -3 & -1 \\ 4 & 4 & -2 & -2 \\ -1 & -1 & 1 & -1 \\ 0 & 2 & 1 & 3 \end{bmatrix}$$

$$76. \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 3 \\ 1 & 2 & 3 & 4 \end{bmatrix}$$

De los problemas 77 al 82 encuentre una base para la imagen y el espacio nulo de la matriz dada.

77. La matriz del problema 59.

78. La matriz del problema 61.

79. La matriz del problema 62.

80. La matriz del problema 66.

81. La matriz del problema 69.

82. La matriz del problema 72.

83. Demuestre que el rango de una matriz diagonal es igual al número de componentes diferentes de cero en la diagonal.

84. Sea A una matriz triangular inferior de $n \times n$ con ceros en la diagonal. Demuestre que $\rho(A) < n$.

- 85.** Demuestre que si A es una matriz de $m \times n$ y $m < n$, entonces
- $\rho(A) \leq m$
 - $\nu(A) \geq n - m$.
- 86.** Demuestre que si una matriz A no es cuadrada, entonces los vectores renglón o los vectores columna de A son linealmente dependientes.
- 87.** Sean A y B matrices de $m \times n$ y $n \times p$, respectivamente. Demuestre que $\rho(AB) \leq \min\{\rho(A), \rho(B)\}$.
- 88.** Sea A una matriz de $m \times n$ y sean B y C matrices invertibles de $m \times m$ y $n \times n$, respectivamente. Pruebe que $\rho(A) = \rho(BA) = \rho(AC)$. Es decir, si se multiplica una matriz por una matriz invertible, el rango no cambia.
- 89.** Sean A y B matrices de $m \times n$. Demuestre que si $\rho(A) = \rho(B)$, entonces existen matrices invertibles C y D tales que $B = CAD$.
- 90.** Con el propósito de este ejercicio, definamos una “matriz tablero de ajedrez” como una matriz cuadrada $A = [a_{ij}]$ tal que

$$a_{ij} = \begin{cases} 1 & \text{si } i + j \text{ es par} \\ 0 & \text{si } i + j \text{ es impar} \end{cases}$$

Determina el rango y la nulidad de las siguientes matrices:

- La matriz tablero de ajedrez de 3×3 .
 - La matriz tablero de ajedrez de 4×4 .
 - La matriz tablero de ajedrez de $n \times n$.
- 91.** Suponga que cualesquiera k renglones de A son linealmente independientes mientras que cualesquiera $k + 1$ renglones de A son linealmente dependientes. Demuestre que $\rho(A) = k$.
- 92.** Si $B = CAD$, donde C y D son invertibles, demuestre que $\rho(A) = \rho(B)$.
- 93.** Sea A una matriz de $m \times n$. Suponga que para todo $y \in \mathbb{R}^m$ existe $x \in \mathbb{R}^n$ tal que $Ax = y$. Demuestre que $\rho(A) = m$.
- 94.** Si A es una matriz de $n \times n$, demuestre que $\rho(A) < n$ si y sólo si existe un vector $x \in \mathbb{R}^n$ tal que $x \neq 0$ y $Ax = 0$.
- 95.** Pruebe que el rango de una matriz es igual al número de pivotes en su forma escalonada por renglones.

Del ejercicio 96 al 109, aplique el método de eliminación de Gauss-Jordan para resolver los siguientes sistemas. Si existe más de una solución, dé tanto la solución general como una solución particular.

- | | |
|--|--|
| <p>96.</p> $\begin{aligned} 2x + 4y &= 6 \\ 3x + 6y &= 5 \end{aligned}$ | <p>97.</p> $\begin{aligned} 2x + 4y &= 6 \\ 3x + 6y &= 9 \end{aligned}$ |
| <p>98.</p> $\begin{aligned} x + y &= -6 \\ 2x - 2y &= 4 \end{aligned}$ | <p>99.</p> $\begin{aligned} 2x + 3y + z &= 3 \\ x + 2y + z &= 1 \\ -x + 4y &= -2 \end{aligned}$ |
| <p>100.</p> $\begin{aligned} 2x + 3y + 4z &= 1 \\ x - 3z &= 4 \\ 4x + y - z &= 6 \end{aligned}$ | <p>101.</p> $\begin{aligned} y + 3z &= 1 \\ -x + 2y - 7z &= 1 \\ 2x + 3y - z &= -2 \end{aligned}$ |

$$\begin{aligned} 102. \quad & 7x - 12y + 2z = 0 \\ & 3x + 8y - 11z = 0 \\ & \frac{1}{2}x + \frac{1}{4}y + 6z = 0 \end{aligned}$$

$$\begin{aligned} 103. \quad & 2x + 6z = 9 \\ & 5x + 8y + 4z = 3 \\ & -3x + 7y = 5 \end{aligned}$$

$$\begin{aligned} 104. \quad & x + 5y + 11z = -5 \\ & 2x + 3y + 8z = 4 \\ & -x + 2y + 3z = -9 \end{aligned}$$

$$\begin{aligned} 105. \quad & x + 5y + 11z = -5 \\ & 2x + 3y + 8z = 4 \\ & -x + 2y + 3z = 3 \end{aligned}$$

$$\begin{aligned} 106. \quad & x + 3y + 5z + 10w = 2 \\ & -x - 2y - 4w = 4 \\ & 2x + 4y + 8z + 16w = 0 \end{aligned}$$

$$\begin{aligned} 107. \quad & 2x - y - z = 4 \\ & 3x - 2y + 4z = 11 \\ & 6x + 8y - 4z = 22 \end{aligned}$$

$$\begin{aligned} 108. \quad & 2x + 2y + 4z + 4w = 3 \\ & 8x + 4z + 3w = 7 \\ & x + z + w = 1 \\ & 5x + 4y + 5z + 6w = 6 \end{aligned}$$

$$\begin{aligned} 109. \quad & 9x + 12y + 17w = 4 \\ & 6x + 3y + 3z + 8w = 4 \\ & 3x + 6y + 6z + 13w = 2 \\ & x + 6y + 6z + 13w = 0 \end{aligned}$$

En los problemas 110 a 131 determine si la matriz dada es invertible. De ser así, calcule la inversa.

$$110. \begin{bmatrix} 3 & -2 \\ 5 & -4 \end{bmatrix}$$

$$111. \begin{bmatrix} 4 & -7 \\ -8 & 14 \end{bmatrix}$$

$$112. \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$113. \begin{bmatrix} 1 & 1 \\ 3 & 3 \end{bmatrix}$$

$$114. \begin{bmatrix} 8 & -2 \\ 16 & -4 \end{bmatrix}$$

$$115. \begin{bmatrix} a & b \\ -a & -b \end{bmatrix}$$

$$116. \begin{bmatrix} 1 & 1 & 1 \\ 0 & 2 & 3 \\ 5 & 5 & 1 \end{bmatrix}$$

$$117. \begin{bmatrix} 0 & 1 & 3 \\ 3 & 4 & -2 \\ -1 & 5 & 8 \end{bmatrix}$$

$$118. \begin{bmatrix} 3 & 2 & 1 \\ 0 & 2 & 2 \\ 0 & 0 & -1 \end{bmatrix}$$

$$119. \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

$$120. \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

$$121. \begin{bmatrix} 2 & 24 & 48 \\ 0 & -3 & 12 \\ 0 & 0 & 2 \end{bmatrix}$$

$$122. \begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 8 \\ 1 & -1 & 1 \end{bmatrix}$$

$$123. \begin{bmatrix} 1 & -2 & -3 \\ 0 & 3 & 4 \\ 0 & 0 & 5 \end{bmatrix}$$

$$124. \begin{bmatrix} 1 & 2 & 3 \\ 1 & 1 & 2 \\ 0 & 1 & 2 \end{bmatrix}$$

$$125. \begin{bmatrix} \alpha & \beta & \gamma \\ \delta & \epsilon & \zeta \\ 3\alpha - 2\delta & 3\beta - 2\epsilon & 3\gamma - 2\zeta \end{bmatrix}$$

$$126. \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & -1 & 2 \\ 1 & -1 & 2 & 1 \\ 1 & 3 & 3 & 2 \end{bmatrix}$$

$$127. \begin{bmatrix} 1 & 0 & 2 & 3 \\ -1 & 1 & 0 & 4 \\ 2 & 1 & -1 & 3 \\ -1 & 0 & 5 & 7 \end{bmatrix}$$

$$128. \begin{bmatrix} 1 & 2 & 0 & 0 \\ 2 & 3 & -3 & 0 \\ 0 & -3 & -2 & 4 \\ 0 & 0 & 4 & 4 \end{bmatrix}$$

$$129. \begin{bmatrix} 1 & 3 & 0 & -2 \\ 3 & 12 & -2 & -6 \\ 2 & 10 & 2 & 5 \\ 1 & 6 & 1 & 3 \end{bmatrix}$$

$$130. \begin{bmatrix} 3 & 4 & 0 & 0 \\ 2 & 3 & 0 & 0 \\ 0 & 0 & 2 & 3 \\ 0 & 0 & 3 & 4 \end{bmatrix}$$

$$131. \begin{bmatrix} 0 & 0 & 3 & 4 \\ 0 & 0 & 2 & 3 \\ 2 & 3 & 0 & 0 \\ 3 & 4 & 0 & 0 \end{bmatrix}$$

132. Sean A y B sean matrices cuadradas del mismo tamaño. Muestre que si AB es invertible, entonces A y B también deben ser invertibles.

133. Si A_1, A_2, \dots, A_m son matrices invertibles de $n \times n$, muestre que $A_1 A_2 \cdots A_m$ es invertible y calcule su inversa.

134. Muestre que la matriz $\begin{bmatrix} 3 & 4 \\ -2 & -3 \end{bmatrix}$ es su propia inversa.

De los problemas 135 a 151 determine cuáles matrices son matrices elementales.

135. $\begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$ **136.** $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ **137.** $\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$ **138.** $\begin{bmatrix} 1 & 3 \\ 0 & 1 \end{bmatrix}$

139. $\begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}$ **140.** $\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$ **141.** $\begin{bmatrix} 0 & 3 \\ 2 & 0 \end{bmatrix}$ **142.** $\begin{bmatrix} -4 & 0 \\ 4 & 3 \end{bmatrix}$

143. $\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ **144.** $\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ **145.** $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -2 & 1 \end{bmatrix}$ **146.** $\begin{bmatrix} 1 & 3 & 4 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix}$

147. $\begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 0 \\ 0 & -2 & 1 \end{bmatrix}$ **148.** $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}$ **149.** $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$

150. $\begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ **151.** $\begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$

Del problema 152 al 162, encuentre la matriz elemental E tal que $EA = B$.

152. $A = \begin{bmatrix} 2 & 3 \\ -1 & 4 \end{bmatrix}, B = \begin{bmatrix} 2 & 3 \\ 2 & -8 \end{bmatrix}$ **153.** $A = \begin{bmatrix} 2 & 3 \\ -1 & 4 \end{bmatrix}, B = \begin{bmatrix} 2 & 3 \\ -5 & -2 \end{bmatrix}$

154. $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, B = \begin{bmatrix} 1 & 2 \\ 4 & 6 \end{bmatrix}$ **155.** $A = \begin{bmatrix} 1 & -3 \\ 2 & -8 \end{bmatrix}, B = \begin{bmatrix} 2 & -8 \\ 1 & -3 \end{bmatrix}$

156. $A = \begin{bmatrix} 2 & 3 \\ -1 & 4 \end{bmatrix}, B = \begin{bmatrix} -1 & 4 \\ 2 & 3 \end{bmatrix}$ **157.** $A = \begin{bmatrix} 1 & 2 \\ 5 & 6 \end{bmatrix}, B = \begin{bmatrix} 5 & 6 \\ 1 & 2 \end{bmatrix}$

158. $A = \begin{bmatrix} 1 & 3 & 4 \\ -1 & 3 & 2 \end{bmatrix}, B = \begin{bmatrix} 3 & -3 & 0 \\ -1 & 3 & 2 \end{bmatrix}$

159. $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, B = \begin{bmatrix} 1 & 2 \\ 0 & -2 \end{bmatrix}$ **160.** $A = \begin{bmatrix} 0 & 5 \\ 1 & 2 \end{bmatrix}, B = \begin{bmatrix} -3 & -1 \\ 1 & 2 \end{bmatrix}$

161. $A = \begin{bmatrix} -2 & 5 & 4 \\ 0 & 5 & 7 \end{bmatrix}, B = \begin{bmatrix} -2 & 5 & 4 \\ 6 & -10 & -5 \end{bmatrix}$

162. $A = \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix}, B = \begin{bmatrix} \alpha - 3\gamma & \beta - 3\delta \\ \gamma & \delta \end{bmatrix}$

De los problemas 163 a 179 encuentre la inversa de la matriz elemental dada.

163. $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ **164.** $\begin{bmatrix} 1 & 3 \\ 0 & 1 \end{bmatrix}$ **165.** $\begin{bmatrix} 1 & 0 \\ -3 & 1 \end{bmatrix}$

166. $\begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}$ **167.** $\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ **168.** $\begin{bmatrix} 1 & 0 & -5 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

$$169. \begin{bmatrix} 1 & -2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$170. \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{bmatrix}$$

$$171. \begin{bmatrix} 1 & 0 & 0 \\ -7 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$172. \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -2 & 0 & 1 \end{bmatrix}$$

$$173. \begin{bmatrix} -\pi & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$174. \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$175. \begin{bmatrix} 1 & 0 & 0 & 5 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$176. \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & -3 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$177. \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$178. \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

$$179. \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -6 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

De los problemas 180 a 189 demuestre que cada matriz es invertible y escríbala como un producto de matrices elementales.

$$180. \begin{bmatrix} 2 & 1 \\ 3 & 2 \end{bmatrix}$$

$$181. \begin{bmatrix} -2 & 8 \\ 5 & 12 \end{bmatrix}$$

$$182. \begin{bmatrix} 1 & 1 & 1 \\ 0 & 2 & 3 \\ 5 & 5 & 1 \end{bmatrix}$$

$$183. \begin{bmatrix} 3 & 2 & 1 \\ 0 & 2 & 2 \\ 0 & 0 & -1 \end{bmatrix}$$

$$184. \begin{bmatrix} 0 & -1 & 0 \\ 0 & 1 & -1 \\ 1 & 0 & 1 \end{bmatrix}$$

$$185. \begin{bmatrix} 1 & 3 & 0 \\ 1 & 1 & 0 \\ 2 & -3 & 1 \end{bmatrix}$$

$$186. \begin{bmatrix} 2 & 0 & 0 \\ 4 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$$

$$187. \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & -4 & 0 \\ 0 & 0 & 0 & 5 \end{bmatrix}$$

$$188. \begin{bmatrix} 2 & 1 & 0 & 0 \\ 0 & 2 & 1 & 0 \\ 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 2 \end{bmatrix}$$

$$189. \begin{bmatrix} 1 & 0 & 0 & -2 \\ 5 & 1 & 0 & -10 \\ -4 & 2 & 1 & 8 \\ 3 & 6 & 0 & -5 \end{bmatrix}$$

190. Sea $A = \begin{bmatrix} a & b \\ 0 & c \end{bmatrix}$ donde $ac \neq 0$. Escriba A como un producto de tres matrices elementales y concluya que A es invertible.

191. Sea A una matriz triangular superior de $n \times n$. Pruebe que si toda componente en la diagonal de A es diferente de cero, entonces A es invertible.

192. Demuestre que si A es una matriz triangular superior de $n \times n$ con componentes diferentes de cero en la diagonal, entonces A^{-1} es triangular superior.

193. Sean

$$A = \begin{bmatrix} 1 & -2 & 2 \\ 4 & 5 & 1 \\ 0 & 3 & -1 \end{bmatrix}, \quad \mathbb{b}_1 = \begin{bmatrix} 0 \\ 1 \\ 7 \end{bmatrix}, \quad \mathbb{b}_2 = \begin{bmatrix} 11 \\ 5 \\ 3 \end{bmatrix}, \quad \mathbb{b}_3 = \begin{bmatrix} 1 \\ -4 \\ 2 \end{bmatrix}.$$

Resuelva los sistemas lineales $A\mathbf{x} = \mathbb{b}_1$, $A\mathbf{x} = \mathbb{b}_2$, $A\mathbf{x} = \mathbb{b}_3$ usando el método del ejemplo 2.6.10.

3

Capítulo

DETERMINANTES

i Nota

Se debe mencionar que durante un tiempo los determinantes jugaron un papel fundamental en el estudio del álgebra lineal; ahora, sin embargo, tienen una importancia mucho menor. Veremos, de hecho, que virtualmente nuestra única utilización de los determinantes será en el cálculo de los “eigenvalores”.

El estudio de los determinantes adquiere relevancia primordial al abordar la noción de invertibilidad de las matrices. Su presencia y magnitud se convierten en criterios decisivos para determinar si una matriz posee una inversa, un concepto esencial en la resolución de sistemas lineales y en el análisis de transformaciones lineales.

Además, la conexión intrínseca entre los determinantes y la geometría añade otra dimensión de fascinación. No se limitan meramente a manipulaciones algebraicas; los determinantes encuentran aplicación directa en la determinación de áreas y volúmenes en espacios euclidianos, estableciendo un puente conceptual entre el álgebra y la geometría. La relación entre los determinantes y la geometría amplía nuestro entendimiento matemático y resalta la belleza de su interrelación. Explorar cómo se entrelazan con la estructura geométrica de los espacios vectoriales nos permite apreciar la profundidad y la elegancia de las matemáticas en su conjunto.

3.1 Determinantes por expansión de cofactores

Definición 3.1.1:

Consideremos una matriz de tamaño 2×2

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

Definimos el determinante de dicha matriz, escribiendo

$$\det(A) = ad - bc \quad \text{o} \quad \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc. \quad (3.1.1)$$

i Nota

Es importante considerar que $\det(A)$ es un número, mientras que A es una matriz.

A continuación, demostraremos un teorema que relaciona el determinante y la inversa de una matriz.

Teorema 3.1.1: Sea A una matriz de tamaño 2×2 .

- i) A es invertible si y solo si $\det(A) \neq 0$.
- ii) Si $\det(A) \neq 0$, entonces

$$A^{-1} = \frac{1}{\det(A)} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}. \quad (3.1.2)$$

Demostración:

- i) Supongamos que A es invertible. Entonces existe A^{-1} tal que $AA^{-1} = I$. Si $\det(A) = 0$, entonces $ad - bc = 0$. Consideremos el sistema homogéneo asociado $A\mathbf{x} = \mathbf{0}$,

$$\begin{aligned} ax_1 + bx_2 &= 0 \\ cx_1 + dx_2 &= 0 \end{aligned}$$

Como $ad - bc = 0$, las filas de A son linealmente dependientes, y el sistema tiene soluciones no triviales. Sin embargo, si A es invertible, el único vector \mathbf{x} que satisface $A\mathbf{x} = \mathbf{0}$ es $\mathbf{x} = \mathbf{0}$. Esta contradicción implica que $\det(A) \neq 0$. Ahora bien, supongamos que $\det(A) \neq 0$ y sea

$$B = \frac{1}{\det(A)} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}. \quad (3.1.3)$$

Entonces,

$$\begin{aligned} BA &= \frac{1}{\det(A)} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} \\ &= \frac{1}{ad - bc} \begin{bmatrix} da - bc & 0 \\ 0 & -cb + ad \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \end{aligned}$$

De manera similar, $AB = I$. De esta manera, $AB = BA = I$, por lo que A es invertible y $B = A^{-1}$.

- ii) Del inciso anterior, si $\det(A) \neq 0$, entonces A es invertible, y la matriz B definida en (3.1.3), satisface $AB = BA = I$. Por tanto,

$$A^{-1} = \frac{1}{\det(A)} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

Uno de nuestros principales objetivos en este capítulo es obtener un análogo de la fórmula (3.1.2) que sea aplicable a matrices cuadradas de todos los órdenes. Para este propósito, nos resultará conveniente usar entradas con subíndices al escribir matrices o determinantes. Así, si denotamos una matriz 2×2 como

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

entonces las dos ecuaciones en (3.1.1) toman la forma

$$\det(A) = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}. \quad (3.1.4)$$

En situaciones donde no sea conveniente asignar un nombre a la matriz, podemos expresar esta fórmula como

$$\det \left(\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \right) = a_{11}a_{22} - a_{12}a_{21}. \quad (3.1.5)$$

Existen varios métodos para definir determinantes de matrices cuadradas de orden superior. En este texto, el determinante de una matriz de $n \times n$ se definirá de manera inductiva. En otras palabras, se usará lo que se sabe sobre un determinante de 2×2 para definir un determinante de 3×3 , que a su vez se usará para definir un determinante de 4×4 , y así sucesivamente. Para iniciar el proceso, definamos el determinante de una matriz 1×1 como

$$\det([a_{11}]) = a_{11}$$

de donde se deduce que la fórmula (3.1.5) puede expresarse como

$$\det \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \det([a_{11}]) \det([a_{22}]) - \det([a_{12}]) \det([a_{21}]).$$

Ahora que hemos establecido un punto de partida, podemos definir determinantes de matrices 3×3 en términos de determinantes de matrices 2×2 , luego determinantes de matrices 4×4 en términos de determinantes de matrices 3×3 , y así sucesivamente. La siguiente terminología y notación ayudarán a hacer este proceso inductivo más eficiente.

Definición 3.1.2:

Si A es una matriz cuadrada, entonces el *menor de la entrada* a_{ij} se denota por M_{ij} y se define como el determinante de la submatriz que permanece después de que la i -ésima fila y la j -ésima columna se eliminen de A . El número $(-1)^{i+j} M_{ij}$ se denota por C_{ij} y se llama *cofactor de la entrada* a_{ij} .

Ejemplo 3.1.1:

Consideremos la matriz

$$A = \begin{bmatrix} 0 & 2 & -1 \\ 3 & 5 & 6 \\ 4 & 4 & 8 \end{bmatrix}.$$

El menor de la entrada a_{11} es

$$M_{11} = \begin{vmatrix} 3 & 1 & -4 \\ 2 & 5 & 6 \\ 1 & 4 & 8 \end{vmatrix} = \begin{vmatrix} 5 & 6 \\ 4 & 8 \end{vmatrix} = 16.$$

El cofactor de a_{11} es

$$C_{11} = (-1)^{1+1} M_{11} = M_{11} = 16.$$

Del mismo modo, el menor de la entrada a_{32} es

$$M_{32} = \begin{vmatrix} 3 & 1 & -4 \\ 2 & 5 & 6 \\ 1 & 4 & 8 \end{vmatrix} = \begin{vmatrix} 3 & -4 \\ 2 & 6 \end{vmatrix} = 26.$$

El cofactor de a_{32} es

$$C_{32} = (-1)^{3+2} M_{32} = -M_{32} = -26.$$

Nota

Hemos seguido la convención estándar de utilizar letras mayúsculas para denotar menores y cofactores aunque sean números, no matrices.

Notemos que un menor M_{ij} y su cofactor correspondiente C_{ij} son iguales o negativos el uno del otro y que la relación de signo $(-1)^{i+j}$ es $+1$ o -1 de acuerdo con el patrón en la “matriz de tablero de ajedrez”

$$\begin{bmatrix} + & - & + & - & \cdots \\ - & + & - & + & \cdots \\ + & - & + & - & \cdots \\ - & + & - & + & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

Por ejemplo,

$$C_{11} = M_{11}, \quad C_{21} = -M_{21}, \quad C_{22} = M_{22}$$

y así sucesivamente. Así, nunca es realmente necesario calcular $(-1)^{i+j}$ para obtener C_{ij} . Se puede simplemente calcular el menor M_{ij} y luego ajustar el signo de acuerdo con el patrón del tablero de ajedrez.

Ejemplo 3.1.2: El patrón del tablero de ajedrez para una matriz A de tamaño 2×2 es

$$\begin{bmatrix} + & - \\ - & + \end{bmatrix}$$

De esta manera: $C_{11} = M_{11} = a_{22}$, $C_{12} = -M_{12} = -a_{21}$, $C_{21} = -M_{21} = -a_{12}$, $C_{22} = M_{22} = a_{11}$. Dejamos al lector usar la fórmula (3.1.4) para verificar que $\det(A)$ puede expresarse en términos de cofactores de las siguientes cuatro maneras:

$$\begin{aligned} \det(A) &= \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \\ &= a_{11}C_{11} + a_{12}C_{12} \\ &= a_{21}C_{21} + a_{22}C_{22} \\ &= a_{11}C_{11} + a_{21}C_{21} \\ &= a_{12}C_{12} + a_{22}C_{22} \end{aligned} \quad (3.1.6)$$

Cada una de las cuatro ecuaciones se llama una *expansión por cofactores* de $\det(A)$. En cada expansión por cofactores, las entradas y los cofactores provienen todos de la misma fila o de la misma columna de A . Por ejemplo, en la primera ecuación las entradas y cofactores provienen todos de la primera fila de A , en la segunda provienen todos de la segunda fila de A , en la tercera provienen todos de la primera columna de A , y en la cuarta provienen todos de la segunda columna de A .

La fórmula (3.1.6) es un caso especial del siguiente resultado general, que presentaremos sin demostración.

Teorema 3.1.2: Si A es una matriz $n \times n$, entonces independientemente de qué fila o columna de A se elija, el número obtenido al multiplicar las entradas de esa fila o columna por los cofactores correspondientes y sumar los productos resultantes es el mismo.

Este resultado nos permite establecer la siguiente definición.

Definición 3.1.3: Si A es una matriz $n \times n$, entonces el número obtenido al multiplicar las entradas de cualquier fila o columna de A por los cofactores correspondientes y sumar los productos resultantes se llama *determinante de A* , y las sumas en sí mismas se denominan *expansiones por cofactores de A* . Es decir, el determinante se puede obtener mediante expansión por cofactores a lo largo de la j -ésima columna

$$\det(A) = a_{1j}C_{1j} + a_{2j}C_{2j} + \cdots + a_{nj}C_{nj} \quad (3.1.7)$$

o bien, mediante expansión por cofactores a lo largo de la i -ésima fila

$$\det(A) = a_{i1}C_{i1} + a_{i2}C_{i2} + \cdots + a_{in}C_{in}. \quad (3.1.8)$$

Ejemplo 3.1.3: Encuentra el determinante de la matriz

$$A = \begin{bmatrix} 3 & 1 & 0 \\ -2 & -4 & 3 \\ 5 & 4 & -2 \end{bmatrix}$$

por expansión de cofactores a lo largo de la primera fila.

Solución:

$$\begin{aligned} \det(A) &= 3 \begin{vmatrix} -4 & 3 \\ 4 & -2 \end{vmatrix} - 1 \begin{vmatrix} -2 & 3 \\ 5 & -2 \end{vmatrix} + 0 \begin{vmatrix} -2 & -4 \\ 5 & 4 \end{vmatrix} \\ &= 3(-4) - (1)(-11) + 0 = -1 \end{aligned}$$

Ejemplo 3.1.4:

Sea A la matriz del ejemplo 3.1.3, evalúe $\det(A)$ mediante la expansión de cofactores a lo largo de la primera columna de A .

Solución:

$$\begin{aligned}\det(A) &= 3 \begin{vmatrix} -4 & 3 \\ 4 & -2 \end{vmatrix} - (-2) \begin{vmatrix} 1 & 0 \\ 4 & -2 \end{vmatrix} + 5 \begin{vmatrix} 1 & 0 \\ -4 & 3 \end{vmatrix} \\ &= 3(-4) - (-2)(-2) + 5(3) = -1\end{aligned}$$

Esto concuerda con el resultado obtenido en el ejemplo 3.1.3.

Tenga en cuenta que en el ejemplo 3.1.4 tuvimos que calcular tres cofactores, mientras que en el ejemplo 3.1.3 solo se necesitaban dos porque el tercero se multiplicó por cero. Como regla general, la mejor estrategia para la expansión de cofactores es expandirse a lo largo de una fila o columna con más ceros.

Ejemplo 3.1.5:

Si A es una matriz de 4×4 dada por

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

Entonces para encontrar $\det(A)$ será más fácil usar la expansión de cofactores a lo largo de la segunda columna, ya que tiene la mayor cantidad de ceros:

$$\det(A) = 1 \cdot \begin{vmatrix} 1 & 0 & -1 \\ 1 & -2 & 1 \\ 2 & 0 & 1 \end{vmatrix} = \begin{vmatrix} 1 & 0 & -1 \\ 1 & -2 & 1 \\ 2 & 0 & 1 \end{vmatrix}.$$

Para el determinante 3×3 , será más fácil utilizar la expansión del cofactor a lo largo de su segunda columna, ya que tiene la mayor cantidad de ceros:

$$\det(A) = -2 \cdot \begin{vmatrix} 1 & -1 \\ 2 & 1 \end{vmatrix} = -2(1 + 2) = -6.$$

Ejemplo 3.1.6:

El siguiente cálculo muestra que el determinante de una matriz triangular inferior de 4×4 es el producto de sus elementos en la diagonal principal. Cada parte del cálculo utiliza una expansión de cofactores a lo largo de la primera fila.

$$\begin{aligned}\begin{vmatrix} a_{11} & 0 & 0 & 0 \\ a_{21} & a_{22} & 0 & 0 \\ a_{31} & a_{32} & a_{33} & 0 \\ a_{41} & a_{42} & a_{43} & a_{44} \end{vmatrix} &= a_{11} \begin{vmatrix} a_{22} & 0 & 0 \\ a_{32} & a_{33} & 0 \\ a_{42} & a_{43} & a_{44} \end{vmatrix} \\ &= a_{11} a_{22} \begin{vmatrix} a_{33} & 0 \\ a_{43} & a_{44} \end{vmatrix} \\ &= a_{11} a_{22} a_{33} |a_{44}| \\ &= a_{11} a_{22} a_{33} a_{44}\end{aligned}$$

El método visto en el ejemplo anterior se puede generalizar fácilmente para demostrar el siguiente teorema.

Teorema 3.1.3:

Si A es una matriz $n \times n$ triangular superior o triangular inferior, entonces

$$\det(A) = a_{11} a_{22} \cdots a_{nn}.$$

Es decir, el determinante de una matriz triangular superior o inferior es igual al producto de sus componentes en la diagonal.

Demostración: La parte triangular inferior del teorema se deja como ejercicio al lector (se puede deducir del ejemplo anterior). Se demostrará la parte triangular superior por inducción matemática sobre n , empezando por $n = 2$. Si A es una matriz triangular superior de 2×2 , entonces

$$A = \begin{bmatrix} a_{11} & a_{12} \\ 0 & a_{22} \end{bmatrix}$$

y $\det(A) = a_{11}a_{22} - a_{12} \cdot 0 = a_{11}a_{22}$ de manera que el teorema se cumple para $n = 2$. Se supondrá que se cumple para $k = n - 1$ y se demostrará para $k = n$. El determinante de una matriz triangular superior de $n \times n$ es

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ 0 & a_{22} & a_{23} & \cdots & a_{2n} \\ 0 & 0 & a_{33} & \cdots & a_{3n} \\ \vdots & & & \ddots & \\ 0 & 0 & 0 & \cdots & a_{nn} \end{vmatrix} = a_{11} \begin{vmatrix} a_{22} & a_{23} & \cdots & a_{2n} \\ 0 & a_{33} & \cdots & a_{3n} \\ \vdots & & \ddots & \\ 0 & 0 & \cdots & a_{nn} \end{vmatrix} - a_{12} \begin{vmatrix} 0 & a_{23} & \cdots & a_{2n} \\ 0 & a_{33} & \cdots & a_{3n} \\ \vdots & & \ddots & \\ 0 & 0 & \cdots & a_{nn} \end{vmatrix} \\ + a_{13} \begin{vmatrix} 0 & a_{22} & \cdots & a_{2n} \\ 0 & 0 & \cdots & a_{3n} \\ \vdots & & \ddots & \\ 0 & 0 & \cdots & a_{nn} \end{vmatrix} + \cdots + (-1)^{1+n} a_{1n} \begin{vmatrix} 0 & a_{22} & \cdots & a_{2n-1} \\ 0 & 0 & \cdots & a_{3n-1} \\ \vdots & & \ddots & \\ 0 & 0 & \cdots & 0 \end{vmatrix}$$

Cada uno de estos determinantes es el determinante de una matriz triangular superior de $(n - 1) \times (n - 1)$ que, de acuerdo con la hipótesis de inducción, es igual al producto de las componentes en la diagonal. Todas las matrices, excepto la primera, tienen una columna de ceros, por lo que por lo menos una de sus componentes diagonales es cero. De este modo, todos los determinantes, excepto el primero, son cero. Por último,

$$\det(A) = a_{11} \begin{vmatrix} a_{22} & a_{23} & \cdots & a_{2n} \\ 0 & a_{33} & \cdots & a_{3n} \\ \vdots & & \ddots & \\ 0 & 0 & \cdots & a_{nn} \end{vmatrix} = a_{11} a_{22} a_{33} \cdots a_{nn}$$

lo que prueba que el teorema se cumple para matrices de $n \times n$.

Nota

La técnica de flechas para una matriz de 3×3 se le conoce como *regla de Sarrus*. Esta solo funciona con determinantes de matrices de 3×3 . No funciona con matrices de tamaño 4×4 o superior.

Los determinantes de matrices 2×2 y 3×3 pueden evaluarse muy eficientemente usando el siguiente patrón:

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \quad \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$$

En el caso 2×2 , el determinante puede calcularse formando el producto de las entradas en la flecha hacia abajo y restando el producto de las entradas en la flecha hacia arriba. En el caso 3×3 , primero copiamos las dos primeras columnas como se muestra en la figura, después de lo cual podemos calcular el determinante sumando los productos de las entradas en las flechas hacia abajo y restando los productos de las flechas hacia arriba. Estos procedimientos nos dan como resultado

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}$$

y

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \\ = a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{21}a_{33} - a_{23}a_{31}) + a_{13}(a_{21}a_{32} - a_{22}a_{31}) \\ = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31} - a_{12}a_{21}a_{33} - a_{11}a_{23}a_{32}$$

lo que concuerda con las expansiones por cofactores a lo largo de la primera fila.

Ejemplo 3.1.7: Aplicando la regla de Sarrus a la matriz del ejemplo 3.1.3, obtenemos que:

$$\det(A) = \begin{vmatrix} 3 & 1 & 0 & 3 & 1 \\ -2 & -4 & 3 & -2 & -4 \\ 5 & 4 & -2 & 5 & 4 \end{vmatrix}$$

$\begin{matrix} & & 0 & 36 & 4 \\ & & & & \\ & & & & \\ & & & & \\ & & 24 & 15 & 0 \end{matrix}$

$$= (24 + 15 + 0) - (0 + 36 + 4)$$

$$= -1$$

lo que concuerda con el resultado obtenido en el ejemplo 3.1.3.

3.2 Evaluación de determinantes por reducción de filas

En esta sección mostraremos cómo evaluar un determinante reduciendo la matriz asociada a forma escalonada por filas. En general, este método requiere menos cálculos que la expansión por cofactores y, por lo tanto, es el método de elección para matrices grandes.

Comenzamos con un teorema fundamental que nos llevará a un procedimiento eficiente para evaluar el determinante de una matriz cuadrada de cualquier tamaño.

Teorema 3.2.1: Si A tiene una fila de ceros o una columna de ceros, entonces $\det(A) = 0$.

Demostración: Dado que el determinante de A puede encontrarse mediante una expansión por cofactores a lo largo de cualquier fila o columna, podemos usar la fila o columna de ceros. Si denotamos por C_1, C_2, \dots, C_n a los cofactores de A a lo largo de esa fila o columna, entonces se sigue de la fórmula (3.1.7) o (3.1.8) que

$$\det(A) = 0 \cdot C_1 + 0 \cdot C_2 + \dots + 0 \cdot C_n = 0.$$

El siguiente teorema útil relaciona el determinante de una matriz con el determinante de su transpuesta.

Teorema 3.2.2: Sea A una matriz cuadrada. Entonces $\det(A) = \det(A^T)$.

Demostración: Dado que trasponer una matriz cambia sus columnas por filas y sus filas por columnas, la expansión por cofactores de A a lo largo de cualquier fila es la misma que la expansión por cofactores de A^T a lo largo de la columna correspondiente. Así, ambos tienen el mismo determinante.

El siguiente teorema muestra cómo una operación elemental en el renglón de una matriz cuadrada afecta el valor de su determinante, proporcionando así una herramienta fundamental para el cálculo eficiente del determinante.

Teorema 3.2.3: Sea A una matriz $n \times n$.

- Si B es la matriz que resulta cuando un solo renglón o una sola columna de A se multiplica por un escalar k , entonces $\det(B) = k \det(A)$.
- Si B es la matriz que resulta cuando dos renglones o dos columnas de A se intercambian, entonces $\det(B) = -\det(A)$.
- Si B es la matriz que resulta cuando se agrega un múltiplo de un renglón de A a otro, o cuando se agrega un múltiplo de una columna a otra, entonces $\det(B) = \det(A)$.

Demostración: Solo demostraremos el primer inciso. Supongamos que B se obtiene de A multiplicando el i -ésimo renglón por k . De esta forma, podemos expandir mediante cofactores a lo largo de ese renglón, entonces se sigue de la fórmula (3.1.8) que

$$\begin{aligned}\det(B) &= ka_{i1}C_{i1} + ka_{i2}C_{i2} + \cdots + ka_{in}C_{in} \\ &= k(a_{i1}C_{i1} + a_{i2}C_{i2} + \cdots + a_{in}C_{in}) \\ &= k\det(A)\end{aligned}$$

En el caso de las columnas se puede hacer una demostración similar.

Será útil considerar el caso especial del teorema 3.2.3 en el que $A = I_n$ es la matriz identidad $n \times n$ y E (en lugar de B) denota la matriz elemental que resulta cuando se realiza la operación elemental. En este caso especial, el teorema 3.2.3 implica el siguiente resultado.

Teorema 3.2.4: Sea E una matriz $n \times n$ elemental.

- a) Si E resulta de multiplicar un renglón de I_n por $k \neq 0$, $\det(E) = k$.
- b) Si E resulta de intercambiar dos renglones de I_n , $\det(E) = -1$.
- c) Si E resulta de sumar un múltiplo de un renglón de I_n a otro, $\det(E) = 1$.

Ejemplo 3.2.1: Los siguientes determinantes de matrices elementales, que se evalúan por inspección, ilustran el teorema anterior.

$$\begin{array}{ccc} \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{vmatrix} = 3 & \begin{vmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{vmatrix} = -1 & \begin{vmatrix} 1 & 0 & 0 & 7 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{vmatrix} = 1 \\ \text{El segundo renglón de } I_4 \text{ fue multiplicado por 3} & \text{El primer y último renglón de } I_4 \text{ fueron intercambiados} & \text{Se sumó 7 veces el último renglón de } I_4 \text{ al último renglón} \end{array}$$

Si una matriz cuadrada A tiene dos filas proporcionales, entonces se puede introducir una fila de ceros sumando un múltiplo adecuado de una de las filas a la otra. Similarmente para columnas. Pero sumar un múltiplo de una fila o columna a otra no cambia el determinante, por lo que, a partir del teorema 3.2.1, debemos tener $\det(A) = 0$. Esto prueba el siguiente teorema.

Teorema 3.2.5: Si A es una matriz cuadrada con dos filas proporcionales o dos columnas proporcionales, entonces $\det(A) = 0$.

Ejemplo 3.2.2: Cada una de las siguientes matrices tiene dos filas o columnas proporcionales; por lo tanto, cada una tiene determinante cero.

$$\begin{bmatrix} 1 & -2 & 7 \\ -4 & 8 & 5 \\ 2 & -4 & 3 \end{bmatrix}, \quad \begin{bmatrix} 3 & -1 & 4 & -5 \\ 6 & -2 & 5 & 2 \\ 5 & 8 & 1 & 4 \\ -9 & 3 & -12 & 15 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & 4 & -5 & 3 \\ 3 & -1 & 1 & 2 & 9 \\ -2 & 9 & 1 & 4 & -6 \\ 4 & 1 & -2 & 0 & 12 \\ -1 & 0 & 8 & 7 & -3 \end{bmatrix}.$$

Presentaremos ahora un método para evaluar determinantes que requiere considerablemente menos cálculos que la expansión por cofactores. La idea del método consiste en reducir la matriz dada a una forma triangular superior mediante operaciones elementales por renglón, luego calcular el determinante de dicha matriz triangular superior (un cálculo sencillo) y, finalmente, relacionar ese determinante con el de la matriz original.

A continuación, mostramos un ejemplo.

Ejemplo 3.2.3: Calcule el determinante de la siguiente matriz:

$$A = \begin{bmatrix} 0 & 1 & 5 \\ 3 & -6 & 9 \\ 2 & 6 & 1 \end{bmatrix}.$$

Solución: Reduciremos A a su forma escalonada (que es triangular superior) y luego aplicaremos el teorema 3.1.3.

$$\begin{aligned} \det(A) &= \begin{vmatrix} 0 & 1 & 5 \\ 3 & -6 & 9 \\ 2 & 6 & 1 \end{vmatrix} = - \begin{vmatrix} 3 & -6 & 9 \\ 0 & 1 & 5 \\ 2 & 6 & 1 \end{vmatrix} && \text{El primer y segundo renglón de } A \text{ fueron intercambiados} \\ &= -3 \begin{vmatrix} 1 & -2 & 3 \\ 0 & 1 & 5 \\ 2 & 6 & 1 \end{vmatrix} && \text{Se tomó un múltiplo común de 3 del primer renglón} \\ &= -3 \begin{vmatrix} 1 & -2 & 3 \\ 0 & 1 & 5 \\ 0 & 10 & -5 \end{vmatrix} && \text{Se sumó } -2 \text{ veces el primer renglón al tercer renglón} \\ &= -3 \begin{vmatrix} 1 & -2 & 3 \\ 0 & 1 & 5 \\ 0 & 0 & -55 \end{vmatrix} && \text{Se sumó } -10 \text{ veces el segundo renglón al tercer renglón} \\ &= (-3)((1)(1)(-55)) && \text{Al ser triangular superior, el determinante es el producto de las componentes de la diagonal} \\ &= 165 \end{aligned}$$

Ejemplo 3.2.4: Calcule el determinante de la siguiente matriz:

$$A = \begin{bmatrix} 1 & 0 & 0 & 3 \\ 2 & 7 & 0 & 6 \\ 0 & 6 & 3 & 0 \\ 7 & 3 & 1 & -5 \end{bmatrix}.$$

Solución: Este determinante podría calcularse como se indicó anteriormente utilizando operaciones elementales para reducir A a su forma escalonada, pero podemos poner A en forma triangular inferior en un solo paso sumando -3 veces la primera columna a la cuarta para obtener

$$\det(A) = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 2 & 7 & 0 & 0 \\ 0 & 6 & 3 & 0 \\ 7 & 3 & 1 & -26 \end{vmatrix} = (1)(7)(3)(-26) = -546.$$

La expansión de cofactores y las operaciones elementales a veces se pueden usar en combinación para proporcionar un método llamado *regla de Chio-Laplace*. Este método se basa en el principio de que se pueden usar las propiedades de los determinantes para crear ceros en un renglón o columna. La idea principal es reducir una matriz de orden $n \times n$ a una de orden $(n-1) \times (n-1)$ de forma sucesiva hasta llegar a una matriz de 2×2 , cuyo determinante se puede calcular directamente.

Para aplicar el método, se selecciona un elemento de la matriz llamado “pivote” que idealmente sea 1. Luego, se realizan operaciones entre renglones o columnas para convertir todos los demás elementos del renglón o columna del pivote en cero. Después de esto, el determinante se calcula a partir del pivote.

Ejemplo 3.2.5: Calcule el determinante de la siguiente matriz:

$$A = \begin{bmatrix} 3 & 5 & -2 & 6 \\ 1 & 2 & -1 & 1 \\ 2 & 4 & 1 & 5 \\ 3 & 7 & 5 & 3 \end{bmatrix}.$$

Solución: Aplicando la regla de Chio-Laplace, obtenemos que

$$\begin{aligned} \det(A) &= \begin{vmatrix} 3 & 5 & -2 & 6 \\ 1 & 2 & -1 & 1 \\ 2 & 4 & 1 & 5 \\ 3 & 7 & 5 & 3 \end{vmatrix} = \begin{vmatrix} 0 & -1 & 1 & 3 \\ 1 & 2 & -1 & 1 \\ 0 & 0 & 3 & 3 \\ 0 & 1 & 8 & 0 \end{vmatrix} && \begin{array}{l} \text{Se sumó } -2 \text{ veces el segundo} \\ \text{renglón al tercer renglón, y se} \\ \text{sumó } -3 \text{ veces el segundo ren-} \\ \text{glón al primer y último renglón} \end{array} \\ &= - \begin{vmatrix} -1 & 1 & 3 \\ 0 & 3 & 3 \\ 1 & 8 & 0 \end{vmatrix} && \begin{array}{l} \text{Expandimos por cofactores a lo} \\ \text{largo de la primer columna} \end{array} \\ &= - \begin{vmatrix} -1 & 1 & 3 \\ 0 & 3 & 3 \\ 0 & 9 & 3 \end{vmatrix} && \begin{array}{l} \text{Se sumó 1 vez el primer} \\ \text{renglón al segundo renglón} \end{array} \\ &= -(-1) \begin{vmatrix} 3 & 3 \\ 9 & 3 \end{vmatrix} && \begin{array}{l} \text{Expandimos por cofactores a lo} \\ \text{largo de la primer columna} \end{array} \\ &= -18 \end{aligned}$$

3.3 Propiedades de los determinantes

En esta sección desarrollaremos algunas propiedades fundamentales de matrices, y utilizaremos estos resultados para derivar una fórmula para la inversa de una matriz no singular.

Supongamos que A y B son matrices $n \times n$ y k es cualquier escalar. Comenzaremos considerando las posibles relaciones entre $\det(A)$, $\det(B)$ y

$$\det(kA), \quad \det(A + B), \quad \det(AB).$$

Dado que un factor común de cualquier fila de una matriz puede eliminarse a través del determinante, y dado que cada una de las n filas en kA tiene un factor común de k , se sigue que

$$\det(kA) = k^n \det(A).$$

Por ejemplo,

$$\begin{vmatrix} ka_{11} & ka_{12} & ka_{13} \\ ka_{21} & ka_{22} & ka_{23} \\ ka_{31} & ka_{32} & ka_{33} \end{vmatrix} = k^3 \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}.$$

Desafortunadamente, no siempre existe una relación simple entre $\det(A)$, $\det(B)$, y $\det(A + B)$. En particular, $\det(A + B)$ usualmente no es igual a $\det(A) + \det(B)$. El siguiente ejemplo ilustra este hecho.

Ejemplo 3.3.1: Consideremos

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 5 \end{bmatrix}, \quad B = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}, \quad A + B = \begin{bmatrix} 4 & 3 \\ 3 & 8 \end{bmatrix}.$$

Tenemos que $\det(A) = 1$, $\det(B) = 8$, y $\det(A + B) = 23$; así

$$\det(A + B) \neq \det(A) + \det(B).$$

A pesar del ejemplo anterior, existe una relación útil concerniente a las sumas de determinantes que es aplicable cuando las matrices involucradas son iguales excepto por una fila (columna). Por ejemplo, consideremos las siguientes dos matrices que difieren solo en la segunda fila:

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \quad \text{y} \quad B = \begin{bmatrix} a_{11} & a_{12} \\ b_{21} & b_{22} \end{bmatrix}.$$

Calculando los determinantes de A y B , obtenemos

$$\begin{aligned} \det(A) + \det(B) &= (a_{11}a_{22} - a_{12}a_{21}) + (a_{11}b_{22} - a_{12}b_{21}) \\ &= a_{11}(a_{22} + b_{22}) - a_{12}(a_{21} + b_{21}) \\ &= \det \left(\begin{bmatrix} a_{11} & a_{12} \\ a_{21} + b_{21} & a_{22} + b_{22} \end{bmatrix} \right) \end{aligned}$$

Así

$$\det \left(\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \right) + \det \left(\begin{bmatrix} a_{11} & a_{12} \\ b_{21} & b_{22} \end{bmatrix} \right) = \det \left(\begin{bmatrix} a_{11} & a_{12} \\ a_{21} + b_{21} & a_{22} + b_{22} \end{bmatrix} \right).$$

Esto es un caso especial del siguiente resultado general.

Teorema 3.3.1: Sean A , B y C matrices $n \times n$ que difieren únicamente en un solo renglón, digamos el i -ésimo, y asumamos que el i -ésimo renglón de C puede obtenerse sumando las entradas correspondientes en los i -ésimos renglones de A y B . Entonces

$$\det(C) = \det(A) + \det(B).$$

El mismo resultado se cumple para columnas.

Ejemplo 3.3.2: Dejamos al lector la tarea de confirmar la siguiente igualdad

$$\begin{vmatrix} 1 & 7 & 5 \\ 2 & 0 & 3 \\ 1+0 & 4+1 & 7+(-1) \end{vmatrix} = \begin{vmatrix} 1 & 7 & 5 \\ 2 & 0 & 3 \\ 1 & 4 & 7 \end{vmatrix} + \begin{vmatrix} 1 & 7 & 5 \\ 2 & 0 & 3 \\ 0 & 1 & -1 \end{vmatrix}.$$

Considerando la complejidad de las fórmulas para determinantes y multiplicación de matrices, parecería poco probable que existiera una relación simple entre ellos. Esto es lo que hace que la simplicidad de nuestro próximo resultado sea tan sorprendente. Mostraremos que si A y B son matrices cuadradas del mismo tamaño, entonces

$$\det(AB) = \det(A) \det(B). \quad (3.3.1)$$

La demostración de este teorema es bastante compleja, por lo que tendremos que desarrollar algunos resultados preliminares primero. Comenzamos con el caso especial de (3.3.1) en el que A es una matriz elemental. Dado que este caso especial es solo un preludio a (3.3.1), lo llamamos un lema.

Lema 3.3.1: Si B es una matriz $n \times n$ y E es una matriz $n \times n$ elemental, entonces

$$\det(EB) = \det(E) \det(B).$$

Demostración: Consideraremos tres casos, cada uno de acuerdo con la operación elemental que produce la matriz E .

- i) Si E resulta de multiplicar un renglón de I_n por k , entonces por el teorema 2.6.5, EB resulta de B multiplicando el renglón correspondiente por k ; así, por el teorema 3.2.3a) tenemos

$$\det(EB) = k \det(B).$$

Pero por el teorema 3.2.4a) tenemos $\det(E) = k$, así

$$\det(EB) = \det(E) \det(B).$$

- ii) Si E se obtiene intercambiando los renglones i y j de I_n , entonces por el teorema 2.6.5, EB resulta de intercambiar los renglones i y j de B ; así, por el teorema 3.2.3b) tenemos

$$\det(EB) = -\det(B).$$

Pero por el teorema 3.2.4b) tenemos $\det(E) = -1$, así

$$\det(EB) = \det(E) \det(B).$$

- iii) Se deja como ejercicio al lector.

Aplicando repetidas veces el lema anterior, se sigue que que si B es una matriz $n \times n$ y E_1, E_2, \dots, E_r son matrices $n \times n$ elementales, entonces

$$\det(E_1 E_2 \cdots E_r B) = \det(E_1) \det(E_2) \cdots \det(E_r) \det(B) \quad (3.3.2)$$

Nuestro siguiente teorema proporciona un criterio importante para determinar si una matriz es invertible.

Teorema 3.3.2: Una matriz cuadrada A es invertible si y solo si $\det(A) \neq 0$.

Demostración: Sea A equivalente por renglones a B , donde B está en forma escalonada reducida. Como paso preliminar, mostraremos que $\det(A)$ y $\det(B)$ son ambos cero o ambos no cero: Sean E_1, E_2, \dots, E_r las matrices elementales que corresponden a las operaciones elementales que producen B a partir de A . Así

$$B = E_r \cdots E_2 E_1 A$$

y por (3.3.2),

$$\det(B) = \det(E_r) \cdots \det(E_2) \det(E_1) \det(A). \quad (3.3.3)$$

Del teorema 3.2.4, se sabe que el determinante de una matriz elemental es no cero. Así, se deduce de la fórmula (3.3.3) que $\det(A)$ y $\det(B)$ son ambos cero o ambos no cero, lo que establece el escenario para la parte principal de la demostración. Si asumimos primero que A es invertible, entonces se deduce del teorema 2.6.7 que

$$B = I \quad \text{y por lo tanto que} \quad \det(B) = 1 \neq 0.$$

Esto, a su vez, implica que $\det(A) \neq 0$, que es lo que queríamos mostrar. Inversamente, asumamos que $\det(A) \neq 0$. Esto implica que $\det(B) \neq 0$, lo que nos dice que B no puede tener un renglón de ceros. Así, se deduce del teorema 2.6.7 que $B = I$ y por lo tanto que A es invertible por el teorema 2.6.7.

Ejemplo 3.3.3: Dado que el primer y el tercer renglón de

$$A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 1 & 0 & 1 & 0 \\ 2 & 4 & 6 & 8 \\ 0 & 1 & 3 & 2 \end{bmatrix}$$

son proporcionales, entonces $\det(A) = 0$. Por lo tanto, A no es invertible.

Ahora bien, estamos listos para el resultado principal con respecto a los productos de matrices.

Teorema 3.3.3: Si A y B son matrices cuadradas del mismo tamaño, entonces

$$\det(AB) = \det(A) \det(B).$$

Demostración: Dividimos la prueba en dos casos que dependen de si A es invertible o no. Si la matriz A no es invertible, entonces por el ejercicio 132 en la página 97, tampoco lo es el producto AB . Así, por el teorema 3.3.2, tenemos $\det(AB) = 0$ y $\det(A) = 0$, por lo que se deduce que $\det(AB) = \det(A) \det(B)$. Ahora asumamos que A es invertible. Por el teorema 2.6.7, la matriz A se puede expresar como un producto de matrices elementales, digamos

$$A = E_1 E_2 \cdots E_r \quad (3.3.4)$$

por lo que

$$AB = E_1 E_2 \cdots E_r B.$$

Aplicando (3.3.2) a esta ecuación da como resultado

$$\det(AB) = \det(E_1) \det(E_2) \cdots \det(E_r) \det(B)$$

y aplicando (3.3.2) nuevamente da

$$\det(AB) = \det(E_1 E_2 \cdots E_r) \det(B)$$

lo que, a partir de (3.3.4), se puede escribir como $\det(AB) = \det(A) \det(B)$.

Ejemplo 3.3.4: Consideremos las matrices

$$A = \begin{bmatrix} 3 & 1 \\ 2 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} -1 & 3 \\ 5 & 8 \end{bmatrix}, \quad AB = \begin{bmatrix} 2 & 17 \\ 3 & 14 \end{bmatrix}.$$

Dejamos al lector la tarea de verificar que

$$\det(A) = 1, \quad \det(B) = -23, \quad \det(AB) = -23.$$

Así, $\det(AB) = \det(A) \det(B)$, como se garantiza por el teorema 3.3.3.

El siguiente teorema da una relación útil entre el determinante de una matriz invertible y el determinante de su inversa.

Teorema 3.3.4: Si A es invertible, entonces

$$\det(A^{-1}) = \frac{1}{\det(A)}.$$

Demostración: Dado que $A^{-1}A = I$, se deduce que

$$\det(A^{-1}A) = \det(I).$$

Por lo tanto, se sigue que

$$\det(A^{-1}) \det(A) = 1.$$

Dado que $\det(A) \neq 0$, podemos dividir entre $\det(A)$ y obtener así,

$$\det(A^{-1}) = \frac{1}{\det(A)}.$$

En una expansión de cofactores, $\det(A)$ se calcula multiplicando las entradas de una fila o columna por sus cofactores correspondientes y sumando los productos. Resulta que si se multiplican las entradas de una fila por los cofactores de otra fila, la suma de los productos es cero (lo mismo ocurre con las columnas).

Ejemplo 3.3.5: Sea

$$A = \begin{bmatrix} 3 & 2 & -1 \\ 1 & 6 & 3 \\ 2 & -4 & 0 \end{bmatrix}.$$

Se deja al lector la tarea de verificar que los cofactores de A son

$$\begin{array}{lll} C_{11} = 12 & C_{12} = 6 & C_{13} = -16 \\ C_{21} = 4 & C_{22} = 2 & C_{23} = 16 \\ C_{31} = 12 & C_{32} = -10 & C_{33} = 16 \end{array}$$

por lo que, por ejemplo, la expansión de cofactores de $\det(A)$ a lo largo de la primera fila es

$$\det(A) = 3C_{11} + 2C_{12} + (-1)C_{13} = 36 + 12 + 16 = 64$$

y a lo largo de la primera columna es

$$\det(A) = 3C_{11} + 1C_{21} + 2C_{31} = 36 + 4 + 24 = 64.$$

Supongamos, sin embargo, que multiplicamos las entradas en la primera fila por los cofactores correspondientes de la segunda fila y sumamos los productos resultantes. El resultado es

$$3C_{21} + 2C_{22} + (-1)C_{23} = 12 + 4 - 16 = 0.$$

O supongamos que multiplicamos las entradas en la primera columna por los cofactores correspondientes de la segunda columna y sumamos los productos resultantes. El resultado es nuevamente cero, ya que

$$3C_{12} + 1C_{22} + 2C_{32} = 18 + 2 - 20 = 0.$$

Definición 3.3.1:

Si A es una matriz $n \times n$ y C_{ij} es el cofactor de a_{ij} , entonces la matriz

$$\begin{bmatrix} C_{11} & C_{12} & \cdots & C_{1n} \\ C_{21} & C_{22} & \cdots & C_{2n} \\ \vdots & \vdots & & \vdots \\ C_{n1} & C_{n2} & \cdots & C_{nn} \end{bmatrix}$$

se llama la *matriz de cofactores de A* . La transpuesta de esta matriz se llama el *adjunto clásico de A* y se denota por $\text{adj}(A)$.

Ejemplo 3.3.6:

Consideremos la matriz del ejemplo 3.3.5. Como se notó en el ejemplo 3.3.5, los cofactores de A son

$$\begin{array}{lll} C_{11} = 12 & C_{12} = 6 & C_{13} = -16 \\ C_{21} = 4 & C_{22} = 2 & C_{23} = 16 \\ C_{31} = 12 & C_{32} = -10 & C_{33} = 16 \end{array}$$

por lo que la matriz de cofactores es

$$\begin{bmatrix} 12 & 6 & -16 \\ 4 & 2 & 16 \\ 12 & -10 & 16 \end{bmatrix}$$

y el adjunto clásico de A es

$$\text{adj}(A) = \begin{bmatrix} 12 & 4 & 12 \\ 6 & 2 & -10 \\ -16 & 16 & 16 \end{bmatrix}.$$

En el teorema 3.1.1 dimos una fórmula para la inversa de una matriz invertible 2×2 . Nuestro próximo teorema extiende ese resultado a matrices invertibles $n \times n$.

Teorema 3.3.5: Si A es una matriz invertible, entonces

$$A^{-1} = \frac{1}{\det(A)} \operatorname{adj}(A). \quad (3.3.5)$$

Demostración: Mostraremos primero que

$$A \operatorname{adj}(A) = \det(A)I.$$

Consideremos el producto

$$A \operatorname{adj}(A) = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} C_{11} & C_{21} & \cdots & C_{j1} & \cdots & C_{n1} \\ C_{12} & C_{22} & \cdots & C_{j2} & \cdots & C_{n2} \\ \vdots & \vdots & & \vdots & & \vdots \\ C_{1n} & C_{2n} & \cdots & C_{jn} & \cdots & C_{nn} \end{bmatrix}.$$

La entrada en la fila i -ésima y columna j -ésima del producto $A \operatorname{adj}(A)$ es

$$a_{i1}C_{j1} + a_{i2}C_{j2} + \cdots + a_{in}C_{jn} \quad (3.3.6)$$

Si $i = j$, entonces (3.3.6) es la expansión de cofactores de $\det(A)$ a lo largo de la fila i -ésima de A (teorema 3.1.2), y si $i \neq j$, entonces las a 's y los cofactores provienen de filas diferentes de A , por lo que el valor de (3.3.6) es cero (como se ilustró en el ejemplo 3.3.5). Por lo tanto,

$$A \operatorname{adj}(A) = \begin{bmatrix} \det(A) & 0 & \cdots & 0 \\ 0 & \det(A) & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \det(A) \end{bmatrix} = \det(A)I.$$

Dado que A es invertible, $\det(A) \neq 0$. Por lo tanto, la expresión anterior puede reescribirse como

$$\frac{1}{\det(A)} [A \operatorname{adj}(A)] = I$$

o bien,

$$A \left[\frac{1}{\det(A)} \operatorname{adj}(A) \right] = I.$$

Multiplicando ambos lados por la izquierda por A^{-1} da como resultado

$$A^{-1} = \frac{1}{\det(A)} \operatorname{adj}(A).$$

Ejemplo 3.3.7: Usa la fórmula (3.3.5) para encontrar la inversa de la matriz A en el ejemplo 3.3.5.

Demostración: Mostramos en el ejemplo 3.3.5 que $\det(A) = 64$. Así,

$$A^{-1} = \frac{1}{\det(A)} \operatorname{adj}(A) = \frac{1}{64} \begin{bmatrix} 12 & 4 & 12 \\ 6 & 2 & -10 \\ -16 & 16 & 16 \end{bmatrix} = \begin{bmatrix} \frac{12}{64} & \frac{4}{64} & \frac{12}{64} \\ \frac{6}{64} & \frac{2}{64} & -\frac{10}{64} \\ -\frac{16}{64} & \frac{16}{64} & \frac{16}{64} \end{bmatrix}.$$

3.4 Otros métodos para calcular determinantes

En el cálculo de determinantes, existen diversos métodos. Una estrategia fundamental consiste en utilizar las propiedades de los determinantes para introducir ceros en una fila o columna. El objetivo es simplificar la matriz para luego calcular el determinante de manera más directa, por ejemplo, mediante una expansión por cofactores a lo largo de esa fila o columna que ahora contiene múltiples ceros. Un método que emplea esta lógica es la regla de Chio-Laplace, que reduce sucesivamente el orden de la matriz.

Si bien la expansión por cofactores es el método que permite definir el determinante para una matriz de cualquier orden de manera inductiva, para matrices grandes este procedimiento puede requerir una gran cantidad de cálculos. Por esta razón, se presentó un método alternativo más eficiente: la evaluación de determinantes mediante la reducción de la matriz a su forma triangular. Una vez en forma triangular, el determinante se calcula multiplicando las componentes de la diagonal.

A continuación, examinaremos en detalle otros métodos a los ya mencionados y se presentarán problemas diseñados para asimilar cada técnica.

i Nota

No existe un único método universal para todos los casos; la elección dependerá de la estructura de la matriz. Para ciertos tipos especiales, se aplican técnicas que resultan en expresiones más simples. Un ejemplo es la regla de Sarrus, un procedimiento de patrones con flechas que es muy eficiente, pero que funciona exclusivamente para determinantes de matrices de 3×3 .

Método de separación de los factores lineales

El método de separación de los factores lineales consiste en considerar el determinante como un polinomio de una o varias variables que intervienen en sus elementos. Al efectuar transformaciones adecuadas sobre sus filas o columnas, dicho polinomio puede factorizarse en el producto de varios factores lineales. Si estos factores resultan ser primos entre sí, entonces el determinante se puede expresar como el producto de esos factores lineales y de un coeficiente constante.

Para determinar dicho coeficiente, se comparan algunos términos del determinante original con los términos correspondientes del producto de los factores lineales previamente obtenidos. De esta manera, se calcula el cociente de la división del determinante por dicho producto y, finalmente, se obtiene una expresión explícita y simplificada para el determinante.

Ejemplo 3.4.1: Calcular el determinante

$$D = \begin{vmatrix} 0 & x & y & z \\ x & 0 & z & y \\ y & z & 0 & x \\ z & y & x & 0 \end{vmatrix}.$$

Solución: Si a la primera columna de D se le añaden todas las demás, se ve que el determinante tiene como factor común $x + y + z$ en la primera columna, es decir,

$$\begin{vmatrix} x+y+z & x & y & z \\ x+y+z & 0 & z & y \\ x+y+z & z & 0 & x \\ x+y+z & y & x & 0 \end{vmatrix} = (x+y+z) \begin{vmatrix} 1 & x & y & z \\ 1 & 0 & z & y \\ 1 & z & 0 & x \\ 1 & y & x & 0 \end{vmatrix}.$$

Si a la primera columna de D se le añade la segunda y se restan la tercera y cuarta columna, aparece el factor $y + z - x$, es decir,

$$\begin{vmatrix} x-y-z & x & y & z \\ x-z-y & 0 & z & y \\ y+z-x & z & 0 & x \\ z+y-x & y & x & 0 \end{vmatrix} = (y+z-x) \begin{vmatrix} -1 & x & y & z \\ -1 & 0 & z & y \\ 1 & z & 0 & x \\ 1 & y & x & 0 \end{vmatrix}.$$

Si a la primer columna de D se le añade la tercera y se sustraen la segunda y cuarta columnas, surge el factor $x - y + z$, es decir,

$$\begin{vmatrix} y-x-z & x & y & z \\ x+z-y & 0 & z & y \\ y-z-x & z & 0 & x \\ z+x-y & y & x & 0 \end{vmatrix} = (x-y+z) \begin{vmatrix} -1 & x & y & z \\ 1 & 0 & z & y \\ -1 & z & 0 & x \\ 1 & y & x & 0 \end{vmatrix}.$$

Por último, si a la primer columna de D se le añade la cuarta y se sustraen la segunda y tercera columnas, sale el factor $x + y - z$, es decir,

$$\begin{vmatrix} z-x-y & x & y & z \\ x+y-z & 0 & z & y \\ y+x-z & z & 0 & x \\ z-y-x & y & x & 0 \end{vmatrix} = (x+y-z) \begin{vmatrix} -1 & x & y & z \\ 1 & 0 & z & y \\ 1 & z & 0 & x \\ -1 & y & x & 0 \end{vmatrix}.$$

Considerando que x, y, z son incógnitas independientes, hacemos la conclusión de que todos esos cuatro factores son primos entre sí de dos en dos lo que significa que el determinante se divide por su producto $(x+y+z)(y+z-x)(x-y+z)(x+y-z)$. Este producto contiene el término z^4 con el coeficiente -1 , mientras que el propio determinante tiene el mismo término z^4 con el coeficiente $+1$. Entonces,

$$\begin{aligned} D &= -(x+y+z)(y+z-x)(x-y+z)(x+y-z) \\ &= x^4 + y^4 + z^4 - 2x^2y^2 - 2x^2z^2 - 2y^2z^2. \end{aligned}$$

Ejemplo 3.4.2:

Aplicando el método de separación de los factores lineales, calcular el determinante de Vandermonde de orden n :

$$D_n = \begin{vmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{n-1} \\ \vdots & & & \ddots & \\ 1 & x_n & x_n^2 & \cdots & x_n^{n-1} \end{vmatrix}.$$

Solución: Considerando D_n como polinomio en x_n de grado $n-1$, se anula cuando $x_n = x_i$ para $i = 1, 2, \dots, n-1$ (pues dos filas serían iguales). Así, $(x_n - x_i)$ es factor para cada i . Todos esos factores son primos entre sí, ya que x_1, x_2, \dots, x_n son independientes desde el punto de vista algebraico. Por consiguiente, D_n se divide por su producto, es decir,

$$D_n = q(x_1, x_2, \dots, x_n) \prod_{k=1}^{n-1} (x_n - x_k). \quad (3.4.1)$$

donde $q(x_1, x_2, \dots, x_n)$ es un polinomio. Al desarrollar D_n por la última fila, el coeficiente de x_n^{n-1} es el cofactor C_{nn} . Este cofactor es el determinante de la matriz obtenida al eliminar la última fila y última columna, que es precisamente el determinante de Vandermonde de orden $n-1$, pues

$$C_{nn} = (-1)^{n+n} D_{n-1} = \begin{vmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{n-2} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{n-2} \\ \vdots & & & \ddots & \\ 1 & x_{n-1} & x_{n-1}^2 & \cdots & x_{n-1}^{n-2} \end{vmatrix}.$$

Como el producto de la expresión (3.4.1) contiene x_n^{n-1} con el coeficiente 1, el polinomio $q(x_1, x_2, \dots, x_n)$ no posee el término x_n y comparando los coeficientes de x_n^{n-1} en ambos miembros de la igualdad, obtenemos

$$D_{n-1} = q(x_1, x_2, \dots, x_n).$$

Sustituyendo $q(x_1, x_2, \dots, x_n)$ en la expresión (3.4.1),

$$D_n = D_{n-1} \prod_{k=1}^{n-1} (x_n - x_k).$$

Aplicando recursivamente esta relación,

$$\begin{aligned} D_{n-1} &= D_{n-2} \prod_{k=1}^{n-2} (x_{n-1} - x_k) \\ D_{n-2} &= D_{n-3} \prod_{k=1}^{n-3} (x_{n-2} - x_k) \\ &\vdots \\ D_2 &= D_1 (x_2 - x_1) = 1 \cdot (x_2 - x_1) \end{aligned}$$

Sustituyendo hacia arriba y observando que $D_1 = |1| = 1$, obtenemos que

$$D_n = \prod_{i=2}^n \prod_{j=1}^{i-1} (x_i - x_j) = \prod_{1 \leq j < i \leq n} (x_i - x_j).$$

Método de relaciones recurrentes

Este método consiste en que el determinante dado se expresa, transformándolo y descomponiéndolo por una fila o columna, mediante los determinantes de la misma forma pero de un orden más inferior. La igualdad obtenida se denomina *relación recurrente*.

Después se calculan directamente, por la forma general del determinante, la cantidad de determinantes de órdenes inferiores tantos, cuantos había en el segundo miembro de la relación recurrente. Los determinantes de orden más elevado se calculan sucesivamente de la relación recurrente. Si hay que obtener una expresión para un determinante de cualquier orden n , entonces, calculando varios determinantes de órdenes inferiores a partir de la relación recurrente, se tiende a advertir la forma general de la expresión buscada y luego se demuestra la validez de ésta para cualquier n mediante la relación recurrente y el método de inducción por n .

La expresión general puede obtenerse también de otra manera. Para ello en la relación recurrente que denota el determinante de orden n se pone la expresión del determinante de orden $(n-1)$ de la misma relación recurrente, sustituyendo n por $n-1$, luego se pone la expresión análoga del determinante de orden $(n-2)$, etc., hasta que se aclara la forma de la expresión general buscada del determinante de orden n . Pueden también combinarse ambas vías, utilizando la segunda para encontrar la expresión buscada y demostrando luego mediante inducción sobre n la validez de esa expresión. El método de relaciones recurrentes es el más eficaz entre los métodos analizados aquí y se aplica a determinantes más complejos.

Antes de pasar a los ejemplos de cálculo de los determinantes empleando el método de relaciones recurrentes, examinemos un caso particular de éste en el que la relación recurrente ofrece el algoritmo para resolver el problema, excluyendo el elemento de suposición que existe en el caso general. Supongamos que la relación recurrente tiene el siguiente aspecto

$$D_n = pD_{n-1} + qD_{n-2}, \quad n > 2, \quad (3.4.2)$$

donde p, q son constantes, o sea, magnitudes independientes de n .

Para $q = 0$, D_n puede calcularse como un término de la progresión geométrica: $D_n = p^{n-1} D_1$; aquí D_1 es el determinante de primer orden de la forma dada.

i Nota

Este método se aplica también a la relación recurrente $D_n = p_1 D_{n-1} + \dots + p_k D_{n-k}$, siendo p_1, \dots, p_k constantes y k cualquiera, pero a causa de que los razonamientos son muy voluminosos, nos limitaremos solo a $k = 2$.

Sea $q \neq 0$ y α, β raíces de la ecuación cuadrática $x^2 - px - q = 0$. Entonces $p = \alpha + \beta$, $q = -\alpha\beta$ y la igualdad (3.4.2) puede escribirse de la siguiente manera:

$$D_n - \beta D_{n-1} = \alpha(D_{n-1} - \beta D_{n-2}) \quad (3.4.3)$$

o

$$D_n - \alpha D_{n-1} = \beta(D_{n-1} - \alpha D_{n-2}) \quad (3.4.4)$$

Supongamos primero que $\alpha \neq \beta$. Partiendo de las igualdades (3.4.3) y (3.4.4) por la fórmula para el término $(n-1)$ de la progresión geométrica, hallamos

$$D_n - \beta D_{n-1} = \alpha^{n-2}(D_2 - \beta D_1)$$

y

$$D_n - \alpha D_{n-1} = \beta^{n-2}(D_2 - \alpha D_1)$$

de donde

$$D_n = \frac{\alpha^{n-1}(D_2 - \beta D_1) - \beta^{n-1}(D_2 - \alpha D_1)}{\alpha - \beta}$$

o bien,

$$D_n = C_1 \alpha^n + C_2 \beta^n$$

donde

$$\begin{aligned} C_1 &= \frac{D_2 - \beta D_1}{\alpha(\alpha - \beta)} \\ C_2 &= -\frac{D_2 - \alpha D_1}{\beta(\alpha - \beta)} \end{aligned} \quad (3.4.5)$$

La última expresión para D_n es fácil de recordar. Fue obtenida para $n > 2$, pero se verifica directamente para $n = 1$ y $n = 2$. Los valores de C_1 y C_2 pueden obtenerse no solo de las expresiones dadas en (3.4.5), sino también a partir de las condiciones iniciales

$$D_1 = C_1 \alpha + C_2 \beta, \quad D_2 = C_1 \alpha^2 + C_2 \beta^2.$$

Supongamos ahora que $\alpha = \beta$. Las igualdades (3.4.3) y (3.4.4) se reducen a la misma:

$$D_n - \alpha D_{n-1} = \alpha(D_{n-1} - \alpha D_{n-2}),$$

de donde

$$D_n - \alpha D_{n-1} = A \alpha^{n-2}, \quad (3.4.6)$$

donde

$$A = D_2 - \alpha D_1.$$

Sustituyendo aquí n por $n-1$, obtenemos

$$D_{n-1} - \alpha D_{n-2} = A \alpha^{n-3},$$

de donde

$$D_{n-1} = \alpha D_{n-2} + A \alpha^{n-3}.$$

Sustituyendo esta expresión en la igualdad (3.4.6), encontramos

$$D_n = \alpha^2 D_{n-2} + 2A \alpha^{n-2}.$$

Repitiendo este procedimiento varias veces, obtenemos

$$D_n = \alpha^{n-1} D_1 + (n-1)A \alpha^{n-2},$$

o bien

$$D_n = \alpha^n [(n-1)C_1 + C_2]$$

donde $C_1 = \frac{A}{\alpha^2}$, $C_2 = \frac{D_1}{\alpha}$ (aquí $\alpha \neq 0$, ya que $q \neq 0$).

Ejemplo 3.4.3: Aplicando el método de relaciones recurrentes, calcular el siguiente determinante:

$$D_n = \begin{vmatrix} a_1 & x & x & \cdots & x \\ x & a_2 & x & \cdots & x \\ x & x & a_3 & \cdots & x \\ \vdots & & & \ddots & \\ x & x & x & \cdots & a_n \end{vmatrix}.$$

Solución: Representando el elemento en la esquina inferior derecha en la forma $a_n = x + (a_n - x)$, podemos descomponer el determinante D_n en la suma de dos determinantes:

$$D_n = \begin{vmatrix} a_1 & x & \cdots & x & x \\ x & a_2 & \cdots & x & x \\ \vdots & & \ddots & & \vdots \\ x & x & \cdots & a_{n-1} & x \\ x & x & \cdots & x & x \end{vmatrix} + \begin{vmatrix} a_1 & x & \cdots & x & 0 \\ x & a_2 & \cdots & x & 0 \\ \vdots & & \ddots & & \vdots \\ x & x & \cdots & a_{n-1} & 0 \\ x & x & \cdots & x & a_n - x \end{vmatrix}.$$

En el primer determinante restamos la última columna de las demás, y el segundo determinante lo desarrollamos por la última columna

$$D_n = x(a_1 - x)(a_2 - x) \cdots (a_{n-1} - x) + (a_n - x)D_{n-1}.$$

Esta es precisamente la relación recurrente. Introduciendo en ella la expresión análoga para D_{n-1} , hallaremos

$$D_n = x(a_1 - x)(a_2 - x) \cdots (a_{n-1} - x) + x(a_1 - x)(a_2 - x) \cdots (a_{n-2} - x)(a_n - x) + D_{n-2}(a_{n-1} - x)(a_n - x)$$

Repitiendo el mismo razonamiento $n - 1$ veces y observando que

$$D_1 = a_1 = x + (a_1 - x),$$

obtenemos que

$$\begin{aligned} D_n &= x(a_1 - x)(a_2 - x) \cdots (a_{n-1} - x) + x(a_1 - x) \cdots (a_{n-2} - x)(a_n - x) + \cdots \\ &\quad \cdots + x(a_2 - x) \cdots (a_n - x) + (a_1 - x)(a_2 - x) \cdots (a_n - x) \\ &= x(a_1 - x)(a_2 - x) \cdots (a_n - x) \left(\frac{1}{x} + \frac{1}{a_1 - x} + \cdots + \frac{1}{a_n - x} \right) \end{aligned}$$

Ejemplo 3.4.4: Calcular el determinante de orden n :

$$D_n = \begin{vmatrix} 5 & 3 & 0 & 0 & \cdots & 0 & 0 \\ 2 & 5 & 3 & 0 & \cdots & 0 & 0 \\ 0 & 2 & 5 & 3 & \cdots & 0 & 0 \\ \vdots & & & & \ddots & & \\ 0 & 0 & 0 & 0 & \cdots & 2 & 5 \end{vmatrix}.$$

Solución: Descomponiendo por la primera fila, hallemos la relación recurrente

$$D_n = 5D_{n-1} - 6D_{n-2}.$$

La ecuación

$$x^2 - 5x + 6 = 0$$

tiene raíces $\alpha = 2$, $\beta = 3$. Aplicando la fórmula (3.4.5),

$$D_n = C_1 \alpha^n + C_2 \beta^n = 3^{n+1} - 2^{n+1}.$$

Método de representación de un determinante como la suma de determinantes

Algunos determinantes se pueden calcular fácilmente descomponiéndolos en la suma de determinantes del mismo orden con respecto a las filas (o columnas).

Ejemplo 3.4.5: Calcular el determinante

$$D_n = \begin{vmatrix} a_1 + b_1 & a_1 + b_2 & \dots & a_1 + b_n \\ a_2 + b_1 & a_2 + b_2 & \dots & a_2 + b_n \\ \vdots & \vdots & \ddots & \vdots \\ a_n + b_1 & a_n + b_2 & \dots & a_n + b_n \end{vmatrix}.$$

Solución: Este determinante, con respecto a la primera fila, se descompone en dos determinantes, cada uno de los cuales, con respecto a la segunda fila, se vuelve a descomponer en dos determinantes, y así sucesivamente. Llegando hasta la última fila, obtenemos 2^n determinantes. Para entenderlo mejor, consideremos los casos particulares donde $n = 2$ y $n = 3$. Para el primer caso tenemos que

$$\begin{aligned} D_2 &= \begin{vmatrix} a_1 & a_1 \\ a_2 + b_1 & a_2 + b_2 \end{vmatrix} + \begin{vmatrix} b_1 & b_2 \\ a_2 + b_1 & a_2 + b_2 \end{vmatrix} \\ &= \begin{vmatrix} a_1 & a_1 \\ a_2 & a_2 \end{vmatrix} + \begin{vmatrix} a_1 & a_1 \\ b_1 & b_2 \end{vmatrix} + \begin{vmatrix} b_1 & b_2 \\ a_2 & a_2 \end{vmatrix} + \begin{vmatrix} b_1 & b_2 \\ b_1 & b_2 \end{vmatrix} \end{aligned}$$

Es decir, obtenemos $2^2 = 4$ determinantes. Análogamente, para el segundo caso tenemos que

$$\begin{aligned} D_3 &= \begin{vmatrix} a_1 & a_1 & a_1 \\ a_2 + b_1 & a_2 + b_2 & a_2 + b_3 \\ a_3 + b_1 & a_3 + b_2 & a_3 + b_3 \end{vmatrix} + \begin{vmatrix} b_1 & b_2 & b_3 \\ a_2 + b_1 & a_2 + b_2 & a_2 + b_3 \\ a_3 + b_1 & a_3 + b_2 & a_3 + b_3 \end{vmatrix} \\ &= \begin{vmatrix} a_1 & a_1 & a_1 \\ a_2 & a_2 & a_2 \\ a_3 + b_1 & a_3 + b_2 & a_3 + b_3 \end{vmatrix} + \begin{vmatrix} a_1 & a_1 & a_1 \\ b_1 & b_2 & b_3 \\ a_3 + b_1 & a_3 + b_2 & a_3 + b_3 \end{vmatrix} \\ &\quad + \begin{vmatrix} b_1 & b_2 & b_3 \\ a_2 & a_2 & a_2 \\ a_3 + b_1 & a_3 + b_2 & a_3 + b_3 \end{vmatrix} + \begin{vmatrix} b_1 & b_2 & b_3 \\ b_1 & b_2 & b_3 \\ a_3 + b_1 & a_3 + b_2 & a_3 + b_3 \end{vmatrix} \\ &= \begin{vmatrix} a_1 & a_1 & a_1 \\ a_2 & a_2 & a_2 \\ a_3 & a_3 & a_3 \end{vmatrix} + \begin{vmatrix} a_1 & a_1 & a_1 \\ a_2 & a_2 & a_2 \\ b_1 & b_2 & b_3 \end{vmatrix} + \begin{vmatrix} a_1 & a_1 & a_1 \\ b_1 & b_2 & b_3 \\ a_3 & a_3 & a_3 \end{vmatrix} + \begin{vmatrix} a_1 & a_1 & a_1 \\ b_1 & b_2 & b_3 \\ b_1 & b_2 & b_3 \end{vmatrix} \\ &\quad + \begin{vmatrix} b_1 & b_2 & b_3 \\ a_2 & a_2 & a_2 \\ a_3 & a_3 & a_3 \end{vmatrix} + \begin{vmatrix} b_1 & b_2 & b_3 \\ a_2 & a_2 & a_2 \\ b_1 & b_2 & b_3 \end{vmatrix} + \begin{vmatrix} b_1 & b_2 & b_3 \\ b_1 & b_2 & b_3 \\ a_3 & a_3 & a_3 \end{vmatrix} + \begin{vmatrix} b_1 & b_2 & b_3 \\ b_1 & b_2 & b_3 \\ b_1 & b_2 & b_3 \end{vmatrix} \end{aligned}$$

Es decir, obtenemos $2^3 = 8$ determinantes. Si en cada descomposición los primeros términos son números a_i , y los segundos b_j , entonces las filas resultantes serán del tipo a_i, a_i, \dots, a_i o bien del tipo b_1, b_2, \dots, b_n . Dos filas del primer tipo son proporcionales, al igual que dos filas del segundo tipo. Para $n > 2$, en cada determinante obtenido aparecerán al menos dos filas del mismo tipo, y por lo tanto será nulo. Así, $D_n = 0$ para $n > 2$. Además,

$$D_1 = a_1 + b_1, \quad D_2 = \begin{vmatrix} a_1 & a_1 \\ b_1 & b_2 \end{vmatrix} + \begin{vmatrix} b_1 & b_2 \\ a_2 & a_2 \end{vmatrix} = (a_1 - a_2)(b_2 - b_1).$$

Método de variación de los elementos de un determinante

Este método se aplica en aquellos casos en que, modificando todos los elementos de un determinante en la misma cantidad x , se logra una forma en la que sea fácil calcular la suma de los cofactores de todos los elementos. El método se basa en la siguiente propiedad: si a todos los elementos de un determinante D se les suma un mismo número x , el determinante aumenta en el producto de x por la suma de los cofactores de todos sus elementos. Empezaremos por un caso particular para $n = 3$ y luego lo generalizaremos. Consideremos

$$D = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}, \quad D' = \begin{vmatrix} a_{11} + x & a_{12} + x & a_{13} + x \\ a_{21} + x & a_{22} + x & a_{23} + x \\ a_{31} + x & a_{32} + x & a_{33} + x \end{vmatrix}.$$

Descomponiendo D' fila por fila, obtenemos que

$$\begin{aligned} D' &= \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} + \begin{vmatrix} x & x & x \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} + \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ x & x & x \\ a_{31} & a_{32} & a_{33} \end{vmatrix} + \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ x & x & x \end{vmatrix} \\ &\quad + \begin{vmatrix} x & x & x \\ x & x & x \\ a_{31} & a_{32} & a_{33} \end{vmatrix} + \begin{vmatrix} x & x & x \\ a_{21} & a_{22} & a_{23} \\ x & x & x \end{vmatrix} + \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ x & x & x \\ x & x & x \end{vmatrix} + \begin{vmatrix} x & x & x \\ x & x & x \\ x & x & x \end{vmatrix} \\ &= D + \begin{vmatrix} x & x & x \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} + \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ x & x & x \\ a_{31} & a_{32} & a_{33} \end{vmatrix} + \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ x & x & x \end{vmatrix} \end{aligned}$$

Ahora calculamos cada uno de esos tres determinantes usando el desarrollo por la fila que contiene los x . Es decir,

$$\begin{aligned} \begin{vmatrix} x & x & x \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} &= x \sum_{j=1}^3 C_{1j}, \quad \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ x & x & x \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = x \sum_{j=1}^3 C_{2j}, \\ \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ x & x & x \end{vmatrix} &= x \sum_{j=1}^3 C_{3j}. \end{aligned}$$

Por lo tanto,

$$D' = D + x \left(\sum_{j=1}^3 C_{1j} + \sum_{j=1}^3 C_{2j} + \sum_{j=1}^3 C_{3j} \right) = D + x \sum_{i=1}^3 \sum_{j=1}^3 C_{ij}.$$

Ahora bien, sea

$$D = \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix}, \quad D' = \begin{vmatrix} a_{11} + x & a_{12} + x & \dots & a_{1n} + x \\ a_{21} + x & a_{22} + x & \dots & a_{2n} + x \\ \vdots & & \ddots & \vdots \\ a_{n1} + x & a_{n2} + x & \dots & a_{nn} + x \end{vmatrix}.$$

Descompongamos D' en dos determinantes con respecto a la primera fila, cada uno de ellos en dos determinantes con respecto a la segunda fila, etc.

Los sumandos que contienen más de una fila de elementos, iguales a x , son iguales a 0. Los sumandos con una fila de elementos iguales a x , los descomponemos por esa fila. Entonces obtenemos

$$D' = D + x \sum_{i,j=1}^n C_{ij}$$

lo que se requería. Así pues, el cálculo del determinante D' se reduce a la deducción del determinante D y de la suma de sus cofactores.

Ejemplo 3.4.6: Calcular el determinante D_n del ejemplo 3.4.3.

Solución: Después de restar el número x de todos sus elementos, obtenemos

$$D = \begin{vmatrix} a_1 - x & 0 & 0 & \cdots & 0 \\ 0 & a_2 - x & 0 & \cdots & 0 \\ 0 & 0 & a_3 - x & \cdots & 0 \\ \vdots & & & \ddots & \\ 0 & 0 & 0 & \cdots & a_n - x \end{vmatrix}.$$

Los cofactores de los elementos D , no yacentes en la diagonal principal, son nulos, y de cada elemento en la diagonal principal son iguales al producto de los demás elementos de la diagonal principal. Por eso

$$\begin{aligned} D_n &= (a_1 - x) \cdots (a_n - x) + x \sum_{i=1}^n (a_1 - x) \cdots (a_{i-1} - x)(a_{i+1} - x) \cdots (a_n - x) \\ &= x(a_1 - x)(a_2 - x) \cdots (a_n - x) \left(\frac{1}{x} + \frac{1}{a_1 - x} + \cdots + \frac{1}{a_n - x} \right) \end{aligned}$$

3.5 Algunas aplicaciones del determinante

En esta sección, analizaremos cómo se pueden emplear los determinantes para resolver sistemas de ecuaciones lineales mediante la regla de Cramer, para estudiar la independencia lineal de funciones, y para calcular áreas y volúmenes en el espacio.

Regla de Cramer

Nuestro siguiente teorema utiliza la fórmula para la inversa de una matriz invertible para producir una fórmula, llamada *regla de Cramer*, para la solución de un sistema lineal $A\mathbf{x} = \mathbf{b}$ de n ecuaciones con n incógnitas en el caso en que la matriz de coeficientes A sea invertible (o, equivalentemente, cuando $\det(A) \neq 0$).

Teorema 3.5.1: Si $A\mathbf{x} = \mathbf{b}$ es un sistema de n ecuaciones lineales en n incógnitas tal que $\det(A) \neq 0$, entonces el sistema tiene una solución única. Esta solución es

$$x_1 = \frac{\det(A_1)}{\det(A)}, \quad x_2 = \frac{\det(A_2)}{\det(A)}, \quad \dots, \quad x_n = \frac{\det(A_n)}{\det(A)}$$

donde A_j es la matriz obtenida reemplazando las entradas en la j -ésima columna de A por las entradas en la matriz

$$\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}.$$

Demostración: Si $\det(A) \neq 0$, entonces A es invertible, y por el teorema 2.6.8, $\mathbf{x} = A^{-1}\mathbf{b}$ es la solución única de $A\mathbf{x} = \mathbf{b}$. Por lo tanto, por el teorema 3.3.5 tenemos que

$$\mathbf{x} = A^{-1}\mathbf{b} = \frac{1}{\det(A)} \operatorname{adj}(A)\mathbf{b} = \frac{1}{\det(A)} \begin{bmatrix} C_{11} & C_{21} & \cdots & C_{n1} \\ C_{12} & C_{22} & \cdots & C_{n2} \\ \vdots & & \ddots & \vdots \\ C_{1n} & C_{2n} & \cdots & C_{nn} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}.$$

Multiplicando las matrices da como resultado

$$\mathbb{x} = \frac{1}{\det(A)} \begin{bmatrix} b_1 C_{11} + b_2 C_{21} + \cdots + b_n C_{n1} \\ b_1 C_{12} + b_2 C_{22} + \cdots + b_n C_{n2} \\ \vdots \\ b_1 C_{1n} + b_2 C_{2n} + \cdots + b_n C_{nn} \end{bmatrix}.$$

La entrada en la j -ésima fila de \mathbb{x} es por lo tanto

$$x_j = \frac{b_1 C_{1j} + b_2 C_{2j} + \cdots + b_n C_{nj}}{\det(A)}. \quad (3.5.1)$$

Ahora sea

$$A_j = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1j-1} & b_1 & a_{1j+1} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2j-1} & b_2 & a_{2j+1} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nj-1} & b_n & a_{nj+1} & \cdots & a_{nn} \end{bmatrix}.$$

Dado que A_j difiere de A solo en la j -ésima columna, se deduce que los cofactores de las entradas b_1, b_2, \dots, b_n en A_j son los mismos que los cofactores de las entradas correspondientes en la j -ésima columna de A . La expansión de cofactores de $\det(A_j)$ a lo largo de la j -ésima columna es por lo tanto

$$\det(A_j) = b_1 C_{1j} + b_2 C_{2j} + \cdots + b_n C_{nj}.$$

Sustituyendo este resultado en (3.5.1) da como resultado

$$x_j = \frac{\det(A_j)}{\det(A)}.$$

Ejemplo 3.5.1: Usa la regla de Cramer para resolver

$$\begin{aligned} x_1 + 2x_3 &= 6 \\ -3x_1 + 4x_2 + 6x_3 &= 30 \\ -x_1 - 2x_2 + x_3 &= 8 \end{aligned}$$

Solución: Consideremos

$$\begin{aligned} A &= \begin{bmatrix} 1 & 0 & 2 \\ -3 & 4 & 6 \\ -1 & -2 & 3 \end{bmatrix} & A_1 &= \begin{bmatrix} 6 & 0 & 2 \\ 30 & 4 & 6 \\ 8 & -2 & 3 \end{bmatrix} \\ A_2 &= \begin{bmatrix} 1 & 6 & 2 \\ -3 & 30 & 6 \\ -1 & 8 & 3 \end{bmatrix} & A_3 &= \begin{bmatrix} 1 & 0 & 6 \\ -3 & 4 & 30 \\ -1 & -2 & 8 \end{bmatrix} \end{aligned}$$

Por lo tanto,

$$\begin{aligned} x_1 &= \frac{\det(A_1)}{\det(A)} = \frac{-40}{44} = -\frac{10}{11}, & x_2 &= \frac{\det(A_2)}{\det(A)} = \frac{72}{44} = \frac{18}{11}, \\ x_3 &= \frac{\det(A_3)}{\det(A)} = \frac{152}{44} = \frac{38}{11}. \end{aligned}$$

Para $n > 3$, generalmente es más eficiente resolver un sistema lineal $A\mathbb{x} = \mathbb{b}$ de n ecuaciones con n incógnitas por eliminación de Gauss-Jordan que por la regla de Cramer. Su uso principal es para obtener propiedades de soluciones de un sistema lineal sin resolver realmente el sistema, como determinar si existe una única solución, infinitas soluciones o si el sistema es inconsistente.

Independencia lineal de funciones

A veces, la dependencia lineal de funciones puede deducirse de identidades conocidas. Por ejemplo, las funciones

$$f_1 = \sin^2 x, \quad f_2 = \cos^2 x, \quad f_3 = 5$$

forman un conjunto linealmente dependiente en $F(-\infty, \infty)$, ya que la ecuación

$$5f_1 + 5f_2 - f_3 = 5\sin^2 x + 5\cos^2 x - 5 = 5(\sin^2 x + \cos^2 x) - 5 = 0$$

expresa 0 como una combinación lineal de f_1, f_2, f_3 con coeficientes que no son todos cero.

Sin embargo, es relativamente raro que la independencia o dependencia lineal de funciones pueda determinarse por métodos algebraicos o trigonométricos. Para empeorar las cosas, no existe un método general para hacerlo. Dicho esto, existe un teorema que puede ser útil para ese propósito en ciertos casos. La siguiente definición es necesaria para ese teorema.

Definición 3.5.1:

Si $f_1 = f_1(x), f_2 = f_2(x), \dots, f_n = f_n(x)$ son funciones que son $n - 1$ veces diferenciables en el intervalo $(-\infty, \infty)$, entonces el determinante

$$W(x) = \begin{vmatrix} f_1(x) & f_2(x) & \cdots & f_n(x) \\ f_1'(x) & f_2'(x) & \cdots & f_n'(x) \\ \vdots & \vdots & & \vdots \\ f_1^{(n-1)}(x) & f_2^{(n-1)}(x) & \cdots & f_n^{(n-1)}(x) \end{vmatrix}$$

se llama el *Wronskiano* de f_1, f_2, \dots, f_n .

Supongamos por el momento que $f_1 = f_1(x), f_2 = f_2(x), \dots, f_n = f_n(x)$ son vectores linealmente dependientes en $C^{n-1}(-\infty, \infty)$. Esto implica que la ecuación vectorial

$$k_1 f_1 + k_2 f_2 + \cdots + k_n f_n = 0$$

se satisface por valores de los coeficientes k_1, k_2, \dots, k_n que no son todos cero, y para estos coeficientes la ecuación

$$k_1 f_1(x) + k_2 f_2(x) + \cdots + k_n f_n(x) = 0$$

se satisface para todo $x \in (-\infty, \infty)$. Usando esta ecuación junto con aquellas que resultan al diferenciarla $n - 1$ veces, obtenemos el sistema lineal

$$\begin{aligned} k_1 f_1(x) + k_2 f_2(x) + \cdots + k_n f_n(x) &= 0 \\ k_1 f_1'(x) + k_2 f_2'(x) + \cdots + k_n f_n'(x) &= 0 \\ &\vdots \\ k_1 f_1^{(n-1)}(x) + k_2 f_2^{(n-1)}(x) + \cdots + k_n f_n^{(n-1)}(x) &= 0 \end{aligned}$$

Así, la dependencia lineal de f_1, f_2, \dots, f_n implica que el sistema lineal

$$\begin{bmatrix} f_1(x) & f_2(x) & \cdots & f_n(x) \\ f_1'(x) & f_2'(x) & \cdots & f_n'(x) \\ \vdots & \vdots & & \vdots \\ f_1^{(n-1)}(x) & f_2^{(n-1)}(x) & \cdots & f_n^{(n-1)}(x) \end{bmatrix} \begin{bmatrix} k_1 \\ k_2 \\ \vdots \\ k_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (3.5.2)$$

tiene una solución no trivial para todo x en el intervalo $(-\infty, \infty)$, y esto a su vez implica que el determinante de la matriz de coeficientes de (3.5.2) es cero para todo x . Dado que este determinante es el Wronskiano de f_1, f_2, \dots, f_n , hemos establecido el siguiente resultado.

Teorema 3.5.2: Si las funciones f_1, f_2, \dots, f_n tienen $n - 1$ derivadas continuas en el intervalo $(-\infty, \infty)$, y si el Wronskiano de estas funciones no es idénticamente cero en $(-\infty, \infty)$, entonces estas funciones forman un conjunto linealmente independiente de vectores en $C^{n-1}(-\infty, \infty)$.

En el ejemplo 1.4.2 mostramos que x y $\sin x$ son funciones linealmente independientes observando que ninguna es un múltiplo escalar de la otra. El siguiente ejemplo ilustra cómo obtener el mismo resultado usando el Wronskiano (aunque es un procedimiento más complicado en este caso particular).

Ejemplo 3.5.2: Usa el Wronskiano para mostrar que $f_1 = x$ y $f_2 = \sin x$ son vectores linealmente independientes en $C^\infty(-\infty, \infty)$.

Solución: El Wronskiano es

$$W(x) = \begin{vmatrix} x & \sin x \\ 1 & \cos x \end{vmatrix} = x \cos x - \sin x.$$

Esta función no es idénticamente cero en $(-\infty, \infty)$ ya que, por ejemplo,

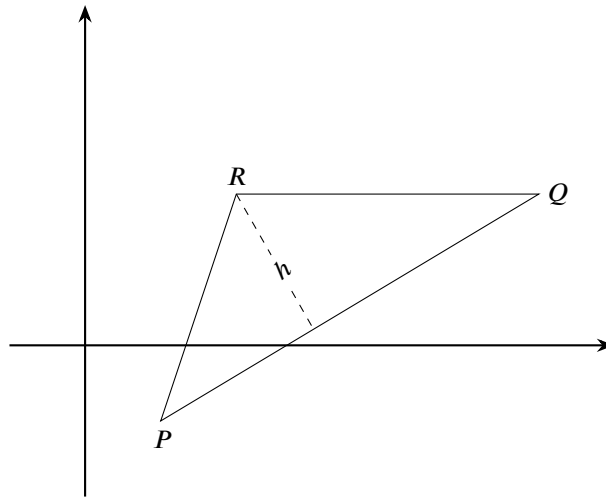
$$W\left(\frac{\pi}{2}\right) = \frac{\pi}{2} \cos\left(\frac{\pi}{2}\right) - \sin\left(\frac{\pi}{2}\right) = -1.$$

Por lo tanto, las funciones son linealmente independientes.

Área de un triángulo en el plano

Consideremos un triángulo cuyos vértices son $P = (x_1, x_2)$, $Q = (x_2, y_2)$ y $R = (x_3, y_3)$, es decir:

Figura 3.1 Triángulo formado por los puntos P , Q y R



La altura h de dicho triángulo es la distancia del punto R a la recta que pasa por los puntos P y Q . La ecuación de esta recta viene dada por

$$\frac{y - y_1}{x - x_1} = \frac{y_2 - y_1}{x_2 - x_1},$$

la cual puede escribirse como

$$(y_1 - y_2)x + (x_2 - x_1)y + (x_1y_2 - x_2y_1) = 0,$$

es decir, como

$$Ax + By + C = 0,$$

donde $A = y_1 - y_2$, $B = x_2 - x_1$ y $C = x_1y_2 - x_2y_1$.

Del primer curso de Geometría Analítica, sabemos que

$$h = \frac{|Ax_3 + By_3 + C|}{\sqrt{A^2 + B^2}}.$$

Puesto que el área de un triángulo viene dada por la fórmula

$$S = \frac{1}{2}bh,$$

donde b es la distancia de P a Q , es decir,

$$\begin{aligned} b &= \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \\ &= \sqrt{(y_1 - y_2)^2 + (x_2 - x_1)^2} \\ &= \sqrt{A^2 + B^2} \end{aligned}$$

entonces

$$\begin{aligned} S &= \frac{1}{2} \left(\sqrt{A^2 + B^2} \right) \left(\frac{|Ax_3 + By_3 + C|}{\sqrt{A^2 + B^2}} \right) \\ &= \frac{1}{2} |Ax_3 + By_3 + C| \\ &= \frac{1}{2} |x_1y_2 + x_2y_3 + x_3y_1 - x_3y_2 - x_2y_1 - x_1y_3| \end{aligned}$$

❶ Nota

Nótese que intercambiando cualesquiera dos renglones de la matriz, no se modifica el valor del área, pues el determinante solo cambia de signo.

Por tanto,

$$S = \frac{1}{2} \text{abs} \begin{pmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{pmatrix},$$

donde abs significa valor absoluto.

Ejemplo 3.5.3: Calcule el área del triángulo cuyos vértices son:

$$P = (-5, 0), \quad Q = (7, 0), \quad R = (0, 6).$$

Solución: El dibujo del triángulo en el plano cartesiano es como sigue:

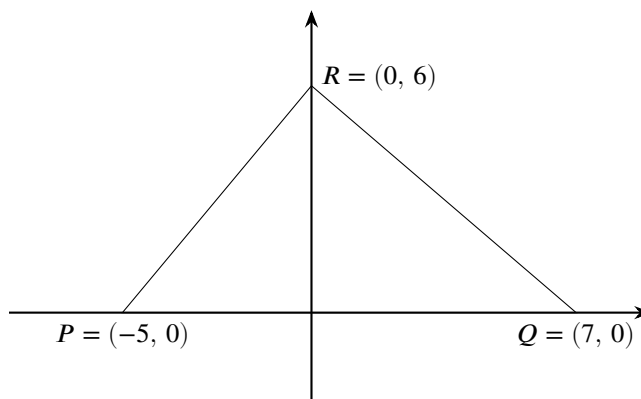


Figura 3.2 Triángulo formado por los puntos P , Q y R

El área del triángulo es:

$$S = \frac{1}{2} \text{abs} \begin{pmatrix} -5 & 0 & 1 \\ 7 & 0 & 1 \\ 0 & 6 & 1 \end{pmatrix} = \frac{1}{2}(72) = 36 \text{ unidades}^2.$$

Ejemplo 3.5.4: Calcule el área del polígono cuyos vértices son:

$$A = (0, 0), \quad B = (2, 4), \quad C = (7, 2), \quad D = (5, -3), \quad E = (3, -5).$$

Solución: Notemos que podemos dividir el área total del polígono en tres triángulos, el primero formado por los puntos A, B, C ; el segundo formado por los puntos A, C, D ; y el tercero formado por los puntos A, D, E . Si S es el área del polígono, entonces dicha área es la suma de las áreas de los triángulos ya mencionados. Dibujando el polígono en el plano cartesiano y triangulando apropiadamente, obtenemos que

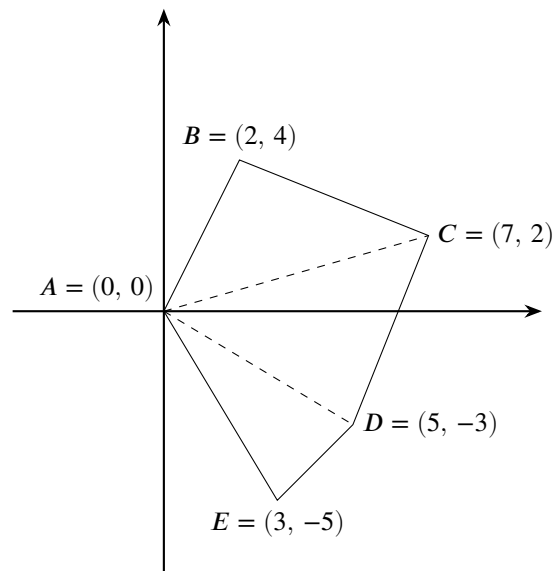


Figura 3.3 Polígono formado por los puntos A, B, C, D y E

El área del polígono es:

$$\begin{aligned} S &= \frac{1}{2} \text{abs} \begin{vmatrix} 0 & 0 & 1 \\ 2 & 4 & 1 \\ 7 & 2 & 1 \end{vmatrix} + \frac{1}{2} \text{abs} \begin{vmatrix} 0 & 0 & 1 \\ 7 & 2 & 1 \\ 5 & -3 & 1 \end{vmatrix} + \frac{1}{2} \text{abs} \begin{vmatrix} 0 & 0 & 1 \\ 5 & -3 & 1 \\ 3 & -5 & 1 \end{vmatrix} \\ &= \frac{1}{2} (24 + 31 + 16) \\ &= \frac{71}{2} \text{ unidades}^2 \end{aligned}$$

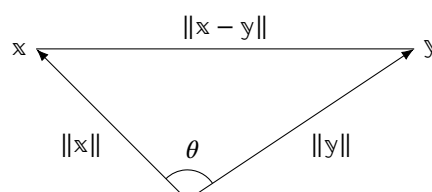
Área de un triángulo en el espacio

Consideremos los vectores

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}$$

como se muestran en la siguiente figura:

Figura 3.4 Gráfica de los vectores \mathbf{x} e \mathbf{y} formando un triángulo



De acuerdo con la ley de cosenos tenemos que

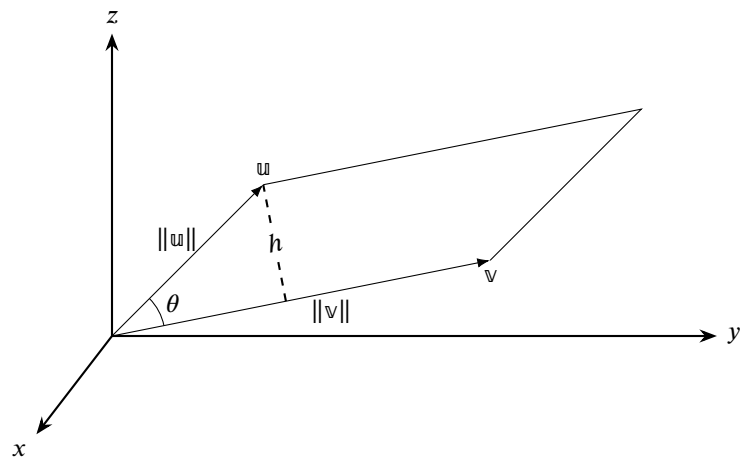
$$\begin{aligned}\cos \theta &= \frac{\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 - \|\mathbf{y} - \mathbf{x}\|^2}{2\|\mathbf{x}\|\|\mathbf{y}\|} \\ &= \frac{2x_1y_1 + 2x_2y_2 + 2x_3y_3}{2\|\mathbf{x}\|\|\mathbf{y}\|} \\ &= \frac{x_1y_1 + x_2y_2 + x_3y_3}{\|\mathbf{x}\|\|\mathbf{y}\|}\end{aligned}$$

Ahora consideremos el paralelogramo formado por los vectores

$$\mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix}, \quad \mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}$$

como se muestran en la siguiente figura:

Figura 3.5 Paralelogramo generado por los vectores \mathbf{u} y \mathbf{v}



El área S de dicho paralelogramo viene dada por

$$S = \|\mathbf{v}\|h$$

donde

$$h = \|\mathbf{u}\| \sin \theta.$$

Puesto que

$$\sin^2 \theta = 1 - \cos^2 \theta,$$

entonces

$$\begin{aligned}S^2 &= \|\mathbf{v}\|^2 h^2 \\ &= \|\mathbf{v}\|^2 \|\mathbf{u}\|^2 \sin^2 \theta \\ &= \|\mathbf{v}\|^2 \|\mathbf{u}\|^2 [1 - \cos^2 \theta] \\ &= \|\mathbf{v}\|^2 \|\mathbf{u}\|^2 \left[1 - \left(\frac{u_1v_1 + u_2v_2 + u_3v_3}{\|\mathbf{u}\|\|\mathbf{v}\|} \right)^2 \right] \\ &= \|\mathbf{v}\|^2 \|\mathbf{u}\|^2 \left[\frac{\|\mathbf{u}\|^2 \|\mathbf{v}\|^2 - (u_1v_1 + u_2v_2 + u_3v_3)^2}{\|\mathbf{u}\|^2 \|\mathbf{v}\|^2} \right] \\ &= (u_1^2 + u_2^2 + u_3^2)(v_1^2 + v_2^2 + v_3^2) - (u_1v_1 + u_2v_2 + u_3v_3)^2 \\ &= (u_2v_3 - u_3v_2)^2 + (u_1v_3 - u_3v_1)^2 + (u_1v_2 - u_2v_1)^2\end{aligned}$$

Por tanto

$$S^2 = \begin{vmatrix} u_2 & u_3 \\ v_2 & v_3 \end{vmatrix}^2 + \begin{vmatrix} u_1 & u_3 \\ v_1 & v_3 \end{vmatrix}^2 + \begin{vmatrix} u_1 & u_2 \\ v_1 & v_2 \end{vmatrix}^2.$$

Ejemplo 3.5.5: Calcular el área S del paralelogramo generado por los vectores

$$\mathbf{u} = \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}, \quad \mathbf{v} = \begin{pmatrix} 0 \\ 5 \\ 0 \end{pmatrix}.$$

Solución: De acuerdo con la fórmula tenemos que

$$\begin{aligned} S^2 &= \begin{vmatrix} 1 & 2 \\ 5 & 0 \end{vmatrix}^2 + \begin{vmatrix} 0 & 2 \\ 0 & 0 \end{vmatrix}^2 + \begin{vmatrix} 0 & 1 \\ 0 & 5 \end{vmatrix}^2 \\ &= (-10)^2 + (0)^2 + (0)^2 \\ &= 100 \end{aligned}$$

Por lo tanto,

$$S = 10 \text{ unidades}^2.$$

El área del triángulo cuyos vértices son

$$\mathbf{0} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix}, \quad \mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}$$

es la mitad del área del paralelogramo generado por los vectores \mathbf{u} y \mathbf{v} . Consecuentemente, el área del triángulo cuyos vértices son los puntos

$$\mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix}, \quad \mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}, \quad \mathbf{w} = \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix}$$

es la mitad del área del paralelogramo generado por los vectores

$$\mathbf{v} - \mathbf{u} = \begin{pmatrix} v_1 - u_1 \\ v_2 - u_2 \\ v_3 - u_3 \end{pmatrix}, \quad \mathbf{w} - \mathbf{u} = \begin{pmatrix} w_1 - u_1 \\ w_2 - u_2 \\ w_3 - u_3 \end{pmatrix}.$$

Ejemplo 3.5.6: Calcular el área del triángulo cuyos vértices son

$$\mathbf{u} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{v} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \mathbf{w} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

Solución: El área S_t del triángulo es la mitad del área S_p del paralelogramo generado por los vectores

$$\mathbf{v} - \mathbf{u} = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \quad \mathbf{w} - \mathbf{u} = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}.$$

De acuerdo con la fórmula

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

Por lo tanto,

$$S_p = \sqrt{3},$$

y en consecuencia,

$$S_t = \frac{1}{2} S_p = \frac{\sqrt{3}}{2} \text{ unidades}^2.$$

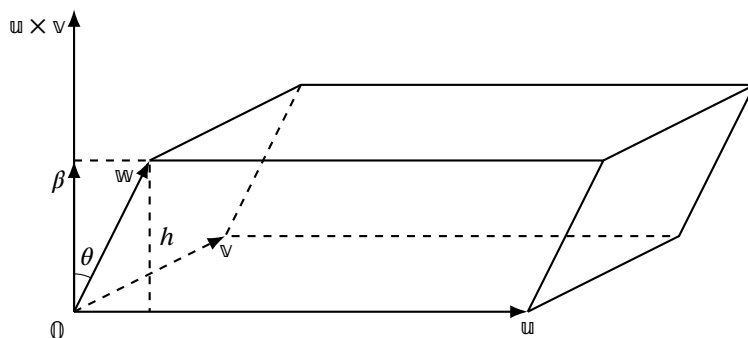
Volumen de un paralelepípedo

Ahora consideremos el paralelepípedo generado por los vectores

$$\mathfrak{u} = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix}, \quad \mathfrak{v} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}, \quad \mathfrak{w} = \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix}$$

como se muestran en la siguiente figura:

Figura 3.6 Paralelepípedo generado por los vectores \mathfrak{u} , \mathfrak{v} y \mathfrak{w}



El volumen de dicho paralelepípedo es

$$V = Sh,$$

donde S es el área de la base y está generada por los vectores \mathfrak{u} y \mathfrak{v} ; y h es la altura. De la subsección anterior, se deduce que

$$S = \sqrt{(u_2v_3 - u_3v_2)^2 + (u_1v_3 - u_3v_1)^2 + (u_1v_2 - u_2v_1)^2}.$$

Calculemos ahora h . De acuerdo con la figura precedente, se requiere un vector

$$\beta = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

perpendicular tanto a \mathfrak{u} como a \mathfrak{v} , así que por la ley de los cosenos tenemos el siguiente sistema de dos ecuaciones lineales con tres incógnitas,

$$\begin{aligned} u_1x + u_2y + u_3z &= 0 \\ v_1x + v_2y + v_3z &= 0 \end{aligned}$$

Sumando las ecuaciones que resultan de multiplicar la primera ecuación del sistema por v_3 y la segunda por $-u_3$, obtenemos

$$(u_1v_3 - u_3v_1)x + (u_2v_3 - u_3v_2)y = 0. \quad (3.5.3)$$

Análogamente, sumando las ecuaciones que resultan de multiplicar la primera ecuación del sistema por $-v_2$ y la segunda por u_2 , obtenemos

$$(u_2v_1 - u_1v_2)x + (u_2v_3 - u_3v_2)z = 0. \quad (3.5.4)$$

Finalmente, si $u_2v_3 - u_3v_2 \neq 0$, de (3.5.3) y (3.5.4) obtenemos que

$$y = \frac{u_3v_1 - u_1v_3}{u_2v_3 - u_3v_2}x$$

y

$$z = \frac{u_1v_2 - u_2v_1}{u_2v_3 - u_3v_2}x.$$

Entonces las soluciones del sistema vienen dadas por

$$\begin{aligned}x &= k(u_2v_3 - u_3v_2) \\y &= k(u_3v_1 - u_1v_3) \\z &= k(u_1v_2 - u_2v_1)\end{aligned}$$

es decir, las soluciones son de la forma

$$\beta = k \begin{pmatrix} u_2v_3 - u_3v_2 \\ u_3v_1 - u_1v_3 \\ u_1v_2 - u_2v_1 \end{pmatrix}.$$

Cuando $k = 1$ la solución del sistema se define como el producto vectorial de \mathfrak{u} y \mathfrak{v} , por tanto

$$\mathfrak{u} \times \mathfrak{v} = \begin{pmatrix} u_2v_3 - u_3v_2 \\ u_3v_1 - u_1v_3 \\ u_1v_2 - u_2v_1 \end{pmatrix}.$$

Ahora queremos un vector α en la dirección de $\mathfrak{u} \times \mathfrak{v}$, esto es $\alpha = \lambda(\mathfrak{u} \times \mathfrak{v})$, tal que $h = \|\alpha\|$. De acuerdo con la figura precedente

$$h = \|\mathfrak{w}\| \cos \theta$$

donde, por la ley de los cosenos, tenemos que

$$\begin{aligned}\cos \theta &= \frac{\|\alpha\|^2 + \|\mathfrak{w}\|^2 - \|\alpha - \mathfrak{w}\|^2}{2\|\alpha\|\|\mathfrak{w}\|} \\&= \frac{2\lambda[(u_2v_3 - u_3v_2)w_1 + (u_3v_1 - u_1v_3)w_2 + (u_1v_2 - u_2v_1)w_3]}{2\|\lambda(\mathfrak{u} \times \mathfrak{v})\|\|\mathfrak{w}\|} \\&= \frac{\pm[u_2v_3w_1 - u_3v_2w_1 + u_3v_1w_2 - u_1v_3w_2 + u_1v_2w_3 - u_2v_1w_3]}{\|\mathfrak{u} \times \mathfrak{v}\|\|\mathfrak{w}\|}\end{aligned}$$

En consecuencia, observando que $S = \|\mathfrak{u} \times \mathfrak{v}\|$, tenemos que

$$\begin{aligned}V &= Sh \\&= \|\mathfrak{u} \times \mathfrak{v}\|\|\mathfrak{w}\| \cos \theta \\&= \|\mathfrak{u} \times \mathfrak{v}\|\|\mathfrak{w}\| \frac{\pm[u_2v_3w_1 - u_3v_2w_1 + u_3v_1w_2 - u_1v_3w_2 + u_1v_2w_3 - u_2v_1w_3]}{\|\mathfrak{u} \times \mathfrak{v}\|\|\mathfrak{w}\|} \\&= \pm[u_2v_3w_1 - u_3v_2w_1 + u_3v_1w_2 - u_1v_3w_2 + u_1v_2w_3 - u_2v_1w_3] \\&= \text{abs}[u_2v_3w_1 - u_3v_2w_1 + u_3v_1w_2 - u_1v_3w_2 + u_1v_2w_3 - u_2v_1w_3] \\&= \text{abs}[u_1v_2w_3 + u_2v_3w_1 + u_3v_1w_2 - u_3v_2w_1 - u_2v_1w_3 - u_1v_3w_2]\end{aligned}$$

Por consiguiente, el volumen del paralelepípedo formado por los vectores \mathfrak{u} , \mathfrak{v} y \mathfrak{w} , está dado por

$$V = \text{abs} \begin{pmatrix} u_1 & v_1 & w_1 \\ u_2 & v_2 & w_2 \\ u_3 & v_3 & w_3 \end{pmatrix}.$$

Ejemplo 3.5.7: Consideremos el paralelepípedo generado por los vectores

$$\mathfrak{x}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \mathfrak{x}_2 = \begin{pmatrix} 0 \\ 2 \\ 0 \end{pmatrix}, \quad \mathfrak{x}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

De acuerdo con la formula, el volumen del paralelepípedo es:

$$V = \text{abs} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix} = 2 \text{ unidades}^3.$$

4

Capítulo

TRANSFORMACIONES LINEALES

En los capítulos precedentes, hemos explorado los conceptos fundamentales de los espacios vectoriales, que nos han proporcionado el marco abstracto para operar con vectores. Posteriormente, estudiamos las matrices como arreglos numéricos que facilitan la resolución de sistemas de ecuaciones lineales y, a su vez, analizamos los determinantes como una herramienta para comprender propiedades clave de las matrices cuadradas, como la invertibilidad.

Ahora, estamos en una posición ideal para unificar estas ideas bajo un concepto más general y profundo: las transformaciones lineales. En este capítulo, dejaremos de ver las matrices como simples arreglos de números y comenzaremos a entenderlas como operadores que actúan sobre vectores para transformarlos en otros vectores. Esta “acción” no es arbitraria; sigue reglas muy precisas que preservan la estructura subyacente del espacio vectorial.

Una transformación lineal es, en esencia, una función entre dos espacios vectoriales que respeta las operaciones de suma de vectores y multiplicación por un escalar. Esta propiedad de preservar la estructura es lo que las convierte en un pilar del álgebra lineal y de sus innumerables aplicaciones. Descubriremos que toda matriz puede definir una transformación lineal y, de manera recíproca, toda transformación lineal entre espacios vectoriales de dimensión finita puede ser representada por una matriz.

Este capítulo servirá como un puente entre el álgebra matricial concreta que hemos estudiado y la geometría de los espacios vectoriales. Exploraremos conceptos cruciales como el núcleo y la imagen de una transformación, que generalizan las ideas de espacio nulo y espacio columna de una matriz.

4.1 Definición y ejemplos

Definición 4.1.1:

Si $T: V \rightarrow W$ es una función de un espacio vectorial V a un espacio vectorial W , entonces T se llama *transformación lineal* de V a W si las siguientes dos propiedades se cumplen para cada $\mathbf{u}, \mathbf{v} \in V$ y para todos los escalares k :

- i) $T(\mathbf{u} + \mathbf{v}) = T\mathbf{u} + T\mathbf{v}$.
- ii) $T(k\mathbf{u}) = kT\mathbf{u}$.

Nota

Se escriben indistintamente $T\mathbf{v}$ y $T(\mathbf{v})$. Denotan lo mismo; las dos se leen “ T de \mathbf{v} ”. Si T tiene como dominio y codominio a \mathbb{R}^n , lo llamaremos *operador en \mathbb{R}^n* .

Existen muchas formas de transformar los espacios \mathbb{R}^2 y \mathbb{R}^3 . Por ejemplo, rotaciones alrededor del origen, reflexiones sobre rectas y planos que pasan por el origen, y proyecciones sobre rectas y planos a través del origen. Los siguientes ejemplos mostrarán uno de los operadores más importantes: *operadores de reflexión*.

Ejemplo 4.1.1:

Consideremos la siguiente función

$$T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$$

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto T \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ -y \end{pmatrix}$$

Geométricamente T toma un vector en \mathbb{R}^2 y lo refleja respecto al eje x como se muestra en la figura 4.1, así se comprueba que T es función. Ahora verifiquemos que T es un operador en \mathbb{R}^2 :

- i) Sea $\mathbf{u}, \mathbf{v} \in \mathbb{R}^2$ con $\mathbf{u} = \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}$, $\mathbf{v} = \begin{pmatrix} x_2 \\ y_2 \end{pmatrix}$. Entonces

$$\begin{aligned} T(\mathbf{u} + \mathbf{v}) &= T \left(\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} + \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} \right) \\ &= T \begin{pmatrix} x_1 + x_2 \\ y_1 + y_2 \end{pmatrix} \\ &= \begin{pmatrix} x_1 + x_2 \\ -(y_1 + y_2) \end{pmatrix} \\ &= \begin{pmatrix} x_1 + x_2 \\ -y_1 + (-y_2) \end{pmatrix} \\ &= \begin{pmatrix} x_1 \\ -y_1 \end{pmatrix} + \begin{pmatrix} x_2 \\ -y_2 \end{pmatrix} \\ &= T\mathbf{u} + T\mathbf{v} \end{aligned}$$

Por tanto $T(\mathbf{u} + \mathbf{v}) = T\mathbf{u} + T\mathbf{v}$.

- ii) Sea $\mathbf{u} \in \mathbb{R}^2$ con $\mathbf{u} = \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}$ y $\alpha \in \mathbb{R}$. Entonces

$$\begin{aligned} T(\alpha\mathbf{u}) &= T \left(\alpha \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} \right) \\ &= T \begin{pmatrix} \alpha x_1 \\ \alpha y_1 \end{pmatrix} \\ &= \begin{pmatrix} \alpha x_1 \\ -\alpha y_1 \end{pmatrix} \\ &= \alpha \begin{pmatrix} x_1 \\ -y_1 \end{pmatrix} \\ &= \alpha T\mathbf{u} \end{aligned}$$

Por tanto, $T(\alpha\mathbf{u}) = \alpha T\mathbf{u}$.

Por tanto, T es un operador en \mathbb{R}^2 .

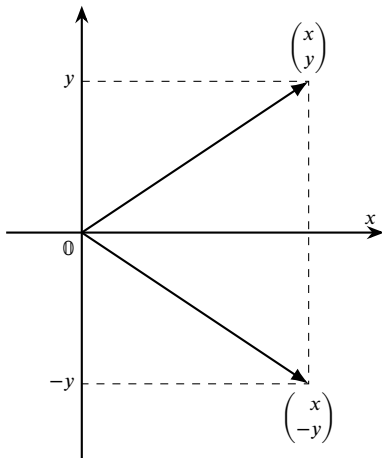


Figura 4.1

Ejemplo 4.1.2:

Consideremos la siguiente función

$$T: \mathbb{R}^2 \longrightarrow \mathbb{R}^2$$

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -x \\ y \end{pmatrix}$$

Geométricamente T toma un vector en \mathbb{R}^2 y lo refleja respecto al eje y como se muestra en la figura 4.2, así se comprueba que T es función. Ahora verifiquemos que T es un operador en \mathbb{R}^2 :

i) Sea $\mathfrak{u}, \mathfrak{v} \in \mathbb{R}^2$ con $\mathfrak{u} = \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}$, $\mathfrak{v} = \begin{pmatrix} x_2 \\ y_2 \end{pmatrix}$. Entonces

$$\begin{aligned} T(\mathfrak{u} + \mathfrak{v}) &= T\left(\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} + \begin{pmatrix} x_2 \\ y_2 \end{pmatrix}\right) \\ &= T\begin{pmatrix} x_1 + x_2 \\ y_1 + y_2 \end{pmatrix} \\ &= \begin{pmatrix} -(x_1 + x_2) \\ y_1 + y_2 \end{pmatrix} \\ &= \begin{pmatrix} -x_1 + (-x_2) \\ y_1 + y_2 \end{pmatrix} \\ &= \begin{pmatrix} -x_1 \\ y_1 \end{pmatrix} + \begin{pmatrix} -x_2 \\ y_2 \end{pmatrix} \\ &= T\mathfrak{u} + T\mathfrak{v} \end{aligned}$$

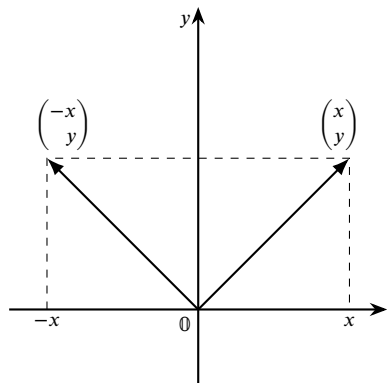
Por tanto $T(\mathfrak{u} + \mathfrak{v}) = T\mathfrak{u} + T\mathfrak{v}$.

ii) Sea $\mathfrak{u} \in \mathbb{R}^2$ con $\mathfrak{u} = \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}$ y $\alpha \in \mathbb{R}$. Entonces

$$\begin{aligned} T(\alpha\mathfrak{u}) &= T\left(\alpha \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}\right) \\ &= T\begin{pmatrix} \alpha x_1 \\ \alpha y_1 \end{pmatrix} \\ &= \begin{pmatrix} -\alpha x_1 \\ \alpha y_1 \end{pmatrix} \\ &= \alpha \begin{pmatrix} -x_1 \\ y_1 \end{pmatrix} \\ &= \alpha T\mathfrak{u} \end{aligned}$$

Por tanto, $T(\alpha\mathfrak{u}) = \alpha T\mathfrak{u}$.

Por tanto, T es un operador en \mathbb{R}^2 .

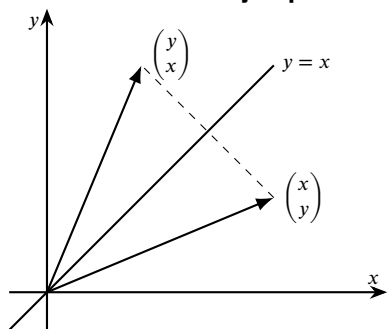
**Figura 4.2****Ejemplo 4.1.3:**

Consideremos la siguiente función

$$T: \mathbb{R}^2 \longrightarrow \mathbb{R}^2$$

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} y \\ x \end{pmatrix}$$

Este operador intercambia las coordenadas de cada vector del plano, de modo que la primera pasa a ser la segunda y viceversa. A nivel geométrico, puede interpretarse como una simetría respecto a la recta $y = x$ como se muestra en la figura 4.3. Dicho de otra manera, T actúa como un espejo cuya “línea de simetría” es precisamente $y = x$, de modo que todo vector que se encuentre sobre esta recta permanece inalterado, mientras que los vectores situados fuera de ella cambian de posición de forma que su distancia a la recta se conserva. Se deja al lector demostrar que T es un operador en \mathbb{R}^2 .

**Figura 4.3**

5

Capítulo

ESPACIOS VECTORIALES CON PRODUCTO INTERNO

En los capítulos anteriores hemos desarrollado una base sólida para comprender la estructura de los espacios vectoriales y sus aplicaciones. Comenzamos con el estudio de los espacios vectoriales reales, donde se establecieron los principios fundamentales que rigen la combinación lineal, la independencia y las bases. Posteriormente, el análisis de las matrices nos permitió disponer de una herramienta algebraica para representar y manipular transformaciones, así como para organizar información de manera sistemática. La teoría de los determinantes añadió un componente crucial al dotarnos de un criterio para la invertibilidad de matrices y el cálculo de volúmenes orientados, ampliando la visión geométrica del álgebra lineal. Finalmente, las transformaciones lineales integraron todos estos conceptos, mostrando cómo las estructuras abstractas se reflejan en aplicaciones concretas, revelando la estrecha relación entre álgebra y geometría.

En este capítulo avanzamos un paso más hacia la comprensión profunda de los espacios vectoriales introduciendo la noción de producto interno. Esta operación enriquece la estructura algebraica con una dimensión geométrica adicional, permitiendo medir longitudes, calcular ángulos y definir ortogonalidad entre vectores. Gracias a ello, conceptos como la proyección ortogonal, la descomposición en subespacios y la noción de ortonormalidad cobran sentido matemático y aplicación práctica. Así, los espacios vectoriales con producto interno constituyen el puente natural entre el álgebra lineal y la geometría euclidiana.

Algunos ejemplos comunes de espacios vectoriales con producto interno incluyen espacios de funciones, donde los vectores son funciones y el producto interno puede ser definido utilizando la integral de su producto puntual, y espacios de matrices, donde los vectores son matrices y el producto interno puede ser definido de diversas maneras, como el producto de Frobenius.

5.1 Producto punto, norma y proyecciones

En este texto denotaremos la longitud de un vector \mathbf{u} por el símbolo $\|\mathbf{u}\|$, que se lee como la *norma* de \mathbf{u} , la *longitud* de \mathbf{u} o la *magnitud* de \mathbf{u} (el término “norma” siendo un sinónimo matemático común para longitud). Como se sugiere en la Figura 3.2.1a, se deduce del Teorema de Pitágoras que la norma de un vector \mathbf{u} en \mathbb{R}^2 es

$$\|\mathbf{u}\| = \sqrt{u_1^2 + u_2^2}. \quad (5.1.1)$$

De manera similar, para un vector \mathbf{u} en \mathbb{R}^3 , se deduce de la Figura 3.2.1b y dos aplicaciones del Teorema de Pitágoras que

$$\|\mathbf{u}\|^2 = (OR)^2 + (RP)^2 = (OQ)^2 + (QR)^2 + (RP)^2 = u_1^2 + u_2^2 + u_3^2$$

y de ahí que

$$\|\mathbf{u}\| = \sqrt{u_1^2 + u_2^2 + u_3^2}. \quad (5.1.2)$$

Motivados por el patrón de las fórmulas (5.1.1) y (5.1.2), hacemos la siguiente definición.

Definición 5.1.1:

Si \mathbf{u} es un vector en \mathbb{R}^n , entonces la *norma* de \mathbf{u} (también llamada la *longitud* de \mathbf{u} o la *magnitud* de \mathbf{u}) se denota por $\|\mathbf{u}\|$ y se define por la fórmula

$$\|\mathbf{u}\| = \sqrt{u_1^2 + u_2^2 + \cdots + u_n^2}.$$

Recordando la definición 2.2.1 del producto punto, si \mathbf{u} es un vector en \mathbb{R}^n , entonces la norma puede expresarse en términos del producto punto como

$$\|\mathbf{u}\| = \sqrt{\mathbf{u} \bullet \mathbf{u}}.$$

En el siguiente teorema, se presentan algunas propiedades de la norma de un vector.

Teorema 5.1.1: Si \mathbf{u} es un vector en \mathbb{R}^n , y si k es un escalar, entonces

- a) $\|\mathbf{u}\| \geq 0$.
- b) $\|\mathbf{u}\| = 0$ si y solo si $\mathbf{u} = \mathbf{0}$.
- c) $\|k\mathbf{u}\| = |k|\|\mathbf{u}\|$.

Demostración: Dado que las propiedades (a) y (b) son demostraciones directas, nos enfocaremos en demostrar la propiedad (c). Sea $\mathbf{u} \in \mathbb{R}^n$ y $k \in \mathbb{R}$, entonces

$$\begin{aligned} \|k\mathbf{u}\|^2 &= (k\mathbf{u}) \bullet (k\mathbf{u}) \\ &= k(\mathbf{u} \bullet (k\mathbf{u})) \\ &= k \cdot k(\mathbf{u} \bullet \mathbf{u}) \\ &= k^2 \|\mathbf{u}\|^2 \end{aligned}$$

De esta forma, $\|k\mathbf{u}\|^2 = k^2 \|\mathbf{u}\|^2$. Por el curso de Cálculo I, sabemos que $\sqrt{k^2} = |k|$. Así pues, se hereda que

$$\|k\mathbf{u}\| = |k| \|\mathbf{u}\|.$$

Finalmente, de la propiedad (a) de este teorema se sigue que

$$\|k\mathbf{u}\| = |k| \|\mathbf{u}\|.$$

Nota

A veces verás la fórmula (5.1.3) expresada como

$$\hat{u} = \frac{u}{\|u\|}.$$

Esta es solo una forma más compacta de escribir dicha fórmula y no pretende dar a entender que el vector u esté siendo dividido por $\|u\|$.

Un vector de norma 1 se llama *vector unitario*. Estos vectores son útiles para especificar una dirección cuando la longitud no es relevante para el problema en cuestión. Se puede obtener un vector unitario en una dirección deseada eligiendo cualquier vector no nulo u en esa dirección y multiplicándolo por el recíproco de su longitud. Es decir, si u es cualquier vector no nulo en \mathbb{R}^n , entonces

$$\hat{u} = \frac{1}{\|u\|} u \quad (5.1.3)$$

define un vector unitario que está en la misma dirección que u . Podemos confirmar que (5.1.3) es un vector unitario aplicando la parte (c) del teorema 5.1.1 con $k = \frac{1}{\|u\|}$ para obtener

$$\|\hat{u}\| = \|k u\| = |k| \|u\| = k \|u\| = \frac{1}{\|u\|} \|u\| = 1.$$

El proceso de multiplicar un vector $u \neq 0$ por el recíproco de su norma para obtener un vector unitario \hat{u} se denomina *normalización*.

Ejemplo 5.1.1: Normalice el siguiente vector:

$$u = \begin{pmatrix} -1 \\ 2 \\ -4 \\ 3 \end{pmatrix}.$$

Solución: La norma del vector u es

$$\|u\| = \sqrt{(-1)^2 + (2)^2 + (-4)^2 + (3)^2} = \sqrt{30}.$$

Por lo tanto, al normalizar el vector u , obtenemos que

$$\hat{u} = \frac{1}{\sqrt{30}} \begin{pmatrix} -1 \\ 2 \\ -4 \\ 3 \end{pmatrix}.$$

Se deja al lector confirmar que $\|\hat{u}\| = 1$.

El siguiente teorema proporciona propiedades adicionales del producto punto. Las demostraciones se dejan al lector y pueden obtenerse expresando los vectores en términos de componentes o utilizando las propiedades algebraicas establecidas en el teorema 2.2.1.

Teorema 5.1.2: Si u, v, w son vectores en \mathbb{R}^n , entonces

$$\text{i) } (u + v) \bullet w = u \bullet w + v \bullet w.$$

$$\text{ii) } u \bullet (v - w) = u \bullet v - u \bullet w.$$

$$\text{iii) } (u - v) \bullet w = u \bullet w - v \bullet w.$$

Ejemplo 5.1.2: Simplifique la siguiente expresión:

$$(u - 2v) \bullet (3u + 4v).$$

Solución: Usando las propiedades de los teoremas 2.2.1 y 5.1.2,

$$\begin{aligned} (u - 2v) \bullet (3u + 4v) &= u \bullet (3u + 4v) - 2v \bullet (3u + 4v) \\ &= 3(u \bullet u) + 4(u \bullet v) - 6(v \bullet u) - 8(v \bullet v) \\ &= 3\|u\|^2 - 2(u \bullet v) - 8\|v\|^2 \end{aligned}$$

Aplicación del producto punto a números ISBN

Aunque el sistema ha cambiado recientemente, la mayoría de los libros antiguos han sido asignados un número único de 10 dígitos llamado Número Internacional Estándar de Libro o por sus siglas en inglés, ISBN (*International Standard Book Number*). Los primeros nueve dígitos de este número se dividen en tres grupos: el primer grupo representa el país o grupo de países en los que se originó el libro, el segundo identifica al editor, y el tercero está asignado al título del libro mismo. El décimo y último dígito, llamado dígito de control, se computa a partir de los primeros nueve dígitos y se usa para asegurar que una transmisión electrónica del ISBN, digamos a través de Internet, ocurra sin error. Para explicar cómo se hace esto, consideremos los primeros nueve dígitos del ISBN como un vector \mathbf{b} en \mathbb{R}^9 , y dejemos que \mathbf{a} sea el vector

$$\mathbf{a} = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \end{pmatrix}.$$

Entonces, el dígito de control c se computa usando el siguiente procedimiento:

- i) Formar el producto punto $\mathbf{a} \bullet \mathbf{b}$.

- ii) Dividir $\mathbf{a} \bullet \mathbf{b}$ entre 11, produciendo un resto c que es un entero entre 0 y 10, inclusive. El dígito de control se toma como c , con la salvedad de que $c = 10$ se escribe como X para evitar dígitos dobles.

Por ejemplo, el ISBN del libro *Álgebra Lineal*, quinta edición, de Stanley I. Grossman es

$$970 - 10 - 0890 - 1$$

que tiene un dígito de control de 1. Esto es consistente con los primeros nueve dígitos del ISBN, ya que

$$\mathbf{a} \bullet \mathbf{b} = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \end{pmatrix} \bullet \begin{pmatrix} 9 \\ 7 \\ 0 \\ 1 \\ 0 \\ 0 \\ 8 \\ 9 \\ 0 \end{pmatrix} = 155.$$

Dividiendo 155 entre 11 produce un cociente de 14 y un resto de 1, por lo que el dígito de control es $c = 1$. Si se realiza un pedido electrónico de un libro con un cierto ISBN, el almacén puede usar el procedimiento anterior para verificar que el dígito de control es consistente con los primeros nueve dígitos, reduciendo así la posibilidad de un costoso error de envío.

Nuestro próximo objetivo es redefinir el producto punto sobre vectores en \mathbb{R}^2 y \mathbb{R}^3 y luego extender esa definición a \mathbb{R}^n . Para hacerlo, primero necesitamos definir exactamente qué queremos decir con el “ángulo” entre dos vectores en \mathbb{R}^2 o \mathbb{R}^3 . Para este propósito, dejemos que \mathbf{u} y \mathbf{v} sean vectores no nulos en \mathbb{R}^2 o \mathbb{R}^3 que han sido posicionados de modo que sus puntos iniciales coincidan. Definimos el ángulo entre \mathbf{u} y \mathbf{v} como el ángulo θ determinado por \mathbf{u} y \mathbf{v} que satisface $0 \leq \theta \leq \pi$.

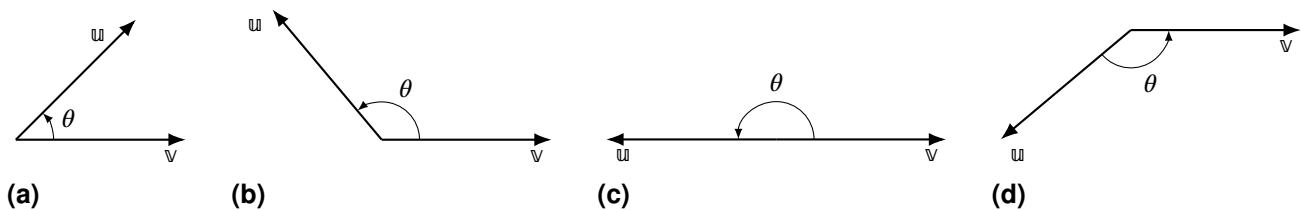


Figura 5.1 El ángulo θ entre \mathbf{u} y \mathbf{v} satisface $0 \leq \theta \leq \pi$

De acuerdo con la ley de cosenos, tenemos que

$$\|\mathbf{u} - \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - 2\|\mathbf{u}\|\|\mathbf{v}\|\cos\theta. \quad (5.1.4)$$

Ahora bien,

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

Por lo tanto,

$$\|\mathbf{u} - \mathbf{v}\|^2 = \|\mathbf{u}\|^2 - 2(\mathbf{u} \bullet \mathbf{v}) + \|\mathbf{v}\|^2. \quad (5.1.5)$$

Igualando ambas expresiones (5.1.4) y (5.1.5),

$$\|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - 2\|\mathbf{u}\|\|\mathbf{v}\|\cos\theta = \|\mathbf{u}\|^2 - 2(\mathbf{u} \bullet \mathbf{v}) + \|\mathbf{v}\|^2.$$

Entonces,

$$\mathbf{u} \bullet \mathbf{v} = \|\mathbf{u}\|\|\mathbf{v}\|\cos\theta. \quad (5.1.6)$$

El signo del producto punto revela información sobre el ángulo θ que podemos obtener reescribiendo la fórmula (5.1.6) como

$$\cos\theta = \frac{\mathbf{u} \bullet \mathbf{v}}{\|\mathbf{u}\|\|\mathbf{v}\|}. \quad (5.1.7)$$

Dado que $0 \leq \theta \leq \pi$, se deduce de la fórmula (5.1.7) y las propiedades de la función coseno estudiadas en trigonometría que

$$\theta \text{ es agudo si } \mathbf{u} \bullet \mathbf{v} > 0, \quad \theta \text{ es obtuso si } \mathbf{u} \bullet \mathbf{v} < 0, \quad \theta = \frac{\pi}{2} \text{ si } \mathbf{u} \bullet \mathbf{v} = 0.$$

Ejemplo 5.1.3:

Encuentre el producto punto de los vectores que se muestran en la figura 5.2.

Solución: La norma de los vectores son

$$\|\mathbf{u}\| = \sqrt{0^2 + 2^2} = 2, \quad \|\mathbf{v}\| = \sqrt{2^2 + 2^2} = \sqrt{8} = 2\sqrt{2}$$

y el coseno del ángulo θ entre ellos es

$$\cos(45^\circ) = \frac{1}{\sqrt{2}}.$$

Por lo tanto, se sigue de la fórmula (5.1.6) que

$$\mathbf{u} \bullet \mathbf{v} = \|\mathbf{u}\|\|\mathbf{v}\|\cos\theta = (2)(2\sqrt{2})\left(\frac{1}{\sqrt{2}}\right) = 4.$$

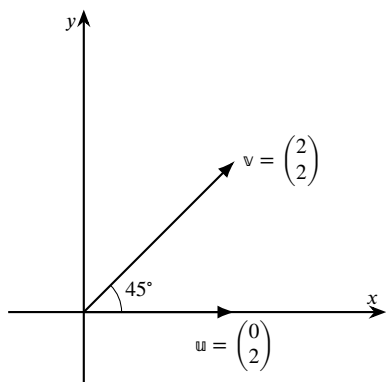


Figura 5.2

Nuestro próximo objetivo es extender a \mathbb{R}^n la noción de “ángulo” entre vectores no nulos \mathbf{u} y \mathbf{v} . Haremos esto comenzando con la fórmula

$$\theta = \arccos\left(\frac{\mathbf{u} \bullet \mathbf{v}}{\|\mathbf{u}\|\|\mathbf{v}\|}\right) \quad (5.1.8)$$

que previamente derivamos para vectores no nulos en \mathbb{R}^2 y \mathbb{R}^3 . Dado que los productos puntos y las normas han sido definidos para vectores en \mathbb{R}^n , parecería que esta fórmula tiene todos los ingredientes para servir como definición del ángulo θ entre dos vectores \mathbf{u} y \mathbf{v} en \mathbb{R}^n . Sin embargo, hay un problema: el arcocoseno en la fórmula (5.1.8) no está definido a menos que su argumento satisfaga

$$-1 \leq \frac{\mathbf{u} \bullet \mathbf{v}}{\|\mathbf{u}\|\|\mathbf{v}\|} \leq 1. \quad (5.1.9)$$

Afortunadamente, la expresión (5.1.9) es cierta para cualesquiera vectores no nulos en \mathbb{R}^n como resultado de un resultado fundamental conocido como la *desigualdad de Cauchy-Schwarz*.

Teorema 5.1.3: Desigualdad de Cauchy-Schwarz: Si \mathbf{u} y \mathbf{v} son vectores en \mathbb{R}^n , entonces

$$|\mathbf{u} \bullet \mathbf{v}| \leq \|\mathbf{u}\|\|\mathbf{v}\| \quad (5.1.10)$$

o en términos de componentes

$$|u_1v_1 + u_2v_2 + \cdots + u_nv_n| \leq (u_1^2 + u_2^2 + \cdots + u_n^2)^{1/2} (v_1^2 + v_2^2 + \cdots + v_n^2)^{1/2}.$$

Omitiremos la demostración de este teorema porque más adelante en este capítulo probaremos una versión más general de la cual este será un caso especial. Nuestro objetivo por ahora será usar este teorema para probar que (5.1.9) se cumple para todos los vectores no nulos en \mathbb{R}^n . Una vez que esto se haya logrado, habremos establecido todos los resultados necesarios para usar la fórmula (5.1.8) como nuestra definición del ángulo entre vectores no nulos \mathbf{u} y \mathbf{v} en \mathbb{R}^n .

Para probar que la expresión (5.1.9) se cumple para todos los vectores no nulos en \mathbb{R}^n , dividimos ambos lados de la fórmula (5.1.10) entre el producto $\|\mathbf{u}\|\|\mathbf{v}\|$ para obtener

$$\frac{|\mathbf{u} \bullet \mathbf{v}|}{\|\mathbf{u}\|\|\mathbf{v}\|} \leq 1 \quad \text{o equivalentemente} \quad \left| \frac{\mathbf{u} \bullet \mathbf{v}}{\|\mathbf{u}\|\|\mathbf{v}\|} \right| \leq 1$$

de lo cual (5.1.9) se deduce.

Las normas y el producto punto en \mathbb{R}^n poseen relaciones fundamentales que conectan la geometría y el álgebra. El siguiente teorema resume identidades clave como la desigualdad del triángulo y la identidad del paralelogramo.

Teorema 5.1.4: Si \mathbf{u} y \mathbf{v} son vectores en \mathbb{R}^n , entonces

- i) $\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$. (Desigualdad del triángulo)
- ii) $\|\mathbf{u} + \mathbf{v}\|^2 + \|\mathbf{u} - \mathbf{v}\|^2 = 2(\|\mathbf{u}\|^2 + \|\mathbf{v}\|^2)$. (Identidad del paralelogramo)
- iii) $\|\mathbf{u} + \mathbf{v}\|\|\mathbf{u} - \mathbf{v}\| \leq \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$.
- iv) $\mathbf{u} \bullet \mathbf{v} = \frac{1}{4}(\|\mathbf{u} + \mathbf{v}\|^2 - \|\mathbf{u} - \mathbf{v}\|^2)$. (Identidad de polarización)

Demostración:

i) Sean \mathbf{u} y \mathbf{v} vectores en \mathbb{R}^n ,

$$\begin{aligned} \|\mathbf{u} + \mathbf{v}\| &= (\mathbf{u} + \mathbf{v}) \bullet (\mathbf{u} + \mathbf{v}) \\ &= (\mathbf{u} \bullet \mathbf{u}) + 2(\mathbf{u} \bullet \mathbf{v}) + (\mathbf{v} \bullet \mathbf{v}) \\ &= \|\mathbf{u}\|^2 + 2(\mathbf{u} \bullet \mathbf{v}) + \|\mathbf{v}\|^2 \\ &\leq \|\mathbf{u}\|^2 + 2|\mathbf{u} \bullet \mathbf{v}| + \|\mathbf{v}\|^2 \\ &\leq \|\mathbf{u}\|^2 + 2\|\mathbf{u}\|\|\mathbf{v}\| + \|\mathbf{v}\|^2 \\ &= (\|\mathbf{u}\| + \|\mathbf{v}\|)^2 \end{aligned}$$

ii) Sean \mathbf{u} y \mathbf{v} vectores en \mathbb{R}^n ,

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

iii) Sean \mathbf{u} y \mathbf{v} vectores en \mathbb{R}^n . Si definimos $a_1 = \|\mathbf{u} + \mathbf{v}\|^2$ y $a_2 = \|\mathbf{u} - \mathbf{v}\|^2$, observemos que podemos usar la *desigualdad entre la media geométrica y la media aritmética* demostrada en el ejemplo A.4.2, pues $a_1, a_2 \geq 0$. De esta forma, obtenemos que

$$\sqrt{\|\mathbf{u} + \mathbf{v}\|^2 \|\mathbf{u} - \mathbf{v}\|^2} \leq \frac{\|\mathbf{u} + \mathbf{v}\|^2 + \|\mathbf{u} - \mathbf{v}\|^2}{2},$$

es decir,

$$\|\mathbf{u} + \mathbf{v}\|\|\mathbf{u} - \mathbf{v}\| \leq \frac{\|\mathbf{u} + \mathbf{v}\|^2 + \|\mathbf{u} - \mathbf{v}\|^2}{2}.$$

Pero por el inciso anterior, se obtiene finalmente que

$$\|\mathbf{u} + \mathbf{v}\|\|\mathbf{u} - \mathbf{v}\| \leq \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2.$$

iv) Se deja como ejercicio al lector.

Definición 5.1.2:

Dados $u_1, u_2, \dots, u_k \in \mathbb{R}^n$.

- i) Diremos que dos vectores u_i y u_j con $i \neq j$ son *ortogonales* si $u_i \bullet u_j = 0$. En general, un conjunto $\{u_1, u_2, \dots, u_k\}$ se llama *conjunto ortogonal* si todos sus vectores son ortogonales entre sí.
- ii) Si además se cumple que cada u_i es un vector unitario, entonces los vectores se denominan *ortonormales*, y el conjunto $\{u_1, u_2, \dots, u_k\}$ se llama *conjunto ortonormal*.

Ejemplo 5.1.4:

- i) Considere los siguientes vectores

$$u = \begin{pmatrix} -2 \\ 3 \\ 1 \\ 4 \end{pmatrix}, \quad v = \begin{pmatrix} 1 \\ 2 \\ 0 \\ -1 \end{pmatrix}.$$

Muestre que u y v son vectores ortogonales en \mathbb{R}^4 .

- ii) Sea $S = \{e_1, e_2, e_3\}$ el conjunto de vectores estándar en \mathbb{R}^3 . Muestre que el conjunto S es un conjunto ortogonal.

Solución:

- i) Los vectores son ortogonales, ya que

$$u \bullet v = (-2)(1) + (3)(2) + (1)(0) + (4)(-1) = 0.$$

- ii) Es suficiente mostrar que

$$e_1 \bullet e_2 = e_1 \bullet e_3 = e_2 \bullet e_3 = 0,$$

porque se sigue del segundo inciso del teorema 2.2.1 que

$$e_2 \bullet e_1 = e_3 \bullet e_1 = e_3 \bullet e_2 = 0.$$

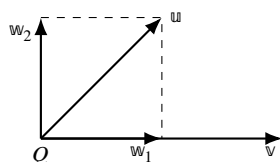
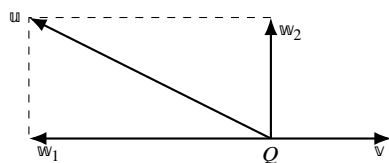
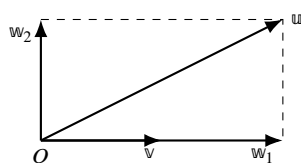
Por lo tanto

$$e_1 \bullet e_2 = (1)(0) + (0)(1) + (0)(0) = 0$$

$$e_1 \bullet e_3 = (1)(0) + (0)(0) + (0)(1) = 0$$

$$e_2 \bullet e_3 = (0)(0) + (1)(0) + (0)(1) = 0$$

lo que prueba que S es un conjunto ortogonal.

**(a)****(b)****(c)****Figura 5.3** Tres posibles casos

En muchas aplicaciones es necesario “descomponer” un vector u en la suma de dos términos, un término siendo un múltiplo escalar de un vector no nulo específico v y el otro término siendo ortogonal a v . Por ejemplo, si u y v son vectores en \mathbb{R}^2 que están posicionados de modo que sus puntos iniciales coinciden en un punto Q , entonces podemos crear tal descomposición como sigue (figura 5.3):

- Primero, consideremos la recta que pasa por el origen en la dirección del vector v . Trazamos una perpendicular desde el extremo de u hasta dicha recta.
- A continuación, definimos el vector w_1 como aquel que parte del punto inicial Q y llega hasta el pie de la perpendicular construida en el paso anterior.
- Finalmente, definimos el vector $w_2 = u - w_1$.

Dado que

$$w_1 + w_2 = w_1 + (u - w_1) = u$$

hemos descompuesto u en la suma de dos vectores ortogonales, el primer término siendo un múltiplo escalar de v y el segundo siendo ortogonal a v .

El siguiente teorema muestra que los resultados anteriores, que ilustramos solo usando vectores en \mathbb{R}^2 , también se aplican en \mathbb{R}^n .

Teorema 5.1.5: Teorema de proyección: Si \mathbf{u} y \mathbf{v} son vectores en \mathbb{R}^n , y si $\mathbf{v} \neq \mathbf{0}$, entonces \mathbf{u} puede expresarse de manera única en la forma $\mathbf{u} = \mathbf{w}_1 + \mathbf{w}_2$, donde \mathbf{w}_1 es un múltiplo escalar de \mathbf{v} y \mathbf{w}_2 es ortogonal a \mathbf{v} .

Demostración: Dado que el vector \mathbf{w}_1 debe ser un múltiplo escalar de \mathbf{v} , debe tener la forma

$$\mathbf{w}_1 = k\mathbf{v} \quad (5.1.11)$$

Nuestro objetivo es encontrar un valor para k y un vector \mathbf{w}_2 que sea ortogonal a \mathbf{v} de modo que

$$\mathbf{u} = \mathbf{w}_1 + \mathbf{w}_2 \quad (5.1.12)$$

Podemos determinar k usando (5.1.11) para reescribir (5.1.12) como

$$\mathbf{u} = \mathbf{w}_1 + \mathbf{w}_2 = k\mathbf{v} + \mathbf{w}_2$$

y luego aplicando propiedades del producto punto para obtener

$$\mathbf{u} \bullet \mathbf{v} = (k\mathbf{v} + \mathbf{w}_2) \bullet \mathbf{v} = k\|\mathbf{v}\|^2 + (\mathbf{w}_2 \bullet \mathbf{v}) \quad (5.1.13)$$

Dado que \mathbf{w}_2 debe ser ortogonal a \mathbf{v} , el último término en (5.1.13) debe ser 0, y por lo tanto k debe satisfacer la ecuación

$$\mathbf{u} \bullet \mathbf{v} = k\|\mathbf{v}\|^2$$

de la cual obtenemos

$$k = \frac{\mathbf{u} \bullet \mathbf{v}}{\|\mathbf{v}\|^2}$$

como el único valor posible para k . La demostración se completa reescribiendo (5.1.12) como

$$\mathbf{w}_2 = \mathbf{u} - \mathbf{w}_1 = \mathbf{u} - k\mathbf{v} = \mathbf{u} - \frac{\mathbf{u} \bullet \mathbf{v}}{\|\mathbf{v}\|^2} \mathbf{v}$$

y luego verificando que \mathbf{w}_2 es ortogonal a \mathbf{v} , ya que

$$\begin{aligned} \mathbf{w}_2 \bullet \mathbf{v} &= \left(\mathbf{u} - \frac{\mathbf{u} \bullet \mathbf{v}}{\|\mathbf{v}\|^2} \mathbf{v} \right) \bullet \mathbf{v} \\ &= \mathbf{u} \bullet \mathbf{v} - \frac{\mathbf{u} \bullet \mathbf{v}}{\|\mathbf{v}\|^2} (\mathbf{v} \bullet \mathbf{v}) \\ &= \mathbf{u} \bullet \mathbf{v} - \frac{\mathbf{u} \bullet \mathbf{v}}{\|\mathbf{v}\|^2} \|\mathbf{v}\|^2 \\ &= \mathbf{u} \bullet \mathbf{v} - \mathbf{u} \bullet \mathbf{v} \\ &= 0 \end{aligned}$$

Los vectores \mathbf{w}_1 y \mathbf{w}_2 en el teorema anterior tienen nombres asociados: el vector \mathbf{w}_1 se llama la *proyección ortogonal de \mathbf{u} sobre \mathbf{v}* o a veces el *componente vectorial de \mathbf{u} a lo largo de \mathbf{v}* , y el vector \mathbf{w}_2 se llama el *componente vectorial de \mathbf{u} ortogonal a \mathbf{v}* . El vector \mathbf{w}_1 se denota comúnmente por el símbolo $\text{proy}_{\mathbf{v}} \mathbf{u}$, en cuyo caso se sigue de (5.1.12) que $\mathbf{w}_2 = \mathbf{u} - \text{proy}_{\mathbf{v}} \mathbf{u}$. En resumen, la *proyección ortogonal de \mathbf{u} sobre \mathbf{v}* está dada por

$$\text{proy}_{\mathbf{v}} \mathbf{u} = \frac{\mathbf{u} \bullet \mathbf{v}}{\|\mathbf{v}\|^2} \mathbf{v} = (\mathbf{u} \bullet \hat{\mathbf{v}}) \hat{\mathbf{v}} \quad (5.1.14)$$

y el *componente vectorial de \mathbf{u} ortogonal a \mathbf{v}* está dado por

$$\mathbf{u} - \text{proy}_{\mathbf{v}} \mathbf{u} = \mathbf{u} - \frac{\mathbf{u} \bullet \mathbf{v}}{\|\mathbf{v}\|^2} \mathbf{v} = \mathbf{u} - (\mathbf{u} \bullet \hat{\mathbf{v}}) \hat{\mathbf{v}}, \quad (5.1.15)$$

que además, es ortogonal al vector \mathbf{v} .

Ejemplo 5.1.5:

Encuentra las proyecciones ortogonales de los vectores e_1 y e_2 sobre la recta L que forma un ángulo θ con el eje x positivo en \mathbb{R}^2 .

Solución: Como se ilustra en la figura 5.4, v es un vector unitario a lo largo de la recta L , por lo que nuestro primer problema es encontrar la proyección ortogonal de e_1 sobre v . Dado que

$$\hat{v} = \frac{1}{\sqrt{\cos^2 \theta + \sin^2 \theta}} \begin{pmatrix} \cos \theta \\ \sin \theta \end{pmatrix} = \begin{pmatrix} \cos \theta \\ \sin \theta \end{pmatrix} \quad \text{y} \quad e_1 \cdot \hat{v} = \cos \theta$$

se deduce de la fórmula (5.1.14) que esta proyección es

$$\text{proy}_v e_1 = (e_1 \cdot \hat{v}) \hat{v} = \cos \theta \begin{pmatrix} \cos \theta \\ \sin \theta \end{pmatrix} = \begin{pmatrix} \cos^2 \theta \\ \sin \theta \cos \theta \end{pmatrix}.$$

De manera similar, dado que $e_2 \cdot \hat{v} = \sin \theta$, se deduce de la fórmula (5.1.14) que

$$\text{proy}_v e_2 = (e_2 \cdot \hat{v}) \hat{v} = \sin \theta \begin{pmatrix} \cos \theta \\ \sin \theta \end{pmatrix} = \begin{pmatrix} \sin \theta \cos \theta \\ \sin^2 \theta \end{pmatrix}.$$

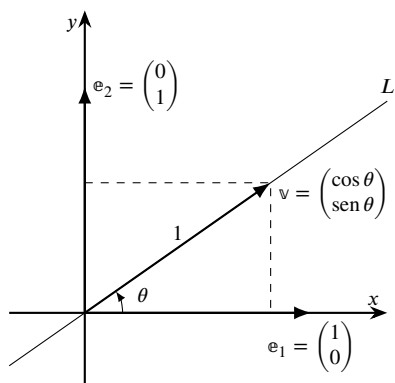


Figura 5.4

A veces estaremos más interesados en la norma de la proyección ortogonal de u sobre v que en la proyección ortogonal de u sobre v en sí. Una fórmula para esta norma se puede derivar como sigue:

$$\|\text{proy}_v u\| = \left\| \frac{u \cdot v}{\|v\|^2} v \right\| = \frac{|u \cdot v|}{\|v\|^2} \|v\| = \frac{|u \cdot v|}{\|v\|}. \quad (5.1.16)$$

Si θ denota el ángulo entre u y v , entonces $u \cdot v = \|u\| \|v\| \cos \theta$, por lo que también se puede escribir como

$$\|\text{proy}_v u\| = \|u\| |\cos \theta|.$$

Ahora bien, mostraremos cómo las proyecciones ortogonales pueden usarse para resolver los siguientes tres problemas de distancia:

- **Problema 1:** Encontrar la distancia entre un punto y una recta en \mathbb{R}^2 .
- **Problema 2:** Encontrar la distancia entre un punto y un plano en \mathbb{R}^3 .
- **Problema 3:** Encontrar la distancia entre dos planos paralelos en \mathbb{R}^3 .

Un método para resolver los dos primeros problemas se proporciona en el siguiente teorema. Dado que las demostraciones de los dos incisos son similares, probaremos el inciso (b) y dejaremos el inciso (a) como ejercicio al lector.

Teorema 5.1.6:

a) En \mathbb{R}^2 , la distancia D entre el punto P_0 y la recta $ax + by + c = 0$ es

$$D = \frac{|ax_0 + by_0 + c|}{\sqrt{a^2 + b^2}}.$$

b) En \mathbb{R}^3 , la distancia D entre el punto P_0 y el plano $ax + by + cz + d = 0$ es

$$D = \frac{|ax_0 + by_0 + cz_0 + d|}{\sqrt{a^2 + b^2 + c^2}}.$$

Demostración: La idea subyacente de la prueba se ilustra en la figura 5.5. Como se muestra en esa figura, sea Q cualquier punto en el plano, y sea n un vector normal al plano que está posicionado con su punto inicial en Q . Ahora es evidente que la distancia D entre P_0 y el plano es simplemente la norma de la proyección ortogonal del vector $P_0 - Q$ sobre n , que por la fórmula (5.1.16) es

$$D = \|\text{proy}_n (P_0 - Q)\| = \frac{|(P_0 - Q) \cdot n|}{\|n\|}.$$

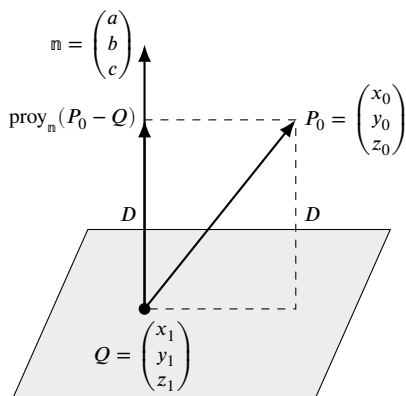


Figura 5.5 La distancia del punto P_0 al plano

Pero

$$P_0 - Q = \begin{pmatrix} x_0 - x_1 \\ y_0 - y_1 \\ z_0 - z_1 \end{pmatrix}, \quad (P_0 - Q) \cdot \mathbf{n} = a(x_0 - x_1) + b(y_0 - y_1) + c(z_0 - z_1),$$

y además, $\|\mathbf{n}\| = \sqrt{a^2 + b^2 + c^2}$. Por lo tanto,

$$D = \frac{|a(x_0 - x_1) + b(y_0 - y_1) + c(z_0 - z_1)|}{\sqrt{a^2 + b^2 + c^2}}. \quad (5.1.17)$$

Dado que el punto Q yace en el plano dado, sus coordenadas satisfacen la ecuación de ese plano; así

$$ax_1 + by_1 + cz_1 + d = 0$$

o bien,

$$d = -ax_1 - by_1 - cz_1.$$

Sustituyendo esta expresión en (5.1.17) se obtiene

$$D = \frac{|ax_0 + by_0 + cz_0 + d|}{\sqrt{a^2 + b^2 + c^2}}.$$

Ejemplo 5.1.6:

Encuentra la distancia D entre el punto

$$P_0 = \begin{pmatrix} 1 \\ -4 \\ 3 \end{pmatrix}$$

y el plano $2x - 3y + 6z = -1$.

Solución: Dado que las fórmulas de distancia en el teorema 5.1.6 requieren que las ecuaciones de la recta y el plano estén escritas con cero en el lado derecho, primero necesitamos reescribir la ecuación del plano como $2x - 3y + 6z + 1 = 0$, de lo cual obtenemos

$$D = \frac{|2(1) + (-3)(-4) + 6(-3) + 1|}{\sqrt{2^2 + (-3)^2 + 6^2}} = \frac{|2 + 12 - 18 + 1|}{\sqrt{4 + 9 + 36}} = \frac{|-3|}{7} = \frac{3}{7}.$$

El tercer problema de distancia planteado anteriormente es encontrar la distancia entre dos planos paralelos en \mathbb{R}^3 . Como se sugiere en la figura 5.6, la distancia entre un plano V y un plano W puede obtenerse al encontrar cualquier punto P_0 en uno de los planos y calcular la distancia entre ese punto y el otro plano.

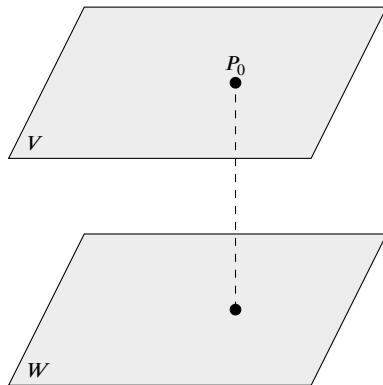


Figura 5.6

Ejemplo 5.1.7:

Los planos $x + 2y - 2z = 3$ y $2x + 4y - 4z = 7$ son paralelos ya que sus vectores normales,

$$\mathbf{n}_1 = \begin{pmatrix} 1 \\ 2 \\ -2 \end{pmatrix} \quad \text{y} \quad \mathbf{n}_2 = \begin{pmatrix} 2 \\ 4 \\ -4 \end{pmatrix},$$

son vectores ortogonales. Encuentra la distancia entre estos planos.

Solución: Para encontrar la distancia D entre los planos, podemos seleccionar un punto arbitrario en uno de los planos y calcular su distancia al otro plano. Al establecer $y_0 = z_0 = 0$ en la ecuación $x + 2y - 2z = 3$, obtenemos que $x_0 = 3$ y, por tanto, el punto P_0 en este plano. Análogo al ejemplo anterior, la distancia entre P_0 y el plano $2x + 4y - 4z = 7$ es

$$D = \frac{|2(3) + 4(0) + (-4)(0) - 7|}{\sqrt{2^2 + 4^2 + (-4)^2}} = \frac{|6 + 0 + 0 - 7|}{\sqrt{4 + 16 + 16}} = \frac{|-1|}{6} = \frac{1}{6}.$$

5.2 Proceso de ortonormalización de Gram-Schmidt

En muchos problemas que involucran espacios vectoriales, quien resuelve el problema tiene la libertad de elegir cualquier base para el espacio vectorial que considere apropiada. En \mathbb{R}^n , la solución de un problema a menudo puede simplificarse eligiendo una base en la que los vectores sean ortonormales entre sí. En esta sección, mostraremos cómo obtener dichas bases.

En \mathbb{R}^2 , dos vectores ortogonales no nulos son linealmente independientes porque ninguno es múltiplo escalar del otro; y en \mathbb{R}^3 , tres vectores mutuamente ortogonales no nulos son linealmente independientes porque ninguno se encuentra en el plano de los otros dos (y, por lo tanto, no se pueden expresar como una combinación lineal de los otros dos). El siguiente teorema generaliza estas observaciones.

Teorema 5.2.1: Sean $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k \in \mathbb{R}^n$. Si $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k\}$ es un conjunto ortogonal de vectores diferentes del vector cero, entonces dicho conjunto es linealmente independiente.

Demostración: Ya que $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k\}$ es un conjunto ortogonal, entonces $\mathbf{u}_i \bullet \mathbf{u}_j = 0$, $\forall i \neq j$ con $i, j = 1, 2, \dots, k$. Sea

$$\mathbf{0} = \alpha_1 \mathbf{u}_1 + \alpha_2 \mathbf{u}_2 + \dots + \alpha_k \mathbf{u}_k$$

con $\alpha_i \in \mathbb{R}$, debemos probar que $\alpha_i = 0$ para $i = 1, 2, \dots, k$. Así,

$$\begin{aligned} 0 &= \mathbf{u}_i \bullet \mathbf{0} = \mathbf{u}_i \bullet (\alpha_1 \mathbf{u}_1 + \alpha_2 \mathbf{u}_2 + \dots + \alpha_k \mathbf{u}_k) \\ &= \mathbf{u}_i \bullet (\alpha_1 \mathbf{u}_1) + \mathbf{u}_i \bullet (\alpha_2 \mathbf{u}_2) + \dots + \mathbf{u}_i \bullet (\alpha_i \mathbf{u}_i) + \dots + \mathbf{u}_i \bullet (\alpha_k \mathbf{u}_k) \\ &= \alpha_1 (\mathbf{u}_i \bullet \mathbf{u}_1) + \alpha_2 (\mathbf{u}_i \bullet \mathbf{u}_2) + \dots + \alpha_i (\mathbf{u}_i \bullet \mathbf{u}_i) + \dots + \alpha_k (\mathbf{u}_i \bullet \mathbf{u}_k) \\ &= \alpha_1 \cdot 0 + \alpha_2 \cdot 0 + \dots + \alpha_i \|\mathbf{u}_i\|^2 + \dots + \alpha_k \cdot 0 \end{aligned}$$

Como $\mathbf{u}_i \neq \mathbf{0}$ por hipótesis, $\|\mathbf{u}_i\|^2 > 0$ y se tiene $\alpha_i = 0$ para $i = 1, 2, \dots, k$. Por tanto, $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k$ son linealmente independientes.

6

Capítulo

DIAGONALIZACIÓN Y FORMAS CUADRÁTICAS

7

Capítulo

ESPACIOS VECTORIALES COMPLEJOS

8

Capítulo

APLICACIONES DEL ÁLGEBRA LINEAL

Este capítulo presenta ocho aplicaciones del álgebra lineal, cada una desarrollada en una sección independiente. Esto permite al lector abordar las secciones en el orden que prefiera, o incluso ignorar aquellas que no considere relevantes, según sus intereses o necesidades. El objetivo principal es mostrar cómo el álgebra lineal puede aplicarse en distintos contextos, destacando su versatilidad y utilidad.

Dado el enfoque práctico del capítulo, algunas demostraciones han sido omitidas para no desviar la atención del lector. Sin embargo, los resultados necesarios de otros campos se enuncian de manera clara y precisa, acompañados de explicaciones que, cuando es posible, ayudan a comprender su importancia en el contexto presentado. De este modo, el capítulo encuentra un equilibrio entre la formalidad y la claridad, ofreciendo una experiencia valiosa tanto para estudiantes que desean aprender los conceptos básicos del álgebra lineal como para profesionales interesados en sus aplicaciones prácticas.

8.1 Construcción de curvas y superficies a través de puntos específicos

En esta sección describimos una técnica que utiliza determinantes para construir rectas, círculos y secciones cónicas generales que pasan por puntos específicos en el plano. Este procedimiento también se emplea para trazar planos y esferas en el espacio tridimensional que pasan por puntos fijos.

Teorema 8.1.1: Un sistema lineal homogéneo con tantas ecuaciones como incógnitas tiene una solución no trivial si y solo si el determinante de la matriz de coeficientes es cero.

Mostraremos ahora cómo este resultado puede utilizarse para determinar ecuaciones de varias curvas y superficies que pasan por puntos específicos.

Una línea a través de dos puntos

Supongamos que (x_1, y_1) y (x_2, y_2) son dos puntos distintos en el plano. Existe una única recta

$$c_1x + c_2y + c_3 = 0 \quad (8.1.1)$$

que pasa por estos dos puntos (figura 8.1). Nótese que c_1, c_2 y c_3 no son todos iguales a cero y que estos coeficientes son únicos hasta una constante multiplicativa. Debido a que (x_1, y_1) y (x_2, y_2) están en la recta, al sustituirlos en (8.1.1) obtenemos las dos ecuaciones

$$c_1x_1 + c_2y_1 + c_3 = 0 \quad (8.1.2)$$

$$c_1x_2 + c_2y_2 + c_3 = 0 \quad (8.1.3)$$

Las tres ecuaciones, (8.1.1), (8.1.2) y (8.1.3), pueden agruparse y reescribirse como

$$xc_1 + yc_2 + c_3 = 0$$

$$x_1c_1 + y_1c_2 + c_3 = 0$$

$$x_2c_1 + y_2c_2 + c_3 = 0$$

Esto constituye un sistema lineal homogéneo de tres ecuaciones para c_1, c_2 y c_3 . Debido a que c_1, c_2 y c_3 no son todos cero, este sistema tiene una solución no trivial, por lo que el determinante de la matriz de coeficientes del sistema debe ser cero. Es decir,

$$\begin{vmatrix} x & y & 1 \\ x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \end{vmatrix} = 0 \quad (8.1.4)$$

En consecuencia, cada punto (x, y) en la recta satisface la ecuación (8.1.4); y, de manera inversa, se puede demostrar que cada punto (x, y) que satisface (8.1.4) pertenece a la recta.

Ejemplo 8.1.1: Encuentra la ecuación de la recta que pasa por los dos puntos $(2, 1)$ y $(3, 7)$.

Solución: Sustituyendo las coordenadas de los dos puntos en la ecuación (8.1.4), obtenemos

$$\begin{vmatrix} x & y & 1 \\ 2 & 1 & 1 \\ 3 & 7 & 1 \end{vmatrix} = 0$$

La expansión de este determinante a lo largo de la primera fila da como resultado

$$-6x + y + 11 = 0.$$

Un círculo a través de tres puntos

Supongamos que hay tres puntos distintos en el plano, (x_1, y_1) , (x_2, y_2) y (x_3, y_3) , que no están todos en una misma línea recta. De la geometría analítica sabemos que existe un círculo único, de la forma,

$$c_1(x^2 + y^2) + c_2x + c_3y + c_4 = 0 \quad (8.1.5)$$

que pasa a través de dichos puntos (figura 8.2). Sustituyendo las coordenadas de los

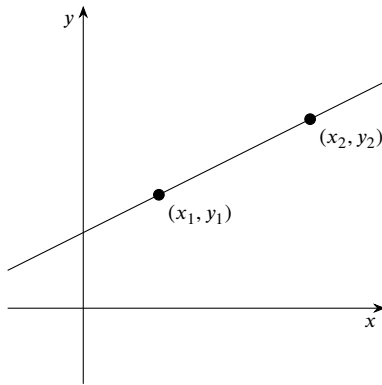


Figura 8.1

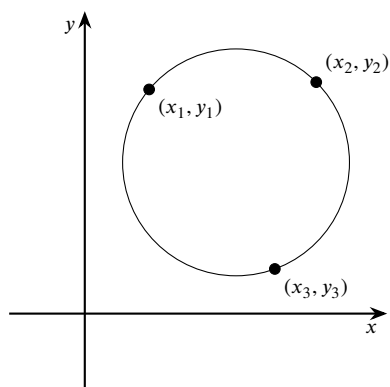


Figura 8.2

tres puntos en esta ecuación, obtenemos

$$c_1 (x_1^2 + y_1^2) + c_2 x_1 + c_3 y_1 + c_4 = 0 \quad (8.1.6)$$

$$c_1 (x_2^2 + y_2^2) + c_2 x_2 + c_3 y_2 + c_4 = 0 \quad (8.1.7)$$

$$c_1 (x_3^2 + y_3^2) + c_2 x_3 + c_3 y_3 + c_4 = 0 \quad (8.1.8)$$

Como antes, las ecuaciones (8.1.5) a (8.1.8) forman un sistema lineal homogéneo con una solución no trivial para c_1, c_2, c_3 y c_4 . Por lo tanto, el determinante de la matriz de coeficientes debe ser cero

$$\begin{vmatrix} x^2 + y^2 & x & y & 1 \\ x_1^2 + y_1^2 & x_1 & y_1 & 1 \\ x_2^2 + y_2^2 & x_2 & y_2 & 1 \\ x_3^2 + y_3^2 & x_3 & y_3 & 1 \end{vmatrix} = 0 \quad (8.1.9)$$

Esta es la forma de obtener la ecuación del círculo.

Ejemplo 8.1.2:

Encuentra la ecuación del círculo que pasa por los tres puntos $(1, 7)$, $(6, 2)$ y $(4, 6)$.

Solución: Sustituyendo las coordenadas de los tres puntos en la ecuación (8.1.9),

$$\begin{vmatrix} x^2 + y^2 & x & y & 1 \\ 50 & 1 & 7 & 1 \\ 40 & 6 & 2 & 1 \\ 52 & 4 & 6 & 1 \end{vmatrix} = 0$$

Al simplificar, esto se reduce a

$$10(x^2 + y^2) - 20x - 40y - 200 = 0$$

En su forma estándar, esta ecuación se escribe como

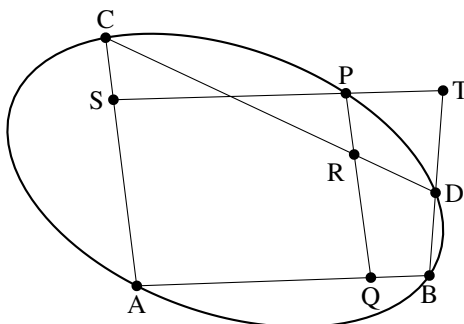
$$(x - 1)^2 + (y - 2)^2 = 5^2$$

Por lo tanto, el círculo tiene centro en $(1, 2)$ y radio 5.

Sección cónica general a través de cinco puntos

En su monumental obra *Principia Mathematica*, Isaac Newton planteó y resolvió el siguiente problema (Libro I, Proposición 22, Problema 14): “Describir una cónica que pase por cinco puntos dados”. Newton resolvió este problema geoméricamente, como se muestra en la figura 8.3, en la que hizo pasar una elipse por los puntos A, B, D, P, C; sin embargo, los métodos de esta sección también pueden aplicarse.

Figura 8.3



La ecuación general de una sección cónica en el plano (una parábola, hipérbola, elipse o formas degeneradas de estas curvas) está dada por

$$c_1 x^2 + c_2 xy + c_3 y^2 + c_4 x + c_5 y + c_6 = 0.$$

Esta ecuación contiene seis coeficientes, pero podemos reducir su número a cinco si dividimos entre alguno de ellos que no sea cero. Por lo tanto, solo se necesitan cinco puntos distintos en el plano para determinar la ecuación de la sección cónica. Como antes, la ecuación puede escribirse en forma de determinante:

$$\begin{vmatrix} x^2 & xy & y^2 & x & y & 1 \\ x_1^2 & x_1y_1 & y_1^2 & x_1 & y_1 & 1 \\ x_2^2 & x_2y_2 & y_2^2 & x_2 & y_2 & 1 \\ x_3^2 & x_3y_3 & y_3^2 & x_3 & y_3 & 1 \\ x_4^2 & x_4y_4 & y_4^2 & x_4 & y_4 & 1 \\ x_5^2 & x_5y_5 & y_5^2 & x_5 & y_5 & 1 \end{vmatrix} = 0 \quad (8.1.10)$$

Ejemplo 8.1.3:

Un astrónomo que desea determinar la órbita de un asteroide alrededor del Sol establece un sistema de coordenadas cartesianas en el plano de la órbita con el Sol en el origen. Se utilizan unidades astronómicas de medida a lo largo de los ejes (1 unidad astronómica = distancia media de la Tierra al Sol = 93 millones de millas). Según la primera ley de Kepler, la órbita debe ser una elipse, por lo que el astrónomo realiza cinco observaciones del asteroide en cinco momentos diferentes y encuentra cinco puntos a lo largo de la órbita

$$(8.025, 8.310), \quad (10.170, 6.355), \quad (11.202, 3.212), \\ (10.736, 0.375), \quad (9.092, -2.267).$$

Determinar la ecuación de la órbita.

Solución: Sustituyendo las coordenadas de los cinco puntos dados en la ecuación (8.1.10) y redondeando a tres cifras decimales obtenemos

$$\begin{vmatrix} x^2 & xy & y^2 & x & y & 1 \\ 64.401 & 66.688 & 69.056 & 8.025 & 8.310 & 1 \\ 103.429 & 64.630 & 40.386 & 10.170 & 6.355 & 1 \\ 125.485 & 35.981 & 10.317 & 11.202 & 3.212 & 1 \\ 115.262 & 4.026 & 0.141 & 10.736 & 0.375 & 1 \\ 82.664 & -20.612 & 5.139 & 9.092 & -2.267 & 1 \end{vmatrix} = 0$$

La expansión del determinante por la primera fila produce

$$386.802x^2 - 102.895xy + 446.029y^2 - 2476.443x - 1427.998y - 17109.375 = 0$$

La figura 8.4 es un diagrama preciso de la órbita, junto con los cinco puntos.

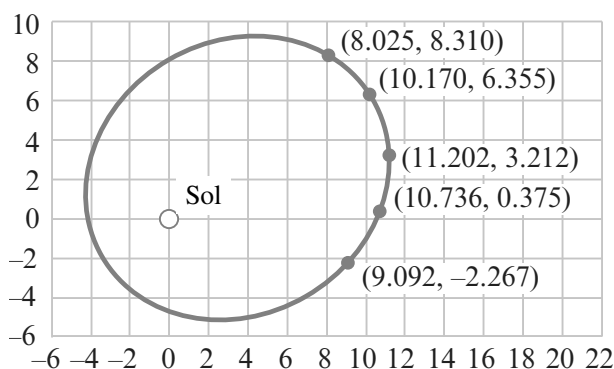


Figura 8.4

Un plano a través de tres puntos

En el ejercicio 8 se deja al lector demostrar lo siguiente: el plano en el espacio tridimensional con ecuación

$$c_1x + c_2y + c_3z + c_4 = 0$$

que pasa por tres puntos no colineales (x_1, y_1, z_1) , (x_2, y_2, z_2) y (x_3, y_3, z_3) está dado por la ecuación determinante

$$\begin{vmatrix} x & y & z & 1 \\ x_1 & y_1 & z_1 & 1 \\ x_2 & y_2 & z_2 & 1 \\ x_3 & y_3 & z_3 & 1 \end{vmatrix} = 0 \quad (8.1.11)$$

Ejemplo 8.1.4: La ecuación del plano que pasa por los tres puntos no colineales $(1, 1, 0)$, $(2, 0, -1)$ y $(2, 9, 2)$ es

$$\begin{vmatrix} x & y & z & 1 \\ 1 & 1 & 0 & 1 \\ 2 & 0 & -1 & 1 \\ 2 & 9 & 2 & 1 \end{vmatrix} = 0,$$

lo que se reduce a

$$2x - y + 3z - 1 = 0.$$

Una esfera a través de cuatro puntos

En el ejercicio 9 se deja al lector demostrar lo siguiente: la esfera en el espacio tridimensional con ecuación

$$c_1(x^2 + y^2 + z^2) + c_2x + c_3y + c_4z + c_5 = 0$$

que pasa por cuatro puntos no coplanarios (x_1, y_1, z_1) , (x_2, y_2, z_2) , (x_3, y_3, z_3) y (x_4, y_4, z_4) está dada por el siguiente determinante

$$\begin{vmatrix} x^2 + y^2 + z^2 & x & y & z & 1 \\ x_1^2 + y_1^2 + z_1^2 & x_1 & y_1 & z_1 & 1 \\ x_2^2 + y_2^2 + z_2^2 & x_2 & y_2 & z_2 & 1 \\ x_3^2 + y_3^2 + z_3^2 & x_3 & y_3 & z_3 & 1 \\ x_4^2 + y_4^2 + z_4^2 & x_4 & y_4 & z_4 & 1 \end{vmatrix} = 0 \quad (8.1.12)$$

Ejemplo 8.1.5: La ecuación de la esfera que pasa por los cuatro puntos no coplanarios $(0, 3, 2)$, $(1, -1, 1)$, $(2, 1, 0)$ y $(5, 1, 3)$ es la que satisface

$$\begin{vmatrix} x^2 + y^2 + z^2 & x & y & z & 1 \\ 13 & 0 & 3 & 2 & 1 \\ 3 & 1 & -1 & 1 & 1 \\ 5 & 2 & 1 & 0 & 1 \\ 35 & 5 & 1 & 3 & 1 \end{vmatrix} = 0,$$

lo que se reduce a

$$x^2 + y^2 + z^2 - 4x - 2y - 6z + 5 = 0,$$

que en su forma estándar es

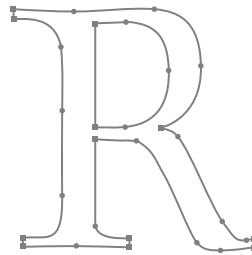
$$(x - 2)^2 + (y - 1)^2 + (z - 3)^2 = 9.$$

8.2 Interpolación mediante splines cúbicos

Ajuste de curvas

Ajustar una curva a través de puntos específicos en el plano es un problema común que se encuentra en el análisis de datos experimentales, en la determinación de las relaciones entre las variables y en el trabajo de diseño. Una aplicación omnipresente está en el diseño y la descripción de las fuentes de ordenador e impresora, como las fuentes PostScript™ y TrueType™.

Figura 8.5 ‘R’ en PostScript™ y TrueType™



En la figura 8.6a se muestran siete puntos en el plano xy , y en la figura 8.6c se ha dibujado una curva suave que pasa a través de ellos. Se dice que una curva que pasa a través de un conjunto de puntos en el plano *interpola* esos puntos, y la curva se llama *curva de interpolación* para esos puntos. La curva de interpolación de la figura 8.6c se dibujó con la ayuda de un *drafting spline* (figura 8.6b). Esta ayuda de dibujo consiste en una tira delgada y flexible de madera u otro material que se dobla para pasar a través de los puntos que se van a interpolar. Los pesos deslizantes adjuntos mantienen el *spline* en posición mientras el diseñador dibuja la curva de interpolación. El *drafting spline* servirá como modelo físico para una teoría matemática de interpolación que discutiremos en esta sección.

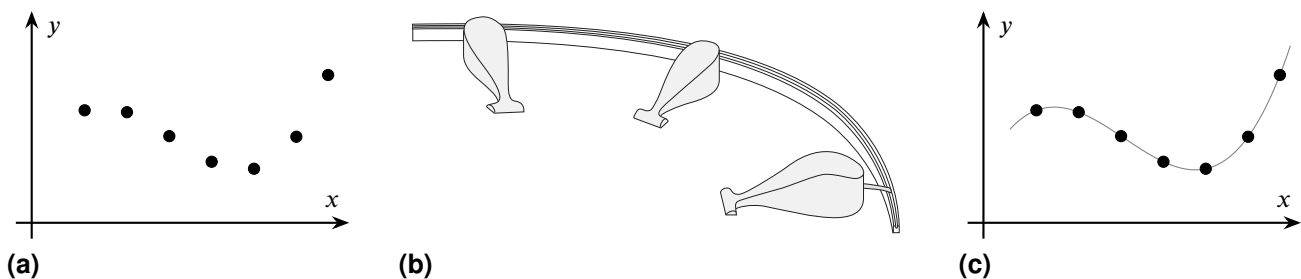


Figura 8.6 Ejemplos de curvas de interpolación

Planteamiento del problema

Supongamos que se nos dan n puntos en el plano xy ,

$$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$$

que deseamos interpolar con una curva “bien comportada” (figura 8.7). Para mayor comodidad, tomamos los puntos equiespaciados en la dirección x , aunque nuestros resultados se pueden extender fácilmente al caso de puntos desigualmente espaciados. Si tomamos la distancia común entre las coordenadas x de los puntos como h , entonces tenemos

$$x_2 - x_1 = x_3 - x_2 = \dots = x_n - x_{n-1} = h$$

Sea $y = S(x)$, $x_1 \leq x \leq x_n$, la curva de interpolación que buscamos. Supondremos que esta curva describe el desplazamiento de una regla flexible que interpola los

n puntos cuando los pesos que sujetan la regla se sitúan precisamente en los n puntos. Se sabe, por la teoría de vigas lineales, que para pequeños desplazamientos, la cuarta derivada del desplazamiento de una viga es cero a lo largo de cualquier intervalo del eje x que no contenga fuerzas externas actuando sobre la viga. Si tratamos nuestra regla flexible como una viga delgada y comprendemos que las únicas fuerzas externas actuando sobre ella provienen de los pesos en los n puntos especificados, entonces se sigue que

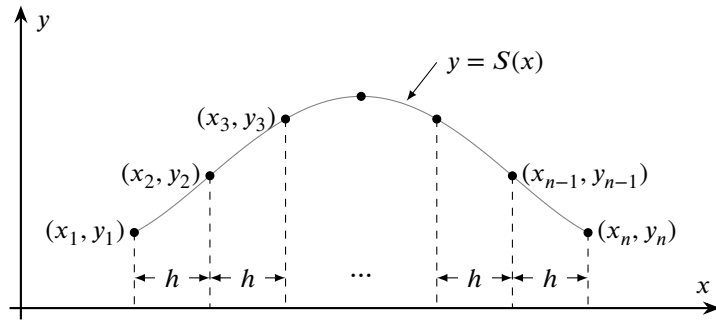
$$S^{(4)}(x) \equiv 0$$

para valores de x que se encuentran en los $n - 1$ intervalos abiertos

$$(x_1, x_2), (x_2, x_3), \dots, (x_{n-1}, x_n)$$

entre los n puntos.

Figura 8.7



También necesitamos el resultado de la teoría de vigas lineales que establece que, para una viga sujeta únicamente a fuerzas externas, el desplazamiento debe tener dos derivadas continuas. En el caso de la curva interpolante $y = S(x)$ construida con la regla flexible, esto significa que $S(x)$, $S'(x)$ y $S''(x)$ deben ser continuas para $x_1 \leq x \leq x_n$.

La condición de que $S''(x)$ sea continua es lo que hace que una regla flexible produzca una curva agradable, ya que resulta en una curvatura continua. El ojo puede percibir cambios bruscos en la curvatura, es decir, discontinuidades en $S''(x)$, pero los cambios bruscos en derivadas de orden superior no son discernibles. Por lo tanto, la condición de que $S''(x)$ sea continua es el requisito mínimo para que la curva interpolante sea perceptible como una curva suave única, en lugar de una serie de curvas separadas unidas.

Para determinar la forma matemática de la función $S(x)$, observamos que, dado que $S^{(4)}(x) \equiv 0$ en los intervalos entre los n puntos especificados, se sigue que al resolver esta ecuación repetidamente, $S(x)$ debe ser un polinomio cúbico en x en cada uno de esos intervalos. Sin embargo, en general, $S(x)$ será un polinomio cúbico diferente en cada intervalo, por lo que $S(x)$ debe tener la forma

$$S(x) = \begin{cases} S_1(x), & x_1 \leq x \leq x_2 \\ S_2(x), & x_2 \leq x \leq x_3 \\ \vdots & \vdots \\ S_{n-1}(x), & x_{n-1} \leq x \leq x_n \end{cases} \quad (8.2.1)$$

donde $S_1(x)$, $S_2(x)$, ..., $S_{n-1}(x)$ son polinomios cúbicos. Para mayor comodidad, los escribiremos de la forma

$$\begin{aligned} S_1(x) &= a_1(x - x_1)^3 + b_1(x - x_1)^2 + c_1(x - x_1) + d_1, & x_1 \leq x \leq x_2 \\ S_2(x) &= a_2(x - x_2)^3 + b_2(x - x_2)^2 + c_2(x - x_2) + d_2, & x_2 \leq x \leq x_3 \\ &\vdots \\ S_{n-1}(x) &= a_{n-1}(x - x_{n-1})^3 + b_{n-1}(x - x_{n-1})^2 \\ &\quad + c_{n-1}(x - x_{n-1}) + d_{n-1}, & x_{n-1} \leq x \leq x_n \end{aligned} \quad (8.2.2)$$

Los coeficientes a_i , b_i , c_i y d_i constituyen un total de $4n - 4$ coeficientes, los cuales debemos determinar para definir completamente la función $S(x)$. Para lograrlo, seleccionamos estos coeficientes de manera que $S(x)$ interpole los n puntos dados en el plano y, además, garantice que $S(x)$, junto con sus derivadas primera y segunda, sean continuas en todo el dominio. Estas condiciones aseguran una transición suave entre los segmentos polinómicos, lo que da lugar a la curva de interpolación conocida como *spline cúbico*.

Derivación de la fórmula de un spline cúbico

De las ecuaciones (8.2.1) y (8.2.2), tenemos que

$$\begin{aligned} S(x) &= S_1(x) = a_1(x - x_1)^3 + b_1(x - x_1)^2 + c_1(x - x_1) + d_1, & x_1 \leq x \leq x_2 \\ S(x) &= S_2(x) = a_2(x - x_2)^3 + b_2(x - x_2)^2 + c_2(x - x_2) + d_2, & x_2 \leq x \leq x_3 \\ &\vdots \\ S(x) &= S_{n-1}(x) = a_{n-1}(x - x_{n-1})^3 + b_{n-1}(x - x_{n-1})^2 \\ &\quad + c_{n-1}(x - x_{n-1}) + d_{n-1}, & x_{n-1} \leq x \leq x_n \end{aligned}$$

Primero, derivando la expresión (8.2.1), es decir, derivando cada polinomio anterior, se sigue que

$$\begin{aligned} S'_1(x) &= 3a_1(x - x_1)^2 + 2b_1(x - x_1) + c_1, & x_1 \leq x \leq x_2 \\ S'_2(x) &= 3a_2(x - x_2)^2 + 2b_2(x - x_2) + c_2, & x_2 \leq x \leq x_3 \\ &\vdots \\ S'_{n-1}(x) &= 3a_{n-1}(x - x_{n-1})^2 + 2b_{n-1}(x - x_{n-1}) + c_{n-1}, & x_{n-1} \leq x \leq x_n \end{aligned} \quad (8.2.3)$$

y derivando nuevamente, obtenemos que

$$\begin{aligned} S''_1(x) &= 6a_1(x - x_1) + 2b_1, & x_1 \leq x \leq x_2 \\ S''_2(x) &= 6a_2(x - x_2) + 2b_2, & x_2 \leq x \leq x_3 \\ &\vdots \\ S''_{n-1}(x) &= 6a_{n-1}(x - x_{n-1}) + 2b_{n-1}, & x_{n-1} \leq x \leq x_n \end{aligned}$$

Ahora utilizaremos estas ecuaciones y las cuatro propiedades de los splines cúbicos que se indican a continuación para expresar los coeficientes desconocidos a_i , b_i , c_i , d_i con $i = 1, 2, \dots, n - 1$, en términos de los valores conocidos y_1, y_2, \dots, y_n , lo que nos permitirá construir explícitamente la curva de interpolación deseada.

$S(x)$ interpola los puntos (x_i, y_i) con $i = 1, 2, \dots, n$.

Debido a que $S(x)$ interpola los puntos (x_i, y_i) , con $i = 1, 2, \dots, n$, tenemos

$$S(x_1) = y_1, S(x_2) = y_2, \dots, S(x_n) = y_n \quad (8.2.4)$$

De las primera $n - 1$ de estas ecuaciones y (8.2.1), obtenemos

$$\begin{aligned} d_1 &= y_1 \\ d_2 &= y_2 \\ &\vdots \\ d_{n-1} &= y_{n-1} \end{aligned} \quad (8.2.5)$$

De la última ecuación en (8.2.4), la última ecuación en (8.2.1), y el hecho de que $x_n - x_{n-1} = h$, obtenemos

$$a_{n-1}h^3 + b_{n-1}h^2 + c_{n-1}h + d_{n-1} = y_n \quad (8.2.6)$$

$S(x)$ es continua en $[x_1, x_n]$.

Debido a que $S(x)$ es continuo para $x_1 \leq x \leq x_n$, se deduce que en cada punto x_i en el conjunto x_2, x_3, \dots, x_{n-1} debemos tener

$$S_{i-1}(x_i) = S_i(x_i), \quad i = 2, 3, \dots, n-1 \quad (8.2.7)$$

De lo contrario, las gráficas de $S_{i-1}(x)$ y $S_i(x)$ no se unirían para formar una curva continua en x_i . Cuando aplicamos la propiedad de interpolación $S_i(x_i) = y_i$, se sigue de (8.2.7) que $S_{i-1}(x_i) = y_i$ donde $i = 2, 3, \dots, n-1$, o de (8.2.1) que

$$\begin{aligned} a_1 h^3 + b_1 h^2 + c_1 h + d_1 &= y_2 \\ a_2 h^3 + b_2 h^2 + c_2 h + d_2 &= y_3 \\ &\vdots \\ a_{n-2} h^3 + b_{n-2} h^2 + c_{n-2} h + d_{n-2} &= y_{n-1} \end{aligned} \quad (8.2.8)$$

$S'(x)$ es continua en $[x_1, x_n]$.

Debido a que $S'(x)$ es continuo para $x_1 \leq x \leq x_n$, se deduce que

$$S'_{i-1}(x_i) = S'_i(x_i), \quad i = 2, 3, \dots, n-1$$

es decir

$$\begin{aligned} 3a_1 h^2 + 2b_1 h + c_1 &= c_2 \\ 3a_2 h^2 + 2b_2 h + c_2 &= c_3 \\ &\vdots \\ 3a_{n-2} h^2 + 2b_{n-2} h + c_{n-2} &= c_{n-1} \end{aligned} \quad (8.2.9)$$

$S''(x)$ es continua en $[x_1, x_n]$.

Debido a que $S''(x)$ es continuo para $x_1 \leq x \leq x_n$, se deduce que

$$S''_{i-1}(x_i) = S''_i(x_i), \quad i = 2, 3, \dots, n-1$$

es decir

$$\begin{aligned} 6a_1 h + 2b_1 &= 2b_2 \\ 6a_2 h + 2b_2 &= 2b_3 \\ &\vdots \\ 6a_{n-2} h + 2b_{n-2} &= 2b_{n-1} \end{aligned} \quad (8.2.10)$$

Las ecuaciones (8.2.5), (8.2.6), (8.2.8), (8.2.9) y (8.2.10) constituyen un sistema de $4n - 6$ ecuaciones lineales con $4n - 4$ coeficientes desconocidos a_i, b_i, c_i, d_i con $i = 1, 2, \dots, n-1$. Por lo tanto, necesitamos dos ecuaciones más para determinar estos coeficientes de manera única. Sin embargo, antes de obtener estas ecuaciones adicionales, podemos simplificar nuestro sistema existente expresando las incógnitas a_i, b_i, c_i y d_i en términos de nuevas cantidades desconocidas:

$$M_1 = S''(x_1), M_2 = S''(x_2), \dots, M_n = S''(x_n)$$

y las cantidades conocidas y_1, y_2, \dots, y_n . Es decir,

$$\begin{aligned} M_1 &= 2b_1 \\ M_2 &= 2b_2 \\ &\vdots \\ M_{n-1} &= 2b_{n-1} \end{aligned}$$

así que

$$b_1 = \frac{1}{2}M_1, b_2 = \frac{1}{2}M_2, \dots, b_{n-1} = \frac{1}{2}M_{n-1}$$

Además, sabemos de (8.2.5) que

$$d_1 = y_1, d_2 = y_2, \dots, d_{n-1} = y_{n-1}$$

Dejamos como ejercicio al lector derivar las expresiones de los a_i y c_i en términos de los M_i y y_i . El resultado final es el siguiente:

Teorema 8.2.1: Consideremos n puntos $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ donde $x_{i+1} - x_i = h$ con $i = 1, 2, \dots, n-1$. El spline cúbico

$$S(x) = \begin{cases} a_1(x-x_1)^3 + b_1(x-x_1)^2 + c_1(x-x_1) + d_1, & x_1 \leq x \leq x_2 \\ a_2(x-x_2)^3 + b_2(x-x_2)^2 + c_2(x-x_2) + d_2, & x_2 \leq x \leq x_3 \\ \vdots \\ a_{n-1}(x-x_{n-1})^3 + b_{n-1}(x-x_{n-1})^2 \\ \quad + c_{n-1}(x-x_{n-1}) + d_{n-1}, & x_{n-1} \leq x \leq x_n \end{cases}$$

que interpola estos puntos tiene como coeficientes a

$$\begin{aligned} a_i &= \frac{M_{i+1} - M_i}{6h} \\ b_i &= \frac{M_i}{2} \\ c_i &= \frac{y_{i+1} - y_i}{h} - \frac{(M_{i+1} + 2M_i)h}{6} \\ d_i &= y_i \end{aligned} \quad (8.2.11)$$

para $i = 1, 2, \dots, n-1$, donde $M_i = S''(x_i)$ con $i = 1, 2, \dots, n$.

A partir de este resultado, vemos que las cantidades M_1, M_2, \dots, M_n determinan de manera única el spline cúbico. Para encontrar estas cantidades, sustituimos las expresiones de a_i, b_i y c_i dadas en (8.2.11) en (8.2.9). Después de algunas simplificaciones algebraicas, obtenemos

$$\begin{aligned} M_1 + 4M_2 + M_3 &= \frac{6(y_1 - 2y_2 + y_3)}{h^2} \\ M_2 + 4M_3 + M_4 &= \frac{6(y_2 - 2y_3 + y_4)}{h^2} \\ &\vdots \\ M_{n-2} + 4M_{n-1} + M_n &= \frac{6(y_{n-2} - 2y_{n-1} + y_n)}{h^2} \end{aligned} \quad (8.2.12)$$

o, en forma matricial,

$$\begin{bmatrix} 1 & 4 & 1 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 1 & 4 & 1 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 4 & \cdots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & & \ddots & \ddots & \ddots & & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 4 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 1 & 4 & 1 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 & 4 & 1 \end{bmatrix} \begin{bmatrix} M_1 \\ M_2 \\ M_3 \\ \vdots \\ M_{n-2} \\ M_{n-1} \\ M_n \end{bmatrix} = \frac{6}{h^2} \begin{bmatrix} y_1 - 2y_2 + y_3 \\ y_2 - 2y_3 + y_4 \\ y_3 - 2y_4 + y_5 \\ \vdots \\ y_{n-4} - 2y_{n-3} + y_{n-2} \\ y_{n-3} - 2y_{n-2} + y_{n-1} \\ y_{n-2} - 2y_{n-1} + y_n \end{bmatrix}$$

Observemos que este es un sistema lineal de $n-2$ ecuaciones con n incógnitas dadas por M_1, M_2, \dots, M_n . Aún necesitamos dos ecuaciones adicionales para determinar estos valores de manera única, ya que hay infinitos splines cúbicos que interpolan los puntos dados. Para obtener un spline cúbico único, se requieren condiciones adicionales. A continuación, presentamos tres formas de especificarlas, resumidas en la siguiente tabla:

Spline natural	$M_1 = 0$ $M_n = 0$	$\begin{bmatrix} 4 & 1 & 0 & \cdots & 0 & 0 & 0 \\ 1 & 4 & 1 & \cdots & 0 & 0 & 0 \\ \vdots & & \ddots & \ddots & & & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 4 & 1 \\ 0 & 0 & 0 & \cdots & 0 & 1 & 4 \end{bmatrix} \begin{bmatrix} M_2 \\ M_3 \\ \vdots \\ M_{n-2} \\ M_{n-1} \end{bmatrix} = \frac{6}{h^2} \begin{bmatrix} y_1 - 2y_2 + y_3 \\ y_2 - 2y_3 + y_4 \\ \vdots \\ y_{n-3} - 2y_{n-2} + y_{n-1} \\ y_{n-2} - 2y_{n-1} + y_n \end{bmatrix}$
Spline de salida parabólica	$M_1 = M_2$ $M_n = M_{n-1}$	$\begin{bmatrix} 5 & 1 & 0 & \cdots & 0 & 0 & 0 \\ 1 & 4 & 1 & \cdots & 0 & 0 & 0 \\ \vdots & & \ddots & \ddots & & & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 4 & 1 \\ 0 & 0 & 0 & \cdots & 0 & 1 & 5 \end{bmatrix} \begin{bmatrix} M_2 \\ M_3 \\ \vdots \\ M_{n-2} \\ M_{n-1} \end{bmatrix} = \frac{6}{h^2} \begin{bmatrix} y_1 - 2y_2 + y_3 \\ y_2 - 2y_3 + y_4 \\ \vdots \\ y_{n-3} - 2y_{n-2} + y_{n-1} \\ y_{n-2} - 2y_{n-1} + y_n \end{bmatrix}$
Spline de salida cúbica	$M_1 = 2M_2 - M_3$ $M_n = 2M_{n-1} - M_{n-2}$	$\begin{bmatrix} 6 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 1 & 4 & 1 & \cdots & 0 & 0 & 0 \\ \vdots & & \ddots & \ddots & & & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 4 & 1 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 6 \end{bmatrix} \begin{bmatrix} M_2 \\ M_3 \\ \vdots \\ M_{n-2} \\ M_{n-1} \end{bmatrix} = \frac{6}{h^2} \begin{bmatrix} y_1 - 2y_2 + y_3 \\ y_2 - 2y_3 + y_4 \\ \vdots \\ y_{n-3} - 2y_{n-2} + y_{n-1} \\ y_{n-2} - 2y_{n-1} + y_n \end{bmatrix}$

Tabla 8.1 Condiciones adicionales para la unicidad del spline cúbico

Es decir, en el spline natural, la segunda derivada del spline es cero en los extremos; en el spline de salida parabólica, el spline se reduce a una curva parabólica en los primeros y últimos intervalos; y en el spline de salida cúbica, el spline es una única curva cúbica en los dos primeros y los dos últimos intervalos.

El spline natural

Las dos condiciones matemáticas más simples que podemos imponer son

$$M_1 = M_n = 0$$

Estas condiciones, junto con (8.2.12), resultan en un sistema lineal de $n \times n$ para M_1, M_2, \dots, M_n , que se puede escribir en forma matricial como

$$\begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 1 & 4 & 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & 4 & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & & \ddots & \ddots & & & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & 4 & 1 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} M_1 \\ M_2 \\ M_3 \\ \vdots \\ M_{n-1} \\ M_n \end{bmatrix} = \frac{6}{h^2} \begin{bmatrix} 0 \\ y_1 - 2y_2 + y_3 \\ y_2 - 2y_3 + y_4 \\ \vdots \\ y_{n-2} - 2y_{n-1} + y_n \\ 0 \end{bmatrix}$$

Para cálculos numéricos, es más conveniente eliminar M_1 y M_n de este sistema y escribir

$$\begin{bmatrix} 4 & 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 1 & 4 & 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & 4 & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & & \ddots & \ddots & & & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & 4 & 1 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 & 4 \end{bmatrix} \begin{bmatrix} M_2 \\ M_3 \\ M_4 \\ \vdots \\ M_{n-2} \\ M_{n-1} \end{bmatrix} = \frac{6}{h^2} \begin{bmatrix} y_1 - 2y_2 + y_3 \\ y_2 - 2y_3 + y_4 \\ y_3 - 2y_4 + y_5 \\ \vdots \\ y_{n-3} - 2y_{n-2} + y_{n-1} \\ y_{n-2} - 2y_{n-1} + y_n \end{bmatrix}$$

junto con

$$M_1 = 0 \quad (8.2.13)$$

$$M_2 = 0 \quad (8.2.14)$$

Así, el sistema lineal de $(n-2) \times (n-2)$ puede resolverse para los $n-2$ coeficientes M_2, M_3, \dots, M_{n-1} , y M_1 y M_n están determinados mediante (8.2.13) y (8.2.14).

Físicamente, el spline natural resulta cuando los extremos de una regla flexible se extienden libremente más allá de los puntos interpolados sin restricción. Las porciones finales del spline fuera de los puntos interpolados caerán en trayectorias de líneas rectas, lo que hace que $S''(x)$ se anule en los extremos x_1 y x_n , y resultando en las condiciones matemáticas $M_1 = M_n = 0$.

El spline natural tiende a aplanar la curva interpolante en los extremos, lo cual puede ser indeseable. Por supuesto, si se requiere que $S''(x)$ se anule en los extremos, entonces se debe usar el spline natural.

El spline de salida parabólica

Las dos restricciones adicionales impuestas para este tipo de spline son

$$M_1 = M_2 \quad (8.2.15)$$

$$M_n = M_{n-1} \quad (8.2.16)$$

Si utilizamos las dos ecuaciones anteriores para eliminar M_1 y M_n de (8.2.12), obtenemos el sistema lineal de $(n-2) \times (n-2)$

$$\begin{bmatrix} 5 & 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 1 & 4 & 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & 4 & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & & \ddots & \ddots & \ddots & & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & 4 & 1 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 & 5 \end{bmatrix} \begin{bmatrix} M_2 \\ M_3 \\ M_4 \\ \vdots \\ M_{n-2} \\ M_{n-1} \end{bmatrix} = \frac{6}{h^2} \begin{bmatrix} y_1 - 2y_2 + y_3 \\ y_2 - 2y_3 + y_4 \\ y_3 - 2y_4 + y_5 \\ \vdots \\ y_{n-3} - 2y_{n-2} + y_{n-1} \\ y_{n-2} - 2y_{n-1} + y_n \end{bmatrix}$$

para M_2, M_3, \dots, M_{n-1} . Una vez que se han determinado estos $n-2$ valores, M_1 y M_n se determinan a partir de (8.2.15) y (8.2.16). De (8.2.11), si $M_1 = M_2$, entonces $a_1 = 0$, y si $M_n = M_{n-1}$, entonces $a_{n-1} = 0$. Así, en (8.2.1), no hay términos cúbicos en los intervalos $[x_1, x_2]$ y $[x_{n-1}, x_n]$, por lo que el spline de salida parabólica se reduce a una curva parabólica en estos extremos.

El spline de salida cúbica

Para este tipo de spline, imponemos las siguientes dos condiciones adicionales

$$M_1 = 2M_2 - M_3 \quad (8.2.17)$$

$$M_n = 2M_{n-1} - M_2 \quad (8.2.18)$$

Usando estas dos ecuaciones para eliminar M_1 y M_n de (8.2.12), resulta en el siguiente sistema lineal de $(n-2) \times (n-2)$ para M_2, M_3, \dots, M_{n-1} :

$$\begin{bmatrix} 6 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 1 & 4 & 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & 4 & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & & \ddots & \ddots & \ddots & & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & 4 & 1 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 6 \end{bmatrix} \begin{bmatrix} M_2 \\ M_3 \\ M_4 \\ \vdots \\ M_{n-2} \\ M_{n-1} \end{bmatrix} = \frac{6}{h^2} \begin{bmatrix} y_1 - 2y_2 + y_3 \\ y_2 - 2y_3 + y_4 \\ y_3 - 2y_4 + y_5 \\ \vdots \\ y_{n-3} - 2y_{n-2} + y_{n-1} \\ y_{n-2} - 2y_{n-1} + y_n \end{bmatrix}$$

Después de resolver este sistema lineal para M_2, M_3, \dots, M_{n-1} , podemos usar (8.2.17) y (8.2.18) para determinar M_1 y M_n .

Si reescribimos (8.2.17) como

$$M_2 - M_1 = M_3 - M_2$$

se sigue de (8.2.11) que $a_1 = a_2$. Dado que $S'''(x) = 6a_1$ en $[x_1, x_2]$ y $S'''(x) = 6a_2$ en $[x_2, x_3]$, vemos que $S'''(x)$ es constante sobre todo el intervalo $[x_1, x_3]$. En consecuencia, $S(x)$ consiste en una única curva cúbica sobre el intervalo $[x_1, x_3]$ en lugar de dos curvas cúbicas diferentes unidas en x_2 . Un análisis similar muestra que $S(x)$ consiste en una única curva cúbica sobre los últimos dos intervalos.

Mientras que el spline natural tiende a aplanar la curva interpolante en los extremos, el spline de salida cúbica tiene el efecto contrario, generando mayor curvatura en los extremos. Si ninguno de estos comportamientos es deseado, el spline de salida parabólica ofrece un equilibrio intermedio.

Ejemplo 8.2.1:

Es bien sabido que la densidad del agua alcanza un máximo a una temperatura ligeramente superior a la de congelación. La tabla 8.2, tomada del *Handbook of Chemistry and Physics* (CRC Press, 2009), proporciona la densidad del agua en gramos por centímetro cúbico para cinco temperaturas igualmente espaciadas desde -10°C hasta 30°C . Interpolaremos estas cinco mediciones de temperatura y densidad utilizando un *spline* parabólico con final ajustado (*runout spline*) e intentaremos encontrar la densidad máxima del agua en este rango al determinar el valor máximo de este *spline* cúbico. En los ejercicios, se le solicita realizar cálculos similares utilizando un *spline* natural y un *spline* cúbico con final ajustado para interpolar los puntos de datos.

Temperatura ($^\circ\text{C}$)	Densidad (g/cm^3)
-10	0.99815
0	0.99987
10	0.99973
20	0.99823
30	0.99567

Tabla 8.2

Vamos a interpolar estas cinco mediciones de temperatura-densidad con un spline de salida parabólica y trataremos de encontrar la densidad máxima del agua en este rango hallando el valor máximo en este spline cúbico. Sean

$$\begin{array}{ll} x_1 = -10 & y_1 = 0.99815 \\ x_2 = 0 & y_2 = 0.99987 \\ x_3 = 10 & y_3 = 0.99973 \\ x_4 = 20 & y_4 = 0.99823 \\ x_5 = 30 & y_5 = 0.99567 \end{array}$$

entonces

$$\begin{aligned} \frac{6(y_1 - 2y_2 + y_3)}{h^2} &= \frac{6(0.99815 - 2(0.99987) + 0.99973)}{10^2} = -0.0001116 \\ \frac{6(y_2 - 2y_3 + y_4)}{h^2} &= \frac{6(0.99987 - 2(0.99973) + 0.99823)}{10^2} = -0.0000816 \\ \frac{6(y_3 - 2y_4 + y_5)}{h^2} &= \frac{6(0.99973 - 2(0.99823) + 0.99567)}{10^2} = -0.0000636 \end{aligned}$$

y el sistema lineal para el spline de salida parabólica se convierte en

$$\begin{bmatrix} 5 & 1 & 4 \\ 1 & 4 & 1 \\ 0 & 1 & 5 \end{bmatrix} \begin{bmatrix} M_2 \\ M_3 \\ M_4 \end{bmatrix} = \begin{bmatrix} -0.0001116 \\ -0.0000816 \\ -0.0000636 \end{bmatrix}$$

Al resolver este sistema, da como resultado

$$M_2 = -0.00001973, \quad M_3 = -0.00001293, \quad M_4 = -0.00001013$$

De (8.2.15) y (8.2.16), tenemos que

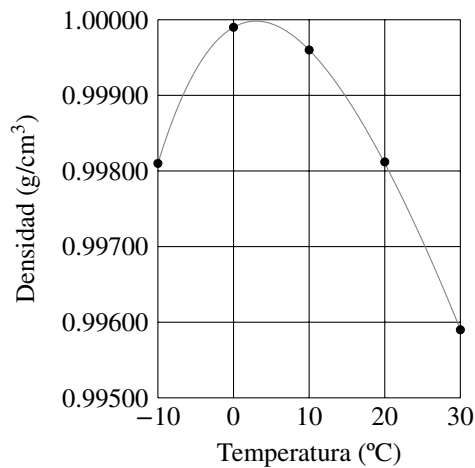
$$M_1 = M_2 = -0.00001973$$

$$M_5 = M_4 = -0.00001013$$

Al resolver para los a_i , b_i , c_i y d_i en (8.2.12), obtenemos la siguiente expresión para el spline de terminación parabólica interpolante:

$$S(x) = \begin{cases} -0.00000987(x+10)^2 + 0.002707(x+10) + 0.99815, & -10 \leq x \leq 0 \\ 0.00000113(x-0)^3 - 0.00000987(x-0)^2 + 0.000733(x-0) + 0.99987, & 0 \leq x \leq 10 \\ 0.00000047(x-10)^3 - 0.00000647(x-10)^2 - 0.000900(x-10) + 0.99973, & 10 \leq x \leq 20 \\ -0.00000507(x-20)^2 - 0.002053(x-20) + 0.99823, & 20 \leq x \leq 30 \end{cases}$$

Este spline se muestra en la siguiente figura:



En dicha figura, vemos que el máximo se alcanza en el intervalo $[0, 10]$. Para encontrar este máximo, igualamos $S'(x)$ a cero en el intervalo $[0, 10]$:

$$S'(x) = 0.000000339x^2 - 0.0001974x + 0.0000733 = 0$$

Con tres cifras significativas, la raíz de esta ecuación cuadrática en el intervalo $[0, 10]$ es $x = 3.99$, y para este valor de x , $S(3.99) = 1.00001$. Así, según nuestra estimación interpolada, la densidad máxima del agua es 1.00001 g/cm^3 alcanzada a 3.99°C . Esto concuerda bien con la densidad máxima experimental de 1.00000 g/cm^3 alcanzada a 3.98°C .

En conclusión, además de producir excelentes curvas de interpolación, los splines cúbicos y sus generalizaciones son útiles para la integración y diferenciación numérica, para la solución numérica de ecuaciones diferenciales e integrales, y en la teoría de la optimización.

8.3 Cadenas de Markov

Supongamos que un sistema físico o matemático atraviesa un proceso de cambio de tal manera que en cualquier momento puede ocupar uno de un número finito de estados. Por ejemplo, el clima en una ciudad determinada podría estar en uno de tres posibles estados: soleado, nublado o lluvioso. O imaginemos que un individuo podría estar en uno de cuatro posibles estados emocionales: feliz, triste, enojado o aprensivo. Supongamos que tal sistema cambia con el tiempo de un estado a otro y que en momentos programados se observa el estado del sistema. Si el estado del sistema no puede predecirse con certeza, pero las probabilidades de los posibles estados dependen únicamente del estado previo, el proceso se denomina *cadena de Markov* o *proceso de Markov*.

Definición 8.3.1:

Si una cadena de Markov tiene k posibles estados, etiquetados como $1, 2, \dots, k$, entonces la probabilidad de que el sistema esté en el estado i en una observación, después de haber estado en el estado j en la observación anterior, se denota por p_{ij} y se llama probabilidad de transición del estado j al estado i . La matriz $P = [p_{ij}]$ se llama *matriz de transición de la cadena de Markov*.

Por ejemplo, en una cadena de Markov de tres estados, la matriz de transición tiene la forma

		Estado precedente		
		1	2	3
Nuevo estado	1	$\begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{bmatrix}$		
	2			
	3			

En esta matriz, p_{32} es la probabilidad de que el sistema cambie del estado 2 al estado 3, p_{11} es la probabilidad de que el sistema permanezca en el estado 1 si estaba previamente en el estado 1, y así sucesivamente.

Ejemplo 8.3.1:

Una agencia de alquiler de coches tiene tres ubicaciones de alquiler, denominadas 1, 2 y 3. Un cliente puede alquilar un coche en cualquiera de las tres ubicaciones y devolverlo a cualquiera de las tres ubicaciones. El gerente encuentra que los clientes devuelven los coches a las diversas ubicaciones según las siguientes probabilidades:

		Alquilado desde la ubicación		
		1	2	3
Regresado a la ubicación	1	$\begin{bmatrix} 0.8 & 0.3 & 0.2 \\ 0.1 & 0.2 & 0.6 \\ 0.1 & 0.5 & 0.2 \end{bmatrix}$		
	2			
	3			

Esta matriz es la matriz de transición del sistema considerado como una cadena de Markov. De esta matriz, se puede ver que la probabilidad de que un coche alquilado en la ubicación 3 sea devuelto a la ubicación 2 es de 0.3, la probabilidad de que un coche alquilado en la ubicación 1 sea devuelto a la ubicación 1 es de 0.8, y así sucesivamente.

Ejemplo 8.3.2:

Al revisar los registros de donaciones, la oficina de exalumnos de una universidad descubre que el 80 % de sus exalumnos que contribuyen al fondo anual un año también contribuirán al siguiente año, y el 30 % de aquellos que no contribuyen un año contribuirán al siguiente. Esto se puede ver como una cadena de Markov con dos estados: el estado 1 corresponde a un exalumno que dona en un año determinado, y el estado 2 corresponde a un exalumno que no dona en ese año. La matriz de transición es

$$P = \begin{bmatrix} 0.80 & 0.30 \\ 0.20 & 0.70 \end{bmatrix}$$

En los ejemplos anteriores, las matrices de transición de las cadenas de Markov tienen la propiedad de que la suma de las entradas en cualquier columna es igual a 1. Esto no es accidental. Si $P = [p_{ij}]$ es la matriz de transición de cualquier cadena de Markov con k estados, entonces para cada j debemos tener

$$p_{1j} + p_{2j} + \dots + p_{kj} = 1 \quad (8.3.1)$$

porque si el sistema está en el estado j en una observación, es seguro que estará en uno de los k estados posibles en la siguiente observación.

Una matriz con la propiedad (8.3.1) se llama *matriz estocástica*, *matriz de probabilidad* o *matriz de Markov*. De la discusión anterior, se sigue que la matriz de transición para una cadena de Markov debe ser una matriz estocástica.

En una cadena de Markov, el estado del sistema en cualquier momento de observación generalmente no puede determinarse con certeza. Lo mejor que se puede hacer normalmente es especificar probabilidades para cada uno de los estados posibles. Por ejemplo, en una cadena de Markov con tres estados, podríamos describir el estado posible del sistema en algún momento de observación mediante un vector columna

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$$

en el cual, x_1 es la probabilidad de que el sistema esté en el estado 1, x_2 es la probabilidad de que esté en el estado 2, x_3 es la probabilidad de que esté en el estado 3, y x_4 es la probabilidad de que esté en el estado 4. En general, tenemos la siguiente definición.

Definición 8.3.2:

El *vector de estado* para una observación de una cadena de Markov con k estados es un vector columna \mathbf{x} , cuya i -ésima componente x_i es la probabilidad de que el sistema esté en el estado i en ese momento.

Observe que las entradas en cualquier vector de estado para una cadena de Markov son no negativas y suman 1 (¿por qué?). Un vector columna que tiene esta propiedad se llama *vector de probabilidad*.

Supongamos ahora que conocemos el vector de estado $\mathbf{x}^{(0)}$ para una cadena de Markov en alguna observación inicial. El siguiente teorema nos permitirá determinar los vectores de estado

$$\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(n)}, \dots$$

en los tiempos de observación subsiguientes.

Teorema 8.3.1:

Si P es la matriz de transición de una cadena de Markov y $\mathbf{x}^{(n)}$ es el vector de estado en la n -ésima observación, entonces $\mathbf{x}^{(n+1)} = P\mathbf{x}^{(n)}$.

La demostración del teorema anterior recurre a conceptos avanzados de la teoría de probabilidad, los cuales exceden el alcance de esta presentación, por lo que no se incluirá aquí. Sin embargo, como consecuencia directa de este teorema, se deduce que

$$\begin{aligned} \mathbf{x}^{(1)} &= P\mathbf{x}^{(0)} \\ \mathbf{x}^{(2)} &= P\mathbf{x}^{(1)} = P^2\mathbf{x}^{(0)} \\ \mathbf{x}^{(3)} &= P\mathbf{x}^{(2)} = P^3\mathbf{x}^{(0)} \\ \mathbf{x}^{(4)} &= P\mathbf{x}^{(3)} = P^4\mathbf{x}^{(0)} \\ &\vdots \\ \mathbf{x}^{(n)} &= P\mathbf{x}^{(n-1)} = P^n\mathbf{x}^{(0)} \end{aligned}$$

De esta manera, el vector de estado inicial $\mathbf{x}^{(0)}$ y la matriz de transición P determinan $\mathbf{x}^{(n)}$ para $n = 1, 2, \dots$.

Ejemplo 8.3.3:

La matriz de transición en el ejemplo 8.3.2 era

$$P = \begin{bmatrix} 0.8 & 0.3 \\ 0.2 & 0.7 \end{bmatrix}$$

Ahora construimos el probable registro futuro de donaciones de un nuevo graduado que no hizo una donación en el año inicial después de la graduación. Para tal graduado, el sistema está inicialmente en el estado 2 con certeza, por lo que el vector de estado inicial es

$$\mathbf{x}^{(0)} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

A partir del teorema 8.3.1, tenemos entonces

$$\begin{aligned}\mathbf{x}^{(1)} &= P\mathbf{x}^{(0)} = \begin{bmatrix} 0.8 & 0.3 \\ 0.2 & 0.7 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.3 \\ 0.7 \end{bmatrix} \\ \mathbf{x}^{(2)} &= P\mathbf{x}^{(1)} = \begin{bmatrix} 0.8 & 0.3 \\ 0.2 & 0.7 \end{bmatrix} \begin{bmatrix} 0.3 \\ 0.7 \end{bmatrix} = \begin{bmatrix} 0.45 \\ 0.55 \end{bmatrix} \\ \mathbf{x}^{(3)} &= P\mathbf{x}^{(2)} = \begin{bmatrix} 0.8 & 0.3 \\ 0.2 & 0.7 \end{bmatrix} \begin{bmatrix} 0.45 \\ 0.55 \end{bmatrix} = \begin{bmatrix} 0.525 \\ 0.475 \end{bmatrix}\end{aligned}$$

Así, después de tres años, se espera que el exalumno haga una donación con una probabilidad de 0.525. Más allá de tres años, encontramos los siguientes vectores de estado (con tres decimales):

$$\begin{aligned}\mathbf{x}^{(4)} &= \begin{bmatrix} 0.563 \\ 0.438 \end{bmatrix}, \quad \mathbf{x}^{(5)} = \begin{bmatrix} 0.581 \\ 0.419 \end{bmatrix}, \quad \mathbf{x}^{(6)} = \begin{bmatrix} 0.591 \\ 0.409 \end{bmatrix}, \quad \mathbf{x}^{(7)} = \begin{bmatrix} 0.595 \\ 0.405 \end{bmatrix}, \\ \mathbf{x}^{(8)} &= \begin{bmatrix} 0.598 \\ 0.402 \end{bmatrix}, \quad \mathbf{x}^{(9)} = \begin{bmatrix} 0.599 \\ 0.401 \end{bmatrix}, \quad \mathbf{x}^{(10)} = \begin{bmatrix} 0.599 \\ 0.401 \end{bmatrix}, \quad \mathbf{x}^{(11)} = \begin{bmatrix} 0.600 \\ 0.400 \end{bmatrix}\end{aligned}$$

Para todos los n mayores de 11, tenemos

$$\mathbf{x}^{(n)} = \begin{bmatrix} 0.600 \\ 0.400 \end{bmatrix}$$

con tres decimales. En otras palabras, los vectores de estado convergen a un vector fijo a medida que aumenta el número de observaciones.

Ejemplo 8.3.4:

La matriz de transición en el ejemplo 8.3.1 era

$$\begin{bmatrix} 0.8 & 0.3 & 0.2 \\ 0.1 & 0.2 & 0.6 \\ 0.1 & 0.5 & 0.2 \end{bmatrix}$$

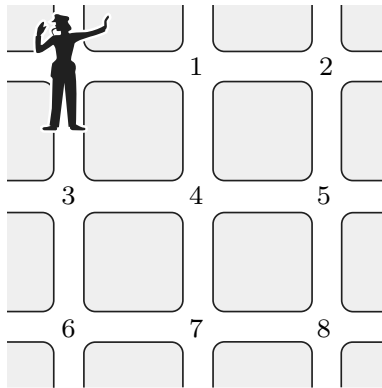
Si un auto se alquila inicialmente desde la ubicación 2, entonces el vector de estado inicial es

$$\mathbf{x}^{(0)} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

Usando este vector y el teorema 8.3.1, se obtienen los siguientes vectores de estado que se enumeran en la siguiente tabla:

$\mathbf{x}^{(n)} \backslash n$	0	1	2	3	4	5	6	7	8	9	10	11
$x_1^{(n)}$	0	0.300	0.400	0.477	0.511	0.533	0.544	0.550	0.553	0.555	0.556	0.557
$x_2^{(n)}$	1	0.200	0.370	0.252	0.261	0.240	0.238	0.233	0.232	0.231	0.230	0.230
$x_3^{(n)}$	0	0.500	0.230	0.271	0.228	0.227	0.219	0.217	0.215	0.214	0.214	0.213

Para todos los valores de n mayores a 11, todos los vectores de estado son iguales a $\mathbf{x}^{(11)}$ con tres decimales. Esto indica que el sistema ha alcanzado un estado estable, en el cual las probabilidades de transición ya no generan cambios significativos en la distribución del estado. Se deben observar dos aspectos clave en este ejemplo. Primero, no era necesario conocer la duración exacta del tiempo que un cliente mantiene el coche. En otras palabras, en un proceso de Markov, el intervalo de tiempo entre observaciones no necesita ser regular, lo que lo hace útil en situaciones donde las transiciones ocurren en momentos aleatorios o no uniformes. Segundo, los vectores de estado se aproximan hacia un vector fijo a medida que n aumenta. Esto sugiere que el sistema alcanza un equilibrio estacionario, lo que es una propiedad fundamental de las cadenas de Markov.

Ejemplo 8.3.5:**Figura 8.8**

Una oficial de tráfico está asignada a controlar el tráfico en las ocho intersecciones indicadas en la figura 8.8. Se le instruye permanecer en cada intersección durante una hora y luego, ya sea permanecer en la misma intersección o moverse a una intersección vecina. Para evitar establecer un patrón, se le indica elegir su nueva intersección de manera aleatoria, con cada elección posible igualmente probable. Por ejemplo, si se encuentra en la intersección 5, su próxima intersección puede ser 2, 4, 5, o 8, cada una con una probabilidad de $1/4$. Cada día, comienza en el lugar donde se detuvo el día anterior. La matriz de transición para esta cadena de Markov es:

		Antigua intersección							
		1	2	3	4	5	6	7	8
Nueva intersección	1	$\frac{1}{3}$	$\frac{1}{3}$	0	$\frac{1}{5}$	0	0	0	0
	2	$\frac{1}{3}$	$\frac{1}{3}$	0	0	$\frac{1}{4}$	0	0	0
	3	0	0	$\frac{1}{3}$	$\frac{1}{5}$	0	$\frac{1}{3}$	0	0
	4	$\frac{1}{3}$	0	$\frac{1}{3}$	$\frac{1}{5}$	$\frac{1}{4}$	0	$\frac{1}{4}$	0
	5	0	$\frac{1}{3}$	0	$\frac{1}{5}$	$\frac{1}{4}$	0	0	$\frac{1}{3}$
	6	0	0	$\frac{1}{3}$	0	0	$\frac{1}{3}$	$\frac{1}{4}$	0
	7	0	0	0	$\frac{1}{5}$	0	$\frac{1}{3}$	$\frac{1}{4}$	$\frac{1}{3}$
	8	0	0	0	0	$\frac{1}{4}$	0	$\frac{1}{4}$	$\frac{1}{3}$

Si la oficial de tráfico comienza en la intersección 5, sus ubicaciones probables, hora por hora, están dadas por los vectores de estado que se muestran en la tabla precedente. Para todos los valores de n mayores que 22, todos los vectores de estado son iguales a $\mathbf{x}^{(22)}$ con tres decimales. Así, al igual que en los dos primeros ejemplos, los vectores de estado se acercan a un vector fijo a medida que n aumenta.

$\mathbf{x}^{(n)}$ \ n	0	1	2	3	4	5	...	10	...	15	...	21	22
$x_1^{(n)}$	0	0.000	0.133	0.116	0.130	0.123	...	0.113	...	0.109	...	0.108	0.107
$x_2^{(n)}$	0	0.250	0.146	0.163	0.140	0.138	...	0.115	...	0.109	...	0.108	0.107
$x_3^{(n)}$	0	0.000	0.050	0.039	0.067	0.073	...	0.100	...	0.106	...	0.107	0.107
$x_4^{(n)}$	0	0.250	0.113	0.187	0.162	0.178	...	0.178	...	0.179	...	0.179	0.179
$x_5^{(n)}$	1	0.250	0.279	0.190	0.190	0.168	...	0.149	...	0.144	...	0.143	0.143
$x_6^{(n)}$	0	0.000	0.000	0.050	0.056	0.074	...	0.099	...	0.105	...	0.107	0.107
$x_7^{(n)}$	0	0.000	0.133	0.104	0.131	0.125	...	0.138	...	0.142	...	0.143	0.143
$x_8^{(n)}$	0	0.250	0.146	0.152	0.124	0.121	...	0.108	...	0.107	...	0.107	0.107

En nuestros ejemplos vimos que los vectores de estado se aproximaban a un vector fijo a medida que aumentaba el número de observaciones. Ahora nos preguntamos si los vectores de estado siempre se aproximan a un vector fijo en una cadena de Markov. Un ejemplo sencillo muestra que este no es siempre el caso.

Ejemplo 8.3.6:

Sean

$$P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad \text{y} \quad \mathbf{x}^{(0)} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

Entonces, dado que $P^2 = I_2$ y $P^3 = P$, tenemos que

$$\mathbf{x}^{(0)} = \mathbf{x}^{(2)} = \mathbf{x}^{(4)} = \dots = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

y

$$\mathbf{x}^{(1)} = \mathbf{x}^{(3)} = \mathbf{x}^{(5)} = \dots = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Este sistema oscila indefinidamente entre los dos vectores $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ y $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$, por lo que no se aproxima a ningún vector fijo.

No obstante, si imponemos una condición simple a la matriz de transición, podemos demostrar que se alcanza un vector de estado fijo. Esta condición se define a continuación.

Definición 8.3.3:

Una matriz de transición se considera *regular* si, al elevarla a alguna potencia entera, todas sus entradas son positivas.

Por lo tanto, para una matriz de transición regular P , existe un entero positivo m tal que todas las entradas de P^m son positivas. Este es el caso de las matrices de transición en los ejemplos 8.3.1 y 8.3.2 con $m = 1$. En el ejemplo 8.3.5, resulta que P^4 tiene todas las entradas positivas. Por lo tanto, en los tres ejemplos, las matrices de transición son regulares. Esto significa que dichas matrices satisfacen una propiedad clave: cualquier estado puede alcanzarse desde cualquier otro estado después de un número finito de pasos, lo cual garantiza una conectividad total en el espacio de estados.

Una cadena de Markov con una matriz de transición regular se llama *cadena de Markov regular*. Este tipo de cadenas exhibe un comportamiento especial: independientemente del estado inicial $\mathbf{x}^{(0)}$, la distribución de probabilidad del estado converge hacia un único vector de estado fijo \mathbf{q} a medida que el número de pasos n tiende a infinito. Este vector \mathbf{q} es invariante bajo la acción de la matriz P , es decir, satisface $P\mathbf{q} = \mathbf{q}$.

La existencia de este vector fijo \mathbf{q} es de suma importancia porque describe el comportamiento a largo plazo del sistema modelado por la cadena de Markov. Por ejemplo, en aplicaciones como redes de comunicación, procesos biológicos, o modelos económicos, este resultado permite predecir el estado estable del sistema, independientemente de su configuración inicial. Además, la unicidad y positividad de los elementos de \mathbf{q} garantizan que el sistema tiene una distribución de probabilidad única en el límite, asegurando estabilidad en el análisis.

El resultado mencionado se basa en el siguiente teorema fundamental:

Teorema 8.3.2:

Si P es una matriz de transición regular, entonces a medida que $n \rightarrow \infty$,

$$P^n \rightarrow \begin{bmatrix} q_1 & q_1 & \cdots & q_1 \\ q_2 & q_2 & \cdots & q_2 \\ \vdots & \vdots & \ddots & \vdots \\ q_k & q_k & \cdots & q_k \end{bmatrix}$$

donde los q_i son números positivos tales que $q_1 + q_2 + \cdots + q_k = 1$.

Nota

No demostraremos este teorema aquí. Para una demostración detallada, consulte un texto más especializado, como J. Kemeny y J. Snell, *Finite Markov Chains* (New York: Springer-Verlag, 1976).

Este teorema formaliza la idea de que, para cadenas de Markov regulares, la matriz de transición elevada a potencias suficientemente grandes se estabiliza en una forma en la que cada fila es idéntica y corresponde al vector de estado fijo \mathbf{q} .

Consideremos

$$Q = \begin{bmatrix} q_1 & q_1 & \cdots & q_1 \\ q_2 & q_2 & \cdots & q_2 \\ \vdots & & \ddots & \vdots \\ q_k & q_k & \cdots & q_k \end{bmatrix} \quad \text{y} \quad \mathbf{q} = \begin{bmatrix} q_1 \\ q_2 \\ \vdots \\ q_k \end{bmatrix}$$

Es decir, Q es una matriz de transición cuyas columnas son todas iguales al vector de probabilidad \mathbf{q} . Q tiene la propiedad de que si \mathbf{x} es cualquier vector de probabilidad, entonces

$$\begin{aligned} Q\mathbf{x} &= \begin{bmatrix} q_1 & q_1 & \cdots & q_1 \\ q_2 & q_2 & \cdots & q_2 \\ \vdots & & \ddots & \vdots \\ q_k & q_k & \cdots & q_k \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{bmatrix} \\ &= \begin{bmatrix} q_1 x_1 + q_1 x_2 + \cdots + q_1 x_k \\ q_2 x_1 + q_2 x_2 + \cdots + q_2 x_k \\ \vdots \\ q_k x_1 + q_k x_2 + \cdots + q_k x_k \end{bmatrix} \\ &= (x_1 + x_2 + \cdots + x_k) \begin{bmatrix} q_1 \\ q_2 \\ \vdots \\ q_k \end{bmatrix} \\ &= (1)\mathbf{q} \\ &= \mathbf{q} \end{aligned}$$

Es decir, Q transforma cualquier vector de probabilidad \mathbf{x} en el vector de probabilidad fijo \mathbf{q} . Este resultado nos lleva al siguiente teorema.

Teorema 8.3.3: Si P es una matriz de transición regular y \mathbf{x} es cualquier vector de probabilidad, entonces, conforme $n \rightarrow \infty$,

$$P^n \mathbf{x} \rightarrow \begin{bmatrix} q_1 \\ q_2 \\ \vdots \\ q_k \end{bmatrix} = \mathbf{q}$$

donde \mathbf{q} es un vector de probabilidad fijo, independiente de n , y cuyos componentes son todos positivos.

Este resultado se cumple ya que el teorema 8.3.2 implica que $P^n \rightarrow Q$ a medida que $n \rightarrow \infty$. Esto, a su vez, implica que $P^n \mathbf{x} \rightarrow Q\mathbf{x} = \mathbf{q}$ a medida que $n \rightarrow \infty$. Así, para una cadena de Markov regular, el sistema se aproxima a un vector de estado fijo \mathbf{q} . Este vector \mathbf{q} se llama el *vector de estado estacionario* de la cadena de Markov regular. Para sistemas con muchos estados, la forma más eficiente de calcular el vector de estado estacionario \mathbf{q} es evaluar $P^n \mathbf{x}$ para un n grande. Nuestros ejemplos muestran este procedimiento, garantizando la convergencia debido a la regularidad del proceso. Alternativamente, \mathbf{q} puede calcularse mediante el siguiente teorema.

Teorema 8.3.4: El vector de estado estacionario \mathbf{q} de una matriz de transición regular P es el único vector de probabilidad que satisface la ecuación $P\mathbf{q} = \mathbf{q}$.

Para entender esto, consideremos la identidad $PP^n = P^{n+1}$. Por el teorema 8.3.2, tanto P^n como P^{n+1} convergen a Q cuando $n \rightarrow \infty$, lo que implica $PQ = Q$ y, por lo tanto, $P\mathbf{q} = \mathbf{q}$. Para demostrar la unicidad de \mathbf{q} , supongamos que existe otro vector de probabilidad \mathbf{r} con $P\mathbf{r} = \mathbf{r}$. Entonces, $P^n \mathbf{r} = \mathbf{r}$ para todo n . Tomando el límite cuando $n \rightarrow \infty$ y aplicando el teorema 8.3.3, se concluye que $\mathbf{q} = \mathbf{r}$.

La ecuación del teorema 8.3.4 también se puede expresar como un sistema lineal homogéneo

$$(I_k - P)\mathbf{q} = \mathbf{0}$$

donde tiene una solución única, el vector \mathbf{q} , con entradas no negativas que satisfacen la condición $q_1 + q_2 + \cdots + q_k = 1$. Podemos aplicar esta técnica para calcular los vectores de estado estacionario en nuestros ejemplos.

Ejemplo 8.3.7: En el ejemplo 8.3.2, la matriz de transición era

$$P = \begin{bmatrix} 0.8 & 0.3 \\ 0.2 & 0.7 \end{bmatrix}$$

Entonces, el sistema lineal $(I_2 - P)\mathbf{q} = \mathbf{0}$ es

$$\begin{bmatrix} 0.2 & -0.3 \\ -0.2 & 0.3 \end{bmatrix} \begin{bmatrix} q_1 \\ q_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Esto lleva a la única ecuación independiente

$$0.2q_1 - 0.3q_2 = 0$$

o bien,

$$q_1 = 1.5q_2$$

Es decir, la solución general al sistema está dada por

$$\mathbf{q} = \begin{bmatrix} 1.5q_2 \\ q_2 \end{bmatrix} = q_2 \begin{bmatrix} 1.5 \\ 1 \end{bmatrix}$$

Para que el vector \mathbf{q} sea un vector de probabilidad, la suma de las entradas de \mathbf{q} debe ser igual a 1. Así,

$$q_2 = \frac{1}{1.5 + 1} = 0.4$$

Por consiguiente,

$$\mathbf{q} = \begin{bmatrix} 0.6 \\ 0.4 \end{bmatrix}$$

es el vector de estado estacionario de esta cadena de Markov regular. Esto significa que a largo plazo, el 60 % de los exalumnos harán una donación en un año determinado, y el 40 % no lo harán. Además, estos porcentajes se mantendrán constantes en el tiempo, independientemente del estado inicial de la población. Observe que esto concuerda con el resultado obtenido numéricamente en el ejemplo 8.3.3.

Ejemplo 8.3.8: En el ejemplo 8.3.1, la matriz de transición era

$$P = \begin{bmatrix} 0.8 & 0.3 & 0.2 \\ 0.1 & 0.2 & 0.6 \\ 0.1 & 0.5 & 0.2 \end{bmatrix}$$

Así que el sistema lineal $(I_3 - P)\mathbf{q} = \mathbf{0}$ está dado por

$$\begin{bmatrix} 0.2 & -0.3 & -0.2 \\ -0.1 & 0.8 & -0.6 \\ -0.1 & -0.5 & 0.8 \end{bmatrix} \begin{bmatrix} q_1 \\ q_2 \\ q_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Utilizando el método de reducción de Gauss-Jordan, obtenemos que

$$\begin{bmatrix} 1 & 0 & -\frac{34}{13} \\ 0 & 1 & -\frac{14}{13} \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} q_1 \\ q_2 \\ q_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Por lo que el sistema lineal original es equivalente al sistema

$$\begin{aligned} q_1 &= \frac{34}{13} q_3 \\ q_2 &= \frac{14}{13} q_3 \end{aligned}$$

Es decir, la solución general al sistema está dada por

$$\mathbf{q} = \begin{bmatrix} \frac{34}{13} q_3 \\ \frac{14}{13} q_3 \\ q_3 \end{bmatrix} = q_3 \begin{bmatrix} \frac{34}{13} \\ \frac{14}{13} \\ 1 \end{bmatrix}$$

Para que el vector \mathbf{q} sea un vector de probabilidad, la suma de las entradas de \mathbf{q} debe ser igual a 1. Así,

$$q_3 = \frac{1}{\frac{34}{13} + \frac{14}{13} + 1} = \frac{13}{61}$$

Por consiguiente, el vector de estado estacionario del sistema es

$$\mathbf{q} = \begin{bmatrix} \frac{34}{61} \\ \frac{14}{61} \\ \frac{13}{61} \end{bmatrix} = \begin{bmatrix} 0.5573 \dots \\ 0.2295 \dots \\ 0.2131 \dots \end{bmatrix}$$

Esto concuerda con el resultado obtenido numéricamente en la tabla del ejemplo 8.3.4. Los valores de \mathbf{q} dan las probabilidades a largo plazo de que cualquier coche sea devuelto a la ubicación 1, 2 o 3, respectivamente. Si la agencia de alquiler de coches tiene una flota de 1 000 coches, debería diseñar sus instalaciones de manera que haya al menos 558 espacios en la ubicación 1, al menos 230 espacios en la ubicación 2, y al menos 214 espacios en la ubicación 3.

Ejemplo 8.3.9:

Tomemos la matriz de transición del ejemplo 8.3.5. No daremos los detalles de los cálculos, pero simplemente afirmaremos que la única solución del vector de probabilidad para el sistema lineal $(I_8 - P)\mathbf{q} = \mathbf{0}$ es

$$\mathbf{q} = \frac{1}{28} \begin{bmatrix} 3 \\ 3 \\ 3 \\ 5 \\ 4 \\ 3 \\ 4 \\ 3 \end{bmatrix} = \begin{bmatrix} 0.1071 \dots \\ 0.1071 \dots \\ 0.1071 \dots \\ 0.1785 \dots \\ 0.1428 \dots \\ 0.1071 \dots \\ 0.1428 \dots \\ 0.1071 \dots \end{bmatrix}$$

Esto concuerda con el resultado obtenido numéricamente en la tabla del ejemplo 8.3.5. Los valores de este vector indican la proporción de tiempo que la oficial de tráfico pasa en cada intersección a largo plazo, mostrando que su presencia no se distribuye de manera uniforme. En otras palabras, algunas intersecciones recibirán más atención que otras de manera natural bajo esta estrategia. Por lo tanto, si se busca que pase la misma cantidad de tiempo en todas las intersecciones, la estrategia de moverse aleatoriamente con probabilidades iguales no es adecuada. En su lugar, sería necesario diseñar una estrategia alternativa que ajuste las probabilidades de movimiento entre intersecciones para alcanzar una distribución uniforme, optimizando así el tiempo dedicado en cada una.

8.4 Teoría de grafos

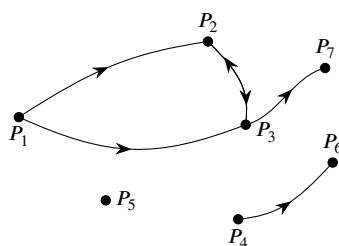
Existen innumerables ejemplos de conjuntos finitos donde sus elementos están relacionados de alguna manera. Por ejemplo, el conjunto podría estar compuesto por personas, animales, países, empresas, equipos deportivos o ciudades. La relación entre dos elementos, A y B , podría ser que la persona A domina a la persona B , el animal A se alimenta del animal B , el país A apoya militarmente al país B , la empresa A vende su producto a la empresa B , el equipo deportivo A siempre vence al equipo deportivo B , o la ciudad A tiene un vuelo directo a la ciudad B .

Ahora mostraremos cómo la teoría de *grafos dirigidos* puede utilizarse para modelar matemáticamente relaciones como estas.

Grafos dirigidos

Un grafo dirigido es un conjunto finito de elementos $\{P_1, P_2, \dots, P_n\}$ junto con una colección finita de pares ordenados (P_i, P_j) de elementos distintos de este conjunto, sin repetir ningún par ordenado. Los elementos del conjunto se llaman vértices, y los pares ordenados se llaman aristas dirigidas del grafo dirigido. Usamos la notación $P_i \rightarrow P_j$ (que se lee “ P_i está conectado a P_j ”) para indicar que la arista dirigida (P_i, P_j) pertenece al grafo dirigido. Geométricamente, podemos visualizar un grafo dirigido (figura 8.9) representando los vértices como puntos en el plano y representando la arista dirigida $P_i \rightarrow P_j$ dibujando una línea o arco desde el vértice P_i hasta el vértice P_j , con una flecha que apunta de P_i a P_j . Si tanto $P_i \rightarrow P_j$ como $P_j \rightarrow P_i$ son ciertos (denotado $P_i \leftrightarrow P_j$), dibujamos una sola línea entre P_i y P_j con dos flechas apuntando en direcciones opuestas (como con P_2 y P_3 en la figura).

Figura 8.9



Como se muestra en la figura 8.9, un grafo dirigido puede tener componentes separadas de vértices que están conectados solo entre sí; y algunos vértices, como P_5 , pueden no estar conectados con ningún otro vértice. Además, dado que $P_i \rightarrow P_i$ no está permitido en un grafo dirigido, un vértice no puede estar conectado consigo mismo por un solo arco que no pase por ningún otro vértice.

La figura 8.10 muestra diagramas que representan tres ejemplos más de grafos dirigidos.

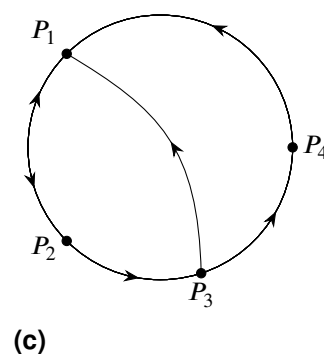
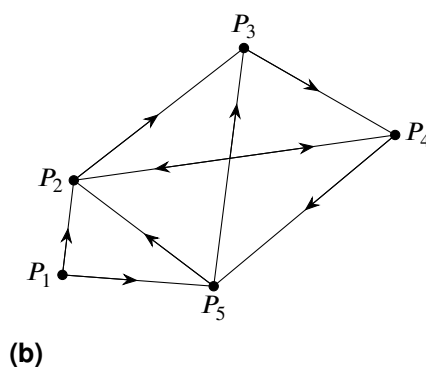
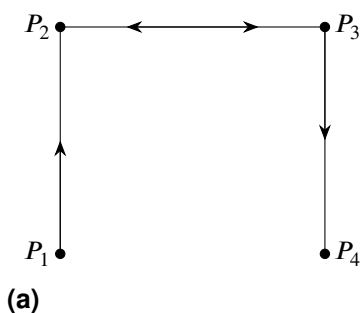


Figura 8.10

Con un grafo dirigido de n vértices, podemos asociar una matriz de tamaño $n \times n$ denotada como $M = \llbracket m_{ij} \rrbracket$, llamada *matriz de vértices* del grafo dirigido. Sus elementos se definen por

$$m_{ij} = \begin{cases} 1, & \text{si } P_i \rightarrow P_j \\ 0, & \text{en cualquier otro caso} \end{cases}$$

para $i, j = 1, 2, \dots, n$. Para los tres grafos dirigidos en la figura 8.10, las matrices de vértices correspondientes son:

Figura 8.10a: $M = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

Figura 8.10b: $M = \begin{bmatrix} 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 \end{bmatrix}$

Figura 8.10c: $M = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}$

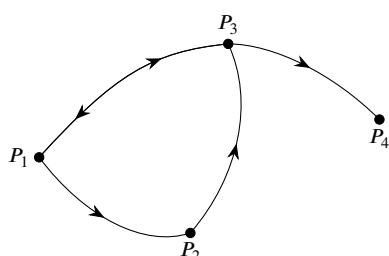


Figura 8.11

Por definición, las matrices de vértices tienen las siguientes dos propiedades:

- i) Todos los elementos son 0 o 1.
- ii) Todos los elementos en la diagonal son 0.

Inversamente, cualquier matriz con estas dos propiedades determina un grafo dirigido único que tiene dicha matriz como su matriz de vértices. Por ejemplo,

$$M = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

determina el grafo dirigido en la figura 8.11.

Ejemplo 8.4.1:

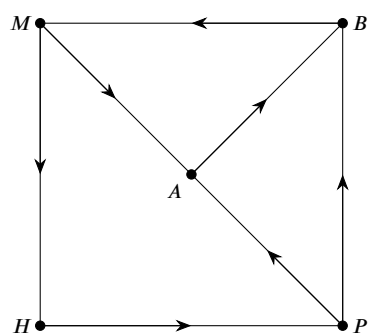


Figura 8.12

Una familia está compuesta por una madre (M), un padre (P), una hija (H) y dos hijos, el mayor (A) y el menor (B). La madre influye en la hija y en el hijo mayor, mientras que el padre influye en ambos hijos. La hija influye en el padre, el hijo mayor en el menor y el hijo menor en la madre. Podemos modelar estas influencias con un grafo dirigido, donde los vértices representan a los miembros de la familia y una flecha $A \rightarrow B$ indica que A influye en B . La figura 8.12 muestra este grafo, y su matriz de vértices es la siguiente

	M	P	H	A	B
M	0	0	1	1	0
P	0	0	0	1	1
H	0	1	0	0	0
A	0	0	0	0	1
B	1	0	0	0	0

Ejemplo 8.4.2:

En ajedrez, el caballo se mueve en un patrón en forma de “L” sobre el tablero. Para el tablero en la figura 8.13, puede moverse horizontalmente dos casillas y luego verticalmente una casilla, o puede moverse verticalmente dos casillas y luego horizontalmente una casilla. Así, desde la casilla central en la figura, el caballo puede moverse a cualquiera de las ocho casillas sombreadas marcadas. Supongamos que el caballo está restringido a las nueve casillas numeradas en la figura 8.14. Si con $i \rightarrow j$ se quiere decir que el caballo puede moverse de la casilla

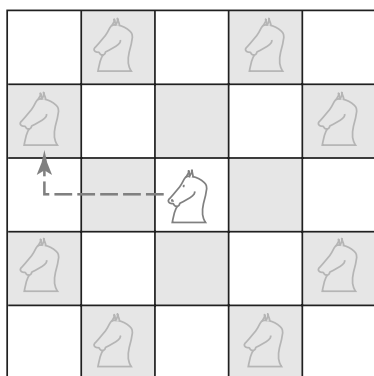


Figura 8.13

1	2	3
4	5	6
7	8	9

Figura 8.14

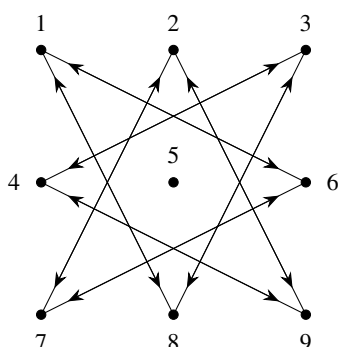


Figura 8.15

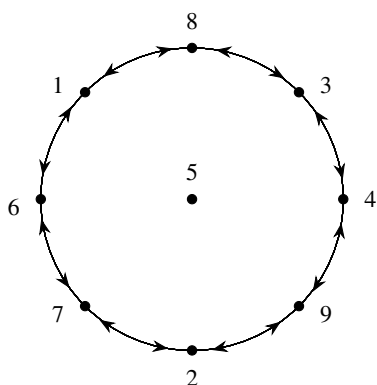


Figura 8.16

i a la casilla j , el grafo dirigido en la figura 8.15 ilustra todos los movimientos posibles que el caballo puede realizar entre estas nueve casillas. En la figura 8.16 hemos “desenredado” la figura 8.15 para hacer más claro el patrón de movimientos posibles. La matriz de vértices de este grafo dirigido es

$$M = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

En el ejemplo 8.4.1, el padre no puede influir directamente a la madre; es decir, la conexión $P \rightarrow M$ no existe. Sin embargo, puede influir en el hijo menor, quien a su vez puede influir en la madre. Esta relación indirecta se representa como $P \rightarrow B \rightarrow M$ y se denomina una conexión de 2 pasos de P a M , que muestra una influencia mediada por un vértice intermedio, en este caso, el hijo menor. De manera similar, la conexión $M \rightarrow H$ representa un camino de 1 paso, indicando una influencia directa de la madre hacia la hija. Por otro lado, un camino más largo como $P \rightarrow A \rightarrow B \rightarrow M$ describe una conexión de 3 pasos, donde el padre influye en la hermana mayor, quien a su vez influye en el hijo menor, y este último finalmente influye en la madre.

Consideremos ahora una técnica para determinar el número de todas las posibles conexiones de r pasos ($r = 1, 2, \dots$) desde un vértice P_i hasta otro vértice P_j en un grafo dirigido arbitrario, incluyendo el caso en que P_i y P_j sean el mismo vértice. Para una conexión de 1 paso desde P_i hasta P_j , basta observar el elemento m_{ij} de la matriz de vértices M . Si $m_{ij} = 1$, existe una conexión directa; en caso contrario, no existe tal conexión. Así, m_{ij} determina el número de conexiones de 1 paso entre P_i y P_j .

Para calcular el número de conexiones de 2 pasos, consideramos el cuadrado de la matriz de vértices, M^2 . Si designamos $m_{ij}^{(2)}$ como el elemento ij -ésimo de M^2 , se tiene que

$$m_{ij}^{(2)} = m_{i1}m_{1j} + m_{i2}m_{2j} + \dots + m_{in}m_{nj} \quad (8.4.1)$$

En esta expresión, cada término $m_{ik}m_{kj}$ representa una posible conexión de 2 pasos que pasa por el vértice intermedio P_k . Por ejemplo, si $m_{i1} = m_{1j} = 1$, entonces existe una conexión de 2 pasos $P_i \rightarrow P_1 \rightarrow P_j$. Si, por el contrario, $m_{i1} = 0$ o $m_{1j} = 0$, dicha conexión no es posible. De manera general, $P_i \rightarrow P_k \rightarrow P_j$ es una conexión válida de 2 pasos si, y solo si, $m_{ik}m_{kj} = 1$. Por lo tanto, la suma de estos términos en el lado derecho de la ecuación representa el número total de conexiones de 2 pasos desde P_i hasta P_j .

Este razonamiento puede extenderse para conexiones de r pasos. Para obtener el número total de conexiones de r pasos, simplemente calculamos la potencia M^r de la matriz de vértices. En este caso, el elemento ij -ésimo de M^r , denotado como $m_{ij}^{(r)}$, representa el número total de caminos de r pasos entre los vértices P_i y P_j .

En general, este enfoque nos lleva al siguiente resultado.

Teorema 8.4.1:

Sea M la matriz de vértices de un grafo dirigido y $m_{ij}^{(r)}$ el elemento ij -ésimo de M^r . Entonces, $m_{ij}^{(r)}$ es igual al número de conexiones de r pasos desde P_i hasta P_j .

Ejemplo 8.4.3:

La figura 8.17 es el mapa de rutas de una pequeña aerolínea que opera en las cuatro ciudades P_1 , P_2 , P_3 , y P_4 . Como un grafo dirigido, su matriz de vértices es

$$M = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix}$$

y además, tenemos que

$$M^2 = \begin{bmatrix} 2 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 0 & 2 & 2 & 0 \\ 2 & 0 & 1 & 1 \end{bmatrix} \quad \text{y} \quad M^3 = \begin{bmatrix} 1 & 3 & 3 & 1 \\ 2 & 2 & 3 & 1 \\ 4 & 0 & 2 & 2 \\ 1 & 3 & 3 & 1 \end{bmatrix}$$

Si estamos interesados en el número de conexiones de la ciudad P_4 a la ciudad P_3 , podemos usar el teorema 8.4.1. Puesto que $m_{43} = 1$, hay una conexión de un paso; dado que $m_{43}^{(2)} = 1$, hay una conexión de dos pasos; y como $m_{43}^{(3)} = 3$, entonces hay tres conexiones de tres pasos. Para verificar lo dicho anteriormente, en la figura 8.17 encontramos que

Conexiones de un paso de P_4 a P_3 :

$$P_4 \rightarrow P_3$$

Conexiones de dos pasos de P_4 a P_3 :

$$P_4 \rightarrow P_2 \rightarrow P_3$$

Conexiones de tres pasos de P_4 a P_3 :

$$P_4 \rightarrow P_3 \rightarrow P_4 \rightarrow P_3$$

$$P_4 \rightarrow P_2 \rightarrow P_1 \rightarrow P_3$$

$$P_4 \rightarrow P_3 \rightarrow P_1 \rightarrow P_3$$

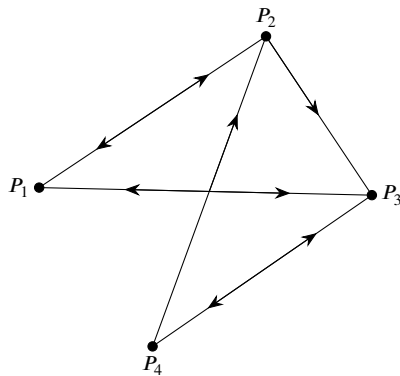


Figura 8.17

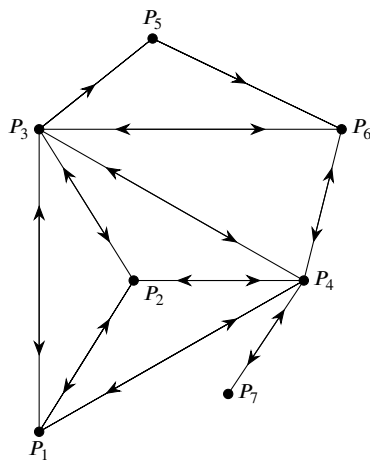


Figura 8.18

Definición 8.4.1:

Un subconjunto de un grafo dirigido se llama “clique” si cumple con las siguientes tres condiciones:

- i) El subconjunto contiene al menos tres vértices.
- ii) Para cada par de vértices P_i y P_j en el subconjunto, se cumplen tanto $P_i \rightarrow P_j$ como $P_j \rightarrow P_i$.
- iii) El subconjunto es lo más grande posible; es decir, no es posible agregar otro vértice al subconjunto y seguir cumpliendo la condición (ii).

Esta definición implica que una clique es un subconjunto máximo de vértices que están en perfecta “comunicación” entre sí, formando una estructura altamente conectada. Por ejemplo, si los vértices representan ciudades y $P_i \rightarrow P_j$ indica que hay un vuelo directo entre la ciudad P_i y la ciudad P_j , entonces dentro de una clique, cada par de ciudades tiene un vuelo directo en ambas direcciones, garantizando una conectividad total entre ellas. Esto significa que cualquier ciudad dentro de una clique puede acceder de manera inmediata a cualquier otra sin necesidad de escalas.

Ejemplo 8.4.4:

El grafo dirigido ilustrado en la figura 8.18 tiene dos cliques:

$$\{P_1, P_2, P_3, P_4\} \quad \text{y} \quad \{P_3, P_4, P_6\}$$

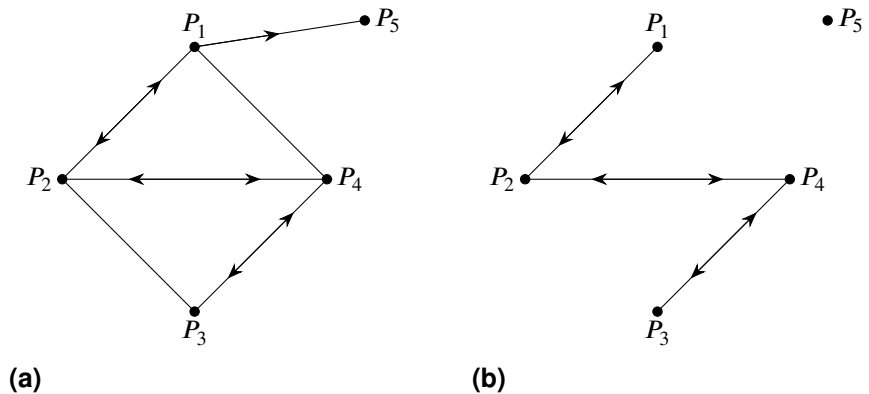
Este ejemplo muestra que un grafo dirigido puede contener varios cliques y que un vértice puede pertenecer a más de un clique simultáneamente.

Para grafos dirigidos simples, las cliques se pueden encontrar por inspección. Sin embargo, en grafos dirigidos grandes, es conveniente contar con un procedimiento sistemático para detectarlas. Para ello, definimos la matriz $S = \llbracket s_{ij} \rrbracket$ relacionada con el grafo dirigido de la siguiente manera:

$$s_{ij} = \begin{cases} 1, & \text{si } P_i \leftrightarrow P_j \\ 0, & \text{en cualquier otro caso} \end{cases}$$

La matriz S determina un grafo dirigido igual al original, pero eliminando las aristas dirigidas con una sola flecha. Por ejemplo, si el grafo original es el de la figura 8.19a, el grafo que tiene a S como matriz de vértices se muestra en la figura 8.19b.

Figura 8.19



La matriz S puede obtenerse de la matriz de vértices M del grafo dirigido original estableciendo $s_{ij} = 1$ si $m_{ij} = m_{ji} = 1$ y estableciendo $s_{ij} = 0$ en caso contrario.

El siguiente teorema, que utiliza la matriz S , es útil para identificar cliques.

Teorema 8.4.2: Sea $s_{ij}^{(3)}$ el elemento ij -ésimo de la matriz S^3 . Entonces, un vértice P_i pertenece a alguna clique si y solo si $s_{ii}^{(3)} \neq 0$.

Demostración: Si $s_{ii}^{(3)} \neq 0$, entonces existe al menos una conexión de tres pasos desde P_i hacia sí mismo en el grafo dirigido modificado determinado por S , por ejemplo, $P_i \rightarrow P_j \rightarrow P_k \rightarrow P_i$. Como todas las relaciones en este grafo son bidireccionales, también se tiene $P_i \leftrightarrow P_j \leftrightarrow P_k \leftrightarrow P_i$. Esto implica que $\{P_i, P_j, P_k\}$ forma una clique o un subconjunto de una, por lo que P_i pertenece a alguna clique. La afirmación recíproca, “si P_i pertenece a una clique, entonces $s_{ii}^{(3)} \neq 0$ ”, se demuestra de forma similar.

Ejemplo 8.4.5: Supongamos que un grafo dirigido tiene como su matriz de vértices

$$M = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

Entonces

$$S = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \quad \text{y} \quad S^3 = \begin{bmatrix} 0 & 3 & 0 & 2 \\ 3 & 0 & 2 & 0 \\ 0 & 2 & 0 & 1 \\ 2 & 0 & 1 & 0 \end{bmatrix}$$

Dado que todas las entradas diagonales de S^3 son cero, se deduce del teorema 8.4.2 que el grafo dirigido no tiene cliques.

Ejemplo 8.4.6:

Supongamos que un grafo dirigido tiene como su matriz de vértices

$$M = \begin{bmatrix} 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \end{bmatrix}$$

Entonces

$$S = \begin{bmatrix} 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix} \quad \text{y} \quad S^3 = \begin{bmatrix} 2 & 4 & 0 & 4 & 3 \\ 4 & 2 & 0 & 3 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 4 & 3 & 0 & 2 & 1 \\ 3 & 1 & 0 & 1 & 0 \end{bmatrix}$$

Las entradas no nulas de la diagonal de S^3 son $s_{11}^{(3)}$, $s_{22}^{(3)}$ y $s_{44}^{(3)}$. En consecuencia, en el grafo dirigido dado, P_1 , P_2 y P_4 pertenecen a cliques. Dado que una clique debe contener al menos tres vértices, el grafo dirigido tiene solo una clique, $\{P_1, P_2, P_4\}$.

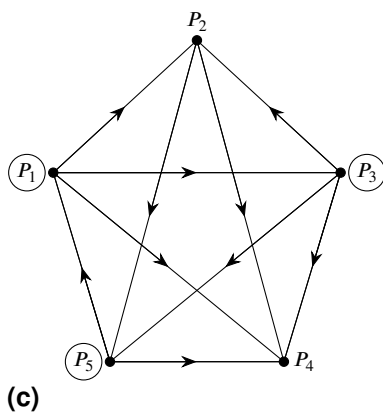
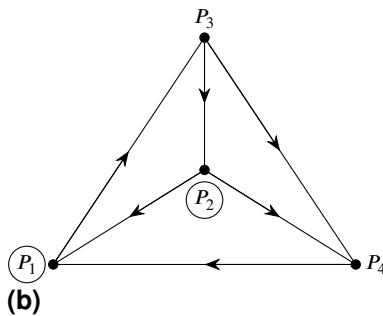
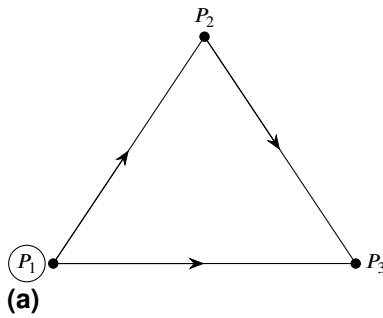


Figura 8.20

Grafos dirigidos de dominancia

En muchos grupos de individuos o animales, existe un “orden jerárquico” o una relación de dominancia que se establece claramente entre cualquier par de miembros del grupo. Esto significa que, dado un par de individuos A y B , necesariamente se cumple que uno domina al otro: o bien A domina a B , o bien B domina a A , pero no es posible que ambos se dominen mutuamente de manera simultánea. Esta relación jerárquica puede representarse gráficamente mediante un grafo dirigido, donde una flecha $P_i \rightarrow P_j$ indica que P_i domina a P_j . En este contexto, la estructura jerárquica implica que para cualquier par de vértices distintos en el grafo, se cumple que o bien $P_i \rightarrow P_j$, o bien $P_j \rightarrow P_i$, pero nunca ambas direcciones de forma simultánea. Esta propiedad refleja un orden total en las relaciones de dominancia dentro del grupo, lo que permite organizar jerárquicamente a los miembros en función de sus interacciones. Tal estructura puede ser de utilidad en diversas áreas, como la etología, la sociología o incluso en el análisis de sistemas de toma de decisiones en grupos. Esta idea se formaliza de manera precisa en la siguiente definición, que nos permite capturar y entender de forma rigurosa este concepto en distintos contextos.

Definición 8.4.2:

Un *grafo dirigido de dominancia* es un grafo dirigido tal que, para cualquier par de vértices distintos P_i y P_j , o bien $P_i \rightarrow P_j$ o $P_j \rightarrow P_i$, pero no ambos.

Un ejemplo clásico de un grafo dirigido que satisface esta definición es una liga de n equipos deportivos que se enfrentan entre sí exactamente una vez, como en una ronda de un torneo de todos contra todos en la que no se permiten empates. Si interpretamos $P_i \rightarrow P_j$ como que el equipo P_i venció al equipo P_j en su único encuentro, resulta evidente que se cumple la definición de un grafo dirigido de dominancia. Por esta razón, a este tipo de grafos dirigidos también se les denomina *torneos*, por su similitud con competiciones deportivas.

La figura 8.20 muestra ejemplos de grafos dirigidos de dominancia con tres, cuatro y cinco vértices, respectivamente. En estos grafos, los vértices encerrados en un círculo poseen una propiedad notable: desde cada uno de ellos es posible llegar a cualquier otro vértice del grafo mediante una conexión de uno o dos pasos. En el contexto de un torneo deportivo, estos vértices representan a los equipos más “dominantes” o “poderosos”. Esto significa que dichos equipos vencen directamente a cualquier otro equipo o, al menos, vencen a un equipo que, a su vez, derrotó al equipo en cuestión.

Teorema 8.4.3: En cualquier grafo dirigido de dominancia, existe al menos un vértice desde el cual hay una conexión de un paso o de dos pasos hacia cualquier otro vértice.

Demostración: Consideremos un vértice (puede haber varios) con el mayor número total de conexiones de un paso y de dos pasos hacia otros vértices en el grafo. Al reenumerar los vértices, podemos asumir que P_1 es tal vértice. Supongamos que existe algún vértice P_i tal que no hay ninguna conexión de un paso o de dos pasos desde P_1 hacia P_i . Entonces, en particular, $P_1 \rightarrow P_i$ no es cierto, por lo que, según la definición de un grafo dirigido de dominancia, debe ser que $P_i \rightarrow P_1$. Luego, sea P_k cualquier vértice tal que $P_1 \rightarrow P_k$ es cierto. Entonces, no podemos tener $P_k \rightarrow P_i$, ya que $P_1 \rightarrow P_k \rightarrow P_i$ sería una conexión de dos pasos desde P_1 hacia P_i . Por lo tanto, debe ser que $P_i \rightarrow P_k$. Es decir, P_i tiene conexiones de un paso con todos los vértices a los que P_1 tiene conexiones de un paso. El vértice P_i debe entonces también tener conexiones de dos pasos con todos los vértices a los que P_1 tiene conexiones de dos pasos. Pero como, además, tenemos que $P_i \rightarrow P_1$, esto significa que P_i tiene más conexiones de un paso y de dos pasos hacia otros vértices que P_1 . Sin embargo, esto contradice la forma en que se eligió P_1 . Por lo tanto, no puede haber un vértice P_i al cual P_1 no tenga ninguna conexión de un paso o de dos pasos.

Esta demostración muestra que un vértice con el mayor número total de conexiones de uno y dos pasos hacia otros vértices cumple con la propiedad del teorema. Para identificar dichos vértices, se utiliza la matriz de vértices M y su cuadrado M^2 . La suma de los elementos en la i -ésima fila de M indica el total de conexiones directas desde P_i , mientras que la suma en la i -ésima fila de M^2 representa las conexiones de dos pasos. Por lo tanto, la suma de los elementos en la i -ésima fila de $A = M + M^2$ da el total de conexiones directas y de dos pasos desde P_i , y la fila con la mayor suma identifica un vértice que satisface el teorema.

Ejemplo 8.4.7: Supongamos que cinco equipos de béisbol juegan entre sí exactamente una vez, y los resultados son los indicados en el grafo dirigido de dominancia de la figura 8.21. La matriz de vértices del grafo es

$$M = \begin{bmatrix} 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \end{bmatrix}$$

así que

$$A = M + M^2 = \begin{bmatrix} 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 2 & 3 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 2 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 1 & 2 & 0 \\ 2 & 0 & 3 & 3 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 2 & 3 & 0 \end{bmatrix}$$

Las sumas de las filas de A son respectivamente

$$\text{Suma de la 1}^\circ \text{ fila: } 0 + 1 + 1 + 2 + 0 = 4$$

$$\text{Suma de la 2}^\circ \text{ fila: } 2 + 0 + 3 + 3 + 1 = 9$$

$$\text{Suma de la 3}^\circ \text{ fila: } 0 + 1 + 0 + 1 + 0 = 2$$

$$\text{Suma de la 4}^\circ \text{ fila: } 1 + 1 + 1 + 0 + 1 = 4$$

$$\text{Suma de la 5}^\circ \text{ fila: } 1 + 1 + 2 + 3 + 0 = 7$$

Debido a que la segunda fila tiene la mayor suma, el vértice P_2 debe tener una conexión de 1 o 2 pasos con cualquier otro vértice. Esto se puede verificar fácilmente en la figura 8.21.

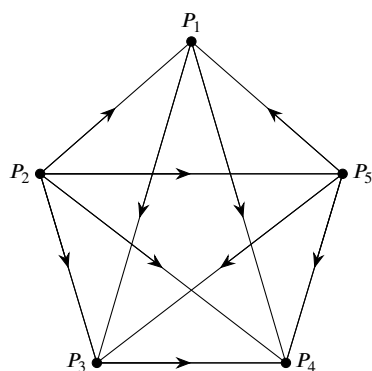


Figura 8.21

Hemos sugerido de manera informal que un vértice con el mayor número de conexiones de 1 y 2 pasos a otros vértices es un vértice “poderoso”. Podemos formalizar este concepto con la siguiente definición.

Definición 8.4.3:

El *poder* de un vértice en un grafo dirigido de dominancia es el número total de conexiones de 1 y 2 pasos desde él hacia otros vértices. Alternativamente, el poder de un vértice P_i es la suma de las entradas de la fila i de la matriz $A = M + M^2$, donde M es la matriz de vértices del grafo dirigido.

Ejemplo 8.4.8:

Clasifiquemos los cinco equipos de béisbol en el ejemplo anterior según sus poderes. De los cálculos de las sumas de filas en ese ejemplo, tenemos:

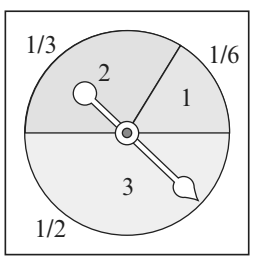
- Poder del equipo $P_1 = 4$
- Poder del equipo $P_2 = 9$
- Poder del equipo $P_3 = 2$
- Poder del equipo $P_4 = 4$
- Poder del equipo $P_5 = 7$

Por lo tanto, la clasificación de los equipos según sus poderes sería:

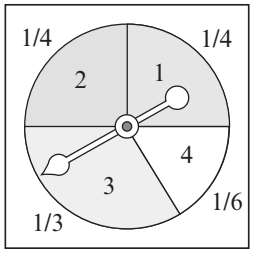
- 1° lugar = P_2
- 2° lugar = P_5
- 3° lugar = P_1 y P_4
- 5° lugar = P_3

8.5 Juegos de estrategia

Para introducir los conceptos básicos en la teoría de juegos, consideraremos el siguiente juego de tipo carnaval al que dos personas acuerdan jugar. Llamaremos a los participantes jugador R y jugador C . Cada jugador tiene una rueda estacionaria con un puntero móvil, como se muestra en la figura 8.22. Por razones que se aclararán más adelante, llamaremos a la rueda del jugador R la rueda de filas y a la rueda del jugador C la rueda de columnas. La rueda de filas está dividida en tres sectores numerados 1, 2 y 3, mientras que la rueda de columnas está dividida en cuatro sectores numerados 1, 2, 3 y 4. Las fracciones del área ocupada por los diferentes sectores están indicadas en la figura. Para jugar, cada jugador gira el puntero de su rueda y lo deja reposar al azar. El número del sector en el que cada puntero se detiene se llama el movimiento de ese jugador. Así, el jugador R tiene tres posibles movimientos y el jugador C tiene cuatro. Dependiendo del movimiento que realice cada jugador, el jugador C hará un pago en dinero al jugador R según la siguiente tabla.



Rueda de filas del jugador R



Rueda de columnas del jugador C

Figura 8.22

		Movimiento del jugador C			
		1	2	3	4
Movimiento del jugador R	1	\$3	\$5	-\$2	-\$1
	2	-\$2	\$4	-\$3	-\$4
	3	\$6	-\$5	\$0	\$3

Tabla 8.3 Pago al jugador R

Por ejemplo, si el puntero de la rueda de filas se detiene en el sector 1 (el jugador R realiza el movimiento 1) y el puntero de la rueda de columnas se detiene en el sector 2 (el jugador C realiza el movimiento 2), entonces el jugador C debe pagar al jugador R la suma de \$5. Algunas de las entradas en esta tabla son negativas, lo

que indica que el jugador C realiza un pago negativo al jugador R . Esto significa que el jugador R realiza un pago positivo al jugador C . Por ejemplo, si la rueda de filas muestra 2 y la rueda de columnas muestra 4, entonces el jugador R paga al jugador C la suma de \$4, porque la entrada correspondiente en la tabla es $-\$4$. De esta manera, las entradas positivas de la tabla representan las ganancias del jugador R y las pérdidas del jugador C , mientras que las entradas negativas representan las ganancias del jugador C y las pérdidas del jugador R .

En este juego, los jugadores no tienen control sobre sus movimientos; cada movimiento está determinado por el azar. Sin embargo, si cada jugador puede decidir si desea o no jugar, entonces cada uno querrá saber cuánto puede esperar ganar o perder a largo plazo si decide jugar.

Juegos de matriz de suma cero para dos personas

El juego descrito anteriormente es un ejemplo de un *juego de matriz de suma cero para dos personas*. El término *suma cero* significa que en cada jugada del juego, la ganancia positiva de un jugador es igual a la pérdida negativa (ganancia negativa) del otro jugador. Es decir, la suma de las dos ganancias es cero. El término *juego de matriz* se utiliza para describir un juego de dos personas en el que cada jugador tiene un número finito de movimientos, de modo que todos los posibles resultados de cada jugada, y las correspondientes ganancias de los jugadores, pueden representarse en forma tabular o matricial, como en la tabla 8.3.

En un juego general de este tipo, sea R el jugador que tiene m posibles movimientos y C el jugador que tiene n posibles movimientos. En una jugada del juego, cada jugador realiza uno de sus posibles movimientos, y luego se realiza un *pago* del jugador C al jugador R , dependiendo de los movimientos. Para $i = 1, 2, \dots, m$ y $j = 1, 2, \dots, n$, definimos

a_{ij} = pago que el jugador C realiza al jugador R si el jugador R
hace el movimiento i y el jugador C hace el movimiento j .

Este pago no necesariamente tiene que ser dinero; puede ser cualquier tipo de mercancía a la cual se le pueda asignar un valor numérico. Como antes, si una entrada a_{ij} es negativa, esto significa que el jugador C recibe un pago de $|a_{ij}|$ del jugador R . Organizamos estos mn pagos posibles en forma de una matriz $m \times n$

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

a la que llamaremos *matriz de pagos* del juego.

Cada jugador realiza sus movimientos sobre una base probabilística. Por ejemplo, en el juego discutido en la introducción, la proporción del área de un sector respecto al área de la rueda sería la probabilidad de que el jugador realice el movimiento correspondiente a ese sector. Así, en la figura 8.22, vemos que el jugador R realizaría el movimiento 2 con probabilidad $1/3$, y el jugador C realizaría el movimiento 2 con probabilidad $1/4$. En el caso general, hacemos las siguientes definiciones:

p_i = probabilidad de que el jugador R realice el movimiento i ($i = 1, 2, \dots, m$),
 q_j = probabilidad de que el jugador C realice el movimiento j ($j = 1, 2, \dots, n$).

Se deduce de estas definiciones que

$$p_1 + p_2 + \cdots + p_m = 1$$

y

$$q_1 + q_2 + \cdots + q_n = 1.$$

Con las probabilidades p_i y q_j formamos dos vectores:

$$\mathbb{p} = [p_1 \quad p_2 \quad \cdots \quad p_m] \quad \text{y} \quad \mathbb{q} = \begin{bmatrix} q_1 \\ q_2 \\ \vdots \\ q_n \end{bmatrix}.$$

Llamamos al vector fila \mathbb{p} la *estrategia del jugador R* y al vector columna \mathbb{q} la *estrategia del jugador C*. Por ejemplo, en el caso del juego de carnaval descrito anteriormente, a partir de la figura 8.22 tenemos

$$\mathbb{p} = \begin{bmatrix} \frac{1}{6} & \frac{1}{3} & \frac{1}{2} \end{bmatrix} \quad \text{y} \quad \mathbb{q} = \begin{bmatrix} \frac{1}{4} \\ \frac{1}{4} \\ \frac{1}{3} \\ \frac{1}{6} \end{bmatrix}.$$

De la teoría de probabilidad, si la probabilidad de que el jugador R haga el movimiento i es p_i , y de manera independiente la probabilidad de que el jugador C haga el movimiento j es q_j , entonces $p_i q_j$ es la probabilidad de que, en una jugada, R haga el movimiento i y C el movimiento j . El pago a R por este par de movimientos es a_{ij} . Multiplicando cada pago posible por su probabilidad y sumando todos los casos, obtenemos la expresión:

$$a_{11}p_1q_1 + a_{12}p_1q_2 + \cdots + a_{1n}p_1q_n + a_{21}p_2q_1 + a_{22}p_2q_2 + \cdots + a_{mn}p_mq_n. \quad (8.5.1)$$

La ecuación (8.5.1) representa un promedio ponderado de los pagos al jugador R , donde cada pago se pondera según su probabilidad de ocurrencia. En probabilidad, este promedio se denomina *pago esperado* al jugador R . Si el juego se repite muchas veces, el pago promedio por jugada a largo plazo para R está dado por esta expresión. Denotamos el pago esperado como $E(\mathbb{p}, \mathbb{q})$ para resaltar su dependencia de las estrategias de ambos jugadores. Usando la matriz de pagos A y las estrategias \mathbb{p} y \mathbb{q} , podemos expresar el pago esperado en forma matricial como

$$E(\mathbb{p}, \mathbb{q}) = [p_1 \quad p_2 \quad \cdots \quad p_m] \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} q_1 \\ q_2 \\ \vdots \\ q_n \end{bmatrix} = \mathbb{p}A\mathbb{q}. \quad (8.5.2)$$

Dado que $E(\mathbb{p}, \mathbb{q})$ es el pago esperado al jugador R , se deduce que $-E(\mathbb{p}, \mathbb{q})$ es el pago esperado al jugador C .

Ejemplo 8.5.1: Para el juego de carnaval descrito anteriormente, tenemos

$$E(\mathbb{p}, \mathbb{q}) = \mathbb{p}A\mathbb{q} = \begin{bmatrix} \frac{1}{6} & \frac{1}{3} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 3 & 5 & -2 & -1 \\ -2 & 4 & -3 & -4 \\ 6 & -5 & 0 & 3 \end{bmatrix} \begin{bmatrix} \frac{1}{4} \\ \frac{1}{4} \\ \frac{1}{3} \\ \frac{1}{6} \end{bmatrix} = \frac{13}{72} \approx 0.1805 \dots$$

Por lo tanto, a largo plazo, el jugador R puede esperar recibir un promedio de aproximadamente 18 centavos del jugador C en cada jugada del juego.

Hasta ahora hemos discutido la situación en la que cada jugador tiene una estrategia predeterminada. Ahora consideraremos una situación más difícil en la que ambos jugadores pueden cambiar sus estrategias de manera independiente. Por ejemplo, en el juego descrito en la introducción, permitiríamos que ambos jugadores alteren las áreas de los sectores de sus ruedas y, de este modo, controlen las probabilidades de sus respectivos movimientos. Esto cambia cualitativamente la naturaleza del problema y nos sitúa firmemente en el campo de la teoría de juegos propiamente dicha. Se entiende que ninguno de los jugadores sabe qué estrategia elegirá el otro. También se asume que cada jugador hará la mejor elección posible de estrategia y que el otro jugador lo sabe. Así, el jugador R intenta elegir una estrategia p tal que $E(p, q)$ sea lo más grande posible para la mejor estrategia q que el jugador C pueda elegir; de manera similar, el jugador C intenta elegir una estrategia q tal que $E(p, q)$ sea lo más pequeño posible para la mejor estrategia p que el jugador R pueda elegir. Para ver que estas elecciones son realmente posibles, necesitaremos el siguiente teorema, llamado el *teorema fundamental de juegos de suma cero de dos jugadores*.

Teorema 8.5.1: Existen estrategias p^* y q^* tales que

$$E(p^*, q) \geq E(p^*, q^*) \geq E(p, q^*) \quad (8.5.3)$$

para todas las estrategias p y q .

Las estrategias p^* y q^* en este teorema son las mejores estrategias posibles para los jugadores R y C , respectivamente. Para ver por qué esto es así, sea $v = E(p^*, q^*)$. La desigualdad del lado izquierdo de la ecuación (8.5.3) se escribe entonces como

$$E(p^*, q) \geq v \quad \text{para todas las estrategias } q.$$

Esto significa que si el jugador R elige la estrategia p^* , entonces, sin importar qué estrategia elija el jugador C , el pago esperado para el jugador R nunca será menor que v . Además, no es posible que el jugador R logre un pago esperado mayor que v . Para ver por qué, supongamos que existe alguna estrategia p^{**} que el jugador R puede elegir tal que

$$E(p^{**}, q) > v \quad \text{para todas las estrategias } q.$$

Entonces, en particular,

$$E(p^{**}, q^*) > v.$$

Pero esto contradice la desigualdad del lado derecho de la ecuación (8.5.3), que requiere que $v \geq E(p^{**}, q^*)$. En consecuencia, lo mejor que puede hacer el jugador R es evitar que su pago esperado caiga por debajo del valor v . De manera similar, lo mejor que puede hacer el jugador C es asegurar que el pago esperado del jugador R no exceda v , y esto se puede lograr utilizando la estrategia q^* .

Definición 8.5.1:

Si p^* y q^* son estrategias tales que

$$E(p^*, q) \geq E(p^*, q^*) \geq E(p, q^*) \quad (8.5.4)$$

para todas las estrategias p y q , entonces:

- i) p^* se llama una *estrategia óptima para el jugador R* .
- ii) q^* se llama una *estrategia óptima para el jugador C* .
- iii) $v = E(p^*, q^*)$ se llama el *valor del juego*.

La redacción de esta definición sugiere que las estrategias óptimas no son necesariamente únicas. Sin embargo, se puede probar que cualquier par de conjuntos de estrategias óptimas siempre resulta en el mismo valor v del juego. Es decir, si p^*, q^* y p^{**}, q^{**} son estrategias óptimas, entonces

$$E(p^*, q^*) = E(p^{**}, q^{**}) \quad (8.5.5)$$

El valor de un juego es, por lo tanto, el pago esperado al jugador R cuando ambos jugadores eligen cualquier estrategia óptima.

Para encontrar estrategias óptimas, debemos encontrar los vectores \mathbf{p}^* y \mathbf{q}^* que satisfacen la ecuación (8.5.4). Esto se realiza generalmente utilizando técnicas de programación lineal. A continuación, discutimos casos especiales en los cuales las estrategias óptimas pueden encontrarse mediante técnicas más elementales.

Definición 8.5.2:

Un elemento a_{rs} en una matriz de pagos A se llama *punto silla* si:

- i) a_{rs} es el elemento más pequeño en su fila, y
- ii) a_{rs} es el elemento más grande en su columna.

Un juego cuya matriz de pagos tiene un punto silla se llama *estrictamente determinado*.

Por ejemplo, el elemento sombreado en cada una de las siguientes matrices de pagos es un punto silla:

$$\begin{bmatrix} 3 & 1 \\ -4 & 0 \end{bmatrix}, \quad \begin{bmatrix} -50 & 30 & -5 \\ 90 & 60 & 75 \\ 60 & -10 & -30 \end{bmatrix}, \quad \begin{bmatrix} 0 & -3 & 5 & -9 \\ 15 & -8 & -2 & 10 \\ 7 & 10 & 6 & 9 \\ 6 & 11 & -3 & 2 \end{bmatrix}.$$

Si una matriz tiene un punto silla a_{rs} , resulta que las siguientes estrategias son estrategias óptimas para los dos jugadores:

$$\mathbf{p}^* = [0 \quad 0 \quad \cdots \quad 1 \quad \cdots \quad 0], \quad \mathbf{q}^* = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix}$$

\nearrow
 r -ésima entrada

\nwarrow
 s -ésima entrada

Es decir, una estrategia óptima para el jugador R es siempre realizar el movimiento r , y una estrategia óptima para el jugador C es siempre realizar el movimiento s . Estas estrategias en las que solo un movimiento es posible se denominan *estrategias puras*. Las estrategias para las cuales más de un movimiento es posible se denominan *estrategias mixtas*. Para demostrar que las estrategias puras anteriores son óptimas, puede verificarse que se cumplen las siguientes tres ecuaciones:

$$E(\mathbf{p}^*, \mathbf{q}^*) = \mathbf{p}^* A \mathbf{q}^* = a_{rs} \quad (8.5.6)$$

$$E(\mathbf{p}^*, \mathbf{q}) = \mathbf{p}^* A \mathbf{q} \geq a_{rs} \quad \text{para cualquier estrategia } \mathbf{q} \quad (8.5.7)$$

$$E(\mathbf{p}, \mathbf{q}^*) = \mathbf{p} A \mathbf{q}^* \leq a_{rs} \quad \text{para cualquier estrategia } \mathbf{p} \quad (8.5.8)$$

En conjunto, estas tres ecuaciones implican que

$$E(\mathbf{p}^*, \mathbf{q}) \geq E(\mathbf{p}^*, \mathbf{q}^*) \geq E(\mathbf{p}, \mathbf{q}^*)$$

para todas las estrategias \mathbf{p} y \mathbf{q} . Dado que esto es exactamente la ecuación (8.5.4), se deduce que \mathbf{p}^* y \mathbf{q}^* son estrategias óptimas.

De la ecuación (8.5.6), el valor de un juego estrictamente determinado es simplemente el valor numérico del punto silla a_{rs} . Esto significa que el punto silla no solo representa una solución óptima, sino que también proporciona una medida cuantitativa del equilibrio del juego. Aunque es posible que una matriz de pagos tenga varios puntos silla, la unicidad del valor de un juego asegura que todos estos puntos comparten el mismo valor numérico. En otras palabras, independientemente de cuál sea el punto silla identificado en la matriz, el resultado del juego será consistente y no variará, lo que refuerza la idea de equilibrio inherente a los juegos estrictamente determinados.

Ejemplo 8.5.2:

Dos cadenas de televisión competidoras, R y C , están programando programas de una hora en el mismo período de tiempo. La cadena R puede programar uno de tres programas posibles, y la cadena C puede programar uno de cuatro programas posibles. Ninguna de las cadenas sabe qué programa programará la otra. Ambas cadenas consultan a la misma agencia de encuestas externa para obtener una estimación de cómo todas las combinaciones posibles de programas dividirán la audiencia televisiva. La agencia les proporciona la tabla 8.4, cuya ij -ésima entrada representa el porcentaje de la audiencia que verá la cadena R si el programa i de R se enfrenta al programa j de C . ¿Qué programa debería programar cada cadena para maximizar su audiencia?

		Programa de la red C			
		1	2	3	4
Programa de la red R	1	60	20	30	55
	2	50	75	45	60
	3	70	45	35	30

Tabla 8.4 Porcentaje de audiencia de la cadena de televisión R

Solución: Restamos 50 de cada entrada en la tabla 8.4 para construir la siguiente matriz:

$$\begin{bmatrix} 10 & -30 & -20 & 5 \\ 0 & 25 & -5 & 10 \\ 20 & -5 & -15 & -20 \end{bmatrix}$$

Esta es la matriz de pagos del juego de suma cero para dos personas en el que se considera que cada cadena comienza con el 50 % de la audiencia, y la entrada ij -ésima de la matriz representa el porcentaje de la audiencia que la cadena C pierde frente a la cadena R si los programas i y j se enfrentan entre sí. Es fácil ver que la entrada

$$a_{23} = -5$$

es un punto silla de la matriz de pagos. Por lo tanto, la estrategia óptima para la cadena R es programar el programa 2, y la estrategia óptima para la cadena C es programar el programa 3. Esto resultará en que la cadena R reciba el 45 % de la audiencia y la cadena C reciba el 55 % de la audiencia.

Juegos de matriz 2×2

Otro caso en el que las estrategias óptimas pueden encontrarse mediante métodos elementales ocurre cuando cada jugador tiene solo dos posibles movimientos. En este caso, la matriz de pagos es una matriz de 2×2 dada por

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

Si el juego es estrictamente determinado, al menos una de las cuatro entradas de A es un punto silla, y las técnicas discutidas anteriormente pueden entonces aplicarse para determinar las estrategias óptimas para los dos jugadores. Si el juego no es estrictamente determinado, primero computamos el pago esperado para estrategias arbitrarias \mathbf{p} y \mathbf{q} :

$$E(\mathbf{p}, \mathbf{q}) = \mathbf{p}A\mathbf{q} = a_{11}p_1q_1 + a_{12}p_1q_2 + a_{21}p_2q_1 + a_{22}p_2q_2 \quad (8.5.9)$$

Dado que

$$p_1 + p_2 = 1 \quad \text{y} \quad q_1 + q_2 = 1 \quad (8.5.10)$$

podemos sustituir $p_2 = 1 - p_1$ y $q_2 = 1 - q_1$ en la ecuación (8.5.9) para obtener:

$$E(\mathbf{p}, \mathbf{q}) = a_{11}p_1q_1 + a_{12}p_1(1 - q_1) + a_{21}(1 - p_1)q_1 + a_{22}(1 - p_1)(1 - q_1) \quad (8.5.11)$$

Si reorganizamos los términos en la ecuación (8.5.11), podemos escribir:

$$E(p, q) = [(a_{11} + a_{22} - a_{12} - a_{21})p_1 - (a_{22} - a_{21})]q_1 + (a_{12} - a_{22})p_1 + a_{22} \quad (8.5.12)$$

Al examinar el coeficiente del término q_1 en (8.5.12), vemos que si definimos

$$p_1 = p_1^* = \frac{a_{22} - a_{21}}{a_{11} + a_{22} - a_{12} - a_{21}} \quad (8.5.13)$$

entonces ese coeficiente es cero, y la ecuación (8.5.12) se reduce a:

$$E(p, q) = \frac{a_{11}a_{22} - a_{12}a_{21}}{a_{11} + a_{22} - a_{12} - a_{21}} \quad (8.5.14)$$

La ecuación (8.5.14) es independiente de q ; es decir, si el jugador R elige la estrategia determinada por (8.5.13), el jugador C no puede cambiar el pago esperado variando su estrategia.

De manera similar, puede verificarse que si el jugador C elige la estrategia determinada por

$$q_1 = q_1^* = \frac{a_{22} - a_{12}}{a_{11} + a_{22} - a_{12} - a_{21}} \quad (8.5.15)$$

entonces, sustituyendo en (8.5.12), obtenemos:

$$E(p, q) = \frac{a_{11}a_{22} - a_{12}a_{21}}{a_{11} + a_{22} - a_{12} - a_{21}} \quad (8.5.16)$$

Las ecuaciones (8.5.14) y (8.5.16) muestran que

$$E(p^*, q) = E(p^*, q^*) = E(p, q^*) \quad (8.5.17)$$

para todas las estrategias p y q . Por lo tanto, las estrategias determinadas por (8.5.13), (8.5.15) y (8.5.10) son estrategias óptimas para los jugadores R y C , respectivamente, y así tenemos el siguiente resultado.

Teorema 8.5.2: Para un juego 2×2 que no está estrictamente determinado, las estrategias óptimas para los jugadores R y C son

$$p^* = \left[\frac{a_{22} - a_{21}}{a_{11} + a_{22} - a_{12} - a_{21}} \quad \frac{a_{11} - a_{12}}{a_{11} + a_{22} - a_{12} - a_{21}} \right]$$

y

$$q^* = \left[\frac{a_{22} - a_{12}}{a_{11} + a_{22} - a_{12} - a_{21}} \quad \frac{a_{11} - a_{21}}{a_{11} + a_{22} - a_{12} - a_{21}} \right]$$

El valor del juego es

$$v = \frac{a_{11}a_{22} - a_{12}a_{21}}{a_{11} + a_{22} - a_{12} - a_{21}}$$

Para ser completos, debemos demostrar que las entradas en los vectores p^* y q^* son números estrictamente entre 0 y 1.

La ecuación (8.5.17) es interesante porque implica que cualquiera de los jugadores, al optar por su estrategia óptima, puede garantizar que el pago esperado sea exactamente igual al valor del juego, independientemente de la estrategia elegida por el otro jugador. Esto resalta una propiedad clave de los juegos con dos posibles movimientos por jugador: el control que cada jugador puede ejercer sobre el resultado esperado del juego. Sin embargo, este resultado no se generaliza a juegos más complejos, en los cuales alguno de los jugadores tiene más de dos movimientos posibles. En tales casos, las estrategias óptimas suelen ser más complicadas de calcular y no ofrecen el mismo nivel de control sobre el valor del juego, lo que añade un mayor grado de incertidumbre y estrategia al análisis.

Ejemplo 8.5.3:

El gobierno federal desea inocular a sus ciudadanos contra un cierto virus de la gripe. El virus tiene dos cepas, y las proporciones en las que las dos cepas ocurren en la población del virus no se conocen. Se han desarrollado dos vacunas y a cada ciudadano se le administra solo una de ellas. La vacuna 1 es 85 % efectiva contra la cepa 1 y 70 % efectiva contra la cepa 2. La vacuna 2 es 60 % efectiva contra la cepa 1 y 90 % efectiva contra la cepa 2. ¿Qué política de inoculación debería adoptar el gobierno?

Solución: Podemos considerar esto como un juego de dos personas en el que el jugador R (el gobierno) desea maximizar el pago (la fracción de ciudadanos resistentes al virus) tanto como sea posible, y el jugador C (el virus) desea minimizar el pago tanto como sea posible. La matriz de pagos es

		Cepa	
		1	2
Vacuna	1	0.85	0.70
	2	0.60	0.90

Esta matriz no tiene puntos silla, por lo que el teorema anterior es aplicable. Por consiguiente,

$$p_1^* = \frac{a_{22} - a_{21}}{a_{11} + a_{22} - a_{12} - a_{21}} = \frac{0.90 - 0.60}{0.85 + 0.90 - 0.70 - 0.60} = \frac{0.30}{0.45} = \frac{2}{3}$$

entonces

$$p_2^* = 1 - p_1^* = 1 - \frac{2}{3} = \frac{1}{3}$$

y

$$q_1^* = \frac{a_{22} - a_{12}}{a_{11} + a_{22} - a_{12} - a_{21}} = \frac{0.90 - 0.70}{0.85 + 0.90 - 0.70 - 0.60} = \frac{0.20}{0.45} = \frac{4}{9}$$

entonces

$$q_2^* = 1 - q_1^* = 1 - \frac{4}{9} = \frac{5}{9}$$

Finalmente,

$$v = \frac{a_{11}a_{22} - a_{12}a_{21}}{a_{11} + a_{22} - a_{12} - a_{21}} = \frac{(0.85)(0.90) - (0.70)(0.60)}{0.85 + 0.90 - 0.70 - 0.60} = \frac{0.345}{0.45} = 0.7666 \dots$$

Por lo tanto, la estrategia óptima para el gobierno es inocular a $2/3$ de los ciudadanos con la vacuna 1 y a $1/3$ de los ciudadanos con la vacuna 2. Esto garantizará que aproximadamente el 76.7 % de los ciudadanos sean resistentes a un ataque del virus, independientemente de la distribución de las dos cepas. En contraste, una distribución del virus de $4/9$ de la cepa 1 y $5/9$ de la cepa 2 resultará en el mismo 76.7 % de ciudadanos resistentes, independientemente de la estrategia de inoculación adoptada por el gobierno.

8.6 Algunos modelos económicos de Leontief

En esta sección analizamos dos modelos lineales para sistemas económicos. Algunos resultados sobre matrices no negativas se aplican para determinar estructuras de precios de equilibrio y productos necesarios para satisfacer la demanda. La teoría de matrices ha sido muy exitosa en describir las interrelaciones entre precios, productos y demandas en los sistemas económicos. En esta sección discutimos algunos modelos simples basados en las ideas del laureado Nobel Wassily Leontief. Examinamos dos modelos diferentes pero relacionados: el modelo cerrado o modelo de entrada-salida, y el modelo abierto o modelo de producción. En cada uno, se nos dan ciertos parámetros económicos que describen las interrelaciones entre

las “industrias” en la economía bajo consideración. Usando la teoría de matrices, evaluamos ciertos parámetros, como precios o niveles de producción, con el fin de satisfacer un objetivo económico deseado. Comenzamos con el modelo cerrado.

Modelo cerrado (o de entrada-salida) de Leontief

Ejemplo 8.6.1: Tres propietarios de viviendas (un carpintero, un electricista y un plomero) acuerdan realizar reparaciones en sus tres casas. Ellos acuerdan trabajar un total de 10 días cada uno según el siguiente horario:

	Trabajo realizado por		
	Carpintero	Electricista	Plomero
Días de trabajo en casa del carpintero	2	1	6
Días de trabajo en casa del electricista	4	5	1
Días de trabajo en casa del plomero	4	4	3

Por propósitos fiscales, deben reportar y pagarse entre sí un salario diario razonable, incluso por el trabajo que cada uno realiza en su propia casa. Sus salarios diarios normales son de aproximadamente \$100, pero acuerdan ajustar sus respectivos salarios diarios de modo que cada propietario quede en igualdad de condiciones; es decir, que la cantidad total pagada por cada uno sea igual a la cantidad total que recibe. Podemos establecer que

p_1 = salario diario del carpintero,

p_2 = salario diario del electricista,

p_3 = salario diario del plomero.

Para satisfacer la condición de “equilibrio” de que cada propietario quede en igualdad, requerimos que

$$\text{gastos totales} = \text{ingresos totales}$$

para cada uno de los propietarios durante el periodo de 10 días. Por ejemplo, el carpintero paga un total de $2p_1 + p_2 + 6p_3$ por las reparaciones en su propia casa y recibe un ingreso total de $10p_1$ por las reparaciones que realiza en las tres casas. Igualando estas dos expresiones, obtenemos la primera de las siguientes tres ecuaciones:

$$2p_1 + p_2 + 6p_3 = 10p_1$$

$$4p_1 + 5p_2 + p_3 = 10p_2$$

$$4p_1 + 4p_2 + 3p_3 = 10p_3$$

Las dos ecuaciones restantes son las ecuaciones de equilibrio para el electricista y el fontanero. Dividiendo estas ecuaciones entre 10 y reescribiéndolas en forma matricial, obtenemos

$$\begin{bmatrix} 0.2 & 0.1 & 0.6 \\ 0.4 & 0.5 & 0.1 \\ 0.4 & 0.4 & 0.3 \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix} = \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix} \quad (8.6.1)$$

La ecuación (8.6.1) puede reescribirse como un sistema homogéneo restando el lado derecho del lado izquierdo para obtener

$$\begin{bmatrix} 0.8 & -0.1 & -0.6 \\ -0.4 & 0.5 & -0.1 \\ -0.4 & -0.4 & 0.7 \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

La solución de este sistema homogéneo es

$$\begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix} = s \begin{bmatrix} 31 \\ 32 \\ 36 \end{bmatrix}$$

donde s es una constante arbitraria. Esta constante es un factor de escala, que los propietarios pueden elegir para su conveniencia. Por ejemplo, pueden establecer $s = 3$ de modo que los salarios diarios correspondientes \$93, \$96 y \$108 sean aproximadamente \$100.

Este ejemplo ilustra las características destacadas del modelo input-output de Leontief en una economía cerrada. En la ecuación básica (8.6.1), la suma de cada columna de la matriz de coeficientes es 1, lo que corresponde al hecho de que la “producción” de trabajo de cada propietario está completamente distribuida entre estos propietarios en las proporciones dadas por las entradas en la columna. Nuestro problema es determinar “precios” adecuados para estos productos de modo que el sistema esté en equilibrio, es decir, que los gastos totales de cada propietario sean iguales a sus ingresos totales.

En el modelo general, tenemos un sistema económico que consta de un número finito de “industrias”, a las cuales numeramos como industrias 1, 2, ..., k . Durante un período de tiempo fijo, cada industria produce una “salida” de algún bien o servicio que es completamente utilizado de manera predeterminada por las k industrias. Un problema importante es encontrar precios adecuados para estos k productos de modo que, para cada industria, los gastos totales sean iguales a los ingresos totales. Tal estructura de precios representa una posición de equilibrio para la economía. Para el período de tiempo fijo en cuestión, definimos:

p_i = precio fijado por la industria i por su producción total,

e_{ij} = fracción de la producción total de la industria j adquirida por la industria i ,

para $i, j = 1, 2, \dots, k$. Por definición, tenemos:

i) $p_i \geq 0$, para $i = 1, 2, \dots, k$,

ii) $e_{ij} \geq 0$, para $i, j = 1, 2, \dots, k$,

iii) $e_{1j} + e_{2j} + \dots + e_{kj} = 1$, para $j = 1, 2, \dots, k$.

Con estas cantidades, formamos el *vector de precios*

$$\mathbb{p} = \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_k \end{bmatrix}$$

y la *matriz de intercambio* o *matriz entrada-salida*

$$E = \begin{bmatrix} e_{11} & e_{12} & \cdots & e_{1k} \\ e_{21} & e_{22} & \cdots & e_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ e_{k1} & e_{k2} & \cdots & e_{kk} \end{bmatrix}$$

La condición (iii) expresa el hecho de que la suma de cada columna de la matriz de intercambio es 1.

Como en el ejemplo, para que los gastos de cada industria sean iguales a sus ingresos, se debe satisfacer la siguiente ecuación matricial [ver (8.6.1)]:

$$E\mathbb{p} = \mathbb{p} \quad (8.6.2)$$

o

$$(I_k - E)\mathbb{p} = \mathbf{0} \quad (8.6.3)$$

La ecuación (8.6.3) es un sistema lineal homogéneo para el vector de precios \mathbb{p} . Tendrá una solución no trivial si y solo si el determinante de la matriz de coeficientes $I_k - E$ es cero. Por lo tanto, la ecuación (8.6.3) siempre tiene soluciones no triviales para el vector de precios \mathbb{p} .

En realidad, para que nuestro modelo económico tenga sentido, necesitamos algo más que el hecho de que la ecuación (8.6.3) tenga soluciones no triviales para \mathbb{p} . También necesitamos que los precios p_i de las k salidas sean números no negativos. Expresamos esta condición como $\mathbb{p} \geq 0$. En general, si A es un vector o matriz, la notación $A \geq 0$ significa que cada entrada de A es no negativa, y la notación $A > 0$ significa que cada entrada de A es positiva. De manera similar, $A \geq B$ significa que $A - B \geq 0$, y $A > B$ significa que $A - B > 0$.

Demostrar que la ecuación (8.6.3) tiene una solución no trivial para la cual $\mathbb{p} \geq 0$ es un poco más difícil que mostrar simplemente que existe alguna solución no trivial. Pero es cierto, y afirmamos este hecho sin prueba en el siguiente teorema.

Teorema 8.6.1: Si E es una matriz de intercambio, entonces $E\mathbb{p} = \mathbb{p}$ siempre tiene una solución no trivial \mathbb{p} cuyas entradas son no negativas.

Consideremos algunos ejemplos simples de este teorema.

Ejemplo 8.6.2: Sea

$$E = \begin{bmatrix} \frac{1}{2} & 0 \\ \frac{1}{2} & 1 \end{bmatrix}$$

Entonces, la ecuación $(I_2 - E)\mathbb{p} = \mathbf{0}$ es

$$\begin{bmatrix} \frac{1}{2} & 0 \\ -\frac{1}{2} & 0 \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

lo que tiene como solución general

$$\mathbb{p} = s \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

donde s es una constante arbitraria. Por lo tanto, tenemos soluciones no triviales $\mathbb{p} \geq 0$ para cualquier $s > 0$.

Ejemplo 8.6.3: Sea

$$E = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Entonces, la ecuación $(I_2 - E)\mathbb{p} = \mathbf{0}$ tiene la solución general

$$\mathbb{p} = s \begin{bmatrix} 1 \\ 0 \end{bmatrix} + t \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

donde s y t son constantes arbitrarias e independientes. Las soluciones no triviales $\mathbb{p} \geq 0$ resultan de cualquier $s \geq 0$ y $t \geq 0$, pero no ambos cero.

El ejemplo 8.6.2 indica que, en algunas situaciones, uno de los precios debe ser cero para satisfacer la condición de equilibrio. El ejemplo 8.6.3 indica que puede haber varias estructuras de precios linealmente independientes disponibles. Ninguna de estas situaciones describe una estructura económica verdaderamente interdependiente. El siguiente teorema proporciona condiciones suficientes para excluir ambos casos.

Teorema 8.6.2: Sea E una matriz de intercambio tal que, para algún entero positivo m , todas las entradas de E^m son positivas. Entonces, existe exactamente una solución linealmente independiente de $(I - E)\mathbb{p} = 0$, y puede elegirse de manera que todas sus entradas sean positivas.

No daremos una demostración de este teorema. Si has leído la sección 8.3 sobre cadenas de Markov, observa que este teorema es esencialmente el mismo que el teorema 8.3.4. Lo que estamos llamando matrices de intercambio en esta sección fueron llamadas matrices estocásticas o de Markov en la sección 8.3.

Ejemplo 8.6.4: La matriz de intercambio en el ejemplo 8.6.1 era

$$E = \begin{bmatrix} 0.2 & 0.1 & 0.6 \\ 0.4 & 0.5 & 0.1 \\ 0.4 & 0.4 & 0.3 \end{bmatrix}$$

Debido a que $E > 0$, entonces la condición $E^m > 0$ del teorema 8.6.2 se satisface para $m = 1$. En consecuencia, se garantiza que existe exactamente una solución linealmente independiente de $(I_3 - E)\mathbb{p} = 0$, y puede elegirse de forma que $\mathbb{p} > 0$. En ese ejemplo, encontramos que

$$\mathbb{p} = \begin{bmatrix} 31 \\ 32 \\ 36 \end{bmatrix}$$

es tal solución.

Modelo abierto (o de producción) de Leontief

A diferencia del modelo cerrado, donde los productos de las k industrias se distribuyen solo entre ellas, el modelo abierto busca satisfacer una demanda externa. Parte de estos productos puede destinarse a mantener las industrias en funcionamiento, pero debe existir un excedente para atender la demanda externa.

En el modelo cerrado, los productos de las industrias son fijos, y el objetivo es determinar sus precios para que se cumpla la condición de equilibrio: que los gastos sean iguales a los ingresos. En el modelo abierto, los precios son fijos, y el objetivo es calcular los niveles de producción necesarios para satisfacer la demanda externa. Estos niveles se medirán en términos de sus valores económicos utilizando los precios fijos. Durante un período determinado, definimos:

x_i = valor monetario de la producción total de la i -ésima industria,

d_i = valor monetario de la producción de la i -ésima industria necesario para satisfacer la demanda externa,

c_{ij} = valor monetario de la producción de la i -ésima industria requerido por la j -ésima industria para producir una unidad monetaria de su producción.

Con estas cantidades, definimos el *vector de producción* como

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{bmatrix},$$

el *vector de demanda* como

$$\mathbf{d} = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_k \end{bmatrix}$$

y la matriz de consumo

$$C = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1k} \\ c_{21} & c_{22} & \cdots & c_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ c_{k1} & c_{k2} & \cdots & c_{kk} \end{bmatrix}$$

Por su naturaleza, tenemos que

$$x \geq 0, \quad d \geq 0, \quad C \geq 0.$$

A partir de la definición de c_{ij} y x_j , se puede observar que la cantidad

$$c_{i1}x_1 + c_{i2}x_2 + \cdots + c_{ik}x_k$$

es el valor del producto de la industria i necesario para que todas las k industrias produzcan el total de productos especificado por el vector de producción x . Dado que esta cantidad es simplemente la entrada i -ésima del vector columna Cx , podemos decir además que la entrada i -ésima del vector columna

$$x - Cx$$

es el valor del producto excedente de la industria i disponible para satisfacer la demanda externa. El valor de la demanda externa del producto de la industria i es la entrada i -ésima del vector de demanda d . En consecuencia, llegamos a la siguiente ecuación

$$x - Cx = d$$

o

$$(I - C)x = d \quad (8.6.4)$$

para que la demanda se cumpla exactamente, sin excedentes ni carencias. Así, dados C y d , nuestro objetivo es encontrar un vector de producción $x \geq 0$ que satisfaga la ecuación (8.6.4).

Ejemplo 8.6.5:

Una ciudad cuenta con tres industrias principales: una operación minera de carbón, una planta generadora de electricidad y un ferrocarril local. Para extraer \$1 de carbón, la operación minera debe comprar \$0.25 de electricidad para operar su equipo y \$0.25 de transporte para sus necesidades de envío. Para producir \$1 de electricidad, la planta generadora requiere \$0.65 de carbón como combustible, \$0.05 de su propia electricidad para operar su equipo auxiliar y \$0.05 de transporte. Para proporcionar \$1 de transporte, el ferrocarril requiere \$0.55 de carbón como combustible y \$0.10 de electricidad para su equipo auxiliar. En una semana determinada, la operación minera de carbón recibe pedidos por \$50 000 de carbón desde fuera de la ciudad, y la planta generadora recibe pedidos por \$25 000 de electricidad desde fuera de la ciudad. No hay demanda externa para el ferrocarril local. ¿Cuánto debe producir cada una de las tres industrias en esa semana para satisfacer exactamente su propia demanda y la demanda externa?

Solución: Para el período de una semana, definamos:

- x_1 = valor de la producción total de la operación minera de carbón,
- x_2 = valor de la producción total de la planta generadora de electricidad,
- x_3 = valor de la producción total del ferrocarril local.

Con la información proporcionada, la *matriz de consumo* del sistema es

$$C = \begin{bmatrix} 0 & 0.65 & 0.55 \\ 0.25 & 0.05 & 0.10 \\ 0.25 & 0.05 & 0 \end{bmatrix}$$

El sistema lineal $(I - C)x = d$ es entonces

$$\begin{bmatrix} 1.00 & -0.65 & -0.55 \\ -0.25 & 0.95 & -0.10 \\ -0.25 & -0.05 & 1.00 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 50\,000 \\ 25\,000 \\ 0 \end{bmatrix}.$$

La matriz de coeficientes a la izquierda es invertible, y la solución está dada por

$$x = (I - C)^{-1}d = \frac{1}{503} \begin{bmatrix} 756 & 542 & 470 \\ 220 & 690 & 190 \\ 200 & 170 & 630 \end{bmatrix} \begin{bmatrix} 50\,000 \\ 25\,000 \\ 0 \end{bmatrix} = \begin{bmatrix} 102\,087 \\ 56\,163 \\ 28\,330 \end{bmatrix}.$$

Por lo tanto, la producción total de la operación minera de carbón debe ser \$102 087, la producción total de la planta generadora de electricidad debe ser \$56 163, y la producción total del ferrocarril debe ser \$28 330.

Volvamos a considerar la ecuación (8.6.4):

$$(I - C)x = d.$$

Si la matriz cuadrada $I - C$ es invertible, podemos escribir

$$x = (I - C)^{-1}d. \quad (8.6.5)$$

Además, si la matriz $(I - C)^{-1}$ tiene únicamente entradas no negativas, entonces estamos garantizados de que para cualquier $d \geq 0$, la ecuación (8.6.5) tiene una solución única y no negativa para x . Esta es una situación particularmente deseable, ya que significa que cualquier demanda externa puede ser satisfecha. La terminología utilizada para describir este caso se da en la siguiente definición:

Definición 8.6.1:

Se dice que una matriz de consumo C es *productiva* si y solo si $(I - C)^{-1}$ existe, y además,

$$(I - C)^{-1} \geq 0.$$

Ahora consideraremos algunos criterios simples que garantizan que una matriz de consumo es productiva. El primero se da en el siguiente teorema:

Teorema 8.6.3:

Una matriz de consumo C es productiva si y solo si existe algún vector de producción $x \geq 0$ tal que $x > Cx$.

La condición $x > Cx$ significa que existe un plan de producción posible tal que cada industria produce más de lo que consume. El teorema 8.6.3 tiene dos corolarios interesantes. Supongamos que la suma de cada fila de C es menor que 1. Si

$$x = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

entonces Cx es un vector columna cuyas entradas son estas sumas de fila. Por lo tanto, $x > Cx$, y se satisface la condición del teorema 8.6.3. Así, llegamos al siguiente corolario:

Corolario 8.6.1:

Una matriz de consumo es productiva si la suma de cada una de sus filas es menor que 1.

Corolario 8.6.2:

Una matriz de consumo es productiva si cada una de las sumas de sus columnas es menor que 1.

Recordando la definición de las entradas de la matriz de consumo C , observamos que la suma de la j -ésima columna de C es el valor total de los productos de todas

las k industrias necesarios para producir una unidad de valor del producto de la j -ésima industria. Por lo tanto, se dice que la j -ésima industria es *rentable* si la suma de su columna es menor que 1. En otras palabras, el corolario 8.6.2 establece que una matriz de consumo es productiva si todas las k industrias en el sistema económico son rentables.

Ejemplo 8.6.6: Usando el corolario 8.6.2, la matriz de consumo en el ejemplo 8.6.5 era

$$C = \begin{bmatrix} 0 & 0.65 & 0.55 \\ 0.25 & 0.05 & 0.10 \\ 0.25 & 0.05 & 0 \end{bmatrix}$$

Todas las sumas de las columnas en esta matriz son menores que 1, por lo que las tres industrias son rentables. En consecuencia, según el corolario 8.6.2, la matriz de consumo C es productiva. Esto también se puede ver en los cálculos del ejemplo 8.6.5, ya que $(I - C)^{-1}$ es no negativa.

8.7 Criptografía

El estudio de codificar y decodificar mensajes secretos se denomina *criptografía*. Aunque los códigos secretos se remontan a los primeros días de la comunicación escrita, ha habido un reciente aumento de interés en el tema debido a la necesidad de mantener la privacidad de la información transmitida a través de líneas públicas de comunicación. En el lenguaje de la criptografía, los códigos se llaman *cifras*, los mensajes sin codificar se llaman *texto plano* y los mensajes codificados se llaman *texto cifrado*. El proceso de convertir texto plano en texto cifrado se llama *cifrado*, y el proceso inverso de convertir texto cifrado en texto plano se llama *descifrado*.

Las cifras más simples, llamadas *cifras de sustitución*, son aquellas que reemplazan cada letra del alfabeto por una letra diferente. Por ejemplo, en la cifra de sustitución

Plano	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Cifra	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C

la letra del texto plano A es reemplazada por D , la letra del texto plano B por E , y así sucesivamente. Con esta cifra, el mensaje en texto plano

LAS ESTRELLAS BRILLAN EN SILENCIO

se convierte en

ODV HVWUHOODV EULOODQ HQ VLOHQFLR

Cifras de Hill

Una desventaja de las cifras de sustitución es que preservan las frecuencias de las letras individuales, lo que hace relativamente fácil romper el código mediante métodos estadísticos. Una forma de superar este problema es dividir el texto plano en grupos de letras y cifrar el texto plano grupo por grupo, en lugar de una letra a la vez. Un sistema de criptografía en el cual el texto plano se divide en conjuntos de n letras, cada una de las cuales es reemplazada por un conjunto de n letras cifradas, se llama *sistema poligráfico*. En esta sección, estudiaremos una clase de sistemas poligráficos basados en transformaciones matriciales. Las cifras que discutiremos se llaman *cifras de Hill* en honor a Lester S. Hill, quien las introdujo en dos artículos: “Cryptography in an Algebraic Alphabet”, *American Mathematical Monthly*, 36 (junio-julio de 1929), pp. 306–312; y “Concerning Certain Linear Transformation Apparatus of Cryptography”, *American Mathematical Monthly*, 38 (marzo de 1931), pp. 135–154.

En la discusión que sigue, asumimos que a cada letra del texto plano y del texto cifrado, excepto a *Z*, se le asigna el valor numérico que especifica su posición en el alfabeto estándar (tabla 8.5). Por razones que se harán claras más adelante, a *Z* se le asigna un valor de cero.

Tabla 8.5 Asignación de valores numéricos al alfabeto

<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>F</i>	<i>G</i>	<i>H</i>	<i>I</i>	<i>J</i>	<i>K</i>	<i>L</i>	<i>M</i>	<i>N</i>	<i>O</i>	<i>P</i>	<i>Q</i>	<i>R</i>	<i>S</i>	<i>T</i>	<i>U</i>	<i>V</i>	<i>W</i>	<i>X</i>	<i>Y</i>	<i>Z</i>
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	0

En los cifrados de Hill más simples, pares sucesivos de texto plano se transforman en texto cifrado mediante el siguiente procedimiento:

Paso 1. Elegir una matriz 2×2 con entradas enteras

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

para realizar el cifrado. Se impondrán ciertas condiciones adicionales sobre *A* más adelante.

Paso 2. Agrupar las letras sucesivas del texto plano en pares, añadiendo una letra “ficticia” arbitraria para completar el último par si el texto tiene un número impar de letras, y reemplazar cada letra del texto plano por su valor numérico.

Paso 3. Convertir sucesivamente cada par de texto plano $p_1 p_2$ en un vector columna

$$\mathbf{p} = \begin{bmatrix} p_1 \\ p_2 \end{bmatrix}$$

y formar el producto $A\mathbf{p}$. Llamaremos \mathbf{p} *vector de texto plano* y $A\mathbf{p}$ el *vector de texto cifrado* correspondiente.

Paso 4. Convertir cada vector de texto cifrado en su equivalente alfabético.

Ejemplo 8.7.1:

Utilizar la matriz

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

para obtener el cifrado de Hill para el mensaje de texto plano: *I AM HIDING*.

Solución: Si agrupamos el texto plano en pares y añadimos la letra ficticia *G* para completar el último par, obtenemos:

I A M H I D I N G G

o, de manera equivalente, según la tabla 8.5,

9 1 13 8 9 4 9 14 7 7

Para cifrar el par *IA*, formamos el producto matricial

$$\begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 9 \\ 1 \end{bmatrix} = \begin{bmatrix} 11 \\ 3 \end{bmatrix},$$

que, según la tabla 8.5, produce el texto cifrado *KC*. Análogamente, para cifrar el par *MH* formamos el producto

$$\begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 13 \\ 8 \end{bmatrix} = \begin{bmatrix} 29 \\ 24 \end{bmatrix}.$$

Sin embargo, aquí hay un problema, porque el número 29 no tiene un equivalente en el alfabeto (tabla 8.5). Para resolver este problema, hacemos el siguiente acuerdo:

Siempre que ocurra un número mayor a 25, será reemplazado por el residuo que resulta al dividir dicho número entre 26.

Dado que el residuo después de dividir entre 26 siempre será uno de los enteros 0, 1, 2, ..., 25, este procedimiento siempre producirá un entero con un equivalente en el alfabeto. Así, en la ecuación anterior reemplazamos 29 por 3, que es el residuo al dividir 29 entre 26. Según la tabla 8.5, el texto cifrado para el par MH es CX . Los cálculos para los vectores cifrados restantes son:

$$\begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 9 \\ 4 \end{bmatrix} = \begin{bmatrix} 17 \\ 12 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 9 \\ 14 \end{bmatrix} = \begin{bmatrix} 37 \\ 42 \end{bmatrix} \quad \text{o bien} \quad \begin{bmatrix} 11 \\ 16 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 7 \\ 7 \end{bmatrix} = \begin{bmatrix} 21 \\ 21 \end{bmatrix}$$

Estos corresponden a los pares de texto cifrado QL , KP y UU , respectivamente. En resumen, el mensaje completo cifrado es

$KC \quad CX \quad QL \quad KP \quad UU$

que usualmente sería transmitido como una cadena única sin espacios:

$KCCXQLKPUU$

Debido a que el texto claro se agrupó en pares y se cifró mediante una matriz 2×2 , el cifrado de Hill en el ejemplo 8.7.1 se denomina *cifrado de Hill 2*. Obviamente, también es posible agrupar el texto plano en tríos y cifrarlo mediante una matriz 3×3 con entradas enteras; esto se conoce como *cifrado de Hill 3*. En general, para un *cifrado de Hill n*, el texto plano se agrupa en conjuntos de n letras y se cifra mediante una matriz $n \times n$ con entradas enteras.

Aritmética modular

En el ejemplo 8.7.1, los enteros mayores que 25 se reemplazaron por sus residuos tras dividir entre 26. Esta técnica de trabajar con residuos está en el núcleo de un área de las matemáticas llamada *aritmética modular*. Debido a su importancia en criptografía, nos desviaremos brevemente para abordar algunas de las ideas principales en esta área.

En la aritmética modular se nos da un entero positivo m , llamado el *módulo*, y cualquier par de enteros cuya diferencia sea un múltiplo entero del módulo se considera “igual” o “equivalente” con respecto al módulo. Más precisamente, hacemos la siguiente definición:

Definición 8.7.1:

Si m es un entero positivo y a y b son cualesquiera enteros, entonces decimos que a es *equivalente* a b módulo m , escrito como

$$a \equiv b \pmod{m}$$

si $a - b$ es un múltiplo entero de m .

Ejemplo 8.7.2:

Algunos ejemplos de equivalencias módulo m son los siguientes:

$$7 \equiv 2 \pmod{5}$$

$$19 \equiv 3 \pmod{2}$$

$$-1 \equiv 8 \pmod{9}$$

$$12 \equiv 0 \pmod{4}$$

$$-2 \equiv 5 \pmod{7}$$

$$6 \equiv 3 \pmod{3}$$

Para cualquier módulo m , se puede demostrar que todo entero a es equivalente, módulo m , a exactamente uno de los enteros

$$0, 1, 2, \dots, m-1.$$

Llamamos a este entero el *residuo* de a módulo m , y escribimos

$$Z_m = \{0, 1, 2, \dots, m-1\}$$

para denotar el conjunto de residuos módulo m .

Si a es un entero no negativo, su residuo módulo m es simplemente el residuo que resulta al dividir a entre m . Para un entero arbitrario a , el residuo puede encontrarse utilizando el siguiente teorema.

Teorema 8.7.1: Para cualquier entero a y módulo m , sea

$$R = \text{el residuo de } \frac{|a|}{m}.$$

Entonces, el residuo r de a módulo m está dado por

$$r = \begin{cases} R & \text{si } a \geq 0, \\ m - R & \text{si } a < 0 \text{ y } R \neq 0, \\ 0 & \text{si } a < 0 \text{ y } R = 0. \end{cases}$$

Ejemplo 8.7.3: Encuentre el residuo módulo 26 de

- a) 87,
- b) -38,
- c) -26.

Solución:

- a) Dividiendo $|87| = 87$ entre 26, se obtiene un residuo de $R = 9$, por lo que $r = 9$. Por lo tanto,

$$87 \equiv 9 \pmod{26}.$$

- b) Dividiendo $|-38| = 38$ entre 26, se obtiene un residuo de $R = 12$, por lo que $r = 26 - 12 = 14$. Por lo tanto,

$$-38 \equiv 14 \pmod{26}.$$

- c) Dividiendo $|-26| = 26$ entre 26, se obtiene un residuo de $R = 0$. Por lo tanto,

$$-26 \equiv 0 \pmod{26}.$$

En aritmética ordinaria, todo número distinto de cero a tiene un recíproco o inverso multiplicativo, denotado por a^{-1} , tal que

$$aa^{-1} = a^{-1}a = 1.$$

En aritmética modular, tenemos el siguiente concepto correspondiente:

Definición 8.7.2:

Si a es un número en Z_m , entonces un número a^{-1} en Z_m se llama *recíproco* o *inverso multiplicativo* de a módulo m si

$$aa^{-1} \equiv a^{-1}a \equiv 1 \pmod{m}.$$

Se puede demostrar que si a y m no tienen factores primos comunes, entonces a tiene un único recíproco módulo m ; de manera contraria, si a y m tienen un factor primo común, entonces a no tiene recíproco módulo m .

Ejemplo 8.7.4:

El número 3 tiene un recíproco módulo 26 porque 3 y 26 no tienen factores primos comunes. Este recíproco puede obtenerse encontrando el número x en Z_{26} que satisface la ecuación modular

$$3x \equiv 1 \pmod{26}.$$

Aunque existen métodos generales para resolver tales ecuaciones modulares, sería desviarnos demasiado estudiarlos aquí. Sin embargo, debido a que 26 es un número relativamente pequeño, esta ecuación puede resolverse probando las posibles soluciones, de 0 a 25, una por una. Con este enfoque encontramos que $x = 9$ es la solución, porque

$$3 \cdot 9 = 27 \equiv 1 \pmod{26}.$$

Por lo tanto,

$$3^{-1} \equiv 9 \pmod{26}.$$

Ejemplo 8.7.5:

El número 4 no tiene recíproco módulo 26, porque 4 y 26 tienen el factor primo 2 en común.

Para referencia futura, en la tabla 8.6 se presentan los siguientes recíprocos módulo 26:

Tabla 8.6 Recíprocos módulo 26

a	1	3	5	7	9	11	15	17	19	21	23	25
a^{-1}	1	9	21	15	3	19	7	23	11	5	17	25

Descifrado

Todo cifrado útil debe incluir un procedimiento para descifrar. En el caso de un cifrado de Hill, el descifrado utiliza el inverso (mód 26) de la matriz de cifrado. Para ser precisos, si m es un número positivo, entonces una matriz cuadrada A con entradas en Z_m se dice *invertible módulo m* si existe una matriz B con entradas en Z_m tal que

$$AB = BA = I \pmod{m}.$$

Supongamos ahora que

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

es invertible módulo 26 y esta matriz se usa en un cifrado de Hill 2. Si

$$\mathbf{p} = \begin{bmatrix} p_1 \\ p_2 \end{bmatrix}$$

es un vector de texto plano, entonces

$$\mathbf{c} = A\mathbf{p} \pmod{26}$$

es el vector de texto cifrado correspondiente y

$$\mathbf{p} = A^{-1}\mathbf{c} \pmod{26}.$$

Por lo tanto, cada vector de texto plano puede recuperarse del vector de texto cifrado correspondiente multiplicándolo por la izquierda con A^{-1} (mód 26).

En criptografía es importante saber qué matrices son invertibles módulo 26 y cómo obtener sus inversos. Ahora investigamos estas preguntas.

En aritmética ordinaria, una matriz cuadrada A es invertible si y solo si $\det(A) \neq 0$, o, equivalentemente, si y solo si $\det(A)$ tiene un recíproco. El siguiente teorema es el análogo de este resultado en aritmética modular.

Teorema 8.7.2: Una matriz cuadrada A con entradas en Z_m es invertible módulo m si y solo si el residuo de $\det(A)$ módulo m tiene un recíproco módulo m .

Dado que el residuo de $\det(A)$ módulo m tendrá un recíproco módulo m si y solo si este residuo y m no tienen factores primos en común, tenemos el siguiente corolario.

Corolario 8.7.1: Una matriz cuadrada A con entradas en Z_m es invertible módulo m si y solo si m y el residuo de $\det(A)$ módulo m no tienen factores primos en común.

Dado que los únicos factores primos de $m = 26$ son 2 y 13, tenemos el siguiente corolario, que es útil en criptografía.

Corolario 8.7.2: Una matriz cuadrada A con entradas en Z_{26} es invertible módulo 26 si y solo si el residuo de $\det(A)$ módulo 26 no es divisible por 2 ni por 13.

Dejamos como ejercicio al lector verificar que si

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

tiene entradas en Z_{26} y el residuo de $\det(A) = ad - bc$ módulo 26 no es divisible por 2 ni por 13, entonces el inverso de A (mód 26) está dado por

$$A^{-1} = (ad - bc)^{-1} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \quad (\text{mód } 26), \quad (8.7.1)$$

donde $(ad - bc)^{-1}$ es el recíproco del residuo de $ad - bc$ (mód 26).

Ejemplo 8.7.6: Hallar el inverso de

$$A = \begin{bmatrix} 5 & 6 \\ 2 & 3 \end{bmatrix}$$

módulo 26.

Solución: El determinante de A es

$$\det(A) = ad - bc = 5 \cdot 3 - 6 \cdot 2 = 3,$$

por lo tanto, de la tabla 8.5, sabemos que

$$(ad - bc)^{-1} = 3^{-1} = 9 \quad (\text{mód } 26).$$

De la ecuación (8.7.1), obtenemos

$$A^{-1} = 9 \begin{bmatrix} 3 & -6 \\ -2 & 5 \end{bmatrix} = \begin{bmatrix} 27 & -54 \\ -18 & 45 \end{bmatrix} = \begin{bmatrix} 1 & 24 \\ 8 & 19 \end{bmatrix} \quad (\text{mód } 26).$$

Para verificar, calculamos

$$AA^{-1} = \begin{bmatrix} 5 & 6 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} 1 & 24 \\ 8 & 19 \end{bmatrix} = \begin{bmatrix} 53 & 234 \\ 26 & 105 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad (\text{mód } 26).$$

De manera similar, se puede comprobar que $A^{-1}A = I$.

Ejemplo 8.7.7: Decodificar el siguiente cifrado de Hill, que fue cifrado con la matriz del ejemplo anterior:

GTNKGKDU SK

Solución: De la tabla 8.5, el equivalente numérico de este texto cifrado es

7 20 14 11 7 11 4 21 19 11

Para obtener los pares del texto plano, multiplicamos cada vector del texto cifrado por el inverso de A (obtenido en el ejemplo anterior):

$$\begin{bmatrix} 1 & 24 \\ 8 & 19 \end{bmatrix} \begin{bmatrix} 7 \\ 20 \end{bmatrix} = \begin{bmatrix} 487 \\ 436 \end{bmatrix} = \begin{bmatrix} 19 \\ 20 \end{bmatrix} \pmod{26}$$

$$\begin{bmatrix} 1 & 24 \\ 8 & 19 \end{bmatrix} \begin{bmatrix} 14 \\ 11 \end{bmatrix} = \begin{bmatrix} 278 \\ 321 \end{bmatrix} = \begin{bmatrix} 18 \\ 9 \end{bmatrix} \pmod{26}$$

$$\begin{bmatrix} 1 & 24 \\ 8 & 19 \end{bmatrix} \begin{bmatrix} 7 \\ 11 \end{bmatrix} = \begin{bmatrix} 271 \\ 265 \end{bmatrix} = \begin{bmatrix} 11 \\ 5 \end{bmatrix} \pmod{26}$$

$$\begin{bmatrix} 1 & 24 \\ 8 & 19 \end{bmatrix} \begin{bmatrix} 4 \\ 21 \end{bmatrix} = \begin{bmatrix} 508 \\ 431 \end{bmatrix} = \begin{bmatrix} 14 \\ 15 \end{bmatrix} \pmod{26}$$

$$\begin{bmatrix} 1 & 24 \\ 8 & 19 \end{bmatrix} \begin{bmatrix} 19 \\ 11 \end{bmatrix} = \begin{bmatrix} 283 \\ 361 \end{bmatrix} = \begin{bmatrix} 23 \\ 23 \end{bmatrix} \pmod{26}$$

De la tabla 8.5, los equivalentes alfabéticos de estos vectores son

STRIKE NOW

Esto da como resultado el mensaje:

STRIKE NOW

Rompiendo un cifrado de Hill

El propósito de cifrar mensajes e información es prevenir que los “oponentes” conozcan sus contenidos. Los criptógrafos se preocupan por la *seguridad* de los cifrados, es decir, qué tan fácil es romperlos (descifrarlos por sus oponentes). La seguridad de un cifrado depende no solo de la complejidad matemática del algoritmo utilizado, sino también de la cantidad de datos que puedan obtener los atacantes y de su capacidad para analizarlos. Concluiremos esta sección discutiendo una técnica para romper cifrados de Hill.

Supongamos que puedes obtener algunos pares correspondientes de vectores de texto plano y cifrado de un mensaje interceptado. Por ejemplo, al examinar algún texto cifrado interceptado, podrías ser capaz de deducir que el mensaje comienza con la frase *DEAR SIR*, basándote en patrones comunes o suposiciones previas sobre el contenido del mensaje. Este tipo de información parcial puede ser crucial, ya que, con una pequeña cantidad de datos conocidos, es posible determinar la matriz de descifrado de un código de Hill y, en consecuencia, obtener acceso al resto del mensaje cifrado. Esto convierte a los cifrados de Hill en sistemas cuya seguridad está directamente relacionada con la dificultad de determinar dicha matriz.

Es un resultado básico en álgebra lineal que una transformación lineal está completamente determinada por sus valores en una base. Este principio sugiere que si tenemos un cifrado de Hill n , y

$$\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_n$$

son vectores linealmente independientes de texto plano cuyos vectores cifrados correspondientes son

$$A\mathbf{p}_1, A\mathbf{p}_2, \dots, A\mathbf{p}_n,$$

entonces hay suficiente información disponible para determinar la matriz A y, por ende, su inversa A^{-1} (mód m).

El siguiente teorema, cuya demostración se discute en los ejercicios, proporciona una manera de hacer esto, ilustrando un método directo para encontrar la matriz de descifrado a partir de un conjunto de datos conocidos.

Teorema 8.7.3: Sea $\mathbb{p}_1, \mathbb{p}_2, \dots, \mathbb{p}_n$ un conjunto de vectores de texto plano linealmente independientes, y sea $\mathbb{c}_1, \mathbb{c}_2, \dots, \mathbb{c}_n$ el conjunto de vectores cifrados correspondientes en un cifrado de Hill n . Si

$$P = \begin{bmatrix} \mathbb{p}_1^T \\ \mathbb{p}_2^T \\ \vdots \\ \mathbb{p}_n^T \end{bmatrix}$$

es la matriz $n \times n$ con los vectores fila $\mathbb{p}_1^T, \mathbb{p}_2^T, \dots, \mathbb{p}_n^T$, y si

$$C = \begin{bmatrix} \mathbb{c}_1^T \\ \mathbb{c}_2^T \\ \vdots \\ \mathbb{c}_n^T \end{bmatrix}$$

es la matriz $n \times n$ con los vectores fila $\mathbb{c}_1^T, \mathbb{c}_2^T, \dots, \mathbb{c}_n^T$, entonces la secuencia de operaciones elementales que transforma C en la identidad I transforma P en la traspuesta de la matriz inversa $(A^{-1})^T = A^{-T}$.

Este teorema nos dice que para encontrar la traspuesta de la matriz de descifrado A^{-1} , debemos encontrar una secuencia de operaciones de fila que transforme C en I y luego realizar esta misma secuencia de operaciones en P . El siguiente ejemplo ilustra un algoritmo simple para hacer esto.

Ejemplo 8.7.8: El siguiente cifrado de Hill 2 es interceptado:

I O S B T G X E S P X H O P D E

Descifre el mensaje, dado que comienza con la palabra *DEAR*.

Solución: A partir de la tabla 8.5, podemos asignar un valor numérico a cada letra del texto plano conocido, según su posición en el alfabeto estándar. Así, el texto plano *DEAR* se convierte en los siguientes valores numéricos correspondientes

$$\begin{array}{cccc} D & E & A & R \\ 4 & 5 & 1 & 18 \end{array}$$

y el equivalente numérico del texto cifrado correspondiente es

$$\begin{array}{cccc} I & O & S & B \\ 9 & 15 & 19 & 2 \end{array}$$

Por lo tanto, los vectores de texto plano y cifrado correspondientes son

$$\begin{aligned} \mathbb{p}_1 &= \begin{bmatrix} 4 \\ 5 \end{bmatrix} \leftrightarrow \mathbb{c}_1 = \begin{bmatrix} 9 \\ 15 \end{bmatrix} \\ \mathbb{p}_2 &= \begin{bmatrix} 1 \\ 18 \end{bmatrix} \leftrightarrow \mathbb{c}_2 = \begin{bmatrix} 19 \\ 2 \end{bmatrix} \end{aligned}$$

Queremos reducir

$$C = \begin{bmatrix} \mathbb{c}_1^T \\ \mathbb{c}_2^T \end{bmatrix} = \begin{bmatrix} 9 & 15 \\ 19 & 2 \end{bmatrix}$$

a la identidad I mediante operaciones elementales por renglones y simultáneamente aplicar estas operaciones a

$$P = \begin{bmatrix} \mathbb{p}_1^T \\ \mathbb{p}_2^T \end{bmatrix} = \begin{bmatrix} 4 & 5 \\ 1 & 18 \end{bmatrix}$$

para obtener A^{-T} , se reduce $[C \mid P]$ hasta formar $[I \mid A^{-T}]$.

$$\left[\begin{array}{cc|cc} 9 & 15 & 4 & 5 \\ 19 & 2 & 1 & 18 \end{array} \right]$$

Formamos la matriz $[C \mid P]$

$$\left[\begin{array}{cc|cc} 1 & 45 & 12 & 15 \\ 19 & 2 & 1 & 18 \end{array} \right]$$

Multiplicamos la primera fila por $9^{-1} = 3$

$$\left[\begin{array}{cc|cc} 1 & 19 & 12 & 15 \\ 19 & 2 & 1 & 18 \end{array} \right]$$

Reemplazamos 45 por su residuo módulo 26

$$\left[\begin{array}{cc|cc} 1 & 19 & 12 & 15 \\ 0 & -359 & -227 & -267 \end{array} \right]$$

Sumamos -19 veces la primera fila a la segunda

$$\left[\begin{array}{cc|cc} 1 & 19 & 12 & 15 \\ 0 & 5 & 7 & 19 \end{array} \right]$$

Reemplazamos las entradas en la segunda fila por sus residuos módulo 26

$$\left[\begin{array}{cc|cc} 1 & 19 & 12 & 15 \\ 0 & 1 & 147 & 399 \end{array} \right]$$

Multiplicamos la segunda fila por $5^{-1} = 21$

$$\left[\begin{array}{cc|cc} 1 & 19 & 12 & 15 \\ 0 & 1 & 17 & 9 \end{array} \right]$$

Reemplazamos las entradas en la segunda fila por sus residuos módulo 26

$$\left[\begin{array}{cc|cc} 1 & 0 & -311 & -156 \\ 0 & 1 & 17 & 9 \end{array} \right]$$

Sumamos -19 veces la segunda fila a la primera

$$\left[\begin{array}{cc|cc} 1 & 0 & 1 & 0 \\ 0 & 1 & 17 & 9 \end{array} \right]$$

Reemplazamos las entradas en la primera fila por sus residuos módulo 26

Por lo tanto,

$$A^{-T} = \begin{bmatrix} 1 & 0 \\ 17 & 9 \end{bmatrix}$$

Entonces, la matriz de descifrado es

$$A^{-1} = \begin{bmatrix} 1 & 17 \\ 0 & 9 \end{bmatrix}$$

Para descifrar el mensaje, primero agrupamos el texto cifrado en pares y encontramos el equivalente numérico de cada letra

$$\begin{array}{cccccccccccccccc} I & O & S & B & T & G & X & E & S & P & X & H & O & P & D & E \\ 9 & 15 & 19 & 2 & 20 & 7 & 24 & 5 & 19 & 16 & 24 & 8 & 15 & 16 & 4 & 5 \end{array}$$

A continuación, multiplicamos sucesivamente los vectores de texto cifrado por la izquierda con A^{-1} y encontramos los equivalentes alfabéticos de los pares de texto plano resultantes:

$$\begin{array}{l} \begin{bmatrix} 1 & 17 \\ 0 & 9 \end{bmatrix} \begin{bmatrix} 9 \\ 15 \end{bmatrix} = \begin{bmatrix} 4 \\ 5 \end{bmatrix} \quad \begin{array}{l} D \\ E \end{array} \\ \begin{bmatrix} 1 & 17 \\ 0 & 9 \end{bmatrix} \begin{bmatrix} 19 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 18 \end{bmatrix} \quad \begin{array}{l} A \\ R \end{array} \\ \begin{bmatrix} 1 & 17 \\ 0 & 9 \end{bmatrix} \begin{bmatrix} 20 \\ 7 \end{bmatrix} = \begin{bmatrix} 9 \\ 11 \end{bmatrix} \quad \begin{array}{l} I \\ K \end{array} \\ \begin{bmatrix} 1 & 17 \\ 0 & 9 \end{bmatrix} \begin{bmatrix} 24 \\ 5 \end{bmatrix} = \begin{bmatrix} 5 \\ 19 \end{bmatrix} \quad \begin{array}{l} E \\ S \end{array} \\ \begin{bmatrix} 1 & 17 \\ 0 & 9 \end{bmatrix} \begin{bmatrix} 19 \\ 16 \end{bmatrix} = \begin{bmatrix} 5 \\ 14 \end{bmatrix} \quad \begin{array}{l} E \\ N \end{array} \\ \begin{bmatrix} 1 & 17 \\ 0 & 9 \end{bmatrix} \begin{bmatrix} 24 \\ 8 \end{bmatrix} = \begin{bmatrix} 4 \\ 20 \end{bmatrix} \quad \begin{array}{l} D \\ T \end{array} \\ \begin{bmatrix} 1 & 17 \\ 0 & 9 \end{bmatrix} \begin{bmatrix} 15 \\ 16 \end{bmatrix} = \begin{bmatrix} 1 \\ 14 \end{bmatrix} \quad \begin{array}{l} A \\ N \end{array} \\ \begin{bmatrix} 1 & 17 \\ 0 & 9 \end{bmatrix} \begin{bmatrix} 4 \\ 5 \end{bmatrix} = \begin{bmatrix} 11 \\ 19 \end{bmatrix} \quad \begin{array}{l} K \\ S \end{array} \end{array} \quad (\text{mód } 26)$$

8.8 Genética

Herencia de rasgos

En esta sección examinamos la herencia de rasgos en animales o plantas. El rasgo hereditario considerado se asume gobernado por un conjunto de dos genes, designados como A y a . Bajo *herencia autosómica*, cada individuo en la población, de cualquier género, posee dos de estos genes, pudiendo formarse las combinaciones AA , Aa y aa . Este par de genes constituye el *genotipo* del individuo y determina cómo el rasgo controlado por los genes se manifiesta en él.

Por ejemplo, en las flores *boca de dragón*, un conjunto de dos genes determina el color de las flores. El genotipo AA produce flores rojas, Aa flores rosadas y aa flores blancas. En los humanos, el color de los ojos está controlado por la herencia autosómica. Los genotipos AA y Aa producen ojos marrones, mientras que aa produce ojos azules. En este caso, decimos que el gen A *domina* al gen a , o que el gen a es *recesivo* al gen A , ya que el genotipo Aa tiene la misma manifestación externa que el genotipo AA .

Además de la herencia autosómica, también discutimos la *herencia ligada al cromosoma X*. En este tipo de herencia, el macho de la especie posee solo uno de los dos posibles genes (A o a), mientras que la hembra posee un par de los dos genes (AA , Aa o aa). En humanos, algunos rasgos controlados por la herencia ligada al cromosoma X incluyen el daltonismo, la calvicie hereditaria, la hemofilia y la distrofia muscular, entre otros.

A continuación, explicamos cómo se transmiten los genes de los padres a su descendencia para ambos tipos de herencia. Construimos modelos matriciales que proporcionan los genotipos probables de la descendencia en términos de los genotipos de los padres. Estos modelos se utilizan para seguir la distribución genotípica de una población a través de generaciones sucesivas.

Herencia autosómica

En la herencia autosómica, un individuo hereda un gen de cada par de genes de sus padres para formar su propio par. Según lo que sabemos, es cuestión de azar cuál de los dos genes transmite un padre a la descendencia. Por ejemplo, si uno de los padres tiene el genotipo Aa , es igualmente probable que la descendencia herede el gen A o el gen a de ese padre. Si uno de los padres tiene genotipo aa y el otro tiene genotipo Aa , la descendencia siempre heredará un gen a del padre aa y recibirá, con igual probabilidad, un gen A o un gen a del padre Aa . Por consiguiente, cada descendiente tiene igual probabilidad de tener genotipo aa o Aa . En la tabla 8.7 listamos las probabilidades de los genotipos de la descendencia para todas las combinaciones posibles de los genotipos de los padres.

Tabla 8.7 Distribución de genotipos de la descendencia en función de los genotipos parentales

		Genotipos de los padres					
		$AA - AA$	$AA - Aa$	$AA - aa$	$Aa - Aa$	$Aa - aa$	$aa - aa$
Genotipo de la descendencia	AA	1	$\frac{1}{2}$	0	$\frac{1}{4}$	0	0
	Aa	0	$\frac{1}{2}$	1	$\frac{1}{2}$	$\frac{1}{2}$	0
	aa	0	0	0	$\frac{1}{4}$	$\frac{1}{2}$	1

Ejemplo 8.8.1: Supongamos que un agricultor tiene una población de plantas con los genotipos AA , Aa y aa , y desea implementar un programa de reproducción donde cada planta sea fertilizada con una de genotipo AA , reemplazándose por uno de sus descendientes. Queremos expresar la distribución de genotipos en la población tras n generaciones. Para $n = 0, 1, 2, \dots$, definamos

a_n = fracción de plantas con genotipo AA en la n -ésima generación

b_n = fracción de plantas con genotipo Aa en la n -ésima generación

c_n = fracción de plantas con genotipo aa en la n -ésima generación

De esta manera, a_0, b_0, c_0 especifican la distribución inicial de los genotipos. También tenemos que

$$a_n + b_n + c_n = 1 \quad \text{para } n = 0, 1, 2, \dots$$

A partir de la tabla 8.7, podemos determinar la distribución de genotipos de cada generación a partir de la distribución de genotipos de la generación previa mediante las siguientes ecuaciones:

$$\begin{aligned} a_n &= a_{n-1} + \frac{1}{2}b_{n-1} \\ b_n &= c_{n-1} + \frac{1}{2}b_{n-1} \quad n = 1, 2, \dots \\ c_n &= 0 \end{aligned} \quad (8.8.1)$$

Por ejemplo, la primera de estas tres ecuaciones establece que todos los descendientes de una planta con genotipo AA tendrán genotipo AA bajo este programa de reproducción y que la mitad de los descendientes de una planta con genotipo Aa tendrán genotipo AA . Las ecuaciones (8.8.1) pueden escribirse en notación matricial como

$$\mathbf{x}^{(n)} = M\mathbf{x}^{(n-1)}, \quad n = 1, 2, \dots \quad (8.8.2)$$

donde

$$\mathbf{x}^{(n)} = \begin{bmatrix} a_n \\ b_n \\ c_n \end{bmatrix}, \quad \mathbf{x}^{(n-1)} = \begin{bmatrix} a_{n-1} \\ b_{n-1} \\ c_{n-1} \end{bmatrix}, \quad M = \begin{bmatrix} 1 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

Obsérvese que las tres columnas de la matriz M son las mismas que las tres primeras columnas de la tabla 8.7. De la ecuación (8.8.2) se deduce que

$$\mathbf{x}^{(n)} = M\mathbf{x}^{(n-1)} = M^2\mathbf{x}^{(n-2)} = \dots = M^n\mathbf{x}^{(0)} \quad (8.8.3)$$

En consecuencia, si podemos encontrar una expresión explícita para M^n , podemos usar (8.8.3) para obtener una expresión explícita para $\mathbf{x}^{(n)}$. Para encontrar una expresión explícita para M^n , primero diagonalizamos M . Es decir, encontramos una matriz invertible P y una matriz diagonal D tales que

$$M = PDP^{-1} \quad (8.8.4)$$

Con dicha diagonalización, tenemos

$$M^n = PD^nP^{-1}, \quad n = 1, 2, \dots$$

donde

$$D^n = \begin{bmatrix} \lambda_1^n & 0 & \dots & 0 \\ 0 & \lambda_2^n & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_k^n \end{bmatrix}$$

La diagonalización de M se realiza encontrando sus valores propios y los vectores propios correspondientes. Estos son los siguientes

$$\begin{aligned}\lambda_1 &= 1 & \lambda_2 &= \frac{1}{2} & \lambda_3 &= 0 \\ \mathbf{v}_1 &= \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} & \mathbf{v}_2 &= \begin{bmatrix} -1 \\ -2 \\ 1 \end{bmatrix} & \mathbf{v}_3 &= \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}\end{aligned}$$

Así, en la ecuación (8.8.4) tenemos

$$D = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

y

$$P = [\mathbf{v}_1 \quad \mathbf{v}_2 \quad \mathbf{v}_3] = \begin{bmatrix} 1 & -1 & 1 \\ 0 & -2 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

Por lo tanto,

$$\mathbf{x}^{(n)} = P D^n P^{-1} \mathbf{x}^{(0)} = \begin{bmatrix} 1 & -1 & 1 \\ 0 & -2 & 1 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \left(\frac{1}{2}\right)^n & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 0 & -1 & -2 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} a_0 \\ b_0 \\ c_0 \end{bmatrix}$$

o bien,

$$\begin{aligned}\mathbf{x}^{(n)} &= \begin{bmatrix} a_n \\ b_n \\ c_n \end{bmatrix} = \begin{bmatrix} 1 & 1 - \left(\frac{1}{2}\right)^n & 1 - \left(\frac{1}{2}\right)^{n-1} \\ 0 & \left(\frac{1}{2}\right)^n & \left(\frac{1}{2}\right)^{n-1} \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} a_0 \\ b_0 \\ c_0 \end{bmatrix} \\ &= \begin{bmatrix} a_0 + b_0 + c_0 - \left(\frac{1}{2}\right)^n b_0 - \left(\frac{1}{2}\right)^{n-1} c_0 \\ \left(\frac{1}{2}\right)^n b_0 + \left(\frac{1}{2}\right)^{n-1} c_0 \\ 0 \end{bmatrix}\end{aligned}$$

Usando el hecho de que $a_0 + b_0 + c_0 = 1$, tenemos que

$$\begin{aligned}a_n &= 1 - \left(\frac{1}{2}\right)^n b_0 - \left(\frac{1}{2}\right)^{n-1} c_0 \\ b_n &= \left(\frac{1}{2}\right)^n b_0 + \left(\frac{1}{2}\right)^{n-1} c_0 \\ c_n &= 0\end{aligned} \quad n = 1, 2, \dots \quad (8.8.5)$$

Estas son fórmulas explícitas para las fracciones de los tres genotipos en la generación n en términos de las fracciones iniciales de los genotipos. Como $(1/2)^n$ tiende a cero cuando $n \rightarrow \infty$, se sigue de estas ecuaciones que

$$a_n \rightarrow 1, \quad b_n \rightarrow 0, \quad c_n = 0.$$

Es decir, en el límite, todas las plantas en la población serán de genotipo AA .

Ejemplo 8.8.2: Podemos modificar el ejemplo anterior para que cada planta sea fertilizada con otra de su mismo genotipo. Con la misma notación, se tiene

$$\mathbf{x}^{(n)} = M^n \mathbf{x}^{(0)}, \quad \text{donde } M = \begin{bmatrix} 1 & \frac{1}{4} & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{4} & 1 \end{bmatrix}$$

Las columnas de M coinciden con las de la tabla 8.7 para los genotipos $AA - AA$, $Aa - Aa$ y $aa - aa$. Sus valores propios son

$$\lambda_1 = 1, \quad \lambda_2 = 1, \quad \lambda_3 = \frac{1}{2}.$$

El valor propio λ_1 tiene multiplicidad dos y un espacio propio bidimensional. Con dos vectores propios linealmente independientes $\mathbf{v}_1, \mathbf{v}_2$ en ese espacio, y un vector propio \mathbf{v}_3 para λ_3 , se tiene:

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$$

Los cálculos para $\mathbf{x}^{(n)}$ son entonces:

$$\begin{aligned} \mathbf{x}^{(n)} &= M^n \mathbf{x}^{(0)} = P D^n P^{-1} \mathbf{x}^{(0)} \\ &= \begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & -2 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \left(\frac{1}{2}\right)^n \end{bmatrix} \begin{bmatrix} 1 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 1 \\ 0 & -\frac{1}{2} & 1 \end{bmatrix} \begin{bmatrix} a_0 \\ b_0 \\ c_0 \end{bmatrix} \\ &= \begin{bmatrix} 1 & \frac{1}{2} - \left(\frac{1}{2}\right)^{n+1} & 0 \\ 0 & \left(\frac{1}{2}\right)^n & 0 \\ 0 & \frac{1}{2} - \left(\frac{1}{2}\right)^{n+1} & 1 \end{bmatrix} \begin{bmatrix} a_0 \\ b_0 \\ c_0 \end{bmatrix} \end{aligned}$$

Así, obtenemos las siguientes expresiones

$$\begin{aligned} a_n &= a_0 + \left[\frac{1}{2} - \left(\frac{1}{2}\right)^{n+1} \right] b_0 \\ b_n &= \left(\frac{1}{2}\right)^n b_0 \\ c_n &= c_0 + \left[\frac{1}{2} - \left(\frac{1}{2}\right)^{n+1} \right] b_0 \end{aligned} \quad n = 1, 2, \dots \quad (8.8.6)$$

Cuando $n \rightarrow \infty$, entonces $\left(\frac{1}{2}\right)^n \rightarrow 0$ y $\left(\frac{1}{2}\right)^{n+1} \rightarrow 0$, y por consiguiente

$$a_n \rightarrow a_0 + \frac{1}{2} b_0, \quad b_n \rightarrow 0, \quad c_n \rightarrow c_0 + \frac{1}{2} b_0.$$

Por lo tanto, fertilizar cada planta con una de su propio genotipo produce una población que en el límite contiene solo los genotipos AA y aa .

Enfermedades recesivas autosómicas

Existen muchas enfermedades genéticas regidas por la herencia autosómica, en la que un gen normal A domina a un gen anormal a . El genotipo AA corresponde a un individuo normal; el genotipo Aa es portador de la enfermedad pero no está afectado por ella; y el genotipo aa está afectado por la enfermedad. En los humanos, tales enfermedades genéticas a menudo están asociadas con ciertos grupos raciales. Por ejemplo, la fibrosis quística (predominante entre caucásicos), la anemia de células falciformes (predominante entre personas de origen africano), la anemia de Cooley (predominante entre personas de origen mediterráneo) y la enfermedad de Tay-Sachs (predominante entre judíos de Europa oriental).

Supongamos que un criador de animales tiene una población de animales que porta una enfermedad recesiva autosómica. Además, supongamos que aquellos animales afectados por la enfermedad (aa) no sobreviven hasta la madurez. Una posible manera de controlar dicha enfermedad es que el criador siempre cruce a una hembra, independientemente de su genotipo, con un macho normal. De esta forma, todos los descendientes tendrán o bien un padre y una madre normales ($AA - AA$) o un padre normal y una madre portadora ($AA - Aa$). No puede haber cruces $AA - aa$ ya que los animales con genotipo aa no sobreviven hasta la madurez. Bajo este programa de cría, ningún descendiente estará afectado por la enfermedad en las generaciones futuras, aunque seguirán existiendo portadores. Ahora determinaremos la fracción de portadores en futuras generaciones. Definimos

$$\mathbf{x}^{(n)} = \begin{bmatrix} a_n \\ b_n \end{bmatrix}, \quad n = 1, 2, \dots$$

donde:

a_n = fracción de la población con genotipo AA en la n -ésima generación

b_n = fracción de la población con genotipo Aa en la n -ésima generación

Debido a que cada descendiente tiene al menos un padre normal, podemos considerar el programa de apareamiento controlado como uno de apareamiento continuo con genotipo AA , como en el primer. Por lo tanto, la transición de las distribuciones genotípicas de una generación a la siguiente está gobernada por la ecuación

$$\mathbf{x}^{(n)} = \mathbf{M}\mathbf{x}^{(n-1)}, \quad n = 1, 2, \dots$$

donde

$$\mathbf{M} = \begin{bmatrix} 1 & \frac{1}{2} \\ 0 & \frac{1}{2} \end{bmatrix}$$

Dado que conocemos la distribución inicial $\mathbf{x}^{(0)}$, la distribución de genotipos en la n -ésima generación está dada por

$$\mathbf{x}^{(n)} = \mathbf{M}^n \mathbf{x}^{(0)}, \quad n = 1, 2, \dots$$

La diagonalización de \mathbf{M} se lleva a cabo fácilmente y conduce a

$$\begin{aligned} \mathbf{x}^{(n)} &= \mathbf{P}\mathbf{D}^n\mathbf{P}^{-1}\mathbf{x}^{(0)} = \begin{bmatrix} 1 & 1 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \left(\frac{1}{2}\right)^n \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} a_0 \\ b_0 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 1 - \left(\frac{1}{2}\right)^n \\ 0 & \left(\frac{1}{2}\right)^n \end{bmatrix} \begin{bmatrix} a_0 \\ b_0 \end{bmatrix} = \begin{bmatrix} a_0 + b_0 - \left(\frac{1}{2}\right)^n b_0 \\ \left(\frac{1}{2}\right)^n b_0 \end{bmatrix} \end{aligned}$$

Dado que $a_0 + b_0 = 1$, tenemos

$$\begin{aligned} a_n &= 1 - \left(\frac{1}{2}\right)^n b_0 \\ b_n &= \left(\frac{1}{2}\right)^n b_0 \end{aligned} \quad n = 1, 2, \dots \quad (8.8.7)$$

Por lo tanto, cuando $n \rightarrow \infty$, tenemos:

$$a_n \rightarrow 1, \quad b_n \rightarrow 0,$$

lo que significa que en el límite no habrá portadores en la población.

De (8.8.7), vemos que

$$b_n = \frac{1}{2} b_{n-1}, \quad n = 1, 2, \dots \quad (8.8.8)$$

Es decir, la fracción de portadores en cada generación es la mitad de la fracción de portadores en la generación anterior. También sería interesante investigar la propagación de portadores bajo apareamiento aleatorio, cuando dos animales se aparean sin tener en cuenta sus genotipos. Desafortunadamente, dicho apareamiento aleatorio lleva a ecuaciones no lineales, y las técnicas de esta sección no son aplicables. Sin embargo, con otras técnicas se puede demostrar que, bajo apareamiento aleatorio, la ecuación (8.8.8) se reemplaza por

$$b_n = \frac{b_{n-1}}{1 + \frac{1}{2} b_{n-1}}, \quad n = 1, 2, \dots \quad (8.8.9)$$

Como ejemplo numérico, supongamos que el criador comienza con una población en la que el 10 % de los animales son portadores. Bajo el programa de apareamiento controlado gobernado por la ecuación (8.8.8), el porcentaje de portadores puede reducirse al 5 % en una generación. Sin embargo, bajo apareamiento aleatorio, la ecuación (8.8.9) predice que el 9.5 % de la población serán portadores después de una generación ($b_n = 0.095$ si $b_{n-1} = 0.10$). Además, bajo apareamiento controlado ningún descendiente estará afectado por la enfermedad, pero con apareamiento aleatorio se puede demostrar que aproximadamente 1 de cada 400 descendientes nacerá con la enfermedad cuando el 10 % de la población sean portadores.

Herencia ligada al cromosoma X

Como se mencionó en la introducción, en la herencia ligada al cromosoma X, el macho posee un solo gen (A o a), mientras que la hembra posee dos genes (AA , Aa o aa). El término *ligado al cromosoma X* se utiliza porque estos genes se encuentran en el cromosoma X, que está presente en una sola copia en los machos y en dos copias en las hembras.

La herencia de estos genes sigue un patrón definido: un macho hereda exclusivamente el cromosoma X de su madre, lo que significa que recibe uno de los dos genes de ella con igual probabilidad. Por otro lado, una hembra hereda el cromosoma X de su padre (portador de un solo gen) y uno de los dos cromosomas X de su madre, seleccionados con igual probabilidad. Este mecanismo asegura una distribución específica de los genotipos entre los descendientes.

Por lo tanto, los patrones de herencia ligados al cromosoma X tienen implicaciones tanto en la probabilidad como en la genética poblacional. Por ejemplo, ciertas enfermedades genéticas como el daltonismo o la hemofilia son más frecuentes en los machos debido a la falta de un segundo cromosoma X que pueda compensar un gen recesivo defectuoso. Los lectores familiarizados con conceptos básicos de probabilidad pueden verificar que este tipo de herencia conduce a las probabilidades de genotipo indicadas en la tabla 8.8, las cuales reflejan las combinaciones posibles de los genes transmitidos por los padres.

Tabla 8.8 Probabilidades de genotipos de la descendencia basadas en la herencia ligada al cromosoma X

			Genotipos de padres (Padre, Madre)					
			(A, AA)	(A, Aa)	(A, aa)	(a, AA)	(a, Aa)	(a, aa)
Descendiente	Macho	A	1	$\frac{1}{2}$	0	1	$\frac{1}{2}$	0
		a	0	$\frac{1}{2}$	1	0	$\frac{1}{2}$	1
	Hembra	AA	1	$\frac{1}{2}$	0	0	0	0
		Aa	0	$\frac{1}{2}$	1	1	$\frac{1}{2}$	0
		aa	0	0	0	0	$\frac{1}{2}$	1

Discutiremos un programa de consanguinidad en relación con la herencia ligada al cromosoma X . Este programa consiste en iniciar con un macho y una hembra como progenitores iniciales; seleccionar al azar dos de sus descendientes, uno de cada género, y cruzarlos para obtener la siguiente generación; repetir el proceso seleccionando dos descendientes de esta nueva generación, nuevamente uno de cada género, y cruzarlos entre sí; y continuar de esta manera en generaciones sucesivas.

Dichos cruces consanguíneos son una práctica común en el manejo genético de animales, especialmente cuando se busca preservar características específicas de una población. (En los humanos, históricamente, este tipo de matrimonios consanguíneos, como entre hermanos o primos cercanos, fueron utilizados por los gobernantes del antiguo Egipto para mantener pura la línea real y reforzar la legitimidad dinástica. Sin embargo, estos también traían consecuencias genéticas significativas debido a la reducción de la diversidad genética).

El par original macho-hembra puede ser uno de los seis tipos, correspondientes a las seis columnas de la tabla 8.8

$$(A, AA), \quad (A, Aa), \quad (A, aa), \quad (a, AA), \quad (a, Aa), \quad (a, aa)$$

Los pares de hermanos cruzados en cada generación sucesiva tienen ciertas probabilidades de ser uno de estos seis tipos. Para calcular estas probabilidades, para $n = 0, 1, 2, \dots$, definamos

a_n = probabilidad de que el par cruzado en la generación n sea del tipo (A, AA)

b_n = probabilidad de que el par cruzado en la generación n sea del tipo (A, Aa)

c_n = probabilidad de que el par cruzado en la generación n sea del tipo (A, aa)

d_n = probabilidad de que el par cruzado en la generación n sea del tipo (a, AA)

e_n = probabilidad de que el par cruzado en la generación n sea del tipo (a, Aa)

f_n = probabilidad de que el par cruzado en la generación n sea del tipo (a, aa)

Con estas probabilidades, formamos un vector columna

$$\mathbf{x}^{(n)} = \begin{bmatrix} a_n \\ b_n \\ c_n \\ d_n \\ e_n \\ f_n \end{bmatrix}, \quad n = 0, 1, 2, \dots$$

De la tabla 8.8 se sigue que

$$\mathbf{x}^{(n)} = \mathbf{M}\mathbf{x}^{(n-1)}, \quad n = 1, 2, \dots \quad (8.8.10)$$

donde

$$M = \begin{matrix} & \begin{matrix} (A, AA) & (A, Aa) & (A, aa) & (a, AA) & (a, Aa) & (a, aa) \end{matrix} \\ \begin{bmatrix} 1 & \frac{1}{4} & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{4} & 0 & 1 & \frac{1}{4} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{4} & 0 \\ 0 & \frac{1}{4} & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{4} & 1 & 0 & \frac{1}{4} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{4} & 1 \end{bmatrix} & \begin{matrix} (A, AA) \\ (A, Aa) \\ (A, aa) \\ (a, AA) \\ (a, Aa) \\ (a, aa) \end{matrix} \end{matrix}$$

Por ejemplo, supongamos que en la generación $n-1$, el par de hermanos cruzados es del tipo (A, Aa) . Entonces, su descendiente macho será de genotipo A o a con igual probabilidad, y su descendiente hembra será de genotipo AA o Aa también con igual probabilidad. Dado que uno de los descendientes machos y uno de las descendientes hembras son elegidos al azar para el cruce, el siguiente par de hermanos será de uno de los tipos (A, AA) , (A, Aa) , (a, AA) o (a, Aa) con igual probabilidad. Por lo tanto, la segunda columna de M contiene un $1/4$ en cada una de las cuatro filas correspondientes a estos cuatro pares de hermanos.

Como en nuestros ejemplos anteriores, se deduce de la ecuación (8.8.10) que

$$\mathbf{x}^{(n)} = M^n \mathbf{x}^{(0)}, \quad n = 1, 2, \dots \quad (8.8.11)$$

Después de cálculos extensos, los valores y vectores propios de M resultan ser

$$\begin{aligned} \lambda_1 &= 1 & \lambda_2 &= 1 & \lambda_3 &= \frac{1}{2} \\ \mathbf{v}_1 &= \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} & \mathbf{v}_2 &= \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} & \mathbf{v}_3 &= \begin{bmatrix} -1 \\ 2 \\ -1 \\ 1 \\ -2 \\ 1 \end{bmatrix} \\ \lambda_4 &= -\frac{1}{2} & \lambda_5 &= \frac{1 + \sqrt{5}}{4} & \lambda_6 &= \frac{1 - \sqrt{5}}{4} \\ \mathbf{v}_4 &= \begin{bmatrix} 1 \\ -6 \\ -3 \\ 3 \\ 6 \\ -1 \end{bmatrix} & \mathbf{v}_5 &= \begin{bmatrix} \frac{-3 - \sqrt{5}}{4} \\ 1 \\ \frac{-1 + \sqrt{5}}{4} \\ \frac{-1 + \sqrt{5}}{4} \\ 1 \\ \frac{-3 - \sqrt{5}}{4} \end{bmatrix} & \mathbf{v}_6 &= \begin{bmatrix} \frac{-3 + \sqrt{5}}{4} \\ 1 \\ \frac{-1 - \sqrt{5}}{4} \\ \frac{-1 - \sqrt{5}}{4} \\ 1 \\ \frac{-3 + \sqrt{5}}{4} \end{bmatrix} \end{aligned}$$

La diagonalización de M conduce a:

$$\mathbf{x}^{(n)} = P D^n P^{-1} \mathbf{x}^{(0)}, \quad n = 1, 2, \dots \quad (8.8.12)$$

donde

$$P = \begin{bmatrix} 1 & 0 & -1 & \frac{-3-\sqrt{5}}{4} & \frac{-3+\sqrt{5}}{4} \\ 0 & 0 & 2 & -6 & 1 \\ 0 & 0 & -1 & \frac{-1+\sqrt{5}}{4} & \frac{-1-\sqrt{5}}{4} \\ 0 & 0 & 1 & \frac{-1+\sqrt{5}}{4} & \frac{-1-\sqrt{5}}{4} \\ 0 & -2 & 6 & 1 & 1 \\ 0 & 1 & 1 & \frac{-3-\sqrt{5}}{4} & \frac{-3+\sqrt{5}}{4} \end{bmatrix}$$

$$D^n = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & \left(\frac{1}{2}\right)^n & 0 & 0 & 0 \\ 0 & 0 & 0 & \left(-\frac{1}{2}\right)^n & 0 & 0 \\ 0 & 0 & 0 & 0 & \left(\frac{1+\sqrt{5}}{4}\right)^n & 0 \\ 0 & 0 & 0 & 0 & 0 & \left(\frac{1-\sqrt{5}}{4}\right)^n \end{bmatrix}$$

$$P^{-1} = \begin{bmatrix} 1 & \frac{2}{3} & \frac{1}{3} & \frac{2}{3} & \frac{1}{3} & 0 \\ 0 & \frac{1}{3} & \frac{2}{3} & \frac{1}{3} & \frac{2}{3} & 1 \\ 0 & \frac{1}{8} & -\frac{1}{4} & \frac{1}{4} & -\frac{1}{8} & 0 \\ 0 & -\frac{1}{24} & -\frac{1}{12} & \frac{1}{12} & \frac{1}{24} & 0 \\ 0 & \frac{5+\sqrt{5}}{20} & \frac{\sqrt{5}}{5} & \frac{\sqrt{5}}{5} & \frac{5+\sqrt{5}}{20} & 0 \\ 0 & \frac{5-\sqrt{5}}{20} & -\frac{\sqrt{5}}{5} & -\frac{\sqrt{5}}{5} & \frac{5-\sqrt{5}}{20} & 0 \end{bmatrix}$$

No escribiremos el producto matricial en la ecuación (8.8.12), ya que es bastante complicado. Sin embargo, si se da un vector específico $\mathbf{x}^{(0)}$, el cálculo para $\mathbf{x}^{(n)}$ no es demasiado engorroso.

Dado que los valores absolutos de las últimas cuatro entradas diagonales de D son menores que 1, observamos que, a medida que n tiende a infinito,

$$D^n \rightarrow \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Por lo tanto, de la ecuación (8.8.12),

$$\mathbb{x}^{(n)} \rightarrow P \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} P^{-1} \mathbb{x}^{(0)}$$

Realizando la multiplicación matricial en el lado derecho, obtenemos

$$\mathbb{x}^{(n)} \rightarrow \begin{bmatrix} a_0 + \frac{2}{3}b_0 + \frac{1}{3}c_0 + \frac{2}{3}d_0 + \frac{1}{3}e_0 \\ 0 \\ 0 \\ 0 \\ 0 \\ f_0 + \frac{1}{3}b_0 + \frac{2}{3}c_0 + \frac{1}{3}d_0 + \frac{2}{3}e_0 \end{bmatrix}$$

Es decir, en el límite, todos los pares de hermanos serán del tipo (A, AA) o del tipo (a, aa) . Por ejemplo, si los padres iniciales son del tipo (A, Aa) (es decir, $b_0 = 1$ y $a_0 = c_0 = d_0 = e_0 = f_0 = 0$), entonces, a medida que n tiende a infinito,

$$\mathbb{x}^{(n)} \rightarrow \begin{bmatrix} \frac{2}{3} \\ 0 \\ 0 \\ 0 \\ 0 \\ \frac{1}{3} \end{bmatrix}$$

Por lo tanto, en el límite, existe una probabilidad de $2/3$ de que los pares de hermanos sean del tipo (A, AA) , y una probabilidad de $1/3$ de que sean del tipo (a, aa) .

8.9 Ejercicios del Capítulo 8

Ejercicios de la sección 8.1

- Encuentra las ecuaciones de las rectas que pasan por los siguientes puntos:
 - $(1, -1), (2, 2)$
 - $(0, 1), (1, -1)$
- Encuentra las ecuaciones de los círculos que pasan por los siguientes puntos:
 - $(2, 6), (2, 0), (5, 3)$
 - $(2, -2), (3, 5), (-4, 6)$
- Encuentra la ecuación de la sección cónica que pasa por los puntos $(0, 0)$, $(0, -1)$, $(2, 0)$, $(2, -5)$ y $(4, -1)$.
- Encuentra las ecuaciones de los planos en el espacio tridimensional que pasan por los siguientes puntos:
 - $(1, 1, -3), (1, -1, 1), (0, -1, 2)$
 - $(2, 3, 1), (2, -1, -1), (1, 2, 1)$
- Modifique la ecuación (8.1.11) para que determine el plano que pasa por el origen y sea paralelo al plano que pasa por tres puntos no colineales especificados.
 - Encuentre los dos planos descritos en el inciso anterior correspondientes a los tripletes de puntos en los ejercicios 4a) y 4b).
- Encuentra las ecuaciones de las esferas en el espacio tridimensional que pasan por los siguientes puntos:
 - $(1, 2, 3), (-1, 2, 1), (1, 0, 1), (1, 2, -1)$
 - $(0, 1, -2), (1, 3, 1), (2, -1, 0), (3, 1, -1)$
- Demuestre que la ecuación (8.1.10) es la ecuación de la sección cónica que pasa por cinco puntos distintos dados en el plano.
- Demuestre que la ecuación (8.1.11) es la ecuación del plano en el espacio tridimensional que pasa por tres puntos no colineales dados.
- Demuestre que la ecuación (8.1.12) es la ecuación de la esfera en el espacio tridimensional que pasa por cuatro puntos no coplanares dados.
- Encuentre el determinante que da la ecuación para la parábola de la forma

$$c_1y + c_2x^2 + c_3x + c_4 = 0$$
 que pase por tres puntos no colineales dados en el plano.
- ¿En qué se convierte la ecuación (8.1.9) si los tres puntos (que deben ser distintos) son colineales?
- ¿En qué se convierte la ecuación (8.1.11) si los tres puntos (que deben ser distintos) son colineales?
- ¿En qué se convierte la ecuación (8.1.12) si los cuatro puntos son coplanares?

Los siguientes ejercicios están diseñados para resolverse utilizando una herramienta tecnológica. Típicamente, esto será MATLAB, MATHEMATICA o MAXIMA, pero también puede ser algún otro tipo de software de álgebra lineal.

- La ecuación general de una superficie cuadrática está dada por

$$a_1x^2 + a_2y^2 + a_3z^2 + a_4xy + a_5xz + a_6yz + a_7x + a_8y + a_9z + a_{10} = 0$$

Dado un conjunto de nueve puntos en esta superficie, es posible determinar su ecuación.

- a) Demuestra que si los nueve puntos (x_i, y_i, z_i) para $i = 1, 2, 3, \dots, 9$ están en esta superficie, y si determinan de manera única la ecuación de esta superficie, entonces su ecuación se puede obtener del siguiente determinante:

$$\begin{vmatrix} x^2 & y^2 & z^2 & xy & xz & yz & x & y & z & 1 \\ x_1^2 & y_1^2 & z_1^2 & x_1 y_1 & x_1 z_1 & y_1 z_1 & x_1 & y_1 & z_1 & 1 \\ x_2^2 & y_2^2 & z_2^2 & x_2 y_2 & x_2 z_2 & y_2 z_2 & x_2 & y_2 & z_2 & 1 \\ x_3^2 & y_3^2 & z_3^2 & x_3 y_3 & x_3 z_3 & y_3 z_3 & x_3 & y_3 & z_3 & 1 \\ x_4^2 & y_4^2 & z_4^2 & x_4 y_4 & x_4 z_4 & y_4 z_4 & x_4 & y_4 & z_4 & 1 \\ x_5^2 & y_5^2 & z_5^2 & x_5 y_5 & x_5 z_5 & y_5 z_5 & x_5 & y_5 & z_5 & 1 \\ x_6^2 & y_6^2 & z_6^2 & x_6 y_6 & x_6 z_6 & y_6 z_6 & x_6 & y_6 & z_6 & 1 \\ x_7^2 & y_7^2 & z_7^2 & x_7 y_7 & x_7 z_7 & y_7 z_7 & x_7 & y_7 & z_7 & 1 \\ x_8^2 & y_8^2 & z_8^2 & x_8 y_8 & x_8 z_8 & y_8 z_8 & x_8 & y_8 & z_8 & 1 \\ x_9^2 & y_9^2 & z_9^2 & x_9 y_9 & x_9 z_9 & y_9 z_9 & x_9 & y_9 & z_9 & 1 \end{vmatrix} = 0$$

- b) Usa el resultado del inciso anterior para determinar la ecuación de la superficie cuadrática que pasa por los puntos $(1, 2, 3)$, $(2, 1, 7)$, $(0, 4, 6)$, $(3, -1, 4)$, $(3, 0, 11)$, $(-1, 5, 8)$, $(9, -8, 3)$, $(4, 5, 3)$ y $(-2, 6, 10)$.

15. a) Un hiperplano en el espacio n -dimensional \mathbb{R}^n está dado como

$$H = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n \mid a_1 x_1 + a_2 x_2 + a_3 x_3 + \dots + a_n x_n + a_{n+1} = 0 \right\}$$

donde a_i , con $i = 1, 2, 3, \dots, n + 1$, son constantes, no todas cero. Demuestre que la ecuación del hiperplano puede escribirse en forma de determinante como

$$\begin{vmatrix} x_1 & x_2 & x_3 & \cdots & x_n & 1 \\ x_{11} & x_{21} & x_{31} & \cdots & x_{n1} & 1 \\ x_{12} & x_{22} & x_{32} & \cdots & x_{n2} & 1 \\ x_{13} & x_{23} & x_{33} & \cdots & x_{n3} & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ x_{1n} & x_{2n} & x_{3n} & \cdots & x_{nn} & 1 \end{vmatrix} = 0$$

- b) Determine la ecuación del hiperplano en \mathbb{R}^9 que pasa a través de los siguientes nueve puntos:

$$\begin{array}{ll} (1, 2, 3, 4, 5, 6, 7, 8, 9) & (2, 3, 4, 5, 6, 7, 8, 9, 1) \\ (3, 4, 5, 6, 7, 8, 9, 1, 2) & (4, 5, 6, 7, 8, 9, 1, 2, 3) \\ (5, 6, 7, 8, 9, 1, 2, 3, 4) & (6, 7, 8, 9, 1, 2, 3, 4, 5) \\ (7, 8, 9, 1, 2, 3, 4, 5, 6) & (8, 9, 1, 2, 3, 4, 5, 6, 7) \\ (9, 1, 2, 3, 4, 5, 6, 7, 8) & \end{array}$$

Ejercicios de la sección 8.2

1. Derive las expresiones para a_i , c_i en las ecuaciones (8.2.11) del teorema 8.2.1.

2. Los seis puntos:

$$\begin{array}{lll} (0.0, 0.00000), & (0.2, 0.19867), & (0.4, 0.38942), \\ (0.6, 0.56464), & (0.8, 0.71736), & (1.0, 0.84147) \end{array}$$

están en el gráfico de $y = \sin(x)$, donde x está en radianes.

- a) Encuentra la porción del spline parabólico que interpola estos seis puntos para $0.4 \leq x \leq 0.6$. Mantén una precisión de cinco decimales en tus cálculos.
- b) Calcula $S(0.5)$ para el spline que encontraste en la parte (a). ¿Cuál es el error porcentual de $S(0.5)$ con respecto al valor “exacto” de $\sin(0.5) = 0.47943$?

3. Los siguientes cinco puntos

$$(0, 1), \quad (1, 7), \quad (2, 27), \quad (3, 79), \quad (4, 181)$$

están en una única curva cúbica.

- a) ¿Cuál de los tres tipos de splines cúbicos (natural, de salida parabólica o de salida cúbica) coincidiría exactamente con la curva cúbica que pasa por estos cinco puntos?
- b) Determina el spline cúbico que interpolará los cinco puntos en la parte (a), y verifica que es una curva cúbica única que interpola estos cinco puntos.
4. Repite los cálculos en el ejemplo 8.2.1 usando un spline natural para interpolar los cinco puntos de datos.
5. Repite los cálculos en el ejemplo 8.2.1 usando un spline de salida cúbica para interpolar los cinco puntos de datos.
6. Considera los cinco puntos

$$(0, 0), \quad (0.5, 1), \quad (1, 0), \quad (1.5, -1), \quad (2.0, 0)$$

en el gráfico de $y = \sin(\pi x)$.

- a) Usa un spline natural para interpolar los puntos $(0, 0)$, $(0.5, 1)$, $(1, 0)$.
- b) Usa un spline natural para interpolar los puntos $(0.5, 1)$, $(1, 0)$, $(1.5, -1)$.
- c) Explica el resultado inusual en la parte (b).
- 7. El spline periódico:** Si se sabe o se desea que los n puntos

$$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$$

a interpolar se encuentren en un solo ciclo de una curva periódica con período $x_n - x_1$, entonces una spline cúbica interpolante $S(x)$ debe satisfacer:

$$\begin{aligned} S(x_1) &= S(x_n) \\ S'(x_1) &= S'(x_n) \\ S''(x_1) &= S''(x_n) \end{aligned}$$

- a) Demuestra que estas tres condiciones de periodicidad requieren que:

$$\begin{aligned} y_1 &= y_n \\ M_1 &= M_n \\ 4M_1 + M_2 + M_{n-1} &= \frac{6(y_{n-1} - 2y_1 + y_2)}{h^2} \end{aligned}$$

- b) Usando las tres ecuaciones en la parte (a) y las ecuaciones (8.2.12), construye un sistema lineal de $(n-1) \times (n-1)$ para M_1, M_2, \dots, M_{n-1} en forma matricial.
- 8. El spline clampeado:** Supongamos que, además de los n puntos a interpolar, se nos dan valores específicos y'_1 y y'_n para las pendientes $S'(x_1)$ y $S'(x_n)$ del spline cúbico interpolante en los extremos x_1 y x_n .

a) Demuestra que:

$$2M_1 + M_2 = \frac{6(y_2 - y_1 - hy'_1)}{h^2}$$

$$2M_n + M_{n-1} = \frac{6(y_{n-1} - y_n + hy'_n)}{h^2}$$

b) Usando las ecuaciones en la parte (a) y las ecuaciones (8.2.12), construye un sistema lineal de $n \times n$ para M_1, M_2, \dots, M_n en forma matricial.

Los siguientes ejercicios están diseñados para resolverse utilizando una herramienta tecnológica. Típicamente, esto será MATLAB, MATHEMATICA o MAXIMA, pero también puede ser algún otro tipo de software de álgebra lineal.

9. En la solución del problema del spline cúbico natural, es necesario resolver un sistema de ecuaciones con la matriz de coeficientes

$$A_n = \begin{bmatrix} 4 & 1 & 0 & \cdots & 0 & 0 & 0 \\ 1 & 4 & 1 & \cdots & 0 & 0 & 0 \\ \vdots & & \ddots & \ddots & \ddots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 4 & 1 \\ 0 & 0 & 0 & \cdots & 0 & 1 & 4 \end{bmatrix}$$

Si podemos presentar una fórmula para esta matriz, entonces la solución para el problema del spline cúbico natural puede ser obtenida fácilmente. En este ejercicio y el siguiente, usaremos una computadora para descubrir esta fórmula. Para eso, primero determinamos una expresión para el determinante de A_n , denotado por D_n . Dado que

$$A_1 = [4] \quad \text{y} \quad A_2 = \begin{bmatrix} 4 & 1 \\ 1 & 4 \end{bmatrix}$$

vemos que

$$D_1 = \det(A_1) = 4, \quad D_2 = \det(A_2) = 15.$$

a) Usa la expansión de cofactores para demostrar que

$$D_n = 4D_{n-1} - D_{n-2}, \quad n = 3, 4, 5, \dots$$

b) Al escribir

$$D_n = 4D_{n-1} - D_{n-2}$$

y la identidad, $D_{n-1} = D_{n-1}$, en forma matricial,

$$\begin{bmatrix} D_n \\ D_{n-1} \end{bmatrix} = \begin{bmatrix} 4 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} D_{n-1} \\ D_{n-2} \end{bmatrix}$$

Demuestra que

$$\begin{bmatrix} D_n \\ D_{n-1} \end{bmatrix} = \begin{bmatrix} 4 & -1 \\ 1 & 0 \end{bmatrix}^{n-2} \begin{bmatrix} D_2 \\ D_1 \end{bmatrix} = \begin{bmatrix} 4 & -1 \\ 1 & 0 \end{bmatrix}^{n-2} \begin{bmatrix} 15 \\ 4 \end{bmatrix}$$

c) Usa diagonalización y una computadora para demostrar que la expresión para calcular

$$\begin{bmatrix} 4 & -1 \\ 1 & 0 \end{bmatrix}^{n-2}$$

es

$$\frac{1}{2\sqrt{3}} \begin{bmatrix} (2 + \sqrt{3})^{n-1} - (2 - \sqrt{3})^{n-1} & (2 - \sqrt{3})^{n-2} - (2 + \sqrt{3})^{n-2} \\ (2 + \sqrt{3})^{n-2} - (2 - \sqrt{3})^{n-2} & (2 - \sqrt{3})^{n-3} - (2 + \sqrt{3})^{n-3} \end{bmatrix}$$

y luego

$$D_n = \frac{(2 + \sqrt{3})^{n+1} - (2 - \sqrt{3})^{n+1}}{2\sqrt{3}}, \quad \text{para } n = 1, 2, 3, \dots$$

d) Usando una computadora, verifica este resultado para $1 \leq n \leq 10$.

10. En este ejercicio, determinamos una fórmula para calcular A_n^{-1} a partir de D_k para $k = 0, 1, 2, 3, n$, asumiendo que D_0 se define como 1.

a) Usa una computadora para calcular A_k^{-1} para $k = 1, 2, 3, 4, 5$.

b) De tus resultados en el apartado (a), demuestra la conjetura de que

$$A_n^{-1} = \llbracket \alpha_{ij} \rrbracket,$$

donde

$$\alpha_{ij} = (-1)^{i+j} \left(\frac{D_{n-j} D_{i-1}}{D_n} \right),$$

para $i \leq j$.

c) Usa el resultado en la parte (b) para calcular A_7^{-1} y compáralo con el resultado obtenido con la computadora.

Ejercicios de la sección 8.3

1. Considera la matriz de transición

$$P = \begin{bmatrix} 0.4 & 0.5 \\ 0.6 & 0.5 \end{bmatrix}$$

a) Calcula $\mathbf{x}^{(n)}$ para $n = 1, 2, 3, 4, 5$ si $\mathbf{x}^{(0)} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$.

b) Explica por qué P es regular y encuentra su vector de estado estacionario.

2. Considera la matriz de transición

$$P = \begin{bmatrix} 0.2 & 0.1 & 0.7 \\ 0.6 & 0.4 & 0.2 \\ 0.2 & 0.5 & 0.1 \end{bmatrix}$$

a) Calcula $\mathbf{x}^{(1)}$, $\mathbf{x}^{(2)}$ y $\mathbf{x}^{(3)}$ con tres decimales si

$$\mathbf{x}^{(0)} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

3. Encuentra los vectores de estado estacionario de las siguientes matrices de transición regulares:

a) $\begin{bmatrix} \frac{1}{3} & \frac{2}{3} \\ \frac{3}{4} & \frac{1}{4} \end{bmatrix}$

b) $\begin{bmatrix} 0.81 & 0.19 \\ 0.26 & 0.74 \end{bmatrix}$

c) $\begin{bmatrix} \frac{1}{3} & \frac{1}{3} & 0 \\ \frac{1}{3} & 0 & \frac{2}{3} \\ \frac{1}{3} & \frac{2}{3} & \frac{1}{4} \end{bmatrix}$

4. Sea P la matriz de transición

$$P = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ 0 & 1 \end{bmatrix}$$

- a) Demuestra que P no es regular.
- b) Demuestra que, a medida que n aumenta, $P^n \mathbf{x}^{(0)}$ se aproxima a $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ para cualquier vector de estado inicial $\mathbf{x}^{(0)}$.
- c) ¿Qué conclusión del teorema 8.3.3 no es válida para el estado estacionario de esta matriz de transición?
5. Verifica que si P es una matriz de transición regular $k \times k$ cuyos sumas de filas son todas iguales a 1, entonces las entradas de su vector de estado estacionario son todas iguales a $1/k$.

6. Demuestra que la matriz de transición

$$P = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix}$$

es regular, y usa el ejercicio 8.3.5 para encontrar su vector de estado estacionario.

7. Juan está feliz o triste. Si está feliz un día, entonces estará feliz al día siguiente cuatro veces de cada cinco. Si está triste un día, entonces estará triste al día siguiente una vez de cada tres. A largo plazo, ¿cuál es la probabilidad de que Juan esté feliz en un día cualquiera?
8. Un país está dividido en tres regiones demográficas. Se sabe que cada año el 5 % de los residentes de la región 1 se trasladan a la región 2 y el 5 % se trasladan a la región 3. De los residentes de la región 2, el 15 % se trasladan a la región 1 y el 10 % se trasladan a la región 3. De los residentes de la región 3, el 10 % se trasladan a la región 1 y el 5 % se trasladan a la región 2. ¿Qué porcentaje de la población reside en cada una de las tres regiones después de un largo período de tiempo?

Los siguientes ejercicios están diseñados para resolverse utilizando una herramienta tecnológica. Típicamente, esto será MATLAB, MATHEMATICA o MAXIMA, pero también puede ser algún otro tipo de software de álgebra lineal.

9. Considera la secuencia de matrices de transición $\{P_2, P_3, P_4, \dots\}$.

$$P_2 = \begin{bmatrix} 0 & \frac{1}{2} \\ 1 & \frac{1}{2} \end{bmatrix},$$

$$P_3 = \begin{bmatrix} 0 & 0 & \frac{1}{3} \\ 0 & \frac{1}{2} & \frac{1}{3} \\ 1 & \frac{1}{2} & \frac{1}{3} \end{bmatrix}$$

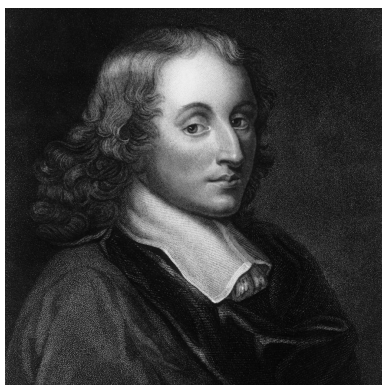
$$P_4 = \begin{bmatrix} 0 & 0 & 0 & \frac{1}{4} \\ 0 & 0 & \frac{1}{3} & \frac{1}{4} \\ 0 & \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ 1 & \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \end{bmatrix}$$

$$P_5 = \begin{bmatrix} 0 & 0 & 0 & 0 & \frac{1}{5} \\ 0 & 0 & 0 & \frac{1}{4} & \frac{1}{5} \\ 0 & 0 & \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \\ 0 & \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \\ 1 & \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \end{bmatrix}$$

A

Apéndice

PRINCIPIOS DE INDUCCIÓN Y DE BUEN ORDEN



Blaise Pascal: Nacido el 19 de junio de 1623 en Clermont-Ferrand, Francia; fue un genio precoz y de clara inteligencia, pues su entusiasmo juvenil por la ciencia se materializó en importantes y precursoras aportaciones a la física y a las matemáticas. Siendo aún niño, con solo 12 años, demostró que la suma de los ángulos de un triángulo es siempre igual a 180° . Pese a su frágil salud y corta vida, murió a los 39 años, pero su huella quedó grabada en la historia de la física y de la informática.

Uno de los métodos más usados para realizar demostraciones es el Método de Inducción Matemática. Este fue creado por Blaise Pascal en el siglo XVII, aunque el primer matemático que ofreció una demostración formal mediante el uso explícito de la inducción matemática fue el italiano Franciscus Maurolicus (1494-1575). Maurolicus utilizó la inducción para demostrar que para todo entero positivo n

$$1 + 3 + 5 + \dots + (2n - 1) = n^2.$$

Dicho método ha servido para demostrar teoremas en distintas áreas de la matemática; como: geometría, teoría de grafos, teoría de números, análisis combinatorio. Una manera informal (pero eficaz) de ver y explicar el Principio de Inducción Matemática es mediante fichas de dominó (véase la figura A.1). Imaginemos que tenemos fichas de dominó puestas en una hilera infinita. Si empujamos la primera ficha, esta empujará la segunda ficha; la segunda ficha empujará la tercera ficha; la tercera ficha empujará la cuarta ficha; y así sucesivamente hasta que caigan todas las fichas. En este caso, cada ficha representa un número natural.

Como preliminar, debemos saber que cualquier proposición se puede clasificar como general o particular. Algunos ejemplos de proposiciones generales son:

1. Todos los ciudadanos de México tienen derecho a la educación.
2. Todos los números que terminan en cero son divisibles entre 5.

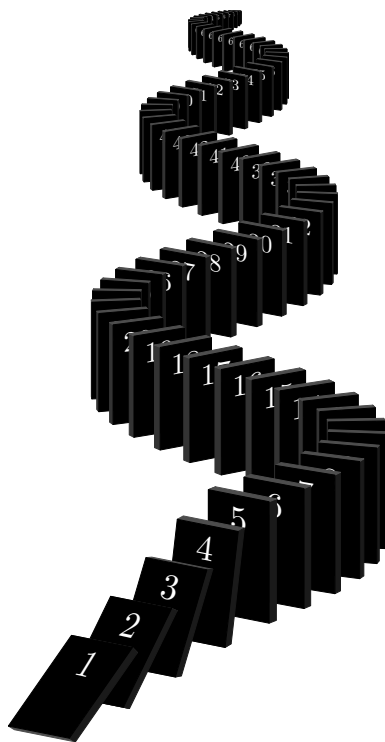


Figura A.1 Representación del Principio de Inducción

Algunos ejemplos de proposiciones particulares son:

1. Guinevere tiene derecho a la educación
2. 300 es divisible entre 5.

El proceso de obtener una proposición particular de una general se llama **deducción**. Por ejemplo, si tenemos

1. Todos los ciudadanos de México tienen derecho a la educación.
2. Guinevere es mexicana.
3. Guinevere tiene derecho a la educación.

entonces de la proposición general (1), junto con la proposición particular (2), se obtiene la proposición particular (3).

El proceso de obtener proposiciones generales de proposiciones particulares se llama **inducción**. El razonamiento inductivo puede conducir a conclusiones falsas así como a verdaderas. Por ejemplo, si tenemos

1. 300 es divisible entre 5.
2. Todos los números que terminan en cero son divisibles entre 5.

entonces la proposición general (2), obtenida de la proposición particular (1), es verdadera. Pero, si consideramos

1. 300 es divisible entre 5.
2. Todos los números con tres cifras son divisibles entre 5.

entonces la proposición general (2), deducida de la proposición particular (1), es falsa.

En resumen, *La inducción* (o sea, la sugerencia de una idea o una hipótesis) sin duda desempeña en las Matemáticas un papel importante, pero puramente heurístico: permite adivinar cuál debe ser, según todas las apariencias, la solución. Pero las proposiciones matemáticas se demuestran siempre deductivamente. Ningún resultado matemático puede considerarse justo, válido, si no ha sido deducido de las proposiciones de partida.

Problema: Calcular la suma de los primeros n números impares.

Solución: Los problemas de este tipo se pueden resolver usando una fórmula probada, pero nos interesa resolver el problema sin recurrir a tal fórmula y aplicando el método de la inducción matemática. Para hacerlo, es necesario establecer primero una hipótesis, es decir, tratar de adivinar la solución basándonos en patrones observados. Esto se puede lograr al analizar las sumas parciales de los primeros números impares. Observemos los cálculos de las primeras sumas parciales:

$$\begin{aligned}
 1 &= 1 \\
 1 + 3 &= 4 \\
 1 + 3 + 5 &= 9 \\
 1 + 3 + 5 + 7 &= 16 \\
 1 + 3 + 5 + 7 + 9 &= 25 \\
 1 + 3 + 5 + 7 + 9 + 11 &= 36 \\
 1 + 3 + 5 + 7 + 9 + 11 + 13 &= 49 \\
 1 + 3 + 5 + 7 + 9 + 11 + 13 + 15 &= 64 \\
 1 + 3 + 5 + 7 + 9 + 11 + 13 + 15 + 17 &= 81
 \end{aligned}$$

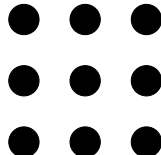
Para $n = 1$:



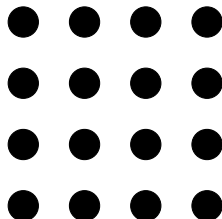
Para $n = 2$:



Para $n = 3$:



Para $n = 4$:



Para $n = 5$:

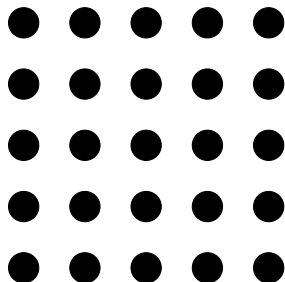


Figura A.2 Representación del modelo

$$1 + 3 + 5 + \dots + (2n - 1) = n^2$$

cuando $n = 1, 2, 3, 4, 5$

Cada suma puede representarse en la forma

$$\alpha + \beta + \gamma + \dots + u = S,$$

donde u representa el último sumando y S la suma. Tanto u como S dependen del número de sumandos, al cual denominaremos con n . Estas dependencias se pueden estudiar para identificar patrones o relaciones matemáticas entre las variables involucradas. Dadas estas convenciones, podemos construir dos tablas. La primera tabla nos permitirá explorar la relación entre n , el número de sumandos, y u , el último término de la suma. La segunda tabla nos ayudará a identificar una posible relación entre n y S , la suma total. Estas tablas nos servirán para formular hipótesis iniciales sobre las dependencias funcionales.

n	u
1	1
2	3
3	5
4	7
5	9
6	11
7	13
8	15
9	17
10	19

n	S
1	1
2	4
3	9
4	16
5	25
6	36
7	49
8	64
9	81
10	100

Por lo que se propone el siguiente modelo:

$$1 + 3 + 5 + 7 + 9 + 11 + \dots + (2n - 1) = n^2.$$

Probemosla por medio de inducción sobre n . Llamemos a la suma, S_n . Es decir,

$$S_n = 1 + 3 + 5 + 7 + 9 + 11 + \dots + (2n - 1).$$

i. Para $n = 1$ es evidente que se cumple, ya que la suma consiste en un solo término, el 1. El valor de la expresión n^2 también es 1.

ii. Supóngase que la hipótesis se cumple para $n = k$, es decir, $S_k = k^2$.

iii. A partir de (ii), probemos que se cumple para $n = k + 1$, es decir,

$$S_{k+1} = (k + 1)^2.$$

En efecto, se tiene

$$S_{k+1} = S_k + (2k + 1)$$

pero $S_k = k^2$, se sigue que

$$S_{k+1} = k^2 + (2k + 1) = (k + 1)^2.$$

En conclusión, se cumple que $S_n = n^2$. Incluso, para este caso, podemos ver la *demonstración visual* de

$$1 + 3 + 5 + 7 + 9 + 11 + \dots + (2n - 1) = n^2$$

como se muestra en la figura A.2. Aunque no se debe olvidar que esto es una manera visual de ver el comportamiento de una proposición particular. No siempre se puede recurrir a una demostración visual, ya que nos puede llevar a conclusiones falsas. Esto se debe a que las demostraciones visuales suelen depender de la interpretación que hagamos de las imágenes, diagramas o esquemas, lo cual puede resultar engañoso en ciertos contextos. Por ello, la matemática se apoya siempre en la demostración formal y deductiva para garantizar la validez de cualquier proposición. En el caso del ejemplo anterior, aunque la representación visual sea una herramienta poderosa para comprender la regularidad de la suma de los números impares consecutivos, es fundamental corroborar dicho resultado utilizando el Método de Inducción Matemática.

i Nota

Una demostración visual implica que la solución de un problema sea accesible de manera clara a través de un diagrama o dibujo, con un mínimo de desarrollo matemático necesario.

A.1 Inducción errónea en las matemáticas

Ejemplo A.1.1:

Consideremos el polinomio $f(x) = x^2 + x + 41$. Si este polinomio se reemplaza x por el número 0, se obtiene el número primo 41. Si se reemplaza x por el número 1, nuevamente se obtiene un número primo, 43. Si se sigue este procedimiento para 1, 2, 3, 4, 5, 6; obtenemos que

$$f(0) = 41$$

$$f(1) = 43$$

$$f(2) = 47$$

$$f(3) = 53$$

$$f(4) = 61$$

$$f(5) = 71$$

$$f(6) = 83$$

Basándonos en estos resultados, podría parecer que para todo entero no negativo x , el polinomio produce números primos. Sin embargo, esto no es cierto, ya que aunque $x^2 + x + 41$ genera números primos para $f(1), f(2), f(3), \dots, f(38), f(39)$, al evaluar $f(40)$ se obtiene 1681, que es un número compuesto.

Ejemplo A.1.2:

El binomio $x^n - 1$, $n \in \mathbb{N}$, fue estudiado por numerosos matemáticos y factorizado con coeficientes enteros. Al analizar estas factorizaciones para diversos valores de n , los matemáticos observaron que los valores absolutos de los coeficientes de los factores nunca superaron el número 1. Es decir,

$$x - 1 = x - 1$$

$$x^2 - 1 = (x - 1)(x + 1)$$

$$x^3 - 1 = (x - 1)(x^2 + x + 1)$$

$$x^4 - 1 = (x - 1)(x + 1)(x^2 + 1)$$

$$\vdots$$

Se construyeron tablas de coeficientes, y en cada caso los coeficientes cumplían la propiedad mencionada. Sin embargo, no se logró probar la validez de la proposición para todos los valores de n hasta que V. Ivanov resolvió el problema en 1941. Si $n < 105$, el binomio $x^n - 1$ cumple la propiedad, pero uno de los factores de $x^{105} - 1$ es el polinomio.

$$\begin{aligned} & x^{48} + x^{47} + x^{46} - x^{43} - x^{42} - 2x^{41} - x^{40} - x^{39} + x^{36} + x^{35} + x^{34} + x^{33} \\ & + x^{32} + x^{31} - x^{28} - x^{26} - x^{24} - x^{22} - x^{20} + x^{17} + x^{16} + x^{15} + x^{14} \\ & + x^{13} + x^{12} - x^9 - x^8 - 2x^7 - x^6 - x^5 + x^2 + x + 1 \end{aligned}$$

que ya no tiene esta propiedad.

Ejemplo A.1.3:

Gottfried W. Leibniz, destacado matemático alemán del siglo XVII, demostró que, para cualquier entero positivo n , el número $n^3 - n$ es divisible por 3, $n^5 - n$ es divisible por 5 y $n^7 - n$ es divisible por 7. Concluyó que para todo k impar y cualquier n natural, $n^k - n$ sería divisible por k . Sin embargo, pronto descubrió que $2^9 - 2 = 510$ no es divisible por 9.

Ejemplo A.1.4:

Un error del mismo carácter cometió D. Aleksandrovich Grave, conocido matemático soviético, al suponer que para todo primo p el número $2^{p-1} - 1$ no es divisible por p^2 . El cálculo directo confirmaba esta hipótesis para todos los números p menores que mil. Sin embargo, pronto se comprobó que $2^{1092} - 1$ es divisible por 1093^2 (1093 es un número primo); o sea, la hipótesis de Grave resultó errónea.

Ejemplo A.1.5: Consideremos los números del tipo

$$2^{2^n} + 1.$$

Para $n = 0, 1, 2, 3, 4$, los números 3, 5, 17, 257, 65 537 son primos. Pierre de Fermat, matemático francés del siglo XVII, afirmó que todos estos números eran primos. Sin embargo, L. Euler descubrió en el siglo XVIII que

$$2^{2^5} + 1 = 4\,294\,967\,297 = (641)(6\,700\,417)$$

es un número compuesto.

A.2 El principio de la inducción matemática

Principio de Inducción:

Sea P una propiedad cualquiera. Tenemos que $P(1), P(2), P(3), P(4), \dots$ es un conjunto de propiedades para cada número natural tal que:

- i. $P(1)$ es cierta.
- ii. Si $P(n)$ es cierta, entonces $P(n+1)$ es cierta.

Entonces $P(n)$ es cierta para toda $n \in \mathbb{N}$.

Ejemplo A.2.1: Demostrar que

$$1 + 2 + \dots + n = \frac{n(n+1)}{2}, \quad \forall n \in \mathbb{N}.$$

Demostración: Claramente $n = 1$, satisface la fórmula, ya que

$$1 = \frac{1(1+1)}{2}$$

$$1 = 1$$

lo cual es verdadero. Supongamos que la fórmula se cumple para $n = k$, es decir, supongamos que

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

Demostraremos a partir de lo dicho anteriormente que la fórmula se cumple para $n = k+1$. Es decir, probaremos que

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

En efecto, por hipótesis de inducción

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

Entonces

$$\begin{aligned} 1 + 2 + \dots + k + k + 1 &= \frac{k(k+1)}{2} + k + 1 \\ &= \frac{k(k+1) + 2(k+1)}{2} \\ &= \frac{(k+1)(k+2)}{2} \end{aligned}$$

es decir

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

Por tanto,

$$1 + 2 + \dots + n = \frac{n(n+1)}{2}, \quad \forall n \in \mathbb{N}.$$

Ejemplo A.2.2: Demostrar que

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

Demostración: Claramente $n = 1$ satisface la fórmula, ya que al sustituir n por 1 se tiene

$$\begin{aligned} 1^3 &= (1)^2 \\ 1 &= 1 \end{aligned}$$

lo cual es verdadero. Notemos por el ejercicio anterior que

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

entonces tenemos que

$$1^3 + 2^3 + \cdots + n^3 = \left(\frac{n(n+1)}{2} \right)^2.$$

Así pues, supongamos que la fórmula cumple para $n = k$, es decir

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

Demostraremos a partir de lo dicho anteriormente que la fórmula se cumple para $n = k + 1$. Es decir, probaremos que

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

En efecto: por hipótesis de inducción

$$\begin{aligned} 1^3 + 2^3 + \cdots + k^3 + (k+1)^3 &= \left(\frac{k(k+1)}{2} \right)^2 + (k+1)^3 \\ &= \frac{k^2(k+1)^2}{4} + (k+1)^3 \\ &= \frac{k^2(k+1)^2 + 4(k+1)^3}{4} \\ &= \frac{(k+1)^2(k^2 + 4(k+1))}{4} \\ &= \frac{(k+1)^2(k^2 + 4k + 4)}{4} \\ &= \frac{(k+1)^2(k+2)^2}{4} \\ &= \left(\frac{(k+1)(k+2)}{2} \right)^2 \end{aligned}$$

Por tanto,

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

i Nota

Varios ejemplos de dar un contraejemplo, fueron mostrados en la sección anterior, donde se observó cómo un contraejemplo puede invalidar una afirmación aparentemente general.

El método de la inducción matemática no es el único procedimiento para demostrar que una propiedad P se cumple para cada número natural. Existen otros métodos que también permiten establecer la validez de propiedades matemáticas, aunque la inducción es particularmente útil por su estructura sistemática y lógica. No obstante, el hecho de no poder demostrar que P se cumple para todos los números naturales no implica que P sea falsa, sino únicamente que, hasta el momento, no se ha encontrado una demostración válida. Demostrar que una propiedad P es falsa requiere una estrategia diferente, y basta con encontrar un *contraejemplo*, es decir, un caso específico donde P no se cumpla. Este procedimiento es directo y concluyente, pues un solo contraejemplo es suficiente para refutar la validez universal de una propiedad.

Ejemplo A.2.3: Demostrar que

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

Demostración: Claramente $n = 1$ satisface la fórmula, ya que

$$\begin{aligned} 1^2 &= \frac{1(1+1)(2 \cdot 1 + 1)}{6} \\ 1 &= 1 \end{aligned}$$

Supongamos que la fórmula se cumple para $n = k$, es decir, supongamos que

$$1^2 + 2^2 + \dots + k^2 = \frac{k(k+1)(2k+1)}{6}$$

Demostraremos a partir de lo dicho anteriormente que la fórmula se cumple para $n = k + 1$. Es decir, probaremos que

$$1^2 + 2^2 + \dots + (k+1)^2 = \frac{(k+1)(k+2)(2k+3)}{6}.$$

En efecto: por hipótesis de inducción

$$1^2 + 2^2 + \dots + k^2 = \frac{k(k+1)(2k+1)}{6}.$$

Entonces

$$\begin{aligned} 1^2 + 2^2 + \dots + k^2 + (k+1)^2 &= \frac{k(k+1)(2k+1)}{6} + (k+1)^2 \\ &= \frac{k(k+1)(2k+1) + 6(k+1)^2}{6} \\ &= \frac{(k+1)(k(2k+1) + 6(k+1))}{6} \\ &= \frac{(k+1)(2k^2 + 7k + 6)}{6} \\ &= \frac{(k+1)(k+2)(2k+3)}{6} \end{aligned}$$

Por tanto,

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

En resumen, para demostrar que una propiedad P se cumple para cada número natural, se emplea un procedimiento llamado *demostración por inducción matemática*. Es decir, si \mathbb{N} es el conjunto de los números naturales y $A \subset \mathbb{N}$ tal que:

- i. $1 \in A$.
- ii. $k \in A \implies k + 1 \in A$. Entonces $A = \mathbb{N}$.

Sin embargo, en ciertas demostraciones por inducción, resulta útil emplear el siguiente principio:

Inducción Modificado:

Si A es un subconjunto de los números naturales tal que

- i. $1 \in A$.
- ii. Si $1, 2, \dots, n \in A$, implica que $n + 1 \in A$.

Entonces $A = \mathbb{N}$.

Sea $P(n)$ una propiedad cualquiera. Si se quiere demostrar que $P(n)$ es cierta para toda $n \in \mathbb{N}$, es decir, $A := \{n \in \mathbb{N} \mid P(n) \text{ es cierta}\} = \mathbb{N}$ es suficiente mostrar que A satisface las dos hipótesis del Principio de Inducción Modificado.

Podría suceder que $P(n)$ no sea verdadera para todos los números naturales, pero existe cierto número natural i tal que $P(k)$ es verdadera para toda $i \leq k$. Se puede utilizar también el Principio de Inducción Modificado considerando al conjunto

$$A = \{n \in \mathbb{N} \mid P(n+i) \text{ es cierta}\}$$

en cuyo caso se reduce a demostrar que:

- i. $P(i)$ es verdadera.
- ii. Si $P(k)$ es cierta para $k = i, i+1, \dots, i+n$, entonces $P(i+(n+1))$ es cierta.

Proposición A.2.1: El Principio de Inducción y el Principio de Inducción Modificado son equivalentes.

Demostración: Se deja como ejercicio al lector.

A.3 El principio de buen orden

Principio de Buen Orden:

Si A es un subconjunto no vacío de los números naturales, entonces existe $a \in A$ tal que $a \leq x, \forall x \in A$, es decir, A tiene elemento mínimo, que es a . Al elemento mínimo de A , también llamado primer elemento de A , se le denota por $\min A$, es decir, $a = \min A$.

Ejemplo A.3.1:

Consideremos el conjunto $A = \{5, 10, 15, 20, 25, 30\}$, un subconjunto no vacío de los números naturales. Por el Principio de Buen Orden, este conjunto tiene un elemento mínimo, que en este caso es 5, ya que es el menor de todos sus elementos. Por lo tanto,

$$a = \min A = 5.$$

Ejemplo A.3.2:

Todo número natural puede ser representado como una suma de unos.

Demostración: Queremos demostrar que para cualquier número natural n ,

$$n = \underbrace{1 + 1 + \dots + 1}_{n \text{ veces}}$$

Consideremos el conjunto A de todos los números naturales que no pueden ser escritos como una suma de unos. Es decir:

$$A = \{n \in \mathbb{N} \mid n \text{ no puede ser escrito como una suma de unos}\}$$

Según el Principio de Buen Orden, si A no es vacío, entonces debe tener un elemento mínimo. Sea m el elemento mínimo de A . Por definición, m no puede ser escrito como una suma de unos. Ahora, consideremos $m-1$. Si $m-1=0$, entonces $m=1$. Sin embargo, 1 puede ser escrito como una suma de un único 1, es decir, $1=1$. Esto es una contradicción porque asumimos que m no puede ser escrito como una suma de unos. Si $m-1$ es un número natural positivo, entonces $m-1$ debe ser escribible como una suma de unos, ya que m es el mínimo elemento de A y $m-1 < m$. Supongamos entonces que $m-1$ puede ser escrito como

$$m-1 = 1 + 1 + \dots + 1 \quad \text{con } m-1 \text{ unos}$$

Al agregar un 1 más a esta suma, obtenemos

$$m = (1 + 1 + \dots + 1) + 1 \quad \text{con } m \text{ unos}$$

Esto demuestra que m puede ser escrito como una suma de unos, lo cual es una contradicción porque m se suponía que no podía ser escrito como una suma de unos. Dado que suponer que A no es vacío nos lleva a una contradicción, debemos concluir que A es vacío. Esto significa que no existe ningún número natural que no pueda ser escrito como una suma de unos. Por lo tanto, hemos demostrado que cualquier número natural n puede ser representado como una suma de unos.

Ejemplo A.3.3: Todo número natural mayor que uno puede ser descompuesto en un producto de números primos.

Demostración: Supongamos que existe un conjunto B de números naturales mayores que 1 que no pueden ser descompuestos de esta forma. Si $B \neq \emptyset$, sea n el elemento mínimo de B . Si n es primo, ya es un producto de un solo número primo, contradiciendo que $n \in B$. Por lo tanto, n debe ser compuesto, es decir, $n = a \cdot b$ con $1 < a, b < n$. Como $a, b < n$ y n es el mínimo de B , a y b no están en B y pueden ser descompuestos en productos de números primos. Así,

$$n = a \cdot b = (p_1 \cdot p_2 \cdots p_k) \cdot (q_1 \cdot q_2 \cdots q_m)$$

donde $a = p_1 \cdot p_2 \cdots p_k$ y $b = q_1 \cdot q_2 \cdots q_m$ son productos de números primos. Esto contradice la suposición de que n no puede ser descompuesto en un producto de números primos, por lo que $B = \emptyset$. Es decir, no existe ningún número natural mayor que 1 que no pueda ser descompuesto en un producto de números primos. Por lo tanto, hemos demostrado que todo número natural $n > 1$ puede ser descompuesto en un producto de números primos, utilizando el Principio del Buen Orden.

Teorema A.3.1: El Principio de Buen Orden es equivalente al Principio de Inducción

Demostración:

\Rightarrow) El Principio de Buen Orden implica el Principio de Inducción: Supongamos que existe $M \subset \mathbb{N}$ tal que

i) $1 \in M$,

ii) $k \in M \implies k + 1 \in M$.

Supongamos que $M \neq \mathbb{N}$. Sea $M^C = \mathbb{N} - M$ el complemento de M . Como $M \neq \mathbb{N}$, entonces $M^C \neq \emptyset$, por lo tanto existe $a \in M^C$ tal que $a \leq x$, $\forall x \in M^C$. Como $1 \in M$, entonces $1 \notin M^C$, así que $a \neq 1$, por lo tanto existe $k \in \mathbb{N}$ tal que $k + 1 = a$, entonces $k \notin M^C$ y por tanto $k \in M$, de donde se sigue, aplicando (ii), que $k + 1 = a \in M$, lo cual es una contradicción, pues $a \in M^C$ y $M \cap M^C = \emptyset$. Por tanto, si $M \subset \mathbb{N}$ y se cumplen (i) y (ii), se debe cumplir que $M = \mathbb{N}$.

\Leftarrow) El Principio de Inducción implica el Principio de Buen Orden: Consideremos $A \subset \mathbb{N}$ con $A \neq \emptyset$. Supongamos que A no tiene elemento mínimo, es decir, supongamos que $\forall k \in A$, existe $x \in A$ tal que $x < k$. Consideremos a $B = \{y \in \mathbb{N} \mid y < x, \forall x \in A\}$. Ya que $x \not\leq x$, $\forall x \in A$, entonces tenemos que $B \cap A = \emptyset$. Demostremos ahora, aplicando el Principio de Inducción Matemática, que $B = \mathbb{N}$. Por como se define, es claro que $B \subset \mathbb{N}$.

i) Afirmamos que $1 \in B$. Esto ya que $1 \leq n$, $\forall n \in \mathbb{N}$, entonces $1 \leq x$, $\forall x \in A$. Además $1 \notin A$, ya que si $1 \in A$, entonces $1 = \min A$, lo que contradice la suposición de que A no tiene elemento mínimo. Por tanto, $1 < x$, $\forall x \in A$, y en consecuencia $1 \in B$.

ii) Supongamos que $n \in B$, es decir, supongamos que $n < x$, $\forall x \in A$.

iii) Probaremos a partir de (ii) que $n + 1 \in B$. Si $n \in B$ entonces $n < x$, $\forall x \in A$, por tanto $n + 1 \leq x$, $\forall x \in A$, y por tanto $n + 1 \in B$.

Por (i), (ii), (iii), $B = \mathbb{N}$, y como $A \subset \mathbb{N} = B$ y $B \cap A = \emptyset$, entonces $A = \emptyset$, lo cual contradice que $A \neq \emptyset$. En consecuencia, si $A \subset \mathbb{N}$ y $A \neq \emptyset$, entonces A debe tener elemento mínimo.

Al utilizar el Principio de Buen Orden, se pueden abordar problemas complejos al enfocarse en el elemento más pequeño de un conjunto, lo cual a menudo conduce a contradicciones que invalidan suposiciones erróneas. Este enfoque simplifica el análisis lógico al reducir un problema abstracto a un caso específico y manejable, facilitando la identificación de inconsistencias en las suposiciones iniciales.

A.4 Más ejemplos de inducción

Ejemplo A.4.1: Demostremos que para toda $n \in \mathbb{N}$,

$$\int_0^{\infty} e^{-x} x^n dx = n!$$

Pero antes, recordemos la siguiente propiedad:

$$\lim_{x \rightarrow \infty} e^{px} = 0, \text{ si } p < 0.$$

Así pues, si $n = 0$, entonces

$$\begin{aligned} \int_0^{\infty} e^{-x} dx &= -e^{-x} \Big|_0^{\infty} \\ &= - \lim_{x \rightarrow \infty} e^{-x} + e^{-0} \\ &= 0! \end{aligned}$$

Si $n = 1$, se procede por integración por partes. Tomando

$$u = x \quad \text{y} \quad dv = e^{-x} dx$$

obtenemos

$$du = dx \quad \text{y} \quad v = -e^{-x}.$$

Entonces

$$\begin{aligned} \int_0^{\infty} e^{-x} x dx &= -xe^{-x} \Big|_0^{\infty} + \int_0^{\infty} e^{-x} dx \\ &= - \lim_{x \rightarrow \infty} xe^{-x} + (0)e^{-0} + 1 \\ &= 1! \end{aligned}$$

Supongamos que se cumple para $n = k$, es decir,

$$\int_0^{\infty} e^{-x} x^k dx = k!.$$

Demostraremos a partir lo dicho anteriormente, que la propiedad se cumple para $n = k + 1$, es decir,

$$\int_0^{\infty} e^{-x} x^{k+1} dx = (k+1)!.$$

Procederemos por integración por partes. Así pues, tomando a

$$u = x^{k+1} \quad \text{y} \quad dv = e^{-x} dx$$

obtenemos

$$du = (k+1)x^k dx \quad \text{y} \quad v = -e^{-x}.$$

Entonces

$$\begin{aligned} \int_0^{\infty} e^{-x} x^{k+1} dx &= -x^{k+1} e^{-x} \Big|_0^{\infty} + (k+1) \int_0^{\infty} e^{-x} x^k dx \\ &= - \lim_{x \rightarrow \infty} x^{k+1} e^{-x} + (0)^{k+1} e^{-0} + (k+1)k! \\ &= - \lim_{x \rightarrow \infty} \frac{(k+1)!}{e^x} + (k+1)k! \\ &= (k+1)! \end{aligned}$$

Por tanto,

$$\int_0^{\infty} e^{-x} x^n dx = n!, \quad \forall n \in \mathbb{N}.$$

Ejemplo A.4.2: Si a_1, a_2, \dots, a_n son números reales positivos, demuestre que

$$\sqrt[n]{a_1 \cdot a_2 \cdots a_n} \leq \frac{a_1 + a_2 + \cdots + a_n}{n}. \quad (\text{A.4.1})$$

Demostración:

i. Para $n = 2$, la desigualdad (A.4.1) da

$$\sqrt{a_1 a_2} \leq \frac{a_1 + a_2}{2}. \quad (\text{A.4.2})$$

Es fácil obtener esta desigualdad a partir de esta otra

$$(\sqrt{a_1} - \sqrt{a_2})^2 \geq 0,$$

válida para cualesquiera a_1, a_2 números reales positivos. La desigualdad (A.4.2) tiene una interpretación geométrica: en una recta AB , tomemos los segmentos a_1 y a_2 , y construyamos una circunferencia con diámetro $a_1 + a_2$. Aquí, $\frac{a_1 + a_2}{2}$ es el radio de la circunferencia, y $\sqrt{a_1 a_2}$ es la mitad de la cuerda perpendicular al diámetro en el punto donde se unen a_1 y a_2 (véase la figura A.3). Esto demuestra (A.4.2).

ii. Supongamos que la desigualdad (A.4.1) se cumple para $n = k$. Demostremos que, entonces, también se cumple para $n = 2k$. En efecto,

$$\begin{aligned} \sqrt[2k]{a_1 \cdot a_2 \cdots a_{2k}} &= \sqrt{\sqrt[k]{a_1 \cdot a_2 \cdots a_k} \sqrt[k]{a_{k+1} \cdots a_{2k}}} \\ &\leq \frac{\sqrt[k]{a_1 \cdot a_2 \cdots a_k} + \sqrt[k]{a_{k+1} \cdots a_{2k}}}{2} \\ &\leq \frac{\frac{a_1 + a_2 + \cdots + a_k}{k} + \frac{a_{k+1} + \cdots + a_{2k}}{k}}{2} \\ &= \frac{a_1 + a_2 + \cdots + a_k + \cdots + a_{2k}}{2k} \end{aligned}$$

Como ya demostramos la desigualdad (A.4.1) para $n = 2$, se cumple también para $n = 4, 8, 16, \dots$, es decir, para $n = 2^s$ con $s \in \mathbb{N}$.

iii. Para demostrar que la desigualdad (A.4.1) se cumple para cualquier número natural n , probemos que si se cumple para $n = k$, también se cumple para $n = k - 1$. Sean pues, $a_1, a_2, \dots, a_{k-1} \in \mathbb{R}^+$. Sea $\lambda \in \mathbb{R}^+$, entonces

$$\sqrt[k]{a_1 \cdot a_2 \cdots a_{k-1} \lambda} \leq \frac{a_1 + a_2 + \cdots + a_{k-1} + \lambda}{k}.$$

Determinemos λ de modo que

$$\frac{a_1 + a_2 + \cdots + a_{k-1} + \lambda}{k} = \frac{a_1 + a_2 + \cdots + a_{k-1}}{k-1},$$

o sea, pongamos

$$\lambda = \frac{a_1 + a_2 + \cdots + a_{k-1}}{k-1}.$$

Tenemos

$$\sqrt[k]{a_1 \cdot a_2 \cdots a_{k-1} (a_1 + a_2 + \cdots + a_{k-1})} \leq \frac{a_1 + a_2 + \cdots + a_{k-1}}{k-1},$$

es decir,

$$\sqrt[k-1]{a_1 \cdot a_2 \cdots a_{k-1}} \leq \frac{a_1 + a_2 + \cdots + a_{k-1}}{k-1}.$$

Sea ahora $m \in \mathbb{N}$. Si $m = 2^s$, la desigualdad se cumple por (ii). Si $m \neq 2^s$, elegimos s tal que $m < 2^s$; entonces, por (ii) y (iii), la desigualdad también se cumple para $n = m$.

Por tanto, se cumple que

$$\sqrt[n]{a_1 \cdot a_2 \cdots a_n} \leq \frac{a_1 + a_2 + \cdots + a_n}{n}, \quad \forall a_1, a_2, \dots, a_n \in \mathbb{R}^+, \quad \forall n \in \mathbb{N}.$$

Nota

La expresión (A.4.1) se le conoce como *desigualdad entre la media geométrica y la media aritmética*. Obsérvese que si

$$a_1 = a_2 = \cdots = a_n$$

entonces

$$\sqrt[n]{a_1 \cdot a_2 \cdots a_n} = \frac{a_1 + a_2 + \cdots + a_n}{n}.$$

En caso contrario,

$$\sqrt[n]{a_1 \cdot a_2 \cdots a_n} < \frac{a_1 + a_2 + \cdots + a_n}{n}.$$

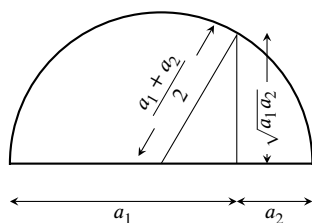


Figura A.3

Ejemplo A.4.3: Hallar una expresión para la suma

$$1 \cdot 1! + 2 \cdot 2! + 3 \cdot 3! + \cdots + n \cdot n!.$$

Demostración: Llamemos a la suma anterior, S_n . Es decir

$$S_n = 1 \cdot 1! + 2 \cdot 2! + 3 \cdot 3! + \cdots + n \cdot n!.$$

Para entender mejor esta suma, calculemos las primeras sumas parciales:

$$1 \cdot 1! = 1$$

$$1 \cdot 1! + 2 \cdot 2! = 5$$

$$1 \cdot 1! + 2 \cdot 2! + 3 \cdot 3! = 23$$

$$1 \cdot 1! + 2 \cdot 2! + 3 \cdot 3! + 4 \cdot 4! = 119$$

$$1 \cdot 1! + 2 \cdot 2! + 3 \cdot 3! + 4 \cdot 4! + 5 \cdot 5! = 719$$

$$1 \cdot 1! + 2 \cdot 2! + 3 \cdot 3! + 4 \cdot 4! + 5 \cdot 5! + 6 \cdot 6! = 5039$$

$$1 \cdot 1! + 2 \cdot 2! + 3 \cdot 3! + 4 \cdot 4! + 5 \cdot 5! + 6 \cdot 6! + 7 \cdot 7! = 40319$$

Examinando los resultados anteriores, podemos notar un patrón que se resume en

$$S_1 = 2! - 1$$

$$S_2 = 3! - 1$$

$$S_3 = 4! - 1$$

$$S_4 = 5! - 1$$

$$S_5 = 6! - 1$$

$$S_6 = 7! - 1$$

$$S_7 = 8! - 1$$

Esto conduce a la hipótesis

$$S_n = (n+1)! - 1.$$

Verifiquemos dicha hipótesis.

i. La hipótesis se cumple para $n = 1$, pues

$$S_1 = 1 \cdot 1! = 2! - 1.$$

ii. Supongamos que la hipótesis se cumple para $n = k$, es decir,

$$S_k = (k+1)! - 1.$$

iii. A partir de (ii), probemos que se cumple para $n = k+1$, es decir,

$$S_{k+1} = (k+2)! - 1.$$

En efecto,

$$\begin{aligned} S_{k+1} &= S_k + (k+1) \cdot (k+1)! \\ &= [(k+1)! - 1] + (k+1) \cdot (k+1)! \\ &= (k+1)! [1 + (k+1)] - 1 \\ &= (k+1)! (k+2) - 1 \\ &= (k+2)! - 1 \end{aligned}$$

Por tanto, se cumple que

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

Ejemplo A.4.4: Sea

$$A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix}.$$

Encuentre A^n para toda $n \in \mathbb{N}$.

Solución: Deseamos calcular A^n . Para esto, primero calculemos varias potencias de esta matriz como sigue:

$$\begin{aligned} A^2 &= \begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 2 \\ 0 & 0 & 0 \\ 2 & 0 & 2 \end{bmatrix} & A^4 = A^3 A = \begin{bmatrix} 8 & 0 & 8 \\ 0 & 0 & 0 \\ 8 & 0 & 8 \end{bmatrix} \\ A^3 &= \begin{bmatrix} 2 & 0 & 2 \\ 0 & 0 & 0 \\ 2 & 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 4 & 0 & 4 \\ 0 & 0 & 0 \\ 4 & 0 & 4 \end{bmatrix} & A^5 = A^4 A = \begin{bmatrix} 16 & 0 & 16 \\ 0 & 0 & 0 \\ 16 & 0 & 16 \end{bmatrix} \end{aligned}$$

Observemos que existe un patrón, así que proponemos

$$A^n = \begin{bmatrix} 2^{n-1} & 0 & 2^{n-1} \\ 0 & 0 & 0 \\ 2^{n-1} & 0 & 2^{n-1} \end{bmatrix}.$$

Probemosla por medio de inducción sobre n .

i. Para $n = 1$ es evidente que se cumple.

ii. Supongamos que la hipótesis se cumple para $n = k$, es decir,

$$A^k = \begin{bmatrix} 2^{k-1} & 0 & 2^{k-1} \\ 0 & 0 & 0 \\ 2^{k-1} & 0 & 2^{k-1} \end{bmatrix}.$$

iii. A partir de (ii), probemos que se cumple para $n = k + 1$, es decir,

$$A^{k+1} = \begin{bmatrix} 2^k & 0 & 2^k \\ 0 & 0 & 0 \\ 2^k & 0 & 2^k \end{bmatrix}.$$

En efecto, sabiendo que $A^{k+1} = A^k A$, se tiene

$$\begin{aligned} A^{k+1} &= \begin{bmatrix} 2^{k-1} & 0 & 2^{k-1} \\ 0 & 0 & 0 \\ 2^{k-1} & 0 & 2^{k-1} \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 2^{k-1} + 2^{k-1} & 0 & 2^{k-1} + 2^{k-1} \\ 0 & 0 & 0 \\ 2^{k-1} + 2^{k-1} & 0 & 2^{k-1} + 2^{k-1} \end{bmatrix} \\ &= \begin{bmatrix} 2^k & 0 & 2^k \\ 0 & 0 & 0 \\ 2^k & 0 & 2^k \end{bmatrix} \end{aligned}$$

Por lo tanto, se cumple que

$$A^n = \begin{bmatrix} 2^{n-1} & 0 & 2^{n-1} \\ 0 & 0 & 0 \\ 2^{n-1} & 0 & 2^{n-1} \end{bmatrix}, \forall n \in \mathbb{N}.$$

Esta fórmula no solo proporciona una expresión general para A^n , sino que también destaca cómo el comportamiento de las potencias de matrices puede describirse de manera sistemática y predecible mediante patrones algebraicos. El siguiente ejemplo, ilustra cómo se aplica este enfoque en un caso general.

Ejemplo A.4.5: Sea $a \in \mathbb{R}$ y sea

$$A = \begin{bmatrix} a & 1 \\ 0 & a \end{bmatrix}.$$

Encuentre A^n para toda $n \in \mathbb{N}$.

Solución: Deseamos calcular A^n . Para ello, comenzaremos calculando varias potencias iniciales de la matriz A y observaremos si se sigue un patrón:

$$A^2 = \begin{bmatrix} a & 1 \\ 0 & a \end{bmatrix} \begin{bmatrix} a & 1 \\ 0 & a \end{bmatrix} = \begin{bmatrix} a^2 & 2a \\ 0 & a^2 \end{bmatrix}$$

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

$$A^4 = \begin{bmatrix} a^3 & 3a^2 \\ 0 & a^3 \end{bmatrix} \begin{bmatrix} a & 1 \\ 0 & a \end{bmatrix} = \begin{bmatrix} a^4 & 4a^3 \\ 0 & a^4 \end{bmatrix}$$

$$A^5 = \begin{bmatrix} a^4 & 4a^3 \\ 0 & a^4 \end{bmatrix} \begin{bmatrix} a & 1 \\ 0 & a \end{bmatrix} = \begin{bmatrix} a^5 & 5a^4 \\ 0 & a^5 \end{bmatrix}$$

$$A^6 = \begin{bmatrix} a^5 & 5a^4 \\ 0 & a^5 \end{bmatrix} \begin{bmatrix} a & 1 \\ 0 & a \end{bmatrix} = \begin{bmatrix} a^6 & 6a^5 \\ 0 & a^6 \end{bmatrix}$$

A partir de estos cálculos iniciales, observamos que las potencias de A parecen seguir un patrón general. Conjeturamos que

$$A^n = \begin{bmatrix} a^n & na^{n-1} \\ 0 & a^n \end{bmatrix}.$$

Probemosla por medio de inducción sobre n .

i. Para $n = 1$ es evidente que se cumple.

ii. Supongamos que la hipótesis se cumple para $n = k$, es decir,

$$A^k = \begin{bmatrix} a^k & ka^{k-1} \\ 0 & a^k \end{bmatrix}.$$

iii. A partir de (ii), probemos que se cumple para $n = k + 1$, es decir,

$$A^{k+1} = \begin{bmatrix} a^{k+1} & (k+1)a^k \\ 0 & a^{k+1} \end{bmatrix}.$$

En efecto, sabiendo que $A^{k+1} = A^k A$, se tiene

$$\begin{aligned} A^{k+1} &= \begin{bmatrix} a^k & ka^{k-1} \\ 0 & a^k \end{bmatrix} \begin{bmatrix} a & 1 \\ 0 & a \end{bmatrix} \\ &= \begin{bmatrix} a^{k+1} & a^k + ka^k \\ 0 & a^{k+1} \end{bmatrix} \\ &= \begin{bmatrix} a^{k+1} & (k+1)a^k \\ 0 & a^{k+1} \end{bmatrix} \end{aligned}$$

Por lo tanto, se cumple que

$$A^n = \begin{bmatrix} a^n & na^{n-1} \\ 0 & a^n \end{bmatrix}, \forall n \in \mathbb{N}.$$

Ejemplo A.4.6:

Demuestra la siguiente igualdad usando inducción:

$$\begin{vmatrix} 1+x^2 & x & 0 & \cdots & 0 & 0 \\ x & 1+x^2 & x & \cdots & 0 & 0 \\ 0 & x & 1+x^2 & \ddots & 0 & 0 \\ \vdots & & & \ddots & & \\ 0 & 0 & 0 & \ddots & 1+x^2 & x \\ 0 & 0 & 0 & \cdots & x & 1+x^2 \end{vmatrix} = 1 + x^2 + x^4 + \cdots + x^{2n}.$$

Demostración: Para $n = 1$ la igualdad es evidente. Para $n = 2$, la matriz es

$$A_2 = \begin{bmatrix} 1+x^2 & x \\ x & 1+x^2 \end{bmatrix}$$

y el determinante de A_2 es

$$D_2 = (1+x^2)(1+x^2) - x^2 = 1 + 2x^2 + x^4 - x^2 = 1 + x^2 + x^4$$

que coincide con la fórmula $1 + x^2 + x^4$. Supongamos que para algún $k \geq 2$, el determinante de A_k es

$$D_k = 1 + x^2 + x^4 + \cdots + x^{2k}.$$

Debemos demostrar, a partir de lo anterior, que

$$D_{k+1} = 1 + x^2 + x^4 + \cdots + x^{2(k+1)}.$$

La matriz A_{k+1} es

$$A_{k+1} = \begin{bmatrix} 1+x^2 & x & 0 & \cdots & 0 & 0 \\ x & 1+x^2 & x & \cdots & 0 & 0 \\ 0 & x & 1+x^2 & \ddots & 0 & 0 \\ \vdots & & & \ddots & & \\ 0 & 0 & 0 & \ddots & 1+x^2 & x \\ 0 & 0 & 0 & \cdots & x & 1+x^2 \end{bmatrix}.$$

El determinante de A_{k+1} se puede calcular expandiendo por la primera fila

$$D_{k+1} = (1+x^2) D_k - x^2 D_{k-1}.$$

Por la hipótesis de inducción, sabemos que

$$D_k = 1 + x^2 + x^4 + \cdots + x^{2k} \quad \text{y} \quad D_{k-1} = 1 + x^2 + x^4 + \cdots + x^{2(k-1)}.$$

Entonces

$$\begin{aligned} D_{k+1} &= (1+x^2)(1+x^2+x^4+\cdots+x^{2k}) - x^2(1+x^2+x^4+\cdots+x^{2(k-1)}) \\ &= 1+x^2+x^4+\cdots+x^{2k} + x^2+x^4+\cdots+x^{2(k+1)} - x^2-x^4-\cdots-x^{2k} \\ &= 1+x^2+x^4+\cdots+x^{2(k+1)} \end{aligned}$$

Por lo tanto, para toda $n \in \mathbb{N}$, se cumple que

$$\begin{vmatrix} 1+x^2 & x & 0 & \cdots & 0 & 0 \\ x & 1+x^2 & x & \cdots & 0 & 0 \\ 0 & x & 1+x^2 & \ddots & 0 & 0 \\ \vdots & & & \ddots & & \\ 0 & 0 & 0 & \ddots & 1+x^2 & x \\ 0 & 0 & 0 & \cdots & x & 1+x^2 \end{vmatrix} = 1 + x^2 + x^4 + \cdots + x^{2n}.$$

Ejemplo A.4.7:

La sucesión numérica 1, 2, 3, 5, 8, 13, 21, ..., se denomina sucesión de Fibonacci. Demostrar que el n -ésimo término de la sucesión mencionada es igual al determinante de orden n :

$$A_n = \begin{bmatrix} 1 & 1 & 0 & 0 & \cdots & 0 & 0 \\ -1 & 1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & 1 & \ddots & 0 & 0 \\ \vdots & & & & \ddots & & \\ 0 & 0 & 0 & 0 & \cdots & -1 & 1 \end{bmatrix}.$$

Demostración: Para $n = 1$, la matriz es simplemente

$$A_1 = [1]$$

El determinante de A_1 es 1, que es el primer término de la serie de Fibonacci. Para $n = 2$, la matriz es:

$$A_2 = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}.$$

El determinante de A_2 es

$$\det(A_2) = (1)(1) - (-1)(1) = 1 + 1 = 2$$

que es el segundo término de la sucesión de Fibonacci. Sea F_k el k -ésimo término de la sucesión de Fibonacci. Supongamos que para algún $k \geq 2$, el determinante de A_k es el k -ésimo término de la sucesión de Fibonacci, es decir

$$\det(A_k) = F_k.$$

Debemos demostrar, a partir de lo anterior, que

$$\det(A_{k+1}) = F_{k+1}.$$

La matriz A_{k+1} es

$$A_{k+1} = \begin{bmatrix} 1 & 1 & 0 & 0 & \cdots & 0 & 0 \\ -1 & 1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & 1 & \ddots & 0 & 0 \\ \vdots & & & & \ddots & & \\ 0 & 0 & 0 & 0 & \cdots & -1 & 1 \end{bmatrix}.$$

El determinante de A_{k+1} se puede calcular expandiendo por la primera fila, es decir,

$$\det(A_{k+1}) = 1 \cdot \det(A_k) - 1 \cdot \det(A_{k-1}).$$

Por la hipótesis de inducción, sabemos que

$$\det(A_k) = F_k \quad \text{y} \quad \det(A_{k-1}) = F_{k-1}.$$

Entonces

$$\det(A_{k+1}) = F_k - (-F_{k-1}) = F_k + F_{k-1}.$$

Por la definición de la sucesión de Fibonacci, sabemos que

$$F_{k+1} = F_k + F_{k-1}.$$

Por lo tanto

$$\det(A_{k+1}) = F_{k+1}.$$

Por lo tanto, el n -ésimo término de la sucesión mencionada es igual al determinante de orden n :

$$A_n = \begin{bmatrix} 1 & 1 & 0 & 0 & \cdots & 0 & 0 \\ -1 & 1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & 1 & \ddots & 0 & 0 \\ \vdots & & & & \ddots & & \\ 0 & 0 & 0 & 0 & \cdots & -1 & 1 \end{bmatrix}.$$

Ejemplo A.4.8: Simplificar el polinomio

$$1 - \frac{x}{1!} + \frac{x(x-1)}{2!} - \dots + (-1)^n \frac{x(x-1) \cdots (x-n+1)}{n!}.$$

Solución: Para $n = 1$, se tiene

$$1 - \frac{x}{1!} = -\frac{x-1}{1}.$$

Para $n = 2$, se tiene

$$1 - \frac{x}{1!} + \frac{x(x-1)}{2!} = -\frac{x-1}{1} + \frac{x(x-1)}{2} = \frac{(x-1)(x-2)}{2!}.$$

Para $n = 3$, se tiene

$$\begin{aligned} 1 - \frac{x}{1!} + \frac{x(x-1)}{2!} - \frac{x(x-1)(x-2)}{3!} &= \frac{(x-1)(x-2)}{2} - \frac{x(x-1)(x-2)}{6} \\ &= -\frac{(x-1)(x-2)(x-3)}{3!} \end{aligned}$$

Esto sugiere la hipótesis de que

$$\begin{aligned} 1 - \frac{x}{1!} + \frac{x(x-1)}{2!} - \dots + (-1)^n \frac{x(x-1) \cdots (x-n+1)}{n!} \\ = (-1)^n \frac{(x-1)(x-2) \cdots (x-n)}{n!}. \end{aligned}$$

Probemosla por medio de inducción sobre n .

i. Para $n = 1$ es evidente que se cumple.

ii. Supongamos que la hipótesis se cumple para $n = k$, es decir,

$$\begin{aligned} 1 - \frac{x}{1!} + \frac{x(x-1)}{2!} - \dots + (-1)^k \frac{x(x-1) \cdots (x-k+1)}{k!} \\ = (-1)^k \frac{(x-1)(x-2) \cdots (x-k)}{k!}. \end{aligned}$$

iii. A partir de (ii), probemos que se cumple para $n = k + 1$, es decir,

$$\begin{aligned} 1 - \frac{x}{1!} + \frac{x(x-1)}{2!} - \dots + (-1)^{k+1} \frac{x(x-1) \cdots (x-k)}{(k+1)!} \\ = (-1)^{k+1} \frac{(x-1)(x-2) \cdots (x-k-1)}{(k+1)!}. \end{aligned}$$

En efecto,

$$\begin{aligned} 1 - \frac{x}{1!} + \frac{x(x-1)}{2!} - \dots + (-1)^k \frac{x(x-1) \cdots (x-k+1)}{k!} \\ + (-1)^{k+1} \frac{x(x-1) \cdots (x-k)}{(k+1)!} \\ = (-1)^k \frac{(x-1)(x-2) \cdots (x-k)}{k!} + (-1)^{k+1} \frac{x(x-1) \cdots (x-k)}{(k+1)!} \\ = (-1)^{k+1} \frac{(x-1)(x-2) \cdots (x-k)}{k!} \left[\frac{x}{k+1} - 1 \right] \\ = (-1)^{k+1} \frac{(x-1)(x-2) \cdots (x-k)(x-k-1)}{(k+1)!} \end{aligned}$$

Por lo tanto, para toda $n \in \mathbb{N}$,

$$\begin{aligned} 1 - \frac{x}{1!} + \frac{x(x-1)}{2!} - \dots + (-1)^n \frac{x(x-1) \cdots (x-n+1)}{n!} \\ = (-1)^n \frac{(x-1)(x-2) \cdots (x-n)}{n!}. \end{aligned}$$

Ejemplo A.4.9:

Entre las muchas sucesiones de números interesantes que se encuentran en las matemáticas discretas y la combinatoria, se encuentran los números armónicos, denotados por H_1, H_2, H_3, \dots , donde se definen como sumas parciales de la serie armónica. Estos números surgen de forma natural en diversas áreas de la matemática, como el análisis de algoritmos, la teoría de números y la probabilidad. La definición de los números armónicos es la siguiente:

$$\begin{aligned} H_1 &= 1 \\ H_2 &= 1 + \frac{1}{2} \\ H_3 &= 1 + \frac{1}{2} + \frac{1}{3} \\ H_4 &= 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} \\ H_5 &= 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} \\ &\vdots \\ H_n &= 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \dots + \frac{1}{n} \end{aligned}$$

La siguiente propiedad de los números armónicos nos ofrece una oportunidad más para aplicar inducción: Para toda $n \in \mathbb{N}$,

$$\sum_{j=1}^n H_j = (n+1)H_n - n.$$

Demostración: Para $n = 1$, tenemos que

$$\sum_{j=1}^k H_j = H_1 = 1$$

y

$$(1+1)H_1 - 1 = 2 \cdot 1 - 1 = 1.$$

Por lo tanto, la propiedad se cumple para $n = 1$. Supongamos que la propiedad se cumple para algún $k \geq 1$, es decir

$$\sum_{j=1}^k H_j = (k+1)H_k - k.$$

Debemos demostrar que se cumple para $k+1$. Es decir, queremos probar que

$$\sum_{j=1}^{k+1} H_j = (k+2)H_{k+1} - (k+1).$$

Entonces

$$\begin{aligned} \sum_{j=1}^{k+1} H_j &= \sum_{j=1}^k H_j + H_{k+1} \\ &= [(k+1)H_k - k] + H_{k+1} \\ &= (k+1) \left(H_{k+1} - \frac{1}{k+1} \right) - k + H_{k+1} \\ &= (k+2)H_{k+1} - 1 - k \\ &= (k+2)H_{k+1} - (k+1) \end{aligned}$$

Por lo tanto, para toda $n \in \mathbb{N}$,

$$\sum_{j=1}^n H_j = (n+1)H_n - n.$$

A.5 Ejercicios del Apéndice A

1. Demostrar que

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

2. Demostrar que

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

3. Demostrar que

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

4. Demostrar que

$$2^n > n, \forall n \in \mathbb{N}.$$

5. Demostrar que

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

6. Demostrar que para toda $n \in \mathbb{N}$

$$1 \cdot 2 \cdot 3 + 2 \cdot 3 \cdot 4 + \cdots + n(n+1)(n+2) = \frac{n(n+1)(n+2)(n+3)}{4}.$$

7. Demostrar que para toda $n \in \mathbb{N}$, $n > 1$

$$\frac{1}{n+1} + \frac{1}{n+2} + \cdots + \frac{1}{2n} > \frac{13}{24}.$$

8. Demostrar que para toda $n \in \mathbb{N}$, $n > 1$

$$\frac{1}{\sqrt{1}} + \frac{1}{\sqrt{2}} + \cdots + \frac{1}{\sqrt{n}} > \sqrt{n}.$$

9. Demostrar que para toda $n \in \mathbb{N}$

$$1 + 2n \leq 3^n.$$

10. Demuestre que $(n-1)^2 \mid n^{n-1} - 1$ para toda $n \in \mathbb{N}$, $n > 1$.

11. Demuestre que $n^2 + n$ es un número par para toda $n \in \mathbb{N}$.

12. Demuestre que $8^n - 3^n$ es múltiplo de 5 para toda $n \in \mathbb{N}$.

13. Demuestre que $n^3 - n$ es múltiplo de 6 para toda $n \in \mathbb{N}$.

14. Demuestre que $2^{2n-1} + 3^{2n-1}$ es múltiplo de 5 para toda $n \in \mathbb{N}$.

15. Demuestre que $11^{n+1} + 12^{2n-1}$ es múltiplo de 19 para toda $n \in \mathbb{N}$.

16. Si $a \neq 1$, entonces

$$1 + a + a^2 + \cdots + a^n = \frac{a^{n+1} - 1}{a - 1}, \forall n \in \mathbb{N}.$$

17. Demuestre que

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

18. Demostrar por inducción

$$1 + nx \leq (1 + x)^n, \forall n \in \mathbb{N}, x \geq -1.$$

Esta desigualdad es conocida como desigualdad de Bernoulli.

19. Demuestre que el número S_n de permutaciones de n elementos está dado por

$$S_n = n!.$$

20. Demuestre la siguiente igualdad para $a \neq b$

$$\begin{vmatrix} a+b & ab & 0 & \cdots & 0 & 0 \\ 1 & a+b & ab & \cdots & 0 & 0 \\ 0 & 1 & a+b & \ddots & 0 & 0 \\ \vdots & & & \ddots & & \\ 0 & 0 & 0 & \ddots & a+b & ab \\ 0 & 0 & 0 & \cdots & 1 & a+b \end{vmatrix} = \frac{a^{n+1} - b^{n+1}}{a - b}$$

donde el determinante es de orden n .

21. Demuestra la siguiente igualdad

$$\begin{vmatrix} 2 \cos \alpha & 1 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 2 \cos \alpha & 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 2 \cos \alpha & 1 & \ddots & 0 & 0 \\ \vdots & & & & \ddots & & \\ 0 & 0 & 0 & 0 & \cdots & 1 & 2 \cos \alpha \end{vmatrix} = \frac{\sin n\alpha}{\sin \alpha}$$

donde el determinante es de orden $n - 1$.

22. Probar la identidad

$$\cos(\varphi) \cos(2\varphi) \cos(4\varphi) \cdots \cos(2^n \varphi) = \frac{\sin(2^{n+1} \varphi)}{2^{n+1} \sin(\varphi)}.$$

23. Probar que para $x \neq m\pi$,

$$\frac{1}{2} \tan \frac{x}{2} + \frac{1}{2^2} \tan \frac{x}{2^2} + \cdots + \frac{1}{2^n} \tan \frac{x}{2^n} = \frac{1}{2^n} \cot \frac{x}{2^n} - \cot x.$$

24. Probar que

$$\frac{1}{a(a+1)} + \frac{1}{(a+1)(a+2)} + \cdots + \frac{1}{(a+n-1)(a+n)} = \frac{n}{a(a+n)}.$$

25. Demuestre que

$$\frac{d^n}{dx^n} \left(\frac{\sin x}{x} \right) = \frac{1}{x^{n+1}} \int_0^x y^n \cos \left(y + \frac{n\pi}{2} \right) dy, \forall n \in \mathbb{N}.$$

26. Demuestre que

$$\binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{n} = 2^n, \forall n \in \mathbb{N}.$$

27. Sea $n, k \in \mathbb{N}$ con $0 \leq k \leq n$, demuestre que

$$\binom{k}{k} + \binom{k+1}{k} + \binom{k+2}{k} + \cdots + \binom{n}{k} = \binom{n+1}{k+1}.$$

B

Apéndice

NÚMEROS COMPLEJOS

La resolución de ecuaciones algebraicas ha sido, a lo largo de la historia de las matemáticas, uno de los problemas más fundamentales y, al mismo tiempo, un motor impulsor del desarrollo teórico de la disciplina. Desde los antiguos matemáticos babilonios, pasando por los matemáticos árabes y europeos del Renacimiento, hasta los avances modernos, el interés por resolver ecuaciones polinómicas ha generado herramientas y conceptos que han transformado profundamente nuestra forma de entender las matemáticas.

Entre las ecuaciones algebraicas, aquellas de una incógnita y con coeficientes reales poseen una importancia particular. Estas ecuaciones se escriben generalmente en la forma:

$$a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0 = 0,$$

donde $a_n, a_{n-1}, \dots, a_1, a_0$ son números reales, y x representa la incógnita o indeterminada. Cada uno de los coeficientes describe una característica particular de la ecuación y, en conjunto, definen las propiedades del polinomio asociado. El término a_n , que corresponde al coeficiente del término de mayor exponente, es de especial relevancia, ya que si $a_n \neq 0$, el valor n se conoce como el grado del polinomio o de la ecuación. Este grado determina el comportamiento del polinomio, tales como su forma general y el número máximo de soluciones posibles en un sistema numérico dado.

Resolver una ecuación polinómica como la anterior implica encontrar todos los valores de x que la satisfagan. Dichos valores se conocen como las *raíces* o *soluciones* de la ecuación. En otras palabras, son aquellos números que, al sustituirlos en lugar de x y realizar todas las operaciones indicadas, reducen el primer miembro de la ecuación a cero. Por ejemplo, si consideramos la ecuación cuadrática simple $x^2 - 5x + 6 = 0$, sus raíces son los valores de x que satisfacen la igualdad, en este caso, $x = 2$ y $x = 3$. Para cada una de estas soluciones, al sustituir x en el polinomio $x^2 - 5x + 6$, obtenemos 0, verificando que son efectivamente soluciones.

El sistema de los números reales \mathbb{R} es extremadamente poderoso y útil. Desde su construcción formal, este conjunto ha servido como el fundamento de gran parte de las matemáticas y ha permitido resolver una amplia variedad de problemas en diversas disciplinas. Sin embargo, hay situaciones en las que \mathbb{R} resulta insuficiente, especialmente cuando tratamos de resolver ciertas ecuaciones algebraicas.

Consideremos, por ejemplo, la ecuación cuadrática

$$x^2 + 1 = 0. \quad (\text{B.0.1})$$

Esta ecuación, aunque simple, nos revela una limitación fundamental del sistema de números reales. Supongamos que $t \in \mathbb{R}$ es una solución de (B.0.1), entonces debe cumplirse que

$$t^2 + 1 = 0.$$

Por lo tanto,

$$t^2 = -1.$$

Aquí radica el problema: dentro del sistema de números reales, no existe ningún número t tal que $t^2 = -1$. Si $t > 0$, entonces $t^2 > 0$. Si $t = 0$, entonces $t^2 = 0$. Si $t < 0$, entonces $t^2 > 0$. En todos los casos, el valor de t^2 es siempre mayor o igual a cero. Por lo tanto, no existe ningún número real t tal que $t^2 = -1$. Esto significa que la ecuación $x^2 + 1 = 0$ no tiene solución en el conjunto de los números reales.

La imposibilidad de resolver ecuaciones como $x^2 + 1 = 0$ dentro del sistema de números reales plantea una pregunta natural: ¿es posible extender \mathbb{R} de alguna manera que permita encontrar soluciones a este tipo de ecuaciones? La respuesta es afirmativa, y lo que haremos enseguida es ampliar el sistema de los números reales, a un sistema de números donde, por lo menos, la ecuación (B.0.1) tenga solución.

Definición B.0.1:

Un número complejo z es una pareja ordenada de \mathbb{R} . El conjunto de números complejos lo denotaremos por \mathbb{C} , donde

$$\mathbb{C} = \{(a, b) \mid a, b \in \mathbb{R}\}.$$

El sistema \mathbb{C} no es una simple extensión arbitraria de \mathbb{R} , sino una construcción cuidadosamente definida que conserva muchas de las propiedades de \mathbb{R} .

Definición B.0.2:

Decimos que dos números complejos $z = (a, b)$ y $w = (c, d)$ son iguales, y lo denotaremos por $z = w$, si $a = c$ y $b = d$.

Definición B.0.3:

Sea $z \in \mathbb{C}$, con $z = (a, b)$. Definimos:

- i. La parte real de z , denotada por $\text{Re}(z)$, como $\text{Re}(z) = a$.
- ii. La parte imaginaria de z , denotada por $\text{Im}(z)$, como $\text{Im}(z) = b$.

B.1 Operaciones con números complejos

El conjunto de los números complejos se caracteriza por la presencia de dos operaciones fundamentales: la suma y el producto. Para definir la suma de dos números complejos z y w , representados como $z = (a, b)$ y $w = (c, d)$ en el conjunto \mathbb{C} , utilizamos la siguiente expresión:

$$z + w = (a, b) + (c, d) = (a + c, b + d).$$

Asimismo, para definir el producto de estos números complejos z y w , expresados como $z = (a, b)$ y $w = (c, d)$ en el conjunto \mathbb{C} , empleamos la siguiente fórmula:

$$zw = (a, b)(c, d) = (ac - bd, ad + bc).$$

Es importante destacar que la suma cumple con las siguientes propiedades:

A1. Cerradura: Si $z, w \in \mathbb{C}$, entonces $z + w \in \mathbb{C}$. En efecto: Sean $z, w \in \mathbb{C}$ con $z = (a, b)$ y $w = (c, d)$, entonces

$$\begin{aligned} z + w &= (a, b) + (c, d) \\ &= (a + c, b + d) \in \mathbb{C}. \end{aligned}$$

A2. Asociatividad: Si $z, w, y \in \mathbb{C}$, entonces

$$(z + w) + y = z + (w + y).$$

En efecto: Sean $z, w \in \mathbb{C}$ con $z = (a, b)$, $w = (c, d)$ e $y = (e, f)$. Entonces

$$\begin{aligned} (z + w) + y &= ((a, b) + (c, d)) + (e, f) \\ &= (a + c, b + d) + (e, f) \\ &= ((a + c) + e, (b + d) + f) \\ &= (a + (c + e), b + (d + f)) \\ &= (a, b) + (c + e, d + f) \\ &= (a, b) + ((c, d) + (e, f)) \\ &= (a, b) + (w + y). \end{aligned}$$

A3. Conmutatividad: Si $z, w \in \mathbb{C}$, entonces

$$z + w = w + z.$$

En efecto: Sean $z, w \in \mathbb{C}$ tal que $z = (a, b)$ y $w = (c, d)$, entonces

$$\begin{aligned} z + w &= (a, b) + (c, d) \\ &= (a + c, b + d) \\ &= (c + a, d + b) \\ &= (c, d) + (a, b) \\ &= w + z. \end{aligned}$$

A4. Elemento neutro: Existe un único número complejo, denotado por 0, que satisface

$$z + 0 = z, \forall z \in \mathbb{C}.$$

En efecto: Sea $z \in \mathbb{C}$, con $z = (a, b)$.

Existencia: Sea $w = (0, 0)$, entonces

$$\begin{aligned} z + w &= (a, b) + (0, 0) \\ &= (a + 0, b + 0) \\ &= (a, b) \\ &= z. \end{aligned}$$

Por lo que $z + w = z$.

Unicidad: Sea $v \in \mathbb{C}$ tal que

$$z + v = z, \forall z \in \mathbb{C},$$

en particular para $z = w$, tenemos que $w + v = w$. Pero $z + w = z, \forall z \in \mathbb{C}$, en particular para $z = v$, tenemos $v + w = v$. Así que

$$\begin{aligned} v &= v + w \\ &= w + v \\ &= w. \end{aligned}$$

Por lo tanto $v = w$.

A5. Inverso aditivo: Dado $z \in \mathbb{C}$, existe un único número complejo, denotado por $-z$, que satisface

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

En efecto: Sea $z \in \mathbb{C}$, con $z = (a, b)$.

Existencia: Sea $w = (-a, -b)$, entonces

$$\begin{aligned} z + w &= (a, b) + (-a, -b) \\ &= (a + (-a), b + (-b)) \\ &= (0, 0) \\ &= 0. \end{aligned}$$

Unicidad: Sea $v \in \mathbb{C}$ tal que $z + v = 0$. Tenemos

$$\begin{aligned} w &= w + 0 \\ &= w + (z + v) \\ &= (w + z) + v \\ &= 0 + v \\ &= v. \end{aligned}$$

Por lo tanto $w = v$.

El producto satisface:

A6. Cerradura: Sea $z, w \in \mathbb{C}$, entonces $zw \in \mathbb{C}$. En efecto: Sea $z, w \in \mathbb{C}$ con $z = (a, b)$ y $w = (c, d)$, entonces

$$\begin{aligned} zw &= (a, b)(c, d) \\ &= (ac - bd, ad + bc) \in \mathbb{C}. \end{aligned}$$

A7. Asociatividad: Si $z, w, y \in \mathbb{C}$, entonces

$$(zw)y = z(wy).$$

En efecto: Sean $z, w, y \in \mathbb{C}$ con $z = (a, b)$, $w = (c, d)$ e $y = (e, f)$. Entonces

$$\begin{aligned} (zw)y &= [(a, b)(c, d)](e, f) \\ &= (ac - bd, ad + bc)(e, f) \\ &= ((ac - bd)e - (ad + bc)f, (ac - bd)f + (ad + bc)e) \\ &= (ace - bde - adf - bcf, acf - bdf + ade + bce) \\ &= (ace - adf - bcf - bde, acf + ade + bce - bdf) \\ &= (a(ce - df) - b(cf + de), a(cf + de) + b(ce - df)) \\ &= (a, b)(ce - df, cf + de) \\ &= (a, b)((c, d)(e, f)) \\ &= z(wy). \end{aligned}$$

A8. Conmutatividad: Si $z, w \in \mathbb{C}$, entonces

$$zw = wz.$$

En efecto: Sean $z, w \in \mathbb{C}$ con $z = (a, b)$ y $w = (c, d)$, entonces

$$\begin{aligned} zw &= (a, b)(c, d) \\ &= (ac - bd, ad + bc) \\ &= (ca - db, cb + da) \\ &= (c, d)(a, b) \\ &= wz. \end{aligned}$$

A9. Elemento identidad: Existe un único complejo, denotado por 1, que satisface

$$1 \cdot z = z, \forall z \in \mathbb{C}.$$

En efecto: Sea $z \in \mathbb{C}$ con $z = (a, b)$.

Existencia: Sea $w = (1, 0)$, entonces

$$\begin{aligned} wz &= (1, 0)(a, b) \\ &= (1a - 0b, 1b + 0a) \\ &= (a, b) \\ &= z. \end{aligned}$$

Unicidad: Sea $v \in \mathbb{C}$ tal que

$$vz = z, \forall z \in \mathbb{C},$$

en particular para $z = w$, entonces $vw = w$. Como $wz = z, \forall z \in \mathbb{C}$, en particular $z = v$, entonces $wv = v$. Así que

$$\begin{aligned} v &= wv \\ &= vw \\ &= w. \end{aligned}$$

Por tanto, $v = w$.

A10. Inverso multiplicativo: Dado $z \in \mathbb{C}$, con $z \neq 0$, existe un único complejo tal que

$$z \cdot z^{-1} = 1.$$

En efecto: Sea $z \in \mathbb{C}$, con $z = (a, b)$ y $z \neq 0$, entonces $a \neq 0$ o $b \neq 0$. Deseamos encontrar $w \in \mathbb{C}$ tal que $zw = (1, 0)$. Expresemos $w = (x, y)$, entonces

$$\begin{aligned} (1, 0) &= zw \\ &= (a, b)(x, y) \\ &= (ax - by, ay + bx). \end{aligned}$$

Luego $ax - by = 1$ y $ay + bx = 0$. Es decir

$$\begin{cases} ax - by = 1 \\ ay + bx = 0 \end{cases} \quad (\text{B.1.1})$$

- i. Si $a = 0$, entonces $b \neq 0$. Por lo tanto, $by = 1$, de donde se sigue que $y = -\frac{1}{b}$. Además, $bx = 0$, lo que implica que $x = 0$.
- ii. Si $b = 0$, entonces $a \neq 0$. Por lo tanto, $ax = 1$, de donde se sigue que $x = \frac{1}{a}$. Además, $ay = 0$, lo que implica que $y = 0$.
- iii. Supongamos que $a, b \neq 0$. Entonces $a^2 + b^2 \neq 0$. Multiplicando la primer ecuación de (B.1.1) por a y la segunda ecuación de (B.1.1) por b , obtenemos que

$$\begin{cases} a^2x - aby = a \\ aby + b^2x = 0 \end{cases}$$

Sumando dichas ecuaciones, se sigue que

$$x(a^2 + b^2) = a.$$

Por lo tanto,

$$x = \frac{a}{a^2 + b^2}.$$

Por otro lado, multiplicando la primera ecuación de (B.1.1) por $-b$ y la segunda ecuación de (B.1.1) por a , obtenemos que

$$\begin{cases} -bax + b^2y = -b \\ bax + a^2y = 0 \end{cases}$$

Sumando dichas ecuaciones, se sigue que

$$y(a^2 + b^2) = -b.$$

Por lo tanto,

$$y = -\frac{b}{a^2 + b^2}.$$

De donde se hereda que,

$$z^{-1} = \left(\frac{a}{a^2 + b^2}, \frac{-b}{a^2 + b^2} \right).$$

La adición y el producto se relacionan con la siguiente propiedad:

A11. Distributividad: Sean $z, w, y \in \mathbb{C}$, entonces

$$z(x + y) = zx + zy.$$

En efecto: Se deja como ejercicio al lector.

Isomorfismo entre los números reales:

Sea $X = \{(a, 0) \mid a \in \mathbb{R}\}$. Sea $f: \mathbb{R} \rightarrow X$ dada por $f(a) = (a, 0)$, para todo $a \in \mathbb{R}$. Sean $a, b \in \mathbb{R}$. Entonces

$$\begin{aligned} f(a+b) &= (a+b, 0) & f(a)f(b) &= (a, 0)(b, 0) \\ &= (a, 0) + (b, 0) & &= (ab, 0) \\ &= f(a) + f(b) & &= f(ab) \end{aligned}$$

De esta forma,

- i. Para demostrar que f es una función inyectiva, tomemos $a, b \in \mathbb{R}$ tal que $f(a) = f(b)$. Esto implica que $(a, 0) = (b, 0)$, lo que a su vez conduce a $a = b$. Por lo tanto, concluimos que f es inyectiva.
- ii. Para demostrar la sobreyectividad de f , tomemos $(a, 0) \in X$. Entonces, se cumple que $f(a) = (a, 0)$.
- iii. Dado que hemos probado que f es inyectiva y sobreyectiva, concluimos que f es biyectiva.

Basándonos en lo expuesto anteriormente, podemos afirmar que X es isomorfo al conjunto de los números reales. Esto implica que, para cualquier elemento $a \in \mathbb{R}$, podemos asociarlo de manera natural con el elemento $(a, 0) \in X$. Por conveniencia, identificaremos directamente cada número real a con el elemento $(a, 0)$ en X , escribiendo $a = (a, 0)$. Esta notación simplifica la representación y mantiene consistencia con las propiedades algebraicas del conjunto. Bajo esta identificación, los elementos neutros de la suma y del producto en X se representan de la siguiente manera: $0 = (0, 0)$ y $1 = (1, 0)$.

Para desarrollar una notación más conveniente, definimos la *unidad imaginaria* como $i = (0, 1)$. Observemos que

$$\begin{aligned} i^2 &= (0, 1)(0, 1) \\ &= (0 \cdot 0 - 1 \cdot 1, 0 \cdot 1 + 1 \cdot 0) \\ &= (-1, 0) \\ &= -1 \end{aligned}$$

Como identificamos $(-1, 0)$ con el número real -1 , hemos demostrado que $i^2 = -1$. Además, si $b \in \mathbb{R}$, entonces

$$\begin{aligned} bi &= (b, 0)(0, 1) \\ &= (b \cdot 0 - 0 \cdot 1, b \cdot 1 + 0 \cdot 0) \\ &= (0, b) \end{aligned}$$

Nota

A la expresión $a + bi$ se le llama *forma binómica*, donde a es la parte real y b es la parte imaginaria.

Un número complejo $z = bi$ cuya parte real es cero se dice que es un *imaginario puro*. De esta forma, cualquier número complejo $z = (a, b)$ se puede reescribir usando la unidad imaginaria i :

$$\begin{aligned} z &= (a, b) \\ &= (a, 0) + (0, b) \\ &= (a, 0) + (b, 0)(0, 1) \\ &= a + bi \end{aligned}$$

Sean $z, w \in \mathbb{C}$ con $z = a + bi$ y $w = c + di$, entonces

$$\begin{aligned} z + w &= (a + bi) + (c + di) \\ &= (a + c) + (bi + di) \\ &= (a + c) + (b + d)i \end{aligned}$$

y

$$\begin{aligned} zw &= (a + bi)(c + di) \\ &= ac + adi + bci + bdi^2 \\ &= ac + (ad + bc)i - bd \\ &= (ac - bd) + (ad + bc)i \end{aligned}$$

El inverso multiplicativo de un número complejo $w = c + di$, con $w \neq 0$ se puede expresar en forma binómica. Multiplicando el numerador y el denominador por $c - di$, obtenemos que

$$\begin{aligned} w^{-1} &= \frac{1}{w} \\ &= \frac{1}{c + di} \\ &= \frac{1}{c + di} \cdot \frac{c - di}{c - di} \\ &= \frac{c - di}{(c + di)(c - di)} \\ &= \frac{c - di}{c^2 + d^2} \\ &= \frac{c}{c^2 + d^2} - \frac{d}{c^2 + d^2}i \end{aligned}$$

De esta manera, la división de dos números complejos z y w se define como:

$$\begin{aligned} \frac{z}{w} &= \frac{a + bi}{c + di} \\ &= \frac{a + bi}{c + di} \cdot \frac{c - di}{c - di} \\ &= \frac{ac + bd + (-ad + bc)i}{c^2 + d^2} \\ &= \frac{ac + bd}{c^2 + d^2} + \frac{-ad + bc}{c^2 + d^2}i. \end{aligned}$$

El siguiente teorema enuncia algunas propiedades.

Teorema B.1.1: Sea $z \in \mathbb{C}$, entonces

- i) $z \cdot 0 = 0$.
- ii) $-z = (-1)z$.
- iii) $-(-z) = z$.

Demostración: Se dejan como ejercicio al lector.

Ejemplo B.1.1:

En la práctica, suele ser más conveniente calcular productos de números complejos por expansión, en lugar de sustituir en alguna fórmula. Por ejemplo,

$$(3 - 2i)(4 + 5i) = 12 + 15i - 8i - 10i^2 = (12 + 10) + 7i = 22 + 7i.$$

Un número complejo $z = a + bi$ puede asociarse con el par ordenado (a, b) de números reales y representarse geoméricamente mediante un punto o un vector en el plano xy (figura B.1a). A esto lo llamamos el *plano complejo*. Los puntos en el eje x tienen una parte imaginaria de cero y, por lo tanto, corresponden a números reales, mientras que los puntos en el eje y tienen una parte real de cero y corresponden a números imaginarios puros. En consecuencia, llamamos al eje x el *eje real* y al eje y el *eje imaginario* (figura B.1b).

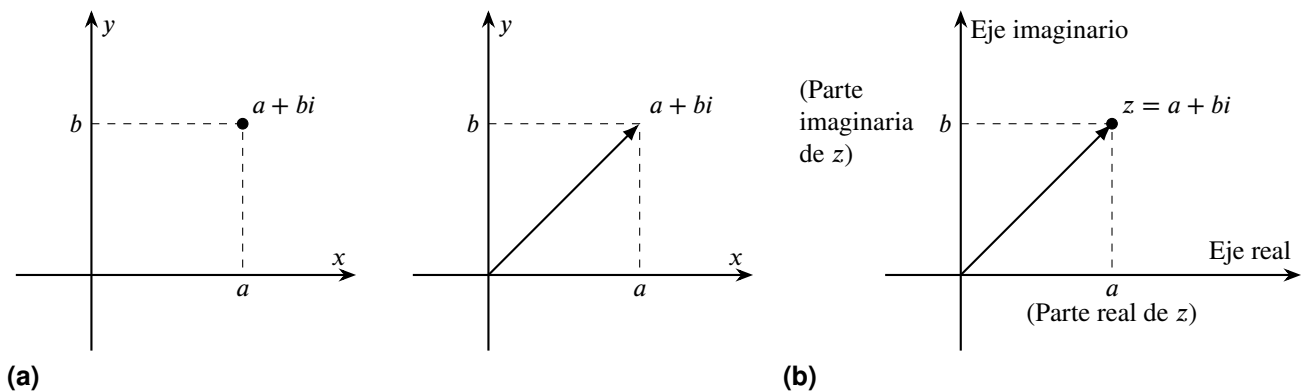


Figura B.1 Representación geométrica de un número complejo como punto y vector en el plano complejo

Los números complejos pueden sumarse, restarse o multiplicarse por números reales geoméricamente al realizar estas operaciones sobre sus vectores asociados (véase, por ejemplo, la figura B.2). En este sentido, el sistema de los números complejos \mathbb{C} está estrechamente relacionado con \mathbb{R}^2 , siendo la principal diferencia que los números complejos pueden multiplicarse para producir otros números complejos, mientras que no existe una operación de multiplicación en \mathbb{R}^2 que produzca otros vectores en \mathbb{R}^2 (el producto punto de la definición 2.2.1 produce un escalar, no un vector en \mathbb{R}^2).

Figura B.2 La suma y resta de dos números complejos

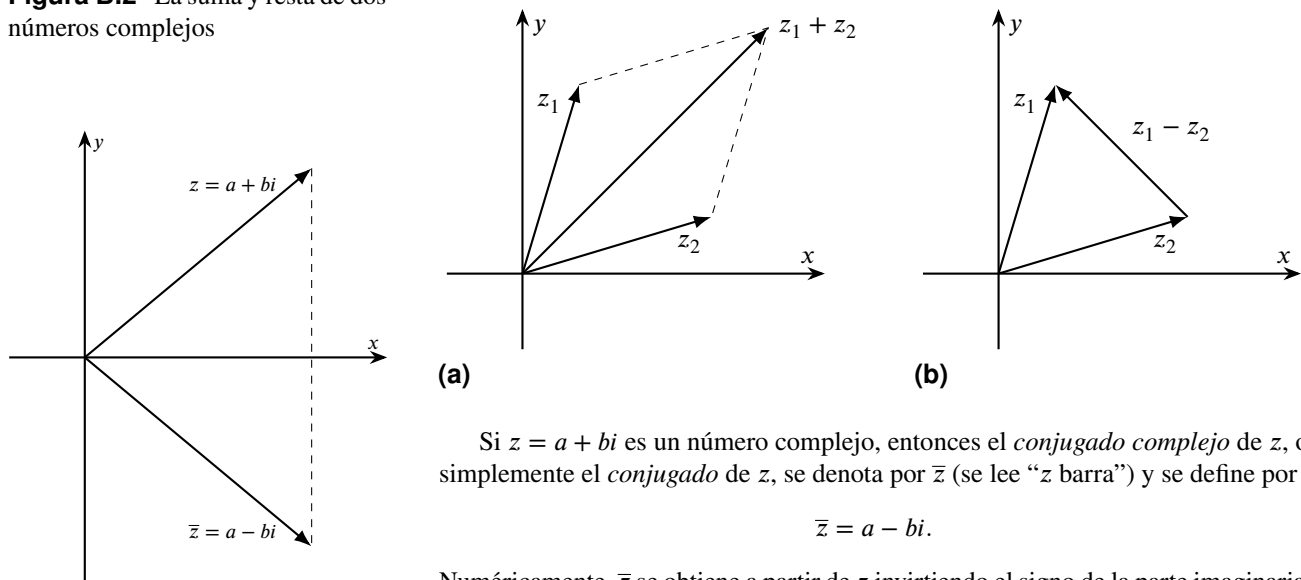


Figura B.3

Si $z = a + bi$ es un número complejo, entonces el *conjugado complejo* de z , o simplemente el *conjugado* de z , se denota por \bar{z} (se lee “ z barra”) y se define por

$$\bar{z} = a - bi.$$

Numéricamente, \bar{z} se obtiene a partir de z invirtiendo el signo de la parte imaginaria, y geoméricamente se obtiene reflejando el vector de z con respecto al eje real (figura B.3).

Ejemplo B.1.2:

Consideremos el número complejo $z = 4 + 7i$. Para encontrar el conjugado, solo cambiamos el signo de la parte imaginaria. Por lo tanto, $\bar{z} = 4 - 7i$.

Observamos que z y su conjugado \bar{z} son diferentes cuando la parte imaginaria de z no es cero. Esto nos lleva a una propiedad importante que caracteriza a los números reales entre los números complejos.

Teorema B.1.2:

Un número complejo z es un número real si y solo si es igual a su conjugado, es decir, si $z = \bar{z}$.

El siguiente cálculo muestra que el producto de un número complejo $z = a + bi$ y su conjugado $\bar{z} = a - bi$ es un número real no negativo:

$$z\bar{z} = (a + bi)(a - bi) = a^2 - abi + bai - b^2i^2 = a^2 + b^2.$$

Observemos que

$$\sqrt{z\bar{z}} = \sqrt{a^2 + b^2}.$$

es la longitud del vector correspondiente a z (figura B.4); a esta longitud la llamamos el *módulo* (o *valor absoluto* de z) y lo denotamos por $|z|$. Así, tenemos que

$$|z| = \sqrt{z\bar{z}} = \sqrt{a^2 + b^2}.$$

Notemos que si $b = 0$, entonces $z = a$ es un número real y $|z| = \sqrt{a^2} = |a|$, lo cual nos indica que el módulo de un número real coincide con su valor absoluto.

Recordemos que si $z \neq 0$, entonces el inverso multiplicativo de z se define mediante la propiedad

$$z \cdot z^{-1} = 1.$$

Esta ecuación tiene una solución única para z^{-1} , la cual se puede obtener multiplicando ambos lados por \bar{z} y usando el hecho de que $z\bar{z} = |z|^2$. Esto da como resultado

$$z^{-1} = \frac{1}{z} = \frac{\bar{z}}{|z|^2}.$$

Si $w \neq 0$, entonces la división entre los números complejos z y w se define como el producto de z y w^{-1} . Esto nos da la fórmula

$$\frac{z}{w} = z \frac{\bar{w}}{|w|^2} = \frac{z\bar{w}}{|w|^2}.$$

Observemos que la expresión en el lado derecho se obtiene si el numerador y el denominador se multiplican por \bar{w} . En la práctica, esta suele ser la mejor forma de realizar divisiones de números complejos.

Ejemplo B.1.3:

Sean $z = 3 + 4i$ y $w = 1 - 2i$. Expresa la división z entre w .

Solución: Multiplicaremos el numerador y denominador por \bar{w} . Por lo tanto,

$$\frac{z}{w} = \frac{z\bar{w}}{w\bar{w}} = \frac{3 + 4i}{1 - 2i} \cdot \frac{1 + 2i}{1 + 2i} = \frac{3 + 6i + 4i + 8i^2}{1 - 4i^2} = -1 + 2i.$$

Los siguientes teoremas enumeran algunas propiedades útiles del módulo y el conjugado de un número complejo.

Teorema B.1.3:

Sean $z, z_1, z_2 \in \mathbb{C}$. Entonces

- i) $\overline{z_1 + z_2} = \bar{z}_1 + \bar{z}_2$.
- ii) $\overline{z_1 - z_2} = \bar{z}_1 - \bar{z}_2$.
- iii) $\overline{z_1 z_2} = \bar{z}_1 \bar{z}_2$.
- iv) $\bar{\bar{z}} = z$.

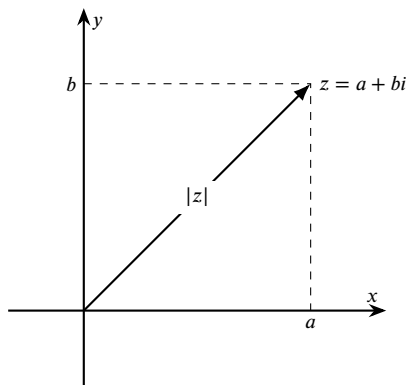


Figura B.4

Teorema B.1.4: Sean $z, z_1, z_2 \in \mathbb{C}$. Entonces

- i) $|\bar{z}| = |z|$.
- ii) $|z_1 z_2| = |z_1| |z_2|$.
- iii) $|z^{-1}| = |z|^{-1}$.
- iv) $\left| \frac{z_1}{z_2} \right| = \frac{|z_1|}{|z_2|}$.
- v) $|z_1 + z_2| \leq |z_1| + |z_2|$.

B.2 Forma polar de un número complejo

Sea $z = a + bi$ un número complejo y sea $r = |z|$. Si $z \neq 0$, considerando su representación geométrica, sea θ la medida del ángulo que forman el eje real positivo y el segmento que une el origen del plano complejo con el punto que representa a z , entonces se tiene que

$$a = r \cos \theta,$$

$$b = r \sin \theta.$$

En consecuencia,

$$\begin{aligned} z &= r \cos \theta + ri \sin \theta \\ &= r (\cos \theta + i \sin \theta) \end{aligned}$$

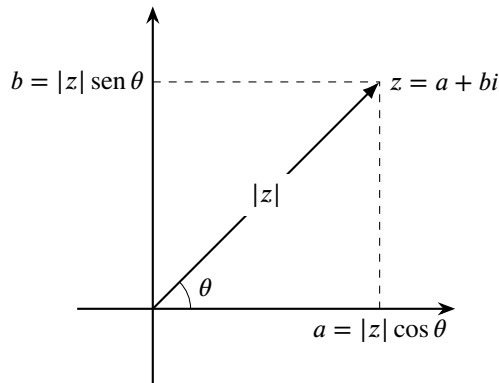
A θ se le llama la *amplitud* o *argumento* de z , y escribimos $\theta = \arg z$. Si $z = 0$, entonces $r = 0$, y por lo tanto $z = r(\cos \theta + i \sin \theta)$ para cualquier θ .

En consecuencia, todo complejo $z = a + bi$ puede expresarse como

$$z = r(\cos \theta + i \sin \theta),$$

donde $r = |z|$ y $\theta = \arg z$, llamada *forma trigonométrica de z* o *forma polar de z* .

Figura B.5 Forma polar de un número complejo



Puesto que $\forall m \in \mathbb{Z}$

$$\cos(2m\pi + \alpha) = \cos(\alpha)$$

y

$$\sin(2m\pi + \alpha) = \sin(\alpha),$$

entonces $\theta = \arg z$ puede tomar muchos valores, difiriendo cada dos por múltiplos de 2π . Será conveniente elegir θ de modo que $-2\pi < \theta < 2\pi$.

Dado $z = a + bi$, con $a \neq 0$ y $b \neq 0$, para determinar un argumento de θ de z podemos emplear la función tangente, pues por definición

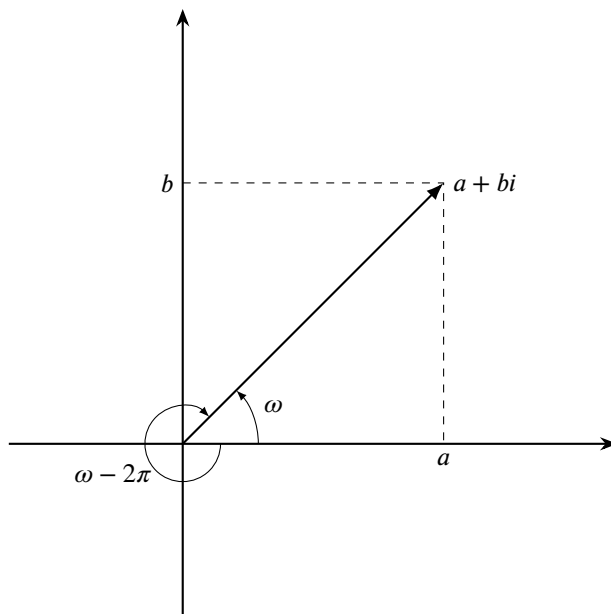
$$\tan(\alpha) = \frac{\sin(\alpha)}{\cos(\alpha)},$$

y las tablas trigonométricas, bajo las siguientes condiciones: Primero determinamos el ángulo agudo ω (positivo) por

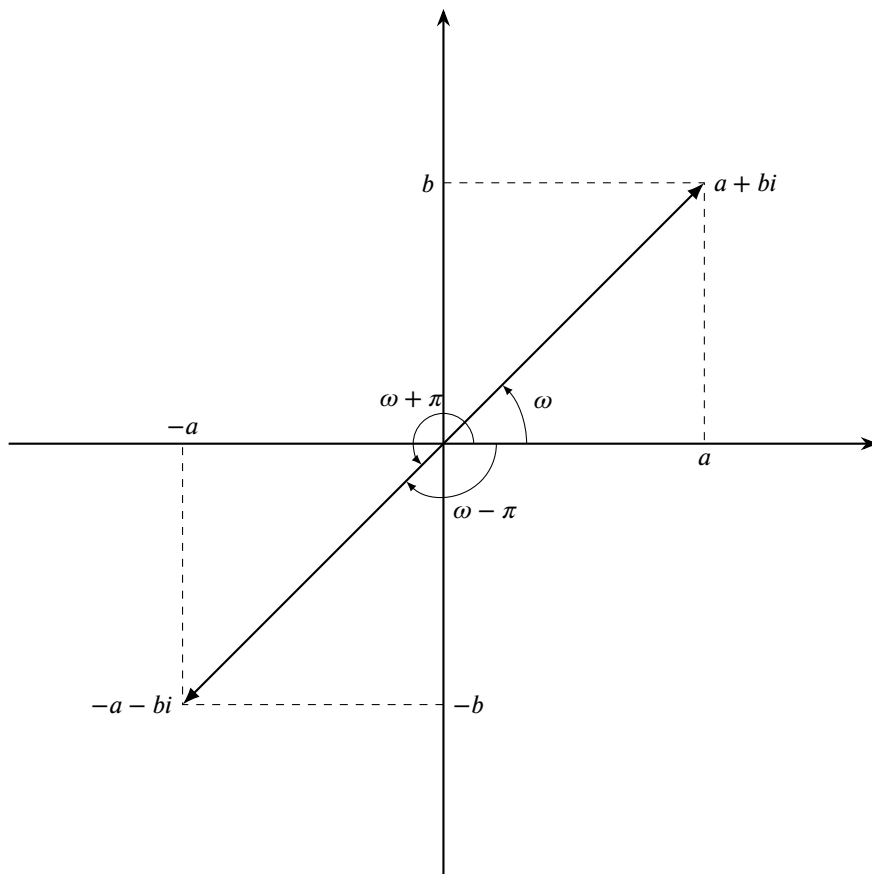
$$\omega = \arctan \frac{|b|}{|a|}$$

y luego

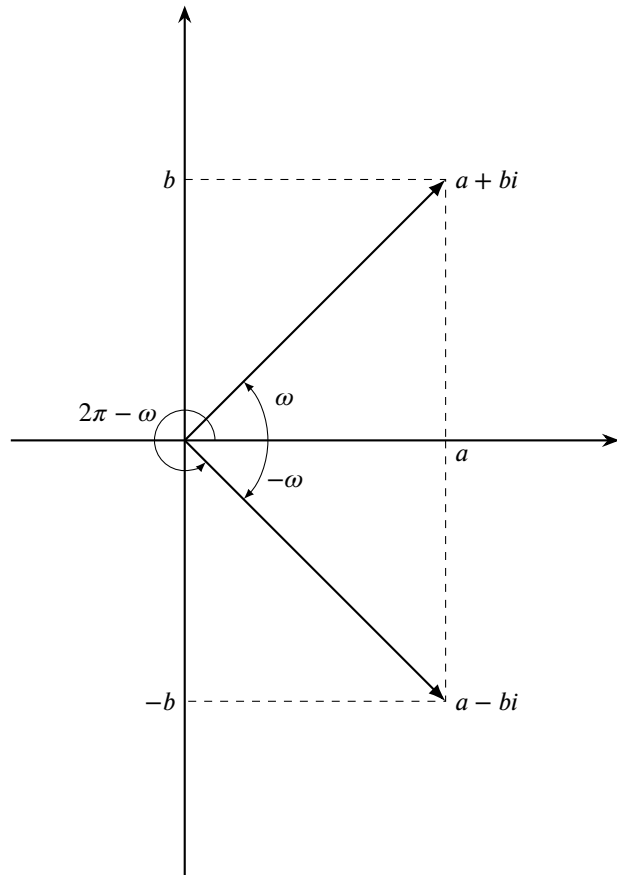
- i) Si $a > 0$ y $b > 0$, elegimos $\theta = \omega > 0$ o $\theta = \omega - 2\pi < 0$.



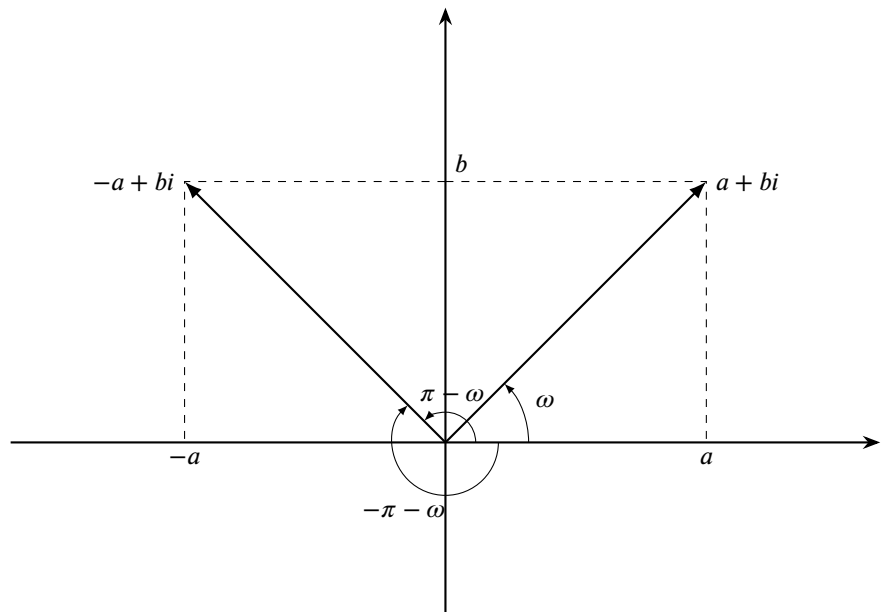
- ii) Si $a < 0$ y $b < 0$, elegimos $\theta = \omega + \pi > 0$ o $\theta = \omega - \pi < 0$.



iii) Si $a > 0$ y $b < 0$, elegimos $\theta = 2\pi - \omega > 0$ o $\theta = -\omega < 0$.



iv) Si $a < 0$ y $b > 0$, elegimos $\theta = \pi - \omega > 0$ o $\theta = -\pi - \omega < 0$.



Sean z, w dos números complejos tal que en su forma polar estén dados por $z = r(\cos \theta + i \operatorname{sen} \theta)$ con $r = |z|$ y $w = s(\cos \varphi + i \operatorname{sen} \varphi)$ con $s = |w|$.

- Si $0 \leq \theta, \varphi < 2\pi$, entonces $z = w$ si y solo si $r = s$ y $\theta = \varphi$.
- Si θ, φ no tienen restricción alguna, entonces $z = w$ si y solo si $r = s$ y $\theta = \varphi + 2m\pi$ con $m \in \mathbb{Z}$.

Ejemplo B.2.1:Expresar a $z = 1 - \sqrt{3}i$ en forma polar.**Solución:** El módulo de z es

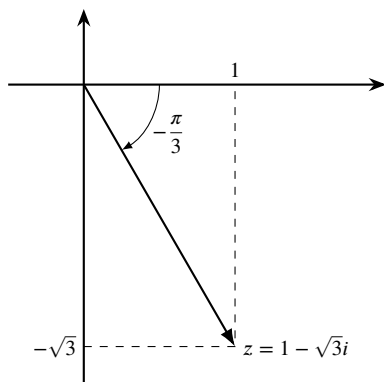
$$|z| = \sqrt{1^2 + (-\sqrt{3})^2} = 2,$$

y el argumento de z es

$$\theta = \arctan \frac{\sqrt{3}}{1} = \frac{\pi}{3}.$$

De la figura B.6, se sigue que la forma polar de z es

$$z = 2 \left(\cos \left(-\frac{\pi}{3} \right) + i \sin \left(-\frac{\pi}{3} \right) \right) = 2 \left(\cos \frac{\pi}{3} - i \sin \frac{\pi}{3} \right).$$

**Figura B.6**

Mostraremos ahora cómo las formas polares de los números complejos proporcionan interpretaciones geométricas de la multiplicación y la división. Sean

$$z_1 = r(\cos \phi_1 + i \sin \phi_1) \quad \text{y} \quad z_2 = s(\cos \phi_2 + i \sin \phi_2)$$

con $r = |z_1|$ y $s = |z_2|$, las formas polares de los números complejos distintos de cero z_1 y z_2 . Al multiplicar, obtenemos

$$z_1 z_2 = rs [(\cos \phi_1 \cos \phi_2 - \sin \phi_1 \sin \phi_2) + i(\sin \phi_1 \cos \phi_2 + \cos \phi_1 \sin \phi_2)].$$

Aplicando ahora las identidades trigonométricas

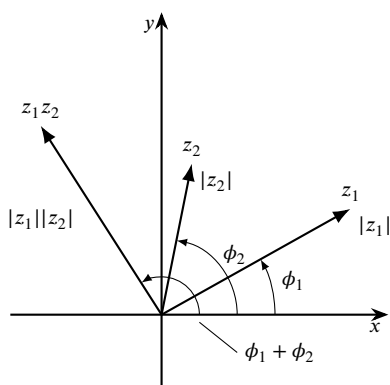
$$\cos(\phi_1 + \phi_2) = \cos \phi_1 \cos \phi_2 - \sin \phi_1 \sin \phi_2$$

$$\sin(\phi_1 + \phi_2) = \sin \phi_1 \cos \phi_2 + \cos \phi_1 \sin \phi_2$$

se obtiene

$$z_1 z_2 = |z_1| |z_2| [\cos(\phi_1 + \phi_2) + i \sin(\phi_1 + \phi_2)] \quad (\text{B.2.1})$$

que es una forma polar del número complejo cuyo módulo es $|z_1| |z_2|$ y cuyo argumento es $\phi_1 + \phi_2$. Así, hemos demostrado que *multiplicar dos números complejos tiene el efecto geométrico de multiplicar sus módulos y sumar sus argumentos* (ver figura B.7).

**Figura B.7**

Cálculos similares muestran que

$$\frac{z_1}{z_2} = \frac{|z_1|}{|z_2|} [\cos(\phi_1 - \phi_2) + i \sin(\phi_1 - \phi_2)] \quad (\text{B.2.2})$$

lo cual nos indica que *dividir números complejos tiene el efecto geométrico de dividir sus módulos y restar sus argumentos* (cada uno en el orden correspondiente).

Ejemplo B.2.2:

Consideremos los números complejos

$$z_1 = 1 + \sqrt{3}i \quad \text{y} \quad z_2 = \sqrt{3} + i,$$

y las formas polares de estos números complejos,

$$z_1 = 2 \left(\cos \frac{\pi}{3} + i \sin \frac{\pi}{3} \right) \quad \text{y} \quad z_2 = 2 \left(\cos \frac{\pi}{6} + i \sin \frac{\pi}{6} \right).$$

Así, se sigue de (B.2.1) que

$$z_1 z_2 = 4 \left[\cos \left(\frac{\pi}{3} + \frac{\pi}{6} \right) + i \sin \left(\frac{\pi}{3} + \frac{\pi}{6} \right) \right] = 4 \left[\cos \left(\frac{\pi}{2} \right) + i \sin \left(\frac{\pi}{2} \right) \right] = 4i$$

y de (B.2.2) que

$$\frac{z_1}{z_2} = 1 \left[\cos \left(\frac{\pi}{3} - \frac{\pi}{6} \right) + i \sin \left(\frac{\pi}{3} - \frac{\pi}{6} \right) \right] = \cos \left(\frac{\pi}{6} \right) + i \sin \left(\frac{\pi}{6} \right) = \frac{\sqrt{3}}{2} + \frac{1}{2}i.$$

Como verificación, tenemos que

$$z_1 z_2 = (1 + \sqrt{3}i)(\sqrt{3} + i) = \sqrt{3} + i + 3i + \sqrt{3}i^2 = 4i$$

y

$$\frac{z_1}{z_2} = \frac{1 + \sqrt{3}i}{\sqrt{3} + i} = \frac{1 + \sqrt{3}i}{\sqrt{3} + i} \cdot \frac{\sqrt{3} - i}{\sqrt{3} - i} = \frac{(1 + \sqrt{3}i)(\sqrt{3} - i)}{3 - i^2} = \frac{\sqrt{3}}{2} + \frac{1}{2}i$$

lo que concuerda con los resultados obtenidos usando formas polares.

B.3 Fórmula de De Moivre y de Euler

Si n es un entero positivo y si z es un número complejo no nulo con forma polar

$$z = |z|(\cos \phi + i \operatorname{sen} \phi)$$

entonces elevar z a la n -ésima potencia da

$$z^n = z z \cdots z = |z|^n [\cos(\phi + \phi + \cdots + \phi) + i \operatorname{sen}(\phi + \phi + \cdots + \phi)]$$

donde hay n factores y n términos, lo que podemos escribir más concisamente como

$$z^n = |z|^n (\cos n\phi + i \operatorname{sen} n\phi).$$

En el caso especial donde $|z| = 1$, esta fórmula se simplifica a

$$z^n = \cos n\phi + i \operatorname{sen} n\phi$$

De hecho, usando la forma polar para z , se convierte en

$$(\cos \phi + i \operatorname{sen} \phi)^n = \cos n\phi + i \operatorname{sen} n\phi. \quad (\text{B.3.1})$$

Este resultado se llama la *fórmula de De Moivre*, nombrada por el matemático francés Abraham de Moivre (1667-1754).

Si θ es un número real, digamos la medida en radianes de algún ángulo, entonces la función exponencial compleja $e^{i\theta}$ se define como

$$e^{i\theta} = \cos \theta + i \operatorname{sen} \theta \quad (\text{B.3.2})$$

lo que a veces se llama la *fórmula de Euler*, nombrada por el matemático suizo Leonhard Euler (1707-1783). Una motivación para esta fórmula proviene de las series de Maclaurin en cálculo. Los lectores que han estudiado series infinitas en cálculo pueden deducir (B.3.2) formalmente sustituyendo $i\theta$ por x en la serie de Maclaurin para e^x y escribiendo

$$\begin{aligned} e^{i\theta} &= 1 + i\theta + \frac{(i\theta)^2}{2!} + \frac{(i\theta)^3}{3!} + \frac{(i\theta)^4}{4!} + \frac{(i\theta)^5}{5!} + \frac{(i\theta)^6}{6!} + \cdots \\ &= 1 + i\theta - \frac{\theta^2}{2!} - i\frac{\theta^3}{3!} + \frac{\theta^4}{4!} + i\frac{\theta^5}{5!} - \frac{\theta^6}{6!} + \cdots \\ &= \left(1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \frac{\theta^6}{6!} + \cdots\right) + i\left(\theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \cdots\right) \\ &= \cos \theta + i \operatorname{sen} \theta \end{aligned}$$

donde el último paso sigue de las series de Maclaurin para $\cos \theta$ y $\operatorname{sen} \theta$.

Si $z = a + bi$ es cualquier número complejo, entonces la exponencial compleja e^z se define como

$$e^z = e^{a+bi} = e^a e^{bi} = e^a (\cos b + i \operatorname{sen} b).$$

Se puede probar que las exponenciales complejas satisfacen las leyes estándar de los exponentes. Así, por ejemplo,

$$e^{z_1} e^{z_2} = e^{z_1+z_2}, \quad \frac{e^{z_1}}{e^{z_2}} = e^{z_1-z_2}, \quad \frac{1}{e^z} = e^{-z}.$$

B.4 Ejercicios del Apéndice B

De los problemas 1 al 10 realice las operaciones indicadas.

1. $(2 - 3i) + (7 - 4i)$

2. $3(4 + i) - 5(-3 + 6i)$

3. $5i(2 + 3i) + 4(6 + 2i)$

4. $(1 + i)(1 - i)$

5. $(2 - 3i)(4 + 7i)$

6. $(6 + 7i)(3 - 7i)$

7. $(-3 + 2i)(7 + 3i)$

8. i^3

9. i^4

10. i^{2025}

11. Sea i la unidad imaginaria. Demuestra que

$$i^n = \begin{cases} 1 & \text{si } n \equiv 0 \pmod{4}, \\ i & \text{si } n \equiv 1 \pmod{4}, \\ -1 & \text{si } n \equiv 2 \pmod{4}, \\ -i & \text{si } n \equiv 3 \pmod{4}. \end{cases}$$

12. Sean $z_1, z_2, z_3 \in \mathbb{C}$. Demuestre las siguientes propiedades:

a) $(-z_1)(z_2) = z_1(-z_2) = -(z_1 z_2)$.

b) $(-z_1)(-z_2) = z_1 z_2$.

c) $z_1(z_2 - z_3) = z_1 z_2 - z_1 z_3$.

De los problemas 13 al 20 escriba a z en forma binómica.

13. $z = 3 - 7i - 8 - 2i$

14. $\frac{1+i}{1-i} - \frac{2-i}{1+i}$

15. $z = \frac{3-2i}{-5+i}$

16. $z = \frac{1+i}{i} - \frac{i}{1-i}$

17. $z = \frac{1}{1+i + \frac{i}{1+i + \frac{i}{1+i}}}$

18. $z = 5 - 2i - (6 - 9i)i$

19. $z = -\frac{i}{(1+i)(2-i)}$

20. $z = \frac{(4+3i)(2-i)}{7-i} + \left(\frac{1}{2} + \frac{3}{2}i\right)^3$

De los problemas 21 al 26 escriba el conjugado de z .

21. $z = 7 + 9i$

22. $z = 128 + 2i$

23. $z = 12 - 189i$

24. $z = 5 - \frac{4}{5}i$

25. $z = \frac{\pi}{9} + \frac{2}{9}i$

26. $1900 - 2022i$

27. Sea $z \in \mathbb{C}$. Demuestre que se cumplen las siguientes propiedades.

a) $\overline{\overline{z}} = z$.

b) $\overline{z^{-1}} = \overline{z}^{-1}$.

c) $z + \overline{z} = 2 \operatorname{Re}(z)$.

d) $z - \overline{z} = 2i \operatorname{Im}(z)$.

De los problemas 28 al 31 represente a z en el plano complejo.

28. $z = 5 + 8i$

29. $z = \pi + i$

30. $z = \frac{5}{2} + \frac{1}{i}$

31. $z = \sqrt{3} + 8i$

C

Apéndice

EL TEOREMA FUNDAMENTAL DEL ÁLGEBRA

Procedamos a probar uno de los más importantes enunciados, el teorema acerca de que el campo de los números complejos es algebraicamente cerrado. Este teorema tiene aplicación en varios campos de las matemáticas. En particular, subyace a toda la teoría posterior de los operadores lineales.

C.1 El teorema fundamental del álgebra

Hemos de probar que todo polinomio de grado $n \geq 1$, con coeficientes complejos tiene al menos una raíz, en general una raíz compleja. Sea $f(z)$ un polinomio arbitrario con coeficientes complejos. Consideremos que es una función compleja con argumento complejo z . Para tales funciones, como para las funciones reales con argumento real, es posible introducir los conceptos de continuidad, de derivada, etc. No todas estas nociones nos serán igualmente necesarias, pero todas se basan en el uso de la completitud del espacio de los números complejos. Se dice que una función compleja $f(z)$ de una variable compleja z es continua en un punto z_0 , si para cualquier $\varepsilon > 0$ podemos encontrar $\delta > 0$ tal que para cualquier número complejo z que satisfaga la desigualdad

$$|z - z_0| < \delta,$$

tenemos

$$|f(z) - f(z_0)| < \varepsilon.$$

Una función $f(z)$ continua en cada punto de su dominio se llama continua en todas partes o simplemente *continua*.

Lema C.1.1:

Un polinomio $f(z)$ con coeficientes complejos es una función continua.

Demostración: Sea

$$f(z) = a_0 + a_1 z + \cdots + a_n z^n$$

y sea z_0 un número complejo arbitrario pero fijo. Sea $h = z - z_0$. Mostremos que para cualquier número $\varepsilon > 0$ tan pequeño, podemos encontrar $\delta > 0$ tal que $|f(z) - f(z_0)| < \varepsilon$ para $|h| < \delta$. Al desarrollar el polinomio $f(z)$ en factores de $(z - z_0)$, tenemos

$$f(z) = A_0 + A_1(z - z_0) + \cdots + A_n(z - z_0)^n.$$

Puesto que $A_0 = f(z_0)$ y $(z - z_0)$ está designado por h , tenemos

$$f(z_0 + h) - f(z_0) = A_1 h + \cdots + A_n h^n.$$

De lo anterior, se sigue que

$$|f(z_0 + h) - f(z_0)| \leq |A_1||h| + \cdots + |A_n||h|^n = A(|h|) \quad (\text{C.1.1})$$

La función real $A(|h|)$ es un polinomio con coeficientes reales $|A_i|$ respecto a la variable real $|h|$. Por el curso de análisis matemático, sabemos que $A(|h|)$ es una función continua en todo punto y, en particular, cuando $|h| = 0$. Como $A(0) = 0$, dado $\varepsilon > 0$ podemos encontrar $\delta > 0$ tal que

$$|h| < \delta. \quad (\text{C.1.2})$$

Entonces $A(|h|) < \varepsilon$. Tomando en consideración la desigualdad (C.1.1), concluimos que si (C.1.2) se cumple, también se cumplirá la desigualdad

$$|f(z_0 + h) - f(z_0)| < \varepsilon.$$

Corolario C.1.1:

El valor absoluto de un polinomio es una función continua.

Demostración: Esta afirmación es inmediata a partir de la siguiente relación:

$$||f(z)| - |f(z_0)|| \leq |f(z) - f(z_0)|.$$

Corolario C.1.2:

Si una sucesión de números complejos $\{z_k\}$ converge a z_0 , entonces para cualquier polinomio $f(z)$

$$\lim_{k \rightarrow \infty} f(z_k) = f(z_0).$$

Lema C.1.2:

Si un polinomio $f(z)$ de grado $n \geq 1$ no se anula cuando $z = z_0$ entonces siempre existe un número complejo h tal que

$$|f(z_0 + h)| < |f(z_0)|$$

Demostración: Tomemos nuevamente la expansión

$$f(z_0 + h) - f(z_0) = A_1 h + \cdots + A_n h^n. \quad (\text{C.1.3})$$

Sea A_k el primer coeficiente no cero entre A_1, A_2, \dots, A_n . Tomemos

$$h = t \sqrt[k]{-\frac{f(z_0)}{A_k}}. \quad (\text{C.1.4})$$

donde tomemos como raíz k -ésima cualquiera de sus valores y

$$0 \leq t \leq 1. \quad (\text{C.1.5})$$

Sea

$$B_p = A_p \left(\sqrt[k]{-\frac{f(z_0)}{A_k}} \right)^p.$$

Ahora, de (C.1.3), teniendo en cuenta (C.1.4) y (C.1.5), encontramos

$$\begin{aligned} |f(z_0 + h)| &= |f(z_0) - t^k f(z_0) + t^{k+1} B_{k+1} + \cdots + t^n B_n| \\ &\leq |(1 - t^k) f(z_0)| + t^{k+1} |B_{k+1}| + \cdots + t^n |B_n| \\ &= (1 - t^k) |f(z_0)| + t^{k+1} |B_{k+1}| + \cdots + t^n |B_n| \\ &= |f(z_0)| + t^k (-|f(z_0)| + t |B_{k+1}| + \cdots + t^{n-k} |B_n|) \\ &= |f(z_0)| + t^k B(t) \end{aligned}$$

Finalmente tenemos que

$$|f(z_0 + h)| \leq |f(z_0)| + t^k B(t).$$

La función $B(t)$ es un polinomio con coeficientes reales y variable independiente t . Es una función continua, pero $B(0) = -|f(z_0)| < 0$ y por lo tanto, en virtud de la continuidad de $B(t)$, existe t_0 dentro de $0 < t_0 \leq 1$ tal que $B(t_0)$ también sea negativo. Para un número complejo h definido por un número t_0 , según (C.1.4), obtenemos

$$|f(z_0 + h)| \leq |f(z_0)| + t_0^k B(t_0) < |f(z_0)|.$$

Lema C.1.3:

Para cualquier polinomio $f(z)$ de grado $n \geq 1$ y cualquier sucesión infinitamente grande $\{z_k\}$ de números complejos, existe una relación límite

$$\lim_{k \rightarrow \infty} |f(z_k)| = +\infty.$$

Demostración: Consideremos el polinomio

$$f(z) = a_0 + a_1 z + \cdots + a_n z^n.$$

Para $z \neq 0$, encontremos

$$|f(z)| \geq |a_n| |z|^n \left(1 - \left| \frac{a_0}{a_n} \right| |z|^{-n} - \cdots - \left| \frac{a_{n-1}}{a_n} \right| |z|^{-1} \right) \quad (\text{C.1.6})$$

Desde que $\{z_k\}$ es infinitamente grande,

$$\lim_{k \rightarrow \infty} |z_k| = +\infty.$$

El lado derecho de la relación (C.1.6) es una función de valor real, y es fácil ver que

$$\lim_{|z_k| \rightarrow \infty} \left(1 - \left| \frac{a_0}{a_n} \right| |z_k|^{-n} - \cdots - \left| \frac{a_{n-1}}{a_n} \right| |z_k|^{-1} \right) = 1.$$

Pero para el otro factor de (C.1.6) tenemos

$$\lim_{|z_k| \rightarrow \infty} |a_n| |z_k|^n = +\infty.$$

En consecuencia,

$$\lim_{k \rightarrow \infty} |f(z_k)| = +\infty.$$

Teorema C.1.1: **Teorema Fundamental del Álgebra (TFA):** Cualquier polinomio $f(z)$ de grado $n \geq 1$ con coeficientes complejos tiene al menos una raíz, en general, una raíz compleja.

Demostración: Consideremos el conjunto de todos los posibles valores absolutos de $f(z)$. Como $|f(z)| \geq 0$, ese conjunto está acotado inferiormente, se sabe a partir del análisis matemático que cualquier límite inferior a un conjunto no vacío de números reales tiene un límite inferior máximo. Sea l para el conjunto de valores $|f(z)|$. Esto significa que para cada k natural podemos encontrar un número complejo z_k tal que

$$0 \leq |f(z_k)| - l \leq 2^{-k}.$$

Se sigue que

$$\lim_{k \rightarrow \infty} |f(z_k)| = l. \quad (\text{C.1.7})$$

Suponiendo que $\{z_k\}$ no esté acotado, sería posible elegir un subsucesión infinitamente grande de ella, y de acuerdo con el lema C.1.3, la relación (C.1.7) no podría cumplirse. Por lo tanto, $\{z_k\}$ está acotado. Elijamos una subsucesión $\{z_{k_v}\}$ y supongamos que

$$\lim_{k_v \rightarrow \infty} z_{k_v} = z_0.$$

Por el corolario C.1.1, tenemos

$$|f(z_0)| = \lim_{k_v \rightarrow \infty} |f(z_{k_v})| = l.$$

Si $l = 0$, entonces del lema C.1.2 se sigue que hay un número z_0 tal que $|f(z_0')| < l$. Esto contradice el hecho de que l es el límite inferior mayor de los valores absolutos del polinomio y, por lo tanto, $l = 0$. Así que hemos demostrado que existe un número complejo z_0 , tal que $|f(z_0)| = 0$ o equivalentemente

$$f(z_0) = 0.$$

Esto significa que z_0 es una raíz de $f(z)$.

C.2 Consecuencias del teorema fundamental

Surgen una variedad de consecuencias de la teorema fundamental. Consideremos el más importante de ellos.

Un polinomio $f(z)$ de grado $n \geq 1$ con coeficientes complejos tiene al menos una raíz z_1 . Por lo tanto $f(z)$ tiene una factorización

$$f(z) = (z - z_1)\varphi(z)$$

donde $\varphi(z)$ es un polinomio de grado $n-1$. Los coeficientes de $\varphi(z)$ son nuevamente números complejos. En consecuencia, $\varphi(z)$ tiene una raíz z_2 (si $n \geq 2$) y

$$\varphi(z) = (z - z_2)\psi(z)$$

donde se sigue que

$$f(z) = (z - z_1)(z - z_2)\psi(z)$$

Continuando con el proceso obtenemos una representación del polinomio como producto de factores lineales:

$$f(z) = b(z - z_1)(z - z_2) \cdots (z - z_n)$$

donde b es algún número. Eliminando los paréntesis a la derecha y comparando los coeficientes de las potencias con los coeficientes a_i de $f(z)$ concluimos que $b = a_n$.

Pueden haber números iguales entre z_1, z_2, \dots, z_n . Supongamos por simplicidad que z_1, \dots, z_r son mutuamente distintos y que cada uno de los números z_{r+1}, \dots, z_n es igual a uno de los primeros números. Entonces $f(z)$ se puede escribir de la siguiente manera:

$$f(z) = a_n(z - z_1)^{k_1}(z - z_2)^{k_2} \dots (z - z_r)^{k_r} \quad (\text{C.2.1})$$

donde $z_i \neq z_j$ para $i \neq j$ y

$$k_1 + k_2 + \dots + k_r = n.$$

La factorización (C.2.1) se denomina *factorización canónica de un polinomio $f(z)$* .

La factorización canónica es única para un polinomio $f(z)$ salvo el orden que se disponen los factores. En efecto, supongamos que junto con la factorización (C.2.1) existe otra factorización canónica, por ejemplo

$$f(z) = a_n(z - v_1)^{l_1}(z - v_2)^{l_2} \dots (z - v_r)^{l_m}.$$

Luego

$$(z - z_1)^{k_1}(z - z_2)^{k_2} \dots (z - z_r)^{k_r} = (z - v_1)^{l_1}(z - v_2)^{l_2} \dots (z - v_r)^{l_m} \quad (\text{C.2.2})$$

Observemos que la colección de números z_1, \dots, z_r debe coincidir con la colección de números v_1, \dots, v_m . Si, por ejemplo, z_1 no es igual a ninguno de los números v_1, \dots, v_m , entonces al sustituir $z = z_1$ en (C.2.2) obtenemos cero a la izquierda y un número distinto de cero a la derecha. Entonces, si hay dos factorizaciones canónicas de $f(z)$, entonces (C.2.2) puede ser solo como sigue:

$$(z - z_1)^{k_1}(z - z_2)^{k_2} \dots (z - z_r)^{k_r} = (z - z_1)^{l_1}(z - z_2)^{l_2} \dots (z - z_r)^{l_r}$$

Supongamos, por ejemplo, que $k_1 \neq l_1$ y sea $k_1 > l_1$ por definición. Al dividir ambos miembros de la última ecuación por el mismo divisor $(z - z_1)^{l_1}$, obtenemos

$$(z - z_1)^{k_1 - l_1}(z - z_2)^{k_2} \dots (z - z_r)^{k_r} = (z - z_2)^{l_2} \dots (z - z_r)^{l_r}$$

Sustituyendo $z = z_1$ vemos de nuevo que hay un cero a la izquierda y un número distinto de cero a la derecha. Queda así demostrada la unicidad de la factorización canónica.

Si $k_i = 1$ en la factorización canónica (C.2.1), entonces una raíz z_i se dice que es simple; si $k_i > 1$, entonces se dice que z_i es una raíz múltiple. Un número k_i es la multiplicidad de una raíz z_i . Ahora podemos enunciar una deducción muy importante:

Cualquier polinomio de grado $n \geq 1$ con coeficientes complejos tiene n raíces, si cada una de las raíces se cuenta tantas veces según su multiplicidad.

Un polinomio de grado cero no tiene raíces. El único polinomio que tiene arbitrariamente muchas raíces mutuamente distintas es el polinomio cero. Estos hechos pueden usarse para sacar la siguiente conclusión:

Si dos polinomios $f(z)$ y $g(z)$ de grado no superior a n tienen valores iguales para más de n valores distintos del argumento, entonces todos los coeficientes correspondientes de dichos polinomios son iguales entre sí.

En efecto, el polinomio $f(z) - g(z)$ tiene, por hipótesis, más de n raíces. Pero su grado no excede de n y por lo tanto $f(z) - g(z) = 0$.

Así pues, un polinomio $f(z)$ cuyo grado no es mayor que n está completamente definido por sus valores para cualquier $n + 1$ valores distintos de la variable independiente. Esto permite restablecer el polinomio por medio de sus valores. No es difícil

mostrar una forma explícita de este polinomio *reconstructor*. Si el polinomio $f(z)$ toma para los valores del argumento $\alpha_1, \dots, \alpha_{n+1}$, los valores $f(\alpha_1), \dots, f(\alpha_{n+1})$, entonces

$$f(z) = \sum_{i=1}^{n+1} f(\alpha_i) \frac{(z - \alpha_1) \cdots (z - \alpha_{i-1})(z - \alpha_{i+1}) \cdots (z - \alpha_{n+1})}{(\alpha_i - \alpha_1) \cdots (\alpha_i - \alpha_{i-1})(\alpha_i - \alpha_{i+1}) \cdots (\alpha_i - \alpha_{n+1})}$$

Es claro que el grado del polinomio de la derecha no excede de n , y en los puntos $z = \alpha_i$ asume los valores $f(\alpha_i)$. Un polinomio construido de esta manera se llama *polinomio de interpolación de Lagrange*.

Consideremos ahora un polinomio $f(z)$ de grado n y sean z_1, z_2, \dots, z_n sus raíces repetidas según su multiplicidad. Entonces

$$f(z) = a_n(z - z_1)(z - z_2) \cdots (z - z_n).$$

Multiplicando los paréntesis de la derecha, reuniendo términos similares y comparando los coeficientes resultantes con los de

$$f(z) = a_0 + a_1 z + \cdots + a_n z^n$$

podemos derivar las siguientes ecuaciones:

$$\begin{aligned} \frac{a_{n-1}}{a_n} &= -(z_1 + z_2 + \cdots + z_n) \\ \frac{a_{n-2}}{a_n} &= +(z_1 z_2 + z_1 z_3 + \cdots + z_1 z_n + \cdots + a_{n-1} z_n) \\ \frac{a_{n-3}}{a_n} &= -(z_1 z_2 z_3 + z_1 z_2 z_4 + \cdots + z_{n-2} z_{n-1} z_n) \\ &\vdots \\ \frac{a_1}{a_n} &= (-1)^{n-1} (z_1 z_2 \cdots z_{n-1} + \cdots + z_2 z_3 \cdots z_n) \\ \frac{a_0}{a_n} &= (-1)^n z_1 z_2 \cdots z_n \end{aligned}$$

Estas se llaman fórmulas de Vieta y expresan los coeficientes del polinomio en términos de sus raíces. A la derecha de la ecuación k -ésima está la suma de todos los productos posibles de k raíces tomadas con un signo más o menos según k sea par o impar.

Para mayor discusión necesitaremos algunas consecuencias del teorema fundamental del álgebra, relacionado con polinomios con coeficientes reales. Sea un polinomio

$$f(z) = a_0 + a_1 z + \cdots + a_n z^n$$

con coeficientes reales tienen una raíz compleja (pero no real) v , es decir

$$a_0 + a_1 v + \cdots + a_n v^n = 0$$

La última ecuación no se viola si todos los números se reemplazan en ella por conjugados complejos. Sin embargo, los coeficientes a_0, \dots, a_n y el número 0, al ser reales, no se verá afectado por el reemplazo. Por lo tanto

$$a_0 + a_1 \bar{v} + \cdots + a_n \bar{v}^n = 0$$

es decir, $f(\bar{v}) = 0$.

Así, si un número complejo (pero no real) es una raíz de un polinomio $f(z)$ con coeficientes reales, entonces también lo es el número complejo conjugado \bar{v} . De ello se deduce que $f(z)$ será divisible por un trinomio cuadrático

$$\varphi(z) = (z - v)(z - \bar{v}) = z^2 - (v + \bar{v})z + v\bar{v}$$

con coeficientes reales. Usando este hecho demostramos que v y \bar{v} tienen la misma multiplicidad. Supongamos que tienen multiplicidades k y l respectivamente y que $k > l$, por ejemplo. Entonces $f(z)$ es divisible por el l -ésimo grado del polinomio $\varphi(z)$, es decir

$$f(z) = \varphi_l^l(z) \cdot q(z)$$

El polinomio $q(z)$, como cociente de dos polinomios con coeficientes reales, también tiene coeficientes reales. Por hipótesis, el número v debe ser la raíz de multiplicidad $(k - l)$ y no debe tener una raíz igual a \bar{v} . Según lo probado anteriormente es imposible y por lo tanto $k = l$. Así todas las raíces complejas de cualquier polinomio con coeficientes reales son complejas conjugadas dos a dos. De la unicidad de la factorización canónica podemos sacar la siguiente conclusión:

Se puede representar cualquier polinomio con coeficientes reales, salvo el orden en que se disponen los factores, de un modo único como producto de su coeficiente principal y polinomios con coeficientes reales.

Dichos polinomios tienen coeficientes principales iguales a la unidad y son lineales, si corresponden a raíces reales, y cuadráticos o si corresponden a un par de raíces complejas conjugadas.

Finalmente, llegamos a la conclusión más importante, aquella para la cual, propiamente hablando, se demostró el teorema fundamental del álgebra. Sea A un operador lineal en un espacio complejo. Los eigenvalores de ese operador, y solo ellos, son las raíces del polinomio característico. Por el teorema fundamental, A tiene al menos un eigenvalor λ . Por consiguiente:

Todo operador lineal en un espacio vectorial complejo tiene al menos un eigenvalor.

Observemos que si A es un operador en un espacio real o racional, esta conclusión ya no es válida.

En relación a los eigenvalores, aplicaremos la misma terminología como en referencia a las raíces de un polinomio. En particular, un eigenvalor se dice que es simple, si es raíz simple de un polinomio característico, y múltiple en caso contrario. La multiplicidad de un eigenvalor λ será la multiplicidad de λ como la raíz del polinomio característico.

D

Apéndice

RESOLUCIÓN DE ECUACIONES BAJO RADICALES

De acuerdo con la versión de diferentes autores, mientras las ecuaciones generales de primer y segundo grados fueron resueltas desde la antigüedad, la ecuación general de tercer grado había resistido todos los esfuerzos de los matemáticos anteriores al italiano Scipio Del Ferro, quien finalmente logró resolverla a principios del siglo XVI, durante el renacimiento en Italia. De acuerdo a la costumbre de su tiempo, Del Ferro no publicó sus resultados, sino que se los comunicó a uno de sus discípulos, mismo que, tras la muerte de aquel, retó al gran matemático Tartaglia (1500-1557), también italiano, a que resolviera un cierto número de ecuaciones de tercer grado. Tartaglia aceptó el reto y, antes del plazo fijado en este, encontró un método para resolver cualquier ecuación cúbica de la forma $x^3 + px + q = 0$. Así como Del Ferro, Tartaglia no publicó su método; pero un profesor de física y matemática de Milán, Cardan (1501-1576), lo convenció de que se lo comunicara, bajo la promesa de mantenerlo en secreto. Cardan violó su promesa y publicó el método de Tartaglia en su trabajo *Ars Magna* (El arte sumo) en 1545. Desde entonces, las fórmulas para resolver una ecuación de tercer grado se conocen como fórmulas de Cardan.

Poco después de la resolución de la ecuación cúbica, el también matemático italiano, alumno de Cardan, Ferrari (1522-1565) resolvió la ecuación general de cuarto grado.

Los procesos que aquí exponemos, para obtener las fórmulas para resolver las

ecuaciones de tercer y cuarto grados, se sustentan en los trabajos de los matemáticos italianos mencionados, y se basan en transformaciones especiales y complicadas de la ecuación respectiva, que bien pueden parecer artificiales y accidentales, pero que ocurrieron en la intensa búsqueda de métodos para resolver dichas ecuaciones.

De la ecuación de primer grado, es decir

$$ax + b = 0$$

solo diremos que su única solución es $x = -\frac{b}{a}$.

D.1 La definición del discriminante

Definición D.1.1:

Consideremos la ecuación de grado $n \geq 2$ y coeficientes complejos dada por

$$a_n x^n + \cdots + a_1 x + a_0 = 0$$

y sean x_1, x_2, \dots, x_n sus raíces (no necesariamente distintas). Definimos el discriminante de esta ecuación como

$$D = a_n^{2n-2} \prod_{1 \leq i < j \leq n} (x_i - x_j)^2.$$

En lugar de la ecuación

$$a_n x^n + \cdots + a_1 x + a_0 = 0$$

de la definición anterior, podemos hablar del polinomio

$$f(x) = a_n x^n + \cdots + a_1 x + a_0.$$

En este caso, se dice que

$$D = a_n^{2n-2} \prod_{1 \leq i < j \leq n} (x_i - x_j)^2$$

es el discriminante del polinomio $f(x)$.

D.2 Ecuación de segundo grado

Consideremos la ecuación general de segundo grado, donde $a, b, c \in \mathbb{C}$ y $a \neq 0$, con la incógnita x :

$$ax^2 + bx + c = 0. \quad (\text{D.2.1})$$

Si r es raíz de (D.2.1), entonces

$$ar^2 + br + c = 0. \quad (\text{D.2.2})$$

Multiplicando ambos miembros de (D.2.2) por $4a$, obtenemos

$$4a^2 r^2 + 4abr + 4ac = 0.$$

Sumando $b^2 - 4ac$ en ambos miembros de (D.2), obtenemos

$$4a^2 r^2 + 4abr + b^2 = b^2 - 4ac$$

es decir,

$$(2ar + b)^2 = b^2 - 4ac$$

por tanto

$$2ar + b = \pm \sqrt{b^2 - 4ac}$$

y por ende,

$$r = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

Nótese que $b^2 - 4ac$ es un número complejo, por lo que tiene dos raíces complejas z y w , donde $w = -z$. Así que $\sqrt{b^2 - 4ac}$ representa a cualquiera, pero solo a una, de z y w .

En resumen, las dos raíces (no necesariamente distintas) de la ecuación

$$ax^2 + bx + c = 0$$

vienen dadas por la fórmula

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

D.3 El discriminante de la ecuación general de segundo grado

Por la sección anterior, sabemos que las raíces de $ax^2 + bx + c = 0$ son

$$x_1 = \frac{-b + \sqrt{b^2 - 4ac}}{2a} \quad \text{y} \quad x_2 = \frac{-b - \sqrt{b^2 - 4ac}}{2a}.$$

Por lo que el discriminante de dicha ecuación es

$$D = a^2(x_1 - x_2)^2,$$

es decir:

$$D = b^2 - 4ac.$$

Observemos que si la ecuación de segundo grado $ax^2 + bx + c = 0$ es de coeficientes reales, entonces:

- i. Tiene dos raíces reales si y solo si $D > 0$.
- ii. Tiene una raíz real doble si y solo si $D = 0$.
- iii. Tienen dos raíces imaginarias diferentes si y solo si $D < 0$.

D.4 La ecuación de tercer grado

Consideremos la ecuación general de tercer grado

$$Ax^3 + Bx^2 + Cx + D = 0 \tag{D.4.1}$$

con $A, B, C, D \in \mathbb{C}$ y $A \neq 0$ y con la incógnita x . Como $A \neq 0$, y puesto que la ecuación

$$\frac{1}{A}(Ax^3 + Bx^2 + Cx + D) = 0$$

tiene las mismas raíces que la ecuación (D.4.1), entonces no se pierde generalidad si en lugar de esta, escribimos

$$x^3 + bx^2 + cx + d = 0, \text{ con } b, c, d \in \mathbb{C} \tag{D.4.2}$$

Sustituyendo $x = y - \frac{b}{3}$ en (D.4.2), se tiene que

$$\left(y - \frac{b}{3}\right)^3 + b\left(y - \frac{b}{3}\right)^2 + c\left(y - \frac{b}{3}\right) + d = 0,$$

por tanto

$$y^3 + \left(c - \frac{b^2}{3}\right)y + \left(d - \frac{bc}{3} + \frac{2b^3}{27}\right) = 0.$$

En consecuencia, resolver la ecuación (D.4.2) se reduce a resolver la ecuación

$$y^3 + py + q = 0$$

donde

$$p = c - \frac{b^2}{3} \quad y \quad q = d - \frac{bc}{3} + \frac{2b^3}{27}.$$

Si y_1, y_2, y_3 son raíces de $y^3 + py + q = 0$, entonces:

$$x_1 = y_1 - \frac{b}{3}, \quad x_2 = y_2 - \frac{b}{3}, \quad x_3 = y_3 - \frac{b}{3}$$

son las raíces de (D.4.2).

Resolveremos ahora la ecuación

$$y^3 + py + q = 0. \quad (\text{D.4.3})$$

Escribiendo $y = u + v$ en (D.4.3), tenemos que

$$(u + v)^3 + p(u + v) + q = 0.$$

Desarrollando la expresión anterior, obtenemos

$$u^3 + v^3 + q + (3uv + p)(u + v) = 0. \quad (\text{D.4.4})$$

Cualquiera que sea el valor numérico de la suma de u y v , siempre podemos determinar a u y v imponiéndoles la condición adicional que su producto uv sea un número prefijado. En efecto: supongamos que $u + v = r$ e impongamos la condición adicional $uv = s$, entonces $v = r - u$ y $uv = s$, por tanto $v = r - u$ y $u(r - u) = s$, por tanto $v = r - u$ y $u^2 - ru + s = 0$. En consecuencia $v = r - u$, y u puede calcularse por la fórmula para resolver una ecuación de segundo grado. Esto demuestra lo que se afirmó.

Enseguida determinaremos a u y v , sabiendo que la suma $u + v$ es raíz de la ecuación $y^3 + py + q = 0$, e imponiendo la condición adicional

$$uv = -\frac{p}{3}. \quad (\text{D.4.5})$$

Sustituyendo (D.4.5) en (D.4.4), tenemos que

$$u^3 + v^3 + q = 0. \quad (\text{D.4.6})$$

De (D.4.5) y (D.4.6), se sigue que

$$u^3 v^3 = -\frac{p^3}{27} \quad (\text{D.4.7})$$

y

$$u^3 + v^3 = -q. \quad (\text{D.4.8})$$

Puesto que

$$(z - u^3)(z - v^3) = z^2 - (u^3 + v^3)z + u^3 v^3$$

entonces por (D.4.7) y (D.4.8), u^3 y v^3 son las dos soluciones de la ecuación de segundo grado

$$z^2 + qz - \frac{p^3}{27} = 0. \quad (\text{D.4.9})$$

Por otro lado, las soluciones de la ecuación (D.4.9), vienen dadas por

$$z_1 = -\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}} \quad y \quad z_2 = -\frac{q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}.$$

Por tanto, podemos escribir

$$u^3 = z_1 \quad y \quad v^3 = z_2. \quad (\text{D.4.10})$$

Nótese que las anteriores ecuaciones son del tipo $x^n = z$, por lo que se pueden resolver mediante el método expuesto en el Apéndice C. Cada una de dichas ecuaciones tiene tres raíces, digamos u_1, u_2, u_3 , para $u^3 = z_1$; y v_1, v_2, v_3 , para $v^3 = z_2$.

Resolviendo la ecuación $x^3 = 1$, tenemos que sus raíces son:

$$w^0 = 1 \quad w = \frac{-1 + \sqrt{3}i}{2} \quad w^2 = \frac{-1 - \sqrt{3}i}{2}$$

Es fácil comprobar que las raíces de

$$u^3 = z_1$$

son:

$$u_1 \quad u_2 = wu_1 \quad u_3 = w^2u_1;$$

y las raíces de

$$v^3 = z_2$$

son:

$$v_1 \quad v_2 = wv_1 \quad v_3 = w^2v_1.$$

No perdamos de vista que estamos determinando valores de u y v , de modo que la suma $u + v$ sea raíz de $y^3 + py + q = 0$, y que además u y v satisfagan la condición adicional $uv = -\frac{p}{3}$. Hemos determinado tres posibles valores u_1, u_2 y u_3 para u ; y tres posibles valores v_1, v_2 y v_3 para v .

Pero observemos que

$$(u_i v_j)^3 = -\frac{p^3}{27}$$

no implica que

$$u_i v_j = -\frac{p}{3},$$

para cada $i, j = 1, 2, 3$. Si elegimos u_1 y v_1 de modo que $u_1 v_1 = -\frac{p}{3}$, y a los que denotaremos por

$$u_1 = \sqrt[3]{\frac{-q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} \quad (\text{D.4.11})$$

y

$$v_1 = \sqrt[3]{\frac{-q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}. \quad (\text{D.4.12})$$

Entonces las raíces de

$$y^3 + py + q = 0$$

son

$$y_1 = u_1 + v_1 \quad y_2 = wu_1 + w^2v_1 \quad y_3 = w^2u_1 + wv_1. \quad (\text{D.4.13})$$

Las expresiones anteriores son conocidas como *fórmulas de Cardano*, para calcular las raíces de la ecuación

$$y^3 + py + q = 0.$$

D.5 El discriminante de la ecuación general de tercer grado

De acuerdo a las fórmulas de Cardano, y a la definición D.1.1, el discriminante de la ecuación $y^3 + py + q = 0$ es

$$D = (y_1 - y_2)^2(y_1 - y_3)^2(y_2 - y_3)^2.$$

Recordemos que

$$w = -\frac{1}{2} + \frac{1}{2}\sqrt{3}i$$

es raíz de la ecuación $x^3 - 1 = 0$, por tanto

$$w^3 = 1 \quad \text{y} \quad w^2 + w + 1 = 0$$

entonces

$$\begin{aligned} y_1 - y_2 &= (u_1 + v_1) - (wu_1 + w^2v_1) \\ &= (1 - w)u_1 - w^2v_1 + w^3v_1 \\ &= (1 - w)u_1 + (1 - w)(-w^2v_1) \\ &= (1 - w)(u_1 - w^2v_1), \end{aligned}$$

$$\begin{aligned} y_1 - y_3 &= (u_1 + v_1) - (w^2u_1 + wv_1) \\ &= (1 - w^2)u_1 - wv_1 + w^3v_1 \\ &= (1 - w^2)u_1 + (1 - w^2)(-wv_1) \\ &= (1 - w^2)(u_1 - wv_1), \end{aligned}$$

$$\begin{aligned} y_2 - y_3 &= (wu_1 + w^2v_1) - (w^2u_1 + wv_1) \\ &= (w - w^2)u_1 - wv_1 + w^2v_1 \\ &= (w - w^2)u_1 + (w - w^2)(-v_1) \\ &= (w - w^2)(u_1 - v_1). \end{aligned}$$

Además

$$\begin{aligned} (1 - w)(1 - w^2) &= \left(\frac{3}{2} - \frac{\sqrt{3}}{2}i\right) \left(\frac{3}{2} + \frac{\sqrt{3}}{2}i\right) \\ &= 3 \end{aligned}$$

y

$$w - w^2 = \sqrt{3}i.$$

Puesto que

$$(x - 1)(x - w)(x - w^2) = x^3 - 1,$$

entonces

$$\left(\frac{u_1}{v_1} - 1\right) \left(\frac{u_1}{v_1} - w\right) \left(\frac{u_1}{v_1} - w^2\right) = \left(\frac{u_1}{v_1}\right)^3 - 1$$

por tanto

$$(u_1 - v_1)(u_1 - wv_1)(u_1 - w^2v_1) = u_1^3 - v_1^3 \quad (\text{D.5.1})$$

Por (D.4.10) de la sección D.4, de (D.5.1) se sigue que

$$(u_1 - v_1)(u_1 - wv_1)(u_1 - w^2v_1) = 2\sqrt{\frac{q^2}{4} + \frac{p^3}{27}}.$$

En consecuencia,

$$\begin{aligned}
 D &= [(y_1 - y_2)(y_1 - y_3)(y_2 - y_3)]^2 \\
 &= [(1 - w)(u_1 - w^2 v_1)(1 - w^2)(u_1 - w v_1)(w - w^2)(u_1 - v_1)]^2 \\
 &= [(1 - w)(1 - w^2)(w - w^2)(u_1 - v_1)(u_1 - w v_1)(u_1 - w^2 v_1)]^2 \\
 &= \left[3(\sqrt{3}i) \left(2\sqrt{\frac{q^2}{4} + \frac{p^3}{27}} \right) \right]^2 \\
 &= -108 \left(\frac{q^2}{4} + \frac{p^3}{27} \right) \\
 &= -27q^2 - 4p^3
 \end{aligned}$$

En resumen, el discriminante de la ecuación

$$y^3 + py + q = 0 \quad (\text{D.5.2})$$

es

$$D = -4p^3 - 27q^2.$$

Si y_1, y_2, y_3 son las raíces de (D.5.2), donde

$$p = c - \frac{b^2}{3}$$

y

$$q = d - \frac{bc}{3} + \frac{2b^3}{27}$$

se sabe que

$$x_1 = y_1 - \frac{b}{3} \quad x_2 = y_2 - \frac{b}{3} \quad x_3 = y_3 - \frac{b}{3}$$

son las raíces de

$$x^3 + bx^2 + cx + d = 0.$$

Por tanto

$$x_1 - x_2 = y_1 - y_2 \quad x_1 - x_3 = y_1 - y_3 \quad x_2 - x_3 = y_2 - y_3.$$

En consecuencia, el discriminante de

$$x^3 + bx^2 + cx + d = 0$$

es

$$\Delta = -4p^3 - 27q^2,$$

es decir,

$$\begin{aligned}
 \Delta &= -4 \left(c - \frac{b^2}{3} \right)^3 - 27 \left(d - \frac{bc}{3} + \frac{2b^3}{27} \right)^2 \\
 &= 18bcd - 4b^3d + b^2c^2 - 4c^3 - 27d^2.
 \end{aligned}$$

Consideremos el escenario en el cual la ecuación cúbica

$$x^3 + bx^2 + cx + d = 0$$

posee *coeficientes reales* (y por tanto, también su ecuación asociada $y^3 + py + q = 0$ tiene coeficientes reales). Dado que esta ecuación es de grado impar, garantiza la existencia de al menos una raíz real. Por lo tanto, se presentan las siguientes tres posibilidades en relación con sus raíces:

- i. Si $\Delta > 0$, entonces tiene una raíz real y dos raíces complejas conjugadas.
- ii. Si $\Delta = 0$, entonces tiene tres raíces reales y al menos dos de ellas son iguales.
- iii. Si $\Delta < 0$, entonces tiene tres raíces reales diferentes.

Ejemplo D.5.1: Resuelva la ecuación

$$y^3 + 3y^2 - 3y - 14 = 0.$$

Solución: Usando la sustitución $y = x - 1$ reducimos esta ecuación a la forma

$$x^3 - 6x - 9 = 0.$$

Notemos que $p = -6$ y $q = -9$, por lo cual

$$\frac{q^2}{4} + \frac{p^3}{27} = \frac{49}{4} > 0.$$

Entonces, la ecuación

$$x^3 - 6x - 9 = 0$$

tiene una raíz real y dos raíces complejas conjugadas. Según las expresiones (D.4.11) y (D.4.12),

$$u_1 = \sqrt[3]{\frac{9}{2} + \frac{7}{2}} = \sqrt[3]{8} \quad y \quad v_1 = \sqrt[3]{\frac{9}{2} - \frac{7}{2}} = \sqrt[3]{1}.$$

Por consiguiente, $u_1 = 2$ y $v_1 = 1$, es decir, $x_1 = 3$. Las otras dos raíces se hallan por las fórmulas (D.4.13):

$$x_2 = -\frac{3}{2} + \frac{\sqrt{3}}{2}i \quad y \quad x_3 = -\frac{3}{2} - \frac{\sqrt{3}}{2}i.$$

De aquí se deduce que las raíces de la ecuación dada son:

$$y_1 = 2, \quad y_2 = -\frac{5}{2} + \frac{\sqrt{3}}{2}i, \quad y_3 = -\frac{5}{2} - \frac{\sqrt{3}}{2}i.$$

Ejemplo D.5.2: Resuelva la ecuación

$$x^3 - 9x^2 + 36x - 80 = 0.$$

Solución: Usando la sustitución $x = y + 3$ reducimos esta ecuación a la forma

$$y^3 + 9y - 26 = 0.$$

Notemos que $p = 9$ y $q = -26$, por lo cual

$$\frac{q^2}{4} + \frac{p^3}{27} = 196 > 0.$$

Entonces, la ecuación

$$y^3 + 9y - 26 = 0$$

tiene una raíz real y dos raíces complejas conjugadas. Según las expresiones (D.4.11) y (D.4.12),

$$u_1 = \sqrt[3]{13 + \sqrt{196}} = \sqrt[3]{27} \quad y \quad v_1 = \sqrt[3]{13 - \sqrt{196}} = \sqrt[3]{-1}.$$

Por consiguiente, $u_1 = 3$ y $v_1 = -1$, es decir, $y_1 = 2$. Las otras dos raíces se hallan por las fórmulas (D.4.13):

$$y_2 = -1 + 2\sqrt{3}i, \quad y_3 = -1 - 2\sqrt{3}i.$$

De aquí se deduce que las raíces de la ecuación dada son:

$$x_1 = 5, \quad x_2 = 2 + 2\sqrt{3}i, \quad x_3 = 2 - 2\sqrt{3}i.$$

Ejemplo D.5.3: Resuelva la ecuación

$$x^3 - 12x + 16 = 0.$$

Solución: Notemos que $p = -12$ y $q = 16$, por lo tanto

$$\frac{q^2}{4} + \frac{p^3}{27} = 0.$$

De esta ecuación, obtenemos $u_1 = \sqrt[3]{-8}$, lo que implica que $u_1 = -2$. En consecuencia, las raíces son $x_1 = -4$ y $x_2 = 2 = x_3$.

D.6 La ecuación de cuarto grado

Consideremos la ecuación general de cuarto grado,

$$Ax^4 + Bx^3 + Cx^2 + Dx + E = 0, \quad (\text{D.6.1})$$

donde $A, B, C, D, E \in \mathbb{C}$ y $A \neq 0$, con la incógnita x .

Como $A \neq 0$, y puesto que la ecuación

$$\frac{1}{A} (Ax^4 + Bx^3 + Cx^2 + Dx + E) = 0$$

tiene las mismas raíces que la ecuación (D.6.1), entonces no se pierde generalidad si en lugar de esta, escribimos

$$x^4 + ax^3 + bx^2 + cx + d = 0 \quad (\text{D.6.2})$$

Si la ecuación (D.6.2) la escribimos en la forma

$$x^4 + ax^3 = -bx^2 - cx - d$$

y sumamos $\frac{a^2}{4}x^2$ a ambos miembros de la ecuación, entonces

$$x^4 + ax^3 + \frac{a^2}{4}x^2 = \frac{a^2}{4}x^2 - bx^2 - cx - d,$$

donde se sigue que

$$\left(x^2 + \frac{a}{2}x\right)^2 = \left(\frac{a^2}{4} - b\right)x^2 - cx - d. \quad (\text{D.6.3})$$

Si el miembro derecho de la ecuación (D.6.3) fuera un cuadrado perfecto, es decir, si fuera de la forma $(ex + f)^2$, entonces resolver dicha expresión y resolver (D.6.2) sería inmediato. Pero en general dicho miembro derecho no es un cuadrado perfecto.

Sumando

$$\left(x^2 + \frac{a}{2}x\right)y + \frac{y^2}{4}$$

a ambos miembros de (D.6.3), tenemos que

$$\left(x^2 + \frac{a}{2}x\right)^2 + \left(x^2 + \frac{a}{2}x\right)y + \frac{y^2}{4} = \left(\frac{a^2}{4} - b\right)x^2 - cx - d + \left(x^2 + \frac{a}{2}x\right)y + \frac{y^2}{4},$$

por tanto

$$\left(x^2 + \frac{a}{2}x + \frac{y}{2}\right)^2 = \left(\frac{a^2}{4} - b + y\right)x^2 + \left(-c + \frac{1}{2}ay\right)x + \left(-d + \frac{1}{4}y^2\right). \quad (\text{D.6.4})$$

Vamos a determinar y de modo que el miembro derecho de la expresión (D.6.4) sea un cuadrado perfecto. Para esto observemos que

$$\alpha x^2 + \beta x + \gamma = (ex + f)^2,$$

si y solo si

$$\beta^2 - 4\alpha\gamma = 0.$$

En efecto

$$\begin{aligned}\alpha x^2 + \beta x + \gamma &= (ex + f)^2 \implies \alpha x^2 + \beta x + \gamma = e^2 x^2 + 2efx + f^2 \\ \implies \alpha &= e^2, \quad \beta = 2ef, \quad \gamma = f^2 \\ \implies \beta^2 - 4\alpha\gamma &= 0.\end{aligned}$$

Recíprocamente, si $\beta^2 - 4\alpha\gamma = 0$, entonces

$$\begin{aligned}\alpha x^2 + \beta x + \gamma &= \alpha \left(x^2 + \frac{\beta}{\alpha}x + \frac{\gamma}{\alpha} \right) \\ &= \alpha \left(x + \frac{\beta}{2\alpha} - \frac{\sqrt{\beta^2 - 4\alpha\gamma}}{2\alpha} \right) \left(x + \frac{\beta}{2\alpha} + \frac{\sqrt{\beta^2 - 4\alpha\gamma}}{2\alpha} \right) \\ &= \alpha \left(x + \frac{\beta}{2\alpha} \right)^2 \\ &= \left(\sqrt{\alpha}x + \frac{\beta}{2\sqrt{\alpha}} \right)^2 \\ &= (ex + f)^2\end{aligned}$$

con $e = \sqrt{\alpha}$ y $f = \frac{\beta}{2\sqrt{\alpha}}$. En consecuencia, el miembro derecho de (D.6.4) será un cuadrado perfecto, es decir,

$$\left(\frac{a^2}{4} - b + y \right) x^2 + \left(-c + \frac{1}{2}ay \right) x + \left(-d + \frac{1}{4}y^2 \right) = (ex + f)^2,$$

si y solo si

$$\left(-c + \frac{1}{2}ay \right)^2 - 4 \left(\frac{a^2}{4} - b + y \right) \left(-d + \frac{1}{4}y^2 \right) = 0,$$

si y solo si

$$c^2 - acy + \frac{1}{4}a^2y^2 - 4 \left(-\frac{a^2d}{4} + \frac{a^2y^2}{16} + bd - \frac{1}{4}by^2 - dy + \frac{1}{4}y^3 \right) = 0,$$

si y solo si

$$-y^3 + by^2 + (4d - ac)y + a^2d - 4bd + c^2 = 0,$$

si y solo si

$$y^3 - by^2 + (ac - 4d)y + (4bd - a^2d - c^2) = 0. \quad (\text{D.6.5})$$

Si y_0 es una solución cualquiera de la ecuación cúbica (D.6.5), la cual es llamada la *resolvente* de la ecuación (D.6.2) de grado cuatro, entonces por (D.6.4) tenemos que:

$$\left(x^2 + \frac{a}{2}x + \frac{y_0}{2} \right)^2 = (ex + f)^2$$

por tanto

$$x^2 + \frac{a}{2}x + \frac{y_0}{2} = ex + f \quad (\text{D.6.6})$$

y

$$x^2 + \frac{a}{2}x + \frac{y_0}{2} = -ex - f. \quad (\text{D.6.7})$$

Las cuatro soluciones de (D.6.6) y (D.6.7) son las raíces de (D.6.2).

Algoritmo para resolver una ecuación de cuarto grado:

- **Paso 1:** Divida toda la ecuación entre A . Esto transforma la ecuación en:

$$x^4 + ax^3 + bx^2 + cx + d = 0.$$

- **Paso 2:** Sustituya y obtenga la expresión:

$$\left(x^2 + \frac{a}{2}x\right)^2 = \left(\frac{a^2}{4} - b\right)x^2 - cx - d.$$

- **Paso 3:** Verifique si el lado derecho de la ecuación es un cuadrado perfecto de la forma $(ex + f)^2$. Si es así, resolver la ecuación principal es inmediato. En caso contrario, forme la ecuación:

$$\left(x^2 + \frac{a}{2}x + \frac{y}{2}\right)^2 = \left(\frac{a^2}{4} - b + y\right)x^2 + \left(-c + \frac{1}{2}ay\right)x + \left(-d + \frac{1}{4}y^2\right).$$

- **Paso 4:** Resuelva la ecuación cúbica auxiliar:

$$y^3 - by^2 + (ac - 4d)y + (4bd - a^2d - c^2) = 0.$$

- **Paso 5:** Si y_0 es una solución cualquiera de la ecuación cúbica auxiliar, basta encontrar las raíces de

$$x^2 + \frac{a}{2}x + \frac{y_0}{2} = ex + f \quad \text{y} \quad x^2 + \frac{a}{2}x + \frac{y_0}{2} = -ex - f$$

que son las soluciones de la ecuación principal, recordando que $e = \sqrt{\alpha}$ y $f = \frac{\beta}{2\sqrt{\alpha}}$ con

$$\alpha = \frac{a^2}{4} - b + y_0 \quad \text{y} \quad \beta = -c + \frac{1}{2}ay_0.$$

Ejemplo D.6.1: Resuelva la ecuación

$$4x^4 + 4x^3 - 12x^2 + 4x + 4 = 0.$$

Solución: Dividiendo entre 4, la ecuación toma la forma

$$x^4 + x^3 - 3x^2 + x + 1 = 0.$$

Entonces, podemos reescribir la ecuación en la forma (D.6.3). En nuestro caso,

$$\left(x^2 + \frac{1}{2}x\right)^2 = 4x^2 - x - 1.$$

Lastimosamente, el lado derecho de la ecuación anterior no es un cuadrado perfecto. Entonces debemos reescribir nuevamente la ecuación a la forma (D.6.4). Así,

$$\left(x^2 + \frac{1}{2}x + \frac{y}{2}\right)^2 = \left(\frac{13}{4} + y\right)x^2 + \left(-1 + \frac{1}{2}y\right)x + \left(-1 + \frac{1}{4}y^2\right).$$

Ahora debemos resolver la ecuación (D.6.5), en nuestro caso, debemos resolver

$$y^3 + 3y^2 - 3y - 14 = 0.$$

Del ejemplo D.5.1, sabemos que una raíz es $y = 2$. Ahora, debemos resolver las ecuaciones (D.6.6) y (D.6.7), es decir,

$$x^2 + \frac{1}{2}x + 1 = \sqrt{\frac{21}{4}}x$$

y

$$x^2 + \frac{1}{2}x + 1 = -\sqrt{\frac{21}{4}}x.$$

Usando la fórmula general, obtenemos las soluciones de las anteriores ecuaciones respectivamente,

$$x_1 = \frac{-1 + \sqrt{21}}{4} + \frac{\sqrt{2\sqrt{21} - 6}}{4}i, \quad x_2 = \frac{-1 + \sqrt{21}}{4} - \frac{\sqrt{2\sqrt{21} - 6}}{4}i$$

y

$$x_3 = \frac{-1 + \sqrt{21} - \sqrt{2\sqrt{21} + 6}}{4}, \quad x_4 = \frac{-1 + \sqrt{21} + \sqrt{2\sqrt{21} + 6}}{4}$$

que son justo, las raíces de la ecuación principal dada.

Ejemplo D.6.2: Resuelva la ecuación

$$4x^4 + 12x^2 + 4i\sqrt{10}x - 3 = 0.$$

Solución: Dividiendo entre 4, la ecuación toma la forma

$$x^4 + 3x^2 + i\sqrt{10}x - \frac{3}{4} = 0.$$

Entonces, podemos reescribir la ecuación en la forma (D.6.3). En nuestro caso,

$$(x^2)^2 = -3x^2 - i\sqrt{10}x + \frac{3}{4}.$$

Lastimosamente, el lado derecho de la ecuación anterior no es un cuadrado perfecto. Entonces debemos reescribir nuevamente la ecuación a la forma (D.6.4). Así,

$$\left(x^2 + \frac{y}{2}\right)^2 = (-3 + y)x^2 - i\sqrt{10}x + \left(\frac{3}{4} + \frac{1}{4}y^2\right).$$

Ahora debemos resolver la ecuación (D.6.5), en nuestro caso, debemos resolver

$$y^3 - 3y^2 + 3y + 1 = 0.$$

Usando la sustitución $y = z + 1$ reducimos esta ecuación a la forma

$$z^3 + 2 = 0.$$

Notemos que $p = 0$ y $q = 2$, por lo cual

$$\frac{q^2}{4} + \frac{p^3}{27} = 1 > 0.$$

Entonces, la ecuación

$$z^3 + 2 = 0$$

tiene una raíz real y dos raíces complejas conjugadas. Según las expresiones (D.4.11) y (D.4.12),

$$u_1 = \sqrt[3]{\frac{-2}{2} + \sqrt{1}} \quad \text{y} \quad v_1 = \sqrt[3]{\frac{-2}{2} - \sqrt{1}}.$$

Por consiguiente, $u_1 = 0$ y $v_1 = -\sqrt[3]{2}$, es decir, $y = 1 - \sqrt[3]{2}$. Ahora, debemos resolver las ecuaciones (D.6.6) y (D.6.7), es decir,

$$x^2 + \frac{1 - \sqrt[3]{2}}{2} = \sqrt{-2 + \sqrt[3]{2}x} - \frac{i\sqrt{10}}{2\sqrt{-2 + \sqrt[3]{2}}}$$

y

$$x^2 + \frac{1 - \sqrt[3]{2}}{2} = -\sqrt{-2 + \sqrt[3]{2}x} + \frac{i\sqrt{10}}{2\sqrt{-2 + \sqrt[3]{2}}}.$$

Usando la fórmula general, obtenemos las soluciones de las anteriores ecuaciones respectivamente,

$$x_{1,2} = \frac{1}{2}i\sqrt{2 + \sqrt[3]{2}} \pm \frac{1}{2}i\sqrt{4 - \sqrt[3]{2} + 2\sqrt{\frac{10}{2 + \sqrt[3]{2}}}}$$

y

$$x_{3,4} = \pm \frac{1}{2}\sqrt{-4 + \sqrt[3]{2} + 2\sqrt{\frac{10}{2 + \sqrt[3]{2}}}} - \frac{1}{2}i\sqrt{2 + \sqrt[3]{2}}$$

que son justo, las raíces de la ecuación principal dada.

D.7 La ecuación de quinto grado

Las fórmulas para la resolución de las ecuaciones de tercero y cuarto grado fueron descubiertas ya en el siglo XVI. Al mismo tiempo comenzaron las búsquedas de fórmulas para la resolución de las ecuaciones de quinto grado y de grados superiores. Señalemos que la forma general de una ecuación de n -ésimo grado es:

$$a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0 = 0$$

Estas búsquedas continuaron sin éxito hasta comienzos del siglo XIX, cuando por fin fue demostrado el siguiente resultado extraordinario:

Para ningún n , mayor o igual a cinco puede hallarse fórmula que exprese las raíces de cualquiera ecuación de n -ésimo grado mediante sus coeficientes por radicales.

Más aún, para cualquier n mayor o igual a cinco se puede indicar una ecuación de n -ésimo grado con coeficientes enteros, cuyas raíces no pueden expresarse mediante radicales. Tal es, por ejemplo, la ecuación

$$x^5 - 4x - 2 = 0.$$

Puede demostrarse que esta ecuación tiene cinco raíces, tres reales y dos complejas, pero ninguna de ellas puede expresarse mediante radicales, es decir, esta ecuación es “irresoluble por radicales”. De este modo la reserva de números, reales o complejos que son raíces de las ecuaciones con coeficientes enteros (estos números se denominan algebraicos en contraposición a los números trascendentes que no son raíces de ninguna ecuación con coeficientes enteros), es mucho más amplia que la reserva de números que se expresan por radicales.

La inexistencia de fórmulas generales para la resolución por radicales de las ecuaciones de n -ésimo grado cuando $n \geq 5$ fue demostrada por Abel (1802-1829). La existencia de ecuaciones con coeficientes enteros irresolubles por radicales fue establecida por Galois (1811-1832), quien también halló las condiciones en las cuales la ecuación puede resolverse por radicales. Todos estos resultados exigieron la creación de una nueva y profunda teoría, la teoría de grupos. El concepto de grupo permitió agotar la cuestión referente a la resolución de ecuaciones por radicales, habiendo hallado más tarde numerosas aplicaciones en diferentes ramas de la matemática y fuera de sus límites, convirtiéndose en uno de los objetos más importantes de estudio en el álgebra.

El hecho de que no existen fórmulas para resolver las ecuaciones de n -ésimo grado cuando $n \geq 5$ no provoca dificultades serias en lo que respecta a la búsqueda práctica de las raíces de las ecuaciones. Esto se compensa totalmente por los numerosos métodos de resolución aproximada de las ecuaciones, que incluso en el caso de las ecuaciones cúbicas conducen al objetivo con mayor rapidez que utilizando la fórmula y extrayendo, a continuación, en forma aproximada los radicales reales. No obstante, la existencia de fórmulas para las ecuaciones de segundo, tercero y cuarto grados permitió demostrar que estas ecuaciones poseen respectivamente dos, tres o cuatro raíces.

En este apéndice, nos hemos referido aquí solamente a las ecuaciones de un cierto grado con una incógnita. El origen de esta teoría se remonta al álgebra elemental, en la que luego del estudio de ecuaciones de primer grado, se pasan al estudio de ecuaciones cuadráticas. Pero en el álgebra elemental también se dio un paso en otra dirección: luego de estudiar una ecuación de primer grado con una incógnita se pasó a considerar el sistema de dos ecuaciones de primer grado con dos incógnitas y el sistema de tres ecuaciones con tres incógnitas. El desarrollo de esta teoría se da en el curso de Álgebra II dado en la ESFM. En el mismo se estudian los métodos de resolución de cualesquiera sistemas de n ecuaciones de primer grado

con n incógnitas, así como también los métodos para hallar la solución de aquellos sistemas de ecuaciones de primer grado en los cuales el número de ecuaciones no es igual al número de incógnitas. La teoría de los sistemas de ecuaciones de primer grado, así como otras, que le son afines, en particular, la teoría de matrices, forman una rama especial del álgebra, el álgebra lineal; de entre todas las partes del álgebra, ésta es la principal debido a sus aplicaciones en geometría y otras ramas de las matemáticas, así como también en física y mecánica teórica.

Por otra parte, tanto la teoría de las ecuaciones algebraicas como el álgebra lineal, en gran medida, pueden ser considerados actualmente como partes acabadas de la ciencia. Las necesidades de ramas contiguas de la matemática y la física condujeron a que en el álgebra pasó a primer plano el estudio de los conjuntos en los cuales están dadas las operaciones algebraicas. Además de la teoría de los campos, dentro de la cual entran la teoría de los números algebraicos y la teoría de las funciones algebraicas, ahora también se desarrolla la teoría de los anillos.