

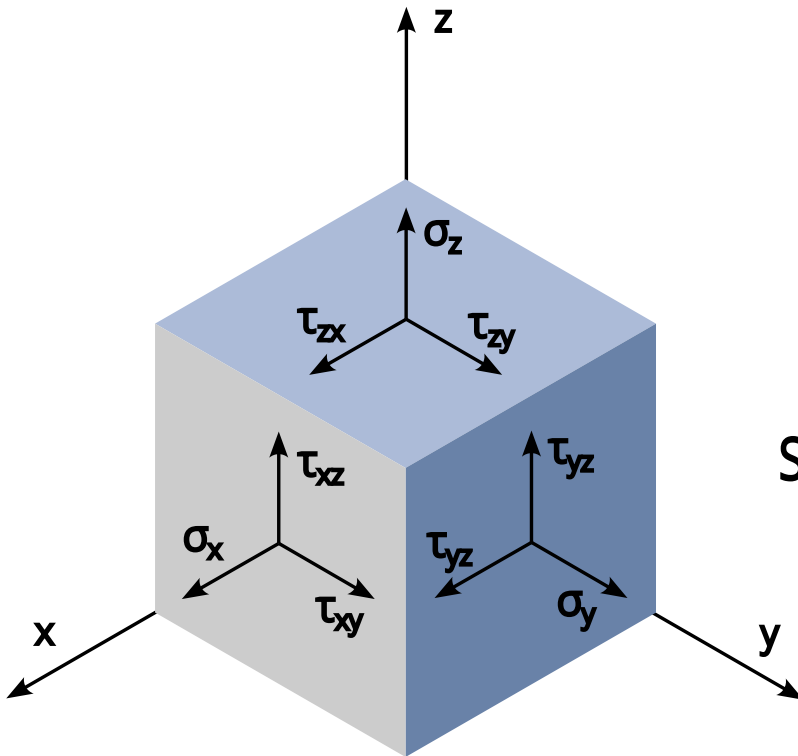


POLITÉCNICA

UNIVERSIDAD
POLITÉCNICA
DE MADRID



ESCUELA TÉCNICA SUPERIOR
DE INGENIEROS DE MINAS
Y ENERGÍA



$$S = \begin{pmatrix} \sigma_x & \tau_{xy} & \tau_{xz} \\ \tau_{yz} & \sigma_y & \tau_{yz} \\ \tau_{zx} & \tau_{zy} & \sigma_z \end{pmatrix}$$

Elementos de álgebra lineal para ingenieros

Alejandro Silva Bernárdez, Ph.D.

Roberto Arranz-Revena, Ph.D.

Esta página está en blanco intencionalmente.

UNIVERSIDAD POLITÉCNICA DE MADRID
ESCUELA TÉCNICA SUPERIOR DE INGENIEROS DE MINAS Y ENERGÍA



POLITÉCNICA

Elementos de álgebra lineal para ingenieros


Alejandro Silva Bernárdez, Ph.D.
Roberto Arranz-Revenga, Ph.D.

DEPARTAMENTO DE INGENIERÍA GEOLÓGICA Y MINERA

AGOSTO | 2025

Esta página está en blanco intencionalmente.

Copyright © 2025 – Alejandro Silva Bernárdez & Roberto Arranz-Revenga.

Open Archives Initiative  La distribución de la obra está bajo los términos de la licencia *Creative Commons* Atribución 4.0 Internacional. Esta licencia permite compartir – copiar y redistribuir el material en cualquier medio o formato, siempre que: se cite debidamente a los autores originales y la fuente, se proporcione un enlace a la licencia *Creative Commons*, se indique si se han realizado cambios. Puede hacerlo de cualquier forma razonable, pero no de forma tal que sugiera que usted o su uso tienen el apoyo de los licenciantes. No puede hacer uso del material con propósitos comerciales. Si remezcla, transforma o crea a partir del material, no podrá distribuir el material modificado. No puede aplicar términos legales ni medidas tecnológicas que restrinjan legalmente a otras a hacer cualquier uso permitido por la licencia. Para más detalles sobre la licencia aplicada, visite: creativecommons.org.

Maquetación

El manuscrito ha sido editado, en formato electrónico, empleando **XeLaTeX**. Se ha usado como base el conjunto de clases y paquetes **KOMA-Script** a través de la clase **kaobook**. Esta clase es de dominio público y ha sido modificada por los autores.

DOI: [10.20868/UPM.book.90391](https://doi.org/10.20868/UPM.book.90391)

Archivo Digital de la UPM: [oa.upm.es:90391](https://oa.upm.es/90391)

Esta página está en blanco intencionalmente.

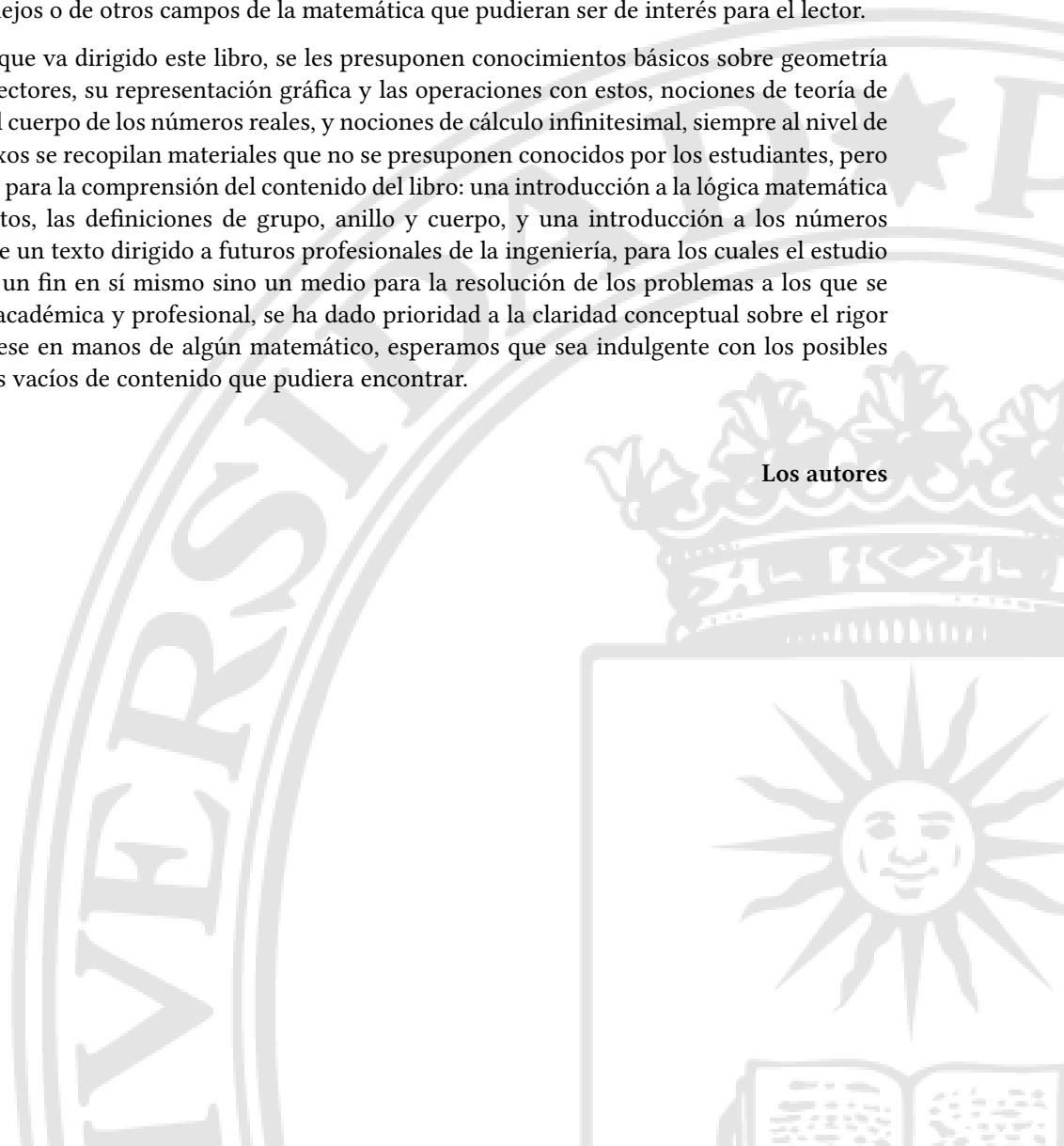
Preámbulo

Este libro nace para dotar a los alumnos del primer año del Grado en Ingeniería Energética de la Universidad Politécnica de Madrid de un texto de referencia para el estudio de la asignatura de Álgebra. Su contenido se dispone y organiza de acuerdo con el temario oficial de la asignatura, salvo por los apartados 5.4.3, 5.6 y 7.2.6 que, a pesar de no formar parte del temario, han sido incluidos por ser de interés para el alumnado, pues en ellos se presentan algunas aplicaciones del álgebra lineal, la estadística y el cálculo numérico, a los que el alumnado habrá de enfrentarse en cursos superiores del grado.

Aunque el libro es fundamentalmente teórico, se han intercalado ejemplos que ilustran cada uno de los conceptos presentados y refuerzan su comprensión y asimilación, todos ellos en espacios de vectores de dos o tres componentes para hacer posible su visualización en el plano o en el espacio *euclidiano*. A diferencia de tantos otros manuales de álgebra lineal dirigidos a estudiantes de ingeniería, que suelen iniciar con un capítulo sobre la geometría de los espacios *euclidianos*, o bien con el estudio y la resolución de los sistemas lineales, este comienza de lleno con los diez axiomas de los espacios vectoriales. Lo hacemos así con la convicción de que la casa debe comenzarse por los cimientos, y no por el tejado. El libro contiene demostraciones de todas las proposiciones que recoge, salvo de las más elementales, que se dejan como ejercicio para el lector, y de los teoremas más complicados, omitidos por requerir para su comprensión un nivel matemático superior al de los nuevos estudiantes. A lo largo del libro, se han añadido observaciones con varias finalidades: aclarar ideas, realizar advertencias o introducir brevemente aplicaciones del álgebra lineal a la ingeniería y conceptos matemáticos más complejos o de otros campos de la matemática que pudieran ser de interés para el lector.

A los estudiantes, a los que va dirigido este libro, se les presuponen conocimientos básicos sobre geometría *euclidiana* y sobre los vectores, su representación gráfica y las operaciones con estos, nociones de teoría de números, en especial del cuerpo de los números reales, y nociones de cálculo infinitesimal, siempre al nivel de Bachillerato. En los anexos se recopilan materiales que no se presuponen conocidos por los estudiantes, pero que son imprescindibles para la comprensión del contenido del libro: una introducción a la lógica matemática y a la teoría de conjuntos, las definiciones de grupo, anillo y cuerpo, y una introducción a los números complejos. Al tratarse de un texto dirigido a futuros profesionales de la ingeniería, para los cuales el estudio de la matemática no es un fin en sí mismo sino un medio para la resolución de los problemas a los que se enfrentarán en su vida académica y profesional, se ha dado prioridad a la claridad conceptual sobre el rigor lógico. Si este libro cayese en manos de algún matemático, esperamos que sea indulgente con los posibles abusos de notación y los vacíos de contenido que pudiera encontrar.

Los autores



Esta página está en blanco intencionalmente.

Índice general

Preámbulo	vii
Índice general	ix
1 Los espacios vectoriales	1
1.1 Definición de espacio vectorial	1
1.1.1 Introducción	1
1.1.2 Definición general de espacio vectorial	1
1.1.3 Definición de vector. Los espacios \mathbb{K}^n	2
1.1.4 Vector nulo	3
1.1.5 Vector opuesto	4
1.1.6 Definición de la suma de vectores	4
1.1.7 Producto de escalar por vector	5
1.1.8 Propiedades adicionales de los espacios vectoriales	7
1.1.9 Otros espacios vectoriales	8
1.2 Clausura, dependencia e independencia lineal. Subespacios vectoriales	11
1.2.1 Introducción	11
1.2.2 Combinaciones lineales	11
1.2.3 Clausura lineal	12
1.2.4 Dependencia e independencia lineal	12
1.2.5 Familias libres y ligadas	14
1.2.6 Subespacios vectoriales	14
1.2.7 Resumen de los apartados anteriores	16
1.3 Bases y dimensión de un subespacio	17
1.3.1 Introducción	17
1.3.2 Sistemas generadores	17
1.3.3 Base	18
1.3.4 Vectores canónicos	18
1.3.5 Bases canónicas	19
1.3.6 Propiedades de las bases	19
1.3.7 Dimensión	21
1.3.8 Propiedades adicionales de las bases	23
1.3.9 Coordenadas respecto de una base	24
1.3.10 Resumen	27
1.4 Intersección y suma de subespacios	28
1.4.1 Introducción	28
1.4.2 Intersección de subespacios	28
1.4.3 Unión de subespacios	30
1.4.4 Suma de subespacios	31
1.4.5 Suma directa de subespacios	33
1.4.6 Subespacios suplementarios	34
1.4.7 La relación de <i>Grassmann</i>	35
1.4.8 Resumen	37

2	Matrices y determinantes	39
2.1	Definición y tipos de matrices	39
2.1.1	Introducción	39
2.1.2	Matrices: definiciones fundamentales	39
2.1.3	Traza y transpuesta de una matriz	42
2.1.4	Opuesta, conjugada y transpuesta conjugada de una matriz	43
2.1.5	Matrices triangulares, trapezoidales y diagonales	44
2.1.6	Simetría y hermiticidad de matrices	45
2.2	Operaciones con matrices	48
2.2.1	Introducción	48
2.2.2	Espacios vectoriales de matrices	48
2.2.3	Producto matricial	50
2.2.4	Propiedades del producto matricial	51
2.2.5	Resumen	54
2.3	Definición y propiedades del determinante	55
2.3.1	Introducción	55
2.3.2	Permutaciones de índices	55
2.3.3	Definición de determinante	56
2.3.4	Propiedades del determinante	57
2.3.5	Resumen	63
2.4	Cálculo de determinantes. Interpretación geométrica	64
2.4.1	Introducción	64
2.4.2	Menores y cofactores	64
2.4.3	Cálculo de determinantes por suma de cofactores	65
2.4.4	Determinantes de matrices triangulares y diagonales	67
2.4.5	Interpretación geométrica del determinante	67
2.4.6	Resumen	70
2.5	Inversa de una matriz cuadrada	71
2.5.1	Introducción	71
2.5.2	Definición de inversa de una matriz	71
2.5.3	Operaciones elementales	74
2.5.4	Algoritmo de <i>Gauss-Jordan</i>	80
2.5.5	Resumen	81
3	Aplicaciones lineales	83
3.1	Definición y propiedades	83
3.1.1	Introducción	83
3.1.2	Aplicaciones	83
3.1.3	Tipos de aplicaciones	85
3.1.4	Aplicaciones lineales	86
3.1.5	Resumen	88
3.2	Aplicación lineal respecto a una base	89
3.2.1	Introducción	89
3.2.2	Aplicación lineal respecto a una base	89
3.2.3	Matriz de una aplicación lineal respecto de las bases canónicas	90
3.2.4	Matriz de una aplicación lineal respecto de dos bases cualesquiera	92
3.2.5	Resumen	94
3.3	Cambio de base. Semejanza y congruencia de matrices	95
3.3.1	Introducción	95
3.3.2	Cambio de base de un vector. Matriz de cambio de base	95
3.3.3	Cambio de base de la matriz de un endomorfismo lineal	97

3.3.4	Matrices semejantes	99
3.3.5	Matrices congruentes	99
3.3.6	Resumen	100
3.4	Imagen, núcleo y rango	101
3.4.1	Introducción	101
3.4.2	Imagen y núcleo de una aplicación	101
3.4.3	Imagen y núcleo de una matriz	102
3.4.4	Subespacios fila y columna de una matriz	104
3.4.5	Imagen y suryectividad	104
3.4.6	Núcleo e inyectividad	105
3.4.7	Rango de una matriz	106
3.4.8	Cálculo del rango mediante reducción <i>gaussiana</i>	109
3.4.9	Rango de una aplicación lineal	112
3.4.10	Teorema del rango-nulidad	112
3.4.11	Resumen	117
3.5	Composición e inversión de aplicaciones lineales	118
3.5.1	Introducción	118
3.5.2	Definición y propiedades de la composición de aplicaciones	118
3.5.3	Composición de aplicaciones lineales	120
3.5.4	Matriz de una composición de aplicaciones lineales	120
3.5.5	Inversa de una aplicación	123
3.5.6	Relación entre la inversa de un isomorfismo y la inversa de una matriz regular	124
3.5.7	Resumen	126
4	Sistemas de ecuaciones lineales	127
4.1	Sistema de ecuaciones lineales	127
4.1.1	Introducción	127
4.1.2	Sistemas de ecuaciones lineales. Definición	127
4.1.3	Tipos de sistemas de ecuaciones lineales	130
4.2	Estructura de las soluciones de un sistema lineal	131
4.2.1	Introducción	131
4.2.2	Estructura de las soluciones	131
4.3	El teorema de <i>Rouché-Frobenius</i>	135
4.3.1	Introducción	135
4.3.2	Teorema de <i>Rouché-Frobenius</i>	135
4.4	Resolución de sistemas por reducción <i>gaussiana</i>	137
4.4.1	Introducción	137
4.4.2	Resolución por reducción <i>gaussiana</i>	137
4.4.3	Resumen de los apartados anteriores	140
5	Producto escalar, norma, ortogonalidad y espacios <i>euclidianos</i>	141
5.1	Producto interno, norma y espacios <i>euclidianos</i>	141
5.1.1	Introducción	141
5.1.2	La norma de un vector en física clásica	141
5.1.3	El producto escalar en física clásica	142
5.1.4	Otra definición para el producto escalar en \mathbb{R}^n	143
5.1.5	Definición de la norma <i>euclidiana</i> en \mathbb{R}^n	144
5.1.6	Equivalencia entre definiciones del producto escalar real	145
5.1.7	Desigualdad de <i>Cauchy-Schwarz</i>	146
5.1.8	Desigualdad triangular	146
5.1.9	Ángulo entre dos vectores reales	147

5.1.10	Definición del producto escalar usual en \mathbb{C}^n	148
5.1.11	Definición de la norma en \mathbb{C}^n	149
5.1.12	Extensión a espacios vectoriales generales. Los espacios <i>euclidianos</i>	149
5.1.13	Definición general de producto interno	150
5.1.14	Definición general de norma. Espacios <i>euclidianos</i>	150
5.1.15	Un ejemplo en un espacio de funciones	151
5.1.16	Resumen	152
5.2	Ortogonalidad	153
5.2.1	Introducción	153
5.2.2	Ortogonalidad entre elementos de un espacio vectorial	153
5.2.3	El teorema de <i>Pitágoras</i>	154
5.2.4	Suplementario ortogonal	156
5.2.5	Familias ortogonales y ortonormales	159
5.2.6	Bases ortogonales y ortonormales	161
5.2.7	Resumen	165
5.3	Matrices unitarias, ortogonales y de giro	166
5.3.1	Introducción	166
5.3.2	Matrices unitarias	166
5.3.3	Matrices ortogonales	168
5.3.4	Matrices de giro	171
5.3.5	Matrices de giro en \mathbb{R}^2	171
5.3.6	Matrices de giro en \mathbb{R}^3	172
5.4	Ortogonalización de <i>Gram-Schmidt</i>	174
5.4.1	Introducción	174
5.4.2	El proceso de <i>Gram-Schmidt</i>	174
5.4.3	La factorización QR	177
5.5	Teorema de la proyección ortogonal. Matrices de proyección ortogonal	179
5.5.1	Introducción	179
5.5.2	Teorema de la proyección ortogonal	179
5.5.3	Cálculo de proyecciones ortogonales	180
5.5.4	Matrices de proyección ortogonal	181
5.5.5	Resumen	186
5.6	El problema de mínimos cuadrados	187
5.6.1	Introducción	187
5.6.2	Soluciones de mínimos cuadrados de un sistema	187
5.6.3	Vector residuo de un sistema lineal incompatible	188
5.6.4	Sistema de ecuaciones normales asociado a un sistema lineal	189
5.6.5	El problema de la regresión lineal	190
5.6.6	Resumen	192
6	Diagonalización de matrices cuadradas	193
6.1	Introducción a la diagonalización	193
6.1.1	Introducción	193
6.1.2	Matrices diagonalizables	193
6.1.3	Interpretación geométrica de la diagonalización	194
6.1.4	Aplicaciones en la ingeniería	195
6.1.5	Resumen	195
6.2	Valores y vectores propios	196
6.2.1	Introducción	196
6.2.2	Valores y vectores propios	196
6.2.3	Espectro y radio espectral	198

6.3	Polinomio característico y subespacios propios de una matriz	199
6.3.1	Introducción	199
6.3.2	Subespacio característico o propio	199
6.3.3	Polinomio característico	201
6.3.4	Resumen de los últimos apartados	204
6.4	Diagonalización de matrices y sus propiedades	205
6.4.1	Introducción	205
6.4.2	Multiplicidad geométrica y algebraica	205
6.4.3	Caracterización de una matriz diagonalizable	207
6.4.4	Propiedades de la diagonalización	208
6.4.5	Polinomios de matrices cuadradas	211
6.4.6	Teorema de <i>Cayley-Hamilton</i>	212
6.4.7	Valores propios de matrices especiales	213
6.4.8	Diagonalización analítica de una matriz cuadrada	214
6.4.9	Resumen	215
6.5	Diagonalización unitaria	216
6.5.1	Introducción	216
6.5.2	Semejanza y diagonalización unitarias	216
6.5.3	Semejanza y diagonalización ortogonales	217
6.5.4	El teorema espectral	217
6.5.5	Descomposición espectral de una matriz real simétrica	220
6.5.6	Descomposición espectral de una matriz normal	223
6.5.7	Resumen	224
7	Formas bilineales y cuadráticas	225
7.1	Definición de forma bilineal y cuadrática	225
7.1.1	Introducción	225
7.1.2	Definición y propiedades de las formas bilineales. Matrices de <i>Gram</i>	225
7.1.3	Cambios de base de una forma bilineal	228
7.1.4	Las formas bilineales en los espacios \mathbb{R}^n	229
7.1.5	Definición de forma cuadrática	230
7.1.6	Acotación de las formas cuadráticas	230
7.2	Cocientes de <i>Rayleigh</i> . Clasificación de matrices reales simétricas	233
7.2.1	Introducción	233
7.2.2	Definición de cociente de <i>Rayleigh</i>	233
7.2.3	Acotación de los cocientes de <i>Rayleigh</i>	233
7.2.4	Clasificación de matrices reales simétricas	234
7.2.5	El criterio de <i>Sylvester</i>	235
7.2.6	El método de la potencia	237
7.2.7	Resumen de los últimos apartados	239
7.3	Clasificación y estudio de las formas cónicas	240
7.3.1	Introducción	240
7.3.2	Las curvas cónicas	240
7.3.3	Clasificación y estudio de cónicas	242
7.3.4	Resumen	245
	APÉNDICES	246
A	Fundamentos de lógica y conjuntos	247
A.1	Elementos de lógica	247

A.2	Teoría básica de conjuntos	253
A.2.1	Relaciones entre conjuntos y elementos	254
A.2.2	Operaciones entre conjuntos	256
A.3	Técnicas fundamentales de demostración	259
A.3.1	Demostración directa	260
A.3.2	Demostración por inducción	261
A.3.3	Demostración por contraposición	262
A.3.4	Demostración por contradicción: reducción al absurdo	263
A.3.5	Demostración constructiva	263
B	Grupos, anillos y cuerpos	265
B.1	Grupos	265
B.2	Anillos	267
B.3	Cuerpos	268
C	El cuerpo de los números complejos	271
C.1	Introducción	271
C.2	Definición de número complejo	272
C.3	Suma y producto de números complejos	273
C.4	El cuerpo de los números complejos	275
C.5	Conjugación de números complejos	276
C.6	Módulo de un número complejo	276
C.7	Inversa y cociente de números complejos	277
C.8	Argumento y forma trigonométrica de un número complejo	278
C.9	Identidad de <i>Euler</i>	280
C.10	Fórmula de <i>De Moivre</i>	281
C.11	Raíces cuadradas y enésimas de un número complejo	282
C.12	Los números complejos en la ingeniería	283
C.13	Resumen	285
	Bibliografía	287

Índice de figuras

1.1	Suma de vectores en el plano euclidiano según la regla del paralelogramo.	4
1.2	Representación geométrica del producto de un vector por un escalar.	6
1.3	Representación gráfica de un subespacio vectorial H generado por los vectores \bar{u}_1 y \bar{u}_2	15
1.4	Representación gráfica del Ejemplo 1.3.12.	26
1.5	Representación gráfica del Ejemplo 1.4.1.	30
1.6	Representación gráfica de la suma directa de los subespacios L Y M	34
2.1	Si se define un paralelogramo contenido en el plano euclidiano mediante dos vectores de \mathbb{R}^2 que representen dos aristas cualesquiera contiguas en un vértice, su área puede calcularse mediante la fórmula (2.163).	68
2.2	Si se define un paralelepípedo contenido en el espacio euclidiano mediante tres vectores de \mathbb{R}^3 que representen tres aristas cualesquiera contiguas en un vértice, su volumen puede calcularse mediante la fórmula (2.164).	69
3.1	La aplicación definida en el Ejemplo 3.1.1.	84
3.2	La aplicación identidad del conjunto E del Ejemplo 3.1.2.	84
3.3	La aplicación inyectiva definida en el Ejemplo 3.1.3.	85
3.4	La aplicación suryectiva definida en el Ejemplo 3.1.4.	86
3.5	La aplicación biyectiva definida en el Ejemplo 3.1.5.	86
3.6	La imagen por la aplicación lineal f del vector $\bar{x} = \bar{u}_1 + \bar{u}_2$, $\bar{u}_1, \bar{u}_2 \in \mathbb{R}^2$ puede descomponerse en la suma de las imágenes por f de \bar{u}_1 y \bar{u}_2 en virtud de la fórmula (3.12).	89
4.1	Todos los puntos de \mathbb{R}^3 que son solución del sistema del Ejemplo 4.2.2 se encuentran contenidos en una recta que es la intersección de los planos $x_1 + 2x_3 = 3$ y $x_2 + x_3 = 2$. Para definirla, basta con conocer un punto cualquiera de esta, que será una solución particular \bar{x}_0 del sistema, y un vector director que determine su dirección (que cumpla las ecuaciones $x_1 + 2x_3 = 0$ y $x_2 + x_3 = 0$, formando una base del núcleo de la matriz de coeficientes del sistema).	134
5.1	Representación gráfica en el plano <i>euclidiano</i> de los vectores del Ejemplo 5.1.1. El ángulo φ relativo entre estos se calculará más adelante, en el Ejemplo 5.1.6. La longitud de su diferencia $\bar{y} - \bar{x}$ determina su distancia <i>euclidiana</i> relativa, definida por la ecuación (5.1.5).	142
5.2	El núcleo de la matriz $A \in \mathbb{R}^{2 \times 3}$, subespacio en \mathbb{R}^3 de dimensión uno, es el suplementario ortogonal de la imagen de su traspuesta. Dos vectores pertenecientes uno a $\ker A$ y otro a $\text{Im } A^t$ son siempre ortogonales: $\langle \lambda_1 \bar{u} + \lambda_2 \bar{v}, \mu \bar{w} \rangle = 0 \forall \lambda_1, \lambda_2, \mu \in \mathbb{R}$	158
5.3	La proyección ortogonal del vector \bar{v} sobre \bar{u}_i (en violeta) es un vector perteneciente a $L[(\bar{u}_i)]$ cuya norma, o longitud, es $\ \bar{v}\ \cdot \cos \varphi$, siendo φ el ángulo entre \bar{v} y \bar{u}_i	162
5.4	Los vectores $\bar{x}, \bar{y}, \bar{z} \in \mathbb{R}^3$ son ortogonales dos a dos, por lo que forman una base ortogonal de \mathbb{R}^3 . Si se normalizan dividiéndolos por su norma, se obtiene una base ortonormal de \mathbb{R}^3	164
5.5	La matriz de giro en \mathbb{R}^2 con $\varphi = \pi/4$ transforma vectores en \mathbb{R}^2 girándolos $\pi/4$ rad en sentido contrario a las agujas del reloj. El eje de giro es perpendicular al plano e intersecta el origen de coordenadas. Se muestra un ejemplo con los vectores canónicos \bar{e}_1 y \bar{e}_2	172
5.6	Al girar un vector cualquiera en \mathbb{R}^3 un ángulo de $ \varphi $ rad alrededor de un eje de giro cuyo vector director es \bar{u} se dibuja un semicono, o un cono si $ \varphi \geq 2\pi$, que tiene su vértice coincidente con el origen de coordenadas y su generatriz coincidente con el vector \bar{x} . El sentido de giro es antihorario respecto a la dirección de \bar{u} , siguiendo la regla de la mano derecha.	172

5.7	Al restarle a \bar{u}_2 su proyección ortogonal sobre \bar{u}_1 se obtiene un vector denotado \tilde{q}_2 que es ortogonal a \bar{u}_1 y que está contenido en el plano que definen \bar{u}_1 y \bar{u}_2 (y por tanto, en el subespacio $L[(\bar{u}_1, \bar{u}_2)]$). Normalizando \bar{u}_1 y \tilde{q}_2 se obtienen dos vectores ortogonales y unitarios cuya clausura lineal coincide con $L[(\bar{u}_1, \bar{u}_2)]$	176
5.8	Representación en el espacio <i>euclidiano</i> del Teorema 5.5.1 de la proyección ortogonal y de su demostración, ecuación (5.213). Dados $\bar{v} \in \mathbb{R}^3$ y un subespacio $M \subseteq \mathbb{R}^3$, la descomposición de \bar{v} en suma de un vector en M denotado \bar{x} más otro vector en M^\perp denotado \bar{y} es única. Al vector \bar{x} se le llama “proyección ortogonal de \bar{v} sobre M ” y es, de todos los vectores contenidos en M , el de menor distancia <i>euclidiana</i> respecto a \bar{v}	180
5.9	Si $\bar{b} \notin \text{Im } A$, el sistema $A\bar{x} = \bar{b}$ es incompatible, Teorema 4.3.1. Pero si \bar{b} se sustituye por su proyección ortogonal sobre $\text{Im } A$, el sistema se vuelve compatible, siendo las soluciones de ese nuevo sistema vectores de \mathbb{K}^n cuya imagen por A se encuentra a una distancia <i>euclidiana</i> mínima de \bar{b}	188
6.1	Los vectores $\bar{v}_1, \bar{v}_2, \bar{v}_3 \in \mathbb{R}^3$ son vectores propios de una cierta matriz real cuadrada A . Sus transformados por A son vectores proporcionales a \bar{v}_1, \bar{v}_2 y \bar{v}_3 con factores de proporcionalidad iguales, respectivamente, a λ_1, λ_2 y λ_3 . Estos coeficientes son los valores propios asociados a cada vector propio. En este ejemplo, el transformado de \bar{v}_3 por A tiene sentido opuesto respecto a \bar{v}_3 . Esto sucede cuando $\lambda_3 < 0$	194
6.2	\bar{v}_1 y \bar{v}_2 son vectores propios de cierta matriz cuadrada asociados al mismo valor propio λ . Cualquier combinación lineal de estos, por ejemplo su suma $\bar{v}_1 + \bar{v}_2$, es también un vector propio de la misma matriz asociado al mismo autovalor.	200
6.3	Vectores pertenecientes a subespacios propios distintos de una misma matriz real simétrica A de orden dos, $\bar{v}_1 \in \ker(\lambda_1 I - A)$, $\bar{v}_2 \in \ker(\lambda_2 I - A)$, $\lambda_1 \neq \lambda_2$, son ortogonales entre sí. Además, la suma directa de ambos subespacios llena todo el espacio \mathbb{R}^2 , lo que permite construir una base ortonormal de \mathbb{R}^2 formada por vectores propios de A y una matriz ortogonal que la diagonalize.	220
6.4	La imagen de $\bar{x} \in \mathbb{R}^3$ por la matriz real simétrica A del Ejemplo 6.5.4 puede escribirse como combinación lineal de las proyecciones ortogonales de \bar{x} sobre los subespacios propios de A , con coeficientes escalares iguales a los valores propios asociados a cada subespacio: $\lambda_1 = -1$ y $\lambda_2 = 2$	222
7.1	Centro, ejes, longitudes de los semiejes y focos (F_1 y F_2) de una elipse. La elipse puede definirse, o bien como la intersección entre un cono y un plano dispuesto con una inclinación menor a la de la generatriz (Figura 7.1a), o bien como el conjunto de todos los puntos del plano euclidiano que cumplen la identidad $l_1 + l_2 = 2a$, donde l_1 y l_2 son las distancias a cada foco de la elipse (Figura 7.1b). La distancia entre cada foco y el centro de la elipse es igual a $\sqrt{a^2 - b^2}$	240
7.2	Centro, ejes, longitudes de los semiejes y focos (F_1 y F_2) de una hipérbola. La hipérbola puede definirse, o bien como la intersección entre un cono y un plano dispuesto verticalmente (Figura 7.2a), o bien como el conjunto de todos los puntos del plano euclidiano que cumplen la identidad $l_1 - l_2 = 2a$, donde l_1 y l_2 son las distancias a cada foco de la hipérbola (Figura 7.2b). La distancia entre cada foco y el centro de la hipérbola es igual a $\sqrt{a^2 + b^2}$	241
7.3	Centro, directriz y foco F de una parábola. La parábola puede definirse, o bien como la intersección entre un cono y un plano dispuesto con una inclinación igual a la de la generatriz (Figura 7.3a) o bien, como el conjunto de todos los puntos del plano euclidiano cuya distancia a una recta llamada “directriz” coincide con su distancia a un punto denominado “foco” (Figura 7.3b). La constante a de la ecuación implícita de una parábola es la mitad de la distancia entre el foco y la directriz.	241
7.4	Al realizar el cambio de variable $(x, y)^t = Q(X, Y)^t$, se sustituyen las coordenadas (x, y) de la cónica (en negro en la figura) por un nuevo sistema de coordenadas (X, Y) (en azul en la figura) que tiene sus ejes de abscisas y ordenadas paralelos a los ejes de simetría de una cónica que, en el ejemplo ilustrado, es una elipse. La matriz A de la cónica en el nuevo sistema de coordenadas será una matriz diagonal.	242

A.1	Diagrama de <i>Venn</i> de los conjuntos A y B y de su unión, $A \cup B$	256
A.2	Diagrama de <i>Venn</i> de los conjuntos A y B y de su intersección, $A \cap B$	257
A.3	Diagrama de <i>Venn</i> de los conjuntos A y B y de su diferencia, $A \setminus B$	258
A.4	Diagrama de <i>Venn</i> del complementario del conjunto A respecto de B , A^c	258
C.1	Representación del número complejo $z = a + bi$	273
C.2	Equivalencia entre el módulo y el argumento del número complejo $z = a + bi$, y la longitud y el ángulo respecto del eje de abscisas de su representación fasorial en \mathbb{R}^2	279
C.3	Las representaciones fasoriales de las tres raíces cúbicas complejas de $z = -1$, denotadas w_1 , w_2 y w_3 . La línea discontinua representa una circunferencia de radio unidad con centro el origen del plano euclidiano. Todos los $z = a + bi \in \mathbb{C}$ que cumplen $ z = 1$ tienen su representación en \mathbb{R}^2 contenida en esta circunferencia.	283

Esta página está en blanco intencionalmente.

Los espacios vectoriales

1.1 Definición de espacio vectorial

1.1.1 Introducción

Comenzaremos este libro definiendo lo que es la base del álgebra lineal: el concepto de espacio vectorial. Se trata de un conjunto de elementos cualesquiera relacionados entre sí mediante dos operaciones que cumplen las mismas propiedades que la suma de vectores y del producto de escalar por vector. Los espacios vectoriales son, por lo tanto, una extensión de las propiedades de los vectores que permite su aplicación en la resolución de problemas que atañen a otros tipos de objetos matemáticos como las matrices, las funciones o los polinomios.

La definición de espacio vectorial es un ejemplo de definición axiomática en la cual no se definen explícitamente los elementos que lo forman ni sus operaciones, sino que se enumeran las propiedades que debe cumplir todo conjunto de elementos y las operaciones que los relacionan para que se pueda denominar un espacio vectorial. De esta forma, cuando se estudien de manera abstracta los espacios vectoriales a partir de esas propiedades axiomáticas, se llegarán a conclusiones que afectan a una multitud de objetos matemáticos distintos, no solo vectores, y que se podrán aplicar para resolver problemas matemáticos de muy diversa naturaleza.

1.1.2 Definición general de espacio vectorial

Definición 1.1.1 — Definición general de espacio vectorial

Sea \mathbb{K} un cuerpo de escalares \dagger , y V un conjunto no vacío dotado de una operación “+” denominada suma (ley interna) y de una operación “·” denominada producto por un escalar (ley externa), se dice que el conjunto V es un espacio vectorial sobre el cuerpo \mathbb{K} si satisface las siguientes diez propiedades:

1. Clausura de la suma de elementos en V :

$$\forall u, v \in V, u + v \in V \quad (1.1)$$

2. Propiedad asociativa de la suma de elementos en V :

$$\forall u, v, w \in V, (u + v) + w = u + (v + w) \quad (1.2)$$

3. Existencia del elemento neutro en V :

$$\exists 0 \in V \text{ tal que } \forall u \in V, u + 0 = 0 + u = u \quad (1.3)$$

- 1.1 Definición de espacio vectorial 1
- 1.2 Clausura, dependencia e independencia lineal. Subespacios vectoriales . 11
- 1.3 Bases y dimensión de un subespacio 17
- 1.4 Intersección y suma de subespacios 28

Observación 1.1.1 \dagger Como, por ejemplo, \mathbb{R} o \mathbb{C} .

4. Existencia del elemento simétrico en V :

$$\forall u \in V, \exists -u \in V \text{ tal que } u + (-u) = -u + u = 0 \quad (1.4)$$

5. Propiedad conmutativa de la suma de elementos de V :

$$\forall u, v \in V, u + v = v + u \quad (1.5)$$

6. Clausura del producto de un elemento de V por un escalar:

$$\forall \lambda \in \mathbb{K}, u \in V, \lambda \cdot u \in V \quad (1.6)$$

7. Propiedad distributiva del producto de un elemento de V por un escalar respecto a la suma de escalares:

$$\forall u \in V, \forall \lambda, \mu \in \mathbb{K}, (\lambda + \mu) \cdot u = \lambda \cdot u + \mu \cdot u \quad (1.7)$$

8. Propiedad distributiva del producto de un elemento de V por un escalar respecto a la suma de elementos de V :

$$\forall u, v \in V, \forall \lambda \in \mathbb{K}, \lambda \cdot (u + v) = \lambda \cdot u + \lambda \cdot v \quad (1.8)$$

9. Propiedad asociativa del producto de un elemento de V por escalares:

$$\forall u \in V, \forall \lambda, \mu \in \mathbb{K}, (\lambda\mu) \cdot u = \lambda \cdot (\mu \cdot u) \quad (1.9)$$

10. Producto de un elemento de V por el escalar unidad:

$$\forall u \in V, 1 \cdot u = u \quad (1.10)$$

De estas diez propiedades axiomáticas emana gran parte del álgebra lineal. De ellas se infieren por deducción nuevos conceptos matemáticos y propiedades aplicables no sólo a vectores (sea cual sea su número de componentes reales o complejas) sino también a matrices, funciones, polinomios, y en general a cualquier conjunto de objetos en los que sea posible definir dos operaciones que cumplan las diez propiedades axiomáticas de la Definición 1.1.1. Gracias a ello, es posible abordar el planteamiento y la resolución de muchos problemas matemáticos de manera geométrica, como si los objetos matemáticos en cuestión fueran vectores en el plano o en el espacio euclidiano. El resto del libro se dedica al estudio y aplicación de todos esos nuevos conceptos que la definición de espacio vectorial engendra, aplicándolos especialmente a los espacios de vectores por brevedad y simplicidad.

Más adelante, en el Tema 5, se definirán de manera axiomática dos nuevas operaciones denominadas “producto interno” y “norma” que permiten extender a los espacios vectoriales generales dos conceptos característicos del plano y del espacio euclidianos: el ángulo entre vectores y la longitud de un vector.

1.1.3 Definición de vector. Los espacios \mathbb{K}^n

Como se indica en el apartado anterior, los espacios vectoriales son una generalización del concepto de vector y de sus operaciones y propiedades.

A continuación, se definirán rigurosamente los vectores, los conjuntos de vectores y las dos operaciones elementales que los relacionan: la suma de vectores y el producto de vector por escalar, operaciones cuyas propiedades elementales coinciden con las diez propiedades axiomáticas de los espacios vectoriales [(1.1) ... (1.10)].

Definición 1.1.2 – Los conjuntos \mathbb{K}^n

Sea n un número entero positivo, se define el conjunto \mathbb{K}^n como el producto cartesiano del cuerpo escalar \mathbb{K} consigo mismo n veces: $\mathbb{K} \times \mathbb{K} \times \dots \times \mathbb{K}$ (n veces). Es decir, como el conjunto de las n -uplas ordenadas de elementos de \mathbb{K} :

$$\mathbb{K}^n = \{(x_1, x_2, \dots, x_n) : x_i \in \mathbb{K}, i = 1, 2, \dots, n\}$$

Los elementos de \mathbb{K}^n se denominan vectores. Los elementos $x_i \in \mathbb{K}$ se denominan componentes del vector.

Dos vectores de n componentes (x_1, x_2, \dots, x_n) y (y_1, y_2, \dots, y_n) son iguales, si y sólo si, $x_1 = y_1, x_2 = y_2, \dots, x_n = y_n$.

Observación 1.1.2 Por lo que respecta a este primer tema, los vectores de un espacio \mathbb{K}^n se pueden representar indistintamente como vectores fila y como vectores columna. Cuando se definen las matrices y en particular el producto matricial, Definiciones 2.1.1 y 2.2.2, se realizará una distinción entre ambos tipos de vectores.

En este texto, los elementos de un espacio vectorial V se denotan mediante letras minúsculas regulares (x, y, z, \dots) para distinguirlos de los números escalares, que se denotan mediante letras minúsculas en cursiva (x, y, z, \dots). Si V es un espacio \mathbb{K}^n , sus elementos se denotarán mediante minúsculas regulares con una barra horizontal para resaltar que se trata de vectores ($\bar{x}, \bar{y}, \bar{z}, \dots$).

Ejemplo 1.1.1 La dupla ordenada de escalares reales $\bar{x} = (2, 3)$ es un vector perteneciente al conjunto de todas las duplas ordenadas, o vectores, de números reales. Conjunto que se denota como \mathbb{R}^2 .

De modo similar, la tripleta ordenada de escalares reales $\bar{y} = (0, 1, -3)$ es un vector perteneciente a \mathbb{R}^3 , mientras que la tripleta ordenada de números complejos $\bar{z} = (-2i, 0, 3 + i)$ pertenece a \mathbb{C}^3 .

Los vectores se definen como n -uplas ordenadas de escalares, por lo que cambian al alterar el orden de los elementos; como en:

$$(2, 3) \neq (3, 2) \text{ y } (-2i, 0, 3 + i) \neq (0, -2i, 3 + i)$$

Observación 1.1.3 Al ser el cuerpo de los números complejos una extensión del cuerpo de los números reales, apartado C.4, la relación $\mathbb{R}^n \subset \mathbb{C}^n$ se cumple para todo $n \geq 1$. Así, en el Ejemplo 1.1.1 se tiene $\bar{x}, \bar{y} \in \mathbb{C}^n$ y $\bar{z} \notin \mathbb{R}^n$.

1.1.4 Definición del vector nulo

Definición 1.1.3 – Vector nulo de \mathbb{K}^n

Considerado el conjunto \mathbb{K}^n . Se llama “vector nulo de \mathbb{K}^n ”, y se denota por $\bar{0}$, al vector $(0, 0, \dots, 0)$. Es decir, al elemento neutro de la suma de vectores.

Ejemplo 1.1.2 La dupla $(0, 0)$ es el vector nulo tanto del conjunto \mathbb{R}^2 como de \mathbb{C}^2 y en ambos casos se denota $\bar{0}$, teniendo presente que el elemento nulo del cuerpo de los números complejos, $0 + 0i$, puede re-

presentarse simplemente como 0. En los conjuntos \mathbb{R}^3 y \mathbb{C}^3 , el vector nulo es la tripleta $\bar{0} = (0, 0, 0)$, mientras que en \mathbb{R}^4 y \mathbb{C}^4 se trata de la 4-upla $\bar{0} = (0, 0, 0, 0)$.

1.1.5 Definición del vector opuesto

Definición 1.1.4 – Vector opuesto en \mathbb{K}^n

Dado cualquier vector $\bar{x} = (x_1, x_2, \dots, x_n) \in \mathbb{K}^n$, se define su opuesto como el vector $-\bar{x} = (-x_1, -x_2, \dots, -x_n)$. Es decir, como su elemento simétrico asociado respecto a la suma de vectores, obtenido cambiando de signo todas sus componentes no nulas.

Ejemplo 1.1.3 El opuesto del vector $\bar{x} = (2, 0) \in \mathbb{R}^2$ es $-\bar{x} = (-2, 0)$. El opuesto de $\bar{y} = (-1, 1, 2) \in \mathbb{R}^3$ es $-\bar{y} = (1, -1, -2)$, y el opuesto de $\bar{z} = (0, i, 3 - 2i) \in \mathbb{C}^3$ es el vector $-\bar{z} = (0, -i, -3 + 2i)$.

1.1.6 Definición y propiedades de la suma de vectores

Definición 1.1.5 – Suma de vectores en \mathbb{K}^n

Dados dos vectores cualesquiera en \mathbb{K}^n con componentes (x_1, \dots, x_n) y (y_1, \dots, y_n) se define su suma como:

$$(x_1, x_2, \dots, x_n) + (y_1, y_2, \dots, y_n) = (x_1 + y_1, x_2 + y_2, \dots, x_n + y_n)$$

El elemento resultante es también un vector de n componentes en el cuerpo \mathbb{K} , por lo que la operación así definida es cerrada.

La suma de vectores en \mathbb{R}^2 y \mathbb{R}^3 puede representarse gráficamente por medio de la regla del paralelogramo, como se muestra en la Figura 1.1.

Observación 1.1.4 Como se muestra en la Figura 1.1, los vectores en \mathbb{R}^2 y \mathbb{R}^3 se representan gráficamente como segmentos dirigidos o flechas. Su punto inicial es coincidente con el origen de coordenadas de un sistema de referencia cartesiano. Su punto final es coincidente con el punto del plano o del espacio cuyas coordenadas cartesianas son iguales a las componentes del vector.

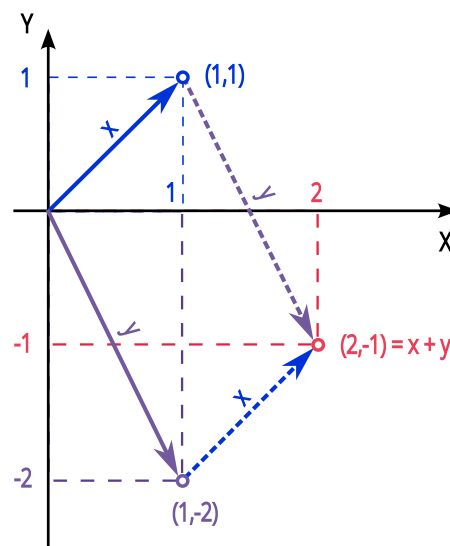


Figura 1.1: La suma de los vectores \bar{x} e \bar{y} en el plano euclidiano \mathbb{R}^2 según la regla del paralelogramo.

Ejemplo 1.1.4 La suma de los vectores $\bar{x} = (1, 1)$, $\bar{y} = (1, -2)$, ambos pertenecientes a \mathbb{R}^2 , es igual al vector $\bar{x} + \bar{y} = (1 + 1, 1 - 2) = (2, -1)$, perteneciente también a \mathbb{R}^2 .

Ejemplo 1.1.5 La suma de los vectores de componentes complejas $\bar{z} = (3, 2i)$, $\bar{w} = (4 + i, 1 - i) \in \mathbb{C}^2$ se denota $\bar{z} + \bar{w}$ y es igual al vector $(3 + 4 + i, 1 + 2i - i) = (7 + i, 1 + i)$, también perteneciente a \mathbb{C}^2 .

Con las últimas definiciones de suma de vectores, vector nulo y vector opuesto, puede comprobarse † que la suma de vectores en \mathbb{K}^n cumple los cinco primeros axiomas de los espacios vectoriales.

Proposición 1.1.1 – Propiedades de la suma de vectores en \mathbb{K}^n

La suma de vectores, Definición 1.1.5, cumple las siguientes propiedades:

1. Propiedad de clausura de la suma de vectores de \mathbb{K}^n :

$$\forall \bar{x}, \bar{y} \in \mathbb{K}^n, \bar{x} + \bar{y} \in \mathbb{K}^n \quad (1.11)$$

2. Propiedad asociativa de la suma de vectores de \mathbb{K}^n :

$$\forall \bar{x}, \bar{y}, \bar{z} \in \mathbb{K}^n, (\bar{x} + \bar{y}) + \bar{z} = \bar{x} + (\bar{y} + \bar{z}) \quad (1.12)$$

3. Existencia del vector neutro en \mathbb{K}^n :

$$\exists \bar{0} \in \mathbb{K}^n \text{ tal que } \forall \bar{x} \in \mathbb{K}^n, \bar{x} + \bar{0} = \bar{0} + \bar{x} = \bar{x} \quad (1.13)$$

4. Existencia del vector simétrico en \mathbb{K}^n :

$$\forall \bar{x} \in \mathbb{K}^n, \exists -\bar{x} \in \mathbb{K}^n \text{ tal que } \bar{x} + (-\bar{x}) = -\bar{x} + \bar{x} = \bar{0} \quad (1.14)$$

5. Propiedad conmutativa de la suma de vectores de \mathbb{K}^n :

$$\forall \bar{x}, \bar{y} \in \mathbb{K}^n, \bar{x} + \bar{y} = \bar{y} + \bar{x} \quad (1.15)$$

Todas ellas coinciden con las propiedades axiomáticas [(1.1) ... (1.5)] de los espacios vectoriales.

Ejercicio 1.1.1 † Demostrar que la suma de vectores en \mathbb{K}^n cumple las cinco primeras propiedades axiomáticas de los espacios vectoriales.

1.1.7 Definición del producto de escalar por vector

Definición 1.1.6 – Producto de escalar por vector en \mathbb{K}^n

Sean el escalar $\alpha \in \mathbb{K}$ y un vector (x_1, x_2, \dots, x_n) perteneciente a \mathbb{K}^n , se define su producto (o ley externa en \mathbb{K}^n) como:

$$\alpha \cdot (x_1, x_2, \dots, x_n) = (\alpha x_1, \alpha x_2, \dots, \alpha x_n)$$

El elemento resultante es también un vector de n componentes en el cuerpo \mathbb{K} , por lo que la operación así definida es cerrada.

Observación 1.1.5 El símbolo “ \cdot ” que denota la ley externa suele omitirse por simplicidad, al igual que en la operación producto de dos escalares.

Observación 1.1.6 No confundir el concepto de producto cartesiano, Definición A.2.8, representado con el símbolo “ \times ”, con el producto o multiplicación de escalares o con la ley externa de un espacio vectorial, representado con el símbolo “ \cdot ”.

Ejemplo 1.1.6 Dados el vector $\bar{x} = (5, -3) \in \mathbb{R}^2$ y el escalar real $\lambda = -2$, su producto es:

$$\lambda \cdot \bar{x} = -2 \cdot (5, -3) = (-2 \cdot 5, -2 \cdot (-3)) = (-10, 6)$$

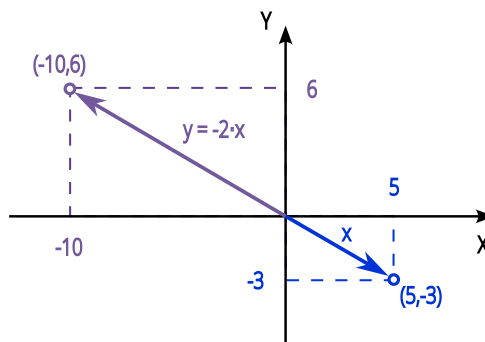


Figura 1.2: Representación geométrica del producto del vector $\bar{x} = (5, -3) \in \mathbb{R}^2$ por el escalar real $\lambda = -2$.

Ejemplo 1.1.7 Dados el vector $\bar{y} = (1, -2i, 1 - i) \in \mathbb{C}^3$ y el escalar complejo $\mu = 1 + i$, su producto es

$$\begin{aligned} \mu \cdot \bar{y} &= (1 + i) \cdot (1, -2i, 1 - i) = ((1 + i) \cdot 1, (1 + i) \cdot (-2i), \dots \\ &\dots (1 + i) \cdot (1 - i)) = (1 + i, 2 - 2i, 2) \end{aligned}$$

Proposición 1.1.2 – Propiedades de la ley externa en \mathbb{K}^n

El producto de escalar por vector, Definición 1.1.6, cumple las siguientes propiedades:

6. Propiedad de clausura del producto de un vector de \mathbb{K}^n por un escalar:

$$\forall \bar{x} \in \mathbb{K}^n \forall \lambda \in \mathbb{K}, \lambda \cdot \bar{x} \in \mathbb{K}^n \quad (1.16)$$

7. Propiedad distributiva del producto de un vector de \mathbb{K}^n por un escalar respecto a la suma de escalares:

$$\forall \bar{x} \in \mathbb{K}^n \forall \lambda, \mu \in \mathbb{K}, (\lambda + \mu) \cdot \bar{x} = \lambda \cdot \bar{x} + \mu \cdot \bar{x} \quad (1.17)$$

8. Propiedad distributiva del producto de un vector de \mathbb{K}^n por un escalar respecto a la suma de vectores de \mathbb{K}^n :

$$\forall \bar{x}, \bar{y} \in \mathbb{K}^n \forall \lambda \in \mathbb{K}, \lambda \cdot (\bar{x} + \bar{y}) = \lambda \cdot \bar{x} + \lambda \cdot \bar{y} \quad (1.18)$$

9. Propiedad asociativa del producto de un vector de \mathbb{K}^n por escalares:

$$\forall \bar{x} \in \mathbb{K}^n \forall \lambda, \mu \in \mathbb{K}, (\lambda\mu) \cdot \bar{x} = \lambda \cdot (\mu \cdot \bar{x}) \quad (1.19)$$

10. Producto de un vector de \mathbb{K}^n por el escalar unidad:

$$\forall \bar{x} \in \mathbb{K}^n, 1 \cdot \bar{x} = \bar{x} \quad (1.20)$$

Todas ellas coinciden con las propiedades axiomáticas [(1.6) ... (1.10)] de los espacios vectoriales.

Las Proposiciones 1.1.1 y 1.1.2 muestran que los conjuntos de vectores \mathbb{R}^n y \mathbb{C}^n , junto con las operaciones de suma de vectores y producto de vector y escalar, cumplen todas las propiedades axiomáticas de los espacios vectoriales [(1.1) ... (1.10)]. En consecuencia, puede decirse que la *cuaterna* $(\mathbb{K}^n, \mathbb{K}, +, \cdot)$ tiene estructura de espacio vectorial para todo cuerpo \mathbb{K} .

Puede comprobarse † que el producto escalar por vector, junto con la suma de vectores en \mathbb{K}^n , cumplen los cinco últimos axiomas de los espacios vectoriales.

1.1.8 Propiedades adicionales de los espacios vectoriales

Mediante simple deducción lógica, de las diez propiedades axiomáticas de los espacios vectoriales, ecuaciones [(1.1) ... (1.10)] se pueden deducir las siguientes:

Proposición 1.1.3 — Otras propiedades de los espacios vectoriales

1. En cada espacio V existe un único elemento neutro.
2. El inverso es único para cada elemento de un espacio V .
3. $\forall \lambda \in \mathbb{K}, \lambda \cdot \mathbf{0} = \mathbf{0}$
4. $\forall \mathbf{x} \in V, \mathbf{0} \cdot \mathbf{x} = \mathbf{0}$
5. $\forall \lambda \in \mathbb{K}, \mathbf{x} \in V, \lambda \cdot \mathbf{x} = \mathbf{0} \Leftrightarrow (\lambda = 0 \vee \mathbf{x} = \mathbf{0})$
6. $\forall \lambda \in \mathbb{K} \forall \mathbf{x} \in V, (-\lambda) \cdot \mathbf{x} = \lambda \cdot (-\mathbf{x}) = -(\lambda \cdot \mathbf{x})$ †
7. $\forall \lambda, \mu \in \mathbb{K} \forall \mathbf{x} \in V, (\lambda \cdot \mathbf{x} = \mu \cdot \mathbf{x} \wedge \mathbf{x} \neq \mathbf{0}) \Rightarrow \lambda = \mu$
8. $\forall \lambda \in \mathbb{K} \forall \mathbf{x}, \mathbf{y} \in V, (\lambda \cdot \mathbf{x} = \lambda \cdot \mathbf{y} \wedge \lambda \neq 0) \Rightarrow \mathbf{x} = \mathbf{y}$

Demostración:

1. Supóngase que un espacio V tiene dos elementos neutros: $\mathbf{0}$ y $\mathbf{0}'$. Aplicando dos veces la propiedad de existencia de elemento neutro, punto 3 de la Definición 1.1.1 de espacio vectorial, se tiene que:

$$\mathbf{0} = \mathbf{0} + \mathbf{0}' = \mathbf{0}'$$

2. Supóngase que un elemento $\mathbf{x} \in V$ tiene dos inversas denotadas $-\mathbf{x}_1, -\mathbf{x}_2$. Aplicando en (1) la propiedad de existencia de elemento neutro, punto 3 de la Definición 1.1.1 de espacio vectorial, en (2) la propiedad de existencia del elemento inverso, punto 4 de la misma definición, y en (3) la propiedad asociativa de los espacios vectoriales, punto 2, se tiene que:

$$\begin{aligned} -\mathbf{x}_1 &\stackrel{(1)}{=} -\mathbf{x}_1 + \mathbf{0} \stackrel{(2)}{=} -\mathbf{x}_1 + (\mathbf{x} + (-\mathbf{x}_2)) \stackrel{(3)}{=} \\ &\stackrel{(3)}{=} (-\mathbf{x}_1 + \mathbf{x}) - \mathbf{x}_2 \stackrel{(2)}{=} \mathbf{0} - \mathbf{x}_2 \stackrel{(1)}{=} -\mathbf{x}_2 \end{aligned}$$

Ejercicio 1.1.2 † Demostrar que el producto de escalar por vector junto con la suma de vectores en \mathbb{K}^n cumplen las cinco últimas propiedades axiomáticas de los espacios vectoriales.

Observación 1.1.7 Cuidado con confundir el escalar nulo $0 \in \mathbb{K}$ con el vector nulo $\mathbf{0} \in \mathbb{K}^n$ o con el elemento nulo de un espacio vectorial genérico V , denotado $\mathbf{0}$.

Observación 1.1.8 † Esta propiedad, junto con la expresión (1.5), permite definir la resta de elementos de un espacio V como $\mathbf{x} - \mathbf{y} = \mathbf{x} + (-1) \cdot \mathbf{y} = \mathbf{x} + (-\mathbf{y}) \forall \mathbf{x}, \mathbf{y} \in V$.

$$3. 0 = \lambda \cdot 0 - \lambda \cdot 0 = \lambda \cdot (0 + 0) - \lambda \cdot 0 = \lambda \cdot 0 + \lambda \cdot 0 - \lambda \cdot 0 = \lambda \cdot 0$$

$$4. 0 = 0 \cdot x - 0 \cdot x = 0 \cdot (x + x) - 0 \cdot x = 0 \cdot x + 0 \cdot x - 0 \cdot x = 0 \cdot x$$

5. El contrapositivo, punto 3 de la Definición A.1.4, de esta proposición es: si λ y x son simultáneamente no nulos, su producto no puede producir en ningún caso el elemento nulo. Para demostrarlo, supóngase que $\lambda \neq 0$ y que $\lambda \cdot x = 0$. Para que la proposición sea verdadera, esto debería implicar $x = 0$; lo que es cierto, pues:

$$\begin{aligned} \lambda^{-1} \cdot x &= \lambda^{-1} \cdot 1 \cdot x = \lambda^{-1}(\lambda^{-1}\lambda) \cdot x = \\ &= \lambda^{-2}(\lambda \cdot x) = \lambda^{-2} \cdot 0 = \lambda^{-1} \cdot 0 \Rightarrow x = 0. \end{aligned}$$

6. Por un lado, se tiene que $(-\lambda) \cdot x = (-\lambda) \cdot x + \lambda \cdot x - \lambda \cdot x = (-\lambda + \lambda) \cdot x - \lambda \cdot x = 0 \cdot x - \lambda \cdot x = -\lambda \cdot x$; y por otro, $\lambda \cdot (-x) = \lambda \cdot (-x) + \lambda \cdot x - \lambda \cdot x = \lambda \cdot (-x + x) - \lambda \cdot x = 0 - \lambda \cdot x = -\lambda \cdot x$, demostrándose así la doble igualdad.

7. $\lambda \cdot x = \mu \cdot x \Rightarrow \lambda \cdot x - \mu \cdot x = 0 \Rightarrow (\lambda - \mu) \cdot x = 0$. Por la propiedad 5 antes demostrada, si $x \neq 0$, $(\lambda - \mu) \cdot x = 0 \Rightarrow \lambda - \mu = 0$, de lo que sigue que $\lambda = \mu$.

8. $\lambda \cdot x = \mu \cdot y \Rightarrow \lambda \cdot x - \lambda \cdot y = 0 \Rightarrow \lambda \cdot (x - y) = 0$. Por el punto 5 antes demostrada, si $\lambda \neq 0$, $\lambda \cdot (x - y) = 0 \Rightarrow x - y = 0$, de lo que sigue que $x = y$. \square

Todos los espacios vectoriales cumplen las propiedades de la Proposición 1.1.3, incluidos naturalmente los espacios de vectores \mathbb{K}^n .

1.1.9 Otros espacios vectoriales

Definir de manera abstracta los espacios vectoriales, enumerando las propiedades axiomáticas que deben cumplir sin especificar qué objetos los componen ni qué operaciones los relacionan, permite construir espacios vectoriales de muy diferente naturaleza con distintos tipos de objetos matemáticos y operaciones, con la única condición de que sus propiedades coincidan con las de la suma de vectores y el producto de escalar por vector. De este modo, se puede abordar la resolución de numerosos problemas matemáticos como si los elementos utilizados en su planteamiento fueran vectores. En este apartado, se definen tres espacios vectoriales compuestos, respectivamente, de matrices, funciones de una variable y polinomios junto con ciertas operaciones cerradas que los relacionan.

Definición 1.1.7 — Espacios Vectoriales de matrices

Sea $\mathbb{K}^{m \times n}$ el conjunto de las matrices de m filas y n columnas con elementos en el cuerpo \mathbb{K} , la suma de dos matrices A y B en $\mathbb{K}^{m \times n}$ se define elemento por elemento como:

$$(A + B)_{ij} = a_{ij} + b_{ij} \quad (1.21)$$

Mientras que la multiplicación de un escalar $\lambda \in \mathbb{K}$ por la matriz $A \in \mathbb{K}^{m \times n}$ se define elemento por elemento como:

$$(\lambda \cdot A)_{ij} = \lambda \cdot a_{ij} \quad (1.22)$$

Puede demostrarse † que la cuaterna $(\mathbb{K}^{m \times n}, \mathbb{K}, +, \cdot)$ tiene estructura de espacio vectorial. Es decir, que la suma de matrices y el producto de escalar por matriz cumplen todas las propiedades axiomáticas de los espacios vectoriales [(1.1) ... (1.10)].

Ejercicio 1.1.3 † Demostrar que la suma de matrices y el producto de escalar por matriz cumplen todas las propiedades axiomáticas de los espacios vectoriales.

Definición 1.1.8 – Espacios vectoriales de funciones

Sea \mathbb{K} un cuerpo de escalares, denótese por $\mathcal{F}(\mathbb{K}, \mathbb{K})$ el conjunto de todas las funciones de una variable en el cuerpo \mathbb{K} . Por ejemplo, $f(x) = x^2 \cdot \cos x \in \mathcal{F}$, donde $x \in \mathbb{R}$.

Dadas dos funciones $f, g \in \mathcal{F}$ y un escalar $\lambda \in \mathbb{K}$, la suma de funciones $f + g$ y el producto de escalar por función λf se definen como:

$$(f + g)(x) = f(x) + g(x) \quad (1.23)$$

$$(\lambda f)(x) = \lambda f(x) \quad (1.24)$$

$\forall x \in \mathbb{R}$. Puede demostrarse ‡ que la cuaterna $(\mathcal{F}(\mathbb{K}, \mathbb{K}), \mathbb{K}, +, \cdot)$ tiene estructura de espacio vectorial. Es decir, que la suma de funciones y el producto de escalar por función cumplen todas las propiedades axiomáticas de los espacios vectoriales [(1.1) ... (1.10)].

Ejercicio 1.1.4 ‡ Demostrar que la suma de funciones de una variable y el producto de escalar por función cumplen todas las propiedades axiomáticas de los espacios vectoriales.

Ejemplo 1.1.8 $f(x) = x^2 \cos x$ y $g(x) = \sqrt{x^2 + x}$ son funciones de una variable real, por lo que $f, g \in \mathcal{F}(\mathbb{R}, \mathbb{R})$. De acuerdo con la definición dada por la ecuación (1.23), su suma es:

$$(f + g)(x) = x^2 \cos x + \sqrt{x^2 + x} \quad (1.25)$$

El producto por f por el escalar real $\lambda = 3$ es, de acuerdo con la definición dada por la ecuación (1.24):

$$\lambda \cdot f(x) = 3x^2 \cos x \quad (1.26)$$

Las funciones resultantes (1.25) y (1.26) son también funciones de una variable real.

Definición 1.1.9 – Espacios vectoriales de polinomios

Denótese por $\mathbb{K}_n[x]$ el conjunto de los polinomios de grado menor o igual que n con coeficientes en el cuerpo \mathbb{K} . Si $P \in \mathbb{K}_n[x]$, entonces:

$$P(x) = a_0 + a_1x + a_2x^2 + \dots + a_nx^n \quad (1.27)$$

Donde $a_i \in \mathbb{K} \forall i = 0 \dots n$ y $n \in \mathbb{N}$. La suma de $P(x)$ con $Q(x) = b_0 + b_1x + b_2x^2 + \dots + b_nx^n \forall i = 0 \dots n$ se define como:

$$P(x) + Q(x) = (a_0 + b_0) + (a_1 + b_1)x + (a_2 + b_2)x^2 + \dots + (a_n + b_n)x^n \quad (1.28)$$

Ejercicio 1.1.5 † Demostrar que la suma de polinomios y el producto de escalar por polinomio cumplen todas las propiedades axiomáticas de los espacios vectoriales.

Mientras que la multiplicación de un escalar $\lambda \in \mathbb{K}$ por $P(x)$ viene dada por:

$$\lambda \cdot P(x) = \lambda \cdot a_0 + \lambda \cdot a_1x + \lambda \cdot a_2x^2 + \cdots + \lambda \cdot a_nx^n \quad (1.29)$$

Puede demostrarse † que la cuaterna $(\mathbb{K}_n[x], \mathbb{K}, +, \cdot)$ tiene estructura de espacio vectorial. Es decir, que la suma de polinomios y el producto de escalar por polinomio cumplen todas las propiedades axiomáticas de los espacios vectoriales [(1.1) ... (1.10)].

1.2 Clausura, dependencia e independencia lineal. Subespacios vectoriales

1.2.1 Introducción

Tras definir los espacios vectoriales y demostrar que los conjuntos de vectores de n componentes \mathbb{R}^n y \mathbb{C}^n con sus operaciones habituales forman espacios vectoriales, se definirán algunos conceptos básicos que se derivan de las propiedades axiomáticas [(1.1) ... (1.10)] de los espacios vectoriales como la combinación lineal de elementos de un espacio, la clausura lineal, la dependencia lineal y la independencia lineal. Más adelante, se introducirá el concepto de subespacio vectorial: un subconjunto de elementos pertenecientes a un espacio vectorial que es cerrado bajo la suma de elementos y el producto de elemento por escalar. Si un subconjunto H de elementos pertenecientes a un espacio vectorial V cumple esta condición, se dirá que H es un subespacio vectorial en V .

1.2.2 Combinaciones lineales

Definición 1.2.1 – Combinación lineal de elementos de un espacio vectorial

Una combinación lineal de los elementos u_1, u_2, \dots, u_k del espacio vectorial V es cualquier elemento en V de la forma:

$$u = \lambda_1 u_1 + \lambda_2 u_2 + \dots + \lambda_k u_k$$

Siendo $\lambda_1, \lambda_2, \dots, \lambda_k$ escalares del cuerpo \mathbb{K} .

El elemento nulo de V es combinación lineal de cualesquiera elementos u_1, \dots, u_k en V . De hecho, $\lambda_1 = \lambda_2 = \lambda_3 = 0 \Rightarrow u = 0$. Además, todo elemento u de un espacio vectorial V es combinación lineal de si mismo, pues si $\lambda = 1 \Rightarrow \lambda u = u$.

Ejemplo 1.2.1 Puede decirse que el vector $(-1/2, 8) \in \mathbb{R}^2$ es combinación lineal de los vectores $\bar{u}_1 = (0, 3)$, $\bar{u}_2 = (2, -1)$ y $\bar{u}_3 = (1/2, -2)$, todos pertenecientes también a \mathbb{R}^2 . De hecho, si $\lambda_1 = 2$, $\lambda_2 = 0$ y $\lambda_3 = -1$, se tiene:

$$\begin{aligned} \lambda_1 \bar{u}_1 + \lambda_2 \bar{u}_2 + \lambda_3 \bar{u}_3 &= \\ &= 2 \cdot (0, 3) + 0 \cdot (2, -1) + (-1) \cdot (1/2, -2) = \quad (1.30) \\ &= (0, 6) + (0, 0) + (-1/2, 2) = (-1/2, 8) \end{aligned}$$

Si $\lambda_1 = \lambda_2 = \lambda_3 = 0$, se obtiene:

$$\lambda_1 \bar{u}_1 + \lambda_2 \bar{u}_2 + \lambda_3 \bar{u}_3 = (0, 0) + (0, 0) + (0, 0) = (0, 0) \quad (1.31)$$

Es decir, el vector nulo de \mathbb{R}^2 .

1.2.3 Clausura lineal

Definición 1.2.2 – Clausura lineal de elementos de un espacio o subespacio

Dado un subconjunto de elementos $M \subset V$, se define la clausura lineal de M , y se denota por $L[M]$, como el conjunto de todas las combinaciones lineales posibles entre los elementos de M .

Ejemplo 1.2.2 Dado el conjunto formado por los vectores de dos componentes complejas $(0, i)$ y $(1 + i, 0)$. Es decir:

$$M = \{(0, i), (1 + i, 0)\} \subset \mathbb{C}^2 \quad (1.32)$$

Se define su clausura lineal como:

$$\begin{aligned} L[M] &= L\{(0, i), (1 + i, 0)\} = \\ &= \{\bar{x} \in \mathbb{C}^2 : \bar{x} = \lambda_1 \cdot (0, i) + \lambda_2 \cdot (1 + i, 0), \lambda_1, \lambda_2 \in \mathbb{C}\} \end{aligned} \quad (1.33)$$

Claramente, $L[M]$ es también un subconjunto del espacio \mathbb{C}^2 .

Si M es el subconjunto vacío de un espacio V , se considera que su clausura lineal produce un conjunto que contiene únicamente el elemento nulo de V . Es decir, $L[\emptyset] = \{\bar{0}\}$. Este es un convenio que la comunidad matemática ha acordado para que sean lógicamente consistentes las demostraciones de las propiedades de las clausuras lineales y de los subespacios vectoriales que se muestran en apartados posteriores.

1.2.4 Dependencia e independencia lineal

Definición 1.2.3 – Definiciones de dependencia e independencia lineal

Se dice que los elementos u_1, u_2, \dots, u_k de un subconjunto $M \subset V$ son linealmente independientes entre sí, si la única combinación lineal de ellos igual al elemento nulo es la que tiene todos los coeficientes escalares nulos. Es decir, si:

$$\lambda_1 u_1 + \lambda_2 u_2 + \dots + \lambda_k u_k = \bar{0} \Leftrightarrow \lambda_i = 0 \forall i = 1, \dots, k \quad (1.34)$$

En caso contrario, se dice que los k elementos son linealmente dependientes entre sí. Es decir, u_1, u_2, \dots, u_k son linealmente dependientes si y sólo si, $\exists \lambda_1, \lambda_2, \dots, \lambda_k \in \mathbb{K}$, con algún $\lambda_j \neq 0$, $j \in \{1, \dots, k\}$, tal que:

$$\sum_{i=1}^k \lambda_i u_i = \bar{0} \quad (1.35)$$

De la ecuación (1.35) puede deducirse la siguiente definición equivalente de dependencia lineal entre k elementos de un espacio V :

Definición 1.2.4 – Elementos linealmente dependientes e independientes

Se dice que los elementos u_1, u_2, \dots, u_k de un subconjunto $M \subset V$ son linealmente dependientes entre sí, si es posible expresar alguno de esos elementos como combinación lineal de los restantes. Es decir, si $\exists u_j, j \in \{1, \dots, k\}$, tal que:

$$u_j = \sum_{i \neq j} \alpha_i u_i, \alpha_i \in \mathbb{K}$$

En tal caso, se dice que u_j es linealmente dependiente de los vectores $u_1, \dots, u_k \in V$. En caso contrario, se dice que u_j es linealmente independiente de los vectores $u_1, \dots, u_k \in V$.

Ejemplo 1.2.3 La combinación lineal de los vectores $\bar{u}_1 = (1, 1)$ y $\bar{u}_2 = (2, 2)$, $\bar{u}_1, \bar{u}_2 \in \mathbb{R}^2$ con los escalares $\lambda_1 = -2$ y $\lambda_2 = 1$ da el vector:

$$\lambda_1 \bar{u}_1 + \lambda_2 \bar{u}_2 = (0, 0) \tag{1.36}$$

Al ser posible obtener el vector nulo de \mathbb{R}^2 combinando linealmente los vectores \bar{u}_1 y \bar{u}_2 con escalares no nulos, puede decirse, aplicando la Definición 1.2.3, que \bar{u}_1 y \bar{u}_2 son linealmente dependientes. También puede decirse, de acuerdo con la Definición 1.2.4, que \bar{u}_1 es dependiente linealmente de \bar{u}_2 , y viceversa, ya que $2 \cdot \bar{u}_1 - \lambda_2 \bar{u}_2 = (0, 0) \Rightarrow \bar{u}_2 = 2 \cdot \bar{u}_1$.

Considérese ahora el vector $\bar{u}_3 = (-1, 1)$. Podría tratarse de encontrar escalares λ_1 y λ_3 no nulos tales que:

$$\lambda_1 \bar{u}_1 + \lambda_3 \bar{u}_3 = (0, 0) \tag{1.37}$$

Sin embargo, esto último ya no sería posible. Obtener el vector nulo a partir de \bar{u}_1 y \bar{u}_3 es posible, si y sólo si, $\lambda_1 = \lambda_3 = 0$ †. Por tanto, \bar{u}_1 y \bar{u}_3 son linealmente independientes entre sí.

Observación 1.2.1 † Esto puede comprobarse aplicando el Teorema 4.3.1 de Rouché-Frobenius al sistema homogéneo de ecuaciones lineales $(\bar{u}_1 | \bar{u}_3)(\lambda_1, \lambda_3)^t = (0, 0)^t$.

Ejemplo 1.2.4 Considérese el conjunto:

$$M = \{\bar{u}_1, \bar{u}_2, \bar{u}_3\} = \{(2, 1), (1, -1), (1, 2)\} \subset \mathbb{R}^2 \tag{1.38}$$

Véase que el conjunto M contiene vectores que se pueden escribir como combinación lineal de los restantes. Concretamente:

$$\bar{u}_1 = \bar{u}_3 + \bar{u}_2 \tag{1.39}$$

$$\bar{u}_2 = \bar{u}_1 - \bar{u}_3 \tag{1.40}$$

$$\bar{u}_3 = \bar{u}_1 - \bar{u}_2 \tag{1.41}$$

Por tanto, \bar{u}_1, \bar{u}_2 y \bar{u}_3 son linealmente dependientes entre sí.

1.2.5 Familias libres y ligadas

Definición 1.2.5 – Definiciones de familia libre y familia ligada

Se dice que el conjunto $M = \{u_1, u_2, \dots, u_k\} \subset V$ es una familia libre si u_1, u_2, \dots, u_k son linealmente independientes. En caso contrario, se dice que M es una familia ligada.

Ejemplo 1.2.5 Como se había mostrado en el Ejemplo 1.2.4, el conjunto:

$$M = \{\bar{u}_1, \bar{u}_2, \bar{u}_3\} = \{(2, 1), (1, -1), (1, 2)\} \subset \mathbb{R}^2 \quad (1.42)$$

contiene vectores que se pueden escribir en función de los restantes y que, por tanto, son linealmente dependientes, por lo que M es una familia ligada.

Por el contrario, los vectores del conjunto:

$$N = \{(1, i), (-1, 1 + i)\} \subset \mathbb{C}^2 \quad (1.43)$$

son linealmente independientes entre sí, ya que:

$$\lambda_1 \cdot (1, i) + \lambda_2 \cdot (-1, 1 + i) = (0, 0) \Leftrightarrow \lambda_1 = \lambda_2 = 0 \quad (1.44)$$

Por tanto, N es una familia libre.

Observación 1.2.2 Por los mismos motivos de consistencia lógica mencionados tras definir la clausura lineal en el apartado 1.2.3, se adopta el convenio de considerar el conjunto vacío, \emptyset , como una familia libre.

1.2.6 Subespacios vectoriales

Definición 1.2.6 – Subespacio vectorial

Se dice que el subconjunto H no vacío de un espacio vectorial V es un subespacio vectorial en V si todas las propiedades axiomáticas de los espacios vectoriales se cumplen en H con las operaciones definidas en V . Es decir, si se cumplen las dos condiciones siguientes:

1. H es cerrado bajo la suma de elementos.

$$\forall u, v \in H, u + v \in H \quad (1.45)$$

2. H es cerrado bajo el producto por un escalar.

$$\forall \lambda \in \mathbb{K}, \forall u \in H, \lambda u \in H \quad (1.46)$$

Nótese que las condiciones (1.45) y (1.46) se pueden escribir conjuntamente como:

$$\forall u, v \in H, \forall \lambda, \mu \in \mathbb{K}, \lambda u + \mu v \in H \quad (1.47)$$

Por lo que la condición necesaria y suficiente que todo subconjunto no vacío $H \subseteq V$ ha de cumplir por definición para ser un subespacio vectorial en V es que sea cerrado bajo combinaciones lineales de sus elementos.

De la condición (1.46) puede deducirse la siguiente condición necesaria para que $H \subseteq V$ sea un subespacio vectorial en V : que contenga el elemento nulo de V . Por tanto, si H no contiene el elemento nulo de V , ya no puede ser un subespacio vectorial en V .

Ejemplo 1.2.6 El conjunto:

$$N = \{(1, i), (-1, 1 + i)\} \subset \mathbb{C}^2 \quad (1.48)$$

no contiene el vector cero. Por tanto, no puede ser un subespacio vectorial. Sin embargo, la clausura lineal de N , $L[N]$ sí que es un subespacio vectorial en \mathbb{C}^2 , ya que:

1. Contiene el vector nulo.

$$\lambda_1 = \lambda_2 = 0 \Rightarrow \lambda_1 \cdot (1, i) + \lambda_2 \cdot (-1, 1 + i) = (0, 0) \quad (1.49)$$

2. $L[N]$ es cerrado para combinaciones lineales de elementos de $L[N]$.

$$\forall \bar{u}_1, \bar{u}_2 \in L[N], \forall \lambda_1, \lambda_2 \in \mathbb{C}, \lambda_1 \bar{u}_1 + \lambda_2 \bar{u}_2 \in L[N] \quad (1.50)$$

Ejemplo 1.2.7 El conjunto:

$$H = \{(x_1, x_2, x_3) \in \mathbb{R}^3 : x_1 - x_2 + x_3 = 1\} \quad (1.51)$$

no contiene el vector cero, pues $\bar{x} = (0, 0, 0) \Rightarrow 0 - 0 + 0 = 0 \neq 1$. Por tanto, no puede ser un subespacio vectorial. Sin embargo:

$$M = \{(x_1, x_2, x_3) \in \mathbb{R}^3 : x_1 - x_2 + x_3 = 0\} \quad (1.52)$$

Sí que es un subespacio vectorial en \mathbb{R}^3 , pues:

1. Contiene el vector nulo:

$$\bar{x} = \bar{0} = (0, 0, 0) \Rightarrow 0 - 0 + 0 = 0 \quad (1.53)$$

2. M es cerrado para combinaciones lineales de elementos de M .

$$\begin{aligned} &(\lambda x_1 + \mu y_1) - (\lambda x_2 + \mu y_2) + (\lambda x_3 + \mu y_3) = \\ &= \lambda \cdot (x_1 - x_2 + x_3) + \mu \cdot (y_1 - y_2 + y_3) = \\ &= 0 + 0 = 0 \end{aligned} \quad (1.54)$$

Para toda pareja de vectores $\bar{x} = (x_1, x_2, x_3) \in M$, $\bar{y} = (y_1, y_2, y_3) \in M$ y de escalares $\lambda, \mu \in \mathbb{R}$.

Observación 1.2.3 En general, toda clausura lineal de un subconjunto de vectores de \mathbb{K}^n es un subespacio vectorial en \mathbb{K}^n , como se muestra en el ejemplo de la Figura 1.3.

Observación 1.2.4 Dado un espacio vectorial V , son subespacios vectoriales de este tanto el propio espacio vectorial V como el subconjunto que contiene únicamente el elemento nulo de V . A este último subespacio se le denota $\{\bar{0}\}$ y se denomina "subespacio nulo en V ".

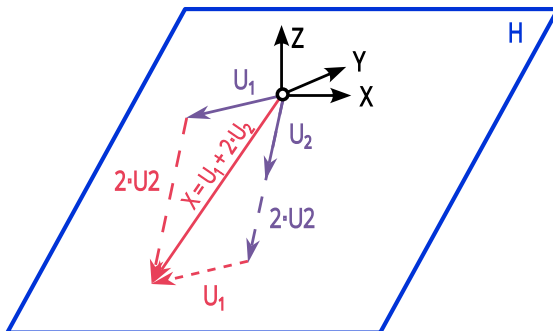


Figura 1.3: Cualquier combinación lineal de dos vectores linealmente independientes \bar{u}_1 y $\bar{u}_2 \in \mathbb{R}^3$, por ejemplo $\bar{x} = \bar{u}_1 + 2\bar{u}_2$, es coplanar a estos. Su clausura lineal genera un plano en el espacio euclidiano que corta el origen de coordenadas. Los vectores que lo contienen forman un subespacio vectorial en \mathbb{R}^3 que se puede denotar como H .

1.2.7 Resumen de los apartados anteriores

1. Los espacios vectoriales constituyen la base del álgebra lineal.
2. Estos están formados por un conjunto de elementos relacionados por dos operaciones denominadas “suma”, o “ley interna”, y “producto por un escalar”, o “ley externa”, que cumplen una serie de propiedades axiomáticas que coinciden con las de la suma de vectores y el producto de vector por escalar.
3. Los espacios \mathbb{R}^n y \mathbb{C}^n son ejemplos de espacios vectoriales.
4. También se pueden construir espacios vectoriales con polinomios, matrices y funciones, por ejemplo.
5. En los espacios vectoriales, y en particular en \mathbb{R}^n y \mathbb{C}^n , se pueden hacer combinaciones lineales de sus elementos.
6. Al conjunto de todas las combinaciones lineales posibles entre elementos de un subconjunto de un espacio vectorial se le conoce como la “clausura lineal” de dicho subconjunto.
7. Si un elemento de un subconjunto de elementos de un espacio vectorial puede representarse como combinación lineal del resto, se dice que los elementos del subconjunto son linealmente dependientes entre si y que forman una familia ligada. En caso contrario, se dice que son linealmente independientes entre si y que forman una familia libre.
8. Un subconjunto de elementos de un espacio vectorial V es un subespacio en V si es cerrado para combinaciones lineales. Es decir, si contiene todas las combinaciones lineales posibles entre los elementos del subconjunto. Toda clausura lineal de un subconjunto de elementos de un espacio V es un subespacio en V .

1.3 Bases y dimensión de un subespacio

1.3.1 Introducción

Se ha terminado el apartado anterior definiendo el concepto de subespacio vectorial y viendo que la clausura lineal de cualquier subconjunto finito de elementos de un espacio vectorial V es un subespacio en V . Resulta además que todo subespacio vectorial H se puede definir como la clausura lineal de un subconjunto de elementos de H al que se denominará “sistema generador de H ”. De entre todos los sistemas generadores de H existen algunos, denominados “bases”, que son particularmente interesantes. Entre las propiedades que cumplen las bases de cualquier subespacio o espacio vectorial H se encuentra la de equicardinalidad: todas ellas tienen el mismo número de elementos. A este número se le conoce como la “dimensión” de H . Todo esto se ve en detalle en este apartado.

1.3.2 Sistemas generadores

Definición 1.3.1 – Sistema generador de un espacio o subespacio

Dados un espacio o subespacio vectorial H y un subconjunto \mathcal{G} de elementos de H , $\mathcal{G} \subset H$, se dice que \mathcal{G} es un sistema generador, o sistema de generadores de H , si:

$$H = L[\mathcal{G}]$$

Es decir, si todo elemento de H puede escribirse como combinación lineal de los de \mathcal{G} .

Todo espacio o subespacio vectorial admite sistemas generadores, o dicho de otra manera, todo espacio o subespacio vectorial H puede definirse como la clausura lineal de algún subconjunto de elementos $\{u_1, \dots, u_k\} \subset H$ que será un sistema generador de este. Además, el número de sistemas generadores que se le puede asociar es infinito; de hecho, realizando combinaciones lineales entre los elementos de un sistema generador \mathcal{G} de H se derivan una infinidad de sistemas generadores de ese mismo espacio o subespacio vectorial H . Esto se verá en mayor detalle cuando estudiemos las propiedades de las bases.

Ejemplo 1.3.1 La clausura lineal del subconjunto:

$$\mathcal{G} = \{\bar{u}_1, \bar{u}_2\} = \{(2, 0, 1), (-1, 1, 0)\} \in \mathbb{R}^3 \quad (1.55)$$

produce un subespacio vectorial en \mathbb{R}^3 definido como:

$$H = L[\mathcal{G}] = \{\bar{x} \in \mathbb{R}^3 : \bar{x} = \lambda_1 \bar{u}_1 + \lambda_2 \bar{u}_2, \lambda_1, \lambda_2 \in \mathbb{R}\} \subset \mathbb{R}^3 \quad (1.56)$$

Por la definición de clausura lineal, los elementos de H son combinaciones lineales de los de \mathcal{G} , por eso \mathcal{G} es un sistema generador de H .

1.3.3 Base

Definición 1.3.2 — Base de un espacio o subespacio

Dado un espacio o subespacio vectorial H , una base \mathcal{B} de H es un subconjunto de vectores de H que cumple estas tres propiedades:

1. Es un sistema generador: todos los vectores de H son combinación lineal de los de \mathcal{B} .
2. Es libre: los elementos de \mathcal{B} son linealmente independientes.
3. Es ordenado: se establece un orden entre los elementos de \mathcal{B} .

Es práctica habitual denotar las bases mediante paréntesis en vez de corchetes para remarcar que se trata de conjuntos ordenados. Por ejemplo, si u_1, u_2, \dots, u_k son k elementos linealmente independientes entre sí que generan H , la base de H constituida por estos elementos en el orden establecido por sus subíndices se denota como (u_1, u_2, \dots, u_k) .

Ejemplo 1.3.2 Dados el subconjunto:

$$\mathcal{B} = \{\bar{u}_1, \bar{u}_2\} = \{(2, 0, 1), (-1, 1, 0)\} \subset \mathbb{R}^3 \quad (1.57)$$

Y el subespacio vectorial $H = L[\mathcal{B}]$, puede verse que:

1. \mathcal{B} es un sistema generador de H , como en el Ejemplo 1.3.1.
2. \mathcal{B} es una familia libre, puesto que:

$$\lambda_1 \bar{u}_1 + \lambda_2 \bar{u}_2 = \bar{0} \Leftrightarrow \lambda_1 = \lambda_2 = 0 \quad (1.58)$$

Ordenando los elementos de \mathcal{B} de acuerdo con sus subíndices, se obtiene un subconjunto ordenado (\bar{u}_1, \bar{u}_2) que es un sistema generador de H y familia libre, por lo que es una base de H .

Una vez más por convenio, se considerará al conjunto vacío \emptyset como la base del subespacio nulo $\{\bar{0}\}$ en V , de modo que todas las propiedades y teoremas de las bases y de la dimensión sean lógicamente consistentes.

1.3.4 Vectores canónicos

Definición 1.3.3 — Vector canónico de \mathbb{K}^n

En \mathbb{K}^n , se define su i -ésimo vector canónico como el vector $\bar{e}_i = (0, \dots, 1, \dots, 0)$ cuyas componentes son todas nulas salvo la i -ésima, que es igual a la unidad.

Ejemplo 1.3.3 Los vectores canónicos de \mathbb{R}^2 son $\bar{e}_1 = (1, 0)$ y $\bar{e}_2 = (0, 1)$, mientras que los vectores canónicos de \mathbb{R}^3 son $\bar{e}'_1 = (1, 0, 0)$, $\bar{e}'_2 = (0, 1, 0)$ y $\bar{e}'_3 = (0, 0, 1)$. Estos vectores son también los vectores canónicos de \mathbb{C}^2 y \mathbb{C}^3 .

1.3.5 Bases canónicas

Definición 1.3.4 — Base canónica de \mathbb{K}^n

La familia formada por todos los vectores canónicos del espacio \mathbb{K}^n , $(\bar{e}_1, \bar{e}_2, \dots, \bar{e}_n)$ es libre, ordenada desde 1 hasta n , y sistema generador de \mathbb{K}^n . Por lo que forma una base a la que se denominará “base canónica de \mathbb{K}^n ”.

Ejemplo 1.3.4 La base canónica del espacio \mathbb{R}^3 es el conjunto ordenado formado por los vectores \bar{e}'_1, \bar{e}'_2 y \bar{e}'_3 definidos en el Ejemplo 1.3.3, $(\bar{e}'_1, \bar{e}'_2, \bar{e}'_3)$. Estos tres vectores son también los vectores canónicos de \mathbb{C}^3 , por lo que ese mismo conjunto ordenado es también la base canónica de \mathbb{C}^3 . Llamemos \mathcal{B} a esta base. Cualquier vector de \mathbb{C}^3 puede expresarse como combinación lineal de los vectores de la base \mathcal{B} . Por ejemplo, el vector $\bar{x} = (0, 2i, -1 + i)$ puede expresarse como:

$$\bar{x} = 0 \cdot \bar{e}_1 + 2i \cdot \bar{e}_2 + (-1 + i) \cdot \bar{e}_3$$

Nótese que los coeficientes escalares de esta descomposición coinciden con las propias componentes del vector \bar{x} . Esto no es una coincidencia, como se remarcará en el Ejemplo 1.3.13, tras definir las coordenadas de un elemento de un espacio o subespacio con respecto a una base de este.

1.3.6 Propiedades de las bases

Supóngase que la clausura lineal del conjunto de elementos ordenados $\mathcal{B} = (u_1, u_2, \dots, u_k)$ produce un subespacio al que se denotará H y que los elementos u_1, u_2, \dots, u_k forman una familia libre. Por la Definición 1.3.4, \mathcal{B} es una base del subespacio H . Cualquier elemento $v \in H$ que no pertenezca a \mathcal{B} puede escribirse como combinación lineal de los elementos de \mathcal{B} :

$$v = \lambda_1 u_1 + \lambda_2 u_2 + \dots + \lambda_k u_k \in H, \lambda_1, \dots, \lambda_k \in \mathbb{K} \quad (1.59)$$

Ejemplo 1.3.5 Defínase el subespacio $H \subset \mathbb{R}^3$ como $L[\mathcal{B}]$, siendo $\mathcal{B} = (\bar{u}_1, \bar{u}_2) = ((2, 0, 1), (-1, 1, 0)) \subset H$ una base de H . El vector $\bar{x} = (1, 1, 1) \in \mathbb{R}^3$ no pertenece a \mathcal{B} pero se puede obtener como combinación lineal de los vectores de \mathcal{B} . De hecho:

$$\begin{aligned} \bar{x} &= 1 \cdot \bar{u}_1 + 2 \cdot \bar{u}_2 = 1 \cdot (2, 0, 1) + 2 \cdot (-1, 1, 0) \Rightarrow \\ &\Rightarrow \bar{x} = (-1, 1, 0) \in H \end{aligned}$$

En general, dado un espacio o subespacio H , todo $x \in H$ puede expresarse como combinación lineal de los elementos de una base dada de H .

Si al conjunto constituido por los elementos u_1, \dots, u_k de una base \mathcal{B} de H se le añade un vector v cualquiera de H se obtiene el conjunto ampliado $\{u_1, u_2, \dots, u_k, v\}$, con $v = \lambda_1 u_1 + \lambda_2 u_2 + \dots + \lambda_k u_k$. Este conjunto ampliado es también un sistema generador de H , pues su clausura lineal coincide con H , pero ya no puede ser una base, pues es una familia linealmente ligada.

De hecho, $\exists \lambda_1, \lambda_2, \dots, \lambda_k \in \mathbb{K}$ con algún $\lambda_i \neq 0$, $i \in \{1, \dots, k\}$ (entre ellos, λ_{k+1}) tales que:

$$\lambda_1 u_1 + \lambda_2 u_2 + \dots + \lambda_k u_k + \lambda_{k+1} v = \bar{0} \quad (1.60)$$

tratándose de una familia ligada, al menos uno de los elementos u_1, \dots, u_k puede ser expresado en función de los restantes. Si u_1 es uno de ellos, este se puede expresar en función de los restantes:

$$u_1 = -\frac{1}{\lambda_1}(\lambda_2 u_2 + \dots + \lambda_k u_k + \lambda_{k+1} v) \quad (1.61)$$

Si se elimina u_1 de $\{u_1, u_2, \dots, u_k, v\}$, se obtiene un conjunto de k vectores linealmente independientes, cuya clausura lineal coincide con H ,

$$\mathcal{B}' = \{u_2, \dots, u_k, v\}, \quad H = L[\mathcal{B}'] \quad (1.62)$$

Estableciendo un orden entre sus elementos se obtiene una nueva base de H .

Mediante el procedimiento anterior es posible obtener tantas bases como se desee de un mismo subespacio o espacio vectorial H . Todas esas bases tienen el mismo número de elementos, lo que lleva al siguiente teorema:

Teorema 1.3.1 — Equicardinalidad de bases

Todas las bases de un mismo subespacio vectorial tienen el mismo número de elementos.

Ejemplo 1.3.6 Considérese el subespacio $H \subset \mathbb{R}^3$ engendrado por la base $\mathcal{B} = (\bar{u}_1, \bar{u}_2) = ((2, 0, 1), (-1, 1, 0)) \subset \mathbb{R}^3$ y el vector $\bar{u}_3 = \bar{u}_1 + \bar{u}_2 = (1, 1, 1)$. Este vector es combinación lineal de los vectores de \mathcal{B} , por lo que $\bar{u}_3 \in H$. Sustituyendo \bar{u}_1 por \bar{u}_3 en \mathcal{B} se tiene

$$\mathcal{B}' = (\bar{u}_3, \bar{u}_2) \quad (1.63)$$

Sustituyendo \bar{u}_2 por \bar{u}_3 en \mathcal{B} se tiene

$$\mathcal{B}'' = (\bar{u}_1, \bar{u}_3) \quad (1.64)$$

Tanto \mathcal{B}' como \mathcal{B}'' son bases de H . Estas tienen el mismo número de elementos que \mathcal{B} : dos.

En todo espacio o subespacio H es posible encontrar un subconjunto de elementos que constituya un sistema generador de este. Eliminando sucesivamente aquellos elementos de \mathcal{G} que sean combinación lineal de los restantes, se llega a una familia libre que es base de H . Esto se resume en el siguiente teorema:

Teorema 1.3.2 — Existencia de bases

Todo subespacio vectorial admite una base.

Ejemplo 1.3.7 Llámese $H \subset \mathbb{R}^3$ a la clausura lineal del conjunto:

$$\mathcal{G} = \{\bar{u}_1, \bar{u}_2, \bar{u}_3\} = \{(2, 0, 1), (-1, 1, 0), (1, 1, 1)\} \subset \mathbb{R}^3 \quad (1.65)$$

\mathcal{G} es sistema generador de H pero no es base, pues $\bar{u}_3 = \bar{u}_1 + \bar{u}_2$. Cualquiera de los tres vectores del conjunto puede escribirse como combinación lineal de los restantes. Eliminando uno de ellos, se obtienen los siguientes subconjuntos:

$$\mathcal{B} = \{\bar{u}_1, \bar{u}_2\} \quad (1.66)$$

$$\mathcal{B}' = \{\bar{u}_2, \bar{u}_3\} \quad (1.67)$$

$$\mathcal{B}'' = \{\bar{u}_1, \bar{u}_3\} \quad (1.68)$$

Todos ellos son sistemas generadores de H y además son familias libres, por lo que son bases del subespacio H , con el orden que establecen sus subíndices.

1.3.7 Dimensión

Según el Teorema 1.3.1, toda base de un espacio o subespacio vectorial dado tiene el mismo número de elementos. A dicho número se le conoce como la “dimensión” del espacio.

Definición 1.3.5 – Dimensión de un espacio vectorial

La dimensión de un espacio o subespacio vectorial H , denotada como $\dim H$, es el número de elementos de cualquiera de sus bases.

Ejemplo 1.3.8 $\dim \mathbb{K}^n = n$, ya que la base canónica que lo genera está formada por n vectores.

Por convenio matemático, al subespacio nulo $\{0\} \subset V$ se le asigna la dimensión cero: $\dim\{0\} = 0$.

Ejemplo 1.3.9 Considérese el subconjunto:

$$\mathcal{G} = \{\bar{u}_1, \bar{u}_2, \bar{u}_3\} = \{(2, 0, 1), (-1, 1, 0), (1, 1, 1)\} \subset \mathbb{R}^3 \quad (1.69)$$

y el subespacio $H = L[\mathcal{G}] \subset \mathbb{R}^3$. Se ha razonado en el Ejemplo 1.3.7 que $\mathcal{B} = \{\bar{u}_1, \bar{u}_2\}$, $\mathcal{B}' = \{\bar{u}_2, \bar{u}_3\}$ y $\mathcal{B}'' = \{\bar{u}_1, \bar{u}_3\}$ son bases de H . Todas ellas tienen el mismo número de elementos. Por la propiedad de equicardinalidad de las bases, toda base de H contendrá dos elementos y, por tanto, la dimensión del subespacio H es igual a dos.

$$\dim H = 2 \quad (1.70)$$

La siguiente proposición establece una relación entre la dimensión de un espacio vectorial V y la dimensión de cualquier subespacio H contenido en V .

Proposición 1.3.3 – Acotación de la dimensión de un subespacio vectorial en V

Dados un espacio vectorial V y un subespacio H contenido en V (de modo que $H \subseteq V$) se cumple la relación $\dim H \leq \dim V$. Además, se tiene que:

$$\dim H = \dim V \Leftrightarrow H = V$$

Demostración:

Se inicia demostrando que $\dim H = \dim V = k$, si y sólo si, $H = V$. Supóngase que $H = V$. Por las definiciones de dimensión y base, $\dim H = k$ elementos cualesquiera linealmente independientes de H engendran H . Al coincidir H con V , esos k elementos también pertenecen a V y generan ese mismo subespacio, por lo que la dimensión de ambos es igual a k .

Supóngase ahora que $\dim H = \dim V = k$ y $H \subseteq V$. k elementos cualesquiera de H linealmente independientes forman una base de este subespacio. Al estar H contenido en V , esos elementos pertenecen también a V , y como la dimensión de V es k , generan el subespacio V . La misma base produce tanto H como V , por lo que ambos subespacios coinciden, siendo $H = V$.

Sólo falta demostrar que $\dim H$ es estrictamente menor que $\dim V$, si y sólo si, H está contenido en V : $H \subset V$. Para ello, puede procederse por reducción al absurdo, apartado A.3.4.

Supóngase que $H \subset V$ y que además $\dim H > \dim V$ (ya se había demostrado que $\dim H = \dim V$ implicaba $H = V$). Escójense $k = \dim H$ elementos linealmente independientes contenidos tanto en H como en V , denotados u_1, \dots, u_k . Estos forman una base del subespacio H a la que se denotará \mathcal{B} . Si $H \subset V$ y $p = \dim V$ son estrictamente menores que $\dim H$, escogiendo p elementos cualesquiera de la base \mathcal{B} y descartando el resto, se obtendría una base de V .

Por simplicidad de notación, se construye una base de V denotada \mathcal{V} con los p primeros elementos de \mathcal{B} : u_1, \dots, u_p . Los $k - p$ elementos descartados pertenecen tanto a H como a V , por lo que en teoría, por la definición de base, se deberían poder expresar como combinación lineal de los de la base \mathcal{V} de V . Se puede, por ejemplo, intentar expresar el elemento $p + 1$ de la base \mathcal{B} de H como combinación lineal de los de la base \mathcal{V} :

$$u_{p+1} = \lambda_1 u_1 + \lambda_2 u_2 + \dots + \lambda_p u_p. \quad (1.71)$$

Pero $u_{p+1} \in V$ es linealmente independiente de u_1, \dots, u_p por su pertenencia a la base \mathcal{B} de H , por lo que, en virtud de la Definición 1.2.4, no se puede expresar como combinación lineal de los elementos u_1, \dots, u_p . Esto contradice la hipótesis de que la dimensión de H sea mayor que la de V .

Así, se demuestra que, si H está contenido en V , su dimensión no puede ser mayor que la de V , como establece la proposición.

□

Ejemplo 1.3.10 El subespacio $H \subset \mathbb{R}^3$, del Ejemplo 1.3.9 tiene dimensión dos. Esta es menor que la dimensión del espacio vectorial en el que H está contenido, que es igual a tres.

1.3.8 Propiedades adicionales de las bases

El apartado 1.3.6 muestra cómo transformar un sistema generador de un subespacio o espacio vectorial H dado en una base de H eliminando sucesivamente elementos que sean linealmente dependientes de los restantes hasta que el conjunto se transforma en una familia libre. También es posible transformar una familia libre \mathcal{S} formada por $p < \dim H$ elementos cualesquiera de H linealmente independientes entre sí en una base de H añadiendo a esta $\dim H - p$ elementos cualesquiera del mismo subespacio, o espacio, linealmente independientes entre sí y de los ya presentes en \mathcal{S} , como afirma el siguiente teorema:

Teorema 1.3.4 – Compleción de bases

Sea H un subespacio o espacio vectorial de dimensión k , y sea $\mathcal{S} = \{u'_1, u'_2, \dots, u'_p\}$ una familia libre de H con $p < k$. Existen elementos $u_{p+1}, \dots, u_k \in H$, tales que $(u'_1, \dots, u'_p, u_{p+1}, \dots, u_k)$ es una base de H .

Demostración:

Dados un espacio o subespacio H de dimensión k y una base cualquiera de este denotada $\mathcal{B} = (u_1, \dots, u_p, u_{p+1}, \dots, u_k)$ con $p < k$, el objetivo es demostrar que la familia libre $\mathcal{S} = \{u'_1, u'_2, \dots, u'_p\} \subset H$ se puede convertir en una base de H añadiéndole $k - p$ elementos tomados de la base \mathcal{B} . En \mathcal{B} hay al menos un elemento linealmente independiente de los de \mathcal{S} , pues de lo contrario \mathcal{S} sería sistema generador de H , lo cual sería imposible porque $p < k$. Tomar este elemento (supóngase que es u_{p+1}) y añadirlo a \mathcal{S} resulta en una familia libre formada por $p + 1$ elementos. Este proceso se puede repetir reiteradamente hasta que la familia libre ampliada \mathcal{S}' contenga k elementos de H que formarán una nueva base de este espacio o subespacio vectorial.

□

Ejemplo 1.3.11 Considérese el plano π contenido en \mathbb{R}^3 cuya ecuación normal es $x_1 - x_2 + x_3 = 0$. El vector $\bar{0} = (0, 0, 0)$ pertenece al plano π pues cumple la ecuación implícita que lo define: $0 - 0 + 0 = 0$. Dos vectores distintos del vector nulo que también cumplen la ecuación implícita son $\bar{u}_1 = (1, 1, 0)$ y $\bar{u}_2 = (0, 1, 1)$. Nótese que son linealmente independientes, pues $\bar{u}_1 \neq \lambda \bar{u}_2 \forall \lambda \in \mathbb{R}$.

Sea $\bar{x} = \lambda_1 \bar{u}_1 + \lambda_2 \bar{u}_2$, $\lambda_1, \lambda_2 \in \mathbb{R}$, puede comprobarse fácilmente que $\bar{x} \in \pi \forall \lambda_1, \lambda_2 \in \mathbb{R}$ al cumplir la ecuación normal del plano π . Por tanto, se cumplen todas las condiciones necesarias para que el conjunto de todos los vectores en π forme un subespacio vectorial en \mathbb{R}^3 . Llámese H a este subespacio. Expresado en su forma implícita:

$$H = \{(x_1, x_2, x_3) \in \mathbb{R}^3 : x_1 - x_2 + x_3 = 0\} \quad (1.72)$$

Observación 1.3.1 † Si un subespacio H formado por vectores pertenecientes a \mathbb{K}^n tiene dimensión k , con $k \leq n$, el vector genérico de dicho espacio puede escribirse como:

$$\bar{x} = \lambda_1 \bar{u}_1 + \dots + \lambda_k \bar{u}_k, \quad \lambda_j \in \mathbb{K} \quad (1.78)$$

Siendo $(\bar{u}_1, \dots, \bar{u}_k)$ cualquier base de H . A esta forma de expresar un vector genérico de un subespacio se la conoce como "forma paramétrica de H ". Para encontrar las $n - k$ ecuaciones de la forma implícita de H se combinan linealmente las k ecuaciones paramétricas de H de modo que se eliminen los parámetros $\lambda_1, \dots, \lambda_k$.

Todo vector perteneciente a H puede descomponerse en combinación lineal de \bar{u}_1 y \bar{u}_2 , por lo que el conjunto ordenado $\mathcal{B} = (\bar{u}_1, \bar{u}_2)$ constituye una base de H y, por tanto, $\dim H = 2$.

$$H = L[\mathcal{B}] \quad (1.73)$$

Conocida una base del subespacio H , este se puede expresar de forma paramétrica † en función de $\dim H = 2$ parámetros:

$$H = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \in \mathbb{R}^3 : \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \lambda_1 \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + \lambda_2 \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \right\} \quad (1.74)$$

$$\forall \lambda_1, \lambda_2 \in \mathbb{R}$$

Una forma alternativa de obtener esta forma paramétrica a partir de la ecuación implícita $x_1 - x_2 + x_3 = 0$ es parametrizar $\dim H$ componentes de esta ecuación, por ejemplo, x_1 y x_2 ($x_1 = \lambda_1$, $x_2 = \lambda_2$), y despejar la componente restante en función de los parámetros establecidos: $x_3 = \lambda_1 + \lambda_2$. Las $n = 3$ ecuaciones paramétricas de H así obtenidas son:

$$\begin{cases} x_1 = \lambda_1 & (1.75a) \\ x_2 = \lambda_1 + \lambda_2 & (1.75b) \\ x_3 = \lambda_2 & (1.75c) \end{cases}$$

De la forma paramétrica de un subespacio H , se puede extraer un conjunto de ecuaciones implícitas que lo definan. El número de ecuaciones implícitas necesarias para definir H es, si $n = 3$ y $k = \dim H = 2$, igual a $n - k = 3 - 2 = 1$. Por tanto, basta con encontrar una ecuación lineal que los dos vectores presentes de la base \mathcal{B} cumplan simultáneamente. Restando la ecuación paramétrica (1.75b) a la suma de las expresiones (1.75a) y (1.75c) se elimina el parámetro λ y se obtiene la ecuación implícita buscada:

$$x_1 - x_2 + x_3 = \lambda_1 - \lambda_1 - \lambda_2 + \lambda_2 = 0 \Rightarrow \quad (1.76)$$

$$x_1 - x_2 + x_3 = 0 \quad (1.77)$$

1.3.9 Coordenadas respecto de una base

Definición 1.3.6 — Coordenadas respecto de una base

Sea $\mathcal{B} = (u_1, \dots, u_k)$ una base del espacio o subespacio H , para todo $x \in H$ existen $\lambda_1, \dots, \lambda_k \in \mathbb{K}$ tal que $x = \lambda_1 u_1 + \dots + \lambda_k u_k$. Los escalares $\lambda_1, \dots, \lambda_k$ se denominan coordenadas del elemento x respecto a \mathcal{B} .

A continuación, se demuestra que las coordenadas de cualquier elemento de un espacio o subespacio vectorial respecto de una base dada son únicas, independientemente de la base escogida.

Supóngase que el elemento $x \in H$ puede descomponerse de dos maneras distintas respecto a la base \mathcal{B} . Es decir:

$$x = \lambda_1 \mathbf{u}_1 + \lambda_2 \mathbf{u}_2 + \cdots + \lambda_k \mathbf{u}_k, \quad \lambda_1, \dots, \lambda_k \in \mathbb{K} \quad (1.79)$$

$$x = \mu_1 \mathbf{u}_1 + \mu_2 \mathbf{u}_2 + \cdots + \mu_k \mathbf{u}_k, \quad \mu_1, \dots, \mu_k \in \mathbb{K} \quad (1.80)$$

Si a la ecuación (1.79) se le resta (1.80) queda:

$$0 = (\lambda_1 - \mu_1) \mathbf{u}_1 + (\lambda_2 - \mu_2) \mathbf{u}_2 + \cdots + (\lambda_k - \mu_k) \mathbf{u}_k \quad (1.81)$$

Los elementos de la base \mathcal{B} son, por definición, linealmente independientes entre sí. Por tanto, para que la combinación lineal anterior produzca el elemento nulo, todos los factores escalares $(\lambda_1 - \mu_1)$, $(\lambda_2 - \mu_2)$, etc. deben ser nulos. O dicho de otro modo, $\lambda_i - \mu_i = 0 \quad \forall i = 1, \dots, k$. Esto implica que $\lambda_i = \mu_i \quad \forall i = 1, \dots, k$, lo cual contradice la hipótesis inicial de que las descomposiciones son distintas ($\lambda_i \neq \mu_i \quad \forall i = 1, \dots, k$).

Queda así demostrado por reducción al absurdo que, fijada una base \mathcal{B} de H , cada elemento de H tiene asociadas unas coordenadas $\lambda_1, \dots, \lambda_k$ respecto de \mathcal{B} únicas.

Proposición 1.3.5 — Unicidad de las coordenadas respecto de una base

Dados un espacio o subespacio H y una base \mathcal{B} de este, las coordenadas de $x \in H$ respecto a la base \mathcal{B} son únicas $\forall x \in H$.

Definición 1.3.7 — Vector de coordenadas

El vector $(\lambda_1, \dots, \lambda_k) \in \mathbb{K}^k$ se denomina vector de coordenadas de x respecto a \mathcal{B} , y se suele denotar por $[x]_{\mathcal{B}}$.

Dado un espacio o subespacio H de dimensión k , y fijada una base \mathcal{B} de este, la relación entre cada elemento de H y su vector de coordenadas asociado respecto de \mathcal{B} es biunívoca †. Ya se ha visto, Proposición 1.3.5, que a cada elemento de H sólo se le puede asociar un vector de coordenadas en \mathbb{K}^k , pues sus coordenadas son únicas.

Falta demostrar que un vector en \mathbb{K}^k no puede ser vector de coordenadas respecto de \mathcal{B} de dos o más elementos de H . De lo contrario, sería posible encontrar dos elementos distintos $x, y \in H$ cuyas coordenadas respecto de \mathcal{B} fuesen las mismas. Es decir, existirían escalares $\lambda_1, \dots, \lambda_k \in \mathbb{K}$ tales que:

$$x = \lambda_1 \mathbf{u}_1 + \lambda_2 \mathbf{u}_2 + \cdots + \lambda_k \mathbf{u}_k \quad (1.82)$$

$$y = \lambda_1 \mathbf{u}_1 + \lambda_2 \mathbf{u}_2 + \cdots + \lambda_k \mathbf{u}_k \quad (1.83)$$

Si a la ecuación (1.82) se le resta (1.83) queda:

$$x - y = 0 \cdot \mathbf{u}_1 + 0 \cdot \mathbf{u}_2 + \cdots + 0 \cdot \mathbf{u}_k = 0 \quad (1.84)$$

Observación 1.3.2 No confundir el concepto de vector de \mathbb{K}^n con el concepto de vector de coordenadas de un elemento de un espacio o subespacio vectorial con respecto a una base de este.

Observación 1.3.3 † Una relación biunívoca entre los elementos de A y los de B es una relación en la que a cada elemento de A le corresponde uno sólo de B ; y viceversa, a cada elemento de B le corresponde uno sólo de A . A este tipo de asociación se le conoce también como "biyectiva", Definición 3.1.5.3.

Este resultado contradice la asunción de que x e y son elementos distintos de H . Así, se demuestra por reducción al absurdo que a cada vector de coordenadas respecto de una misma base de H le corresponde un único elemento de H , por lo que la relación entre elementos de H y vectores de \mathbb{K}^k es biunívoca.

Proposición 1.3.6 – Biunivocidad de los vectores de coordenadas respecto de una base

Dados un espacio o subespacio H y una base \mathcal{B} de este, la relación entre los elementos de H y sus vectores de coordenadas asociados respecto de \mathcal{B} es biunívoca.

Ejemplo 1.3.12 Se tiene el subespacio:

$$H = L[\mathcal{B}] = L \left[\left(\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \right) \right] \subset \mathbb{R}^3 \quad (1.85)$$

Las coordenadas del vector $\bar{x} = (2, 1, -1) \in H$ respecto de la base \mathcal{B} se pueden obtener resolviendo el sistema lineal de ecuaciones:

$$\begin{pmatrix} 2 \\ 1 \\ -1 \end{pmatrix} = \lambda_1 \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + \lambda_2 \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \Rightarrow \begin{cases} 2 = \lambda_1 \\ 1 = \lambda_1 + \lambda_2 \\ -1 = \lambda_2 \end{cases} \quad (1.86)$$

Este sistema es compatible y determinado por la unicidad de las coordenadas respecto a una base, Proposición 1.3.5. Sus soluciones son $\lambda_1 = 2$ y $\lambda_2 = -1$. Estas son las dos coordenadas del vector \bar{x} con respecto a la base \mathcal{B} .

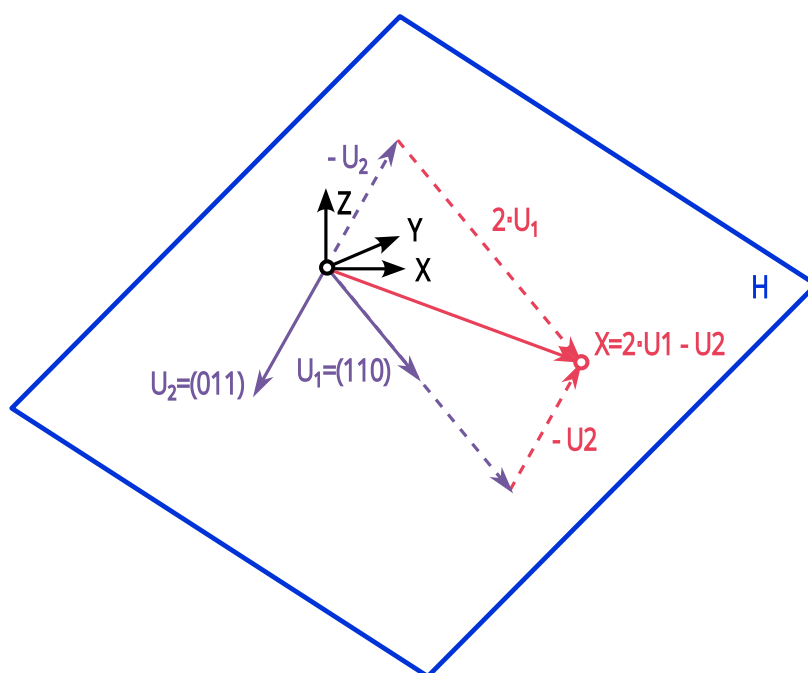


Figura 1.4: Las coordenadas de $\bar{x} \in H$ respecto de la base $\mathcal{B} = (\bar{u}_1, \bar{u}_2)$ de $H \subset \mathbb{R}^3$ son los coeficientes de la combinación lineal de \bar{u}_1 y \bar{u}_2 que produce \bar{x} . En el Ejemplo 1.3.12, las coordenadas de \bar{x} respecto de \mathcal{B} son 2 y -1 , pues $\bar{x} = 2\bar{u}_1 - \bar{u}_2$.

Ejemplo 1.3.13 Las coordenadas del vector $\bar{x} = (2, 1, -1) \in H$ respecto de la base $\mathcal{B} = ((1, 1, 0), (0, 1, 1)) \subset H$ son, como se ha visto en el Ejemplo 1.3.12, $\lambda_1 = 2$ y $\lambda_2 = -1$. Por tanto, el vector de coordenadas de \bar{x} respecto de \mathcal{B} es el vector:

$$[\bar{x}]^{\mathcal{B}} = (2, -1) \in \mathbb{R}^2 \quad (1.87)$$

Por otra parte, las coordenadas de $\bar{x} = (2, 1, -1) \in \mathbb{R}^3$ respecto de la base canónica de \mathbb{R}^3 , a la que se denotará como \mathcal{B}_0 , son, respectivamente, 2, 1 y -1 , pues:

$$(2, 1, -1) = 2 \cdot (1, 0, 0) + 1 \cdot (0, 1, 0) - 1 \cdot (0, 0, 1) \quad (1.88)$$

Por lo que: †

$$[\bar{x}]^{\mathcal{B}_0} = (2, 1, -1) = \bar{x} \quad (1.89)$$

Observación 1.3.4 † En los espacios \mathbb{K}^n , las coordenadas de un vector con respecto a la base canónica de \mathbb{K}^n coinciden con las propias componentes del vector.

1.3.10 Resumen

1. El sistema generador de un espacio vectorial es un subconjunto de elementos de este, en particular vectores, si se trabaja en \mathbb{R}^n o en \mathbb{C}^n , que engendra el espacio por clausura lineal.
2. Una base de un espacio es un sistema generador de este que es libre y ordenado.
3. Todo espacio tiene infinito número de sistemas generadores e infinitas bases.
4. Todas las bases de un espacio tienen el mismo número de elementos. A este número se le conoce como la “dimensión” del espacio o subespacio.
5. Cuando un elemento x de un espacio dado se descompone en combinación lineal de los elementos de una base \mathcal{B} de este, a los coeficientes de la descomposición se les llama “coordenadas del elemento x respecto de \mathcal{B} ”. Al vector cuyas componentes son las coordenadas de x se le llama “vector de coordenadas de x respecto de \mathcal{B} ”.
6. La forma paramétrica de un subespacio es su expresión como una combinación lineal genérica de los elementos de una de sus bases. En \mathbb{R}^n o \mathbb{C}^n , dicho subespacio puede ser caracterizado también por $n - k$ ecuaciones implícitas que se pueden derivar de su forma paramétrica, donde k es tanto la dimensión del subespacio como el número de parámetros de la forma paramétrica.

1.4 Intersección y suma de subespacios

1.4.1 Introducción

Los subespacios vectoriales son conjuntos de elementos, por lo que es posible definir operaciones de intersección y de unión de subespacios. En este nuevo apartado, se verá que el conjunto intersección de dos subespacios cualesquiera de un mismo espacio V es siempre un subespacio en V , y que la unión de dos subespacios vectoriales es un subespacio sólo si se cumple una determinada condición.

Observación 1.4.1 Los apéndices contienen las Definiciones A.2.4 y A.2.5 de unión e intersección de conjuntos. Por lo que respecta a la aplicación de la teoría de conjuntos en el álgebra lineal, los conjuntos consistirán en subespacios vectoriales, bases o sistemas generadores.

Una operación entre subespacios que no tiene parangón en otros conjuntos matemáticos es la suma de subespacios. Tras definirla, se enunciarán las condiciones que deben cumplir los subespacios sumados para que su suma sea directa y para que estos sean suplementarios.

Dados dos subespacios vectoriales en un mismo espacio vectorial, las dimensiones de estos, su suma y su intersección cumplen la denominada “relación de *Grassmann*”. Con ella se pondrá fin a este primer tema.

1.4.2 Intersección de subespacios

Vamos a preguntarnos si la intersección de dos subespacios cualesquiera del espacio vectorial V , denotados $L, M \subseteq V$, es también un subespacio. Para que la respuesta sea afirmativa, el conjunto intersección ha de ser cerrado para la suma de elementos y para el producto por un escalar, lo que se resume en la siguiente condición de clausura para combinaciones lineales de sus elementos:

$$\forall u, v \in H, \forall \lambda, \mu \in \mathbb{K}, \lambda u + \mu v \in H \quad (1.90)$$

Una condición necesaria para que se cumpla la clausura para combinaciones lineales es que H contenga el elemento nulo de V , por lo que cualquier par de subespacios de V contiene el elemento nulo, y su intersección también contendrá el elemento nulo.

Por otro lado, cada uno de los dos elementos u, v del conjunto intersección de L con M pertenece, por la definición de intersección de conjuntos, tanto a L como a M ; y por la condición (1.90), cualquier combinación lineal de u y v pertenece tanto a L como a M , por separado: $\lambda_1 u + \lambda_2 v \in L$ y $\lambda_1 u + \lambda_2 v \in M \forall \lambda_1, \lambda_2 \in \mathbb{K}$.

Esto último implica que cualquier combinación lineal de ambos elementos pertenece también a $L \cap M$. Se concluye que $\lambda_1 u + \lambda_2 v \in L \cap M \forall \lambda_1, \lambda_2 \in \mathbb{K} \forall u, v \in L \cap M$, terminando así la demostración de que la intersección de dos subespacios cualesquiera es un subespacio vectorial.

Proposición 1.4.1 — Intersección de subespacios vectoriales

La intersección de dos subespacios vectoriales contenidos en un mismo espacio V es también un subespacio vectorial en V .

Ejemplo 1.4.1 La intersección de los subespacios vectoriales en \mathbb{R}^3 $A = L[(1, 0, -1), (0, 1, 1)]$ y $B = L[(1, 1, 0)]$, denotada $A \cap B$, es el subespacio en \mathbb{R}^3 formado por todos los vectores pertenecientes simultáneamente a ambos subespacios. A continuación, se procede a calcular el subespacio $A \cap B$.

Para definir A basta con $n - \dim A = 3 - 2 = 1$ ecuación implícita. Una ecuación que los dos vectores de la base dada de A satisfacen es $x_1 - x_2 + x_3 = 0$, $(x_1, x_2, x_3) \in \mathbb{R}^3$. Esta es una ecuación implícita que define A .

Para definir B son necesarias $n - \dim B = 3 - 1 = 2$ ecuaciones implícitas linealmente independientes entre sí. Una primera ecuación que el vector $(1, 1, 0) \in A$ satisface es $x_3 = 0$. Otra ecuación linealmente independiente de la anterior que el vector $(1, 1, 0)$ también cumple es $x_1 - x_2 = 0$. Estas son dos ecuaciones implícitas que definen B .

Por tanto, los conjuntos de ecuaciones implícitas que definen los subespacios A y B son:

$$A = \{(x_1, x_2, x_3) \in \mathbb{R}^3 : x_1 - x_2 + x_3 = 0\} \quad (1.91)$$

$$B = \{(x_1, x_2, x_3) \in \mathbb{R}^3 : x_3 = 0, x_1 - x_2 = 0\} \quad (1.92)$$

Al unir todas las ecuaciones implícitas de A y B , se obtiene un conjunto de ecuaciones implícitas que define el subespacio intersección $A \cap B$.

$$\begin{cases} x_1 - x_2 + x_3 = 0 & (1.93a) \\ x_3 = 0 & (1.93b) \\ x_1 - x_2 = 0 & (1.93c) \end{cases}$$

Todo vector de \mathbb{R}^3 cuyas componentes cumplan simultáneamente las tres ecuaciones implícitas pertenece a $A \cap B$.

Nótese que la ecuación (1.93b) puede obtenerse sustrayendo (1.93c) a (1.93a) por lo que cada una de las tres ecuaciones es redundante y se puede eliminar. Tras eliminar (1.93a), quedan dos ecuaciones implícitas linealmente independientes que definen el subespacio $A \cap B$.

$$\begin{cases} x_3 = 0 \\ x_1 - x_2 = 0 \end{cases} \quad (1.94)$$

La dimensión de $A \cap B$ es:

$$n - \dim A \cap B = 2 \Rightarrow \dim A \cap B = n - 2 = 1 \quad (1.95)$$

Un vector que cumple ambas ecuaciones implícitas es el $(1, 1, 0) \in A \cap B$. Por lo tanto, el conjunto ordenado $((1, 1, 0))$ es una base del subespacio $A \cap B$, que puede definirse como:

$$A \cap B = L[(1, 1, 0)] \quad (1.96)$$

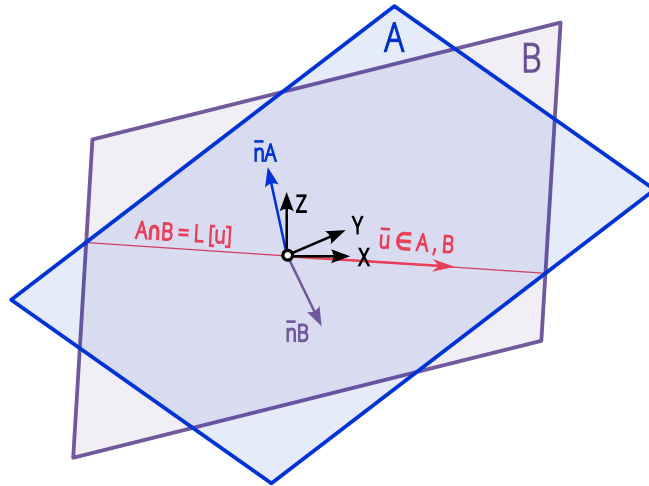


Figura 1.5: Todos los vectores pertenecientes simultáneamente a los subespacios de dimensión dos $A, B \subset \mathbb{R}^3$ del Ejemplo 1.4.1 están contenidos en una recta que corta el origen de coordenadas y que tiene como vector director el vector $\bar{u} \in A, B$. Por tanto, $A \cap B = L[\bar{u}]$.

1.4.3 Unión de subespacios

Al contrario que la intersección de subespacios vectoriales, la unión de subespacios no siempre cumple la condición de clausura (1.90) que, por definición, ha de cumplir todo subconjunto de elementos de un espacio V para ser un subespacio en V .

La condición necesaria y suficiente que toda pareja de subespacios vectoriales debe cumplir para que su unión sea también subespacio viene dada por la siguiente proposición:

Proposición 1.4.2 – Unión de subespacios vectoriales

La unión de dos subespacios vectoriales L y M contenidos en un mismo espacio V es también un subespacio vectorial en V , si y sólo si, se trata de subespacios anidados. Es decir, si uno está contenido en el otro:

$$L \cup M \text{ subespacio en } V \Leftrightarrow (L \subseteq M \vee M \subseteq L)$$

Además, $L \subseteq M \Rightarrow L \cup M = M$, y $M \subseteq L \Rightarrow L \cup M = L$.

Demostración:

Comiencese demostrando que la unión de dos subespacios $L, M \subseteq V$ anidados produce siempre un subespacio en V . Supóngase que uno de los dos subespacios se encuentra contenido en el otro, por ejemplo, L en M . Todo elemento perteneciente a L pertenece también a M , por lo que la unión de ambos conjuntos coincide con L . Siendo L un subespacio, se demuestra que $L \cup M$ es un subespacio en V .

Si en la demostración anterior se invierten los roles de L y M . Es decir, si $M \subseteq L$, se tiene que la unión de ambos conjuntos coincide con el subespacio M , por lo que $L \cup M$ sería también un subespacio en V .

Falta demostrar que la anidación de los subespacios es una condición no sólo suficiente sino también necesaria para que su unión sea un subespacio. Supóngase que $L \cup M$ es un subespacio en V y que $M \not\subseteq L$. Para que la proposición sea correcta esto debería implicar que $L \subseteq M$. Tómense un $x \in L$ y un $y \in M$, $y \notin L$. Tanto x como y pertenecen a $L \cup M$.

Al tratarse $L \cup M$ de un subespacio, $x + y$ pertenece a $L \cup M$, y por la Definición A.2.4 de unión de conjuntos, $x + y$ se encuentra o en L , o en M , o en ambos.

Supóngase que $x + y \in L$. La combinación lineal $x + y + (-x)$ debería pertenecer también a L . Sin embargo, queda $x + y + (-x) = y$, resultado que contradice la suposición de que $y \notin L$, por lo que $x + y \notin L$. Supóngase ahora que $x + y \in M$. La combinación lineal $x + y + (-y)$ debería pertenecer también a M . De hecho, queda $x + y + (-y) = x \in M$, por lo que $x + y \in M$ al igual que el elemento y . Por tanto, x debe pertenecer tanto a L como a M y se concluye que $L \subseteq M$.

Intercambiando en la demostración anterior los roles de L y M se demuestra que, si $L \cup M$ es un subespacio en V y $L \not\subseteq M$, se cumple $M \subseteq L$. Queda así demostrado también el converso de la Proposición 1.4.2. \square

Ejemplo 1.4.2 El subespacio $B = L[(0, 1, 1)] \subset \mathbb{R}^3$ está claramente contenido en $A = L[(1, 0, -1), (0, 1, 1)] \in \mathbb{R}^3$. De hecho, todo vector de \mathbb{R}^3 proporcional al $(0, 1, 1)$ pertenece tanto a A como a B . Al cumplirse $B \subset A$, su unión es un subespacio vectorial coincidente con el subespacio de mayor dimensión de los dos: $A \cup B = A$.

1.4.4 Suma de subespacios

Definición 1.4.1 — Suma de subespacios

Dados dos subespacios L, M contenidos en V , se define su suma como:

$$L + M = \{u + v : u \in L, v \in M\}$$

Es decir, como el conjunto de todas las sumas posibles entre los elementos de L y los de M .

Proposición 1.4.3 — La suma de subespacios es siempre un subespacio

La suma de dos subespacios vectoriales cualesquiera contenidos en un espacio V es también un subespacio vectorial en V .

Demostración:

Como se razonó en el apartado 1.2.6, una condición necesaria para que la suma de dos subespacios L y M en V sea un subespacio en V es que contenga el elemento nulo. L y M contienen el elemento nulo de V . Además, se tiene que $0 + 0 = 0$. Por tanto, el elemento $0 \in V$ pertenece a $L + M$ para toda pareja de subespacios $L, M \subseteq V$. Falta sólo comprobar que $L + M$ sea cerrado para combinaciones lineales de sus elementos.

Tómense $u, v \in L + M$ cualesquiera. Por la Definición 1.4.1 de suma de subespacios, ambos elementos se pueden descomponer en la suma de elementos pertenecientes a L y M :

$$u = u_L + u_M, \quad u_L \in L, \quad u_M \in M \quad (1.97)$$

$$v = v_L + v_M, \quad v_L \in L, \quad v_M \in M \quad (1.98)$$

Sean $\alpha, \beta \in \mathbb{K}$ dos escalares cualesquiera:

$$\begin{aligned} \alpha u + \beta v &= \alpha(u_L + u_M) + \beta(v_L + v_M) = \\ &= (\alpha u_L + \beta v_L) + (\alpha u_M + \beta v_M) \end{aligned} \quad (1.99)$$

Como $\alpha u_L + \beta v_L \in L$, mientras que $\alpha u_M + \beta v_M \in M$, toda combinación lineal de elementos de $L + M$ se encuentra en $L + M$. Queda así demostrado que la suma de subespacios en V produce un subespacio vectorial. \square

Ejemplo 1.4.3 Defínase un subespacio vectorial A mediante la forma paramétrica:

$$A = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \in \mathbb{R}^3 : \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \lambda_1 \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} + \lambda_2 \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \right\} \quad (1.100)$$

$\forall \lambda_1, \lambda_2 \in \mathbb{R}$

Y un segundo subespacio B mediante la forma paramétrica:

$$B = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \in \mathbb{R}^3 : \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \lambda_1 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \right\} \quad \forall \lambda_1 \in \mathbb{R} \quad (1.101)$$

La unión de sistemas generadores de A y B produce un sistema generador del subespacio suma $A + B$:

$$\mathcal{G} = \{(1, 0, -1), (0, 1, 1), (1, 0, 0)\} = \{\bar{u}_1, \bar{u}_2, \bar{u}_3\} \quad (1.102)$$

Para extraer una base \mathcal{B} de $A + B$ a partir del sistema generador \mathcal{G} , se eliminan uno a uno aquellos vectores que sean combinación lineal de los restantes. Para comprobar si \bar{u}_1, \bar{u}_1 y \bar{u}_2 son linealmente independientes o no, se resuelve el sistema:

$$\lambda_1 \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} + \lambda_2 \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} + \lambda_3 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \Rightarrow \begin{cases} \lambda_1 + \lambda_3 = 0 \\ \lambda_2 = 0 \\ \lambda_2 + \lambda_1 = 0 \end{cases} \quad (1.103)$$

Este sistema homogéneo de ecuaciones lineales es compatible determinado, Teorema 4.3.1, por lo que su única solución es $\lambda_1 = \lambda_2 = \lambda_3 = 0$. Por tanto, los tres vectores son linealmente independientes y \mathcal{G} es, además de sistema generador, una base de $A + B$.

Al contener dicha base tres vectores, se deduce que $\dim A + B = 3$, valor que coincide con el número de componentes de los vectores de $A + B$. De hecho, es la dimensión del espacio \mathbb{R}^3 al que pertenecen los vectores de $L + M$. Por la Proposición 1.3.3, se deduce que $A + B = \mathbb{R}^3$.

1.4.5 Suma directa de subespacios

Definición 1.4.2 – Suma directa de subespacios

Dados dos subespacios vectoriales L, M en V , si $L \cap M = \{0\}$ su suma se denota $L \oplus M$ y se dice que forman suma directa.

Ejemplo 1.4.4 Por definición, para que los subespacios $A, B \subset \mathbb{R}^3$ del Ejemplo 1.4.3 formen suma directa, su intersección ha de coincidir con el subespacio nulo de \mathbb{R}^3 , $A \cap B = \{\bar{0}\}$.

La unión de ecuaciones implícitas que definen por separado A y B constituye el sistema:

$$\begin{cases} x_1 - x_2 + x_3 = 0 & (1.104a) \\ x_2 = 0 & (1.104b) \\ x_3 = 0 & (1.104c) \end{cases}$$

Al resolver este sistema de ecuaciones lineales se encuentra una única solución: $x_1 = x_2 = x_3 = 0$, por lo que el único vector que A y B tienen en común es el $(0, 0, 0) \in \mathbb{R}^3$.

Se concluye que la intersección de A con B es el subespacio nulo en \mathbb{R}^3 , por lo que forman suma directa, denotándose $A \oplus B$.

Se puede comprobar que si dos subespacios forman suma directa, cada elemento de $L \cap M$ se descompone de manera unívoca en suma de dos elementos de L y M ; y viceversa, si cada elemento de $L + M$ se descompone en suma de dos elementos en L y M de manera unívoca, se tiene que L y M forman suma directa. Esto último se resume en la siguiente proposición:

Proposición 1.4.4 – Caracterización de la suma directa

Sean L, M subespacios vectoriales en V . Estos forman suma directa, si y sólo si, $\forall x \in L + M$ existen elementos únicos $u \in L$, $v \in M$ tales que $x = u + v$.

Demostración:

Para demostrar que la suma directa de subespacios $L, M \subseteq V$ implica descomposición única de todos los elementos de $L + M$ en suma de elementos en L y M , supóngase que existe algún $x \in L + M$ que se pueda descomponer de dos maneras distintas en suma de elementos en L y M no nulos:

$$\begin{aligned} x &= u_1 + v_1, \quad u_1 \in L, \quad u_1 \neq 0, \quad v_1 \in M, \quad v_1 \neq 0 \\ x &= u_2 + v_2, \quad u_2 \in L, \quad u_2 \neq 0, \quad v_2 \in M, \quad v_2 \neq 0 \\ &u_1 \neq u_2, \quad v_1 \neq v_2 \end{aligned} \quad (1.105)$$

Substrayendo ambas ecuaciones, se obtiene $0 = (u_1 - u_2) + (v_1 - v_2)$.

Denotando $u = u_1 - u_2 \in L$ y $v = v_1 - v_2 \in M$, $0 = u + v \Rightarrow u = -v$, por lo que ambos elementos pertenecen al mismo subespacio: la intersección de L con M . Al ser u y v no nulos, se tiene que la intersección es distinta del subespacio nulo. Se demuestra así que suma directa implica descomposición única.

Falta demostrar la implicación inversa: que la descomposición única de todos los elementos de V en suma de elementos de L y M implica suma directa. El contrapositivo de esa implicación, punto 3 de la Definición A.1.4, es: si L y M no forman suma directa, la descomposición de los elementos de V no es única. Supóngase que existe un $w \in V$ no nulo que pertenezca tanto a L como a M . Es decir, que la intersección de ambos subespacios es distinta del subespacio nulo. Tómesese un $x \in V$ y descompóngase en suma de elementos no nulos en L y M :

$$x = u + v, \quad u \in L, \quad u \neq 0, \quad v \in M, \quad v \neq 0 \tag{1.106}$$

Sumando y restando $w \in L \cap M$ a la ecuación (1.106) y reordenando, se obtienen dos descomposiciones distintas de x en suma de elementos en L y M :

$$\begin{aligned} x &= u + v + w - w = (u + w) + (v - w) = u' + v' \\ u' &\in L, \quad v' \in M, \quad u' \neq u, \quad v' \neq v \end{aligned} \tag{1.107}$$

Por lo que L y M no formarían suma directa. Así, se demuestra que la suma directa es condición no sólo suficiente, sino también necesaria para la unicidad de las descomposiciones.

□

$$M = L[(0 \ -1 \ 1)] = L[u_2] \quad L = L[(1 \ 1 \ 1)] = L[u_1]$$

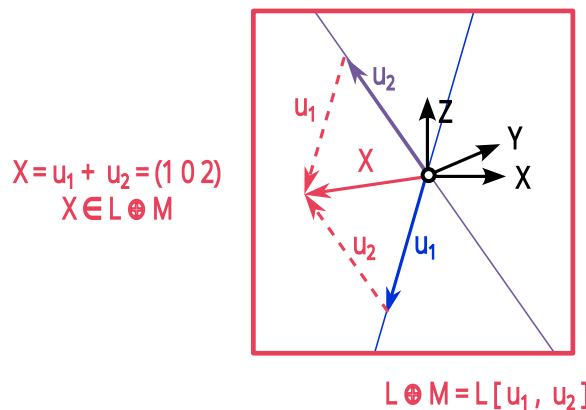


Figura 1.6: El único vector que los subespacios de dimensión uno $L = L[\bar{u}_1]$ y $M = L[\bar{u}_2]$ comparten es el vector nulo. Gráficamente, se pueden representar en el espacio euclidiano como dos rectas no coincidentes que se intersectan en el origen de coordenadas. L y M forman suma directa: un subespacio en dimensión dos igual a la clausura lineal de \bar{u}_1 y \bar{u}_2 . Al formar suma directa, la única descomposición posible del vector $\bar{x} = (1, 0, 2) \in L \oplus M$ en suma de un vector en L más otro en M es $\bar{u}_1 + \bar{u}_2$.

1.4.6 Subespacios suplementarios

Si la suma de dos subespacios vectoriales $L, M \subseteq V$ es igual a V y además es directa, siendo su intersección el subespacio nulo en V , se dice que L y M son suplementarios en V .

Definición 1.4.3 – Subespacios suplementarios

Dos subespacios $L, M \subseteq V$ se dicen suplementarios en V si $L \cap M = \{0\}$ y su suma cubre todo el espacio V . Es decir, si $L \oplus M = V$.

Por la Proposición 1.4.4, si dos subespacios L y M son suplementarios en V , cualquier elemento en V se puede descomponer en la suma de elementos en L y M , y además esta descomposición es única.

Ejemplo 1.4.5 En el Ejemplo 1.4.4, se ha visto que los subespacios A y B forman suma directa; y en el Ejemplo 1.4.3, se había visto que $A \oplus B = \mathbb{R}^3$, por lo que puede decirse que A y B son suplementarios en \mathbb{R}^3 .

1.4.7 La relación de Grassmann

Como se ha discutido en el Ejemplo 1.4.3, la unión de las bases \mathcal{B} y \mathcal{V} de dos subespacios L y M en V produce un sistema generador, pero no necesariamente una base, del subespacio suma $L + M$. Para convertir este sistema generador en una base se pueden eliminar sucesivamente aquellos elementos que sean linealmente dependientes del resto hasta que el sistema generador se convierta en una familia libre, como en el Ejemplo 1.3.7, coincidiendo el número de elementos eliminados con la dimensión de $L \cap M$. Si L y M formasen suma directa, Definición 1.4.2, su intersección contendría solo el vector nulo y no habría elementos que eliminar al unir las bases: la unión sería una base de $L \cap M$.

Estas ideas se encuentran sintetizadas en la relación de *Grassmann*. Se trata de una fórmula que relaciona, dados dos subespacios cualesquiera L y M en V , sus dimensiones, la de su suma y la de su intersección.

Teorema 1.4.5 – Relación de Grassmann

Dados dos subespacios vectoriales L, M en V , se tiene que:

$$\dim L + M = \dim L + \dim M - \dim L \cap M$$

Demostración:

Se sabe por el Teorema 1.3.4 que una base cualquiera de $L \cap M$, $\mathcal{B} = (u_1, \dots, u_r)$ puede completarse con $p - r$ elementos u_{r+1}, \dots, u_p hasta obtener una base de L que se denotará: \mathcal{B}_L . A la familia libre formada por los $p - r$ vectores añadidos se la denotará como S_L .

De la misma forma, la base \mathcal{B} puede completarse con $q - r$ elementos u'_{r+1}, \dots, u'_q hasta obtener una base de M que se denotará: \mathcal{B}_M . A la familia libre formada por los $q - r$ vectores añadidos se la denotará como S_M .

Para demostrar el Teorema 1.4.5, basta con comprobar que el conjunto $\mathcal{S} = \mathcal{B} \cup S_L \cup S_M$ constituye una base de $L + M$, pues de esta forma se tendría que $\dim L + M = r + (p - r) + (q - r) = p + q - r = \dim L + \dim M - \dim L \cap M$ y se cumpliría la ecuación (1.4.5).

En primer lugar, se necesita comprobar que \mathcal{S} es un sistema generador de $L + M$. Es decir, que todo $x \in L + M$ se puede descomponer como com-

binación lineal de elementos en \mathcal{S} . En segundo lugar, se debe comprobar que \mathcal{S} es una familia libre.

Lo primero claramente se cumple, pues en \mathcal{S} se encuentran todos los elementos de las bases \mathcal{B}_L y \mathcal{B}_M , lo que permite descomponer cualquier elemento de la clausura lineal de \mathcal{S} en la suma de un elemento en L más otro elemento en M . Para demostrar lo segundo, se construye la ecuación:

$$\begin{aligned} \mathbf{x} = \lambda_1 \mathbf{u}_1 + \cdots + \lambda_r \mathbf{u}_r + \lambda_{r+1} \mathbf{u}_{r+1} + \cdots + \lambda_p \mathbf{u}_p + \\ + \mu_{r+1} \mathbf{u}'_{r+1} + \cdots + \mu_q \mathbf{u}'_q = 0 \end{aligned} \quad (1.108)$$

Y aplicando la Definición 1.2.3, se comprueba que la identidad $\mathbf{x} = 0$ (1.108) se cumple, si y sólo si, todos los coeficientes escalares son nulos. Así, $\mathbf{x} \in L+M$ se podría escribir como la suma de un elemento $\mathbf{u} = \lambda_1 \mathbf{u}_1 + \cdots + \lambda_r \mathbf{u}_r$ perteneciente a L más otro elemento $\mathbf{v} = \mu_{r+1} \mathbf{u}'_{r+1} + \cdots + \mu_q \mathbf{u}'_q$ perteneciente a M y esta descomposición sería única por la Proposición 1.3.5.

Si la ecuación (1.108) se cumple, despejando \mathbf{v} se tiene $\mathbf{v} = \mathbf{x} - \mathbf{u} = 0 - \mathbf{u} = -\mathbf{u}$, por lo que \mathbf{v} pertenecería también a L . En consecuencia, pertenecería a $L \cap M$ y se podría escribir como combinación lineal de los elementos de la base \mathcal{B} :

$$\mathbf{v} = \mu_{r+1} \mathbf{u}'_{r+1} + \cdots + \mu_q \mathbf{u}'_q = \mu_1 \mathbf{u}_1 + \cdots + \mu_r \mathbf{u}_r \quad (1.109)$$

Por tanto:

$$\mu_{r+1} \mathbf{u}'_{r+1} + \cdots + \mu_q \mathbf{u}'_q - \mu_1 \mathbf{u}_1 - \cdots - \mu_r \mathbf{u}_r = 0 \quad (1.110)$$

Los r vectores presentes en (1.110) son los de la base \mathcal{B}_M , por lo que son linealmente independientes. Por esta razón, la identidad (1.110) se cumple, si y sólo si, μ_1, \dots, μ_q son todos nulos. Sustituyendo $\mu_i = 0 \forall i = 1, \dots, r$ en la ecuación (1.108) se obtiene:

$$\lambda_1 \mathbf{u}_1 + \cdots + \lambda_r \mathbf{u}_r + \lambda_{r+1} \mathbf{u}_{r+1} + \cdots + \lambda_p \mathbf{u}_p = 0 \quad (1.111)$$

Pero estos p vectores son también linealmente independientes por constituir la base \mathcal{B}_L , por lo que los escalares $\lambda_1, \dots, \lambda_p$ son todos nulos. Queda así demostrado que la familia \mathcal{S} es base de $L+M$, y con ello, el Teorema 1.4.5.

□

Ejemplo 1.4.6 Considérense los subespacios $L = L[(1, 0, -1), (0, 1, 1)]$ y $M = L[(1, 2, 1)]$ de \mathbb{R}^3 . Un sistema generador del subespacio suma $L+M$ es:

$$\mathcal{G} = \{(1, 0, -1), (0, 1, 1), (1, 2, 1)\} = \{\bar{\mathbf{u}}_1, \bar{\mathbf{u}}_2, \bar{\mathbf{u}}_3\} \quad (1.112)$$

Como $\bar{\mathbf{u}}_3 = \bar{\mathbf{u}}_1 + 2\bar{\mathbf{u}}_2$, eliminando $\bar{\mathbf{u}}_3$ se obtiene una base de $L+M$:

$$\mathcal{B} = (\bar{\mathbf{u}}_1, \bar{\mathbf{u}}_2) = ((1, 0, -1), (0, 1, 1)) \quad (1.113)$$

Siendo $\dim L = 2$, $\dim M = 1$, y $\dim L+M = 2$. Aplicando *Grassmann*, Teorema 1.4.5, se obtiene:

$$\dim L \cap M = \dim L + M - \dim L - \dim M = 1 \quad (1.114)$$

El vector que se había eliminado de \mathcal{G} para obtener la base \mathcal{B} pertenece tanto a L como a M , por lo que forma un sistema generador de $L \cap M$ que además es base de ese subespacio: $((1, 2, 1))$.

Corolario 1.4.6 Dados dos subespacios vectoriales L, M en V que forman suma directa ($L \cup M = \{0\}$) se tiene que:

$$\dim L \oplus M = \dim L + \dim M$$

El Corolario 1.4.6 deriva de aplicar la Definición 1.4.2 de suma directa a la relación de *Grassmann*, Teorema 1.4.5, teniendo en cuenta que, por convenio, el subespacio nulo tiene dimensión cero, como se justificó en el apartado 1.3.7.

Ejemplo 1.4.7 Los subespacios A y B de los Ejemplos 1.4.3 y 1.4.4 tienen $\dim A = 2$ y $\dim B = 1$. Además, $\dim A \oplus B = 3$ y $\dim A \cap B = 0$, puesto que $\dim\{0\} = 0$. Puede comprobarse que se cumple la identidad (1.4.6):

$$3 = \dim L + M = \dim L + \dim M - \dim L \cap M = 2 + 1 - 0$$

1.4.8 Resumen

1. La intersección y la suma de subespacios producen siempre un subespacio. La unión lo es, si y sólo si, los subespacios están anidados.
2. Si la intersección de dos subespacios contiene sólo el elemento nulo de V , se dice que forman suma directa.
3. Si la suma directa de dos subespacios en V llena todo el espacio V en el que están contenidos, se dice que son suplementarios.
4. Todo elemento en V se puede descomponer en la suma de dos elementos pertenecientes a dos subespacios suplementarios en V , y la descomposición es única.
5. La dimensión de la suma de subespacios es igual a la suma de sus dimensiones menos la dimensión de su intersección por la relación de *Grassmann*.



Esta página está en blanco intencionalmente.

Matrices y determinantes

2.1 Definición y tipos de matrices

2.1.1 Introducción

Se había iniciado el Tema 1 con la definición general de espacio vectorial, aplicándola a los conjuntos de vectores \mathbb{K}^n y demostrando que estos, junto con la suma de vectores y el producto de escalar por vector, cumplen todas las propiedades de los espacios vectoriales. También se definieron conceptos característicos de los espacios vectoriales como el de base y el de subespacio, enunciando y demostrando sus propiedades elementales. Además, se vio que los conjuntos de matrices, junto con la suma y el producto por un escalar, cumplen todas las propiedades axiomáticas de los espacios vectoriales.

A continuación, se procederá a un estudio mucho más detallado de las matrices por su enorme importancia en el álgebra lineal. Tras definir formalmente qué es una matriz, se mostrarán sus tipos fundamentales y las operaciones elementales con matrices: la traza, la conjugación, la transposición, la suma de matrices, el producto de escalar por matriz y el producto matricial. Posteriormente, se definirá el determinante de una matriz cuadrada y el concepto de matriz invertible. Así como un método para el cálculo, si existe, de la inversa de una matriz cuadrada.

2.1.2 Matrices: definiciones fundamentales

Definición 2.1.1 – Definición de matriz

Sean m y n números enteros positivos, una matriz de tamaño $m \times n$ es una tabla rectangular formada por $m \cdot n$ escalares en el cuerpo \mathbb{K} que se disponen como se indica:

$$A = (a_{ij}) = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

Definición 2.1.2 – Elemento, fila, columna, elementos diagonales y diagonal principal de una matriz

Cada escalar $a_{ij} = (A)_{ij}$ se llama elemento, coeficiente o entrada de la matriz.

Se dice que el elemento a_{ij} está situado en la fila i y en la columna j de la matriz. A los elementos a_{ii} de igual índice de fila que de columna se les llama elementos diagonales y constituyen la diagonal principal de la matriz.

2.1	Definición y tipos de matrices	39
2.2	Operaciones con matrices	48
2.3	Definición y propiedades del determinante	55
2.4	Cálculo de determinantes. Interpretación geométrica	64
2.5	Inversa de una matriz cuadrada	71

Ejemplo 2.1.1 A es una matriz de elementos reales de tamaño 2×2 . Es decir, con $m = 2$ filas y $n = 2$ columnas:

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

Ejemplo 2.1.2 B es una matriz de elementos reales de tamaño 2×3 . Es decir, con $m = 2$ filas y $n = 3$ columnas:

$$B = \begin{pmatrix} 4 & 0 & 7 \\ -1/2 & 3 & 9 \end{pmatrix}$$

Ejemplo 2.1.3 C es una matriz de elementos complejos de tamaño 4×1 . Es decir, con $m = 4$ filas y $n = 1$ columnas:

$$C = \begin{pmatrix} 2 + (1/2)i \\ 3i \\ 4 \\ 0 \end{pmatrix}$$

Ejemplo 2.1.4 Algunos elementos de las matrices de los Ejemplos 2.1.1, 2.1.2 y 2.1.3 son:

$$a_{11} = (A)_{11} = 0 \quad (2.1)$$

$$a_{12} = (A)_{12} = 2 \neq -1 = a_{21} = (A)_{21} \quad (2.2)$$

$$b_{23} = (B)_{23} = 9 \quad (2.3)$$

$$c_{11} = (C)_{11} = 2 + i/2 \quad (2.4)$$

$$c_{31} = (C)_{31} = 4 \quad (2.5)$$

Definición 2.1.3 — Los conjuntos $\mathbb{K}^{m \times n}$

Se denota $\mathbb{K}^{m \times n}$ al conjunto de todas las matrices con m filas y n columnas con coeficientes en el cuerpo \mathbb{K} .

Ejemplo 2.1.5 La matriz A pertenece tanto a $\mathbb{R}^{2 \times 2}$ como a $\mathbb{C}^{2 \times 2}$:

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

Ejemplo 2.1.6 La matriz B pertenece tanto a $\mathbb{R}^{2 \times 3}$ como a $\mathbb{C}^{2 \times 3}$:

$$B = \begin{pmatrix} 4 & 0 & 7 \\ -1/2 & 3 & 9 \end{pmatrix}$$

Ejemplo 2.1.7 La matriz C pertenece al conjunto de matrices $\mathbb{C}^{4 \times 1}$:

$$C = \begin{pmatrix} 2 + (1/2)i \\ 3i \\ 4 \\ 0 \end{pmatrix}$$

Definición 2.1.4 – Filas y columnas de una matriz

El vector:

$$F_i(A) = (a_{i1}, a_{i2}, \dots, a_{in}) \quad (2.6)$$

Constituye la fila i -ésima de la matriz A , mientras que el vector:

$$C_j(A) = \begin{pmatrix} a_{1j} \\ a_{2j} \\ \dots \\ a_{mj} \end{pmatrix} \quad (2.7)$$

Constituye la columna j -ésima de la matriz A .

Ejemplo 2.1.8 La matriz:

$$B = \begin{pmatrix} 4 & 0 & 7 \\ -1/2 & 3 & 9 \end{pmatrix} \quad (2.8)$$

Contiene las filas:

$$F_1(B) = (4, 0, 7) \quad (2.9)$$

$$F_2(B) = (-1/2, 3, 9) \quad (2.10)$$

Mientras que, sus columnas son:

$$C_1(B) = \begin{pmatrix} 4 \\ -(1/2) \end{pmatrix} \quad (2.11)$$

$$C_2(B) = \begin{pmatrix} 0 \\ 3 \end{pmatrix} \quad (2.12)$$

$$C_3(B) = \begin{pmatrix} 7 \\ 9 \end{pmatrix} \quad (2.13)$$

Definición 2.1.5 – Vectores fila y columna

Si $A \in \mathbb{K}^{m \times n}$ tiene una única fila, $m = 1$, se dice que es un vector fila de n componentes, reales o complejos. Si $A \in \mathbb{K}^{m \times n}$ tiene una única columna, $n = 1$, se dice que es un vector columna, de m componentes reales o complejos.

Observación 2.1.1 Los vectores fila y columna se pueden sumar y multiplicar por un escalar como si fuesen vectores de \mathbb{K}^n , Definiciones 1.1.5 y 1.1.6 y las propiedades que cumplen son las mismas que los espacios \mathbb{K}^n , Proposiciones 1.1.1 y 1.1.2.

Ejemplo 2.1.9 La matriz A del Ejemplo 2.1.3 es una matriz fila, mientras que las columnas de la matriz (2.8) del Ejemplo 2.1.8 son vectores columna pertenecientes al conjunto $\mathbb{R}^{2 \times 1}$.

Definición 2.1.6 – Matrices cuadradas y rectangulares

Si $m = n$, se dice que la matriz es cuadrada de orden n . De lo contrario, se dice que es rectangular.

Ejemplo 2.1.10 La matriz A tiene $m = 2$ filas y $n = 2 = m$ columnas, por lo que es una matriz cuadrada:

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

Ejemplo 2.1.11 La matriz B tiene $m = 2$ filas y $n = 3$ columnas, por lo que es una matriz rectangular:

$$B = \begin{pmatrix} 4 & 0 & 7 \\ -1/2 & 3 & 9 \end{pmatrix}$$

Definición 2.1.7 – Matriz identidad

Una matriz cuadrada que tiene todos los elementos de la diagonal principal iguales a 1 y el resto nulos se llama matriz identidad, o matriz unidad, y se denota con la letra I .

$$I = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

Definición 2.1.8 – Matriz nula

Una matriz cuadrada que tiene todos sus elementos iguales a cero se llama matriz nula, o matriz neutra, y se denota con la letra O .

$$O = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}$$

2.1.3 Traza y transpuesta de una matriz**Definición 2.1.9 – Traza de una matriz cuadrada**

Dada una matriz $A = (a_{ij})$ cuadrada de orden n , se define su traza como la suma de los elementos de su diagonal principal:

$$\text{tr}(A) = \sum_{k=1}^n a_{kk}$$

Definición 2.1.10 – Transpuesta de una matriz

Dada una matriz $A = (a_{ij})$ de tamaño $m \times n$, se define su transpuesta como la matriz de tamaño $n \times m$ obtenida a partir de A cambiando filas por columnas:

$$(A^t)_{ij} = a_{ji} \quad \forall i = 1, \dots, m, j = 1, \dots, n$$

Ejemplo 2.1.12 La transpuesta de la matriz cuadrada A de orden dos:

$$A = \begin{pmatrix} 0 & 2 \\ -1 & 3 \end{pmatrix} \quad (2.14)$$

Se denota como A^t , mientras que su traza es $\text{tr}(A) = 0 + 3 = 3$.

$$A^t = \begin{pmatrix} 0 & -1 \\ 2 & 3 \end{pmatrix} \quad (2.15)$$

Matriz cuya traza es $\text{tr}(A^t) = 0 + 3 = 3$. †

Observación 2.1.2 † Las trazas de una matriz cuadrada y de su transpuesta coinciden siempre.

Ejemplo 2.1.13 La traza de la matriz B de tamaño 2×3 :

$$B = \begin{pmatrix} 4 & 0 & 7 \\ -1/2 & 3 & 9 \end{pmatrix} \quad (2.16)$$

Es la siguiente matriz, denotada B^t y de tamaño 3×2 :

$$B^t = \begin{pmatrix} 4 & 1/2 \\ 0 & 3 \\ 7 & 9 \end{pmatrix} \quad (2.17)$$

Ejemplo 2.1.14 El transpuesto del vector columna C :

$$C = \begin{pmatrix} 2 + (1/2)i \\ 3i \\ 4 \\ 0 \end{pmatrix} \quad (2.18)$$

Es el siguiente vector fila, denotado C^t :

$$C^t = (2 + \frac{i}{2}, 3i, 4, 0) \quad (2.19)$$

2.1.4 Opuesta, conjugada y transpuesta conjugada de una matriz

Definición 2.1.11 – Opuesta de una matriz

La opuesta de una matriz A de tamaño $m \times n$, denotada $-A$, es la matriz del mismo tamaño obtenida cambiando el signo a todos sus elementos:

$$(-A)_{ij} = -a_{ij} \quad \forall i = 1, \dots, m, j = 1, \dots, n$$

Definición 2.1.12 – Conjugada de una matriz

La conjugada de una matriz A de tamaño $m \times n$, denotada A^* , es la matriz del mismo tamaño cuyos elementos son los conjugados de los de A :

$$(A^*)_{ij} = a_{ij}^* \quad \forall i = 1, \dots, m, j = 1, \dots, n$$

Definición 2.1.13 – Transpuesta conjugada de una matriz

La transpuesta conjugada o adjunta de una matriz A de tamaño $m \times n$, denotada A^h , es la transpuesta de la conjugada de A , o de forma equivalente la conjugada de la transpuesta de A :

$$(A^h)_{ij} = a_{ji}^* \quad \forall i = 1, \dots, m, j = 1, \dots, n$$

Ejemplo 2.1.15 Sea la matriz A , de tamaño 2×3 :

$$A = \begin{pmatrix} 4 & -i & 3-i \\ 0 & (1/2) + i & \pi \end{pmatrix} \quad (2.20)$$

Su opuesta es la matriz $-A$, de tamaño 2×3 :

$$-A = \begin{pmatrix} -4 & i & -3+i \\ 0 & -(1/2) - i & -\pi \end{pmatrix} \quad (2.21)$$

Su conjugada es la matriz A^* , de tamaño 2×3 :

$$A^* = \begin{pmatrix} 4 & i & 3+i \\ 0 & (1/2) - i & \pi \end{pmatrix} \quad (2.22)$$

Su transpuesta conjugada es la matriz A^h , de tamaño 3×2 :

$$A^h = \begin{pmatrix} 4 & 0 \\ i & (1/2) - i \\ 3+i & \pi \end{pmatrix} \quad (2.23)$$

2.1.5 Matrices triangulares, trapezoidales y diagonales**Definición 2.1.14 – Matriz triangular y matriz diagonal**

Dada una matriz A cuadrada de orden n :

1. Se dice que es triangular superior si los elementos que se encuentran por debajo de la diagonal principal son todos nulos. Es decir, si $a_{ij} = 0, \forall i, j = 1, \dots, n, i > j$.
2. Se dice que es triangular inferior si los elementos que se encuentran por encima de la diagonal principal son todos nulos. Es decir, si $a_{ij} = 0, \forall i, j = 1, \dots, n, i < j$.
3. Se dice que es diagonal si es triangular superior e inferior simultáneamente. Es decir, si todos los elementos que se encuentran fuera de su diagonal principal son nulos.

Ejemplo 2.1.16 La matriz cuadrada C es a la vez triangular superior e inferior, por lo que es una matriz diagonal:

$$C = \begin{pmatrix} -i & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 1+i \end{pmatrix}$$

Definición 2.1.15 – Matriz trapezoidal

Dada una matriz A rectangular de tamaño $m \times n$, con $m \neq n$.

1. Se dice que es trapezoidal superior si los elementos que se encuentran por debajo de la diagonal principal son todos nulos. Es decir, si $a_{ij} = 0 \forall i, j = 1, \dots, n \quad i > j$.
2. Se dice que es trapezoidal inferior si los elementos que se encuentran por encima de la diagonal principal son todos nulos. Es decir, si $a_{ij} = 0 \forall i, j = 1, \dots, n \quad i < j$.

Ejemplo 2.1.17 La matriz rectangular A es trapezoidal superior:

$$A = \begin{pmatrix} 0 & 4 & -\pi & -7 \\ 0 & 3 & 0 & 3 \\ 0 & 0 & -2 & 4 \end{pmatrix}$$

Ejemplo 2.1.18 La matriz rectangular B es trapezoidal inferior:

$$B = \begin{pmatrix} -\pi & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 4 & 5 & 1/2 & 0 \end{pmatrix}$$

2.1.6 Simetría y hermiticidad de matrices

Definición 2.1.16 – Matrices simétricas, antisimétricas, hermíticas y antihermíticas

Dada una matriz A cuadrada de orden n :

1. Se dice que es simétrica si $A^t = A$.

$$a_{ji} = a_{ij} \quad \forall i, j = 1, \dots, n \quad (2.24)$$

2. Se dice que es antisimétrica si $A^t = -A$.

$$a_{ji} = -a_{ij} \quad \forall i, j = 1, \dots, n \quad (2.25)$$

3. Se dice que es hermítica si $A^h = A$.

$$a_{ji} = a_{ij}^* \quad \forall i, j = 1, \dots, n \quad (2.26)$$

Observación 2.1.3 En el cuerpo de los números reales, los conjuntos de las matrices hermíticas y antihermíticas coinciden con los de las matrices simétricas y antisimétricas, pues, por la Definición C.5.1 del conjugado de un número complejo, $z = z^* \Leftrightarrow z \in \mathbb{R}, \forall z \in \mathbb{C}$.

4. Se dice que es antihermítica si $A^h = -A$.

$$a_{ji} = -a_{ij}^* \quad \forall i, j = 1, \dots, n \quad (2.27)$$

Ejemplo 2.1.19 La matriz cuadrada A es simétrica:

$$A = \begin{pmatrix} 4 & 1 & -2 \\ 1 & 3 & 5 \\ -2 & 5 & 0 \end{pmatrix} \quad (2.28)$$

Puesto que:

$$A^t = \begin{pmatrix} 4 & 1 & -2 \\ 1 & 3 & 5 \\ -2 & 5 & 0 \end{pmatrix} = A \quad (2.29)$$

Ejemplo 2.1.20 La matriz cuadrada B es antisimétrica:

$$B = \begin{pmatrix} 0 & 1 & -2 \\ -1 & 0 & 5 \\ 2 & -5 & 0 \end{pmatrix} \quad (2.30)$$

Puesto que:

$$B^t = \begin{pmatrix} 0 & -1 & 2 \\ 1 & 0 & -5 \\ -2 & 5 & 0 \end{pmatrix} = -B \quad (2.31)$$

Ejemplo 2.1.21 La matriz cuadrada C es hermítica:

$$C = \begin{pmatrix} 0 & i & 1+i \\ -i & 4 & 2-i \\ 1-i & 2+i & 3 \end{pmatrix} \quad (2.32)$$

Puesto que:

$$C^h = \begin{pmatrix} 0 & i & 1+i \\ -i & 4 & 2-i \\ 1-i & 2+i & 3 \end{pmatrix} = C \quad (2.33)$$

Ejemplo 2.1.22 La matriz cuadrada D es antihermítica:

$$D = \begin{pmatrix} i & i & 1+i \\ i & 2i & 2-i \\ -1+i & -2-i & -i \end{pmatrix} \quad (2.34)$$

Puesto que:

$$D^h = \begin{pmatrix} -i & -i & -1-i \\ -i & -2i & -2+i \\ 1-i & 2+i & i \end{pmatrix} = -D \quad (2.35)$$

La siguiente proposición muestra cuatro propiedades muy sencillas que cumplen las matrices hermíticas y antihermíticas.

Proposición 2.1.1 — Propiedades de las matrices hermíticas y antihermíticas

Sea $A \in \mathbb{K}^{n \times n}$ una matriz cuadrada hermítica.

1. Los elementos presentes en su diagonal principal son todos reales: $a_{ii} \in \mathbb{R} \forall i = 1, \dots, n$.
2. Las partes reales e imaginarias de sus elementos cumplen:

$$\operatorname{Re} a_{ij} = \operatorname{Re} a_{ji} \quad (2.36)$$

$$\operatorname{Im} a_{ij} = -\operatorname{Im} a_{ji} \quad (2.37)$$

$$\forall i, j = 1, \dots, n$$

Sea $A \in \mathbb{K}^{n \times n}$ una matriz cuadrada antihermítica.

3. Los elementos presentes en su diagonal principal son todos imaginarios puros: $\operatorname{Im} a_{ii} = 0 \forall i = 1, \dots, n$. Si se trata de una matriz real antisimétrica, estos son todos nulos.
4. Las partes reales e imaginarias de sus elementos cumplen:

$$\operatorname{Re} a_{ij} = -\operatorname{Re} a_{ji} \quad (2.38)$$

$$\operatorname{Im} a_{ij} = \operatorname{Im} a_{ji} \quad (2.39)$$

$$\forall i, j = 1, \dots, n$$

Nótese que las matrices de los Ejemplos 2.1.19 hasta el 2.1.22 cumplen estas propiedades.

De esta proposición se derivan dos condiciones necesarias que toda matriz cuadrada ha de cumplir para ser hermítica: los puntos 1 y 2; y dos condiciones necesarias que toda matriz cuadrada ha de cumplir para ser antihermítica: los puntos 3 y 4.

Demostración:

Por aplicación directa de la propiedad 6 de existencia de opuesto de un número complejo, Proposición C.3.1, y la Definición C.5.1 de conjugado de un número complejo:

$$z = z^* \Leftrightarrow z \in \mathbb{R} \quad (2.40)$$

$$z = -z^* \Leftrightarrow \operatorname{Re} z = 0 \quad (2.41)$$

Los elementos en la diagonal principal de una matriz hermítica cumplen, por la Definición 2.26, $a_{ii} = a_{ii}^*$. Por tanto, estos han de ser coeficientes reales. Los elementos en la diagonal principal de una matriz antihermítica cumplen, por la Definición 2.27, $a_{ii} = -a_{ii}^*$. Por tanto, estos han de tener parte real nula.

□

2.2 Suma de matrices, producto de matriz por escalar y producto matricial

2.2.1 Introducción

Como se había indicado en la Definición 1.1.7, los conjuntos de matrices, junto con las operaciones de suma de matrices y producto de matriz por escalar, cumplen todas las propiedades axiomáticas de los espacios vectoriales. Habiendo tomado la matriz nula, Definición 2.1.8, como elemento neutro de la suma de matrices.

Los conjuntos $\mathbb{K}^{m \times n}$ constituyen espacios vectoriales, al igual que los conjuntos \mathbb{K}^n de vectores de n componentes en el cuerpo \mathbb{K} . En esta sección, se presentan definiciones, ejemplos y propiedades de estas dos operaciones y de una tercera operación entre matrices denominada “producto matricial”.

2.2.2 Espacios vectoriales de matrices

Definición 2.2.1 — Suma de matrices

Sea $\mathbb{K}^{m \times n}$ el conjunto de todas las matrices de m filas y n columnas con elementos en el cuerpo \mathbb{K} . La suma de dos matrices A y B en $\mathbb{K}^{m \times n}$ se define, elemento por elemento, como:

$$(A + B)_{ij} = a_{ij} + b_{ij}$$

Definición 2.2.2 — Producto de una matriz por un escalar

Sea $\mathbb{K}^{m \times n}$ el conjunto de todas las matrices de m filas y n columnas con elementos en el cuerpo \mathbb{K} . La multiplicación de un escalar $\lambda \in \mathbb{K}$ por la matriz $A \in \mathbb{K}^{m \times n}$ se define, elemento por elemento, como:

$$(\lambda \cdot A)_{ij} = \lambda \cdot a_{ij}$$

Observación 2.2.1 Como ocurre con el producto de un escalar por un vector, el símbolo “ \cdot ” suele omitirse por simplicidad de notación.

Ejemplo 2.2.1 Dadas las matrices A y B y el escalar λ :

$$A = \begin{pmatrix} 3 & -2 \\ 0 & 4 \end{pmatrix} \quad B = \begin{pmatrix} -1 & 0 \\ 7 & -2 \end{pmatrix} \quad \lambda = 1/3 \quad (2.42)$$

Se tiene que:

$$A + B = \begin{pmatrix} 3-1 & -2+0 \\ 0+7 & 4-2 \end{pmatrix} = \begin{pmatrix} 2 & -2 \\ 7 & 2 \end{pmatrix} \quad (2.43)$$

$$\lambda \cdot A = \begin{pmatrix} 1/3 \cdot 3 & 1/3 \cdot (-2) \\ 1/3 \cdot 0 & 1/3 \cdot 4 \end{pmatrix} = \begin{pmatrix} 1 & -2/3 \\ 0 & 4/3 \end{pmatrix} \quad (2.44)$$

$$\lambda \cdot B = \begin{pmatrix} 1/3 \cdot (-1) & 1/3 \cdot 0 \\ 1/3 \cdot 7 & 1/3 \cdot (-2) \end{pmatrix} = \begin{pmatrix} -1/3 & 0 \\ 7/3 & -2/3 \end{pmatrix} \quad (2.45)$$

La siguiente proposición muestra las diez propiedades axiomáticas de los espacios vectoriales [(1.1) ... (1.10)] reescritas para adaptarlas a las operaciones de suma de matrices y producto de escalar por matriz, Definiciones 2.2.1 y 2.2.2.

Proposición 2.2.1 — Propiedades de la suma de matrices y el producto por un escalar

Sean los espacios vectoriales constituidos por la cuaterna $(\mathbb{K}^{m \times n}, \mathbb{K}, +, \cdot)$, donde m denota el número de filas de las matrices, n su número de columnas, “+” la suma de matrices, y “ \cdot ” el producto de escalar por matriz.

Las operaciones “+” y “ \cdot ” cumplen estas propiedades:

1. Clausura de la suma de matrices:

$$\forall A, B \in \mathbb{K}^{m \times n}, A + B \in \mathbb{K}^{m \times n} \quad (2.46)$$

2. Propiedad asociativa de la suma de matrices:

$$\forall A, B, C \in \mathbb{K}^{m \times n}, (A + B) + C = A + (B + C) \quad (2.47)$$

3. Existencia del elemento neutro, Definición 2.1.8: †

$$\exists O \in \mathbb{K}^{m \times n} : \forall A \in \mathbb{K}^{m \times n}, A + O = O + A = A \quad (2.48)$$

4. Existencia del elemento simétrico, Definición 2.1.11: †

$$\forall A \in \mathbb{K}^{m \times n}, \exists -A \in \mathbb{K}^{m \times n} : A + (-A) = -A + A = O \quad (2.49)$$

5. Propiedad conmutativa de la suma de matrices:

$$\forall A, B \in \mathbb{K}^{m \times n}, A + B = B + A \quad (2.50)$$

6. Clausura del producto de una matriz por un escalar:

$$\forall \lambda \in \mathbb{K}, \forall A \in \mathbb{K}^{m \times n}, \lambda \cdot A \in \mathbb{K}^{m \times n} \quad (2.51)$$

7. Propiedad distributiva del producto de una matriz por un escalar respecto a la suma de escalares:

$$\forall A \in \mathbb{K}^{m \times n}, \forall \lambda, \mu \in \mathbb{K}, (\lambda + \mu) \cdot A = \lambda \cdot A + \mu \cdot A \quad (2.52)$$

8. Propiedad distributiva del producto de una matriz por un escalar respecto a la suma de matrices:

$$\forall A, B \in \mathbb{K}^{m \times n}, \forall \lambda \in \mathbb{K}, \lambda \cdot (A + B) = \lambda \cdot A + \lambda \cdot B \quad (2.53)$$

9. Propiedad asociativa del producto de una matriz por escalares:

$$\forall A \in \mathbb{K}^{m \times n}, \forall \lambda, \mu \in \mathbb{K}, (\lambda\mu) \cdot A = \lambda \cdot (\mu \cdot A) \quad (2.54)$$

10. Producto de una matriz por el escalar unidad:

$$\forall A \in \mathbb{K}^{m \times n}, 1 \cdot A = A \quad (2.55)$$

Observación 2.2.2 † El símbolo “:” que aparece en (2.48) y (2.49) se lee como la conjunción “tal que”, Observación A.1.1.

De la misma forma, las propiedades de los espacios vectoriales contenidas en la Proposición 1.1.3 pueden reescribirse para adaptarlas a los espacios de matrices $\mathbb{K}^{m \times n}$:

Proposición 2.2.2 — Propiedades adicionales de la suma de matrices y el producto por un escalar

1. La matriz nula O es única en cada conjunto $\mathbb{K}^{m \times n}$.
2. La matriz opuesta $-A$ es única para cada $A \in \mathbb{K}^{m \times n}$.
3. $\forall \lambda \in \mathbb{K}, \lambda \cdot O = O$
4. $\forall A \in \mathbb{K}^{m \times n}, 0 \cdot A = O$
5. $\forall \lambda \in \mathbb{K}, \forall A \in \mathbb{K}^{m \times n}, \lambda \cdot A = O \Leftrightarrow (\lambda = 0 \vee A = O)$
6. $\dagger \forall \lambda \in \mathbb{K}, \forall A \in \mathbb{K}^{m \times n}, (-\lambda) \cdot A = \lambda \cdot (-A) = -(\lambda \cdot A)$
7. $\forall \lambda, \mu \in \mathbb{K}, \forall A \in \mathbb{K}^{m \times n}, (\lambda \cdot A = \mu \cdot A \wedge A \neq O) \Rightarrow \lambda = \mu$
8. $\forall \lambda \in \mathbb{K}, \forall A, B \in \mathbb{K}^{m \times n}, (\lambda \cdot A = \lambda \cdot B \wedge \lambda \neq O) \Rightarrow A = B$

Observación 2.2.3 \dagger Esta propiedad, junto con el punto 10 de la Proposición 2.2.1, permite definir la resta de matrices como $A - B = A + (-1) \cdot B = A + (-B) \forall A, B \in \mathbb{K}^{m \times n}$.

2.2.3 Producto matricial

A continuación, se define una nueva operación llamada “producto matricial” que puede realizarse entre matrices de distintas dimensiones, bajo ciertas condiciones.

Definición 2.2.3 — Producto matricial

Dadas dos matrices $A \in \mathbb{K}^{m \times n}$ y $B \in \mathbb{K}^{n \times r}$, se define su producto AB como aquella matriz $C \in \mathbb{K}^{m \times r}$ cuyos elementos son iguales a:

$$c_{ij} = \sum_{k=1}^n a_{ik} b_{kj} \quad \forall i = 1, 2, \dots, m, \quad j = 1, 2, \dots, r$$

Cada elemento c_{ij} de la matriz producto se obtiene sumando el producto de los elementos de la fila i de la primera matriz por los elementos de la columna j de la segunda.

Como consecuencia de la definición del producto matricial, la condición, necesaria y suficiente, que dos matrices deben cumplir para que se puedan multiplicar es que el número de columnas de la primera coincida con el número de filas de la segunda.

Nótese que si dos matrices A y B cumplen $m \neq r$, el producto AB puede realizarse, mientras que el producto BA no puede realizarse.

Ejemplo 2.2.2 Dadas las matrices A y B :

$$A = \begin{pmatrix} 3 & -1 & 0 \\ 4 & 2 & -4 \end{pmatrix} \in \mathbb{R}^{2 \times 3} \quad B = \begin{pmatrix} 1 & 5 \\ 0 & -7 \\ -3 & 0 \end{pmatrix} \in \mathbb{R}^{3 \times 2} \quad (2.56)$$

Su producto AB es una matriz real con 2 filas y 2 columnas. Siendo sus elementos los siguientes:

$$(AB)_{11} = a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31} \quad (2.57)$$

$$(AB)_{12} = a_{11}b_{12} + a_{12}b_{22} + a_{13}b_{32} \quad (2.58)$$

$$(AB)_{21} = a_{21}b_{11} + a_{22}b_{21} + a_{23}b_{31} \quad (2.59)$$

$$(AB)_{22} = a_{21}b_{12} + a_{22}b_{22} + a_{23}b_{32} \quad (2.60)$$

Sustituyendo y operando:

$$\begin{aligned} AB &= \begin{pmatrix} 3 \cdot 1 - 1 \cdot 0 - 3 \cdot 0 & 3 \cdot 5 - 1 \cdot (-7) + 0 \cdot 0 \\ 4 \cdot 1 + 2 \cdot 0 - 4 \cdot (-3) & 4 \cdot 5 + 2 \cdot (-7) - 4 \cdot 0 \end{pmatrix} = \\ &= \begin{pmatrix} 3 & 22 \\ 16 & 6 \end{pmatrix} \end{aligned} \quad (2.61)$$

2.2.4 Propiedades del producto matricial

Proposición 2.2.3 – Propiedades elementales del producto matricial

1. En general, el producto matricial no es conmutativo, ni siquiera entre matrices cuadradas de igual orden.

Dadas tres matrices A , B y C cualesquiera cuyas dimensiones sean compatibles con el producto matricial, se tiene que:

2. El producto matricial es asociativo.

$$(AB)C = A(BC) \quad (2.62)$$

3. El producto de matrices es distributivo respecto de la suma a izquierda y derecha. Es decir:

$$A(B+C) = AB+AC \quad (2.63)$$

$$(B+C)A = BA+CA \quad (2.64)$$

4. Sean A y B dos matrices para las que el producto AB tenga sentido, entonces:

$$(AB)^t = B^t A^t \quad (2.65)$$

$$(AB)^h = B^h A^h \quad (2.66)$$

Ejercicio 2.2.1 † Demostrar, aplicando la Definición 2.2.3 de producto matricial, las propiedades de la Proposición 2.2.3. Pista: para demostrar que el producto matricial no es conmutativo, basta con encontrar dos matrices cualesquiera A y B tales que $AB \neq BA$. A esto se le conoce como “demostración por contraejemplo”, apartado A.3.5.

Demostración:

Se demuestra únicamente la propiedad 4 para el caso de la transpuesta, ecuación (2.65). Las demostraciones de las propiedades restantes se dejan como ejercicios † para el lector.

Aplicando la Definición 2.2.3 de producto matricial y la Definición 2.1.10 de matriz transpuesta, se obtiene:

$$\begin{aligned} ((AB)^t)_{ij} &= (AB)_{ji} = \sum_{k=1}^n a_{jk}b_{ki} = \sum_{k=1}^n b_{ki}a_{jk} = \\ &= \sum_{k=1}^n (B)_{ki} (A)_{jk} = \sum_{k=1}^n (B^t)_{ik} (A^t)_{kj} = (B^t A^t)_{ij} \end{aligned} \quad (2.67)$$

□

Supóngase ahora que se tiene un producto, no de dos sino de tres matrices cuyas dimensiones hacen viable el producto matricial AB_1B_2 . Llamando B al producto B_1B_2 y aplicando la Proposición 2.2.3, ecuación (2.65), se obtiene:

$$(AB_1B_2)^t = AB^t = B^t A^t = (B_1B_2)^t A^t = B_2^t B_1^t A^t \quad (2.68)$$

Puede demostrarse por inducción, apartado A.3.2, que para todo número k de matrices, su transpuesta, o transpuesta conjugada, es igual al producto de las transpuestas, o transpuestas conjugadas, de las matrices en orden inverso.

Proposición 2.2.4 La inversa del producto de k matrices de dimensiones compatibles cumple:

$$(A_1 A_2 \dots A_k)^t = A_k^t A_{k-1}^t \dots A_1^t \quad (2.69)$$

$$(A_1 A_2 \dots A_k)^h = A_k^h A_{k-1}^h \dots A_1^h \quad (2.70)$$

Ejemplo 2.2.3 La transpuesta de la matriz A:

$$A = \begin{pmatrix} 3 & -1 & 0 \\ 4 & 2 & -4 \end{pmatrix} \in \mathbb{R}^{2 \times 3} \quad (2.71)$$

Es la matriz A^t :

$$A^t = \begin{pmatrix} 3 & 4 \\ -1 & 2 \\ 0 & -4 \end{pmatrix} \in \mathbb{R}^{3 \times 2} \quad (2.72)$$

Y la transpuesta de la matriz B:

$$B = \begin{pmatrix} 1 & 5 \\ 0 & -7 \\ -3 & 0 \end{pmatrix} \in \mathbb{R}^{3 \times 2} \quad (2.73)$$

Es la matriz B^t :

$$B^t = \begin{pmatrix} 1 & 0 & -3 \\ 5 & -7 & 0 \end{pmatrix} \in \mathbb{R}^{2 \times 3} \quad (2.74)$$

El producto $B^t A^t$ es igual a:

$$B^t A^t = \begin{pmatrix} 1 & 0 & -3 \\ 5 & -7 & 0 \end{pmatrix} \begin{pmatrix} 3 & 4 \\ -1 & 2 \\ 0 & -4 \end{pmatrix} = \begin{pmatrix} 3 & 16 \\ 22 & 6 \end{pmatrix} = (AB)^t \quad (2.75)$$

Las filas y columnas de la matriz C obtenida al multiplicar dos matrices con dimensiones compatibles son combinaciones lineales de las filas y columnas de las matrices multiplicadas, tal y como indican las siguientes propiedades. Su demostración se deja como ejercicio † para el lector.

Proposición 2.2.5 – Propiedades adicionales del producto matricial

Sean $A \in \mathbb{K}^{m \times n}$ y $B \in \mathbb{K}^{n \times r}$, y denotando la columna k -ésima de una matriz como C_k y la fila k -ésima como F_k :

1. Las columnas de AB son combinaciones lineales de las columnas de A , con coeficientes iguales a los elementos de cada columna de B .

$$C_k(AB) = AC_k(B) \quad (2.76)$$

2. Las filas de AB son combinaciones lineales de las filas de B , con coeficientes iguales a los elementos de cada fila de A .

$$F_k(AB) = F_k(A)B \quad (2.77)$$

3. La matriz AB puede expresarse como suma de las matrices resultantes de multiplicar cada columna de A por cada fila de A .

$$AB = \sum_{k=1}^n C_k(A)F_k(B) \quad (2.78)$$

Ejercicio 2.2.2 † Demostrar estas propiedades utilizando la Definición 2.2.3 de producto matricial.

Ejemplo 2.2.4 Dadas las matrices A y B del Ejemplo 2.2.2:

1. La primera columna de AB puede escribirse como combinación lineal de las columnas de A , con coeficientes iguales a los elementos de la primera columna de B .

$$\begin{aligned} AC_1(B) &= \begin{pmatrix} 3 & -1 & 0 \\ 4 & 2 & -4 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ -3 \end{pmatrix} = 1 \cdot \begin{pmatrix} 3 \\ 4 \end{pmatrix} + \\ &+ 0 \cdot \begin{pmatrix} -1 \\ 2 \end{pmatrix} - 3 \cdot \begin{pmatrix} 0 \\ -4 \end{pmatrix} = \begin{pmatrix} 3 \\ 16 \end{pmatrix} = C_1(AB). \end{aligned} \quad (2.79)$$

2. La primera fila de AB puede escribirse como combinación lineal de las filas de B , con coeficientes iguales a los elementos de la primera fila de A .

$$\begin{aligned} F_1(A)B &= (3 \quad -1 \quad 0) \begin{pmatrix} 1 & 5 \\ 0 & -7 \\ -3 & 0 \end{pmatrix} = 3 \cdot (1 \quad 5) - \\ &- 1 \cdot (0 \quad -7) + 0 \cdot (-3 \quad 0) = (-1 \quad 0) = F_1(AB) \end{aligned} \quad (2.80)$$

3. La matriz AB puede descomponerse en la suma de las columnas de A por las filas de B .

$$AB = \begin{pmatrix} 3 \\ 4 \end{pmatrix} (1 \quad 5) + \begin{pmatrix} -1 \\ 2 \end{pmatrix} (0 \quad -7) + \begin{pmatrix} 0 \\ -4 \end{pmatrix} (-3 \quad 0) \quad (2.81)$$

2.2.5 Resumen

1. La suma de matrices y el producto de una matriz por un escalar cumplen todas las propiedades de los espacios vectoriales.
2. El producto matricial no es, en general, conmutativo, pero cumple las propiedades asociativa y distributiva.
3. La transpuesta, o transpuesta conjugada, de un producto de matrices con dimensiones compatibles es igual al producto de las transpuestas, o transpuestas conjugadas, de las matrices multiplicadas en orden inverso.
4. Las columnas de la matriz C resultante de multiplicar dos matrices compatibles son combinaciones lineales de las columnas de la primera matriz del producto, mientras que las filas de C son combinaciones lineales de las filas de la segunda matriz.

2.3 Definición y propiedades del determinante de una matriz cuadrada

2.3.1 Introducción

En este apartado, se definirá un operador llamado “determinante” que asocia cada matriz cuadrada con un escalar real o complejo. Las propiedades de este operador son muy útiles en aplicaciones de álgebra lineal como el estudio y la resolución de sistemas lineales, Tema 4, y la diagonalización de matrices, Tema 6.

En la Definición 2.3.3 del determinante aparecen permutaciones de las columnas o filas de la matriz. Por ello, se comenzará la sección definiendo el concepto de permutación.

2.3.2 Permutaciones de índices

Definición 2.3.1 – Permutación de índices

Una permutación de un conjunto ordenado de índices $(1, 2, 3, \dots, n)$ es una operación consistente en reordenar los elementos del conjunto sin omisiones ni repeticiones.

$$(1, 2, 3, \dots, n) \mapsto (m_1, m_2, m_3, \dots, m_n)$$

Toda permutación es descomponible en una secuencia de intercambios de pares de índices, también llamados “inversiones”.

Ejemplo 2.3.1 La permutación de índices $(1, 2, 3, 4) \mapsto (3, 2, 4, 1)$ puede descomponerse en dos intercambios entre pares de índices:

$$(1, 2, 3, 4) \mapsto (3, 2, 1, 4) \mapsto (3, 2, 4, 1) \quad (2.82)$$

En el primer paso se intercambian los índices 1 y 3, y en el segundo paso los índices 1 y 4.

Aplicando por columnas la permutación de índices $(1, 2, 3, 4) \mapsto (3, 2, 4, 1)$ a la matriz A :

$$A = (A_1, A_2, A_3, A_4) = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{pmatrix} \quad (2.83)$$

Se obtiene la matriz:

$$(A_3, A_2, A_4, A_1) = \begin{pmatrix} a_{13} & a_{12} & a_{14} & a_{11} \\ a_{23} & a_{22} & a_{24} & a_{21} \\ a_{33} & a_{32} & a_{34} & a_{31} \\ a_{43} & a_{42} & a_{44} & a_{41} \end{pmatrix} \quad (2.84)$$

Una permutación de índices puede ser par o impar en función del número mínimo de inversiones de índices necesarias para efectuarla.

Definición 2.3.2 – Permutaciones pares e impares

Una permutación se dice par si el número mínimo de inversiones de índices necesarias para efectuarla es par, e impar si el número mínimo de inversiones de índices necesarias para efectuarla es impar.

Ejemplo 2.3.2 La permutación de índices del Ejemplo 2.3.1, ecuación (2.82), es par, pues el número mínimo de intercambios de índices necesarios para efectuarlo es dos, un número par.

2.3.3 Definición de determinante**Definición 2.3.3 – Determinante de una matriz cuadrada**

Dada una matriz cuadrada $A \in \mathbb{K}^{n \times n}$ cuyas columnas se denotan A_1, A_2, \dots, A_n :

$$A = (A_1, A_2, \dots, A_n) = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \quad (2.85)$$

Se define el determinante de A como un operador $\mathbb{K}^{n \times n} \mapsto \mathbb{K}$ que asocia a cada matriz A un único escalar mediante la fórmula:

$$\begin{aligned} \det(A) = |A| &= \det(A_1, A_2, \dots, A_n) = \\ &= \sum (-1)^p \cdot a_{1m_1} a_{2m_2} \cdots a_{nm_n}. \end{aligned} \quad (2.86)$$

Donde:

1. (m_1, m_2, \dots, m_n) es una permutación del vector de índices $(1, 2, \dots, n)$.
2. $a_{1m_1} a_{2m_2} \cdots a_{nm_n}$ es el producto de los elementos de la diagonal principal de la matriz que resulta de permutar las columnas de la matriz A de acuerdo con la permutación de índices (m_1, m_2, \dots, m_n) . †
3. El exponente p es el número mínimo de inversiones necesarias para llevar a cabo la permutación (m_1, m_2, \dots, m_n) .
4. El sumatorio es sobre todas las permutaciones posibles de las columnas de A . ‡

Observación 2.3.1 † A este producto de n elementos de una matriz A pertenecientes cada uno a una fila y columna distintas se le suele llamar “producto elemental de coeficientes de A ”.

Observación 2.3.2 ‡ El número de permutaciones posibles de un conjunto de n índices es igual a $n! = 1 \cdot 2 \cdot \dots \cdot n$. Por lo que la fórmula que define el determinante de una matriz cuadrada de orden n , ecuación (2.86), contiene $n!$ términos. La expresión $n!$ se lee “ n factorial”.

Cada uno de los términos del sumatorio de la ecuación (2.86) consiste en un producto elemental de n elementos de la matriz multiplicado por $+1$, si la permutación de columnas es par, o por -1 , si la permutación es impar.

Ejemplo 2.3.3 † Sea una matriz cuadrada cualquiera de orden dos:

$$A = (A_1, A_2) = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \quad (2.87)$$

Existen sólo dos permutaciones posibles de sus columnas: la permutación $(1, 2) \mapsto (1, 2)$, compuesta por $p = 0$ intercambios de índices, y la permutación $(1, 2) \mapsto (2, 1)$, compuesta por $p = 1$ intercambio de índices.

Aplicando la permutación $(1, 2) \mapsto (1, 2)$ sobre las columnas de A , se obtiene la matriz:

$$(A_1, A_2) = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \quad (2.88)$$

Aplicando la permutación $(1, 2) \mapsto (2, 1)$ sobre las columnas A , se obtiene la matriz:

$$(A_2, A_1) = \begin{pmatrix} a_{12} & a_{11} \\ a_{22} & a_{21} \end{pmatrix} \quad (2.89)$$

El producto de los elementos en la diagonal principal de la primera matriz permutada (A_1, A_2) es:

$$a_{11}a_{22} \quad (2.90)$$

Mientras que el producto de los elementos en la diagonal principal de la segunda matriz permutada (A_2, A_1) es:

$$a_{12}a_{21} \quad (2.91)$$

Estos son los $2! = 1 \cdot 2 = 2$ productos elementales de A . La permutación de columnas del primero es par, y la permutación del segundo es impar.

Aplicando la definición de determinante, ecuación (2.86), resulta:

$$\det(A) = (-1)^0 \cdot a_{11}a_{22} + (-1)^1 \cdot a_{12}a_{21} = a_{11}a_{22} - a_{12}a_{21} \quad (2.92)$$

Ejercicio 2.3.1 † Utilizar la definición de determinante, ecuación (2.86), como en el Ejemplo 2.3.3 para derivar la fórmula del determinante de una matriz cuadrada de orden tres, conocida como "fórmula de Sarrus".

2.3.4 Propiedades del determinante

A continuación, se enumeran las cuatro propiedades fundamentales que se derivan de la Definición 2.3.3 del operador determinante.

Proposición 2.3.1 - Propiedades elementales del determinante

Sea A una matriz cuadrada cualquiera de orden n , su determinante cumple:

1. Propiedad de transposición: transponer la matriz no cambia el valor de su determinante.

$$\det(A^t) = \det(A) \quad (2.93)$$

2. Antisimetría: intercambiar dos columnas cambia el signo del determinante. Si $1 \leq i < j \leq n$:

$$\begin{aligned} \det(A_1, \dots, A_i, \dots, A_j, \dots, A_n) &= \\ &= -\det(A_1, \dots, A_j, \dots, A_i, \dots, A_n). \end{aligned} \quad (2.94)$$

3. Determinante de un producto de matrices: Dadas k matrices cuadradas de orden n , A_1, A_2, \dots, A_k , se cumple:

$$\det(A_1 A_2 \dots A_k) = \det(A_1) \det(A_2) \dots \det(A_k) \quad (2.95)$$

4. Multilinealidad: si una columna es combinación lineal de dos vectores columna:

$$A_i = \alpha \cdot B_i + \beta \cdot C_i \quad \alpha, \beta \in \mathbb{K} \quad (2.96)$$

Entonces:

$$\begin{aligned} \det(A_1, \dots, \alpha \cdot B_i + \beta \cdot C_i, \dots, A_n) &= \\ &= \alpha \cdot \det(A_1, \dots, B_i, \dots, A_n) + \\ &+ \beta \cdot \det(A_1, \dots, C_i, \dots, A_n) \end{aligned} \quad (2.97)$$

Demostración:

1. Aplicar la Definición 2.1.10 de transpuesta de una matriz cuadrada:

$$(A^t)_{ij} = (A)_{ji} \quad \forall i, j = 1, \dots, n \quad (2.98)$$

al sumatorio (2.86) es tan simple como intercambiar los índices:

$$\det(A^t) = \sum (-1)^p \cdot a_{m_1 1} a_{m_2 2} \dots a_{m_n n} \quad (2.99)$$

Reordenando los factores en cada término de la suma de modo que el primer índice se disponga ordenadamente desde 1 hasta n se obtiene:

$$\det(A^t) = \sum (-1)^p \cdot a_{1m'_1} a_{2m'_2} \dots a_{nm'_n} \quad (2.100)$$

donde (m'_1, \dots, m'_n) es una permutación de los segundos índices de los factores de cada término del sumatorio, no necesariamente coincidente con la permutación (m_1, \dots, m_n) inicial.

Resulta que en cada término de (2.100) aparece una permutación (m'_1, \dots, m'_n) distinta, lo que implica que:

$$\begin{aligned} \det(A^t) &= \sum (-1)^p \cdot a_{1m'_1} a_{2m'_2} \dots a_{nm'_n} = \\ &= \sum (-1)^p \cdot a_{1m_1} a_{2m_2} \dots a_{nm_n} = \det(A) \end{aligned} \quad (2.101)$$

Demostrándose así que el determinante de toda matriz cuadrada coincide con el de su transpuesta.

2. Si se intercambian dos componentes cualesquiera de una permutación de índices par, esta se transforma en una permutación impar, y viceversa, Definición 2.3.2.

Si se intercambiasen dos columnas o filas de la matriz, las permutaciones pares de columnas se convertirían en permutaciones impares, y viceversa; por lo que el factor $(-1)^p$ que aparece en cada

término del sumatorio de la definición del determinante (2.86) cambiaría de signo en cada uno de sus términos, provocando un cambio de signo en el determinante de la matriz.

- Recuérdese que las columnas de AB son, por la propiedad 1 de la Proposición 2.2.5, combinaciones lineales de las columnas de A . Denotando las columnas de B como B_i , $i = 1, \dots, n$, esto implica que el producto AB puede escribirse como

$$AB = (AB_1 | AB_2 | \dots | AB_n) \quad (2.102)$$

Aplicando la definición del determinante (2.86) a la expresión (2.102):

$$\det(AB) = \sum (-1)^p \cdot (AB_{m_1})_1 (AB_{m_2})_2 \dots (AB_{m_n})_n \quad (2.103)$$

Por la propiedad 2 de la Proposición 2.2.5, las filas de AB son combinaciones lineales de las de B . Así, denotando las filas de A como A_i , $i = 1, \dots, n$, el sumatorio (2.102) puede escribirse como:

$$\det(AB) = \sum (-1)^p \cdot A_1 B_{m_1} A_2 B_{m_2} \dots A_n B_{m_n} \quad (2.104)$$

Esta última suma se puede descomponer en el siguiente producto de sumatorios:

$$\det(AB) = \left(\sum (-1)^p \cdot A_{m_1 1} \dots A_{m_n n} \right) \cdot \left(\sum (-1)^p \cdot B_{1 m_1} \dots B_{n m_n} \right) \quad (2.105)$$

La primera suma coincide con $\det(A)$, ecuación (2.107), mientras que la segunda coincide con $\det(B)$. Así se demuestra que el determinante del producto de dos matrices coincide con el producto de sus determinantes. Esta propiedad puede extenderse por inducción, apartado A.3.2, al producto de n matrices cuadradas.

- Supóngase que la i -ésima columna de A es combinación lineal de dos columnas denotadas $B_i = (b_j)$ y $C_i = (c_j)$, $j = 1, \dots, n$. Sustituyendo en la definición del determinante (2.86) y separando términos:

$$\begin{aligned} \det(A) &= \sum (-1)^p \cdot a_{1m_1} a_{2m_2} \dots a_{nm_n} = \\ &= \sum_{j=1}^n (-1)^p \cdot a_{1m_1} \dots (\alpha \cdot b_j + \beta \cdot c_j) \dots a_{nm_n} = \\ &= \sum_{j=1}^n (-1)^p \cdot a_{1m_1} \dots \alpha \cdot b_j \dots a_{nm_n} + \\ &+ \sum_{j=1}^n (-1)^p \cdot a_{1m_1} \dots \beta \cdot c_j \dots a_{nm_n} = \\ &= \alpha \cdot \sum_{j=1}^n (-1)^p \cdot a_{1m_1} \dots b_j \dots a_{nm_n} + \\ &+ \beta \cdot \sum_{j=1}^n (-1)^p \cdot a_{1m_1} \dots c_j \dots a_{nm_n} \end{aligned} \quad (2.106)$$

El primer sumatorio coincide con $\det(A_1, \dots, B_i, \dots, A_n)$, mientras que el segundo coincide con $\det(A_1, \dots, C_i, \dots, A_n)$. Se demuestra así la propiedad de multilinealidad.

□

Como consecuencia de la propiedad 1 de transposición del determinante, Proposición 2.3.1, la Definición 2.3.3 puede reescribirse en función de permutaciones de filas:

$$\det(A) = \det \begin{pmatrix} A_1 \\ A_2 \\ \vdots \\ A_n \end{pmatrix} = \sum (-1)^p \cdot a_{m_1 1} a_{m_2 2} \dots a_{m_n n} \quad (2.107)$$

Otra consecuencia de la propiedad de transposición del determinante es que las propiedades 2 y 4 de la Proposición 2.3.1 se cumplen también para permutaciones y combinaciones lineales de las filas, respectivamente.

Ejemplo 2.3.4 Sea la matriz A:

$$A = \begin{pmatrix} -1 & 2 \\ 3 & 4 \end{pmatrix} \quad (2.108)$$

Su determinante es:

$$\det(A) = a_{11}a_{22} - a_{12}a_{21} = -1 \cdot (-4) - 3 \cdot 2 = 4 - 6 = -2 \quad (2.109)$$

La transpuesta de A es:

$$A^t = \begin{pmatrix} -1 & 3 \\ 2 & 4 \end{pmatrix} \quad (2.110)$$

Los determinantes de ambas matrices coinciden, en virtud de la propiedad 1 de la Proposición 2.3.1:

$$\begin{aligned} \det(A^t) &= (A^t)_{11}(A^t)_{22} - (A^t)_{12}(A^t)_{21} = \\ &= a_{11}a_{22} - a_{21}a_{12} = -1 \cdot (-4) - 2 \cdot 3 = 4 - 6 = \\ &= -2 = \det(A) \end{aligned} \quad (2.111)$$

Ejemplo 2.3.5 Aplicando la fórmula del determinante de orden tres sobre la matriz B:

$$B = \begin{pmatrix} -1 & 3+i & i \\ 2i & 5-i & 4 \\ 2 & -1-i & i \end{pmatrix} \quad (2.112)$$

Se obtiene:

$$\det(B) = 25 - 7i \quad (2.113)$$

Si se aplica la misma fórmula sobre su transpuesta B^t :

$$B^t = \begin{pmatrix} -1 & 2i & 2 \\ 3+i & 5-i & -1-i \\ i & 4 & i \end{pmatrix} \quad (2.114)$$

Se obtiene:

$$\det(B^t) = 25 - 7i = \det(B) \quad (2.115)$$

Ejemplo 2.3.6 Intercambiando las columnas de la matriz A:

$$A = \begin{pmatrix} -1 & 2 \\ 3 & 4 \end{pmatrix} = (A_1, A_2) \quad (2.116)$$

se tiene que:

$$\det(A_2, A_1) = a_{12}a_{21} - a_{11}a_{22} = 6 - 4 = 2 = -\det(A) \quad (2.117)$$

Cumpléndose la propiedad 2 de antisimetría del determinante, Proposición 2.3.1.

Ejemplo 2.3.7 Permutando las filas uno y dos de la matriz B:

$$B = \begin{pmatrix} -1 & 3+i & i \\ 2i & 5-i & 4 \\ 2 & -1-i & i \end{pmatrix} = \begin{pmatrix} B_1 \\ B_2 \\ B_3 \end{pmatrix} \quad (2.118)$$

Se vuelve a comprobar que se cumple la propiedad de antisimetría del determinante:

$$\det \begin{pmatrix} B_2 \\ B_1 \\ B_3 \end{pmatrix} = -25 + 7i = -\det(B) \quad (2.119)$$

Ejemplo 2.3.8 Se descompone la segunda columna de A en el producto de un escalar por un vector columna:

$$A = \begin{pmatrix} -1 & 2 \\ 3 & 4 \end{pmatrix} = (A_1, A_2) = (A_1, 2 \cdot A'_2), A'_2 = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \quad (2.120)$$

Se tiene que, aplicando la propiedad 4 de multilinealidad del determinante, Proposición 2.3.1:

$$\begin{aligned} \det(A_1, 2 \cdot A'_2) &= 2 \cdot \det(A_1, A'_2) = 2 \cdot \det \begin{pmatrix} -1 & 1 \\ 3 & -2 \end{pmatrix} = \\ &= 2 \cdot (-1) = -2 = \det(A) \end{aligned} \quad (2.121)$$

Descomponiendo la segunda columna de A en suma de dos vectores columna:

$$A = \begin{pmatrix} -1 & 2 \\ 3 & 4 \end{pmatrix} = (A_1, A_2) = (A_1, A'_2 + A''_2) \quad (2.122)$$

Donde $A'_2 = (1, -1)^t$ y $A''_2 = (1, 3)^t$, se tiene, aplicando la misma propiedad, que:

$$\begin{aligned} \det(A_1, A'_2 + A''_2) &= \det(A_1, A'_2) + \det(A_1, A''_2) = \\ &= \det \begin{pmatrix} -1 & 1 \\ 3 & -1 \end{pmatrix} + \det \begin{pmatrix} -1 & 1 \\ 3 & 3 \end{pmatrix} = \\ &= -2 + 0 = -2 = \det(A) \end{aligned} \quad (2.123)$$

De la propiedad 4 de multilinealidad, Proposición 2.3.1, se derivan algunas propiedades adicionales del determinante:

Proposición 2.3.2 - Otras propiedades del determinante

1. Si una columna está formada sólo por ceros, $\det(A) = 0$.
2. Si una columna es igual o múltiplo de otra columna, $\det(A) = 0$.
3. Si una columna es combinación lineal de dos o más columnas, $\det(A) = 0$.
4. Si a una columna se le suma un múltiplo de otra columna de la misma matriz, su determinante no varía.

Estas propiedades son aplicables también a las filas.

Ejercicio 2.3.2 † Demostrar la cuatro propiedades contenidas en la Proposición 2.3.2 aplicando la propiedad 4 de multilinealidad del determinante, Proposición 2.3.1.

Su demostración se deja como ejercicio † para el lector.

Ejemplo 2.3.9 A continuación, se enumeran ejemplos de cada una de las propiedades 1 a 4 de la Proposición 2.3.2.

1. $\det \begin{pmatrix} -1 & 0 \\ 3 & 0 \end{pmatrix} = 2 \cdot 0 - 0 \cdot 3 = 0 - 0 = 0$

2. La segunda fila de la matriz A es múltiplo de la primera:

$$A = \begin{pmatrix} 3 & -2 \\ -6 & 3 \end{pmatrix} \quad (2.124)$$

$$F_2(A) = -2 \cdot F_1(A) \quad (2.125)$$

Como consecuencia:

$$\begin{aligned} \det(A) &= -2 \cdot \det \begin{pmatrix} 3 & -2 \\ 3 & -2 \end{pmatrix} = \\ &= -2 \cdot (-6 + 6) = -2 \cdot 0 = 0 \end{aligned} \quad (2.126)$$

3. La tercera fila de la matriz B es combinación lineal de las otras dos:

$$B = \begin{pmatrix} 1 & 0 & -2 \\ -4 & 1 & -1 \\ -2 & 1 & -5 \end{pmatrix} = \begin{pmatrix} B_1 \\ B_2 \\ B_3 \end{pmatrix} \quad (2.127)$$

En concreto:

$$F_3(B) = F_2(B) + 2 \cdot F_1(B) \quad (2.128)$$

Por lo que:

$$\det(B) = \det \begin{pmatrix} B_1 \\ B_2 \\ B_2 \end{pmatrix} + 2 \cdot \det \begin{pmatrix} B_1 \\ B_2 \\ B_1 \end{pmatrix} = 0 + 2 \cdot 0 = 0 \quad (2.129)$$

4. La matriz B' resulta de tomar la matriz B , ecuación (2.127), y sumarle la primera columna a la tercera de acuerdo con la fórmula $C_3(B') = C_3(B) + C_1(B)$:

$$B' = \begin{pmatrix} 1 & 0 & -1 \\ -4 & 1 & -5 \\ -2 & 1 & -7 \end{pmatrix} \quad (2.130)$$

Esta operación no altera el determinante de la matriz:

$$\begin{aligned} \det(B') &= \det(C_1(B), C_2(B), C_3(B) + C_1(B)) = \\ &= \det(C_1(B), C_2(B), C_3(B)) + \\ &+ \det(C_1(B), C_2(B), C_1(B)) = \\ &= \det(B) + 0 = 0 + 0 = 0 = \det(B) \end{aligned} \quad (2.131)$$

2.3.5 Resumen

1. El determinante asocia a cada matriz cuadrada la suma de los productos de los elementos en las diagonales principales de todas las permutaciones posibles de las columnas de la matriz, multiplicados por un factor $+1$ ó -1 en función del orden de cada permutación.
2. El determinante de una matriz no varía al transponer la matriz. Además, es un operador antisimétrico ante intercambios de filas o columnas y cumple una propiedad llamada "multilinealidad" que establece que es lineal para cada una de las columnas de la matriz: si se descompone una columna cualquiera de la matriz en combinación lineal de ciertos vectores columna, el determinante de la matriz puede descomponerse en combinación lineal de determinantes de matrices que contienen esos vectores entre sus columnas.
3. Si una columna de una matriz cuadrada puede expresarse como combinación lineal de otras columnas de la matriz, su determinante es cero.
4. Los puntos anteriores son aplicables, también, a las filas.

2.4 Cálculo de determinantes. Interpretación geométrica del determinante

2.4.1 Introducción

En la práctica, para ahorrar tiempo de cálculo y operaciones, es preferible calcular los determinantes de matrices de orden mayor de tres, o con un gran número de coeficientes nulos, mediante formulaciones alternativas a la Definición 2.3.3 de determinante como, por ejemplo, los desarrollos por cofactores respecto de una fila o columna cualquiera de la matriz.

Tras definir los cofactores de una matriz cuadrada, se mostrará la equivalencia entre la definición de determinante y estos desarrollos y se emplearán para enunciar una nueva propiedad relacionada con los determinantes de matrices triangulares o diagonales.

Finalmente, se establecerá una analogía entre el determinante de una matriz real cuadrada de orden dos o tres y el área, en \mathbb{R}^2 , o el volumen, en \mathbb{R}^3 , de un paralelogramo o paralelepípedo cuyas aristas vienen representadas por los vectores fila o columna de la matriz.

2.4.2 Menores y cofactores

Definición 2.4.1 — Menores y cofactores

Dada una matriz A cuadrada de orden n , se definen:

1. El menor del elemento a_{ij} de la matriz A como el determinante de la matriz resultante de eliminar la fila i y la columna j de A . Se le denotará como M_{ij} .
2. El cofactor del elemento a_{ij} de la matriz A como el producto $(-1)^{i+j} \cdot M_{ij}$, donde M_{ij} es el menor del elemento a_{ij} de A .

Ejemplo 2.4.1 La matriz A tiene los siguientes menores:

$$A = \begin{pmatrix} 3 & 0 & 2 \\ -1 & 1 & -3 \\ 0 & -1 & 4 \end{pmatrix} \quad (2.132)$$

$$M_{11} = \det \begin{pmatrix} 1 & -3 \\ -1 & 4 \end{pmatrix} = 1 \quad (2.133)$$

$$M_{12} = \det \begin{pmatrix} 1 & -3 \\ 0 & 4 \end{pmatrix} = -4 \quad (2.134)$$

$$M_{13} = \det \begin{pmatrix} -1 & 1 \\ 0 & -1 \end{pmatrix} = 1 \quad (2.135)$$

$$M_{21} = \det \begin{pmatrix} 0 & 2 \\ -1 & 4 \end{pmatrix} = 2 \quad (2.136)$$

$$M_{22} = \det \begin{pmatrix} 3 & 2 \\ 0 & 4 \end{pmatrix} = 12 \quad (2.137)$$

$$M_{23} = \det \begin{pmatrix} 3 & 0 \\ 0 & -1 \end{pmatrix} = -3 \quad (2.138)$$

$$M_{31} = \det \begin{pmatrix} 0 & 2 \\ 1 & -3 \end{pmatrix} = -2 \quad (2.139)$$

$$M_{32} = \det \begin{pmatrix} 3 & 2 \\ -1 & -3 \end{pmatrix} = -7 \quad (2.140)$$

$$M_{33} = \det \begin{pmatrix} 3 & 0 \\ -1 & 1 \end{pmatrix} = 3 \quad (2.141)$$

Y los siguientes cofactores asociados a cada elemento:

$$(-1)^{1+1} \cdot M_{11} = 1 \quad (2.142)$$

$$(-1)^{1+2} \cdot M_{12} = 4 \quad (2.143)$$

$$(-1)^{1+3} \cdot M_{13} = 1 \quad (2.144)$$

$$(-1)^{2+1} \cdot M_{21} = -2 \quad (2.145)$$

$$(-1)^{2+2} \cdot M_{22} = 12 \quad (2.146)$$

$$(-1)^{2+3} \cdot M_{23} = 3 \quad (2.147)$$

$$(-1)^{3+1} \cdot M_{31} = -2 \quad (2.148)$$

$$(-1)^{3+2} \cdot M_{32} = 7 \quad (2.149)$$

$$(-1)^{3+3} \cdot M_{33} = 3 \quad (2.150)$$

2.4.3 Cálculo de determinantes por suma de cofactores

Tómese un elemento a_{ij} de la matriz A . Si se agrupan todos los términos de la definición de $\det(A)$ que contienen a_{ij} y se saca factor común este elemento, se obtiene la expresión:

$$a_{ij} \cdot (-1)^{i+j} \cdot M_{ij} \quad (2.151)$$

Donde M_{ij} es el menor del elemento a_{ij} de A .

Si se sacan factor común todos los elementos de la columna j a la que pertenece a_{ij} y se suman los resultados, se obtiene una expresión que permite calcular $\det(A)$ como suma ponderada de los cofactores de cualquier columna j de A :

$$\det(A) = \sum_{i=1}^m a_{ij} \cdot (-1)^{i+j} \cdot M_{ij} \quad (2.152)$$

Por la equivalencia entre las definiciones por columnas y por filas del determinante de una matriz cuadrada A , ecuación (2.107), este podría también expresarse como suma ponderada de los cofactores de cualquier fila i de la matriz:

$$\det(A) = \sum_{j=1}^n a_{ij} \cdot (-1)^{i+j} \cdot M_{ij} \quad (2.153)$$

Ejemplo 2.4.2 Considérese una matriz cuadrada genérica de orden tres:

$$A = (A_1, A_2, A_3) = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \quad (2.154)$$

Calculando $\det(A)$ por desarrollo de cofactores respecto de la primera columna, se tiene:

$$\det(A) = a_{11} \cdot (-1)^2 \cdot M_{11} + a_{21} \cdot (-1)^3 \cdot M_{21} + a_{31} \cdot (-1)^4 \cdot M_{31} \quad (2.155)$$

Si el determinante se desarrolla respecto de la segunda columna:

$$\det(A) = a_{12} \cdot (-1)^3 \cdot M_{12} + a_{22} \cdot (-1)^4 \cdot M_{22} + a_{32} \cdot (-1)^5 \cdot M_{32} \quad (2.156)$$

Y respecto de la segunda fila:

$$\det(A) = a_{21} \cdot (-1)^3 \cdot M_{21} + a_{22} \cdot (-1)^4 \cdot M_{22} + a_{23} \cdot (-1)^5 \cdot M_{23} \quad (2.157)$$

Se deja al lector † comprobar que mediante cualquiera de estos desarrollos se obtiene la fórmula de *Sarrus* para el cálculo del determinante de una matriz de orden tres, conocida como “fórmula de *Sarrus*”:

$$\det(A) = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{21}a_{32}a_{13} - a_{13}a_{22}a_{31} - a_{12}a_{21}a_{33} - a_{23}a_{32}a_{11} \quad (2.158)$$

Por ejemplo, calculando el determinante de la matriz A del Ejemplo 2.4.1 según la ecuación (2.132) por suma de cofactores respecto de la primera fila se obtiene:

$$\det(A) = 3 \cdot 1 + 0 \cdot 4 + 2 \cdot 1 = 5 \quad (2.159)$$

Si se calcula respecto de la segunda columna, el resultado es:

$$\det(A) = 0 \cdot 4 + 1 \cdot 12 - 1 \cdot 7 = 5 \quad (2.160)$$

Y si se calcula respecto de la segunda fila:

$$\det(A) = -1 \cdot (-2) + 1 \cdot 12 - 3 \cdot 3 = 5 \quad (2.161)$$

El resultado es siempre $\det(A) = 5$, independientemente de la fila o columna que se tome para el desarrollo.

Ejercicio 2.4.1 † Comprobar que mediante cualquiera de estos desarrollos se obtiene la fórmula de *Sarrus* para el cálculo del determinante de una matriz de orden tres.

2.4.4 Determinantes de matrices triangulares y diagonales

Aplicando las fórmulas (2.152) y (2.153) del desarrollo por cofactores del determinante a una matriz cuadrada triangular o diagonal genérica, puede demostrarse † que sus determinantes son iguales al producto de los elementos de sus diagonales principales:

Proposición 2.4.1 — Determinante de una matriz triangular o diagonal

1. El determinante de una matriz cuadrada triangular es igual al producto de todos los elementos de su diagonal principal.
2. El determinante de una matriz cuadrada diagonal es igual al producto de todos los elementos de su diagonal principal.

$$\det(A) = a_{11}a_{22}a_{33} \dots a_{nn}$$

Ejercicio 2.4.2 † Demostrar estas dos propiedades desarrollando de manera recursiva el determinante de una matriz triangular genérica de orden n como suma ponderada de cofactores de la primera columna o fila de la matriz, apartado 2.4.3, y recordando que las matrices diagonales son un caso particular de matriz triangular, Definición 2.1.14.

Ejemplo 2.4.3 El determinante de la matriz A:

$$A = \begin{pmatrix} i & 0 & 4 \\ 0 & -i & -2 \\ 0 & 0 & 1+i \end{pmatrix}$$

Es igual a $\det(A) = a_{11}a_{22}a_{33} = i \cdot (-i) \cdot (1+i) = 1+i$.

Ejemplo 2.4.4 El determinante de la matriz B:

$$B = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 1-i \end{pmatrix}$$

Es igual a $\det(B) = b_{11}b_{22}b_{33} = 3 \cdot 4 \cdot (1-i) = 12 - 12i$.

2.4.5 Interpretación geométrica del determinante

Considérese un rectángulo contenido en el plano real euclidiano \mathbb{R}^2 cuyas aristas tienen longitudes a y b . Su área es igual al producto ab . Si el rectángulo se posicionase en \mathbb{R}^2 de modo tal que sus aristas sean paralelas a los ejes cartesianos x e y y uno de sus vértices se sitúe en el origen de coordenadas, dos de sus cuatro aristas coincidirán con los vectores fijos $(\pm a, 0)$ y $(0, \pm b)$.

Constrúyase ahora una matriz A cuyas filas, o columnas, coincidan con estos vectores:

$$A = \begin{pmatrix} \pm a & 0 \\ 0 & \pm b \end{pmatrix} \quad (2.162)$$

Los signos de los elementos de la diagonal dependen del vértice que se tome como origen de coordenadas. El determinante de esta matriz es claramente $\pm ab$. Su valor absoluto es ab . Es decir, el área de un rectángulo, cuyos lados tienen longitudes a y b .

En general, si las longitudes y direcciones de dos aristas contiguas de un paralelogramo dado coinciden con las de los vectores fijos (a_1, b_1) y (a_2, b_2) , el área del paralelogramo coincide con el valor absoluto del determinante de una matriz de orden dos que tenga por filas o columnas esos vectores (Figura 2.1):

$$\text{Área} = |\det(A)| = \left| \det \begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \end{pmatrix} \right| = \left| \det \begin{pmatrix} a_1 & a_2 \\ b_1 & b_2 \end{pmatrix} \right| \quad (2.163)$$

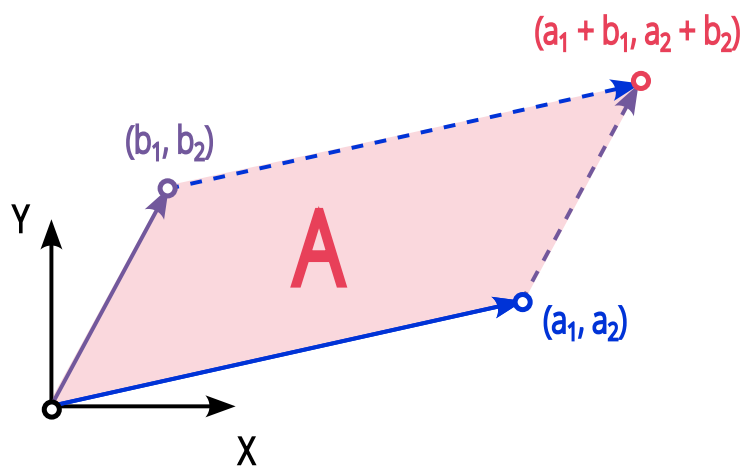


Figura 2.1: Si se define un paralelogramo contenido en el plano euclidiano mediante dos vectores de \mathbb{R}^2 que representen dos aristas cualesquiera contiguas en un vértice, su área puede calcularse mediante la fórmula (2.163).

Ejemplo 2.4.5 Sean $A = (3, -2)$ y $B = (1, 5)$ dos vectores que representan dos aristas contiguas de un paralelogramo dado, su área es:

$$\text{Área} = \left| \det \begin{pmatrix} 3 & 1 \\ -2 & 5 \end{pmatrix} \right| = \left| \det \begin{pmatrix} 3 & -2 \\ 1 & 5 \end{pmatrix} \right| = 13 \text{ unidades.}$$

De forma similar, el volumen de un paralelepípedo en \mathbb{R}^3 coincide con el valor absoluto del determinante de una matriz cuyas filas o columnas representen tres aristas contiguas del mismo (Figura 2.2).

$$\text{Volumen} = \left| \det \begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{pmatrix} \right| = \left| \det \begin{pmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{pmatrix} \right| \quad (2.164)$$

La demostración rigurosa de la equivalencia entre determinantes de matrices de orden dos y tres y los áreas y volúmenes de paralelogramos y paralelepípedos, ecuaciones (2.163) y (2.164) puede consultarse en algunos libros avanzados de álgebra lineal y geometría. En ellas, se recurre a la ayuda de proyecciones ortogonales sobre los ejes cartesianos de los vectores que representan las aristas de las figuras y a las fórmulas del área de un rectángulo y del volumen de un octaedro.

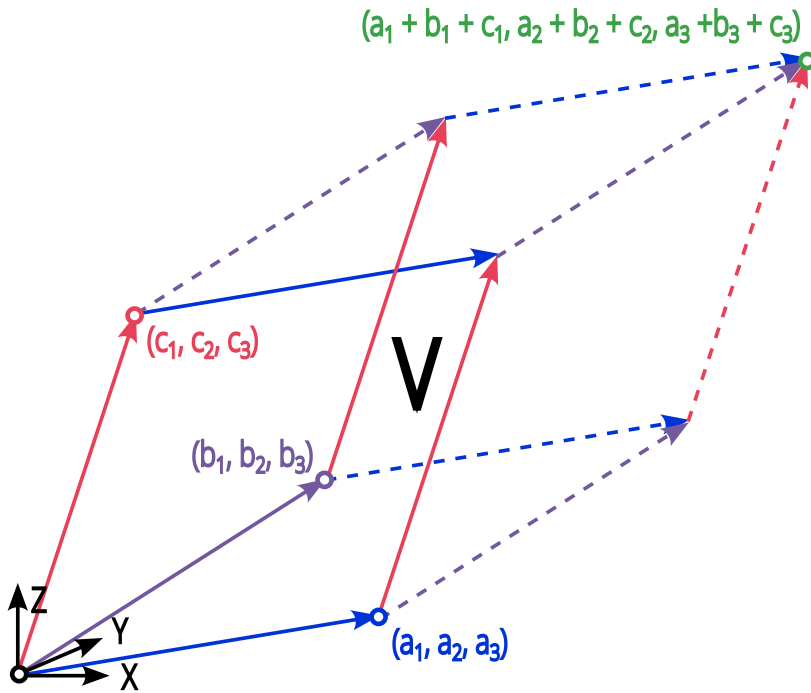


Figura 2.2: Si se define un paralelepípedo contenido en el espacio euclidiano mediante tres vectores de \mathbb{R}^3 que representen tres aristas cualesquiera contiguas en un vértice, su volumen puede calcularse mediante la fórmula (2.164).

Ejemplo 2.4.6 Sean $A = (3, 0, 2)$, $B = (-1, 1, -3)$ y $C = (0, -1, 4)$ tres vectores que representan tres aristas contiguas de un paralelepípedo dado, el volumen de este es:

$$\text{Volumen} = \left| \det \begin{pmatrix} 3 & 0 & 2 \\ -1 & 1 & -3 \\ 0 & -1 & 4 \end{pmatrix} \right| = 5 \text{ unidades}$$

Volviendo al caso en \mathbb{R}^2 , el producto de la matriz:

$$A = \begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \end{pmatrix} \tag{2.165}$$

Por el primer vector canónico $\bar{e}_1 = (1, 0)^t$ da como resultado la primera columna de A , el vector $(a_1, a_2)^t$, que representa una de las aristas del paralelogramo de la Figura 2.1. El producto de A por el segundo vector canónico $\bar{e}_2 = (0, 1)^t$ da como resultado la segunda columna de A , que representa la arista contigua a la primera. Multiplicar A por un vector contenido en el cuadrado que tiene por aristas \bar{e}_1 y \bar{e}_2 produce un vector contenido en el paralelogramo definido por las columnas de A . El área del cuadrado es de uno, mientras que el área del paralelogramo es, de acuerdo con la ecuación (2.4.5), igual a $|\det(A)|$. Lo mismo sucede en \mathbb{R}^3 : los tres vectores canónicos de este espacio forman un cubo de volumen unidad, y el producto con A de los vectores contenidos en ese cubo los transforma en el paralelepípedo cuyas aristas coinciden con las columnas de la matriz y cuyo volumen es $|\det(A)|$.

En general, multiplicar una matriz cuadrada $A \in \mathbb{K}^n$ por el conjunto de los vectores (o puntos) que conforman una figura geométrica en \mathbb{K}^n resulta en otra figura cuyo volumen es $|\det(A)|$ veces el de la figura inicial, por lo que el determinante también puede interpretarse como un factor de ampliación que determina cuánto aumenta, en promedio, la longitud de los vectores cuando vienen transformados por la matriz. El signo del

determinante está relacionado con la orientación de los vectores resultantes de las transformaciones. Si la orientación de los vectores no cambia, el determinante es positivo; si la orientación cambia y las figuras se invierten al ser transformadas por la matriz como si se reflejasen en un espejo, el determinante es negativo.

2.4.6 Resumen

1. El determinante de una matriz cuadrada puede calcularse como la suma ponderada de los cofactores de una de sus filas o columnas.
2. El determinante de una matriz cuadrada triangular o diagonal es simplemente el producto de las componentes de su diagonal principal.
3. El área, en \mathbb{R}^2 , y el volumen, en \mathbb{R}^3 , de un paralelogramo o paralelepípedo son iguales al valor absoluto del determinante de una matriz cuadrada cuyas filas o columnas caracterizan sus aristas.
4. El determinante puede entenderse también como un factor de ampliación que determina cuánto aumenta o disminuye el volumen de una figura en \mathbb{K}^n al ser multiplicada por la matriz, y si esta conserva o no su orientación inicial.

2.5 Inversa de una matriz cuadrada: existencia, propiedades y cálculo

2.5.1 Introducción

Existen ciertas matrices cuadradas, llamadas “invertibles”, que son de particular interés por sus aplicaciones y sus propiedades matemáticas. Son matrices a las cuales se les asocia biunívocamente otra matriz cuadrada del mismo orden, conocida como su “inversa”. Tras definir formalmente el concepto de matriz cuadrada invertible y de inversa de una matriz cuadrada, se enunciarán las propiedades de estas matrices y se propondrá un método para su cálculo conocido como “algoritmo de *Gauss-Jordan*”. Este algoritmo se basa en transformar la matriz cuadrada a invertir multiplicándola por unas matrices especiales llamadas “matrices elementales”.

2.5.2 Definición de inversa de una matriz

Definición 2.5.1 – Matriz invertible e inversa de una matriz

Una matriz cuadrada $A \in \mathbb{K}^{n \times n}$, se dice que es invertible, o regular, si $\exists A^{-1} \in \mathbb{K}^{n \times n}$ tal que $AA^{-1} = A^{-1}A = I$.

La matriz A^{-1} se denomina inversa de A .

Es fácil demostrar que la inversa, de existir, es única. Supóngase que la matriz regular A tiene dos inversas denotadas A_1^{-1}, A_2^{-1} . Aplicando en (1) la Definición 2.1.7 de matriz identidad, en (2) la Definición 2.5.1 de matriz inversa, y en (3) la propiedad asociativa del producto matricial, punto 2 de la Proposición 2.2.3, se tiene que:

$$\begin{aligned} A_1^{-1} &\stackrel{(1)}{=} A_1^{-1}I \stackrel{(2)}{=} A_1^{-1}(AA_2^{-1}) \stackrel{(3)}{=} \\ &\stackrel{(3)}{=} (A_1^{-1}A)A_2^{-1} \stackrel{(2)}{=} IA_2^{-1} \stackrel{(1)}{=} A_2^{-1} \end{aligned} \quad (2.166)$$

Aunque es posible definir la inversa de una matriz rectangular, también llamada “inversa generalizada” o “pseudoinversa”, este apartado se centrará únicamente en el estudio y cálculo de las inversas de matrices cuadradas, dejando la definición y el cálculo de pseudoinversas de matrices rectangulares para el apartado 5.6.4.

Definición 2.5.2 – Matriz singular

Se dice que una matriz cuadrada es singular si no es invertible.

La siguiente proposición enuncia una condición, necesaria y suficiente, para que una matriz cuadrada sea invertible:

Proposición 2.5.1 – Determinantes de matrices regulares y singulares

Una matriz cuadrada es regular, si y sólo si, tiene un determinante distinto de cero.

En virtud de esta proposición, para comprobar si una matriz cuadrada es invertible basta con calcular su determinante. Si es distinto de cero, la matriz es invertible. Si el determinante es nulo, la matriz es singular. Esta proposición se demostrará en el Apartado 3.4.10, cuando se enuncien otras dos condiciones equivalentes a esta, también necesarias y suficientes, para que una matriz cuadrada sea regular, en el Corolario 3.4.6.

Ejemplo 2.5.1 El determinante de la matriz A:

$$A = \begin{pmatrix} 3 & 0 & 2 \\ -1 & 1 & -3 \\ 0 & -1 & 4 \end{pmatrix} \quad (2.167)$$

Es igual a $\det(A) = 12 + 2 - 9 = 5 \neq 0$, por lo que es invertible. Su inversa es la matriz B:

$$B = \frac{1}{5} \begin{pmatrix} 1 & -2 & -2 \\ 4 & 12 & 7 \\ 1 & 3 & 3 \end{pmatrix} \quad (2.168)$$

De hecho, $AB = BA = I = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$.

Ejemplo 2.5.2 El determinante de la matriz C:

$$C = \begin{pmatrix} i & 2i & 3i \\ 1+i & -i & 1 \\ -3 & 4i & -3+4i \end{pmatrix}$$

Es nulo, pues sus filas y columnas son linealmente dependientes, propiedad 3 de la Proposición 2.3.2. Véase, por ejemplo, que la tercera columna es suma de las otras dos.

Por tanto, C es una matriz singular.

La relación entre el determinante de una matriz regular y el de su inversa viene dada por la siguiente proposición:

Proposición 2.5.2 – Determinante de la inversa de una matriz

Dadas una matriz cuadrada regular cualquiera A y su inversa A^{-1} , sus determinantes cumplen:

$$\det(A^{-1}) = \frac{1}{\det(A)}$$

Demostración:

Considerando que, según la propiedad 2 de la Proposición 2.4.1, el determinante de la matriz identidad es uno, sea cual sea su orden, y aplicando la propiedad 3 de la Proposición 2.3.1, se tiene:

$$1 = \det(I) = \det(AA^{-1}) = \det(A)\det(A^{-1}) \quad (2.169)$$

□

Ejemplo 2.5.3 El determinante de la matriz A del Ejemplo 2.5.1, ecuación (2.167), es igual a $\det(A) = 5$, mientras que su inversa, la matriz B , ecuación (2.168), tiene por determinante $\det(B) = 1/5 = 1/\det(A)$ en virtud de la Proposición 2.5.2.

Como cabría esperar, al volver a invertir la inversa de una matriz regular se recupera la matriz inicial, tal y como se enuncia a continuación:

Proposición 2.5.3 – Inversa de la inversa de una matriz

Toda matriz cuadrada regular A cumple:

$$(A^{-1})^{-1} = A$$

Demostración:

Por la Definición 2.5.1 de inversa de una matriz regular, y por las propiedades del producto matricial, Proposición 2.2.3:

$$AA^{-1} = I \Rightarrow AA^{-1}(A^{-1})^{-1} = (A^{-1})^{-1} \Rightarrow A = (A^{-1})^{-1} \quad (2.170)$$

$$A^{-1}A = I \Rightarrow (A^{-1})^{-1}A^{-1}A = (A^{-1})^{-1} \Rightarrow A = (A^{-1})^{-1} \quad (2.171)$$

□

Aplicando la propiedad asociativa del producto matricial, punto 2 de la Proposición 2.2.3, al producto de dos matrices cuadradas regulares A y B de igual orden y de sus inversas, se tiene:

$$(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AA^{-1} = I \quad (2.172)$$

$$(B^{-1}A^{-1})(AB) = B^{-1}(A^{-1}A)B = B^{-1}B = I \quad (2.173)$$

Así, se demuestra que la inversa del producto de dos matrices AB coincide con el producto $B^{-1}A^{-1}$, como enuncia la siguiente proposición:

Proposición 2.5.4 – Inversa del producto de dos matrices regulares

Sean A y B dos matrices cuadradas de igual orden, ambas invertibles, su producto AB también es invertible y cumple:

$$(AB)^{-1} = B^{-1}A^{-1}$$

Supóngase ahora que se tiene el siguiente producto de tres matrices invertibles: AB_1B_2 . Llamando B al producto B_1B_2 y aplicando la Proposición 2.5.4 se obtiene:

$$(AB_1B_2)^{-1} = AB^{-1} = B^{-1}A^{-1} = (B_1B_2)^{-1}A^{-1} = B_2^{-1}B_1^{-1}A^{-1} \quad (2.174)$$

En general, puede demostrarse por inducción que, sea cual sea el número k de matrices, todas cuadradas, regulares y del igual orden, presentes en el producto, su inversa es igual al producto de las inversas de las matrices en orden inverso, como indica esta proposición.

Proposición 2.5.5 – Inversa del producto de $k \geq 2$ matrices regulares

La inversa del producto de k matrices cuadradas regulares de igual orden cumple:

$$(A_1 A_2 \dots A_k)^{-1} = A_k^{-1} A_{k-1}^{-1} \dots A_1^{-1}$$

Ejemplo 2.5.4 Las inversas de las matrices A y B :

$$A = \begin{pmatrix} -3 & 1 \\ -1 & 0 \end{pmatrix} \quad B = \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix} \quad (2.175)$$

Son, respectivamente, A^{-1} y B^{-1} :

$$A^{-1} = \begin{pmatrix} 0 & -1 \\ 1 & -3 \end{pmatrix} \quad B^{-1} = \begin{pmatrix} -3 & 2 \\ 2 & -1 \end{pmatrix} \quad (2.176)$$

Puede comprobarse que su producto, $AB = \begin{pmatrix} -1 & -3 \\ -1 & -2 \end{pmatrix}$ cumple:

$$(AB)^{-1} = \begin{pmatrix} 2 & -3 \\ -1 & 1 \end{pmatrix} = B^{-1} A^{-1} \quad (2.177)$$

2.5.3 Operaciones elementales

Un conocido algoritmo para calcular la inversa de una matriz cuadrada regular es el llamado “método de los cofactores”. Este método tiene el grave inconveniente de que el número de operaciones necesarias para llevar a cabo la inversión de una matriz crece muy rápidamente con su orden. En concreto, para invertir una matriz de orden n se requeriría el cálculo de $n^2 + 1$ determinantes. Una alternativa al método de los cofactores especialmente ventajosa en matrices de grandes dimensiones es el algoritmo de *Gauss-Jordan*, que se basa en efectuar ciertas operaciones sobre las filas y las columnas de la matriz a invertir.

A continuación, como paso previo a la descripción del algoritmo de *Gauss-Jordan*, se definirán las llamadas “operaciones elementales” y “matrices elementales”:

Definición 2.5.3 – Operaciones elementales

Dada una matriz $A \in \mathbb{K}^{m \times n}$ cualquiera, cuadrada o rectangular, se definen las siguientes operaciones elementales por filas sobre la matriz A :

1. Suma de una fila de la matriz, multiplicada por un escalar, a otra fila distinta.
2. Multiplicación de una fila de la matriz por un escalar no nulo.
3. Intercambio de dos filas.

También puede realizarse empleando columnas en lugar de filas.

Definición 2.5.4 – Matrices elementales

Una matriz elemental por filas es una matriz cuadrada resultante de efectuar una determinada operación elemental por filas a la matriz identidad. Igual sucede si se aplica a las columnas.

Las matrices y operaciones elementales cumplen las siguientes propiedades básicas:

Proposición 2.5.6 – Propiedades de las matrices y las operaciones elementales

1. Toda matriz elemental por filas lo es por columnas y viceversa, aunque no necesariamente asociadas a la misma operación por filas que por columnas.
2. Toda matriz elemental, denotada E , es invertible, y su inversa es la matriz elemental asociada a la operación que deshace la realizada sobre la matriz identidad para transformarla en E . Las operaciones inversas son:
 - a) De la suma de una fila i -ésima de la matriz, multiplicada por un escalar λ , a otra fila distinta j -ésima: la resta de la fila i -ésima, multiplicada por λ , a la fila j -ésima. Igual para columnas.
 - b) De la multiplicación de una fila de la matriz por un escalar λ no nulo: la multiplicación de esa misma fila por $1/\lambda$. Igual para columnas.
 - c) Del intercambio de dos filas: ese mismo intercambio. Igual para columnas.
3. Las operaciones y matrices elementales, en general, no conmutan. Es decir, el orden con el que se efectúa una secuencia de operaciones elementales influye en el resultado final.

Se propone una demostración sólo del punto 3, considerándose triviales las propiedades restantes.

Demostración: (sólo de la propiedad 3)

Si se toman dos matrices elementales E_1, E_2 cualesquiera asociadas a operaciones elementales distintas y se multiplican entre sí, lo más probable es que la matriz resultante sea distinta en función del orden en el que se multipliquen. Es decir, que:

$$E_1 E_2 \neq E_2 E_1 \quad (2.178)$$

Así, con un contraejemplo puede demostrarse la no conmutabilidad de las operaciones elementales.

□

Ejemplo 2.5.5 Al aplicar la operación elemental F'_3 :

$$F'_3 = F_3 + 2 \cdot F_1 \quad (2.179)$$

Sobre la matriz identidad de orden tres, esta se transforma en la matriz elemental E_1 :

$$E_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 0 & 1 \end{pmatrix} \quad (2.180)$$

La inversa de la operación elemental $F'_3 = F_3 + 2 \cdot F_1$ es aquella que la deshace. Es decir, la operación $F'_3 = F_3 - 2 \cdot F_1$. Su matriz elemental asociada es E'_1 :

$$E'_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -2 & 0 & 1 \end{pmatrix} \quad (2.181)$$

Puede comprobarse fácilmente que $E'_1 E_1 = E_1 E'_1 = I$, de lo que se deduce que $E'_1 = E_1^{-1}$.

Ejemplo 2.5.6 Al aplicar la operación elemental C'_2 :

$$C'_2 = -3 \cdot C_2 \quad (2.182)$$

Sobre la matriz identidad de orden tres, esta se transforma en la matriz elemental E_2 :

$$E_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (2.183)$$

La inversa de la operación elemental $C'_2 = -3 \cdot C_2$ es aquella que la deshace. Es decir, la operación $C'_2 = -3 \cdot C_2$. Su matriz elemental asociada es E'_2 :

$$E'_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1/3 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (2.184)$$

Puede comprobarse fácilmente que $E'_2 E_2 = E_2 E'_2 = I$, de lo que se deduce que $E'_2 = E_2^{-1}$.

Ejemplo 2.5.7 Al aplicar la operación elemental F_1 :

$$F_1 \leftrightarrow F_2 \quad (2.185)$$

Sobre la matriz identidad de orden tres, esta se transforma en la matriz elemental E_3 :

$$E_3 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (2.186)$$

La inversa de la operación elemental $F_1 \leftrightarrow F_2$ es aquella que la deshace. Es decir, la operación $F_1 \leftrightarrow F_2$. Su matriz elemental asociada es:

$$E'_3 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (2.187)$$

Matriz que coincide con E_3 . En general, toda matriz elemental de tipo tres, intercambio de filas y columnas, es su propia inversa, propiedad conocida como “idempotencia”. Además, son matrices simétricas.

Para efectuar operaciones elementales sobre las filas o columnas de una matriz A , esta puede multiplicarse por matrices elementales como las calculadas en los Ejemplos 2.5.5 a 2.5.7, tal como establece esta proposición:

Proposición 2.5.7 – Relación entre matrices elementales y operaciones elementales

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

1. Realizar operaciones elementales sobre sus filas equivale a pre-multiplicar por sucesivas matrices elementales de orden m :

$$E_k \dots E_2 E_1 A \quad (2.188)$$

2. Realizar operaciones elementales sobre sus columnas equivale a pos-multiplicar por sucesivas matrices elementales de orden n :

$$A E_1 E_2 \dots E_k \quad (2.189)$$

Las propiedades contenidas en la Proposición 2.5.7 resultan triviales si se recuerda, Proposición 2.2.5, que las columnas del producto de dos matrices son combinaciones lineales de las columnas de la primera matriz, propiedad 1, y que las filas del producto de dos matrices son combinaciones lineales de las filas de la segunda matriz, propiedad 2.

Ejemplo 2.5.8 Puede comprobarse que:

$$A = \begin{pmatrix} 3 & 0 & 2 \\ -1 & 1 & -3 \\ 0 & -1 & 4 \end{pmatrix} \xrightarrow[\substack{F_1' = F_3 + 2 \cdot F_1 \\ F_1 \leftrightarrow F_2}]{} \begin{pmatrix} -1 & 1 & -3 \\ 3 & 0 & 2 \\ -2 & 1 & -2 \end{pmatrix} = E_1 E_3 A$$

Siendo E_1 y E_3 las matrices elementales, ya calculadas en los Ejemplos 2.5.5 y 2.5.7, asociadas a cada una de las operaciones elementales por filas indicadas.

Ejemplo 2.5.9 De manera análoga al Ejemplo 2.5.8, se tiene que:

$$A = \begin{pmatrix} 3 & 0 & 2 \\ -1 & 1 & -3 \\ 0 & -1 & 4 \end{pmatrix} \xrightarrow{C_2' = -3 \cdot C_2} \begin{pmatrix} 3 & 0 & 2 \\ -1 & -3 & -3 \\ 0 & 3 & 4 \end{pmatrix} = A E_2$$

Donde E_2 es la matriz elemental, ya calculada en el Ejemplo 2.5.6, asociada a la operación elemental por columnas indicada.

Ejemplo 2.5.10 La operación elemental $C_2 \leftrightarrow C_3$ tiene la siguiente matriz elemental asociada:

$$I = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \xrightarrow{C_2 \leftrightarrow C_3} E_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \quad (2.190)$$

Puede comprobarse que:

$$A = \begin{pmatrix} 3 & 0 & 2 \\ -1 & 1 & -3 \\ 0 & -1 & 4 \end{pmatrix} \xrightarrow{C_2 \leftrightarrow C_3} \begin{pmatrix} 3 & 2 & 0 \\ -1 & -3 & 1 \\ 0 & 4 & -1 \end{pmatrix} = AE_3 \quad (2.191)$$

La matriz E_3 también se puede obtener intercambiando las filas dos y tres de la matriz identidad I . De hecho, pre-multiplicando A por E_3 se permutan estas dos filas de A :

$$A = \begin{pmatrix} 3 & 0 & 2 \\ -1 & 1 & -3 \\ 0 & -1 & 4 \end{pmatrix} \xrightarrow{F_2 \leftrightarrow F_3} \begin{pmatrix} 3 & 0 & 2 \\ 0 & -1 & 4 \\ -1 & 1 & -3 \end{pmatrix} = E_3A \quad (2.192)$$

Las matrices elementales que efectúan intercambios de parejas de filas o columnas, punto 3 de la Definición 2.5.3, se pueden multiplicar entre sí para obtener matrices que, multiplicadas por una matriz cualquiera A , permutan sus filas o columnas. A estas nuevas matrices se les conoce como “matrices de permutación”.

Observación 2.5.1 † El concepto de permutación se define en 2.3.1.

Definición 2.5.5 – Matrices de permutación

A una matriz cuadrada P que resulta de efectuar una secuencia de intercambios de filas o columnas se la conoce como matriz de permutación †.

Al pre-multiplicar P por una matriz A cualquiera se permutan sus filas, y al pos-multiplicar A por P se permutan sus columnas. Algunas propiedades básicas de las matrices de permutación son:

Proposición 2.5.8 – Propiedades de las matrices de permutación

1. Toda matriz de permutación por filas lo es también por columnas y viceversa.
2. Cualquier matriz de permutación puede expresarse como el producto de matrices elementales de tipo tres, es decir, de intercambios de filas y columnas, punto 3 de la Definición 2.5.3.
3. Al igual que los intercambios de filas o columnas, propiedad 3 de la Proposición 2.5.6, el producto de matrices de permutación, en general, no conmuta.

4. La inversa de toda matriz P de permutación coincide con su transpuesta.

$$P^T = P^{-1}$$

Las propiedades 1 y 2 se han justificado ya en apartados anteriores. Se propondrán sólo demostraciones de las propiedades 3 y 4.

Demostración: (sólo de las propiedades 3 y 4)

3. Esto se puede demostrar mediante un contraejemplo, al igual que la Proposición 2.5.3, punto 3 de las matrices elementales. Si se toman dos matrices de permutación cualesquiera y se multiplican, lo más probable es que el orden de la multiplicación influya en el resultado final. Es decir, que:

$$P_1 P_2 \neq P_2 P_1 \tag{2.193}$$

4. Sea $E_1 E_2 \dots E_k$ un producto de k matrices elementales de tipo tres, intercambios de filas o columnas, cualesquiera de orden n . La matriz P resultante es una matriz de permutación, Definición 2.5.5. Invertiendo P y aplicando las Proposiciones 2.69 y 2.5.5 se obtiene:

$$\begin{aligned} P^t &= (E_1 E_2 \dots E_k)^t = E_k^t E_{k-1}^t \dots E_1^t = \\ &\stackrel{(*)}{=} E_k E_{k-1} \dots E_1 \stackrel{(**)}{=} E_k^{-1} E_{k-1}^{-1} \dots E_1^{-1} = \\ &= (E_1 E_2 \dots E_k)^{-1} = P^{-1} \end{aligned} \tag{2.194}$$

En (*) y (**) se aplican, respectivamente, dos propiedades que todas las matrices elementales de tipo tres cumplen: la simetría ($E^t = E$) y la idempotencia ($E^2 = I$, que implica que $E^{-1} = E$). Una muestra del cumplimiento de estas propiedades puede verse en el Ejemplo 2.5.7.

□

Ejemplo 2.5.11 La operación elemental $F_1 \leftrightarrow F_2$ tiene como matriz elemental asociada:

$$I = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \xrightarrow{F_1 \leftrightarrow F_2} E_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{2.195}$$

Mientras que la operación elemental $F_2 \leftrightarrow F_3$ tiene como matriz elemental asociada:

$$I = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \xrightarrow{F_2 \leftrightarrow F_3} E_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \tag{2.196}$$

Multiplicando ambas matrices en orden inverso, se obtiene la siguiente matriz de permutación:

$$P = E_2 E_1 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \quad (2.197)$$

pre-multiplicando A del ejemplo anterior por P , se permutan las filas de A , de modo tal que la primera fila pasa a ser la tercera, la segunda pasa a ser la primera, y la tercera pasa a ser la segunda:

$$A = \begin{pmatrix} 3 & 0 & 2 \\ -1 & 1 & -3 \\ 0 & -1 & 4 \end{pmatrix} \rightarrow \begin{pmatrix} -1 & 1 & -3 \\ 0 & -1 & 4 \\ 3 & 0 & 2 \end{pmatrix} = PA \quad (2.198)$$

Si se multiplicasen las matrices elementales en el orden opuesto, se obtendría una matriz de permutación distinta, lo que implica que los intercambios de filas $F_1 \leftrightarrow F_2$ y $F_2 \leftrightarrow F_3$ no conmutan.

2.5.4 Algoritmo de Gauss-Jordan

Este método de inversión de matrices se basa en la siguiente idea: transformar la matriz A en la matriz identidad I mediante operaciones elementales que pueden ser por filas o por columnas, pero sin mezclar ambos tipos de operaciones en la misma inversión.

Por ejemplo, supóngase que A se reduce a la matriz identidad mediante operaciones elementales por filas. Como se ha visto en la Proposición 2.5.7, esto equivale a pre-multiplicar la matriz A por una sucesión de matrices elementales que efectúan ordenadamente cada una de las operaciones desde 1 hasta k :

$$E_k \dots E_2 E_1 A = I \quad (2.199)$$

Por la Definición 2.5.1 de inversa de una matriz cuadrada, esta, de existir, cumple $A^{-1}A = AA^{-1} = I$. Comparando esta expresión con la ecuación (2.199), se tiene que:

$$E_k \dots E_2 E_1 = A^{-1} \quad (2.200)$$

Es decir, que la inversa de A es igual al producto de k matrices elementales correspondientes a operaciones elementales por filas que transforman A en la matriz identidad, desde 1 hasta k , en orden inverso.

Si se utilizaran operaciones elementales por columnas para transformar la matriz, esta reducción se podría representar matemáticamente como una pos-multiplicación de A por una secuencia de matrices elementales que efectúan ordenadamente cada una de las operaciones desde 1 hasta k , Proposición 2.5.7:

$$A E_1 E_2 \dots E_k = I \Rightarrow E_1 E_2 \dots E_k = A^{-1} \quad (2.201)$$

Por lo que la inversa de A también podría obtenerse como el producto de k matrices elementales correspondientes a operaciones elementales por columnas que transforman A en la matriz identidad, ordenadas desde 1 hasta k .

En la práctica, este algoritmo se implementa del siguiente modo: en primer lugar, se transforma la matriz cuadrada a invertir en la matriz identidad por una secuencia de operaciones elementales por filas o por columnas, pero sin mezclar ambos tipos de operaciones. Y después, se aplican esas mismas operaciones elementales sobre la matriz identidad, en el mismo orden, hasta obtener la matriz inversa buscada.

Véase con el siguiente ejemplo:

Ejemplo 2.5.12 En primer lugar, se transforma A en la matriz identidad mediante operaciones elementales, en este caso por filas:

$$\begin{aligned}
 A &= \begin{pmatrix} 1+i & 1-i \\ -1-i & -1 \end{pmatrix} \xrightarrow{F'_2 = F_2 + F_1} \begin{pmatrix} 1+i & 1-i \\ 0 & -i \end{pmatrix} \\
 &\xrightarrow{F'_1 = F_1 - (1+i) \cdot F'_2} \begin{pmatrix} 1+i & 0 \\ 0 & -i \end{pmatrix} \\
 &\xrightarrow{\substack{F''_1 = \frac{1}{1+i} \cdot F'_1 \\ F''_2 = \frac{1}{-i} \cdot F'_2}} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I.
 \end{aligned} \tag{2.202}$$

Aplicando estas mismas operaciones elementales en el mismo orden sobre la matriz identidad, se obtiene la inversa de A :

$$\begin{aligned}
 I &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \xrightarrow{F'_2 = F_2 + F_1} \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \\
 &\xrightarrow{F'_1 = F_1 - (1+i) \cdot F'_2} \begin{pmatrix} -i & -1+i \\ 1 & 1 \end{pmatrix} \\
 &\xrightarrow{\substack{F''_1 = \frac{1}{1+i} \cdot F'_1 \\ F''_2 = \frac{1}{-i} \cdot F'_2}} \begin{pmatrix} -\frac{1}{2} - \frac{1}{2}i & -1 \\ i & i \end{pmatrix} = A^{-1}
 \end{aligned} \tag{2.203}$$

2.5.5 Resumen

1. Si una matriz cuadrada admite inversa, es regular. Si no la admite, se dice que es singular.
2. Una matriz cuadrada es regular, si y sólo si, su determinante es distinto de cero.
3. La inversa de un producto de matrices invertibles es igual al producto de sus inversas en orden contrario.
4. La inversa de una matriz invertible puede calcularse multiplicando entre sí las matrices elementales que la transforman en la matriz identidad.



Esta página está en blanco intencionalmente.

3.1 Definición, tipos y propiedades de las aplicaciones lineales

3.1.1 Introducción

Una aplicación o función entre dos conjuntos llamados “inicial” y “final” es una relación entre sus elementos que asocia a cada elemento del conjunto inicial un único elemento del conjunto final. Algunos tipos de aplicaciones, como las inyectivas, las suryectivas y las biyectivas, cumplen algunas propiedades que las hacen muy especiales. En particular, el álgebra lineal se centra en el estudio de las denominadas “aplicaciones lineales”. En apartados posteriores, se establecerá una relación entre las aplicaciones lineales que relacionan espacios vectoriales y las matrices.

3.1.2 Aplicaciones

Definición 3.1.1 – Aplicación entre conjuntos

Sean E, F conjuntos no vacíos. Una aplicación, o función, f de E en F es una regla que asigna a cada elemento de E un único elemento de F , de modo que $\forall x \in E$ existe un único $y \in F$ tal que $f(x) = y$. Esta asignación entre los dos conjuntos se denota por:

$$f : E \mapsto F$$

Definición 3.1.2 – Conjuntos inicial y final de una aplicación. Imagen y preimagen

Dada una aplicación $f : E \mapsto F$ de E en F :

1. Se llama conjunto inicial, o dominio, de f al conjunto E .
2. Se llama conjunto final de f al conjunto F .
3. Si $f(x) = y$, se dice que y es la imagen por f de x y que x es el antecedente o preimagen de y por f .

Ejemplo 3.1.1 Dados los conjuntos $E = \{\text{perro, gato, cabra, oveja}\}$ y $F = \{\text{guau, miau, bee, muu}\}$, una aplicación $f : E \mapsto F$ que los asocia es la siguiente:

$$\left\{ \begin{array}{l} f(\text{perro}) = \text{guau}, \\ f(\text{gato}) = \text{miau}, \\ f(\text{cabra}) = \text{bee}, \\ f(\text{oveja}) = \text{bee}. \end{array} \right.$$

3.1	Definición y propiedades	83
3.2	Aplicación lineal respecto a una base	89
3.3	Cambio de base. Similitud y congruencia de matrices	95
3.4	Imagen, núcleo y rango	101
3.5	Composición e inversión de aplicaciones lineales	118

Así, puede decirse que:

1. “guau” es la imagen de “perro” por f .
2. “perro” es la preimagen de “guau” por f .
3. “bee” es la imagen de “cabra” y “oveja” por f .
4. “cabra” y “oveja” son las preimágenes de “bee” por f .

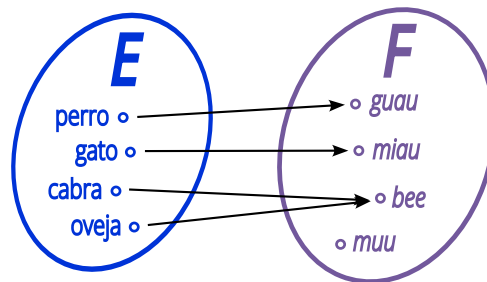


Figura 3.1: La aplicación definida en el Ejemplo 3.1.1.

Definición 3.1.3 – Endomorfismo de un conjunto en sí mismo

Si f es una aplicación cuyos conjuntos inicial y final coinciden. Es decir, si $E = F$, se dice que f es un endomorfismo de E en sí mismo:

$$f : E \mapsto E$$

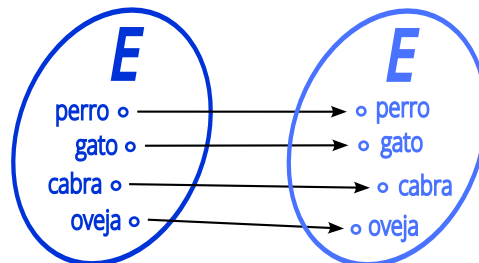


Figura 3.2: La aplicación identidad del conjunto E del Ejemplo 3.1.2.

Definición 3.1.4 – Aplicación identidad de un conjunto

Dado un conjunto E , se define la aplicación identidad de E , y se denota por I_E , como el endomorfismo $I_E : E \mapsto E$ que cumple $I_E(x) = x \forall x \in E$.

Ejemplo 3.1.2 La aplicación identidad del conjunto $E = \{\text{perro, gato, cabra, oveja}\}$, $I_E : E \mapsto E$, se define así:

$$\begin{cases} I_E(\text{perro}) = \text{perro}, \\ I_E(\text{gato}) = \text{gato}, \\ I_E(\text{cabra}) = \text{cabra}, \\ I_E(\text{oveja}) = \text{oveja}. \end{cases}$$

3.1.3 Tipos de aplicaciones

Definición 3.1.5 – Aplicaciones inyectivas, suryectivas y biyectivas

Una aplicación $f : E \mapsto F$ se dice:

1. Inyectiva, si:

$$\forall x, y \in E, f(x) = f(y) \Rightarrow x = y \quad (3.1)$$

Es decir, si cada elemento del conjunto final F es imagen de uno o ningún elemento del dominio E , pero no de más de uno.

2. Suryectiva, sobreyectiva, suprayectiva o subyectiva † si:

$$\forall y \in F, \exists x \in E : f(x) = y \quad (3.2)$$

Es decir, si cada elemento del conjunto final F es imagen de al menos un elemento del dominio E .

3. Biyectiva, si es a la vez inyectiva y suryectiva. Es decir, si cada elemento del conjunto final F es imagen de exactamente un elemento del dominio E .

Observación 3.1.1 † Estos cuatro términos son sinónimos. Se pueden utilizar indistintamente.

Ejemplo 3.1.3 Dados los conjuntos $E = \{\text{perro, gato, cabra}\}$ y $F = \{\text{guau, miau, bee, muu}\}$, la aplicación $f : E \mapsto F$ siguiente es una aplicación inyectiva:

$$\begin{cases} f(\text{perro}) = \text{guau} \\ f(\text{gato}) = \text{miau} \\ f(\text{cabra}) = \text{bee} \end{cases}$$

Ejemplo 3.1.4 Dados los conjuntos $E = \{\text{perro, gato, cabra, oveja}\}$ y $F = \{\text{guau, miau, bee}\}$, la aplicación $f : E \mapsto F$ siguiente es una aplicación suryectiva:

$$\begin{cases} f(\text{perro}) = \text{guau} \\ f(\text{gato}) = \text{miau} \\ f(\text{cabra}) = \text{bee} \\ f(\text{oveja}) = \text{bee} \end{cases}$$

Observación 3.1.2 Para poder establecer una aplicación biyectiva entre dos conjuntos E, F dados, es condición necesaria y suficiente que estos tengan el mismo número de elementos, o si son espacios vectoriales, la misma dimensión.

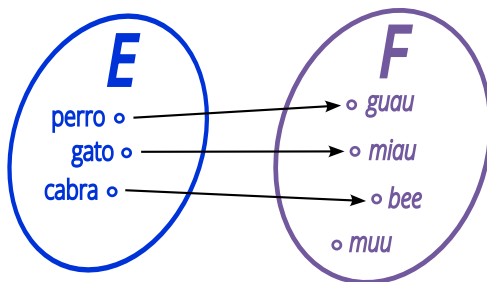


Figura 3.3: La aplicación inyectiva definida en el Ejemplo 3.1.3.

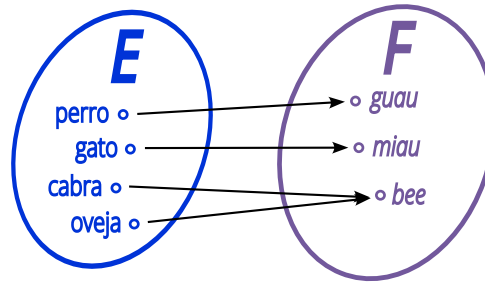


Figura 3.4: La aplicación suryectiva definida en el Ejemplo 3.1.4.

Ejemplo 3.1.5 Dados los conjuntos $E = \{\text{perro, gato, cabra, vaca}\}$ y $F = \{\text{guau, miau, bee, muu}\}$, la aplicación $f : E \mapsto F$ siguiente es una aplicación biyectiva:

$$\begin{cases} f(\text{perro}) = \text{guau} \\ f(\text{gato}) = \text{miau} \\ f(\text{cabra}) = \text{bee} \\ f(\text{vaca}) = \text{muu} \end{cases}$$

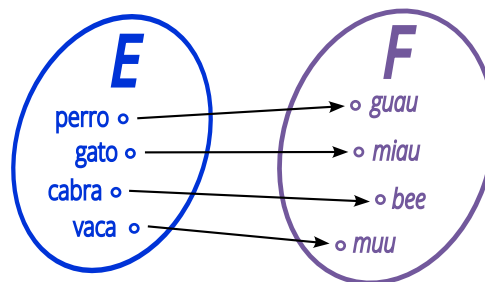


Figura 3.5: La aplicación biyectiva definida en el Ejemplo 3.1.5.

3.1.4 Aplicaciones lineales

Definición 3.1.6 – Aplicación lineal

Se dice que una aplicación $f : E \mapsto F$ entre dos espacios vectoriales es lineal, o un homomorfismo, si verifica las dos condiciones siguientes:

1. La imagen de la suma de elementos del dominio coincide con la suma de las imágenes.

$$\forall \mathbf{u}, \mathbf{v} \in E, f(\mathbf{u} + \mathbf{v}) = f(\mathbf{u}) + f(\mathbf{v}) \quad (3.3)$$

2. La imagen del producto por escalar de un elemento del dominio coincide con el producto por escalar de la imagen.

$$\forall \mathbf{u} \in E, \forall \lambda \in \mathbb{K}, f(\lambda \mathbf{u}) = \lambda f(\mathbf{u}) \quad (3.4)$$

Combinando las condiciones 1 y 2 en una sola se tiene que f es lineal si:

$$\forall \mathbf{u}, \mathbf{v} \in E, \forall \lambda, \mu \in \mathbb{K}, f(\lambda \mathbf{u} + \mu \mathbf{v}) = \lambda f(\mathbf{u}) + \mu f(\mathbf{v}) \quad (3.5)$$

O sea, si la imagen por f de cualquier combinación lineal de elementos del dominio coincide con la combinación lineal de sus imágenes.

E se denomina “espacio inicial”, o “dominio”, de la aplicación lineal f , mientras que F se denomina “espacio final” de f .

Ejemplo 3.1.6 La aplicación $f : \mathbb{R}^3 \mapsto \mathbb{R}^2$ definida mediante la regla de asociación:

$$(x, y, z) \mapsto (2y - z, x + y) \quad (3.6)$$

es lineal, pues para todo $\bar{u}_1 = (x_1, y_1, z_1) \in \mathbb{R}^3$, $\bar{u}_2 = (x_2, y_2, z_2) \in \mathbb{R}^3$ y $\lambda_1, \lambda_2 \in \mathbb{R}$ se cumple $f(\lambda_1 \bar{u}_1 + \lambda_2 \bar{u}_2) = \lambda_1 f(\bar{u}_1) + \lambda_2 f(\bar{u}_2)$.

Sustituyendo, evaluando f y reordenando términos puede comprobarse que:

$$\begin{aligned} f(\lambda_1 \bar{u}_1 + \lambda_2 \bar{u}_2) &= \\ &= f(\lambda_1 x_1 + \lambda_2 x_2, \lambda_1 y_1 + \lambda_2 y_2, \lambda_1 z_1 + \lambda_2 z_2) = \\ &= f(2(\lambda_1 y_1 + \lambda_2 y_2) - (\lambda_1 z_1 + \lambda_2 z_2), \\ &\quad , (\lambda_1 x_1 + \lambda_2 x_2) + (\lambda_1 y_1 + \lambda_2 y_2)) = \\ &= \lambda_1 (2y_1 - z_1, x_1 + y_1) + \lambda_2 (2y_2 - z_2, x_2 + y_2) = \\ &= \lambda_1 f(\bar{u}_1) + \lambda_2 f(\bar{u}_2) \end{aligned} \quad (3.7)$$

Definición 3.1.7 – Isomorfismos, espacios isomorfos y automorfismos

Dados dos espacios vectoriales E, F y la aplicación lineal $f : E \mapsto F$:

1. Si f es biyectiva se dice que f es un isomorfismo y que E y F son espacios isomorfos.
2. Si f es a la vez un isomorfismo y un endomorfismo. Es decir, si es biyectiva y sus espacios inicial y final coinciden, se dice que f es un automorfismo.

Es fácil demostrar que, dado un homomorfismo f cualquiera, la imagen por f del elemento nulo del espacio inicial es el elemento nulo del espacio final:

Proposición 3.1.1 – Imagen del elemento nulo por un homomorfismo

Toda aplicación lineal $f : E \mapsto F$ entre espacios vectoriales verifica:

$$f(0_E) = 0_F$$

Donde $0_E, 0_F$ denotan los elementos nulos de E y F .

Demostración:

Por simple deducción lógica:

$$f(0_E) \stackrel{(*)}{=} f(0 \cdot 0_E) \stackrel{(**)}{=} 0 \cdot f(0_E) = 0_F \quad (3.8)$$

En $(*)$ se aplica la propiedad 3 de la Proposición 1.1.3, y en $(**)$ se aplica la condición (3.5) de linealidad de f .

□

Ejemplo 3.1.7 La aplicación lineal $f : \mathbb{R}^3 \mapsto \mathbb{R}^2$ definida mediante la regla de asociación:

$$(x, y, z) \mapsto (2y - z, x + y) \quad (3.9)$$

cumple:

$$f(\bar{0}_{\mathbb{R}^3}) = f(0, 0, 0) = (2 \cdot 0 - 0, 0 + 0) = (0, 0) = \bar{0}_{\mathbb{R}^2} \quad (3.10)$$

3.1.5 Resumen

1. Una aplicación es una regla de asociación que asigna a cada elemento del conjunto inicial un único elemento del conjunto final. Si los conjuntos inicial y final coinciden, la aplicación es un endomorfismo.
2. Algunas aplicaciones pueden ser inyectivas, suryectivas o biyectivas, siendo estas últimas tanto inyectivas como suryectivas.
3. Se dice que una aplicación es lineal, o un homomorfismo, si esta cumple la condición de linealidad. Es decir, si la imagen de toda combinación lineal de elementos del espacio inicial es igual a la combinación lineal de las imágenes de los elementos.
4. Una aplicación lineal y biyectiva es un isomorfismo. Un isomorfismo cuyos espacios inicial y final coinciden es un automorfismo.

3.2 Expresión de una aplicación lineal respecto de una base

3.2.1 Introducción

En este apartado, se verá que, para definir una aplicación lineal f , basta con conocer las imágenes por f de los elementos de cualquier base \mathcal{B} del espacio inicial de f . Como consecuencia de esta propiedad, puede establecerse una equivalencia entre una aplicación lineal y el producto de un vector de coordenadas respecto de \mathcal{B} de un elemento del espacio inicial por una matriz que, dada una base \mathcal{V} del espacio final de f , caracteriza unívocamente el homomorfismo. A esta se la conoce como la matriz de la aplicación f respecto a las bases \mathcal{B} y \mathcal{V} .

3.2.2 Aplicación lineal respecto a una base

Dados dos espacios vectoriales E, F con dimensiones n y m respectivamente, una aplicación lineal f de E a F y una base $\mathcal{B} = (u_1, \dots, u_n)$ de E . Recordemos por el apartado 1.3.9 que a cada elemento x del espacio inicial se le asocia unívocamente un vector de coordenadas respecto a la base \mathcal{B} denotado $[x]^{\mathcal{B}} = (x_1, \dots, x_n)$, de modo que, por la Definición 1.3.7:

$$x = \sum_{j=1}^n x_j u_j \quad (3.11)$$

Evaluando f en x y teniendo en cuenta la propiedad de linealidad, ecuación (3.5), se obtiene la expresión:

$$f(x) = f\left(\sum_{j=1}^n x_j u_j\right) = \sum_{j=1}^n x_j f(u_j) \quad (3.12)$$

Que permite calcular la imagen por f de un vector x cualquiera del espacio inicial E como combinación lineal de las imágenes por f de los elementos de una base cualquiera de E (Figura 3.6).

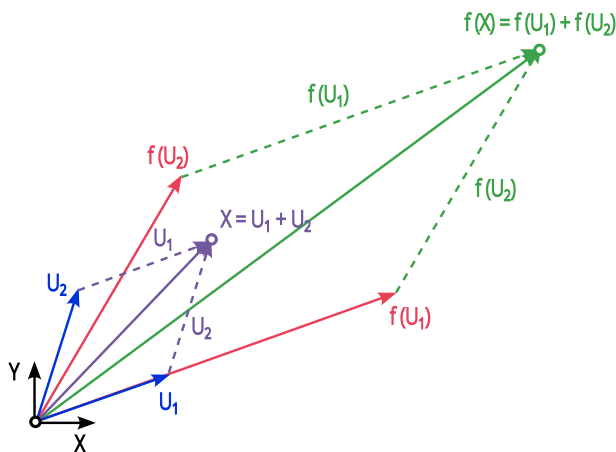


Figura 3.6: La imagen por la aplicación lineal f del vector $\bar{x} = \bar{u}_1 + \bar{u}_2$, $\bar{u}_1, \bar{u}_2 \in \mathbb{R}^2$ puede descomponerse en la suma de las imágenes por f de \bar{u}_1 y \bar{u}_2 en virtud de la fórmula (3.12).

Proposición 3.2.1 – Aplicación lineal respecto a una base

Dadas una aplicación lineal $f : E \mapsto F$ y una base $\mathcal{B} = (\mathbf{u}_1, \dots, \mathbf{u}_n)$ cualquiera de E :

$$f(\mathbf{x}) = \sum_{j=1}^n x_j f(\mathbf{u}_j)$$

Para todo $\mathbf{x} \in E$, $\mathbf{x} = \sum_{j=1}^n x_j \mathbf{u}_j$.

Ejemplo 3.2.1 Las imágenes de los elementos de la base de \mathbb{R}^3 :

$$\mathcal{B} = (\bar{\mathbf{u}}_1, \bar{\mathbf{u}}_2, \bar{\mathbf{u}}_3) = \left(\begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix} \right) \subset \mathbb{R}^3 \quad (3.13)$$

Por la aplicación lineal $f : \mathbb{R}^3 \mapsto \mathbb{R}^2$ definida mediante la regla de asociación $f(x, y, z) = (2y - z, x + y)$ son:

$$f(\bar{\mathbf{u}}_1) = \begin{pmatrix} -2 \\ 1 \end{pmatrix} \quad (3.14)$$

$$f(\bar{\mathbf{u}}_2) = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad (3.15)$$

$$f(\bar{\mathbf{u}}_3) = \begin{pmatrix} 0 \\ -1 \end{pmatrix} \quad (3.16)$$

En virtud de la Proposición 3.2.1, la imagen por f de un $\bar{\mathbf{x}} \in \mathbb{R}^3$ cualquiera puede calcularse según la expresión:

$$f(\bar{\mathbf{x}}) = \lambda_1 \begin{pmatrix} -2 \\ 1 \end{pmatrix} + \lambda_2 \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \lambda_3 \begin{pmatrix} 0 \\ -1 \end{pmatrix} \quad (3.17)$$

Siendo $\lambda_1, \lambda_2, \lambda_3 \in \mathbb{R}$ las coordenadas de $\bar{\mathbf{x}}$ respecto de \mathcal{B} .

3.2.3 Matriz de una aplicación lineal respecto de las bases canónicas

Como se demostrará en este apartado, el producto matricial puede emplearse para representar matemáticamente una aplicación lineal cualquiera f del espacio inicial E en el espacio final F y calcular imágenes por f de los elementos de E . Para ello, es necesario obtener la matriz de la aplicación lineal $f : E \mapsto F$ respecto de unas bases \mathcal{B} y \mathcal{V} de E y F respectivamente.

En primer lugar, se demostrará la expresión de la matriz de una aplicación lineal f de \mathbb{K}^n en \mathbb{K}^m respecto de sus bases canónicas. Hecho esto, no será difícil generalizar el resultado para poder calcular matrices de aplicaciones lineales que relacionen espacios vectoriales cualesquiera E y F dadas unas bases \mathcal{B} y \mathcal{V} de esos espacios.

Dada una aplicación lineal f de \mathbb{K}^n en \mathbb{K}^m , recuérdese que, por la Proposición 3.2.1, conocidas las imágenes por f de los vectores de la base canónica \mathbb{K}^n , la imagen por f de un vector $\bar{x} = \sum_{j=1}^n x_j \bar{e}_j \in \mathbb{K}^n$, donde \bar{e}_j es el j -ésimo vector canónico de \mathbb{K}^n , puede expresarse de la forma:

$$f(\bar{x}) = \sum_{j=1}^n x_j f(\bar{e}_j) \quad (3.18)$$

Con $f(\bar{e}_1), \dots, f(\bar{e}_n)$ pertenecientes a \mathbb{K}^m .

Por otro lado, el vector $f(\bar{e}_j)$ es descomponible como $f(\bar{e}_j) = \sum_{i=1}^m a_{ij} \bar{e}'_i$, donde \bar{e}'_i es el vector canónico i -ésimo del espacio \mathbb{K}^m .

Sustituyendo en la expresión (3.18):

$$\bar{y} = f(\bar{x}) = \sum_{j=1}^n x_j \left(\sum_{i=1}^m a_{ij} \bar{e}'_i \right) = \sum_{i=1}^m \left(\sum_{j=1}^n a_{ij} x_j \right) \bar{e}'_i \quad (3.19)$$

Como se ve arriba, cada una de las m coordenadas de \bar{y} respecto de la base canónica $\mathcal{V} = (\bar{e}'_1, \dots, \bar{e}'_m) \subset \mathbb{R}^m$ se puede expresar mediante la siguiente ecuación algebraica:

$$y_i = \sum_{j=1}^n a_{ij} x_j, \quad 1 \leq i \leq m, \quad 1 \leq j \leq n \quad (3.20)$$

donde (y_1, \dots, y_m) es el vector de coordenadas de la imagen $\bar{y} = f(\bar{x})$ respecto a \mathcal{V} .

Desarrollando, resulta el sistema de ecuaciones siguiente:

$$\begin{aligned} y_1 &= a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n \\ y_2 &= a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n \\ &\vdots \\ y_m &= a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n \end{aligned} \quad (3.21)$$

La representación matricial de este sistema es:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \quad (3.22)$$

Por lo que f puede escribirse como:

$$\bar{y} = f(\bar{x}) = A\bar{x} \quad (3.23)$$

Es decir, como una ecuación matricial que relaciona los vectores del espacio inicial \mathbb{R}^n con sus imágenes por f a través de una matriz A que caracteriza unívocamente f .

Definición 3.2.1 – Matriz de una aplicación lineal respecto de las bases canónicas de \mathbb{R}^n y \mathbb{R}^m

Dada una aplicación lineal $f : \mathbb{K}^n \mapsto \mathbb{K}^m$, se llama matriz de f respecto a las bases canónicas a aquella cuyas columnas son, ordenadamente, las imágenes por f de los vectores canónicos de \mathbb{K}^n :

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

Tal que $C_j(A) = (a_{1j}, \dots, a_{mj})^t = f(\bar{e}_j) = \sum_{i=1}^m a_{ij} \bar{e}_i$.

El número de filas de A coincide con la dimensión m del espacio final de f , y su número de columnas coincide con la dimensión n del espacio inicial de f .

La Definición 3.2.1 establece una correspondencia biunívoca entre aplicaciones lineales de \mathbb{K}^n en \mathbb{K}^m y matrices de m filas y n columnas, de modo que, consideradas las bases canónicas, a cada matriz le corresponde una única aplicación lineal, y viceversa.

Ejemplo 3.2.2 Las imágenes de los vectores de la base canónica del espacio \mathbb{R}^3 por la aplicación lineal $f : \mathbb{R}^3 \mapsto \mathbb{R}^2$ definida mediante la regla de asociación $f(x, y, z) = (2y - z, x + y)$ son:

$$f(\bar{e}_1) = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad (3.24)$$

$$f(\bar{e}_2) = \begin{pmatrix} 2 \\ -1 \end{pmatrix} \quad (3.25)$$

$$f(\bar{e}_3) = \begin{pmatrix} -1 \\ 0 \end{pmatrix} \quad (3.26)$$

Por lo que la matriz de f respecto de las bases canónicas es:

$$A = (f(\bar{e}_1) | f(\bar{e}_2) | f(\bar{e}_3)) = \begin{pmatrix} 0 & 2 & -1 \\ 1 & -1 & 0 \end{pmatrix} \quad (3.27)$$

Así, la aplicación lineal f puede expresarse como:

$$f(\bar{x}) = A\bar{x} = \begin{pmatrix} 0 & 2 & -1 \\ 1 & -1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \quad f(\bar{x}) \in \mathbb{R}^2 \quad (3.28)$$

Para todo $\bar{x} \in \mathbb{R}^3$.

3.2.4 Matriz de una aplicación lineal respecto de dos bases cualesquiera

Supóngase ahora que los espacios inicial y final de f son dos espacios vectoriales cualesquiera E y F y que las bases $\mathcal{B} = (u_1, \dots, u_n)$ y $\mathcal{V} = (v_1, \dots, v_m)$ son dos bases cualesquiera de estos. Con un razonamiento análogo al del apartado 3.2.3, se tiene que cada columna de la matriz A

que caracteriza la aplicación f respecto de las bases \mathcal{B} y \mathcal{V} es igual al vector de coordenadas respecto de \mathcal{V} de la imagen de cada elemento de la base \mathcal{B} :

$$C_j(A) = (a_{1j}, \dots, a_{mj})^t = [f(\mathbf{u}_j)]^{\mathcal{V}} \quad (3.29)$$

La matriz A resultante establece una relación entre el vector de coordenadas respecto a \mathcal{B} de cada elemento en E con el vector de coordenadas respecto a \mathcal{V} de su imagen por f :

$$[f(\mathbf{x})]^{\mathcal{V}} = A[\mathbf{x}]^{\mathcal{B}} \quad (3.30)$$

Definición 3.2.2 — Matriz de una aplicación lineal respecto de dos bases cualesquiera

Dados una aplicación lineal $f : E \mapsto F$ y dos bases \mathcal{B} y \mathcal{V} de E y F respectivamente, se llama matriz de f respecto a las bases \mathcal{B} y \mathcal{V} a aquella cuyas columnas son, ordenadamente, los vectores de coordenadas respecto de \mathcal{V} de las imágenes por f de los elementos de \mathcal{B} :

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \quad (3.31)$$

Tal que $C_j(A) = (a_{1j}, \dots, a_{mj})^t = [f(\mathbf{u}_j)]^{\mathcal{V}}$.

Ejemplo 3.2.3 Dadas las bases:

$$\mathcal{B} = (\bar{\mathbf{u}}_1, \bar{\mathbf{u}}_2, \bar{\mathbf{u}}_3) = \left(\begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix} \right) \subset \mathbb{R}^3 \quad (3.32)$$

$$\mathcal{V} = \left(\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ -2 \end{pmatrix} \right) \subset \mathbb{R}^2 \quad (3.33)$$

Las imágenes por f de los vectores de \mathcal{B} por la aplicación lineal $f : \mathbb{R}^3 \mapsto \mathbb{R}^2$ definida mediante la regla de asociación $f(x, y, z) = (2y - z, x + y)$ son:

$$f(\bar{\mathbf{u}}_1) = \begin{pmatrix} -2 \\ 1 \end{pmatrix} \quad (3.34)$$

$$f(\bar{\mathbf{u}}_2) = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad (3.35)$$

$$f(\bar{\mathbf{u}}_3) = \begin{pmatrix} 0 \\ -1 \end{pmatrix} \quad (3.36)$$

Los vectores de coordenadas de estas imágenes respecto de la base \mathcal{V} son:

$$[f(\bar{u}_1)]^{\mathcal{V}} = \begin{pmatrix} -2 \\ -3/2 \end{pmatrix} \quad (3.37)$$

$$[f(\bar{u}_2)]^{\mathcal{V}} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad (3.38)$$

$$[f(\bar{u}_3)]^{\mathcal{V}} = \begin{pmatrix} 0 \\ 1/2 \end{pmatrix} \quad (3.39)$$

Por lo que la matriz de f respecto de las bases \mathcal{B} y \mathcal{V} queda así:

$$A = ([f(\bar{u}_1)]^{\mathcal{V}} | [f(\bar{u}_2)]^{\mathcal{V}} | [f(\bar{u}_3)]^{\mathcal{V}}) = \begin{pmatrix} -2 & 1 & 0 \\ -3/2 & 0 & 1/2 \end{pmatrix} \quad (3.40)$$

La matriz A asocia vectores de coordenadas respecto de \mathcal{B} de los elementos de \mathbb{R}^3 con los vectores de coordenadas respecto de \mathcal{V} de sus imágenes a través de f :

$$[f(\bar{x})]^{\mathcal{V}} = A[\bar{x}]^{\mathcal{B}} \quad (3.41)$$

3.2.5 Resumen

1. Dadas una aplicación lineal f y una base \mathcal{B} del espacio inicial E , la imagen por f de $x \in E$ puede expresarse como combinación lineal de las imágenes por f de los elementos de la base \mathcal{B} , con coeficientes iguales a las coordenadas de x respecto de \mathcal{B} .
2. Dadas dos bases de dos espacios vectoriales E y F de dimensiones n y m , existe una correspondencia biunívoca entre aplicaciones lineales de E a F y matrices de tamaño $m \times n$. Es decir, a cada aplicación lineal f se le asocia una única matriz A asociada, y viceversa.
3. La matriz A de una aplicación lineal f respecto de dos bases \mathcal{B} , espacio inicial, y \mathcal{V} , espacio final, es una matriz cuyas columnas son los vectores de coordenadas respecto de \mathcal{V} de las imágenes por f de los elementos de la base \mathcal{B} .

3.3 Cambios de base. Semejanza y congruencia de matrices

3.3.1 Introducción

Como se había demostrado en el apartado anterior, dada una aplicación lineal f y dos bases para sus espacios inicial y final se obtiene una matriz asociada a f que es única para cada pareja de bases y que permite calcular las imágenes por f de los elementos del espacio inicial. Por cada pareja de bases escogida, se obtiene una matriz asociada distinta.

A continuación, se verá cómo pasar de una matriz de f a otra matriz de la misma aplicación lineal respecto de distintas bases mediante un operador matricial denominado “matriz de cambio de base”. Entre las distintas matrices asociadas a una misma aplicación lineal se pueden establecer las relaciones de semejanza o congruencia que se definirán al final de este apartado. Cualquier homomorfismo entre espacios vectoriales admite cambios de base, pero por brevedad nos centraremos en los cambios de base de matrices asociadas a endomorfismos lineales: aplicaciones lineales que asocian elementos de un espacio vectorial con elementos de ese mismo espacio vectorial, Definición 3.1.3.

3.3.2 Cambio de base de un vector. Matriz de cambio de base

Sean $\mathcal{B} = (u_1, \dots, u_n)$ y $\mathcal{V} = (v_1, \dots, v_n)$ dos bases cualesquiera de un mismo espacio vectorial V . Por la Definición 1.3.2 de base, es posible expresar los vectores de la base \mathcal{V} como combinaciones lineales de los vectores de \mathcal{B} :

$$\begin{aligned} v_1 &= p_{11}u_1 + p_{21}u_2 + \dots + p_{n1}u_n \\ v_2 &= p_{12}u_1 + p_{22}u_2 + \dots + p_{n2}u_n \\ &\vdots \\ v_n &= p_{1n}u_1 + p_{2n}u_2 + \dots + p_{nn}u_n \end{aligned} \quad (3.42)$$

Con coeficientes escalares a los que, por conveniencia, se denotan p_{ij} , siendo i y j los índices de los vectores de las bases \mathcal{B} y \mathcal{V} correspondientes a cada uno de ellos, respectivamente.

Todo elemento $x \in V$ se puede descomponer como combinación lineal de los elementos de cada una de las bases, de modo tal que $x = \sum_{i=1}^n x_i u_i = \sum_{i=1}^n x'_i v_i$. Sustituyendo los v_i por las descomposiciones de la ecuación (3.42) y denotando $[x]^{\mathcal{B}}$ y $[x]^{\mathcal{V}}$ a los vectores de coordenadas de x respecto de cada base, puede comprobarse que:

$$[x]^{\mathcal{B}} = P_{\mathcal{V}}^{\mathcal{B}} [x]^{\mathcal{V}} \quad (3.43)$$

Donde $P_{\mathcal{V}}^{\mathcal{B}}$ denota la matriz de cambio de base de \mathcal{B} a \mathcal{V} . Sus columnas contienen los coeficientes p_{1i}, \dots, p_{ni} de la ecuación (3.42); las coordenadas de los vectores de la base \mathcal{V} respecto de los de \mathcal{B} .

Definición 3.3.1 — Matriz de cambio de base

Se llama matriz de cambio de base (o matriz de paso) de \mathcal{B} a \mathcal{V} a una matriz de la forma:

$$P_{\mathcal{V}}^{\mathcal{B}} = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{pmatrix} = ([v_1]^{\mathcal{B}} | \dots | [v_n]^{\mathcal{B}})$$

Cuyas columnas son los vectores de coordenadas respecto de la base \mathcal{B} de cada elemento de la base \mathcal{V} .

Ejemplo 3.3.1 Dadas las bases:

$$\mathcal{B} = (\bar{u}_1, \bar{u}_2) = \left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right) \quad (3.44)$$

$$\mathcal{V} = (\bar{v}_1, \bar{v}_2) = \left(\begin{pmatrix} 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right) \quad (3.45)$$

Los vectores de coordenadas de los elementos de \mathcal{V} respecto de la base \mathcal{B} se calculan resolviendo estos sistemas lineales:

$$\begin{pmatrix} 0 \\ -1 \end{pmatrix} = p_{11} \begin{pmatrix} 1 \\ 0 \end{pmatrix} + p_{21} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \Rightarrow p_{11} = 0, p_{21} = -1 \quad (3.46)$$

$$\begin{pmatrix} 1 \\ 1 \end{pmatrix} = p_{12} \begin{pmatrix} 1 \\ 0 \end{pmatrix} + p_{22} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \Rightarrow p_{12} = 1, p_{22} = 1 \quad (3.47)$$

\mathcal{B} es la base canónica de \mathbb{R}^2 , por lo que las coordenadas de los vectores de \mathcal{V} coinciden con las componentes de los vectores que contiene. La matriz de cambio de base de \mathcal{B} a \mathcal{V} es, por tanto, la siguiente:

$$P_{\mathcal{V}}^{\mathcal{B}} = ([\bar{v}_1]^{\mathcal{B}} | [\bar{v}_2]^{\mathcal{B}}) = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1 & 1 \end{pmatrix} \quad (3.48)$$

Esta matriz transforma vectores de coordenadas de los elementos de \mathbb{R}^2 respecto de \mathcal{V} en sus vectores de coordenadas respecto de \mathcal{B} :

$$[\bar{x}]^{\mathcal{B}} = P_{\mathcal{V}}^{\mathcal{B}} [\bar{x}]^{\mathcal{V}} \quad \forall \bar{x} \in \mathbb{R}^2 \quad (3.49)$$

Su inversa \dagger es matriz de cambio de base de \mathcal{V} a \mathcal{B} :

$$P_{\mathcal{B}}^{\mathcal{V}} = (P_{\mathcal{V}}^{\mathcal{B}})^{-1} = \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix} \quad (3.50)$$

Esta transforma vectores de coordenadas de los elementos de \mathbb{R}^2 respecto de \mathcal{B} en sus vectores de coordenadas respecto de \mathcal{V} :

$$[\bar{x}]^{\mathcal{V}} = P_{\mathcal{B}}^{\mathcal{V}} [\bar{x}]^{\mathcal{B}} \quad \forall \bar{x} \in \mathbb{R}^2 \quad (3.51)$$

Observación 3.3.1 \dagger La inversa de $P_{\mathcal{V}}^{\mathcal{B}}$, Definición 3.3.1, realiza el cambio de base opuesto:

$$\begin{aligned} [\bar{x}]^{\mathcal{B}} &= P_{\mathcal{V}}^{\mathcal{B}} [\bar{x}]^{\mathcal{V}} \Rightarrow \\ \Rightarrow [\bar{x}]^{\mathcal{V}} &= P^{-1} [\bar{x}]^{\mathcal{B}} \Rightarrow \\ \Rightarrow P^{-1} &= P_{\mathcal{B}}^{\mathcal{V}}. \end{aligned}$$

Las columnas de P^{-1} contienen las coordenadas de cada elemento de la base \mathcal{B} respecto de la base \mathcal{V} .

3.3.3 Cambio de base de la matriz de un endomorfismo lineal

Dado un endomorfismo lineal $f : E \mapsto E$ y una base \mathcal{B} de E , supóngase que $A^{\mathcal{B}}$ es la matriz de f (cuadrada de orden n) respecto de la base \mathcal{B} y que $A^{\mathcal{B}'}$ es la matriz de f , también cuadrada de orden n , respecto de la base \mathcal{B}' . Se tiene que:

$$[f(x)]^{\mathcal{B}} = A^{\mathcal{B}}[x]^{\mathcal{B}} \quad (3.52)$$

$$[f(x)]^{\mathcal{B}'} = A^{\mathcal{B}'}[x]^{\mathcal{B}'} \quad (3.53)$$

Llamando $P_{\mathcal{B}'}^{\mathcal{B}}$ a la matriz de cambio de base de \mathcal{B} a \mathcal{B}' :

$$[x]^{\mathcal{B}} = P_{\mathcal{B}'}^{\mathcal{B}}[x]^{\mathcal{B}'} \quad (3.54)$$

$$[f(x)]^{\mathcal{B}} = P_{\mathcal{B}'}^{\mathcal{B}}[f(x)]^{\mathcal{B}'} \quad (3.55)$$

Sustituyendo (3.54) y (3.55) en la ecuación (3.52):

$$P_{\mathcal{B}'}^{\mathcal{B}}[f(x)]^{\mathcal{B}'} = A^{\mathcal{B}}P_{\mathcal{B}'}^{\mathcal{B}}[x]^{\mathcal{B}'} \quad (3.56)$$

Introduciendo ahora la expresión (3.53):

$$P_{\mathcal{B}'}^{\mathcal{B}}A^{\mathcal{B}'}[x]^{\mathcal{B}'} = A^{\mathcal{B}}P_{\mathcal{B}'}^{\mathcal{B}}[x]^{\mathcal{B}'} \quad (3.57)$$

Eliminando $[x]^{\mathcal{B}'}$ y despejando $A^{\mathcal{B}'}$, se obtiene:

$$A^{\mathcal{B}'} = (P_{\mathcal{B}'}^{\mathcal{B}})^{-1}A^{\mathcal{B}}P_{\mathcal{B}'}^{\mathcal{B}} \quad (3.58)$$

La identidad (3.58) relaciona dos matrices del mismo endomorfismo lineal f expresadas respecto de bases diferentes. Esta permite transformar la matriz $A^{\mathcal{B}}$ de f respecto de la base \mathcal{B} en la matriz $A^{\mathcal{B}'}$ de la misma aplicación f respecto de otra base \mathcal{B}' , siendo $P_{\mathcal{B}'}^{\mathcal{B}}$ una matriz de paso cuyas columnas contienen las coordenadas de cada elemento de la nueva base \mathcal{B}' respecto de los de la base \mathcal{B} original.

Proposición 3.3.1 – Cambio de base de la matriz de un endomorfismo lineal

Dadas dos bases $\mathcal{B}, \mathcal{B}'$ del espacio vectorial E , la matriz $A^{\mathcal{B}}$ asociada al endomorfismo lineal $f : E \mapsto E$ respecto de la base \mathcal{B} y la matriz $P_{\mathcal{B}'}^{\mathcal{B}}$ de cambio de base de \mathcal{B} a \mathcal{B}' , la matriz:

$$A^{\mathcal{B}'} = (P_{\mathcal{B}'}^{\mathcal{B}})^{-1}A^{\mathcal{B}}P_{\mathcal{B}'}^{\mathcal{B}}$$

es la matriz asociada al endomorfismo lineal f respecto de la base \mathcal{B}' .

Ejemplo 3.3.2 El endomorfismo lineal $f : \mathbb{R}^3 \mapsto \mathbb{R}^3$, $(x, y, z) \mapsto (y - z, x, -z)$ tiene como matriz asociada respecto de la base canónica de \mathbb{R}^3 , $\mathcal{B} = (\bar{e}_1, \bar{e}_2, \bar{e}_3)$, la siguiente:

$$A = \begin{pmatrix} 0 & 1 & -1 \\ 1 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix} \quad (3.59)$$

Considérese ahora la siguiente base de \mathbb{R}^3 :

$$\mathcal{B}' = (\bar{u}_1, \bar{u}_2, \bar{u}_3) = \left(\begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \\ 0 \end{pmatrix} \right) \subset \mathbb{R}^3 \quad (3.60)$$

Los vectores de coordenadas de los elementos de \mathcal{B}' respecto de la base canónica \mathcal{B} pueden calcularse resolviendo un sistema lineal por cada elemento de \mathcal{B}' .

$$\bar{u}_1 = p_{11}\bar{e}_1 + p_{21}\bar{e}_2 + p_{31}\bar{e}_3 \Rightarrow p_{11} = 0, p_{21} = 1, p_{31} = -1 \quad (3.61)$$

$$\bar{u}_2 = p_{12}\bar{e}_1 + p_{22}\bar{e}_2 + p_{32}\bar{e}_3 \Rightarrow p_{12} = 1, p_{22} = 0, p_{32} = 0 \quad (3.62)$$

$$\bar{u}_3 = p_{13}\bar{e}_1 + p_{23}\bar{e}_2 + p_{33}\bar{e}_3 \Rightarrow p_{13} = 0, p_{23} = -1, p_{33} = 0 \quad (3.63)$$

Como en el Ejemplo 3.3.1, las coordenadas coinciden con los componentes de cada vector de \mathcal{B}' debido a que \mathcal{B} es la base canónica de \mathbb{R}^3 .

Con estos coeficientes, la matriz de cambio de base de \mathcal{B} a \mathcal{B}' queda:

$$P_{\mathcal{B}'}^{\mathcal{B}} = ([\bar{u}_1]^{\mathcal{B}} | [\bar{u}_2]^{\mathcal{B}} | [\bar{u}_3]^{\mathcal{B}}) = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & -1 \\ -1 & 0 & 0 \end{pmatrix} \quad (3.64)$$

La inversa de esta última matriz es:

$$(P_{\mathcal{B}'}^{\mathcal{B}})^{-1} = P_{\mathcal{B}}^{\mathcal{B}'} = \begin{pmatrix} 0 & 0 & -1 \\ 1 & 0 & 0 \\ 0 & -1 & -1 \end{pmatrix} \quad (3.65)$$

Para obtener la matriz de f respecto de la base \mathcal{B}' se aplica la identidad (3.58).

$$A^{\mathcal{B}'} = (P_{\mathcal{B}'}^{\mathcal{B}})^{-1} A^{\mathcal{B}} P_{\mathcal{B}'}^{\mathcal{B}} = \begin{pmatrix} -1 & 0 & 0 \\ 2 & 0 & -1 \\ -1 & -1 & 0 \end{pmatrix} \quad (3.66)$$

También es posible efectuar cambios de base de matrices que no son de endomorfismos. Es decir, de matrices asociadas a aplicaciones lineales de E en F con $E \neq F$, pero estas se encuentran fuera del alcance de este libro. El lector interesado encontrará en muchos libros de texto de álgebra lineal una formulación más general para cambiar de base matrices asociadas a aplicaciones lineales de todo tipo.

3.3.4 Matrices semejantes

Dos matrices asociadas a una misma aplicación lineal respecto de diferentes bases cumplen una relación de equivalencia † que se conoce como de “semejanza”. En general, dadas dos matrices cuadradas A y B del mismo orden, si existe una matriz invertible P tal que el producto $P^{-1}AP$ coincide con B, se dice que A y B son matrices semejantes la una respecto a la otra.

Definición 3.3.2 – Matrices semejantes

Dos matrices cuadradas A y B del mismo orden se dicen semejantes si existe una matriz invertible P tal que $P^{-1}AP = B$

Ejemplo 3.3.3 Puede comprobarse que las matrices

$$A = \begin{pmatrix} 0 & 1 & -1 \\ 1 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix} \quad (3.67)$$

$$B = \begin{pmatrix} -1 & 0 & 0 \\ 2 & 0 & -1 \\ -1 & -1 & 0 \end{pmatrix} \quad (3.68)$$

$$P = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & -1 \\ -1 & 0 & 0 \end{pmatrix} \quad (3.69)$$

cumplen la relación:

$$P^{-1}AP = B \quad (3.70)$$

Por lo que A y B son matrices semejantes. Es decir, son matrices asociadas al mismo endomorfismo lineal $f : \mathbb{R}^3 \mapsto \mathbb{R}^3$ respecto de dos bases de \mathbb{R}^3 diferentes relacionadas por la matriz de cambio de base P.

3.3.5 Matrices congruentes

Existe una relación más restrictiva, denominada “de congruencia”, que algunas parejas de matrices cuadradas del mismo orden cumplen. Esta es su definición:

Definición 3.3.3 – Matrices congruentes

Dos matrices cuadradas A y B del mismo orden se dicen congruentes si existe una matriz cuadrada P tal que $P^tAP = B$.

Ejemplos de matrices congruentes son las matrices asociadas a una misma forma bilineal respecto de bases diferentes, apartado 7.1.3.

Observación 3.3.2 † Dado un conjunto A en el que se define una relación entre sus elementos, denotada con el símbolo “ \sim ”, se dice que esta relación es de equivalencia si cumple las tres propiedades siguientes:

1. Reflexividad: $x \sim x \forall x \in A$
2. Simetría: $x \sim y \Leftrightarrow y \sim x$
 $\forall x, y \in A$
3. Transitividad: $(x \sim y \wedge y \sim z) \Rightarrow x \sim z \forall x, y, z \in A$

La relación de igualdad es una relación de equivalencia. Sin embargo, las relaciones mayor que y menor que no lo son al no cumplir las propiedades de reflexividad y simetría.

3.3.6 Resumen

1. Dos vectores de coordenadas del mismo elemento de un espacio vectorial respecto de bases diferentes están relacionados entre sí por una matriz de cambio de base.
2. La columna i -ésima de la matriz de cambio de base de \mathcal{B} a \mathcal{V} es el vector de coordenadas del i -ésimo elemento de \mathcal{V} respecto de la base \mathcal{B} .
3. Dos matrices asociadas a un mismo endomorfismo lineal respecto de diferentes bases están ligadas por una relación de semejanza a través de una matriz de cambio de base que relaciona ambas matrices. Algunas parejas de matrices cumplen una relación más restrictiva llamada “de congruencia” donde la inversión matricial se sustituye por la transposición.

3.4 Definición y propiedades de la imagen, núcleo y rango de una aplicación lineal o matriz. Teorema del rango-nulidad

3.4.1 Introducción

Toda aplicación entre espacios vectoriales tiene asociados dos subconjuntos denominados “imagen” y “núcleo”. Si la aplicación es lineal, se denomina “rango” a la dimensión de su imagen. La relación entre aplicaciones lineales y matrices demostrada en la sección 3.2 permite definir la imagen, el núcleo y el rango de una matriz y asociarlo con la imagen, el núcleo y el rango de la aplicación lineal correspondiente.

Si una aplicación lineal es inyectiva o suryectiva, su imagen, núcleo y rango, o los de cualquiera de sus matrices asociadas, cumplen ciertas propiedades que se verán en este apartado. También se demostrará una fórmula que relaciona entre sí el rango, la dimensión del núcleo y la dimensión del espacio final de una aplicación lineal cualquiera.

3.4.2 Imagen y núcleo de una aplicación

Recordemos que, por la Definición 3.1.1, una aplicación $f : E \mapsto F$ transforma elementos de E en elementos de F , de modo tal que para todo $x \in E$:

$$y = f(x) \in F \quad (3.71)$$

Si el conjunto final F es un espacio vectorial, probablemente existan elementos del dominio E cuya imagen por f sea igual al elemento nulo o cero de F :

$$y = f(x) = 0_F \text{ para algún } x \in E \quad (3.72)$$

Así pues, toda aplicación f lleva asociados dos subconjuntos: su imagen en F y su núcleo en E .

Definición 3.4.1 – Imagen y núcleo de una aplicación

Dada una aplicación $f : E \mapsto F$.

1. Se define la imagen de f como el conjunto:

$$\text{Im } f = \{f(x) : x \in E\} \subseteq F \quad (3.73)$$

2. Se define el núcleo de f como el conjunto:

$$\ker f = \{u \in E : f(u) = 0_F\} \subseteq E \quad (3.74)$$

3.4.3 Imagen y núcleo de una matriz

La correspondencia biunívoca entre aplicaciones lineales y matrices demostrada en la sección 3.2 permite definir los conjuntos imagen y núcleo de una matriz A y relacionarlos con la imagen y el núcleo de una aplicación lineal f asociada a A .

Toda matriz $A \in \mathbb{K}^{m \times n}$ transforma vectores columna de $\mathbb{K}^{1 \times n}$ en vectores columna de $\mathbb{K}^{1 \times m}$ mediante el producto matricial, asociando a cada $\bar{x} \in \mathbb{K}^{1 \times n}$ el vector:

$$\bar{y} = A\bar{x} \in \mathbb{K}^{1 \times m} \quad (3.75)$$

Por tanto, toda matriz $A \in \mathbb{K}^{m \times n}$ lleva asociados dos conjuntos: su imagen en \mathbb{K}^m y su núcleo en \mathbb{K}^n .

Definición 3.4.2 — Imagen y núcleo de una matriz

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

1. Se define la imagen de A como el conjunto:

$$\text{Im } A = \{A\bar{x} : \bar{x} \in \mathbb{K}^n\} \subseteq \mathbb{K}^m \quad (3.76)$$

2. Se define el núcleo de A como el conjunto:

$$\ker A = \{\bar{u} \in \mathbb{K}^n : A\bar{u} = \bar{0}\} \subseteq \mathbb{K}^n \quad (3.77)$$

Si A es la matriz de cierta aplicación lineal f , los conjuntos imagen y núcleo de A coinciden con los de f .

Ejercicio 3.4.1 † Demostrar que los subconjuntos imagen y núcleo de una matriz o aplicación lineal entre espacios vectoriales son siempre subespacios.

Es fácil demostrar † que la imagen y el núcleo de una matriz son subespacios vectoriales de \mathbb{K}^m y \mathbb{K}^n , respectivamente. Para ello, basta con comprobar que, para toda matriz A , cualquier combinación lineal de vectores de la imagen o al núcleo de A , Definición 3.4.2, pertenece también a la imagen o al núcleo de A .

Por ello, y por la correspondencia entre matrices y homomorfismos discutida en la sección 3.2, la imagen y el núcleo de una aplicación lineal f entre los espacios vectoriales E y F son subespacios, respectivamente, de F y de E .

En el siguiente ejemplo se muestra que la imagen de una matriz coincide siempre con la clausura lineal de sus columnas:

Ejemplo 3.4.1 Se tiene el endomorfismo lineal $f : \mathbb{R}^3 \mapsto \mathbb{R}^3$ definido como $f(x, y, z) = (y - x, x + 2z, y + 2z)$. Las imágenes por f de los vectores de la base canónica de \mathbb{R}^3 son:

$$f(1, 0, 0) = (-1, 1, 0) \quad (3.78)$$

$$f(0, 1, 0) = (1, 0, 1) \quad (3.79)$$

$$f(0, 0, 1) = (0, 2, 2) \quad (3.80)$$

Por lo que su matriz asociada respecto de la base canónica de \mathbb{R}^3 es:

$$A = \begin{pmatrix} -1 & 1 & 0 \\ 1 & 0 & 2 \\ 0 & 1 & 2 \end{pmatrix} \quad (3.81)$$

La base de referencia es la base canónica, por lo que $\text{Im } A$ coincide con $\text{Im } f$.

$$\begin{aligned} \text{Im } A &= \text{Im } f = \{A\bar{x} : \bar{x} \in \mathbb{R}^3\} = \\ &= \left\{ \begin{pmatrix} -1 & 1 & 0 \\ 1 & 0 & 2 \\ 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} : (x_1, x_2, x_3) \in \mathbb{R}^3 \right\} \stackrel{(*)}{=} \\ &\stackrel{(*)}{=} \left\{ x_1 \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} + x_3 \begin{pmatrix} 0 \\ 2 \\ 2 \end{pmatrix} : (x_1, x_2, x_3) \in \mathbb{R}^3 \right\} \end{aligned} \quad (3.82)$$

En (*) se aplica la propiedad 1 de la Proposición 2.2.5, según la cual el producto de una matriz por un vector columna se puede escribir como una combinación lineal de las columnas de la matriz, con coeficientes iguales a los elementos del vector.

Ejemplo 3.4.2 Sean la aplicación lineal f del Ejemplo 3.4.1 y su matriz asociada A respecto de las bases canónicas, ecuación (3.81), se tiene que el núcleo de A , $\ker A$, coincide con el de la aplicación f :

$$\begin{aligned} \ker A &= \ker f = \{\bar{u} \in \mathbb{R}^3 : A\bar{u} = \bar{0}\} = \\ &= \left\{ (x_1, x_2, x_3) \in \mathbb{R}^3 : \begin{pmatrix} -1 & 1 & 0 \\ 1 & 0 & 2 \\ 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \right\} \end{aligned} \quad (3.83)$$

Para calcularlo, se resuelve el sistema lineal homogéneo:

$$\begin{cases} x_2 - x_1 = 0 & (3.84a) \end{cases}$$

$$\begin{cases} 2x_3 + x_1 = 0 & (3.84b) \end{cases}$$

$$\begin{cases} 2x_3 + x_2 = 0 & (3.84c) \end{cases}$$

Estas son tres ecuaciones implícitas del núcleo de A . Nótese que la primera ecuación puede obtenerse restandole a la tercera ecuación la segunda, por lo que la dimensión del subespacio núcleo de A es igual a $3 - 2 = 1$.

Parametrizando x_1 :

$$x_1 = \lambda \in \mathbb{R} \Rightarrow x_2 = \lambda \Rightarrow x_3 = -\lambda/2 \Rightarrow \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \lambda \begin{pmatrix} 2 \\ 2 \\ -1 \end{pmatrix} \quad (3.85)$$

Por lo que $\ker A = \ker f = L[(2, 2, -1)]$.

3.4.4 Subespacio fila y subespacio columna de una matriz

Como se ve en el Ejemplo 3.4.1, el producto $A\bar{x}$ entre una matriz $A \in \mathbb{K}^{m \times n}$ y un vector columna $\bar{x} \in \mathbb{K}^{1 \times n}$ puede escribirse como una combinación lineal de las columnas de A . Es decir, como:

$$A\bar{x} = A \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = x_1 C_1(A) + \cdots + x_n C_n(A) \quad (3.86)$$

De la ecuación (3.86) se deduce que un vector en $\mathbb{K}^{1 \times n}$ puede representarse de la forma $A\bar{x}$, si y sólo si, es combinación lineal de las columnas de A . En consecuencia, la imagen de A coincide con el subespacio generado por las columnas de A , llamado también “subespacio columna” de A . Además, al transponer A , sus filas se convierten en columnas y viceversa, lo cual permite definir el subespacio fila de una matriz A como aquel generado por sus filas o, equivalentemente, como la imagen de su transpuesta. Es decir, como $\text{Im } A^t$.

Definición 3.4.3 – Subespacios columna y fila de una matriz

Sea la matriz $A \in \mathbb{K}^{m \times n}$:

1. El subespacio columna de una matriz $A \in \mathbb{K}^{m \times n}$ es el subespacio en \mathbb{K}^m engendrado por las columnas de A . Coincide con $\text{Im } A$.
2. El subespacio fila de una matriz $A \in \mathbb{K}^{m \times n}$ es el subespacio en \mathbb{K}^n engendrado por las filas de A . Coincide con $\text{Im } A^t$.

Ejemplo 3.4.3

1. El subespacio columna de la matriz A del Ejemplo 3.4.1 es la clausura lineal de sus columnas. Es decir, el subespacio:

$$\begin{aligned} & L[\{C_1(A), C_2(A), C_3(A)\}] = \\ & = L\left[\left\{\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 2 \\ 2 \end{pmatrix}\right\}\right] = \text{Im } A. \end{aligned} \quad (3.87)$$

2. El subespacio fila de A es la clausura lineal de sus filas. Es decir, el subespacio:

$$\begin{aligned} & L[\{F_1(A), F_2(A), F_3(A)\}] = \\ & = L[\{(-1, 1, 0), (1, 0, 2), (0, 1, 2)\}] = \text{Im } A^t. \end{aligned} \quad (3.88)$$

3.4.5 Imagen y suryectividad

Recordemos por la Definición 3.1.5 que una aplicación $f : E \mapsto F$ se dice sobreyectiva si $f(E) = F$ o, lo que es lo mismo, si $\forall y \in F, \exists x \in E$ tal que $f(x) = y$. Como consecuencia de esta definición y de la Proposición 1.3.3 que afirma que la dimensión de $H \subseteq V$ coincide con la de V , si y sólo si, $H = V$, toda aplicación lineal sobreyectiva entre espacios vectoriales cumple la siguiente propiedad:

Proposición 3.4.1 – Imagen de una aplicación suryectiva

Una aplicación lineal entre espacios vectoriales es suryectiva, si y sólo si, la dimensión de su imagen coincide con la de su espacio final. Es decir, sea $f : E \mapsto F$ lineal, con E, F espacios vectoriales:

$$f \text{ es suryectiva} \Leftrightarrow \dim \operatorname{Im} f = \dim F = m$$

Naturalmente, si f es suryectiva, la imagen de cualquier matriz $A \in \mathbb{K}^{m \times n}$ asociada a f tiene dimensión m , y viceversa, si la dimensión de la imagen de una matriz coincide con su número de filas m , se trata de una matriz asociada a una aplicación lineal suryectiva.

3.4.6 Núcleo e inyectividad

Recuérdese que, por la Definición 3.1.5, una aplicación $f : E \mapsto F$ se dice inyectiva si cada elemento del conjunto final F es imagen de, como máximo, un elemento del dominio E . Es decir, si $\forall x, y \in E, f(x) = f(y)$ implica $x = y$. En este apartado, se exploran las propiedades del núcleo de una aplicación lineal inyectiva entre espacios vectoriales.

Dada una aplicación lineal inyectiva $f : E \mapsto F$, supóngase que $\ker f \neq \{0_E\}$. Es decir, que su núcleo tiene dimensión mayor que uno. Entonces existirían elementos no nulos del espacio inicial cuya imagen es $0_F, \exists u \in \ker f, u \neq 0_E$, tal que $0_F = f(0_E) = f(u)$, por lo que f no cumpliría la condición de inyectividad. Con este razonamiento se demuestra que si f es inyectiva, su núcleo coincide con el subespacio $\{0_E\}$.

Para demostrar el converso de esta última implicación. Es decir, para probar que si el núcleo de una aplicación lineal f es el subespacio nulo $\{0_E\}$ esta es inyectiva, supóngase que existen $u, v \in E$ tales que $f(u) = f(v)$. Entonces, por linealidad, $f(u - v) = f(u) - f(v) = 0$, luego $u - v \in \ker f = \{0\}$, de donde se deduce que $u - v = 0$, así que $u = v$ y por tanto f es inyectiva. De este modo, se demuestra que si el núcleo de una aplicación lineal es el subespacio nulo, esta es inyectiva.

Se tiene así que para ver si una aplicación lineal entre espacios vectoriales es inyectiva, basta con comprobar si su núcleo es el subespacio nulo:

Proposición 3.4.2 – Núcleo de una aplicación inyectiva

Sea la aplicación lineal $f : E \mapsto F$ entre espacios vectoriales, f es inyectiva, si y sólo si $\ker f = \{0\}$.

Ejemplo 3.4.4 El núcleo de la aplicación lineal f del Ejemplo 3.4.1 y de su matriz asociada A es, como se ve en el Ejemplo 3.4.2, $\ker A = \ker f = L[(2, 2, -1)]$. Por tanto, $\dim \ker f = 1 \neq 0$ y f no son una aplicación inyectiva por la Proposición 3.4.2.

Ejemplo 3.4.5 La matriz asociada a la aplicación lineal:

$$g(x, y, z) = (y - x, x + 2z, z - y) \quad (3.89)$$

Respecto de las bases canónicas es la siguiente:

$$B = \begin{pmatrix} -1 & 1 & 0 \\ 1 & 0 & 2 \\ 0 & -1 & 1 \end{pmatrix} \quad (3.90)$$

Su núcleo coincide con el conjunto de soluciones del sistema lineal homogéneo:

$$\begin{pmatrix} -1 & 1 & 0 \\ 1 & 0 & 2 \\ 0 & -1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \Rightarrow \begin{cases} x_2 - x_1 = 0 \\ 2x_3 - x_2 = 0 \\ x_3 - x_2 = 0 \end{cases} \quad (3.91)$$

Las tres ecuaciones del sistema (3.91) son independientes entre sí, por lo que tiene una única solución: el vector $(0, 0, 0)$. Por tanto, el núcleo de f es el subespacio nulo de \mathbb{R}^3 . Como consecuencia de la Proposición 3.4.2, se tiene que g es una aplicación lineal inyectiva.

Ejemplo 3.4.6 La matriz asociada a la aplicación lineal:

$$h : \mathbb{R}^3 \mapsto \mathbb{R}^2, h(x, y, z) = (2y - z, x + y) \quad (3.92)$$

Respecto de las bases canónicas es la siguiente:

$$C = \begin{pmatrix} 0 & 2 & -1 \\ 1 & 1 & 0 \end{pmatrix} \quad (3.93)$$

Su núcleo coincide con el conjunto de soluciones del sistema lineal homogéneo:

$$\begin{pmatrix} 0 & 2 & -1 \\ 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Rightarrow \begin{cases} x_2 - x_1 = 0 \\ x_1 + x_2 = 0 \end{cases} \quad (3.94)$$

El conjunto de soluciones de este sistema forma un subespacio en \mathbb{R}^3 de dimensión dos. Parametrizando $x_2 = \lambda \in \mathbb{R}$, se tiene que $\ker C = \ker g = L[(1, -1, 2)]$, por lo que la aplicación lineal h no es inyectiva por la Proposición 3.4.2.

3.4.7 Rango de una matriz

A la dimensión de la imagen de una matriz se la conoce como su “rango por columnas”:

Definición 3.4.4 – Rango por columnas de una matriz

El rango por columnas de una matriz $A \in \mathbb{K}^{m \times n}$ es la dimensión de su subespacio columna. Es decir, de su imagen:

$$r(A) = \dim \operatorname{Im} A = \dim (L\{C_1(A), \dots, C_n(A)\})$$

Dicho de otra forma, el rango de A es el número máximo de columnas de A linealmente independientes.

Por otro lado, se llama “rango por filas de una matriz” a la dimensión de su subespacio fila. Es decir, al número máximo de filas de A linealmente independientes, que coincide con la dimensión de la imagen de su transpuesta A^t :

Definición 3.4.5 – Rango por filas de una matriz

El rango por filas de una matriz $A \in \mathbb{K}^{m \times n}$ es la dimensión de su subespacio fila. Es decir, la dimensión de la imagen de su transpuesta:

$$r(A^t) = \dim \operatorname{Im} A^t = \dim (\mathcal{L}[\{F_1(A), \dots, F_n(A)\}])$$

Considérese una matriz genérica $A \in \mathbb{K}^{m \times n}$ cuyo rango por columnas sea $r \leq n$. Para explorar la posible relación que guardan ambos rangos, se construye una matriz $C \in \mathbb{K}^{m \times r}$ que contenga r columnas cualesquiera de A linealmente independientes de modo que ambas matrices tengan el mismo subespacio columna y, en consecuencia, un rango por columnas igual a r . Las n columnas de A son combinaciones lineales de las r columnas de C . Con los coeficientes escalares de esas combinaciones lineales se puede construir una matriz $F \in \mathbb{K}^{r \times n}$ tal que:

$$A = CF \tag{3.95}$$

Además, en virtud de las propiedades del producto matricial, Proposición 2.2.5, se tiene que las filas de A son combinaciones lineales de las filas de F . F tiene r filas, por lo que el rango por filas de A se encuentra superiormente acotado (ha de ser igual o menor) por r .

Como el rango por filas de A coincide, por definición, con el rango por columnas de A^t , puede aplicarse el mismo razonamiento a la matriz A^t . Descomponiéndola como en la ecuación (3.95), como el producto de una matriz $C \in \mathbb{K}^{m \times r}$ que contenga r columnas linealmente independientes de A^t por una matriz $F \in \mathbb{K}^{r \times n}$, se concluye que el número de columnas linealmente independientes de A^t debe ser igual a r , pues de lo contrario A^t tendría rango por filas menor que r y, por tanto, el rango por columnas de A sería menor que r .

Se demuestra así que el rango por filas de cualquier matriz A coincide con su rango por columnas:

Proposición 3.4.3 – Relación entre los rangos por filas y columnas de una matriz

Los rangos por filas y columnas de toda matriz $A \in \mathbb{K}^{m \times n}$ coinciden:

$$\begin{aligned} r(A) &= \dim \operatorname{Im} A = \dim (\mathcal{L}[\{C_1(A), \dots, C_n(A)\}]) = \\ &= \dim (\mathcal{L}[\{F_1(A), \dots, F_m(A)\}]) = \dim \operatorname{Im} A^t = r(A^t) \end{aligned}$$

Al coincidir el rango por filas con el rango por columnas, puede hablarse simplemente del rango de una matriz, definido como la dimensión de su imagen, o equivalentemente, de su subespacio fila asociado:

Definición 3.4.6 – Rango de una matriz

A la dimensión de la imagen de una matriz $A \in \mathbb{K}^{m \times n}$ se la conoce como el rango de A . Este se denota como $r(A)$.

Definición 3.4.7 – Matrices de rango máximo y de rango deficiente

Si el rango de la matriz $A \in \mathbb{K}^{m \times n}$ coincide con su máximo teórico, $r(A) = \min\{m, n\}$, se dice que A es de rango máximo. De lo contrario, si $r(A) < \min\{m, n\}$, se dice que A es de rango deficiente.

Ejemplo 3.4.7 Sean la aplicación lineal f del Ejemplo 3.4.1 y su matriz asociada A , la imagen de A es:

$$\begin{aligned} \text{Im } A &= L[\{C_1(A), C_2(A), C_3(A)\}] = \\ &= L\left[\left\{\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 2 \\ 2 \end{pmatrix}\right\}\right] = L[\{\bar{u}_1, \bar{u}_2, \bar{u}_3\}] \end{aligned} \quad (3.96)$$

Véase que $\bar{u}_3 = 2\bar{u}_1 + 2\bar{u}_2$, mientras que \bar{u}_1 y \bar{u}_2 son linealmente independientes, por lo que estos forman una base de $\text{Im } A$, y en consecuencia:

$$r(A) = \dim \text{Im } A = \dim L[(\bar{u}_1, \bar{u}_2)] = 2 \quad (3.97)$$

Ejemplo 3.4.8 Sean la aplicación lineal g del Ejemplo 3.4.5 y su matriz asociada B , la imagen de B es:

$$\begin{aligned} \text{Im } B &= L[\{C_1(B), C_2(B), C_3(B)\}] = \\ &= L\left[\left\{\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 2 \\ 1 \end{pmatrix}\right\}\right] = L[\{\bar{u}_1, \bar{u}_2, \bar{u}_3\}] \end{aligned} \quad (3.98)$$

En este caso, $\det(B)$ es distinto de cero, por lo que los vectores \bar{u}_1 , \bar{u}_2 y \bar{u}_3 forman una familia libre y, por tanto, una base de $\text{Im } B$, por lo que:

$$r(B) = \dim \text{Im } B = \dim L[(\bar{u}_1, \bar{u}_2, \bar{u}_3)] = 3 \quad (3.99)$$

Ejemplo 3.4.9 Sean la aplicación lineal h del Ejemplo 3.4.6 y su matriz asociada C , la imagen de C es:

$$\text{Im } C = L\left[\left\{\begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \end{pmatrix}\right\}\right] = L[\{\bar{u}_1, \bar{u}_2, \bar{u}_3\}] \quad (3.100)$$

Como consecuencia de la Proposición 1.3.3, la dimensión de la imagen de C no puede ser mayor que dos. Cualquier pareja de columnas de la matriz, por ejemplo \bar{u}_1 y \bar{u}_3 , constituye una familia libre que es base de $\text{Im } C$, por tanto:

$$r(C) = \dim \text{Im } C = \dim L[(\bar{u}_1, \bar{u}_3)] = 2 \quad (3.101)$$

3.4.8 Cálculo del rango mediante reducción *gaussiana*

Recordemos que, como se mostró en el apartado 2.5.3, realizar operaciones elementales sobre las filas de una matriz $A \in \mathbb{K}^{m \times n}$ equivale a pre-multiplicar por sucesivas matrices elementales:

$$E_k \dots E_2 E_1 A \quad (3.102)$$

Por el razonamiento expuesto al demostrar el Teorema 1.3.1, la clausura lineal de las filas de A no cambia cuando se efectúan permutaciones de filas, o se multiplican por escalares no nulos, o a unas se les suman los múltiplos de otras. En consecuencia, las operaciones elementales por filas dejan invariante el subespacio fila de A y, por tanto, su rango. Ocurre lo mismo respecto a las columnas.

Si mediante este proceso se transforma la matriz en una de tipo trapecoidal, da igual si superior o inferior, contando el número de filas no nulas de la matriz reducida se obtendría de manera inmediata el rango de la matriz A .

Igualmente puede procederse efectuando operaciones elementales sobre las columnas, las cuales conservan invariante el subespacio columna de A , o incluso combinando ambos tipos de operaciones, transformando la matriz en otra cuyo rango sea trivial de obtener. A este proceso se le suele conocer como reducción *gaussiana*.

Ejemplo 3.4.10 Sean la aplicación lineal f del Ejemplo 3.4.1 y su matriz A asociada respecto de las bases canónicas, su imagen puede calcularse reduciéndola mediante operaciones elementales por filas:

$$A = \begin{pmatrix} -1 & 1 & 0 \\ 1 & 0 & 2 \\ 0 & 1 & 2 \end{pmatrix} \xrightarrow{F'_2 = F_2 + F_1} \begin{pmatrix} -1 & 1 & 0 \\ 0 & 1 & 2 \\ 0 & 1 & 2 \end{pmatrix}$$

Véase que las dos últimas filas de la matriz reducida coinciden, mientras que la primera fila es linealmente independiente de estas. Por tanto, $r(A) = 2$.

Ejemplo 3.4.11 Sean la aplicación lineal g del Ejemplo 3.4.5 y su matriz B asociada respecto de las bases canónicas, su imagen puede calcularse reduciéndola mediante operaciones elementales por filas:

$$B = \begin{pmatrix} -1 & 1 & 0 \\ 1 & 0 & 2 \\ 0 & -1 & 1 \end{pmatrix} \xrightarrow{\substack{F'_2 = F_2 + F_1 \\ F'_3 = F_3 + F'_2}} \begin{pmatrix} -1 & 1 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 3 \end{pmatrix}$$

Ninguna de las filas de su forma triangular reducida es nula, por lo que $r(B) = 3$.

A continuación, se enumeran algunas propiedades básicas del rango:

Proposición 3.4.4 – Propiedades del rango de una matriz

1. El rango de una matriz $A \in \mathbb{K}^{m \times n}$ no puede exceder nunca el mínimo entre m y n .

$$r(A) \leq \min(m, n) \quad (3.103)$$

2. El rango de la suma de dos matrices $A, B \in \mathbb{K}^{m \times n}$ no puede exceder nunca la suma de sus rangos.

$$r(A + B) \leq r(A) + r(B) \quad (3.104)$$

Sean A, B dos matrices cualesquiera tales que el producto matricial AB tenga sentido:

3. El rango del producto no puede exceder el rango de ninguna de las matrices multiplicadas.

$$r(AB) \leq \min(r(A), r(B)) \quad (3.105)$$

4. Si A tiene rango máximo por columnas:

$$r(AB) = r(B) \quad (3.106)$$

5. Si B tiene rango máximo por filas:

$$r(AB) = r(A) \quad (3.107)$$

6. Para cualquier matriz A , se verifica:

$$r(A^h A) = r(AA^h) = r(A) \quad (3.108)$$

Se demostrarán todas las propiedades salvo la 6, que se demostrará tras haber enunciado el Teorema 3.4.5 del rango-nulidad.

Demostración:

1. Esto es una consecuencia directa de la definición de rango, Definición 3.4.6, y de la Proposición 1.3.3. El rango de una matriz se define como la dimensión de su subespacio columna, que no puede exceder el número de columnas de la matriz. De acuerdo con la Proposición 3.4.3, esa coincide con la dimensión de su subespacio fila, que no puede exceder el número de filas de la matriz. En consecuencia, toda matriz cumple la acotación (3.103).
2. Denótese los subespacios columna de A, B y $A + B$, respectivamente, como C_A, C_B y C_{A+B} . Las columnas de $A + B$ son la suma de las columnas de A más las de B , por lo que el subespacio C_{A+B} está contenido en la suma de C_A y C_B .

El mismo razonamiento puede realizarse sobre las filas de las tres matrices: denotando los subespacios fila de A, B y $A + B$, respecti-

vamente, como F_A , F_B y F_{A+B} , las filas de $A + B$ son la suma de las filas de A más las de B , por lo que el subespacio F_{A+B} está contenido en la suma de F_A y F_B .

Como consecuencia del Teorema 1.4.5 de *Grassmann*, $\dim C_A + C_B \leq \dim C_A + \dim C_B$ y $\dim F_A + F_B \leq \dim F_A + \dim F_B$. De estas acotaciones, de la Definición 3.4.6 de rango y de la igualdad de los rangos por filas y por columnas, Proposición 3.4.3, se deduce la identidad (3.104).

3. Según la propiedad 1 de la Proposición 2.2.5, las columnas de AB son combinaciones lineales de las de A , por lo que el subespacio columna de AB está contenido en el de A . Por otro lado, por la propiedad 2 de la Proposición 2.2.5, las filas de AB son combinaciones lineales de las de B , por lo que el subespacio fila de AB está contenido en el de B . Aplicando la Definición 3.4.6 de rango y por la igualdad entre los rangos por filas y por columnas, Proposición 3.4.3, se deduce la desigualdad (3.105).
4. A tiene rango máximo por columnas, por lo que $r(A)$ coincide con su número de columnas n , que por la propiedad 1 es mayor o igual que su número de filas m . Para que su producto matricial con B sea compatible, el número de filas de B debe coincidir con n , mientras que para su número de columnas r no hay restricciones. Como consecuencia de ello y de la propiedad 1, el rango de B puede variar entre 1 y n ; por lo que, por la propiedad 3, se tiene que $r(AB) \leq n$.

Para demostrar que el rango de AB es estrictamente igual al rango de B , es necesario comprobar que el subespacio fila de AB coincide con el de B . El producto $\bar{x}^t AB$, donde \bar{x} es un vector columna genérico de $\mathbb{K}^{m \times 1}$, es combinación lineal de las filas de AB , por lo que pertenece al subespacio fila de AB , mientras que el producto $\bar{x}^t A$ genera un vector perteneciente al subespacio fila de A . La acotación $r(A) = n \leq m$ implica que el subespacio fila de A coincide con \mathbb{K}^m , mientras que $\{\bar{x}^t B : \bar{x} \in \mathbb{K}^{m \times 1}\} \subseteq \mathbb{R}^n$ es el conjunto de todas las combinaciones lineales posibles de las filas de B , que por la Definición 3.4.3 es el subespacio fila de B .

De este modo, se demuestra la igualdad entre los subespacios fila de B y AB , y con ello la igualdad entre los rangos de ambas matrices.

5. B tiene rango máximo por filas, por lo que $r(A)$ coincide con su número de filas m , que por la propiedad 1 es igual o mayor que su número de columnas n . Para que su producto matricial con A sea compatible, el número de columnas de A debe coincidir con m , mientras que para su número de filas r no hay restricciones. Como consecuencia de ello y de la propiedad 1, el rango de A puede variar entre 1 y m ; por lo que, por la propiedad 3, se tiene que $r(AB) \leq m$. Para demostrar que el rango de AB es estrictamente igual al rango de A , es necesario comprobar que el subespacio columna de AB coincide con el de A . El producto $AB\bar{x}$, donde \bar{x} es un vector columna genérico de $\mathbb{K}^{n \times 1}$, es combinación lineal de las columnas de AB , por lo que pertenece al subespacio columna de AB , mientras que el

producto $B\bar{x}^t$ genera un vector perteneciente al subespacio columna de B . La acotación $r(B) = m \leq n$ implica que el subespacio columna de B coincide con \mathbb{K}^m , mientras que $\{A\bar{x} : \bar{x} \in \mathbb{K}^{n \times 1}\} \subseteq \mathbb{R}^m$ es el conjunto de todas las combinaciones lineales posibles de las columnas de A , que por la Definición 3.4.3 es el subespacio columna de A .

Así se demuestra la igualdad entre los subespacios columna de A y AB , y con ello la igualdad entre los rangos de ambas matrices.

□

3.4.9 Rango de una aplicación lineal

En el apartado 3.4.3, se razonó que el subespacio imagen de una aplicación lineal f coincide con la imagen de una matriz A asociada a f , lo que implica que la dimensión del subespacio imagen de f coincide con la dimensión del subespacio imagen de A , que por definición es el rango de A . Esto permite definir el concepto de rango de una aplicación lineal y relacionarlo con el rango de una matriz asociada cualquiera.

Definición 3.4.8 – Rango de una aplicación lineal

Se define el rango de una aplicación lineal f entre espacios vectoriales como la dimensión de su subespacio imagen: $r(f) = \dim f$.

Ejemplo 3.4.12 El rango de la matriz A asociada a la aplicación lineal f de la ecuación (3.59) del Ejemplo 3.4.1 es $r(A) = 2$, por lo que $r(f) = 2$.

Ejemplo 3.4.13 El rango de la matriz B asociada a la aplicación lineal g de la ecuación (3.90) del Ejemplo 3.4.5 es $r(B) = 3$, por lo que $r(g) = 3$.

Ejemplo 3.4.14 El rango de la matriz C asociada a la aplicación lineal h de la ecuación (3.93) del Ejemplo 3.4.6 es $r(C) = 2$, por lo que $r(h) = 2$.

3.4.10 Teorema del rango-nulidad

A continuación, se mostrará que existe una relación entre el rango de cualquier matriz y la dimensión de su núcleo. Además, por la equivalencia entre matrices y aplicaciones lineales demostrada en el apartado 3.2.4, estas últimas cumplen una relación análoga.

Sea A una matriz con n columnas, su núcleo, $\ker A$, es un subespacio en \mathbb{K}^n cuya dimensión es $k \leq n$. Denótese $(\bar{u}_1, \dots, \bar{u}_k)$ a un conjunto ordenado que es base de $\ker A$. Por el Teorema 1.3.4 de completión de bases, añadiendo al conjunto $n - k$ vectores $\bar{u}_{k+1}, \dots, \bar{u}_n$ linealmente independientes entre si y a los contenidos en el conjunto ordenado, este se convierte en una base de \mathbb{K}^n a la que se denotará en adelante \mathcal{B} .

$$\mathbb{K}^n = L[(\bar{u}_1, \dots, \bar{u}_k, \bar{u}_{k+1}, \dots, \bar{u}_n)] = L[\mathcal{B}] \quad (3.109)$$

Por la Definición 3.4.2, el conjunto de todos los transformados por A de los vectores columna de \mathbb{K}^n constituye el subespacio imagen de A . Al pre-multiplicar A por cada uno de los vectores de la base \mathcal{B} se obtiene un sistema generador de $\text{Im } A$ que puede, o no, ser base.

$$\text{Im } A = L[\{A\bar{u}_1, \dots, A\bar{u}_k, A\bar{u}_{k+1}, \dots, A\bar{u}_n\}] \quad (3.110)$$

Los k primeros vectores de \mathcal{B} pertenecen a $\ker A$, por lo que $A\bar{u}_1, \dots, A\bar{u}_k = \bar{0}$ y la expresión (3.110) se reduce a:

$$\text{Im } A = L[\{\bar{0}, \dots, \bar{0}, A\bar{u}_{k+1}, \dots, A\bar{u}_n\}] = L[\{A\bar{u}_{k+1}, \dots, A\bar{u}_n\}] \quad (3.111)$$

Los transformados por A de los últimos $n - k$ vectores de \mathcal{B} son linealmente independientes entre sí. Esto puede demostrarse por reducción al absurdo mediante el siguiente razonamiento: Supóngase que los $n - k$ transformados por A son linealmente dependientes entre sí. Por la Definición 1.2.3, existirían coeficientes escalares $\lambda_{k+1}, \dots, \lambda_n \in \mathbb{K}$ no todos nulos tales que:

$$\lambda_{k+1}A\bar{u}_{k+1} + \dots + \lambda_n A\bar{u}_n = \bar{0} \quad (3.112)$$

Sacando factor común A , se tiene:

$$A(\lambda_{k+1}\bar{u}_{k+1} + \dots + \lambda_n\bar{u}_n) = \bar{0} \quad (3.113)$$

Por lo que la combinación lineal $\lambda_{k+1}\bar{u}_{k+1} + \dots + \lambda_n\bar{u}_n$ pertenecería a $\ker A$. Pero este resultado contradice la hipótesis inicial de que la dimensión de $\ker A$ es k , pues al ser los vectores $\bar{u}_{k+1}, \dots, \bar{u}_n$ linealmente independientes de $\bar{u}_1, \dots, \bar{u}_k$, no sería posible expresar los primeros como combinación lineal de los últimos. Se concluye que $\bar{u}_{k+1}, \dots, \bar{u}_n$ son vectores linealmente independientes que forman una base de $\text{Im } A$, y que la imagen y el núcleo de A cumplen la relación:

$$\dim \text{Im } A = r(A) = n - k = n - \dim \ker A \quad (3.114)$$

Toda matriz $A \in \mathbb{K}^{m \times n}$ cumple la identidad (3.114) que se acaba de demostrar, lo que permite enunciar el siguiente teorema:

Teorema 3.4.5 – Teorema del rango-nulidad

Sea A una matriz compuesta por n columnas, entonces:

$$\dim \ker A + r(A) = n \quad (3.115)$$

Y sea $f : E \mapsto F$ una aplicación lineal entre espacios vectoriales con $\dim E = n$, entonces:

$$\dim \ker f + r(f) = n \quad (3.116)$$

El teorema del rango-nulidad permite demostrar fácilmente la propiedad 6 de la Proposición 3.4.4:

Demostración: (propiedad 6)

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

6. De acuerdo con la Definición 3.4.1, el núcleo del producto de A con su adjunta, la matriz AA^h , es el conjunto $\{\bar{x} \in \mathbb{K}^m : AA^h\bar{x} = \bar{0}\}$. Si $\bar{x} \in \mathbb{K}^m$ cumple la identidad $A^h\bar{x} = \bar{0}$, este cumple también $AA^h\bar{x} = \bar{0}$, por lo que $\ker A^h \subseteq \ker AA^h$.

Por otro lado, si $\bar{x} \in \ker AA^h$, entonces $AA^h\bar{x} = \bar{0} \Rightarrow \bar{x}^h AA^h\bar{x} = \bar{0} \Rightarrow (A^h\bar{x})^h A^h\bar{x} = \bar{0} \Rightarrow A^h\bar{x} = \bar{0}$, por lo que $\ker AA^h \subseteq \ker A^h$.

Se tiene así que $\ker AA^h = \ker A^h$. Por tanto, la dimensión de $\ker AA^h$ coincide con la de $\ker A^h$. Aplicando tanto la Definición 3.4.6 como el Teorema 3.4.5 del rango-nulidad se obtiene la relación $r(AA^h) = r(A^h) = r(A)$.

Con un razonamiento análogo, se demuestra que $r(A^h A) = r(A)$.

□

Los conceptos de imagen, núcleo y rango de una matriz y el teorema del rango-nulidad permiten extender la Proposición 3.5.5 de forma que existan varias formas equivalentes de comprobar si una matriz cuadrada es regular:

Corolario 3.4.6 – Caracterización de las matrices invertibles

Sea A una matriz cuadrada de orden n , las siguientes afirmaciones son equivalentes:

1. A es invertible.
2. $\det(A) \neq 0$.
3. $r(A) = n$.
4. $\ker A = \{\bar{0}\}$.

Demostración:

La equivalencia de las afirmaciones [(1) ... (4)] puede demostrarse mediante la siguiente cadena de implicaciones: Primero, que (2) implica (3). Después, que (3) implica (4). A continuación, que (4) implica (1). Y por último, para cerrar el círculo, que (1) implica (2).

- (2) \Rightarrow (3): sea A una matriz cuadrada de orden n , el contrapositivo de la proposición “ $\det(A) \neq 0 \Rightarrow r(A) = n$ ” es la proposición “ $r(A) < n \Rightarrow \det(A) = 0$ ”.

Para demostrar la veracidad de esta última proposición, supóngase que el rango de A es menor que n . Esto significa que los subespacios fila y columna de la matriz tienen dimensión menor que n , por lo que al menos una de sus filas, o columnas, es combinación lineal de las restantes.

Por la propiedad 3 de la Proposición 2.3.2, esto implica que $\det(A) = 0$. Se demuestra así que si $\det(A) \neq 0$, el rango de A es n .

- (3) \Rightarrow (4): esto es una consecuencia directa del Teorema 3.4.5 del rango-nulidad. Sea A una matriz cuadrada de orden n y de rango completo, sustituyendo en (3.115) se obtiene:

$$\begin{aligned} \dim \ker A + r(A) &= \dim \ker A + n = n \Rightarrow \\ &\Rightarrow \dim \ker A = n - n = 0. \end{aligned} \quad (3.117)$$

- (4) \Rightarrow (1): sea A una matriz cuadrada de orden n cuyo núcleo es el subespacio nulo de \mathbb{K}^n . Supóngase que existe algún $\bar{x} \in \mathbb{K}^n$ no nulo tal que:

$$A\bar{x} = \bar{0} \quad (3.118)$$

Multiplicando ambos miembros de la identidad (3.118) por la inversa de A :

$$A^{-1}A\bar{x} = A^{-1}\bar{0} \Rightarrow I\bar{x} = \bar{0} \Rightarrow \bar{x} = \bar{0} \quad (3.119)$$

Sin embargo, este último resultado contradice la suposición de que exista algún $\bar{x} \neq \bar{0}$ tal que $A\bar{x} = \bar{0}$. Así, se demuestra por reducción al absurdo que $\ker A = \{\bar{0}\}$, lo que a su vez implica que A es invertible.

- (1) \Rightarrow (2): esta es la implicación más difícil de demostrar. Como se razonó en el apartado 2.5.4, si A es una matriz invertible, existe una secuencia de operaciones elementales por filas que la transforma en la matriz identidad:

$$\begin{aligned} E_k \dots E_2 E_1 A &= I \Rightarrow E_k \dots E_2 E_1 = A^{-1} \\ A E'_1 E'_2 \dots E'_k &= I \Rightarrow E'_1 E'_2 \dots E'_k = A^{-1} \end{aligned} \quad (3.120)$$

Las operaciones elementales mantienen invariante el rango de la matriz en la que se efectúan. El rango de la matriz identidad de orden n es igual a n . De acuerdo con la implicación (4) \Rightarrow (1), si el rango de A fuese menor que n , en cuyo caso el determinante de A sería cero, no sería posible encontrar una secuencia de operaciones elementales por filas, o por columnas, que reduzca A a la matriz identidad, por lo que A no sería invertible. El contrapositivo de esta última implicación es lo que se quería demostrar: que si A es invertible, su determinante es distinto de cero.

□

En virtud del corolario que se acaba de demostrar, existen tres modos de comprobar que una matriz cuadrada es invertible: comprobar que su rango sea n , que su núcleo tenga dimensión cero, o bien que su determinante sea distinto de cero.

La siguiente proposición enumera algunas condiciones necesarias para que un homomorfismo sea inyectivo y/o suryectivo:

Proposición 3.4.7 – Condiciones necesarias para la inyectividad o suryectividad de un homomorfismo

Dados dos espacios E, F con dimensiones n y m , respectivamente, y la aplicación lineal $f : E \mapsto F$:

1. Una condición necesaria para que f sea inyectiva es que $m \geq n$.

$$f \text{ inyectiva} \Rightarrow m \geq n \quad (3.121)$$

2. Una condición necesaria para que f sea suryectiva es que $m \leq n$.

$$f \text{ suryectiva} \Rightarrow m \leq n \quad (3.122)$$

3. En el caso de que $m = n$, si f es inyectiva, también es suryectiva y viceversa.

$$f \text{ inyectiva} \Leftrightarrow f \text{ suryectiva} \quad (3.123)$$

Ejercicio 3.4.2 † Demostrar las tres propiedades enumeradas en la Proposición 3.4.7.

Estas propiedades pueden demostrarse fácilmente † a partir del Teorema 3.4.5 del rango-nulidad y de las dos propiedades que cumplen el rango y el núcleo de las aplicaciones lineales, suryectivas e inyectivas. A saber, las Proposiciones 3.4.1 y 3.4.2.

Ejemplo 3.4.15 Sean la aplicación lineal f del Ejemplo 3.4.1 y su matriz asociada A respecto de las bases canónicas. Se tiene:

$$r(A) = r(f) = 2 \quad (3.124)$$

$$\dim \ker A = \dim \ker f = 1 \quad (3.125)$$

Se comprueba que se cumple el teorema del rango-nulidad:

$$\dim \ker f + r(f) = n \Rightarrow 1 + 2 = 3 \quad (3.126)$$

Ejemplo 3.4.16 Sean la aplicación lineal g del Ejemplo 3.4.5 y su matriz asociada B respecto de las bases canónicas. Se tiene:

$$r(B) = r(g) = 3 \quad (3.127)$$

$$\dim \ker A = \dim \ker f = 0 \quad (3.128)$$

Se comprueba que se cumple el teorema del rango-nulidad:

$$\dim \ker f + r(f) = n \Rightarrow 0 + 3 = 3 \quad (3.129)$$

Ejemplo 3.4.17 Sea la aplicación lineal h del Ejemplo 3.4.6 y su matriz asociada C respecto de las bases canónicas. Se tiene:

$$r(C) = r(h) = 2 \quad (3.130)$$

$$\dim \ker A = \dim \ker f = 1 \quad (3.131)$$

Se comprueba que se cumple el teorema del rango-nulidad:

$$\dim \ker f + r(f) = n \Rightarrow 1 + 2 = 3 \quad (3.132)$$

3.4.11 Resumen

1. Se define la imagen de una aplicación como el subconjunto de elementos del espacio final que son imagen de al menos algún elemento del espacio inicial.
2. Se define el núcleo de una aplicación como el subconjunto de elementos del espacio inicial cuya imagen es el elemento cero del espacio final.
3. Si f es una aplicación lineal entre espacios vectoriales, ambos subconjuntos son subespacios.
4. Debido a la correspondencia entre matrices y aplicaciones lineales, a toda matriz se le puede asociar un subespacio núcleo y un subespacio imagen.
5. Los subespacios columna y fila de una matriz coinciden, respectivamente, con la imagen de la matriz y de su transpuesta.
6. El espacio final de una aplicación lineal suryectiva coincide con su imagen, por lo que ambos tienen la misma dimensión.
7. El núcleo de una aplicación lineal inyectiva es el subespacio nulo, y toda aplicación lineal cuyo núcleo sea el subespacio nulo es inyectiva.
8. El rango por filas de una matriz es la dimensión de su subespacio fila. El rango por columnas es la dimensión de su subespacio columna. Ambos rangos coinciden siempre.
9. Las operaciones elementales sobre matrices dejan inalterado su rango. Por ello, un modo de calcular el rango de una matriz es reducirla a forma trapezoidal. El número de filas o columnas no nulas de la matriz reducida coincide con su rango. A este método se le conoce como “reducción *gaussiana*”.
10. La suma de la dimensión del núcleo de una matriz más su rango coincide siempre con su número de columnas. Como consecuencia, una matriz cuadrada es regular, si y sólo si, su rango coincide con su orden, o equivalentemente, si su núcleo tiene dimensión cero o su determinante no es nulo.

3.5 Composición e inversas de aplicaciones lineales

3.5.1 Introducción

En matemáticas, componer funciones consiste en concatenarlas de manera que el resultado, o imagen, de una se convierta en la entrada de la otra. Si las funciones concatenadas son aplicaciones lineales entre espacios vectoriales, y a cada una de ellas se le asocia una matriz, el producto de las matrices asociadas a cada aplicación es una matriz asociada a la composición de las aplicaciones.

Del mismo modo, y como cabría esperar, la inversa de una matriz regular asociada a cierta aplicación lineal biyectiva es una matriz asociada a la inversa de esa misma aplicación.

3.5.2 Definición y propiedades de la composición de aplicaciones

Definición 3.5.1 — Composición de aplicaciones

Dadas las aplicaciones $f : E \mapsto F$ y $g : F \mapsto G$, se define una nueva aplicación mediante este esquema:

$$E \xrightarrow{f} F \xrightarrow{g} G$$

A la aplicación resultante $E \mapsto G$, denotada $g \circ f$, se le llama composición de g con f .

Como resulta obvio por la definición anterior, una condición necesaria y suficiente para que dos aplicaciones se puedan componer es que el conjunto final de la primera aplicación coincida con el conjunto inicial de la segunda.

Ejemplo 3.5.1 Considerando las aplicaciones lineales siguientes:

$$f : \mathbb{R}^2 \mapsto \mathbb{R}^3, (x, y) \mapsto (x - y, y/2, x) \quad (3.133)$$

$$g : \mathbb{R}^3 \mapsto \mathbb{R}^2, (x, y, z) \mapsto (2y - z, x) \quad (3.134)$$

La composición de g con f , denotada $g \circ f : \mathbb{R}^2 \mapsto \mathbb{R}^2$, es:

$$\bar{x} \in \mathbb{R}^2 \xrightarrow{f} f(\bar{x}) \in \mathbb{R}^3 \xrightarrow{g} (g \circ f)(\bar{x}) \in \mathbb{R}^2 \quad (3.135)$$

La imagen del vector $\bar{u} = (1, 1) \in \mathbb{R}^2$ por $g \circ f$ es igual a:

$$\begin{aligned} (g \circ f)(\bar{u}) &= g(f(\bar{u})) = g(1 - 1, 1/2, 1) = \\ &= g(0, 1/2, 1) = (2 \cdot 1/2 - 1, 0) = \\ &= (0, 0) \in \mathbb{R}^2 \end{aligned} \quad (3.136)$$

Proposición 3.5.1 – Propiedades de la composición de aplicaciones

1. Dadas tres aplicaciones $f : E \mapsto F$, $g : F \mapsto G$ y $h : G \mapsto H$, se tiene que:

$$(h \circ g) \circ f = h \circ (g \circ f) \quad (3.137)$$

por lo que la composición de aplicaciones es asociativa.

2. En general, la composición de aplicaciones no es conmutativa. Si $f, g : E \mapsto E$ son dos endomorfismos definidos en un mismo conjunto E , $g \circ f$ no siempre coincidirá con $f \circ g$.

3. Para toda aplicación $f : E \mapsto F$ se tiene que:

$$f \circ I_E = f \quad (3.138)$$

$$I_F \circ f = f \quad (3.139)$$

Siendo I_E, I_F las funciones identidad asociadas a los conjuntos E y F , respectivamente.

4. La composición de dos aplicaciones inyectivas es siempre inyectiva.

$$g, f \text{ inyectivas} \Rightarrow g \circ f \text{ inyectiva} \quad (3.140)$$

5. La composición de dos aplicaciones suryectivas es siempre suryectiva.

$$g, f \text{ suryectivas} \Rightarrow g \circ f \text{ suryectiva} \quad (3.141)$$

6. La composición de dos aplicaciones biyectivas es siempre biyectiva.

$$g, f \text{ biyectivas} \Rightarrow g \circ f \text{ biyectiva} \quad (3.142)$$

7. Una condición necesaria para que $g \circ f$ sea inyectiva es que f sea inyectiva. Es decir:

$$g \circ f \text{ inyectiva} \Rightarrow f \text{ inyectiva} \quad (3.143)$$

8. Una condición necesaria para que $g \circ f$ sea suryectiva es que g sea suryectiva. Es decir:

$$g \circ f \text{ suryectiva} \Rightarrow g \text{ suryectiva} \quad (3.144)$$

9. Dos condiciones necesarias para que $g \circ f$ sea biyectiva es que f sea inyectiva y que g sea suryectiva. Por tanto:

$$g \circ f \text{ biyectiva} \Rightarrow f \text{ inyectiva y } g \text{ suryectiva} \quad (3.145)$$

En particular, todas las aplicaciones lineales cumplen estas propiedades. Esto se demostrará rigurosamente más adelante, tras haber probado la equivalencia entre la composición de aplicaciones lineales y el producto de sus matrices asociadas.

3.5.3 Composición de aplicaciones lineales

Suponiendo que E , F y G son tres espacios vectoriales y que $f : E \mapsto F$ y $g : F \mapsto G$ son aplicaciones lineales, para toda pareja de elementos $u, v \in E$ y de coeficientes escalares $\lambda, \mu \in \mathbb{K}$ se cumple:

$$\begin{aligned} (g \circ f)(\lambda u + \mu v) &= g(f(\lambda u + \mu v)) = g(\lambda f(u) + \mu f(v)) = \\ &= \lambda g(f(u)) + \mu g(f(v)) = \\ &= \lambda(g \circ f)(u) + \mu(g \circ f)(v) \end{aligned} \quad (3.146)$$

De este modo, se demuestra que toda composición de aplicaciones lineales es también lineal:

$$(g \circ f)(\lambda u + \mu v) = \lambda(g \circ f)(u) + \mu(g \circ f)(v) \quad (3.147)$$

Proposición 3.5.2 – Linealidad de una composición de homomorfismos

Toda composición de aplicaciones lineales es lineal.

Ejemplo 3.5.2 Las aplicaciones f y g del Ejemplo 3.5.1 son lineales. Por lo que su composición $g \circ f$ es también lineal, en virtud de la Proposición 3.5.2.

3.5.4 Matriz de una composición de aplicaciones lineales

Se tienen:

1. Tres espacios vectoriales E , F y G de dimensiones r , n y m de los que se extraen, respectivamente, las bases \mathcal{B} , \mathcal{V} y \mathcal{W} .
2. Una aplicación lineal $f : E \mapsto F$ que tiene asociada la matriz $B \in \mathbb{K}^{n \times r}$ respecto de las bases \mathcal{B} y \mathcal{V} .
3. Una aplicación lineal $g : F \mapsto G$ que tiene asociada la matriz $A \in \mathbb{K}^{m \times n}$ respecto de las bases \mathcal{V} y \mathcal{W} .
4. La composición de ambas aplicaciones $g \circ f$, que tiene asociada la matriz $C \in \mathbb{K}^{m \times r}$ respecto de las bases \mathcal{B} y \mathcal{W} .

Denótese por $y \in F$ la imagen por f de un elemento x cualquiera de E . La imagen por g de y se denota como $z \in G$. Los vectores de coordenadas de x , y y z respecto de \mathcal{B} , \mathcal{V} y \mathcal{W} están relacionados entre sí a través de las matrices A y B :

$$[y]^{\mathcal{V}} = B[x]^{\mathcal{B}} \Rightarrow y_k = \sum_{j=1}^r b_{kj} x_j, \quad 1 \leq k \leq n \quad (3.148)$$

$$[z]^{\mathcal{W}} = A[y]^{\mathcal{V}} \Rightarrow z_i = \sum_{k=1}^n a_{ik} y_k, \quad 1 \leq i \leq m \quad (3.149)$$

Sustituyendo (3.148) en la ecuación (3.149):

$$\begin{aligned}
 z = ABx \Rightarrow z_i &= \sum_{k=1}^n a_{ik} \left(\sum_{j=1}^r b_{kj} x_j \right) = \sum_{k=1}^n \sum_{j=1}^r a_{ik} b_{kj} x_j = \\
 &= \sum_{j=1}^r \sum_{k=1}^n a_{ik} b_{kj} x_j = \sum_{j=1}^r c_{ij} x_j \Rightarrow c_{ij} = \sum_{k=1}^n a_{ik} b_{kj}.
 \end{aligned}
 \tag{3.150}$$

Comparando esta última expresión con la Definición 2.2.3 del producto matricial, se tiene que los coeficientes escalares c_{ij} que asocian las coordenadas de z con las de x son el resultado de realizar el producto AB . En consecuencia, la matriz C , $(C)_{ij} = c_{ij}$, que asocia $[x]^{\mathcal{B}}$ con el vector de coordenadas respecto de \mathcal{W} de su imagen respecto de la composición de aplicaciones $g \circ f$ coincide con el producto matricial AB :

$$[z]^{\mathcal{W}} = AB[x]^{\mathcal{B}} = C[x]^{\mathcal{B}} \tag{3.151}$$

Proposición 3.5.3 – Matriz de una composición de aplicaciones lineales

Sea $A \in \mathbb{K}^{m \times n}$ la matriz asociada a la aplicación lineal $f : E \mapsto F$ respecto de las bases \mathcal{B} y \mathcal{V} , y sea $B \in \mathbb{K}^{n \times r}$ la matriz asociada a la aplicación lineal $g : F \mapsto G$ respecto de las bases \mathcal{V} y \mathcal{W} . La matriz:

$$C = AB$$

Es la matriz asociada a $g \circ f$ respecto de las bases \mathcal{B} y \mathcal{W} .

Ejemplo 3.5.3 Considérense las aplicaciones f, g del Ejemplo 3.5.1. La matriz de g respecto de las bases canónicas de \mathbb{R}^3 y \mathbb{R}^2 es: †

$$A = \begin{pmatrix} 0 & 2 & -1 \\ 1 & 0 & 0 \end{pmatrix} \tag{3.152}$$

La matriz de f respecto de las bases canónicas de \mathbb{R}^2 y \mathbb{R}^3 es: †

$$B = \begin{pmatrix} 1 & -1 \\ 0 & 1/2 \\ 1 & 0 \end{pmatrix} \tag{3.153}$$

La matriz de $g \circ f$ respecto de la base canónica de \mathbb{R}^2 es el producto matricial de la matriz asociada a g , ecuación (3.152) con la matriz asociada a f , ecuación (3.153):

$$C = AB = \begin{pmatrix} 0 & 2 & -1 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 0 & 1/2 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} -1 & 1/2 \\ 1 & -1 \end{pmatrix} \tag{3.154}$$

Ejercicio 3.5.1 † Calcular estas matrices asociadas.

Tras haber enunciado la equivalencia entre la composición de aplicaciones lineales y el producto de sus matrices asociadas, es fácil demostrar que la composición de aplicaciones lineales cumple las propiedades de la Proposición 3.5.1.

Demostración:

Sean $g : F \mapsto G$ y $f : E \mapsto F$ dos aplicaciones lineales que tienen entre sus matrices asociadas $A \in \mathbb{K}^{m \times n}$ y $B \in \mathbb{K}^{n \times r}$ respectivamente:

1. Recuérdese que el producto matricial no es, en general, conmutativo, por la propiedad 1 de la Proposición 2.2.3. Por analogía con la composición de aplicaciones lineales, se tiene que esta última tampoco lo es.
2. El producto matricial sí que cumple la propiedad asociativa como enuncia la propiedad 2 de la Proposición 2.2.3, por lo que:

$$(AB)C = A(BC) \Leftrightarrow (h \circ g) \circ f = h \circ (g \circ f) \quad (3.155)$$

3. Toda matriz permanece invariante cuando se multiplica por una matriz identidad \dagger . Por tanto:

$$f \circ I_E = f \Leftrightarrow AI = A \quad (3.156)$$

$$I_F \circ f = f \Leftrightarrow IA = A \quad (3.157)$$

4. Si g y f son inyectivas, sus matrices asociadas cumplen, por el Teorema 3.4.5 de rango-nulidad y por la Proposición 3.4.2, $m \geq n$, $n \geq r$, $r(A) = n$ y $r(B) = r$.

Por tanto, B es una matriz de rango máximo por columnas, por lo que cumple la propiedad 3 de la Proposición 3.4.4: $r(AB) = r(B) = r$. En consecuencia, $r(C) = r(AB) = r$. Además, $m \geq r$, por lo que $r(g \circ f) = r$ y $g \circ f$ es una aplicación lineal inyectiva.

5. Si g y f son suryectivas, sus matrices asociadas cumplen, por el Teorema 3.4.5 de rango-nulidad y por la Proposición 3.4.1, $m \leq n$, $n \leq r$, $r(A) = m$ y $r(B) = n$.

A es, por tanto, una matriz de rango máximo por filas, por lo que cumple la propiedad 3 de la Proposición 3.4.4, $r(AB) = r(A) = m$. En consecuencia, $r(C) = r(AB) = m$. Además, $m \leq r$, por lo que $r(g \circ f) = m$ y $g \circ f$ es una aplicación lineal suryectiva.

6. Una aplicación se dice biyectiva, Definición 3.1.5, si es tanto inyectiva como suryectiva. Como consecuencia directa de las propiedades 4 y 5, si g y f son biyectivas, estas son tanto inyectivas como suryectivas, y la aplicación $g \circ f$ será también inyectiva y suryectiva. Es decir, biyectiva.

7. La proposición “ $g \circ f$ inyectiva $\Rightarrow f$ inyectiva” equivale a su contrapositivo, que es la proposición:

$$f \text{ no es inyectiva} \Rightarrow g \circ f \text{ no es inyectiva} \quad (3.158)$$

Demostrando el contrapositivo, se demostrará la proposición original. Si f no es inyectiva, por el Teorema 3.4.5 del rango-nulidad y por la Proposición 3.4.2, $r(B) < r$.

Observación 3.5.1 \dagger Demostrar por aplicación directa de la definición de producto matricial, ecuación (2.2.3), que el producto de cualquier matriz por la matriz identidad es invariante, sea cual sea el orden del producto.

Por la propiedad 3 de la Proposición 3.4.4, $r(AB) \leq \min(r(A), r(B)) \Rightarrow r(C) = r(AB) < r$, por lo que $r(g \circ f) < r$.

Una condición necesaria para que una aplicación lineal sea inyectiva es, por el Teorema 3.4.5 y por la Proposición 3.4.2, que su rango coincida con la dimensión de su espacio inicial, por lo que $g \circ f$ no puede ser inyectiva, demostrándose la proposición.

8. De nuevo, se demostrará esta proposición haciéndolo, en su lugar, con su contrapositivo. Es decir, la proposición:

$$g \text{ no es suryectiva} \Rightarrow g \circ f \text{ no es suryectiva} \quad (3.159)$$

Si g no es suryectiva, por el Teorema 3.4.5 del rango-nulidad y por la Proposición 3.4.2, $r(A) < m$.

Por la propiedad 3 de la Proposición 3.4.4, $r(AB) \leq \min(r(A), r(B)) \Rightarrow r(C) = r(AB) < m$, por lo que $r(g \circ f) < m$.

Una condición necesaria para que una aplicación lineal sea suryectiva es, por el Teorema 3.4.5 del rango-nulidad y por la Proposición 3.4.1, que su rango coincida con la dimensión de su espacio final, por lo que $g \circ f$ no puede ser suryectiva, demostrándose la proposición.

9. Esta propiedad es una consecuencia inmediata de las dos propiedades anteriormente demostradas. Si $g \circ f$ es biyectiva, esta es tanto inyectiva como suryectiva, por lo que f será inyectiva, mientras que g será suryectiva.

□

3.5.5 Inversa de una aplicación

Definición 3.5.2 – Invertibilidad e inversa de una aplicación (I)

Una aplicación $f : E \mapsto F$ se dice que es invertible si existe una aplicación $f^{-1} : F \mapsto E$ tal que $f(x) = y \ \forall x \in E$, implica que $f^{-1}(y) = x \ \forall y \in F$.

La aplicación f^{-1} , de existir, es única y se denomina “inversa de f ”.

Teniendo en cuenta las Definiciones 3.1.4 y 3.5.1 de aplicación identidad y de composición de aplicaciones, es posible simplificar la definición anterior, dando lugar a la siguiente definición equivalente:

Definición 3.5.3 – Invertibilidad e inversa de una aplicación (II)

Una aplicación $f : E \mapsto F$ se dice que es invertible si existe una aplicación $f^{-1} : F \mapsto E$ tal que la composición de ambas aplicaciones produce la aplicación identidad. Es decir, $f^{-1} \circ f = I_E$ y $f \circ f^{-1} = I_F$.

La aplicación f^{-1} , de existir, es única y se denomina inversa de f .

Recuérdese que, por la Definición 3.1.1, una aplicación o función se define como una asociación en la que a cada elemento del conjunto inicial E se le asocia un único elemento del conjunto final F . Por lo tanto, para que exista un f^{-1} que cumpla con las Definiciones 3.1.1 y 3.5.3 de aplicación e inversa de una aplicación, todos y cada uno de los elementos del conjunto final de f han de tener al menos una preimagen en E . En consecuencia, la aplicación f debe ser suryectiva. Por otro lado, cada elemento del conjunto final F de f debe tener exactamente una preimagen en E , por lo que f debe ser inyectiva. Las aplicaciones que son tanto inyectivas como suryectivas se llaman “biyectivas”, punto 3 de la Definición 3.1.5. Se concluye así que toda aplicación biyectiva es invertible y viceversa:

Proposición 3.5.4 – Equivalencia entre biyectividad e invertibilidad

Una aplicación es invertible si y solo si es biyectiva, y además su inversa es también una aplicación biyectiva.

Por tanto, para comprobar si una aplicación lineal es invertible, basta comprobar que sea biyectiva.

A continuación, se indican dos propiedades que cumplen todas las aplicaciones invertibles.

Proposición 3.5.5 – Propiedades de las aplicaciones invertibles

1. $\forall f : E \mapsto F$ invertible:

$$(f^{-1})^{-1} = f \quad (3.160)$$

2. $\forall f : E \mapsto F \forall g : F \mapsto G$, f, g invertibles:

$$(g \circ f)^{-1} = f^{-1} \circ g^{-1} \quad (3.161)$$

Más adelante, tras haber estudiado la equivalencia entre las inversas de aplicaciones lineales entre espacios vectoriales, también llamadas “isomorfismos” por la Definición 3.1.7, y las inversas de sus matrices asociadas, se demostrará que las aplicaciones lineales invertibles cumplen las propiedades enunciadas por la Proposición 3.5.5.

3.5.6 Relación entre la inversa de un isomorfismo y la inversa de una matriz regular

Dado un isomorfismo entre espacios vectoriales $f : E \mapsto F$ y su inversa $f^{-1} : F \mapsto E$, fijese en una base \mathcal{B} de E y otra base \mathcal{V} de F y llámese A a la matriz de f respecto de esas bases y B a la matriz de f^{-1} respecto de esas mismas bases. Por la Definición 3.5.3 de aplicación inversa y por la relación que guardan el producto matricial y la composición de aplicaciones lineales, Proposición 3.5.3, se cumple:

$$f^{-1} \circ f = I_E \wedge f \circ f^{-1} = I_F \Leftrightarrow AB = BA = I \quad (3.162)$$

Así, se demuestra que B es la matriz inversa de $A \forall f : E \mapsto F$ invertible respecto de las bases $\mathcal{B} \subset E$ y $\mathcal{V} \subset F$. Por tanto, si A es la matriz asociada a un isomorfismo f respecto de las bases \mathcal{B} y \mathcal{V} , A^{-1} es la matriz asociada a la inversa de f respecto de esas mismas bases.

Proposición 3.5.6 – Matriz de la inversa de un isomorfismo

Sea $A \in \mathbb{K}^{m \times n}$ la matriz asociada a un isomorfismo $f : E \mapsto F$ respecto de las bases \mathcal{B} y \mathcal{V} , la matriz:

$$B = A^{-1}$$

es la matriz asociada a f^{-1} .

Ejemplo 3.5.4 La matriz asociada a la aplicación lineal:

$$f : \mathbb{R}^3 \mapsto \mathbb{R}^3, (x, y, z) \mapsto (y - z, x, x - y) \tag{3.163}$$

Respecto de la base canónica de \mathbb{R}^3 es:

$$A = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix} \tag{3.164}$$

$\det(A) = 1 \neq 0$, por lo tanto, A es regular por el Corolario 3.4.6 y f es un isomorfismo que admite inversa \dagger . La matriz B asociada a f^{-1} respecto de la base canónica de \mathbb{R}^3 es la inversa de la matriz A : \ddagger

$$B = A^{-1} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 0 & -1 \end{pmatrix} \tag{3.165}$$

Cumpléndose $AB = BA = I$.

Observación 3.5.2 \dagger Obviamente, esto supone que toda matriz asociada a un isomorfismo es regular.

Ejercicio 3.5.2 \ddagger Calcular la matriz A y su inversa.

Vista la equivalencia entre la inversa de un isomorfismo y la inversa de una de sus matrices asociadas, se demostrará que todos los isomorfismos cumplen las dos propiedades de la Proposición 3.5.5.

Demostración:

Sean g, f dos isomorfismos en un mismo espacio vectorial E , y sean A, B dos matrices cuadradas regulares del mismo orden asociadas a cada uno de ellos:

1. La matriz asociada a la aplicación $(f^{-1})^{-1}$ es la matriz $(A^{-1})^{-1}$. En virtud de la Proposición 2.5.3, esta última coincide con A , matriz cuya aplicación lineal asociada es f . Por tanto $(f^{-1})^{-1} = f$.
2. El producto matricial AB es una matriz asociada a la inversa de $g \circ f$. Por la Proposición 2.5.4, $(AB)^{-1} = B^{-1}A^{-1}$. El producto $B^{-1}A^{-1}$ está asociado a la composición de aplicaciones $f^{-1} \circ g^{-1}$, por lo que se concluye que $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$.

□

3.5.7 Resumen

1. La composición de aplicaciones lineales equivale al producto de las matrices asociadas a cada una de ellas en el orden inverso.
2. Dada una secuencia de aplicaciones lineales concatenables, en el sentido de que el espacio final de una coincide con el espacio inicial de otra, las matrices asociadas a su composición son productos matriciales de las matrices asociadas a cada una de esas aplicaciones.
3. Una aplicación es invertible si y sólo si es biyectiva, y su inversa es única.
4. Invertir un isomorfismo. Es decir, una aplicación lineal biyectiva, equivale a invertir una cualquiera de sus matrices asociadas.
5. Las matrices asociadas a un isomorfismo cualquiera y a su inversa son una la inversa de la otra.



Sistemas de ecuaciones lineales

4.1 Sistema de ecuaciones lineales

4.1.1 Introducción

Con todo lo visto sobre las matrices, los subespacios imagen y núcleo y sobre el concepto de rango, ya puede procederse a afrontar uno de los grandes objetivos del álgebra lineal: la resolución de sistemas de ecuaciones lineales.

A la hora de estudiar y resolver sistemas lineales resulta de ayuda tener en mente que el producto de una matriz por un vector columna puede reescribirse como una combinación lineal de las columnas de la matriz, con los elementos del vector columna como coeficientes, propiedad 1 de la Proposición 2.2.5. Para todo $\bar{x} = (x_1, x_2, \dots, x_n)^t$:

$$A\bar{x} = x_1C_1(A) + x_2C_2(A) + \dots + x_nC_n(A) \quad (4.1)$$

De esta representación del producto se deducen fácilmente las condiciones que ha de cumplir un sistema para que sea compatible, es decir, para que tenga solución; o incompatible, es decir, sin solución; y en caso de que el sistema sea compatible, para que tenga una única o infinitas soluciones.

Existen muchos métodos para resolver sistemas de ecuaciones lineales: algoritmos iterativos como el método de *Jacobi* y otros basados en la descomposición de una matriz en el producto de matrices más simples como las factorizaciones LU o QR. Unos algoritmos serán más apropiados o eficientes que otros en función de las propiedades del sistema a resolver y de las capacidades computacionales disponibles para resolverlo.

La descripción y el estudio de estos métodos y su implementación se adapta más a un curso de cálculo numérico y programación que a un curso introductorio sobre álgebra lineal, por lo que se deja fuera del alcance de este libro. En su lugar, se pondrá el foco en las propiedades generales de los sistemas de ecuaciones lineales y en su estudio y resolución mediante el método de reducción *gaussiana*.

4.1.2 Sistemas de ecuaciones lineales. Definición

Definición 4.1.1 — Resolución de un sistema de ecuaciones lineales

El problema de resolver un sistema de ecuaciones lineales consiste en encontrar, si existen, todas las *n-uplas* de valores x_1, x_2, \dots, x_n en el cuerpo \mathbb{K} que satisfagan simultáneamente las m ecuaciones:

4.1 Sistema de ecuaciones lineales	127
4.2 Estructura de las soluciones de un sistema lineal	131
4.3 El teorema de <i>Rouché-Frobenius</i>	135
4.4 Resolución de sistemas por reducción <i>gaussiana</i>	137

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

Donde a_{ij} y b_i son elementos de \mathbb{K} prefijados.

Las ecuaciones anteriores, todas ellas lineales, constituyen un sistema lineal de m ecuaciones y n incógnitas x_1, \dots, x_n .

Ejemplo 4.1.1 Estos son algunos sistemas de ecuaciones lineales:

$$\begin{cases} x_1 + 2x_3 = 3 \\ x_2 + x_3 = 2 \\ x_1 - 2x_2 + x_3 = 0 \end{cases} \quad (4.2)$$

$$\begin{cases} x_1 + 2x_3 = 3 \\ x_2 + x_3 = 2 \\ x_1 - x_2 + x_3 = 1 \end{cases} \quad (4.3)$$

$$\begin{cases} x_1 + 2x_3 = 3 \\ x_2 + x_3 = 2 \\ x_1 - x_2 + x_3 = 0 \end{cases} \quad (4.4)$$

Todos ellos tienen $m = 3$ ecuaciones y $n = 3$ incógnitas, denotadas x_1 , x_2 y x_3 , que pueden ser reales o complejas en función del problema que se defina.

Una definición alternativa surge cuando se reescribe de manera matricial el sistema de ecuaciones lineales de la Definición 4.1.1.

Definición 4.1.2 – Resolución de un sistema de ecuaciones lineales - Definición matricial

El problema de resolver un sistema de ecuaciones lineales consiste en encontrar, si existen, todos los vectores $\bar{x} = (x_1, x_2, \dots, x_n)^t \in \mathbb{K}^n$ que cumplen:

$$A\bar{x} = \bar{b} \quad (4.5)$$

Donde A es la matriz de coeficientes del sistema:

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \in \mathbb{K}^{m \times n} \quad (4.6)$$

\bar{x} es el vector de incógnitas del sistema:

$$\bar{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{K}^{n \times 1} \quad (4.7)$$

$\bar{\mathbf{b}}$ es el vector de coeficientes, o términos independientes, del sistema:

$$\bar{\mathbf{b}} = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix} \in \mathbb{K}^{m \times 1} \quad (4.8)$$

A los vectores $\bar{\mathbf{x}}_0 \in \mathbb{K}^n$ que satisfacen la expresión (4.5) se les denominan “vectores solución del sistema”.

Ejemplo 4.1.2 A continuación, se reescriben en formato matricial los tres sistemas del Ejemplo 4.1.1:

$$A_1 = \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \\ 1 & -2 & 1 \end{pmatrix}, \quad \bar{\mathbf{b}}_1 = \begin{pmatrix} 3 \\ 2 \\ 0 \end{pmatrix} \quad (4.9)$$

$$A_2 = \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \\ 1 & -1 & 1 \end{pmatrix}, \quad \bar{\mathbf{b}}_2 = \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix} \quad (4.10)$$

$$A_3 = \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \\ 1 & -1 & 1 \end{pmatrix}, \quad \bar{\mathbf{b}}_3 = \begin{pmatrix} 3 \\ 2 \\ 0 \end{pmatrix} \quad (4.11)$$

El vector de incógnitas se denota igual en los tres sistemas: $\bar{\mathbf{x}} = (x_1, x_2, x_3)^t$.

A menudo, resulta útil unir la matriz de coeficientes y el vector de términos independientes en una única matriz:

Definición 4.1.3 – Matriz ampliada de un sistema de ecuaciones lineales

Dado un sistema de ecuaciones lineales de la forma:

$$A\bar{\mathbf{x}} = \bar{\mathbf{b}} \quad (4.12)$$

con $A \in \mathbb{K}^{m \times n}$, $\bar{\mathbf{b}} \in \mathbb{K}^{1 \times m}$ y $\bar{\mathbf{x}} \in \mathbb{K}^{1 \times n}$, se denomina matriz ampliada del sistema a la yuxtaposición de la matriz de coeficientes y el vector de términos independientes:

$$(A|\bar{\mathbf{b}}) \in \mathbb{K}^{m \times (n+1)} \quad (4.13)$$

La matriz ampliada de un sistema de ecuaciones lineales contiene toda la información necesaria para definirlo y resolverlo.

Ejemplo 4.1.3 Las matrices ampliadas de los tres sistemas del Ejemplo 4.1.1 son:

$$(A_1|\bar{\mathbf{b}}_1) = \left(\begin{array}{ccc|c} 1 & 0 & 2 & 3 \\ 0 & 1 & 1 & 2 \\ 1 & -2 & 1 & 0 \end{array} \right) \quad (4.14)$$

$$(A_2|\bar{b}_2) = \left(\begin{array}{ccc|c} 1 & 0 & 2 & 3 \\ 0 & 1 & 1 & 2 \\ 1 & -1 & 1 & 1 \end{array} \right) \quad (4.15)$$

$$(A_3|\bar{b}_3) = \left(\begin{array}{ccc|c} 1 & 0 & 2 & 3 \\ 0 & 1 & 1 & 2 \\ 1 & -1 & 1 & 0 \end{array} \right) \quad (4.16)$$

4.1.3 Tipos de sistemas de ecuaciones lineales

Tras haber definido los sistemas de ecuaciones lineales, se propone clasificarlos en función de la existencia y el número de soluciones de estos. Dado un sistema lineal definido en forma matricial:

$$A\bar{x} = \bar{b}, \quad A \in \mathbb{K}^{m \times n}, \quad \bar{b} \in \mathbb{K}^{1 \times m}, \quad \bar{x} \in \mathbb{K}^{1 \times n} \quad (4.17)$$

Se tienen las siguientes definiciones:

Definición 4.1.4 — Sistemas lineales incompatibles, compatibles determinados y compatibles indeterminados

Un sistema de ecuaciones lineales se denomina:

1. Compatible: si admite al menos una solución.
 - a) Compatible determinado: si dicha solución es única.
 - b) Compatible indeterminado: si admite más de una solución.

2. Incompatible: si no admite ninguna solución.

4.2 Estructura de las soluciones de un sistema de ecuaciones lineales

4.2.1 Introducción

En el apartado anterior, se definieron los sistemas de ecuaciones lineales y se clasificaron en función de la existencia y del número de soluciones. En este nuevo apartado, se estudiará el modo en el que están matemáticamente relacionadas entre sí las soluciones de un sistema de ecuaciones lineales compatible.

4.2.2 Estructura de las soluciones

En primer lugar, se definen los denominados sistemas lineales homogéneos:

Definición 4.2.1 — Sistema de ecuaciones lineales homogéneo

Un sistema de ecuaciones lineales se dice homogéneo si sus términos independientes son todos nulos. Es decir, si el sistema es de la forma:

$$A\bar{x} = \bar{0}$$

Nótese que el conjunto de todas las soluciones del sistema homogéneo $A\bar{x} = \bar{0}$ coincide precisamente con el núcleo de la matriz de coeficientes del sistema, $\ker A$.

Definición 4.2.2 — Sistema homogéneo asociado a un determinado sistema lineal

A cada sistema de ecuaciones lineales de la forma $A\bar{x} = \bar{b}$, denominado sistema completo, se le asocia un sistema homogéneo con su misma matriz de coeficientes. Es decir:

$$A\bar{x} = \bar{0}$$

A este último sistema se le denomina sistema homogéneo asociado al sistema completo original $A\bar{x} = \bar{b}$.

Ejemplo 4.2.1 Dado el sistema:

$$\begin{cases} x_1 + 2x_3 = 3 \\ x_2 + x_3 = 2 \\ x_1 - x_2 + x_3 = 1 \end{cases} \quad (4.18)$$

su sistema homogéneo asociado es:

$$A\bar{x} = \bar{0}, \quad A = \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \\ 1 & -1 & 1 \end{pmatrix} \quad (4.19)$$

Las soluciones de este último sistema coinciden con $\ker A$. Una forma fácil de calcularlo es reducir por filas la matriz:

$$A \begin{array}{l} F'_3 = F_1 + F_3 \\ F''_3 = F_2 + F'_3 \end{array} \rightarrow \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix} \Rightarrow r(A) = 2 \Rightarrow \quad (4.20)$$

$$\Rightarrow \dim \ker A = 1, \quad \begin{cases} x_1 + 2x_3 = 0 \\ x_2 + x_3 = 0 \end{cases}$$

Estas dos ecuaciones, linealmente independientes, definen $\ker A$. Para obtener una base de este subespacio se puede parametrizar, por ejemplo, la variable x_2 y despejar x_1 en función del parámetro definido:

$$x_2 = \lambda \in \mathbb{R} \Rightarrow \ker A = L \left[\begin{pmatrix} 2 \\ 1 \\ -1 \end{pmatrix} \right] \quad (4.21)$$

Con las definiciones anteriores ya se puede proceder a estudiar la estructura del conjunto de soluciones de un sistema lineal cualquiera. Supóngase que \bar{x}_0 y \bar{x}_1 son vectores solución de un sistema de ecuaciones lineales representado matricialmente como $A\bar{x} = \bar{b}$. Restando entre sí ambos vectores y pre-multiplicando por la matriz de coeficientes:

$$A(\bar{x}_0 - \bar{x}_1) = A\bar{x}_1 - A\bar{x}_2 = \bar{b} - \bar{b} = \bar{0} \quad (4.22)$$

Luego $\bar{x}_0 - \bar{x}_1$ es una solución del sistema homogéneo asociado al sistema $A\bar{x} = \bar{b}$.

Por otra parte, si \bar{x} es una solución del sistema completo original y \bar{u} es una solución de su sistema homogéneo asociado. Es decir, si $\bar{u} \in \ker A$, entonces:

$$A(\bar{u} + \bar{x}) = A\bar{u} + A\bar{x} = \bar{0} + \bar{b} = \bar{b} \quad (4.23)$$

Así, se demuestra que el vector suma $\bar{u} + \bar{x}$ es también una solución del sistema completo original.

En resumen, se tiene que $\bar{x}_0 - \bar{x}_1 \in \ker A$ y toda solución \bar{x} del sistema $A\bar{x} = \bar{b}$ puede escribirse como la suma de una solución cualquiera \bar{x}_0 del sistema completo denominada "solución particular" más algún vector \bar{u} perteneciente a $\ker A$.

Sea $(\bar{u}_1, \dots, \bar{u}_{n-r})$ una base cualquiera de $\ker A$, donde r denota el rango de A , cualquier solución del sistema $A\bar{x} = \bar{b}$ se puede escribir de la forma:

$$\bar{x} = \bar{x}_0 + \lambda_1 \bar{u}_1 + \dots + \lambda_{n-r} \bar{u}_{n-r} \quad (4.24)$$

Donde \bar{x}_0 es una solución particular cualquiera del sistema y $\lambda_1, \dots, \lambda_{n-r}$ son parámetros escalares en el cuerpo \mathbb{K} . Por cada solución $\bar{x} \in \mathbb{K}^n$ del sistema, los parámetros $\lambda_1, \dots, \lambda_{n-r}$ adoptan unos valores únicos que depen-

derán de la base escogida para el núcleo de A y de la solución particular considerada.

Se concluye que para poder definir el conjunto de todas las soluciones de un sistema lineal cualquiera basta con encontrar una solución particular \bar{x}_0 del sistema y una base del núcleo de la matriz de coeficientes del sistema (Figura 4.1).

Proposición 4.2.1 – Conjunto de soluciones de un sistema de ecuaciones lineales

El conjunto que contiene todas las soluciones del sistema de ecuaciones lineales expresado matricialmente como $A\bar{x} = \bar{b}$ es, dada una solución particular \bar{x}_0 del sistema y una base $(\bar{u}_1, \dots, \bar{u}_{n-r})$ de $\ker A$:

$$\{\bar{x} \in \mathbb{K}^n : \bar{x} = \bar{x}_0 + \lambda_1 \bar{u}_1 + \dots + \lambda_{n-r} \bar{u}_{n-r}, \lambda_1, \dots, \lambda_{n-r} \in \mathbb{K}\}$$

Donde $r = r(A)$.

Nótese que si el rango de la matriz de coeficientes A coincide con su número de columnas, por el Teorema 3.4.5 del rango-nulidad, $\ker A = \{\bar{0}\}$ y la solución particular \bar{x}_0 es la única solución que tiene el sistema.

Ejemplo 4.2.2 Para obtener una solución particular \bar{x}_0 del sistema:

$$\begin{cases} x_1 + 2x_3 = 3 \\ x_2 + x_3 = 2 \\ x_1 - x_2 + x_3 = 1 \end{cases} \quad (4.25)$$

se sustituye un número de variables igual a la dimensión de $\ker A$ por unos valores arbitrarios y se despeja el resto. Por ejemplo:

$$x_2 = 1 \Rightarrow x_3 = 2 - 1 = 1 \Rightarrow x_1 = 3 - 2 \cdot 1 = 1 \Rightarrow \bar{x}_0 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \quad (4.26)$$

Por otro lado, el núcleo de A , calculado en el Ejemplo 4.2.1, es:

$$\ker A = L \left[\begin{pmatrix} 2 \\ 1 \\ -1 \end{pmatrix} \right] \quad (4.27)$$

Así, el conjunto de las soluciones del sistema puede escribirse como:

$$\left\{ (x_1, x_2, x_3) \in \mathbb{R}^3 : \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \lambda \begin{pmatrix} 2 \\ 1 \\ -1 \end{pmatrix}, \lambda \in \mathbb{R} \right\} \quad (4.28)$$

Este ejemplo se esquematiza en la Figura 4.1.

El conjunto de todas las soluciones de un sistema lineal forma lo que en álgebra se conoce como una “variedad lineal”: es una traslación de todos los elementos de un subespacio vectorial, en este caso del núcleo de su matriz de coeficientes, de modo que ya no contenga el vector cero. En \mathbb{R}^2 y \mathbb{R}^3 son variedades lineales todas aquellas rectas y planos que no cortan el origen de coordenadas.

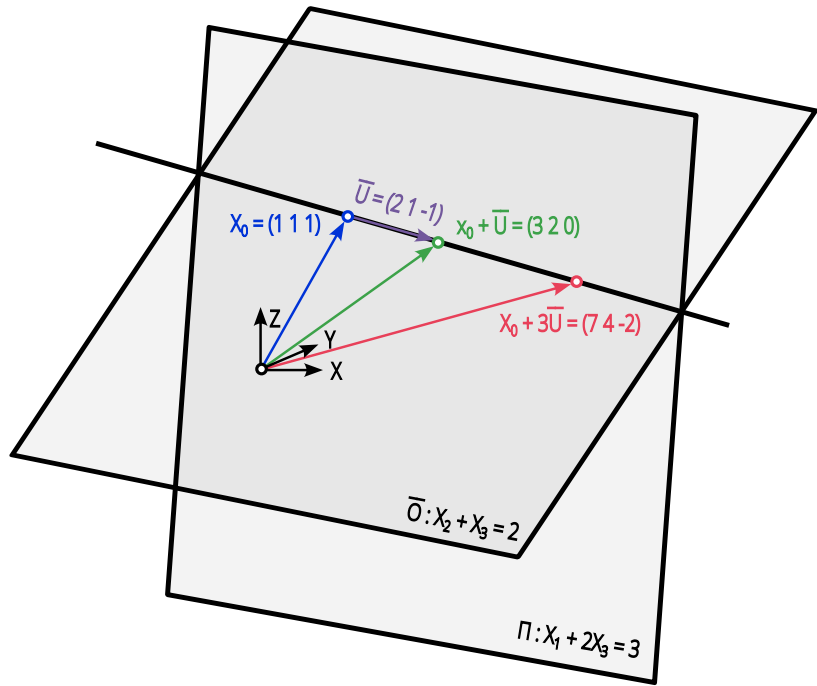


Figura 4.1: Todos los puntos de \mathbb{R}^3 que son solución del sistema del Ejemplo 4.2.2 se encuentran contenidos en una recta que es la intersección de los planos $x_1 + 2x_3 = 3$ y $x_2 + x_3 = 2$. Para definirla, basta con conocer un punto cualquiera de esta, que será una solución particular \bar{x}_0 del sistema, y un vector director que determine su dirección (que cumpla las ecuaciones $x_1 + 2x_3 = 0$ y $x_2 + x_3 = 0$, formando una base del núcleo de la matriz de coeficientes del sistema).

4.3 El teorema de Rouché-Frobenius

4.3.1 Introducción

Tras haber aprendido a expresar el conjunto de soluciones de un sistema de ecuaciones compatible, en este breve apartado se estudiarán las condiciones necesarias y suficientes que ha de cumplir todo sistema de ecuaciones lineales para que sea compatible. El llamado “teorema de Rouché-Frobenius” † enuncia esta condición y, además, en el caso de que el sistema sea compatible, las condiciones que determinan si la solución es única o múltiple.

4.3.2 Teorema de Rouché-Frobenius

Teorema 4.3.1 – Teorema de Rouché-Frobenius

Dado el sistema de ecuaciones lineales $A\bar{x} = \bar{b}$ con la matriz de coeficientes $A \in \mathbb{K}^{m \times n}$, $r(A) = r$, y el vector de incógnitas $\bar{x} \in \mathbb{K}^{1 \times n}$, caben las siguientes posibilidades:

1. Si $r(A|\bar{b}) > r$, el sistema es incompatible y no tiene solución.
2. Si $r(A|\bar{b}) = r = n$, el sistema es compatible determinado y su solución es única.
3. Si $r(A|\bar{b}) = r < n$, el sistema es compatible indeterminado y el conjunto de sus soluciones depende de $n - r$ parámetros.

Demostración:

Todo sistema de ecuaciones lineales de la forma $A\bar{x} = \bar{b}$ puede reescribirse, por la propiedad 1 de la Proposición 2.2.5, como:

$$x_1 C_1(A) + x_2 C_2(A) + \dots + x_n C_n(A) = \bar{b} \quad (4.29)$$

En consecuencia, para que el sistema $A\bar{x} = \bar{b}$ admita solución, el vector de términos independientes \bar{b} debe ser combinación lineal de las n columnas de la matriz de coeficientes A . Los coeficientes escalares que hacen que se cumpla la identidad (4.29) constituyen el vector o vectores solución del sistema.

Si el vector \bar{b} fuese combinación lineal de las columnas de A , este pertenecería a la imagen de A , en cuyo caso el subespacio imagen de A coincidiría con el de $(A|\bar{b})$ y ambas matrices tendrían el mismo rango. Por tanto, si el sistema es compatible, los rangos de A y $(A|\bar{b})$ coinciden.

El converso de esta implicación también se cumple: si el rango de $(A|\bar{b})$ coincidiese con el de A , \bar{b} pertenecería al subespacio columna de A y existirían coeficientes escalares que permitan su expresión como combinación lineal de las columnas de A . Se tiene así una condición necesaria y suficiente para la existencia de solución.

Por último, suponiendo que $A\bar{x} = \bar{b}$ es compatible. Es decir, que el rango de $(A|\bar{b})$ coincide con el de A . Si $r(A|\bar{b}) = r(A) = n$, por el Teorema

Observación 4.3.1 † El nombre de este teorema es fuente de controversia. En países de habla hispana se le suele conocer como de “Rouché-Frobenius”; pero en los países anglosajones, así como en Italia, se prefiere el nombre de “Rouché-Capelli”. Otros países, como Austria y Rusia, prefieren asignar la autoría del teorema al matemático alemán *Leopold Kroneker*. En otros lugares es más conocido como el “teorema de Frobenius” en honor de *Ferdinand Georg Frobenius*, también alemán. Parece ser que fue descubierto de manera independiente por varios matemáticos en torno a la misma época.

3.4.5 del rango-nulidad, $\dim \ker A = 0$, por lo que $\ker A = \{\bar{0}\}$ y, por la Proposición 4.2.1, la solución del sistema sería única.

Y si $r(A|\bar{b}) = r(A) < n$, $\dim \ker A > 0$, por lo que las soluciones del sistema (4.2.1) dependerían de uno o varios parámetros en \mathbb{K} y, por la Proposición 4.2.1, existirían infinitas soluciones.

□

4.4 Resolución de sistemas por reducción gaussiana

4.4.1 Introducción

Existen muchos métodos para resolver sistemas de ecuaciones lineales. Por ejemplo, si la matriz de coeficientes A es cuadrada e invertible, el (único) vector solución del sistema podría despejarse invirtiendo A :

$$A\bar{x} = \bar{b} \Rightarrow A^{-1}A\bar{x} = Ix = A^{-1}\bar{b} \Rightarrow \bar{x} = A^{-1}\bar{b} \quad (4.30)$$

Otro método elemental que puede aplicarse si la matriz de coeficientes A es cuadrada e invertible es la llamada “regla de *Cramer*”, que expresa las componentes de la solución del sistema en función de ciertos determinantes.

Estos dos métodos tienen los siguientes inconvenientes: en primer lugar, sólo se pueden aplicar si la matriz de coeficientes es cuadrada y regular; y en segundo lugar, son métodos muy ineficientes a causa del gran número de operaciones necesarias para su implementación †, especialmente la regla de *Cramer*, que aplicada a un sistema de n ecuaciones e incógnitas requiere de la realización de $n + 1$ determinantes que contienen $n!$ términos, Observación 2.3.2, con n factores cada uno, por lo que su interés es puramente teórico.

En este apartado, se describe un método basado en la realización de operaciones elementales por filas sobre la matriz ampliada del sistema. Es más eficiente que los dos métodos arriba mencionados, sobre todo en sistemas con un elevado número de incógnitas, y además no requiere que la matriz de coeficientes A sea cuadrada ni invertible. Se trata del método de resolución por reducción *gaussiana*.

Observación 4.4.1 † Afortunadamente, en las matrices que aparecen en modelos de física e ingeniería predominan los elementos nulos, lo que ayuda a disminuir drásticamente el tiempo de resolución de los sistemas de ecuaciones lineales y de otros problemas numéricos como la diagonalización.

A las matrices que tienen casi todos sus elementos, o la mayoría de estos, iguales a cero se les conoce como “matrices dispersas”.

4.4.2 Resolución de sistemas lineales por reducción gaussiana

La base de este método de resolución de sistemas se encuentra en la siguiente propiedad:

Proposición 4.4.1 — Sistemas de ecuaciones lineales equivalentes

Dado un sistema de ecuaciones lineales $A\bar{x} = \bar{b}$ con la matriz de coeficientes $A \in \mathbb{K}^{m \times n}$ y el vector de incógnitas $\bar{x} \in \mathbb{K}^{1 \times n}$, el conjunto de todas sus soluciones coincide con el del sistema siguiente:

$$QAx = Qb$$

Para toda matriz cuadrada $Q \in \mathbb{K}^{m \times m}$ invertible, por lo que puede decirse que el sistema $QAx = Qb$ es equivalente al sistema original $A\bar{x} = \bar{b}$.

Demostración:

Despejando $A\bar{x}$ en la identidad (4.4.1) y teniendo en cuenta que Q es regular,

$$QAx = Qb \Rightarrow A\bar{x} = Q^{-1}Q\bar{b} = I\bar{b} = \bar{b} \quad (4.31)$$

□

La Proposición 4.4.1 puede aplicarse para transformar el sistema $A\bar{x} = \bar{b}$ en un sistema equivalente cuyas soluciones se puedan extraer de modo trivial. Recuérdese que toda matriz puede reducirse mediante operaciones elementales por filas a una forma triangular o trapezoidal.

$$E_k \dots E_2 E_1 A = QA \quad (4.32)$$

Aplicando esta reducción a la matriz ampliada del sistema se obtiene la matriz ampliada de un sistema equivalente cuyas soluciones son triviales de obtener:

$$E_k \dots E_2 E_1 (A|\bar{b}) = Q(A|\bar{b}) = (QA|Q\bar{b}) \quad (4.33)$$

En ello consiste el método de reducción *gaussiana*, que también permitiría, al mismo tiempo que se resuelve el sistema, estudiar la compatibilidad y el número de soluciones del mismo, pues como se vio en el apartado 3.4.8, el rango de una matriz triangular o trapezoidal es simplemente el número de filas o columnas no nulas.

Suponiendo, a modo de ejemplo, un sistema de ecuaciones lineales cuya matriz ampliada es de tamaño 6×6 . Si la matriz ampliada reducida tiene la forma:

$$\left(\begin{array}{cccccc|c} \blacksquare & * & * & * & * & * & * \\ 0 & \blacksquare & * & * & * & * & * \\ 0 & 0 & \blacksquare & * & * & * & * \\ 0 & 0 & 0 & \blacksquare & * & * & * \\ 0 & 0 & 0 & 0 & \blacksquare & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & \blacksquare \end{array} \right) \quad (4.34)$$

El sistema es incompatible, pues $r(A) = 5$ y $r(A|\bar{b}) = 6$.

Ejemplo 4.4.1 Reduciendo la matriz ampliada del sistema:

$$\begin{cases} x_1 + 2x_3 = 3 \\ x_2 + x_3 = 2 \\ x_1 - x_2 + x_3 = 0 \end{cases} \quad (4.35)$$

Resulta:

$$(A|\bar{b}) = \left(\begin{array}{ccc|c} 1 & 0 & 2 & 3 \\ 0 & 1 & 1 & 2 \\ 1 & -1 & 1 & 0 \end{array} \right) \begin{array}{l} F'_3 = F_3 - F_1 \\ F''_3 = F'_3 + F_2 \end{array} \quad (4.36)$$

$$\begin{array}{l} F'_3 = F_3 - F_1 \\ F''_3 = F'_3 + F_2 \end{array} \rightarrow \left(\begin{array}{ccc|c} 1 & 0 & 2 & 3 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & 0 & -1 \end{array} \right)$$

$r(A|\bar{b}) = 3$ y $r(A) = 2$. Por lo que el sistema es incompatible.

Si la matriz ampliada reducida tiene la forma:

$$\left(\begin{array}{cccccc|c} \blacksquare & * & * & * & * & * & * \\ 0 & \blacksquare & * & * & * & * & * \\ 0 & 0 & \blacksquare & * & * & * & * \\ 0 & 0 & 0 & \blacksquare & * & * & * \\ 0 & 0 & 0 & 0 & \blacksquare & * & * \\ 0 & 0 & 0 & 0 & 0 & \blacksquare & * \end{array} \right) \quad (4.37)$$

El sistema es compatible determinado, pues $r(A) = 6$ y $r(A|\bar{b}) = 6$.

Ejemplo 4.4.2 Reduciendo la matriz ampliada del sistema:

$$\begin{cases} x_1 + 2x_3 = 3 \\ x_2 + x_3 = 2 \\ -x_1 + 2x_2 - x_3 = 0 \end{cases} \quad (4.38)$$

Se obtiene:

$$(A|\bar{b}) = \left(\begin{array}{ccc|c} 1 & 0 & 2 & 3 \\ 0 & 1 & 1 & 2 \\ -1 & 2 & -1 & 0 \end{array} \right) \begin{array}{l} F'_3 = F_3 + F_1 \\ F''_3 = F'_3 - 2F_2 \end{array} \quad (4.39)$$

$$\begin{array}{l} F'_3 = F_3 + F_1 \\ F''_3 = F'_3 - 2F_2 \end{array} \rightarrow \left(\begin{array}{ccc|c} 1 & 0 & 2 & 3 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & -1 & -1 \end{array} \right)$$

$r(A|\bar{b}) = 3$ y $r(A) = 3$, por lo que el sistema es compatible determinado.

El sistema asociado a la matriz ampliada reducida es equivalente al sistema original, por lo que sus soluciones coinciden. Estas pueden extraerse fácilmente de las ecuaciones del sistema equivalente:

$$\begin{cases} x_1 + 2x_3 = 3 \\ x_2 + x_3 = 2 \\ -x_3 = -1 \Rightarrow x_3 = 1 \Rightarrow x_2 = 2 - x_3 = 2 - 1 = 1 \Rightarrow \\ \Rightarrow x_1 = 3 - 2x_3 = 3 - 2 \cdot 1 = 1. \end{cases} \quad (4.40)$$

Por lo tanto, la solución única del sistema es el vector $\bar{x}_0 = (1, 1, 1)^t$.

Si la matriz ampliada reducida tiene la forma:

$$\left(\begin{array}{cccccc|c} \blacksquare & * & * & * & * & * & * \\ 0 & \blacksquare & * & * & * & * & * \\ 0 & 0 & \blacksquare & * & * & * & * \\ 0 & 0 & 0 & \blacksquare & * & * & * \\ 0 & 0 & 0 & 0 & \blacksquare & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right) \quad (4.41)$$

El sistema es compatible indeterminado, pues $r(A) = 5$ y $r(A|\bar{b}) = 5$.

Ejemplo 4.4.3 Reduciendo la matriz ampliada del sistema:

$$\begin{cases} x_1 + 2x_3 = 3 \\ x_2 + x_3 = 2 \\ x_1 - x_2 + x_3 = 1 \end{cases} \quad (4.42)$$

Se obtiene:

$$(A|\bar{b}) = \left(\begin{array}{ccc|c} 1 & 0 & 2 & 3 \\ 0 & 1 & 1 & 2 \\ 1 & -1 & 1 & 1 \end{array} \right) \xrightarrow{\substack{F'_3 = F_3 - F_1 \\ F''_3 = F'_3 + F_2}} \quad (4.43)$$

$$\xrightarrow{\substack{F'_3 = F_3 - F_1 \\ F''_3 = F'_3 + F_2}} \left(\begin{array}{ccc|c} 1 & 0 & 2 & 3 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{array} \right)$$

$r(A|\bar{b}) = 2$ y $r(A) = 2$. Por lo que el sistema es compatible indeterminado. Este sistema ya se resolvió en el Ejemplo 4.2.2 de la sección 4.2.

4.4.3 Resumen de los apartados anteriores

1. Cualquier solución de un sistema de ecuaciones lineales puede descomponerse en la suma de una solución particular y un vector del núcleo de la matriz del coeficientes.
2. El conjunto de todas las soluciones de un sistema de ecuaciones lineales es una variedad lineal resultante de trasladar el núcleo de la matriz de coeficientes sumando a todos sus vectores una solución cualquiera del sistema.
3. Un sistema de ecuaciones lineales es compatible si el vector de coeficientes pertenece al subespacio columna de la matriz de coeficientes. Dicho de otro modo, si el rango de la matriz ampliada del sistema coincide con el de la matriz de coeficientes.
4. Un modo de resolver un sistema compatible de cualquier número de ecuaciones e incógnitas consiste en reducir su matriz ampliada a forma trapezoidal o triangular mediante operaciones elementales por filas, obteniendo un sistema equivalente cuyas soluciones son triviales de extraer. Este es el método de reducción *gaussiana*.



Producto escalar, norma, ortogonalidad y espacios euclidianos

5

5.1 Producto escalar y de la norma en \mathbb{R}^n y \mathbb{C}^n . Espacios euclidianos

5.1.1 Introducción

El objetivo primordial de este Tema es extender a los espacios vectoriales generales, sección 1.1, los conceptos de longitud y ángulo propios de los vectores del plano real euclidiano \mathbb{R}^2 y del espacio real euclidiano \mathbb{R}^3 . Para ello, se definirán axiomáticamente dos operadores denominados producto interno y norma, que tienen las mismas propiedades del producto escalar y la longitud de un vector, respectivamente.

Así, se podrán resolver problemas matemáticos como, por ejemplo, aproximar funciones mediante polinomios o series trigonométricas por analogía con la resolución de problemas geométricos en el espacio euclidiano como la proyección ortogonal.

Gran parte del foco se pondrá precisamente en el concepto de ortogonalidad, o perpendicularidad, entre los elementos de un espacio vectorial dotado de producto interno. Dos aplicaciones muy importantes son las proyecciones ortogonales y la resolución del problema de mínimos cuadrados.

En esta primera sección, se recuerdan las definiciones y propiedades elementales del producto escalar y del módulo, también conocido como “norma euclidiana”, de un vector en \mathbb{R}^2 o \mathbb{R}^3 . A continuación, se definirá un producto escalar y una norma euclidiana que se puedan aplicar también a vectores complejos, y se estudiarán propiedades que cumplen estos nuevos operadores como el teorema de *Cauchy-Schwartz*. Finalmente, se presentarán definiciones axiomáticas del producto interno y de la norma para los espacios vectoriales generales.

5.1.2 La norma de un vector en física clásica

Definición 5.1.1 – Norma de un vector en física clásica

En física clásica, también llamada física newtoniana o no relativista, la norma euclidiana o módulo de un vector real en \mathbb{R}^2 o \mathbb{R}^3 se define como la longitud de dicho vector, considerando sus coordenadas cartesianas como longitudes de los catetos de un triángulo rectángulo. Generalizando para un número n cualquiera de componentes, se tiene esta expresión:

$$\|\bar{x}\| = \sqrt{\sum_{i=1}^n x_i^2}$$

Donde x_i son las componentes del vector.

5.1	Producto interno, norma y espacios euclidianos	141
5.2	Ortogonalidad	153
5.3	Matrices unitarias, ortogonales y de giro .	166
5.4	Ortogonalización de <i>Gram-Schmidt</i>	174
5.5	Teorema de la proyección ortogonal. Matrices de proyección ortogonal	179
5.6	El problema de mínimos cuadrados	187

Observación 5.1.1 A partir de este Tema se considerará que los vectores de \mathbb{R}^n y \mathbb{C}^n son, por defecto, vectores columna, Definición 2.1.5.

Ejemplo 5.1.1 Los vectores reales:

$$\bar{x} = (1, 1)^t \quad (5.1)$$

$$\bar{y} = (1, \sqrt{3})^t \quad (5.2)$$

$\bar{x}, \bar{y} \in \mathbb{R}^2$, tienen por normas *euclidianas*:

$$\|\bar{x}\| = \sqrt{1^2 + 1^2} = \sqrt{2} \quad (5.3)$$

$$\|\bar{y}\| = \sqrt{1^2 + (\sqrt{3})^2} = \sqrt{4} = 2 \quad (5.4)$$

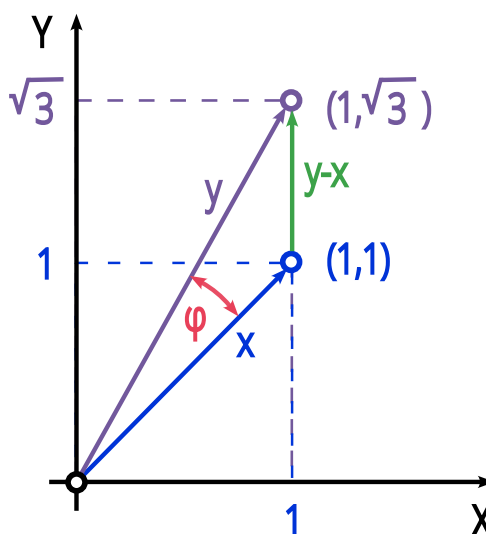


Figura 5.1: Representación gráfica en el plano *euclidiano* de los vectores del Ejemplo 5.1.1. El ángulo φ relativo entre estos se calculará más adelante, en el Ejemplo 5.1.6. La longitud de su diferencia $\bar{y} - \bar{x}$ determina su distancia *euclidiana* relativa, definida por la ecuación (5.1.5).

5.1.3 El producto escalar en física clásica

Definición 5.1.2 – Producto escalar en física clásica

En física clásica, el producto escalar de dos vectores reales en \mathbb{R}^2 o \mathbb{R}^3 suele definirse mediante la siguiente expresión:

$$\langle \bar{x}, \bar{y} \rangle = \|\bar{x}\| \|\bar{y}\| \cdot \cos \varphi$$

Donde φ es el ángulo entre los dos vectores.

El resultado de aplicar la fórmula (5.1.2) se interpreta geoméricamente como la longitud de la proyección de un vector sobre el otro. Es decir, la longitud de la sombra que arroja un vector sobre el que lo multiplica.

Ejemplo 5.1.2 El producto escalar de los vectores reales $\bar{x}, \bar{y} \in \mathbb{R}^2$:

$$\bar{x} = (1, 1)^t \quad (5.5)$$

$$\bar{y} = (1, \sqrt{3})^t \quad (5.6)$$

Se calcula aplicando la definición (5.1.2):

$$\langle \bar{x}, \bar{y} \rangle = \|\bar{x}\| \|\bar{y}\| \cdot \cos \varphi \stackrel{(*)}{=} 2\sqrt{2} \cdot \cos 15^\circ \stackrel{(**)}{=} 1 + \sqrt{3} \quad (5.7)$$

En (*) se calcula así el ángulo φ entre los dos vectores:

$$\varphi = \arctg \frac{\sqrt{3}}{1} - \arctg \frac{1}{1} = \frac{\pi}{12} \equiv 15^\circ \quad (5.8)$$

En (**) se aplica la fórmula del coseno del ángulo mitad para calcular $\cos 15^\circ$:

$$\cos 15^\circ = \sqrt{\frac{1 + \cos 30^\circ}{2}} = \sqrt{\frac{1 + \sqrt{3}/2}{2}} \quad (5.9)$$

5.1.4 Otra definición para el producto escalar en \mathbb{R}^n

A continuación, se presentará una definición alternativa de producto escalar aplicable a vectores reales de cualquier número de componentes.

Definición 5.1.3 – Producto escalar de vectores en \mathbb{R}^n

Dados dos vectores \bar{x}, \bar{y} de \mathbb{R}^n , su producto escalar, o producto interior, se define como el escalar real:

$$\langle \bar{x}, \bar{y} \rangle = \bar{x}^t \bar{y} = \sum_{i=1}^n x_i y_i = \bar{y}^t \bar{x}$$

Ejemplo 5.1.3 El producto escalar de los vectores reales $\bar{x}, \bar{y} \in \mathbb{R}^2$:

$$\bar{x} = (1, 1)^t \quad (5.10)$$

$$\bar{y} = (1, \sqrt{3})^t \quad (5.11)$$

Es, aplicando la fórmula (5.1.3):

$$\langle \bar{x}, \bar{y} \rangle = \bar{x}^t \bar{y} = (1 \ 1) \begin{pmatrix} 1 \\ \sqrt{3} \end{pmatrix} = 1 + \sqrt{3} \quad (5.12)$$

Este valor coincide con el obtenido en el Ejemplo 5.1.2 aplicando la fórmula (5.1.2).

El producto escalar real, Definiciones 5.1.2 y 5.1.3, tiene las siguientes propiedades:

Proposición 5.1.1 – Propiedades del producto escalar en \mathbb{R}^n

1. Distributividad: $\forall \bar{x}, \bar{y}, \bar{z} \in \mathbb{R}^n, \forall \lambda, \mu \in \mathbb{R}$

$$\begin{aligned} \langle \lambda \bar{x} + \mu \bar{y}, \bar{z} \rangle &= \lambda \langle \bar{x}, \bar{z} \rangle + \mu \langle \bar{y}, \bar{z} \rangle \\ \langle \bar{x}, \lambda \bar{y} + \mu \bar{z} \rangle &= \lambda \langle \bar{x}, \bar{y} \rangle + \mu \langle \bar{x}, \bar{z} \rangle \end{aligned} \quad (5.13)$$

Ejercicio 5.1.1 † Demostrar estas tres propiedades del producto escalar real a partir de la Definición 5.1.3.

2. Simetría: $\forall \bar{x}, \bar{y} \in \mathbb{R}^n$

$$\langle \bar{x}, \bar{y} \rangle = \langle \bar{y}, \bar{x} \rangle \quad (5.14)$$

3. Positividad: $\forall \bar{x} \in \mathbb{R}^n, \bar{x} \neq \bar{0}$,

$$\langle \bar{x}, \bar{x} \rangle = \bar{x}^t \bar{x} = \sum_{i=1}^n x_i^2 > 0 \quad (5.15)$$

$$\langle \bar{x}, \bar{x} \rangle = 0 \Leftrightarrow \bar{x} = \bar{0}$$

Se deja al lector † la demostración de estas propiedades.

Como consecuencia de la propiedad 1, el producto escalar es lineal respecto de ambas componentes.

5.1.5 Definición de la norma euclidiana en \mathbb{R}^n

La Definición 5.1.3 del producto escalar real permite redefinir la norma euclidiana de un vector en \mathbb{R}^n , ecuación (5.1.4), del modo siguiente:

Definición 5.1.4 — Norma euclidiana de un vector en \mathbb{R}^n

Para cada vector \bar{x} de \mathbb{R}^n , el número:

$$\|\bar{x}\| = \sqrt{\sum_{i=1}^n x_i^2} = \sqrt{\langle \bar{x}, \bar{x} \rangle}$$

recibe el nombre de norma euclidiana, o euclídea, del vector.

La norma euclidiana puede también emplearse para cuantificar la distancia entre dos puntos en \mathbb{R}^n , entendida como la longitud del segmento que los une:

Definición 5.1.5 — Distancia euclidiana entre vectores en \mathbb{R}^n

La distancia euclidiana entre dos vectores $\bar{x}, \bar{y} \in \mathbb{R}^n$ se define como:

$$d(\bar{x}, \bar{y}) = \|\bar{x} - \bar{y}\|$$

Ejemplo 5.1.4 Aplicando la Definición 5.1.4 de norma euclidiana de un vector real sobre los vectores $\bar{x}, \bar{y} \in \mathbb{R}^2$:

$$\bar{x} = (1, 1)^t \quad (5.16)$$

$$\bar{y} = (1, \sqrt{3})^t \quad (5.17)$$

Se obtiene:

$$\|\bar{x}\| = \sqrt{\langle \bar{x}, \bar{x} \rangle} = \sqrt{1 + 1} = \sqrt{2} \quad (5.18)$$

$$\|\bar{y}\| = \sqrt{\langle \bar{y}, \bar{y} \rangle} = \sqrt{1 + 3} = \sqrt{4} = 2 \quad (5.19)$$

La distancia *euclidiana* entre \bar{x} e \bar{y} es, según la Definición 5.1.5:

$$d(\bar{x}, \bar{y}) = \|\bar{x} - \bar{y}\| = \|\bar{y} - \bar{x}\| = \sqrt{0^2 + (\sqrt{3} - 1)^2} = \sqrt{3} - 1 \quad (5.20)$$

Puede comprobarse † que la norma *euclidiana* de un vector real, Definición 5.1.4, cumple las dos propiedades elementales siguientes:

Proposición 5.1.2 — Propiedades de la norma *euclidiana*

1. Positividad: $\forall \bar{x} \in \mathbb{R}^n$

$$\|\bar{x}\| \geq 0, \quad \|\bar{x}\| = 0 \Leftrightarrow \bar{x} = \bar{0} \quad (5.21)$$

2. Homogeneidad: $\forall \lambda \in \mathbb{R}, \forall \bar{x} \in \mathbb{R}^n$

$$\|\lambda \bar{x}\| = |\lambda| \|\bar{x}\| \quad (5.22)$$

Ejercicio 5.1.2 † Demostrar que la norma *euclidiana* real, Definición 5.1.4, cumple estas dos propiedades.

5.1.6 Equivalencia entre definiciones del producto escalar real

En este apartado se demostrará que las dos formulaciones propuestas para el producto escalar real, Definiciones 5.1.2 y 5.1.3, son equivalentes.

El vector diferencia entre \bar{x} e $\bar{y} \in \mathbb{R}^n$, denotado \bar{z} , se calcula como $\bar{z} = \bar{x} - \bar{y}$. Los tres vectores del conjunto $\{\bar{x}, \bar{y}, \bar{z}\}$ constituyen los lados de un triángulo.

Aplicando las propiedades del producto escalar, Proposición 5.1.1, y la Definición 5.1.4 de norma *euclidiana* se obtiene:

$$\begin{aligned} \langle \bar{z}, \bar{z} \rangle &= \|\bar{z}\|^2 = \langle \bar{x} - \bar{y}, \bar{x} - \bar{y} \rangle = \langle \bar{x}, \bar{x} - \bar{y} \rangle - \langle \bar{y}, \bar{x} - \bar{y} \rangle = \\ &= \langle \bar{x}, \bar{x} \rangle + \langle \bar{y}, \bar{y} \rangle - 2\langle \bar{x}, \bar{y} \rangle = \|\bar{x}\|^2 + \|\bar{y}\|^2 - 2\langle \bar{x}, \bar{y} \rangle. \end{aligned} \quad (5.23)$$

La ley del coseno para triángulos determina que:

$$\|\bar{z}\|^2 = \|\bar{x}\|^2 + \|\bar{y}\|^2 - 2 \cdot \|\bar{x}\| \|\bar{y}\| \cdot \cos \varphi \quad (5.24)$$

Comparando esta última expresión con la (5.23) se concluye que:

$$\langle \bar{x}, \bar{y} \rangle = \|\bar{x}\| \|\bar{y}\| \cdot \cos \varphi \quad (5.25)$$

Demostrándose así la equivalencia entre ambas definiciones.

5.1.7 Desigualdad de *Cauchy-Schwarz*

Este apartado se dedica al estudio de la relación que existe entre las normas *euclidianas* y el producto escalar de dos vectores cualesquiera \bar{x}, \bar{y} en \mathbb{R}^n . Desarrollando la expresión $\|\bar{x} + \lambda\bar{y}\|^2$, $\lambda \in \mathbb{R}$ y aplicando las propiedades de distributividad y simetría, se tiene que, para todo $\lambda \in \mathbb{R}$:

$$\|\bar{x} + \lambda\bar{y}\|^2 = \langle \bar{x} + \lambda\bar{y}, \bar{x} + \lambda\bar{y} \rangle = \|\bar{x}\|^2 + 2\lambda\langle \bar{x}, \bar{y} \rangle + \lambda^2\|\bar{y}\|^2 \quad (5.26)$$

Supóngase que $\bar{y} \neq \bar{0}$. La expresión (5.26) es un polinomio de segundo grado en λ que debe ser mayor o igual que cero para todo λ . De lo contrario, el producto escalar podría producir un valor negativo, contradiciendo la propiedad 3 de positividad del producto escalar, Proposición 5.1.1. Por tanto, su discriminante tiene que ser menor o igual que cero:

$$\langle \bar{x}, \bar{y} \rangle^2 - \|\bar{x}\|^2\|\bar{y}\|^2 \leq 0 \Rightarrow \langle \bar{x}, \bar{y} \rangle^2 \leq \|\bar{x}\|^2\|\bar{y}\|^2 \quad (5.27)$$

Tomando raíces cuadradas positivas en ambos miembros queda demostrada una importante propiedad conocida como la “desigualdad de *Cauchy-Schwarz*”.

Proposición 5.1.3 – Desigualdad de *Cauchy-Schwarz*

Sean \bar{x}, \bar{y} dos vectores cualesquiera de \mathbb{R}^n . Su producto escalar y sus normas *euclidianas* cumplen:

$$|\langle \bar{x}, \bar{y} \rangle| \leq \|\bar{x}\|\|\bar{y}\|$$

La igualdad entre ambos términos se da, si y sólo si, \bar{x}, \bar{y} son proporcionales, pues en este caso existiría algún valor de λ tal que $\|\bar{x} + \lambda\bar{y}\|^2 = 0$, y esto último es posible sólo si el discriminante del polinomio (5.27), $\langle \bar{x}, \bar{y} \rangle^2 - \|\bar{x}\|^2\|\bar{y}\|^2$, es igual a cero.

Ejemplo 5.1.5 Se comprueba que los vectores $\bar{x}, \bar{y} \in \mathbb{R}^2$:

$$\bar{x} = (1, 1)^t \quad (5.28)$$

$$\bar{y} = (1, \sqrt{3})^t \quad (5.29)$$

Cumplen la desigualdad de *Cauchy-Schwarz* (5.1.3):

$$|\langle \bar{x}, \bar{y} \rangle| = |1 + \sqrt{3}| = 1 + \sqrt{3} < 2 \cdot \sqrt{2} = \|\bar{x}\|\|\bar{y}\| \quad (5.30)$$

5.1.8 Desigualdad triangular

A continuación, se estudiará la relación existente entre la norma *euclidiana* de la suma de dos vectores reales y la suma de sus normas. Desarrollando la expresión $\|\bar{x} + \bar{y}\|^2$ y aplicando la desigualdad de *Cauchy-Schwarz*:

$$\begin{aligned} \|\bar{x} + \bar{y}\|^2 &= \|\bar{x}\|^2 + 2\langle \bar{x}, \bar{y} \rangle + \|\bar{y}\|^2 \leq \\ &\leq \|\bar{x}\|^2 + 2\|\bar{x}\|\|\bar{y}\| + \|\bar{y}\|^2 = (\|\bar{x}\| + \|\bar{y}\|)^2 \end{aligned} \quad (5.31)$$

Al extraer raíces cuadradas positivas de ambos miembros de la expresión (5.31), se obtiene lo siguiente:

$$\|\bar{x} + \bar{y}\| \leq \|\bar{x}\| + \|\bar{y}\| \quad (5.32)$$

A esta expresión se la conoce como “desigualdad triangular”. La igualdad se da, si y sólo si, $\langle \bar{x}, \bar{y} \rangle = \|\bar{x}\| \|\bar{y}\|$. Por la desigualdad de *Cauchy-Schwartz*, Proposición 5.1.3, esto último ocurre, si y sólo si, \bar{x} y \bar{y} son proporcionales. Es decir, si $\bar{x} = \lambda \bar{y}$ para algún λ real que además debe ser estrictamente no negativo.

Proposición 5.1.4 – Desigualdad triangular

Sean \bar{x} , \bar{y} dos vectores cualesquiera de \mathbb{R}^n , la suma de sus normas euclidianas y la norma euclidiana de su suma cumplen:

$$\|\bar{x} + \bar{y}\| \leq \|\bar{x}\| + \|\bar{y}\|$$

La igualdad se da si y solo si, los vectores son proporcionales con constante de proporcionalidad no negativa.

5.1.9 Ángulo entre dos vectores reales

La equivalencia entre las Definiciones 5.1.2 y 5.1.3 del producto escalar real permite expresar el ángulo entre dos vectores en función de su producto escalar y de sus normas euclidianas y generalizar el concepto de ángulo a los espacios de vectores reales de cualquier dimensión \mathbb{R}^n .

Definición 5.1.6 – Ángulo entre dos vectores en \mathbb{R}^n

El ángulo que forman dos vectores \bar{x} , \bar{y} no nulos cualesquiera de \mathbb{R}^n se define como el escalar real $\varphi \in [0, \pi]$ cuyo coseno es:

$$\cos \varphi = \frac{\langle \bar{x}, \bar{y} \rangle}{\|\bar{x}\| \|\bar{y}\|}$$

Si los vectores \bar{x} , \bar{y} son perpendiculares, se tiene $\varphi = \pi/2$, en cuyo caso $\langle \bar{x}, \bar{y} \rangle = 0$; mientras que si son proporcionales, se tiene $\varphi = 0$, en cuyo caso $\langle \bar{x}, \bar{y} \rangle = \|\bar{x}\| \|\bar{y}\|$.

Ejemplo 5.1.6 El ángulo entre los vectores reales $\bar{x}, \bar{y} \in \mathbb{R}^2$:

$$\bar{x} = (1, 1)^t \quad (5.33)$$

$$\bar{y} = (1, \sqrt{3})^t \quad (5.34)$$

Es aplicando la Definición 5.1.6:

$$\cos \varphi = \frac{\langle \bar{x}, \bar{y} \rangle}{\|\bar{x}\| \|\bar{y}\|} = \frac{1 + \sqrt{3}}{2\sqrt{2}} = \frac{\pi}{2} \equiv 15^\circ \quad (5.35)$$

5.1.10 Definición del producto escalar usual en \mathbb{C}^n

La formulación del producto escalar presentada en la Definición 5.1.3 es aplicable sólo a vectores de componentes reales. A continuación, se presenta una definición generalizada de producto escalar aplicable tanto a vectores reales como complejos. A este producto escalar generalizado se le suele conocer como “producto escalar usual”, Observación 5.1.2.

Observación 5.1.2 Una definición de producto escalar aún más general es la siguiente: Sean una matriz hermitica $A \in \mathbb{C}^{n \times n}$ y dos vectores $x, y \in \mathbb{C}^n$, su producto escalar asociado a A viene dado por:

$$\langle \bar{x}, \bar{y} \rangle = \bar{x}^t A \bar{y}^* = \bar{y}^h A \bar{x} \quad (5.36)$$

La fórmula (5.37) es un caso particular en el que se toma $A = I$.

Definición 5.1.7 — Producto escalar usual

Dados $\bar{x}, \bar{y} \in \mathbb{C}^n$, su producto escalar usual se define como:

$$\langle \bar{x}, \bar{y} \rangle = \bar{x}^t \bar{y}^* = \bar{y}^h \bar{x} \quad (5.37)$$

Si $\bar{x} \in \mathbb{R}^n$, la fórmula (5.37) coincide con la del producto escalar de vectores reales, Definición 5.1.2.

La presencia del conjugado en la definición (5.37) garantiza que el producto escalar de cualquier vector real o complejo consigo mismo sea un número real positivo que cuantifica la longitud del mismo, entendida esta en un sentido *euclidiano*.

Se enumeran a continuación sus propiedades elementales:

Proposición 5.1.5 — Propiedades del producto escalar usual

1. Linealidad respecto de la primera componente, pero no respecto de la segunda: $\forall \bar{x}, \bar{y}, \bar{z} \in \mathbb{C}^n, \forall \lambda, \mu \in \mathbb{C}$

$$\langle \lambda \bar{x} + \mu \bar{y}, \bar{z} \rangle = \lambda \langle \bar{x}, \bar{z} \rangle + \mu \langle \bar{y}, \bar{z} \rangle \quad (5.38)$$

2. Simetría hermítica: $\forall \bar{x}, \bar{y} \in \mathbb{C}^n$

$$\langle \bar{x}, \bar{y} \rangle = \langle \bar{y}, \bar{x} \rangle^* \quad (5.39)$$

3. Positividad: $\forall \bar{x} \in \mathbb{C}^n, \bar{x} \neq \bar{0}$

$$\langle \bar{x}, \bar{x} \rangle \in \mathbb{R} \quad (5.40)$$

$$\langle \bar{x}, \bar{x} \rangle \geq 0, \langle \bar{x}, \bar{x} \rangle = 0 \Leftrightarrow \bar{x} = \bar{0}$$

Observación 5.1.3 † Las propiedades 1 y 2 del producto escalar usual, Proposición 5.1.5, implican una propiedad denominada antilinealidad respecto de la segunda componente: $\forall \bar{x}, \bar{y}, \bar{z} \in \mathbb{C}^n, \forall \lambda, \mu \in \mathbb{C}$:

$$\begin{aligned} \langle \bar{x}, \lambda \bar{y} + \mu \bar{z} \rangle &= \\ &= \lambda^* \langle \bar{x}, \bar{y} \rangle + \mu^* \langle \bar{x}, \bar{z} \rangle. \end{aligned}$$

En general, si un operador $\mathbb{K}^n \times \mathbb{K}^n \mapsto \mathbb{K}$ es lineal respecto de la primera componente y antilineal respecto de la segunda, se dice que es sesquilineal.

A diferencia del producto escalar real, Definición 5.1.3, el producto escalar complejo no es ni simétrico, pues en general $\langle \bar{x}, \bar{y} \rangle \neq \langle \bar{y}, \bar{x} \rangle$, ni lineal respecto del segundo vector †.

Ejemplo 5.1.7 El producto escalar de los vectores complejos $\bar{x}, \bar{y} \in \mathbb{C}^2$:

$$\bar{x} = (1 + i, 2)^t \quad (5.41)$$

$$\bar{y} = (-1 + i, -i)^t \quad (5.42)$$

Es, aplicando la expresión (5.37):

$$\langle \bar{x}, \bar{y} \rangle = \bar{x}^t \bar{y}^* = (1 + i \quad 2) \begin{pmatrix} -1 - i \\ i \end{pmatrix} = \bar{y}^h \bar{x} = -2i + 2i = 0. \quad (5.43)$$

5.1.11 Definición de la norma en \mathbb{C}^n

La propiedad 3 de positividad del producto escalar usual, Proposición 5.1.5, permite dotar a los conjuntos de vectores complejos \mathbb{C}^n de una norma definida de manera análoga a la norma *euclidiana* de los vectores reales. Es decir, como la raíz cuadrada del producto escalar del vector consigo mismo:

Definición 5.1.8 — Norma de un vector en \mathbb{C}^n

Se define la norma de un vector $\bar{x} \in \mathbb{C}^n$ de componentes complejas como:

$$\|\bar{x}\| = \sqrt{\langle \bar{x}, \bar{x} \rangle} = \sqrt{\bar{x}^h \bar{x}} = \sqrt{\sum_{i=1}^n |u_i|^2} \quad (5.44)$$

Si $\bar{x} \in \mathbb{R}^n$, la fórmula (5.44) coincide con la de la norma *euclidiana* de un vector real, ecuación (5.1.4).

Ejercicio 5.1.3 Compruébese que este operador norma para vectores en \mathbb{C}^n , Definición 5.1.8, cumple todas las propiedades de la norma *euclidiana* real, Proposición 5.1.2. Así como las desigualdades triangular, Proposición 5.1.4, y de *Cauchy-Schwarz*, Proposición 5.1.3.

Ejemplo 5.1.8 La norma de los vectores complejos $\bar{x}, \bar{y} \in \mathbb{C}^2$:

$$\bar{x} = (1 + i, 2)^t \quad (5.45)$$

$$\bar{y} = (-1 + i, -i)^t \quad (5.46)$$

Es, aplicando la Definición 5.1.8:

$$\|\bar{x}\| = \sqrt{\langle \bar{x}, \bar{x} \rangle} = \sqrt{6} \quad (5.47)$$

$$\|\bar{y}\| = \sqrt{\langle \bar{y}, \bar{y} \rangle} = \sqrt{3} \quad (5.48)$$

Donde:

$$\langle \bar{x}, \bar{x} \rangle = \bar{x}^h \bar{x} = (1 - i \quad 2) \begin{pmatrix} 1 + i \\ 2 \end{pmatrix} = 6 \quad (5.49)$$

$$\langle \bar{y}, \bar{y} \rangle = \bar{y}^h \bar{y} = (-1 - i \quad i) \begin{pmatrix} -1 + i \\ -i \end{pmatrix} = 3 \quad (5.50)$$

5.1.12 Extensión a espacios vectoriales generales. Los espacios *euclidianos*

Las Definiciones 5.1.7 y 5.1.8, de producto escalar y norma *euclidiana* son aplicables a vectores tanto reales como complejos.

Para generalizar los conceptos geométricos de ángulo y longitud a otros espacios vectoriales, es necesario definir operadores que posean las mismas propiedades que el producto escalar y la norma de vectores. A este efecto, se definirán de manera axiomática dos operadores denominados producto interno y norma. De esta manera, dado un espacio vectorial V , si se define un operador que tenga V como conjunto inicial y que cumpla

esas propiedades, este podrá constituir un producto interno o norma en ese espacio.

En particular, dado un espacio vectorial V , si se dota a este espacio de un producto interno y de una norma definida como la raíz cuadrada del producto interno de un elemento de V consigo mismo, el espacio V se convierte en un espacio *euclidiano*.

5.1.13 Definición general de producto interno

Estas son las propiedades que, por definición, debe cumplir un operador $V \times V \mapsto \mathbb{K}$, siendo V un espacio vectorial, para que este pueda ser considerado un producto interno en V :

Observación 5.1.4 La linealidad respecto del primer componente y la simetría hermitica del producto interno, puntos 1 y 2 de la Definición 5.1.9 implican que este es antilineal respecto de la segunda componente: $\forall x, y, z \in V, \forall \lambda, \mu \in \mathbb{K}$,

$$\begin{aligned} \langle x, \lambda y + \mu z \rangle &= \\ &= \lambda^* \langle x, y \rangle + \mu^* \langle x, z \rangle. \end{aligned} \tag{5.51}$$

En el caso real, $\mathbb{K} = \mathbb{R}$, el producto interno resulta ser simétrico: $\langle x, y \rangle = \langle y, x \rangle$, y en consecuencia, es lineal también respecto de la segunda componente: $\forall x, y, z \in V, \forall \lambda, \mu \in \mathbb{R}$,

$$\begin{aligned} \langle x, \lambda y + \mu z \rangle &= \\ &= \lambda \langle x, y \rangle + \mu \langle x, z \rangle \end{aligned} \tag{5.52}$$

Definición 5.1.9 – Producto interno en un espacio vectorial general

Dado un espacio vectorial V sobre el cuerpo \mathbb{K} , la función $\langle \cdot, \cdot \rangle : V \times V \mapsto \mathbb{K}$ es un producto interno si, $\forall x, y, z \in V, \forall \lambda, \mu \in \mathbb{K}$, la función $\langle \cdot, \cdot \rangle$ verifica las siguientes propiedades:

1. Linealidad respecto de la primera componente.

$$\langle \lambda x + \mu y, z \rangle = \lambda \langle x, z \rangle + \mu \langle y, z \rangle \tag{5.53}$$

2. Simetría hermitica.

$$\langle x, y \rangle = \langle y, x \rangle^* \tag{5.54}$$

3. Positividad.

$$\begin{aligned} \langle x, x \rangle &\geq 0 \\ \langle x, x \rangle = 0 &\Leftrightarrow x = 0 \end{aligned} \tag{5.55}$$

Véase que estas propiedades coinciden con las del producto escalar usual, Proposición 5.1.5.

5.1.14 Definición general de norma. Espacios euclidianos

Estas son las propiedades que, por definición, debe cumplir un operador $V \mapsto \mathbb{K}$, siendo V un espacio vectorial, para que este pueda ser considerado una norma en V :

Definición 5.1.10 – Norma en un espacio vectorial general

Dado un espacio vectorial V sobre el cuerpo \mathbb{K} , la función $\| \cdot \| : V \mapsto \mathbb{K}$ es una norma si, $\forall x, y \in V, \forall \lambda \in \mathbb{K}$, la función $\| \cdot \|$ verifica las siguientes propiedades:

1. Positividad.

$$\|x\| \geq 0, \|x\| = 0 \Leftrightarrow x = 0 \tag{5.58}$$

2. Homogeneidad.

$$\|\lambda x\| = |\lambda| \|x\| \tag{5.59}$$

3. Desigualdad triangular.

$$\|x + y\| \leq \|x\| + \|y\| \tag{5.60}$$

Observación 5.1.5 Las posibilidades a la hora de diseñar una norma o producto interno con el que dotar un espacio son infinitas. Además de la norma *euclidiana*, Definición 5.1.8, son muy empleadas la llamada “norma infinito”, que en \mathbb{K}^n se define como el máximo de los módulos de las componentes del vector:

$$\|\bar{x}\|_\infty = \max_{i \in \{1, \dots, n\}} |x_i| \tag{5.56}$$

Y la llamada “norma uno”, definida en \mathbb{K}^n como la suma de los módulos de las componentes del vector:

$$\|\bar{x}\|_1 = \sum_{i=1}^n |x_i| \tag{5.57}$$

Esta última se conoce también como la “norma *Manhattan*” o la “norma del taxista”. Se deja como ejercicio al lector averiguar con un poco de imaginación el porqué de estos dos nombres.

A la norma que se define como $\|\cdot\| = \sqrt{\langle \cdot, \cdot \rangle}$ dado un determinado producto interno previamente definido en V se le conoce como la “norma canónica del espacio V ”. †

A los espacios vectoriales sobre el cuerpo \mathbb{K} dotados de un producto interno y de la norma canónica a él asociada se les llama también “espacios euclídeos” o “euclidianos”.

Definición 5.1.11 – Espacio euclidiano

Dado un espacio vectorial V , se dice que la tríada $(V, \langle \cdot, \cdot \rangle, \|\cdot\|)$ es un espacio euclidiano si $\langle \cdot, \cdot \rangle$ es un producto interno con todas sus propiedades axiomáticas, Definición 5.1.9, y $\|\cdot\|$ es una norma definida como $\sqrt{\langle \cdot, \cdot \rangle}$.

Como ejemplos, se tienen los espacios euclidianos de vectores de n componentes en el cuerpo \mathbb{K} , ‡ construidos dotando a los espacios \mathbb{K}^n del producto escalar usual $\langle \cdot, \cdot \rangle$, Definición 5.1.7, y de la norma canónica $\|\cdot\| = \sqrt{\langle \cdot, \cdot \rangle}$, Definición 5.1.8.

5.1.15 Un ejemplo en un espacio de funciones

En este apartado, se muestra cómo transformar el conjunto de las funciones en una variable real continuas en el intervalo cerrado $[a, b]$, $a, b \in \mathbb{R}$, denotado como $C[a, b]$, en un espacio euclidiano.

Una función que es continua para todo \mathbb{R} es $f(x) = x^2 \cdot \cos x$, por lo que pertenece al conjunto $C[a, b]$ para cualesquiera cotas superior e inferior $a, b \in \mathbb{R}$ que se establezcan.

Sean $f, g \in C[a, b]$ y $\lambda \in \mathbb{R}$, se definen las operaciones suma y producto de función por escalar del siguiente modo: $\forall x \in \mathbb{R}$

$$(f + g)(x) = f(x) + g(x) \quad (5.61)$$

$$(\lambda f)(x) = \lambda f(x) \quad (5.62)$$

Claramente, la terna $(C[a, b], +, \cdot)$ forma un espacio vectorial denominado V . Además, puede demostrarse que las siguientes operaciones:

$$\langle f, g \rangle = \int_a^b f(t)g(t) dt \quad (5.63)$$

$$\|f\| = \sqrt{\langle f, f \rangle} = \sqrt{\int_a^b f(t)^2 dt} \quad (5.64)$$

Observación 5.1.6 † A veces, a la norma canónica, o euclidiana en \mathbb{K}^n , se la denomina “norma dos” y se la denota como $\|\cdot\|_2$. Esta norma, junto con las dos mencionadas en la Observación 5.1.5, forman parte de una familia de normas llamadas “normas-p” (en los espacios \mathbb{K}^n) o “normas de Lebesgue” (en los espacios de funciones de una variable acotadas en un determinado intervalo cerrado).

Observación 5.1.7 ‡ Desde el punto de vista geométrico, los espacios euclidianos \mathbb{R}^2 y \mathbb{R}^3 son aquellos en los cuales se cumplen los cinco postulados de la geometría de Euclides. Dos consecuencias de estos postulados son que la distancia más corta entre dos puntos es una línea recta, y que los ángulos entre lados contiguos de un triángulo suman 180° .

Modificando el quinto postulado (dados una recta y un punto externo a ella, existe una recta, y sólo una, que lo corte y que sea paralela a la primera) se construyen las llamadas “geometrías no euclidianas”. En estos espacios, la distancia más corta entre dos puntos ya no es una línea recta y los ángulos de los triángulos ya no suman 180 grados. Un ejemplo de geometría no euclidiana es la superficie de una esfera.

De acuerdo con la teoría de la relatividad general, el espacio en el que nos encontramos, al contrario de lo que intuitivamente nos pueda parecer, no es un espacio euclidiano a causa de la curvatura del continuo espacio-tiempo, de la misma forma que La Tierra, pareciéndonos plana, no lo es. Como consecuencia, el camino más corto entre dos puntos cualesquiera del cosmos (y de La Tierra) no es una línea recta, aunque si la distancia es pequeña se puede aproximar como tal.

Ejercicio 5.1.4 † Demostrar que las siguientes operaciones:

$$\langle f, g \rangle = \int_a^b f(t)g(t) dt \quad (5.65)$$

$$\|f\| = \sqrt{\int_a^b f(t)^2 dt} \quad (5.66)$$

donde f, g son dos funciones de una variable real continuas en el intervalo $[a, b]$, $a, b \in \mathbb{R}$, cumplen todos los axiomas de producto interno y norma, Definiciones 5.1.9 y 5.1.10.

Cumplen todas las propiedades axiomáticas del producto interno y de la norma, Definiciones 5.1.9 y 5.1.10. † Por lo que la terna $(V, \langle f, g \rangle, \|f\|)$ forma un espacio *euclidiano*, Definición 5.1.11.

Ejemplo 5.1.9 Las funciones:

$$f(x) = \cos x \quad (5.67)$$

$$g(x) = \operatorname{sen} x \quad (5.68)$$

Pertenecen al espacio de las funciones de una variable real continuas en el intervalo cerrado $[-\pi, \pi]$. Aplicando las definiciones de norma y producto interno dadas por las ecuaciones (5.63) y (5.64) se tiene:

$$\|f\| = \int_{-\pi}^{\pi} \cos^2 x dx = \frac{\cos x \operatorname{sen} x}{2} \Big|_{-\pi}^{\pi} + \frac{x}{2} \Big|_{-\pi}^{\pi} = \pi \quad (5.69)$$

$$\|g\| = \int_{-\pi}^{\pi} \operatorname{sen}^2 x dx = -\frac{\cos x \operatorname{sen} x}{2} \Big|_{-\pi}^{\pi} + \frac{x}{2} \Big|_{-\pi}^{\pi} = \pi \quad (5.70)$$

$$\langle f, g \rangle = \int_{-\pi}^{\pi} \cos x \operatorname{sen} x dx = \frac{\operatorname{sen}^2 x}{2} \Big|_{-\pi}^{\pi} = 0 \quad (5.71)$$

5.1.16 Resumen

1. El producto escalar de dos vectores proporciona una medida del ángulo que forman el uno respecto del otro.
2. La norma *euclidiana* de un vector real, definida como la raíz cuadrada de su producto escalar consigo mismo, proporciona una medida de la longitud de su representación geométrica.
3. Tanto los espacios de vectores reales como los espacios de vectores complejos, equipados con el producto escalar usual y la norma *euclidiana*, Definiciones 5.1.7 y 5.1.8, cumplen las desigualdades de *Cauchy-Schwarz* y triangular.
4. El producto escalar, o producto interno, y la norma de vectores, así como sus propiedades, se pueden extender a los espacios vectoriales generales, sean estos reales o complejos. Para ello, estos se definen de manera axiomática, de acuerdo con las propiedades que debe cumplir todo operador para que pueda ser considerado producto interno o norma.

5.2 Ortogonalidad

5.2.1 Introducción

De las infinitas bases que puede tener un espacio o subespacio vectorial V equipado con un producto interno y una norma con sus propiedades axiomáticas correspondientes, Definiciones 5.1.9 y 5.1.10, existen algunas cuyos elementos son ortogonales, o perpendiculares, dos a dos. Estas bases, denominadas ortogonales, u ortonormales si sus elementos tienen norma unidad, cumplen unas propiedades que las hacen muy útiles e interesantes.

Dichas bases y sus propiedades se estudiarán en esta sección, pero antes se definirá el concepto de ortogonalidad entre elementos de un espacio vectorial, se enunciará y demostrará el teorema de *Pitágoras*, y se definirá el suplementario ortogonal de un subespacio $H \subseteq V$.

5.2.2 Ortogonalidad entre elementos de un espacio vectorial

Definición 5.2.1 — Elementos ortogonales de un espacio vectorial

Sea V un espacio vectorial sobre el cuerpo \mathbb{K} dotado de un producto interno $\langle \cdot, \cdot \rangle$. Los elementos $x, y \in V$ son ortogonales o perpendiculares entre sí, denotándose en tal caso $x \perp y$, si su producto interno es nulo:

$$\langle x, y \rangle = \langle y, x \rangle = 0$$

Ejemplo 5.2.1 Los vectores de \mathbb{R}^3 :

$$\bar{x} = \begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix} \quad (5.72)$$

$$\bar{y} = \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix} \quad (5.73)$$

$$\bar{z} = \begin{pmatrix} 1 \\ 2 \\ 5 \end{pmatrix} \quad (5.74)$$

Cumplen:

$$\langle \bar{x}, \bar{y} \rangle = (2 \quad -1 \quad 0) \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix} = 0 \quad (5.75)$$

$$\langle \bar{x}, \bar{z} \rangle = 0 \quad (5.76)$$

$$\langle \bar{y}, \bar{z} \rangle = 0 \quad (5.77)$$

Por tanto, los vectores del conjunto $\{\bar{x}, \bar{y}, \bar{z}\}$ son ortogonales entre sí dos a dos.

5.2.3 El teorema de *Pitágoras*

Sean x, y dos elementos cualesquiera de un espacio *euclidiano* definido sobre el cuerpo de los números reales y dotado de un producto interno $\langle \cdot, \cdot \rangle$ y de la norma canónica $\| \cdot \| = \sqrt{\langle \cdot, \cdot \rangle}$. Se tiene que:

$$\|x + y\|^2 = \langle x + y, x + y \rangle = \|x\|^2 + 2\langle x, y \rangle + \|y\|^2 \quad (5.78)$$

Si $\langle x, y \rangle = x^t y = 0$, la expresión (5.78) se reduce a:

$$\|x + y\|^2 = \|x\|^2 + \|y\|^2 \quad (5.79)$$

Que constituye la versión clásica del teorema de *Pitágoras*.

Proposición 5.2.1 — Teorema de *Pitágoras* en los espacios *euclidianos* reales

Sea V un espacio vectorial definido sobre el cuerpo de los números reales y dotado de un producto interno $\langle \cdot, \cdot \rangle$ y de la norma canónica $\| \cdot \| = \sqrt{\langle \cdot, \cdot \rangle}$. Dos elementos $x, y \in V$ son ortogonales, si y sólo si:

$$\|x + y\|^2 = \|x\|^2 + \|y\|^2$$

Ejemplo 5.2.2 Los vectores \bar{x}, \bar{y} del Ejemplo 5.2.1 son ortogonales respecto del producto escalar en \mathbb{R}^3 . Sus normas son:

$$\|\bar{x}\| = \sqrt{2^2 + (-1)^2 + 0^2} = \sqrt{5} \quad (5.80)$$

$$\|\bar{y}\| = \sqrt{1^2 + 2^2 + (-1)^2} = \sqrt{6} \quad (5.81)$$

Sumando ambos vectores, se obtiene $\bar{x} + \bar{y} = (2, -1, 0)^t + (1, 2, -1)^t = (3, 1, -1)^t$, por lo que:

$$\|\bar{x} + \bar{y}\| = \sqrt{11} \quad (5.82)$$

Finalmente, se comprueba que los vectores ortogonales \bar{x}, \bar{y} cumplen el teorema de *Pitágoras*:

$$\|\bar{x} + \bar{y}\|^2 = 11 = 5 + 6 = \|\bar{x}\|^2 + \|\bar{y}\|^2 \quad (5.83)$$

Si se tiene un espacio *euclidiano* V definido sobre el cuerpo de los números complejos, es evidente que toda pareja de elementos $x, y \in V$ ortogonales entre si cumple la identidad (5.2.1) pues:

$$\begin{aligned} \|x + y\|^2 &= \|x\|^2 + \|y\|^2 + \langle x, y \rangle + \langle y, x \rangle = \\ &= \|x\|^2 + \|y\|^2 + 0 + 0 = \\ &= \|x\|^2 + \|y\|^2 \end{aligned} \quad (5.84)$$

Sin embargo, es posible encontrar parejas de elementos $x, y \in V$ que sin ser ortogonales cumplen la identidad (5.2.1). Aplicando la propiedad 2 de simetría hermítica del producto interno, Definición 5.1.9, se tiene:

$$\begin{aligned}\|x + y\|^2 &= \|x\|^2 + \|y\|^2 + \langle x, y \rangle + \langle y, x \rangle = \\ &= \|x\|^2 + \|y\|^2 + \langle x, y \rangle + \langle x, y \rangle^* \end{aligned} \quad (5.85)$$

Por lo que basta que los elementos x, y cumplan la identidad $\langle x, y \rangle = -\langle x, y \rangle^*$ para que se cumpla la igualdad (5.2.1).

Como se razonó en la Proposición 2.41, para todo $z \in \mathbb{C}$, $z = -z^* \Leftrightarrow z$ es imaginario puro. Esta última es la condición necesaria y suficiente que el producto interno $\langle x, y \rangle$ debe cumplir para que $\|x + y\|^2 = \|x\|^2 + \|y\|^2$.

Proposición 5.2.2 – Teorema de Pitágoras en los espacios euclidianos complejos

Sean un espacio vectorial V definido sobre el cuerpo de los números reales y dotado de un producto interno $\langle \cdot, \cdot \rangle$ y de la norma canónica $\|\cdot\| = \sqrt{\langle \cdot, \cdot \rangle}$, y sean $x, y \in V$. Se tiene que:

$$\langle x, y \rangle \text{ es imaginario puro} \Leftrightarrow \|x + y\|^2 = \|x\|^2 + \|y\|^2$$

En concreto, si $x, y \in V$ son ortogonales, se cumple la identidad $\|x + y\|^2 = \|x\|^2 + \|y\|^2$.

Supóngase que x_1, \dots, x_k son k elementos de un espacio vectorial V dotado de un producto interno $\langle \cdot, \cdot \rangle$ y de la norma canónica $\sqrt{\langle \cdot, \cdot \rangle}$. Desarrollando $\|x_1 + x_2 + \dots + x_k\|^2$, se tiene que:

$$\begin{aligned}\|x_1 + x_2 + \dots + x_k\|^2 &= \\ &= \langle x_1 + x_2 + \dots + x_k, x_1 + x_2 + \dots + x_k \rangle = \\ &= \langle x_1, x_1 \rangle + \dots + \langle x_k, x_k \rangle + \sum_{i \neq j} \langle x_i, x_j \rangle \end{aligned} \quad (5.86)$$

Si x_1, \dots, x_k son ortogonales dos a dos, por la Definición 5.2.1 de ortogonalidad, $\langle x_i, x_j \rangle = 0 \forall i, j = 1, \dots, k, i \neq j$, por lo que:

$$\begin{aligned}\|x_1 + \dots + x_k\|^2 &= \langle x_1, x_1 \rangle + \dots + \langle x_k, x_k \rangle \\ &= \|x_1\|^2 + \dots + \|x_k\|^2 \end{aligned} \quad (5.87)$$

Combinando la expresión (5.87) con (5.2.1), y de la Proposición 5.2.2, se deduce una versión general de *Pitágoras* aplicable a $k > 2$ elementos de V ortogonales dos a dos respecto a un determinado producto interno:

Teorema 5.2.3 – Generalización del teorema de Pitágoras

Sean k elementos x_1, \dots, x_k de un espacio vectorial V sobre el cuerpo \mathbb{K} dotado de un producto interno $\langle \cdot, \cdot \rangle$ y de la norma canónica $\sqrt{\langle \cdot, \cdot \rangle}$:

$$\begin{aligned}x_i \perp x_j \forall i, j = 1, \dots, k, i \neq j &\Rightarrow \\ \Rightarrow \|x_1 + \dots + x_k\|^2 &= \|x_1\|^2 + \dots + \|x_k\|^2. \end{aligned}$$

Si $k = 2$ y V es un espacio sobre el cuerpo de los números reales, la implicación es doble.

5.2.4 Suplementario ortogonal

Definición 5.2.2 — Suplementario ortogonal de un subespacio vectorial

Sea H un subespacio en V dotado del producto interno $\langle \cdot, \cdot \rangle$, se define el suplementario ortogonal de H como el conjunto formado por todos los elementos de V que son ortogonales a los de $H \subseteq V$:

$$H^\perp = \{y \in V : \forall x \in H, \langle x, y \rangle = 0\}$$

El suplementario ortogonal de todo subespacio en V es siempre un subespacio vectorial en V , sea cual sea el espacio vectorial V que los contiene. Se propone al lector † demostrarlo.

Ejercicio 5.2.1 † Demuéstrese aplicando las propiedades axiomáticas del producto interno, Definición 5.1.9, y la Definición 1.2.6 de subespacio vectorial, que el suplementario ortogonal de un subespacio vectorial en V es también un subespacio en V .

Ejemplo 5.2.3 Llámese H a la clausura lineal de los vectores $\bar{u} = (2, -1, 0)^t$ y $\bar{v} = (1, 2, -1)^t$ de \mathbb{R}^3 . Por aplicación directa de la Definición 5.2.2 de suplementario ortogonal se tiene que $\bar{x} = (x_1, x_2, x_3)^t \in \mathbb{R}^3$ pertenece al suplementario ortogonal de H si y sólo si $\langle \bar{x}, \lambda \bar{u} + \mu \bar{v} \rangle = 0 \forall \lambda, \mu \in \mathbb{R}$, lo que lleva a estas dos condiciones que han de cumplirse simultáneamente para que $\bar{x} \in H^\perp$:

$$\langle \bar{u}, \bar{x} \rangle = 0 \quad (5.88)$$

$$\langle \bar{v}, \bar{x} \rangle = 0 \quad (5.89)$$

Aplicando la Definición 5.1.3 del producto escalar de vectores reales, las condiciones (5.88) y (5.89) pueden reescribirse del siguiente modo:

$$\begin{cases} 2x_1 - x_2 = 0 & (5.90a) \\ x_1 + 2x_2 - x_3 = 0 & (5.90b) \end{cases}$$

Estas ecuaciones implícitas definen un subespacio que coincide precisamente con el suplementario ortogonal de H . Geométricamente, puede interpretarse H como un plano en \mathbb{R}^3 que tiene por ecuación normal $x_1 + 2x_2 + 5x_3 = 0$, mientras que H^\perp es una recta perpendicular a H formada por la intersección de los planos de ecuaciones normales $2x_1 - x_2 = 0$ y $x_1 + 2x_2 - x_3 = 0$.

Dados un espacio vectorial V de dimensión n dotado de producto interno y un subespacio $H \subseteq V$ de dimensión $k \leq n$, H y su suplementario ortogonal guardan esta importante relación respecto de V :

Proposición 5.2.4 — Suplementariedad de H y H^\perp respecto de V

Sea V un espacio vectorial en el cuerpo \mathbb{K} dotado de producto interno y H un subespacio en V , H y H^\perp son suplementarios en V :

$$\dim H \oplus H^\perp = V$$

Demostración:

Sea $\mathcal{B} = (u_1, \dots, u_k)$ una base de H . Como se ha mostrado en el Ejemplo 5.2.3, todo elemento $x \in H^\perp$ ha de cumplir $\langle x, u_1 \rangle = 0, \dots, \langle x, u_i \rangle = 0, \dots, \langle x, u_k \rangle = 0 \forall i = 1, \dots, k$. Cada una de estas k restricciones es lineal, pues el producto interno es lineal respecto de la primera componente, propiedad 1 de la Definición 5.1.9, y además son linealmente independientes entre si por formar los elementos de \mathcal{B} una familia libre, por lo que las k condiciones determinan un conjunto de k ecuaciones implícitas de un subespacio en dimensión $n - k$ que es precisamente H^\perp . Por la relación de *Grassmann*, Teorema 1.4.5 se tiene que V, H y H^\perp cumplen:

$$\begin{aligned} \dim H + H^\perp &= \dim H + \dim H^\perp - \dim H \cap H^\perp = \\ &= k + (n - k) - \dim H \cap H^\perp = \quad (5.91) \\ &= n - \dim H \cap H^\perp \end{aligned}$$

Falta sólo calcular el subespacio $H \cap H^\perp$. Las Definiciones A.2.5 y 5.2.2 de conjunto intersección y de suplementario ortogonal, así como la Proposición 1.4.1, implican que todo elemento de $H \subseteq V$ que pertenezca también a H^\perp ha de ser ortogonal consigo mismo, $\langle x, x \rangle = 0$. Pero por la propiedad 3 de positividad del producto interno, Definición 5.1.9, el único elemento de un espacio o subespacio vectorial equipado de producto interno que cumple $\langle x, x \rangle = 0$ es el elemento nulo. Por tanto, $H \cap H^\perp = \{0\}$.

Por otro lado, se tiene que, por la Proposición 1.3.3, $\dim H + H^\perp = n \Leftrightarrow H + H^\perp = V$, por lo que se cumplen las dos condiciones que en virtud de la Definición 1.4.3 han de cumplir dos subespacios para que sean suplementarios en V :

1. $H \cap H^\perp = \{0\}$.
2. $H + H^\perp = V$.

Se demuestra así la Proposición 5.2.4.

□

Ejemplo 5.2.4 Sea H el subespacio del Ejemplo 5.2.3 las ecuaciones (5.90a) y (5.90b) son ecuaciones implícitas linealmente independientes que definen su suplementario ortogonal H^\perp . Por esto y por la Proposición 5.2.4, $\dim H^\perp = 3 - 2 = 1$.

Un vector de \mathbb{R}^3 que cumple ambas ecuaciones implícitas es el $(1, 2, 5)^t$. Por tanto:

$$H^\perp = L \left[\left(\begin{pmatrix} 1 \\ 2 \\ 5 \end{pmatrix} \right) \right] \subset \mathbb{R}^3 \quad (5.92)$$

Los vectores \bar{u} , \bar{v} y \bar{w} son linealmente independientes entre si, por lo que forman una base del espacio \mathbb{R}^3 compuesta por tres vectores reales de tres componentes:

$$H \oplus H^\perp = L[(\bar{u}, \bar{v}, \bar{w})] = \mathbb{R}^3 \quad (5.93)$$

Si $V = \mathbb{K}^n$, la Definición 5.2.2 del suplementario ortogonal de un subespacio y el Teorema 3.4.5 del rango-nulidad implican una relación muy interesante entre la imagen de una matriz cualquiera $A \in \mathbb{K}^{m \times n}$ y el núcleo de su transpuesta conjugada:

Proposición 5.2.5 – Relación entre la imagen de una matriz y el núcleo de su adjunta

Toda matriz $A \in \mathbb{K}^{m \times n}$ cumple:

$$(\text{Im } A)^\perp = \ker A^h \tag{5.94}$$

O, equivalentemente:

$$(\text{Im } A^h)^\perp = \ker A \tag{5.95}$$

Demostración:

Llámesse H a la clausura lineal de las n columnas de A , de modo que $H = \text{Im } A$. Por la definición de núcleo de una matriz, ecuación (3.74), $\bar{x} \in \mathbb{K}^n$ pertenece a $\ker A^h \Leftrightarrow A^h \bar{x} = \bar{0}$. Esta es una ecuación matricial que puede reescribirse como un conjunto de m ecuaciones lineales:

$$(C_1(A))^h \cdot \bar{x} = 0 \tag{5.96}$$

$$(C_2(A))^h \cdot \bar{x} = 0 \tag{5.97}$$

\vdots

$$(C_n(A))^h \cdot \bar{x} = 0 \tag{5.98}$$

Por la Definición 5.1.7 del producto escalar usual, las n ecuaciones [(5.96) ... (5.98)] pueden reescribirse como:

$$\langle \bar{x}, C_1(A) \rangle = 0 \tag{5.99}$$

$$\langle \bar{x}, C_2(A) \rangle = 0 \tag{5.100}$$

\vdots

$$\langle \bar{x}, C_n(A) \rangle = 0 \tag{5.101}$$

Por lo que todo vector de \mathbb{K}^n perteneciente a $\ker A^h = \bar{0}$ también pertenece a $H^\perp = (\text{Im } A)^\perp$. □

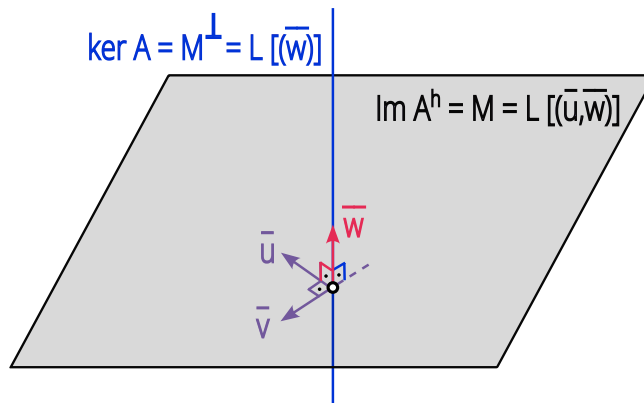


Figura 5.2: El núcleo de la matriz $A \in \mathbb{R}^{2 \times 3}$, subespacio en \mathbb{R}^3 de dimensión uno, es el suplementario ortogonal de la imagen de su transpuesta. Dos vectores pertenecientes uno a $\ker A$ y otro a $\text{Im } A^t$ son siempre ortogonales: $\langle \lambda_1 \bar{u} + \lambda_2 \bar{v}, \mu \bar{w} \rangle = 0 \forall \lambda_1, \lambda_2, \mu \in \mathbb{R}$.

Ejemplo 5.2.5 Se define el subespacio H como la imagen de la transpuesta de la matriz:

$$A = \begin{pmatrix} 2 & -1 & 0 \\ 1 & 2 & -1 \end{pmatrix} \quad (5.102)$$

El núcleo de A es el subespacio siguiente, Ejercicio 5.2.2:

$$\ker A = L \left[\left(\begin{pmatrix} 1 \\ 2 \\ 5 \end{pmatrix} \right) \right] \subset \mathbb{R}^3 \quad (5.103)$$

Que en virtud de la Proposición 5.2.5 coincide con el suplementario ortogonal de H :

$$H^\perp = (\operatorname{Im} A^t)^\perp = \ker A \quad (5.104)$$

De hecho, el vector $(1, 2, 5)^t \in \ker A$ es ortogonal respecto a cada una de las filas de A (Figura 5.2).

Ejercicio 5.2.2 Comprobar que el vector $(1, 2, 5)^t \in \mathbb{R}^3$ forma una base de $\ker A$.

5.2.5 Familias ortogonales y ortonormales

Sea V un espacio vectorial dotado de un producto interno y de una norma, la Definición 5.2.1 y el Ejemplo 5.2.1 muestran que k elementos u_1, \dots, u_k de V son ortogonales dos a dos si $\langle u_i, u_j \rangle = 0 \forall i, j = 1, \dots, k, i \neq j$. En tal caso, los k elementos constituirían una familia ortogonal. Si además la norma de los k vectores es igual a uno, puede decirse que estos constituyen una familia ortonormal.

Definición 5.2.3 – Familias ortogonales y ortonormales

Un conjunto $F = \{u_1, \dots, u_k\}$ de elementos de un espacio V dotado de producto interno y de norma es una:

1. Familia ortogonal: si $\langle u_i, u_j \rangle = 0 \forall i, j = 1, \dots, k, i \neq j$. Es decir, si sus elementos son ortogonales dos a dos.
2. Familia ortonormal: si los elementos de F son ortogonales dos a dos y además unitarios: $\|u_i\| = 1 \forall i = 1, \dots, k$.

Observación 5.2.1 Nótese que todas las familias ortonormales son familias ortogonales, pero no todas las familias ortogonales son ortonormales.

Ejemplo 5.2.6 El conjunto de \mathbb{R}^3 formado por los vectores [(5.72) ... (5.74)] del Ejemplo 5.2.1 y denotado como $F = \{\bar{x}, \bar{y}, \bar{z}\}$, es una familia ortogonal. Puesto que, como se ha visto en el Ejemplo 5.2.1, son ortogonales dos a dos respecto al producto escalar de vectores reales.

Las normas *euclidianas* de los vectores [(5.72) ... (5.74)] son:

$$\|\bar{x}\| = \sqrt{\langle \bar{x}, \bar{x} \rangle} = \sqrt{5} \quad (5.105)$$

$$\|\bar{y}\| = \sqrt{\langle \bar{y}, \bar{y} \rangle} = \sqrt{6} \quad (5.106)$$

$$\|\bar{z}\| = \sqrt{\langle \bar{z}, \bar{z} \rangle} = \sqrt{30} \quad (5.107)$$

Observación 5.2.2 † A esto se le llama “normalizar un elemento de un espacio vectorial respecto de una determinada norma”, que en este ejemplo es la norma canónica de \mathbb{R}^3 .

Dividiendo cada uno de los vectores de F por su norma † se obtiene un nuevo subconjunto de \mathbb{R}^3 al que se denotará como F' y que está compuesto por los siguientes vectores:

$$\bar{x} = \frac{1}{\sqrt{5}} \cdot \begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix} \quad (5.108)$$

$$\bar{y} = \frac{1}{\sqrt{6}} \cdot \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix} \quad (5.109)$$

$$\bar{z} = \frac{1}{\sqrt{30}} \cdot \begin{pmatrix} 1 \\ 2 \\ 5 \end{pmatrix} \quad (5.110)$$

Los vectores de F' son ortogonales dos a dos y unitarios, por lo que F' es una familia ortonormal, punto 2 de la Definición 5.2.3.

Elementos de un espacio vectorial dotado de producto interno que son ortogonales dos a dos son linealmente independientes entre sí, por lo que forman familias libres.

Proposición 5.2.6 – Independencia lineal y ortogonalidad

Toda familia ortogonal es libre. Es decir:

$$F \text{ es una familia ortogonal} \Rightarrow F \text{ es una familia libre.} \quad (5.111)$$

Demostración:

Sea $F = \{u_1, x_2, \dots, u_k\}$ una familia ortogonal contenida en un espacio V dotado de un producto escalar $\langle \cdot, \cdot \rangle$, iguálense a cero la combinación lineal:

$$\lambda_1 u_1 + \lambda_2 x_2 + \dots + \lambda_s u_k = 0 \quad (5.112)$$

Y multiplíquense ambos miembros por un elemento i -ésimo cualquiera de F mediante el producto interno $\langle \cdot, \cdot \rangle$:

$$\lambda_1 \langle u_i, u_1 \rangle + \lambda_2 \langle u_i, x_2 \rangle + \dots + \lambda_k \langle u_i, u_k \rangle = 0 \quad (5.113)$$

Por la ortogonalidad entre los elementos de F la expresión anterior se reduce a $\lambda_i \langle u_i, u_i \rangle = 0$. Los elementos de F no pueden ser nulos, pues de lo contrario, F ya no podría ser una familia ortogonal. Por lo que $\langle u_i, u_i \rangle$ ha de ser distinto de cero por la propiedad 3 de positividad del producto interno, Definición 5.1.9, y así se tiene que $\lambda_k \langle u_i, u_i \rangle = 0 \Leftrightarrow \lambda_i = 0$.

Haciendo variar i desde 1 hasta k se demuestra que:

$$\lambda_1 \langle u_i, u_1 \rangle + \lambda_2 \langle u_i, u_2 \rangle + \dots + \lambda_k \langle u_i, u_k \rangle = 0 \Leftrightarrow \lambda_1, \dots, \lambda_k = 0 \quad (5.114)$$

$\forall i = 1, \dots, k$. Por tanto, por la Definición 1.2.3, los vectores u_1, x_2, \dots, u_k son linealmente independientes y se tiene que toda familia ortogonal es una familia libre.

□

Ejemplo 5.2.7 La familia $F \subset \mathbb{R}^3$ del Ejemplo 5.2.6 es ortogonal, tal como se ve en el Ejemplo 5.2.1, por lo que $F = \{\bar{x}, \bar{y}, \bar{z}\}$ es también una familia libre. Los tres elementos que lo contienen son vectores linealmente independientes de tres componentes, por lo que forman una base ortogonal de \mathbb{R}^3 .

5.2.6 Bases ortogonales y ortonormales

Si una familia ortogonal u ortonormal engendra un espacio o subespacio vectorial V , se dice que constituye una base ortogonal u ortonormal del mismo.

Definición 5.2.4 – Bases ortogonales y ortonormales

Una base \mathcal{B} de un espacio o subespacio vectorial V equipado con un producto interno $\langle \cdot, \cdot \rangle$ es:

1. Ortogonal: si sus elementos forman una familia ortogonal respecto de $\langle \cdot, \cdot \rangle$.
2. Ortonormal: si sus elementos forman una familia ortonormal respecto de $\langle \cdot, \cdot \rangle$.

Una gran ventaja de las bases ortogonales es que resulta muy simple calcular las coordenadas de un elemento del espacio o subespacio que generan respecto de estas bases. Sea $\mathcal{B} = (u_1, \dots, u_n)$ una base ortogonal de un espacio o subespacio V dotado del producto interno $\langle \cdot, \cdot \rangle$, y sea v un elemento perteneciente a V :

$$v = \sum_{i=1}^n \lambda_i u_i \quad (5.115)$$

Multiplicando ambos miembros de esta identidad por un elemento k -ésimo de la base \mathcal{B} en función del producto interno $\langle \cdot, \cdot \rangle$, y teniendo en cuenta que $\langle u_i, u_j \rangle = 0 \forall i, j = 1, \dots, n, i \neq j$:

$$\langle v, u_k \rangle = \left\langle \sum_{i=1}^n \lambda_i u_i, u_k \right\rangle = \lambda_k \langle u_k, u_k \rangle \quad (5.116)$$

Despejando λ_k se obtiene $\lambda_k = \langle v, u_k \rangle / \langle u_k, u_k \rangle$. Repitiendo desde $k = 1$ hasta n se concluye que v puede expresarse como:

$$v = \sum_{i=1}^n \lambda_i u_i, \quad \lambda_i = \frac{\langle v, u_i \rangle}{\langle u_i, u_i \rangle} \quad (5.117)$$

Proposición 5.2.7 – Coordenadas respecto de una base ortogonal

Sean un espacio o subespacio vectorial V dotado de un producto interno $\langle \cdot, \cdot \rangle$, una base \mathcal{B} ortogonal de V y un elemento $v \in V$, la i -ésima coordenada de v respecto de \mathcal{B} es:

$$\lambda_i = \frac{\langle v, u_i \rangle}{\langle u_i, u_i \rangle}$$

Cada término de la descomposición (5.117) puede interpretarse geoméricamente como la proyección ortogonal de v sobre el subespacio engendrado por el i -ésimo elemento de la base ortogonal (u_1, u_2, \dots, u_n) .

Para comprender esta analogía, supóngase que $\bar{v} \in \mathbb{R}^2$ o \mathbb{R}^3 , $\langle \cdot, \cdot \rangle$ es el producto escalar real, Definición 5.1.3, y considérese la norma canónica $\| \cdot \| = \sqrt{\langle \cdot, \cdot \rangle}$. Dado que, como se vio en el apartado 5.1.3, $\langle \bar{v}, \bar{u}_i \rangle = \|\bar{v}\| \|\bar{u}_i\| \cdot \cos \varphi$, se tiene:

$$\begin{aligned} \frac{\langle \bar{v}, \bar{u}_i \rangle}{\langle \bar{u}_i, \bar{u}_i \rangle} \bar{u}_i &= \frac{\langle \bar{v}, \bar{u}_i \rangle}{\|\bar{u}_i\|^2} \cdot \bar{u}_i = \\ &= \frac{\|\bar{v}\| \|\bar{u}_i\| \cos \varphi}{\|\bar{u}_i\|^2} \cdot \bar{u}_i = \|\bar{v}\| \cos \varphi \cdot \frac{\bar{u}_i}{\|\bar{u}_i\|} \end{aligned} \tag{5.118}$$

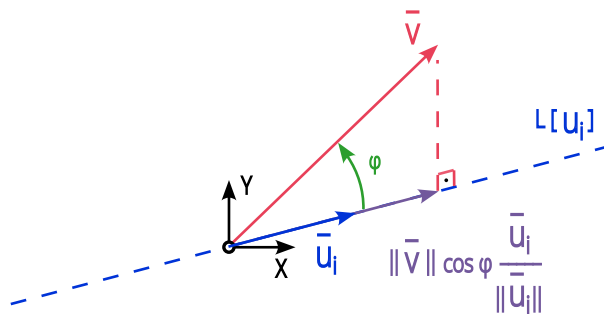
Que es un vector que tiene la dirección de \bar{u}_i y una longitud igual a la de la sombra que arroja perpendicularmente el vector \bar{v} sobre el vector \bar{u}_i (Figura 5.3).

En la sección 5.5, se demuestra que la proyección ortogonal de $v \in V$ sobre el subespacio $L[(u_i)]$ es, de todos los elementos que el subespacio $L[(u_i)]$ contiene, aquel cuya distancia euclidiana a V es mínima:

$$P_V v = \min_{u \in L[(u_i)]} d(v, u) = \min_{u \in L[(u_i)]} \|v - u\| \tag{5.119}$$

Donde $\| \cdot \|$ denota la norma canónica $\sqrt{\langle \cdot, \cdot \rangle}$.

Figura 5.3: La proyección ortogonal del vector \bar{v} sobre \bar{u}_i (en violeta) es un vector perteneciente a $L[(\bar{u}_i)]$ cuya norma, o longitud, es $\|\bar{v}\| \cdot \cos \varphi$, siendo φ el ángulo entre \bar{v} y \bar{u}_i .



Ejemplo 5.2.8 Sea la base ortogonal del espacio euclidiano \mathbb{R}^3 formada por los vectores [(5.72) ... (5.74)] del Ejemplo 5.2.1, denotada $\mathcal{B} = (\bar{x}, \bar{y}, \bar{z})$. Las normas euclidianas de los vectores que contiene son:

$$\|\bar{x}\| = \sqrt{\langle \bar{x}, \bar{x} \rangle} = \sqrt{5} \tag{5.120}$$

$$\|\bar{y}\| = \sqrt{\langle \bar{y}, \bar{y} \rangle} = \sqrt{6} \tag{5.121}$$

$$\|\bar{z}\| = \sqrt{\langle \bar{z}, \bar{z} \rangle} = \sqrt{30} \tag{5.122}$$

Las coordenadas del vector $\bar{v} = (0, 1, 3)^t$ con respecto a la base \mathcal{B} pueden calcularse aplicando la Proposición 5.2.8:

$$\lambda_1 = \frac{\langle \bar{v}, \bar{x} \rangle}{\|\bar{x}\|^2} = -\frac{1}{5} \quad (5.123)$$

$$\lambda_2 = \frac{\langle \bar{v}, \bar{y} \rangle}{\|\bar{y}\|^2} = -\frac{1}{6} \quad (5.124)$$

$$\lambda_3 = \frac{\langle \bar{v}, \bar{z} \rangle}{\|\bar{z}\|^2} = \frac{17}{30} \quad (5.125)$$

Por lo tanto, el vector de coordenadas de \bar{v} respecto de la base \mathcal{B} es el siguiente:

$$[\bar{v}]^{\mathcal{B}} = (\lambda_1, \lambda_2, \lambda_3) = \left(-\frac{1}{5}, -\frac{1}{6}, \frac{17}{30} \right) \quad (5.126)$$

Ejemplo 5.2.9 Los vectores complejos $\bar{x}, \bar{y}, \bar{z}$ siguientes:

$$\bar{x} = \begin{pmatrix} 1+i \\ -i \\ 0 \end{pmatrix} \quad \bar{y} = \begin{pmatrix} 1 \\ 1-i \\ 0 \end{pmatrix} \quad \bar{z} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \quad (5.127)$$

Son ortogonales dos a dos † con respecto al producto escalar usual de los vectores complejos, por lo que forman una base ortogonal de \mathbb{C}^3 a la que llamaremos \mathcal{B} .

Las coordenadas del vector $\bar{v} = (1 - 2i, 0, 4)^t$ respecto de la base \mathcal{B} se calculan aplicando la ecuación (5.117) con el producto escalar usual $\langle \bar{x}, \bar{y} \rangle = \bar{y}^t \bar{v}^* = \bar{y}^h \bar{x}$ y la norma canónica $\|\cdot\| = \sqrt{\langle \cdot, \cdot \rangle}$.

$$\lambda_1 = \frac{\langle \bar{v}, \bar{x} \rangle}{\|\bar{x}\|^2} = 1 - \frac{1}{3}i \quad (5.128)$$

$$\lambda_2 = \frac{\langle \bar{v}, \bar{y} \rangle}{\|\bar{y}\|^2} = \frac{1}{3} - \frac{2}{3}i \quad (5.129)$$

$$\lambda_3 = \frac{\langle \bar{v}, \bar{z} \rangle}{\|\bar{z}\|^2} = 4 \quad (5.130)$$

Supóngase que al espacio o subespacio V se le dota de la norma canónica $\sqrt{\langle \cdot, \cdot \rangle}$. Si se normaliza cada elemento de la base \mathcal{B} mediante la fórmula $q_i = u_i / \|u_i\|$ (Figura 5.4) se obtiene una nueva base $\mathcal{B}_o = (q_1, \dots, q_n)$ que es ortonormal.

La descomposición de $v \in V$ en combinación lineal de los elementos de la nueva base ortonormal \mathcal{B}_o puede llevarse a cabo con una versión simplificada de la expresión (5.117):

$$v = \sum_{i=1}^n \lambda_i q_i, \quad \lambda_i = \langle v, q_i \rangle \quad (5.131)$$

De modo que el vector de coordenadas de V respecto de una base ortonormal \mathcal{B}_o resulta muy sencillo de calcular:

$$[v]^{\mathcal{B}_o} = \begin{pmatrix} \langle v, q_1 \rangle \\ \vdots \\ \langle v, q_n \rangle \end{pmatrix} \quad (5.132)$$

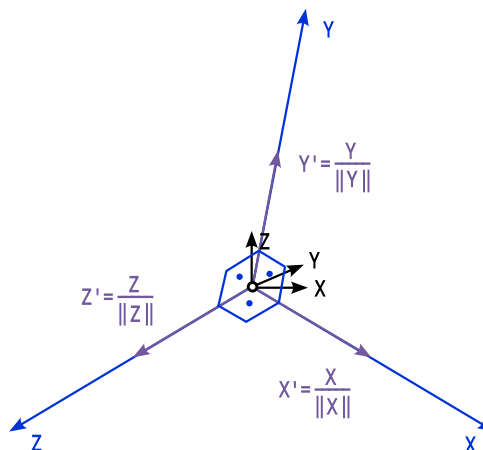
Ejercicio 5.2.3 † Comprobar que los vectores de \mathcal{B} forman una familia ortogonal.

Proposición 5.2.8 – Coordenadas respecto de una base ortonormal

Sean un espacio o subespacio vectorial V dotado de un producto interno $\langle \cdot, \cdot \rangle$, una base \mathcal{B} ortonormal de V y un elemento $v \in V$, la i -ésima coordenada de v respecto de \mathcal{B} es:

$$\lambda_i = \langle v, q_i \rangle$$

Figura 5.4: Los vectores $\bar{x}, \bar{y}, \bar{z} \in \mathbb{R}^3$ son ortogonales dos a dos, por lo que forman una base ortogonal de \mathbb{R}^3 . Si se normalizan dividiéndolos por su norma, se obtiene una base ortonormal de \mathbb{R}^3 .



Ejemplo 5.2.10 Para transformar la base \mathcal{B} del espacio euclideo \mathbb{R}^3 formada por los vectores [(5.72) ... (5.74)] del Ejemplo 5.2.1 en una base ortonormal de \mathbb{R}^3 , se normalizan sus vectores respecto de la norma euclidea:

$$\|\bar{x}\| = \sqrt{\langle \bar{x}, \bar{x} \rangle} = \sqrt{5} \Rightarrow \bar{x}' = \frac{\bar{x}}{\|\bar{x}\|} = \frac{1}{\sqrt{5}} \cdot (2, -1, 0)^t \quad (5.133)$$

$$\|\bar{y}\| = \sqrt{\langle \bar{y}, \bar{y} \rangle} = \sqrt{6} \Rightarrow \bar{y}' = \frac{\bar{y}}{\|\bar{y}\|} = \frac{1}{\sqrt{6}} \cdot (1, 2, -1)^t \quad (5.134)$$

$$\|\bar{z}\| = \sqrt{\langle \bar{z}, \bar{z} \rangle} = \sqrt{30} \Rightarrow \bar{z}' = \frac{\bar{z}}{\|\bar{z}\|} = \frac{1}{\sqrt{30}} \cdot (1, 2, 5)^t \quad (5.135)$$

La base ortogonal de \mathbb{R}^3 obtenida se denota como $\mathcal{B}_o = (\bar{x}', \bar{y}', \bar{z}')$.

Las coordenadas del vector $\bar{v} = (0, 1, 3)^t$ con respecto de la base \mathcal{B}_o se calculan aplicando la Proposición 5.2.7:

$$\lambda_1 = \langle \bar{v}, \bar{x}' \rangle = -\frac{1}{\sqrt{5}} \quad (5.136)$$

$$\lambda_2 = \langle \bar{v}, \bar{y}' \rangle = -\frac{1}{\sqrt{6}} \quad (5.137)$$

$$\lambda_3 = \langle \bar{v}, \bar{z}' \rangle = \frac{17}{\sqrt{30}} \quad (5.138)$$

Por lo tanto, el vector de coordenadas de \bar{v} respecto de la base \mathcal{B}_o es:

$$[\bar{v}]^{\mathcal{B}_o} = (\lambda_1, \lambda_2, \lambda_3) = \left(-\frac{1}{\sqrt{5}}, -\frac{1}{\sqrt{6}}, \frac{17}{\sqrt{30}} \right) \quad (5.139)$$

5.2.7 Resumen

1. Dos elementos de un espacio o subespacio vectorial V dotado de producto interno son ortogonales entre sí si su producto interno es nulo. Un conjunto de elementos ortogonales dos a dos forman una familia ortogonal. Si además son unitarios, forman una familia ortonormal. Todas estas familias son libres.
2. Elementos de V ortogonales entre sí cumplen el teorema de *Pitágoras*: la norma canónica de su suma es igual a la raíz cuadrada de la suma de sus normas canónicas al cuadrado.
3. A cada subespacio H de V se le asocia otro subespacio formado por todos aquellos elementos de V que son perpendiculares a los de H . Dicho subespacio es suplementario a H en V .
4. Las coordenadas de cada elemento v de V respecto de una base ortogonal \mathcal{B} son calculables realizando el producto interno de v con cada elemento de la base. Cada término de la descomposición de v como combinación lineal de los elementos de \mathcal{B} es la proyección ortogonal de v sobre cada elemento de \mathcal{B} .

5.3 Matrices unitarias, ortogonales y de giro

5.3.1 Introducción

Algunos tipos de matrices, como las matrices simétricas y hermíticas, aparecen con cierta asiduidad en la modelización matemática de los sistemas físicos y en el análisis de esos modelos. Esto es así porque estas tipologías de matrices reflejan propiedades y características del mundo físico que nos rodea, como la simetría y la conservación de la energía y del momento, y porque tienen propiedades matemáticas que hacen más estables † y eficientes los cálculos numéricos.

Observación 5.3.1 † La cuantificación de la estabilidad de las matrices se basa en los llamados “números de condición”. Un número de condición bajo indica una matriz numérica estable, o “bien condicionada”. Un número alto indica una matriz menos estable, o numéricamente inestable, en cuyo caso se dice que es una matriz “mal condicionada”. De las matrices de rango deficiente, Definición 3.4.7, se dice que tienen número de condición “infinito”.

Por esos mismos motivos, es habitual encontrar matrices cuadradas cuyas filas y columnas son ortogonales dos a dos y además de norma unidad. Estas son las llamadas “matrices unitarias”, o si todas sus entradas son reales, matrices ortogonales. Esta sección se dedicará a la definición y el estudio de las propiedades de estas matrices y de un tipo particular de matriz ortogonal denominado matriz de giro con importantes aplicaciones en física y en mecánica como, por ejemplo, los cambios de sistema de coordenadas.

5.3.2 Matrices unitarias

Definición 5.3.1 — Matrices unitarias

Una matriz $U \in \mathbb{C}^{n \times n}$ se denomina unitaria si cumple:

$$U^h U = U U^h = I \quad (5.140)$$

O, equivalente, si su inversa coincide con su transpuesta conjugada:

$$U^h = U^{-1} \quad (5.141)$$

Ejemplo 5.3.1 Las siguientes matrices de orden dos son unitarias, como se puede comprobar:

$$U_1 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad (5.142)$$

$$U_2 = \begin{pmatrix} 1 & 1 \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ i & -i \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \quad (5.143)$$

$$U_3 = \begin{pmatrix} 1-i & -1-i \\ \frac{2-i}{2} & \frac{2+i}{2} \end{pmatrix} \quad (5.144)$$

La siguiente matriz de orden tres es también unitaria:

$$U_4 = \begin{pmatrix} -1 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & \frac{-i}{\sqrt{2}} \\ 0 & \frac{-i}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \quad (5.145)$$

Como consecuencia directa de la Definición 5.3.1, las filas y las columnas de una matriz unitaria de orden n forman bases ortonormales de \mathbb{C}^n respecto al producto escalar usual, Definición 5.1.7, y la norma canónica que este define, Definición 5.1.8.

La siguiente proposición muestra que las matrices unitarias conservan tanto los ángulos entre los vectores transformados por la matriz como sus longitudes, y que toda matriz en $\mathbb{C}^{n \times n}$ que conserve los ángulos entre pares de vectores transformados y sus longitudes es una matriz unitaria:

Proposición 5.3.1 — Conservación del producto escalar entre vectores transformados por matrices unitarias

Sea la matriz $U \in \mathbb{C}^{n \times n}$, las siguientes afirmaciones son equivalentes:

1. U es unitaria.
2. U conserva el producto escalar usual.

$$\forall \bar{x}, \bar{y} \in \mathbb{C}^n, \langle U\bar{x}, U\bar{y} \rangle = \langle \bar{x}, \bar{y} \rangle \quad (5.146)$$

3. U conserva la norma canónica de los vectores complejos.

$$\forall \bar{x} \in \mathbb{C}^n, \|U\bar{x}\| = \|\bar{x}\| \quad (5.147)$$

Observación 5.3.2 El concepto de unitariedad de una matriz cuadrada puede generalizarse a los isomorfismos, Definición 3.1.7, de la siguiente manera: Se dice que un isomorfismo f entre dos espacios euclidianos E y F sobre el cuerpo \mathbb{C} es una aplicación unitaria si conserva tanto las normas canónicas como los productos internos entre los elementos transformados. Es decir, si cumple $\|f(x) - f(y)\| = \|x - y\| \forall x, y \in E$. A estos operadores se les suele llamar “isometrías”.

Si una aplicación entre espacios en \mathbb{C} es una isometría, sus matrices asociadas son unitarias, y viceversa.

Demostración:

Para comprobar que las afirmaciones [(1) ... (3)] son equivalentes se demostrará: 1) Que toda matriz unitaria conserva el producto escalar usual. 2) Que toda matriz que conserva el producto escalar usual conserva también la norma canónica. 3) Que toda matriz que conserva la norma canónica de los vectores complejos es unitaria.

- (1) \Rightarrow (2): Sea $U \in \mathbb{C}^{n \times n}$ una matriz unitaria, por la Definición 5.1.7 del producto escalar usual, se tiene, $\forall \bar{x}, \bar{y} \in \mathbb{C}^n$:

$$\langle U\bar{x}, U\bar{y} \rangle = (U\bar{y})^h U\bar{x} \stackrel{(*)}{=} \bar{y}^h U^h U\bar{x} = \bar{y}^h I\bar{x} = \bar{y}^h \bar{x} = \langle \bar{x}, \bar{y} \rangle \quad (5.148)$$

Aplicando en (*) la propiedad 4 de la Proposición 2.2.3.

- (2) \Rightarrow (3): Si en el razonamiento anterior se fija $\bar{x} = \bar{y}$, y teniendo en cuenta que, por la Definición 5.1.8, la norma canónica de un vector complejo es la raíz cuadrada de su producto escalar consigo mismo, se tiene:

$$\|U\bar{x}\| = \sqrt{\langle U\bar{x}, U\bar{x} \rangle} = \sqrt{\langle \bar{x}, \bar{x} \rangle} = \|\bar{x}\| \quad (5.149)$$

- (3) \Rightarrow (1): Supóngase que $U \in \mathbb{C}^{n \times n}$ es una matriz que cumple $\|U\bar{x}\| = \|\bar{x}\|$ para todo $\bar{x} \in \mathbb{C}$. Por la definición de la norma canónica de un vector complejo, Definición 5.1.8:

$$\|U\bar{x}\| = \sqrt{\bar{x}^h U^h U\bar{x}} = \sqrt{\bar{x}^h \bar{x}} = \|\bar{x}\| \quad (5.150)$$

La identidad $\bar{x}^h U^h U\bar{x} = \bar{x}^h \bar{x}$ implica que $U^h U = I$, por lo que U es una matriz unitaria. □

Proposición 5.3.2 – Determinante de una matriz unitaria

El determinante de una matriz $U \in \mathbb{C}^{n \times n}$ unitaria tiene módulo uno:

$$|\det(U)| = 1$$

Demostración:

Aplicando la Definición 5.3.1 de matriz unitaria, las propiedades 1 y 3 de la Proposición 2.3.1, y la Definición C.6.1 de módulo de un número complejo, se obtiene:

$$\begin{aligned} 1 &= \det(I) = \det(U^h U) = \det(U^h) \det(U) = \\ &= \det(U^t)^* \det(U) = \det(U)^* \det(U) = \\ &= |\det(U)|^2 \Rightarrow |\det(U)| = 1 \end{aligned} \quad (5.151)$$

□

Como consecuencia de su unitariedad (de ahí su nombre) de sus determinantes y de las propiedades de la Proposición 5.3.1, estas matrices transforman los puntos de \mathbb{K}^n sin deformar las figuras geométricas, ni agrandarlas ni disminuir su tamaño. A este tipo de transformaciones se las llama “unitarias” en \mathbb{C} y “ortogonales” en \mathbb{R}^n .

Ejemplo 5.3.2 Las tres matrices del Ejemplo 5.3.1 tienen determinante unitario. Puede comprobarse † que:

$$\det(U_1) = -1 \quad (5.152)$$

$$\det(U_2) = -i \quad (5.153)$$

$$\det(U_3) = 1 \quad (5.154)$$

$$\det(U_4) = -i \quad (5.155)$$

Ejercicio 5.3.1 † Calcular estos determinantes.

Observación 5.3.3 Nótese que todas las matrices ortogonales en $\mathbb{R}^{n \times n}$ son matrices unitarias en $\mathbb{C}^{n \times n}$. De hecho, las matrices ortogonales pueden definirse como aquellas matrices en $\mathbb{C}^{n \times n}$ que son unitarias y tienen todos sus coeficientes reales.

5.3.3 Matrices ortogonales**Definición 5.3.2 – Matrices ortogonales**

Una matriz $Q \in \mathbb{R}^{n \times n}$ se dice ortogonal si cumple:

$$Q^t Q = Q Q^t = I \quad (5.156)$$

O, equivalentemente, si su inversa coincide con su transpuesta:

$$Q^t = Q^{-1} \quad (5.157)$$

Ejemplo 5.3.3 Las siguientes matrices de orden dos son ortogonales, como se puede comprobar:

$$Q_1 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad (5.158)$$

$$Q_2 = \begin{pmatrix} \frac{-\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix} \quad (5.159)$$

$$Q_3 = \begin{pmatrix} \frac{1}{2} & \frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & \frac{-1}{2} \end{pmatrix} \quad (5.160)$$

La siguiente matriz de orden tres es también ortogonal:

$$Q_4 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ 0 & \frac{\sqrt{2}}{2} & \frac{-\sqrt{2}}{2} \end{pmatrix} \quad (5.161)$$

Como consecuencia directa de la Definición 5.3.2, las filas y las columnas de una matriz ortogonal de orden n forman bases ortonormales de \mathbb{R}^n respecto al producto escalar de vectores reales, Definición 5.1.3, y la norma canónica o *euclidiana* que este define, Definición 5.1.4.

La siguiente proposición muestra que las matrices ortogonales, al igual que las unitarias, conservan tanto los ángulos entre los vectores transformados por la matriz como las longitudes de estos. Además, toda matriz en $\mathbb{R}^{n \times n}$ que conserve tanto los ángulos entre pares de vectores transformados como sus longitudes es una matriz ortogonal.

Proposición 5.3.3 – Conservación del producto escalar entre vectores transformados por matrices ortogonales

Dada una matriz $Q \in \mathbb{R}^{n \times n}$, las siguientes afirmaciones son equivalentes:

1. Q es ortogonal.
2. Q conserva el producto escalar.

$$\forall \bar{x}, \bar{y} \in \mathbb{R}^n, \langle Q\bar{x}, Q\bar{y} \rangle = \langle \bar{x}, \bar{y} \rangle \quad (5.162)$$

3. Q conserva la norma *euclidiana* de los vectores reales.

$$\forall \bar{x} \in \mathbb{R}^n, \|Q\bar{x}\| = \|\bar{x}\| \quad (5.163)$$

Demostración:

Para comprobar que las afirmaciones [(1) \dots (3)] son equivalentes se demostrará 1) Que toda matriz ortogonal conserva el producto escalar real. 2) Que toda matriz que conserva el producto escalar real conserva también la norma *euclidiana*. 3) Que toda matriz que conserva la norma *euclidiana* de los vectores reales es ortogonal.

Observación 5.3.4 Al igual que las matrices unitarias, Observación 5.3.2, el concepto de matriz ortogonal también se puede extender a otros operadores: se dice que un isomorfismo f , Definición 3.1.7, entre dos espacios *euclidianos* E y F sobre el cuerpo \mathbb{R} es una aplicación ortogonal o una “isometría real” si conserva tanto las normas canónicas como los productos internos entre los elementos transformados. Es decir, si cumple $\|f(x) - f(y)\| = \|x - y\| \forall x, y \in E$. Si una aplicación entre espacios en \mathbb{R} es una isometría, sus matrices asociadas son ortogonales, y viceversa.

- (1) \Rightarrow (2): Sea $Q \in \mathbb{C}^{n \times n}$ una matriz ortogonal, por la definición del producto escalar de vectores reales, Definición 5.1.3, se tiene, $\forall \bar{x}, \bar{y} \in \mathbb{R}^n$:

$$\langle Q\bar{x}, Q\bar{y} \rangle = (Q\bar{x})^t Q\bar{y} \stackrel{(*)}{=} \bar{x}^t Q^t Q\bar{y} = \bar{x}^t I \bar{y} = \bar{x}^t \bar{y} = \langle \bar{x}, \bar{y} \rangle \quad (5.164)$$

Aplicando en (*) la propiedad 4 del producto matricial, Proposición 2.2.3.

- (2) \Rightarrow (3): Si en el razonamiento anterior se fija $\bar{x} = \bar{y}$, y teniendo en cuenta que la norma *euclidiana* de un vector real es la raíz cuadrada de su producto escalar consigo mismo, Definición 5.1.4, se tiene:

$$\|Q\bar{x}\| = \sqrt{\langle Q\bar{x}, Q\bar{x} \rangle} = \sqrt{\langle \bar{x}, \bar{x} \rangle} = \|\bar{x}\| \quad (5.165)$$

- (3) \Rightarrow (1): Supóngase que $Q \in \mathbb{R}^{n \times n}$ es una matriz que cumple $\|Q\bar{x}\| = \|\bar{x}\|$ para todo $\bar{x} \in \mathbb{R}^n$. Por la definición de la norma *euclidiana* de un vector real, Definición 5.1.4:

$$\|Q\bar{x}\| = \sqrt{\bar{x}^t Q^t Q \bar{x}} = \sqrt{\bar{x}^t \bar{x}} = \|\bar{x}\| \quad (5.166)$$

La identidad $\bar{x}^t Q^t Q \bar{x} = \bar{x}^t \bar{x}$ implica que $Q^t Q = I$, por lo que Q es una matriz ortogonal.

□

La Proposición 5.3.2 establece que el determinante de una matriz unitaria es siempre unitaria. Teniendo en cuenta que las matrices ortogonales pueden definirse como aquellas matrices unitarias que tienen todos sus coeficientes reales, Observación 5.3.3, esto implica que el determinante de una matriz ortogonal tiene dos únicos valores posibles: 1 y -1 .

Proposición 5.3.4 – Determinante de una matriz ortogonal

El determinante de Q sólo puede valer 1 ó -1 .

Demostración:

Aplicando la Definición 5.3.2 de matriz ortogonal, la propiedad 1 de la Proposición 2.3.1, y la propiedad 3 de la Proposición 2.3.1:

$$\begin{aligned} 1 &= \det(I) = \det(Q^t Q) = \det(Q^t) \det(Q) = \det(Q) \det(Q) = \\ &= \det(Q)^2 \Rightarrow \det(Q) = \pm 1 \end{aligned} \quad (5.167)$$

□

Ejemplo 5.3.4 Puede comprobarse † que las tres matrices del Ejemplo 5.3.3 tienen determinante unitario:

$$\det(Q_1) = -1 \quad (5.168)$$

$$\det(Q_2) = 1 \quad (5.169)$$

$$\det(Q_3) = -1 \quad (5.170)$$

$$\det(Q_4) = -1 \quad (5.171)$$

Ejercicio 5.3.2 † Calcular estos determinantes.

5.3.4 Matrices de giro

De acuerdo con la Proposición 5.3.3, las matrices ortogonales mantienen invariante la norma de los vectores transformados y los ángulos relativos entre estos. Si además su determinante es igual a uno, estas matrices tienen una interpretación geométrica muy interesante: el producto $Q\bar{x} \forall \bar{x} \in \mathbb{R}^{n \times 1}$, siendo Q una matriz ortogonal, genera un vector que es el resultado de girar \bar{x} un cierto ángulo alrededor de un cierto eje de giro en el espacio \mathbb{R}^n . Estas son las llamadas “matrices de giro en \mathbb{R}^n ”. A continuación, se definirán las matrices de giro y se mostrará cómo calcular matrices de giro en \mathbb{R}^2 y en \mathbb{R}^3 dados un ángulo y un eje de giro.

Definición 5.3.3 – Matriz de giro

Se dice que $G \in \mathbb{R}^{n \times n}$ es una matriz de giro si es ortogonal y su determinante es igual a uno.

5.3.5 Matrices de giro en \mathbb{R}^2

Todas las matrices de giro de orden dos tienen la forma:

$$G = \begin{pmatrix} \cos \varphi & -\operatorname{sen} \varphi \\ \operatorname{sen} \varphi & \cos \varphi \end{pmatrix} \quad (5.172)$$

con $\varphi \in [0, 2\pi)$.

El producto matricial de G por un vector $\bar{x} \in \mathbb{R}^{2 \times 1}$ efectúa un giro del vector en sentido contrario a las agujas del reloj con centro en el origen de coordenadas del plano *euclidiano* \mathbb{R}^2 y con un ángulo de giro igual a $|\varphi|$. Al ser G ortogonal, conserva los ángulos relativos entre los vectores que transforma, así como sus longitudes (Figura 5.5).

Ejemplo 5.3.5 Sustituyendo $\varphi_1 = \pi/4$ en la fórmula (5.172) se obtiene la matriz:

$$G_1 = \begin{pmatrix} \sqrt{2}/2 & -\sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \end{pmatrix} \quad (5.173)$$

Que es una matriz que transforma vectores de \mathbb{R}^2 girándolos sobre el plano *euclidiano* un ángulo igual a $\pi/4$ rad en sentido antihorario con centro de giro en el origen de coordenadas.

Si se sustituye $\varphi_2 = \pi/6$ y $\varphi_3 = 2\pi/3$ se obtienen las matrices:

$$G_2 = \begin{pmatrix} \sqrt{3}/2 & -1/2 \\ 1/2 & \sqrt{3}/2 \end{pmatrix} \quad (5.174)$$

$$G_3 = \begin{pmatrix} -1/2 & -\sqrt{3}/2 \\ \sqrt{3}/2 & -1/2 \end{pmatrix} \quad (5.175)$$

El transformado por G_2 del vector $\bar{x} = (1, 0)^t$ es igual al vector $(\sqrt{3}/2, 1/2)^t$, vector unitario cuyo ángulo respecto de \bar{x} medido en sentido antihorario es de $\pi/6$ rad.

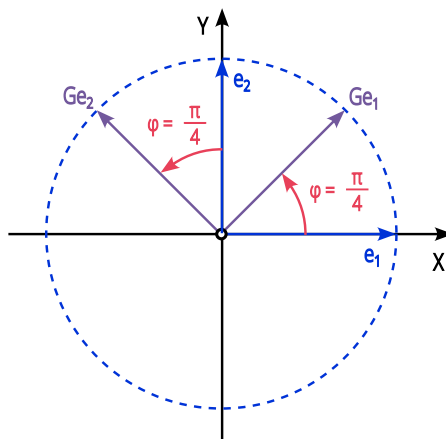


Figura 5.5: La matriz de giro en \mathbb{R}^2 con $\varphi = \pi/4$ transforma vectores en \mathbb{R}^2 girándolos $\pi/4$ rad en sentido contrario a las agujas del reloj. El eje de giro es perpendicular al plano e intersecta el origen de coordenadas. Se muestra un ejemplo con los vectores canónicos \bar{e}_1 y \bar{e}_2 .

5.3.6 Matrices de giro en \mathbb{R}^3

Dada una recta en \mathbb{R}^3 que corta el origen de coordenadas de este espacio y cuya dirección viene dada por el vector director unitario \bar{u} , y dado un ángulo $\varphi \in [0, 2\pi)$, la aplicación G que asocia a cada vector de \mathbb{R}^3 su girado un ángulo φ alrededor de la recta o eje de giro $L[(\bar{u})]$ es una transformación lineal cuyas matrices asociadas son matrices de giro en \mathbb{R}^3 (Figura 5.6).

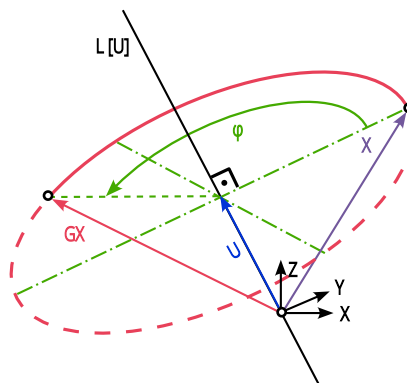


Figura 5.6: Al girar un vector cualquiera en \mathbb{R}^3 un ángulo de $|\varphi|$ rad alrededor de un eje de giro cuyo vector director es \bar{u} se dibuja un semicono, o un cono si $|\varphi| \geq 2\pi$, que tiene su vértice coincidente con el origen de coordenadas y su generatriz coincidente con el vector \bar{x} . El sentido de giro es antihorario respecto a la dirección de \bar{u} , siguiendo la regla de la mano derecha.

Observación 5.3.5 † El producto vectorial, o cruzado, de dos vectores $\bar{x} = (x_1, x_2, x_3), \bar{y} = (y_1, y_2, y_3)$ en \mathbb{R}^3 se define como:

$$\bar{x} \times \bar{y} = \det \begin{pmatrix} \bar{e}_1 & \bar{e}_2 & \bar{e}_3 \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{pmatrix} \quad (5.176)$$

Donde $\bar{e}_1, \bar{e}_2, \bar{e}_3$ son los tres vectores canónicos de \mathbb{R}^3 . El vector resultante es perpendicular al plano que contiene los dos vectores multiplicados, y si estos son unitarios y ortogonales, el vector resultante también es unitario.

$$\bar{w} = \bar{u} \times \bar{v}, \quad (5.177)$$

donde el símbolo \times denota el producto vectorial †. Estos tres vectores columna conforman la siguiente matriz de cambio de la base canónica a la base \mathcal{B} :

$$Q = (\bar{u}|\bar{v}|\bar{w}) \in \mathbb{R}^{n \times n} \quad (5.178)$$

La matriz de giro alrededor del vector \bar{u} expresada en la base $(\bar{u}, \bar{v}, \bar{w})$ puede interpretarse como una matriz de giro alrededor del eje local z' (cuya

dirección viene dada por \bar{u}) sobre el plano definido por ejes cartesianos locales x' e y' . Se trata de la matriz:

$$G^* = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \varphi & -\operatorname{sen} \varphi \\ 0 & \operatorname{sen} \varphi & \cos \varphi \end{pmatrix} \quad (5.179)$$

Efectuando un cambio de base de la matriz G^* a la base canónica, Proposición 3.3.1, se obtiene la expresión genérica de la matriz de giro alrededor de \bar{u} respecto de la base canónica de \mathbb{R}^3 :

$$G = (\bar{u}|\bar{v}|\bar{w}) \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \varphi & -\operatorname{sen} \varphi \\ 0 & \operatorname{sen} \varphi & \cos \varphi \end{pmatrix} (\bar{u}|\bar{v}|\bar{w})^t \quad (5.180)$$

Ejemplo 5.3.6 Dado el vector unitario $\bar{u} = (1/\sqrt{2}, 1/\sqrt{2}, 0)^t \in \mathbb{R}^3$, dos vectores también unitarios que forman junto con \bar{u} una base ortonormal de \mathbb{R}^3 y que además cumplen $\bar{w} = \bar{u} \times \bar{v}$ son $\bar{v} = (1/\sqrt{2}, -1/\sqrt{2}, 0)^t$ y $\bar{w} = (0, 0, 1)^t$. Con estos tres vectores se construye la siguiente matriz de cambio de base:

$$Q = (\bar{u}|\bar{v}|\bar{w}) = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (5.181)$$

Sustituyendo $\varphi = \pi/4$ en la identidad (5.180) se tiene:

$$G = Q \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} Q^t = \begin{pmatrix} \frac{1+\sqrt{2}}{2\sqrt{2}} & \frac{\sqrt{2}-1}{2\sqrt{2}} & \frac{-1}{2} \\ \frac{\sqrt{2}-1}{2\sqrt{2}} & \frac{1}{2} & \frac{1+\sqrt{2}}{2\sqrt{2}} \\ \frac{1}{2} & \frac{-1}{2} & \frac{\sqrt{2}}{2} \end{pmatrix} \quad (5.182)$$

Que es la matriz de giro en \mathbb{R}^3 respecto de la base canónica de \mathbb{R}^3 correspondiente a un giro $\pi/4$ rad en sentido antihorario alrededor del eje de giro definido por el vector $\bar{u} = (1/\sqrt{2}, 1/\sqrt{2}, 0)^t$.

5.4 Ortogonalización de *Gram-Schmidt*

5.4.1 Introducción

Las bases ortogonales y ortonormales, Definición 5.2.4, tienen propiedades que las hacen muy deseables de entre las infinitas bases que un espacio o subespacio vectorial tiene. Una ventaja de estas bases es su estabilidad, que las hace muy aptas para la computación numérica. Los motivos de esta estabilidad son analizables con métodos del álgebra lineal, pero su estudio se deja para textos especializados en cálculo numérico. Otra ventaja de disponer de una base ortogonal $\mathcal{B} = (u_1, \dots, u_n)$ de un espacio o subespacio V dotado de un producto interno $\langle \cdot, \cdot \rangle$ es que descomponer un elemento cualquiera de V en combinación de los de \mathcal{B} es tan simple como aplicar la fórmula:

$$x = \sum_{i=1}^n \lambda_i u_i, \quad \lambda_i = \frac{\langle x, u_i \rangle}{\langle u_i, u_i \rangle} \quad (5.183)$$

Donde cada término de la suma es la proyección ortogonal de x sobre el subespacio en dimensión uno engendrado por el i -ésimo elemento de la base ortogonal \mathcal{B} .

Afortunadamente, todo espacio o subespacio admite bases ortogonales y ortonormales, y toda base de un espacio o subespacio puede convertirse en una base ortonormal del mismo. En este apartado se verá cómo.

5.4.2 El proceso de *Gram-Schmidt*

Dada una base $\mathcal{B} = (u_1, \dots, u_n)$ no ortonormal de un espacio o subespacio vectorial $V = L[\mathcal{B}]$ de dimensión n dotado de un producto interno $\langle \cdot, \cdot \rangle$ y de la norma canónica $\|\cdot\| = \sqrt{\langle \cdot, \cdot \rangle}$, se comienza normalizando el primer elemento de la base respecto de la norma canónica:

$$q_1 = \frac{u_1}{\|u_1\|} \quad (5.184)$$

La proyección ortogonal de u_2 sobre q_1 es $\langle u_2, q_1 \rangle q_1$. Si a u_2 se le sustrae su proyección ortogonal sobre q_1 se obtiene:

$$\tilde{q}_2 = u_2 - \langle u_2, q_1 \rangle q_1 \quad (5.185)$$

\tilde{q}_2 es ortogonal a q_1 :

$$\begin{aligned} \langle \tilde{q}_2, q_1 \rangle &= \langle u_2 - \langle u_2, q_1 \rangle q_1, q_1 \rangle = \langle u_2, q_1 \rangle - \langle u_2, q_1 \rangle \langle q_1, q_1 \rangle = \\ &= \langle u_2, q_1 \rangle - \langle u_2, q_1 \rangle = \\ &= 0 \end{aligned} \quad (5.186)$$

Y si se normaliza \tilde{q}_2 se obtiene el elemento:

$$q_2 = \frac{\tilde{q}_2}{\|\tilde{q}_2\|} \quad (5.187)$$

Que engendra, junto con q_1 , un subespacio en dimensión dos que coincide con el generado por los dos primeros elementos de \mathcal{B} (Figura 5.7)

pues, de acuerdo con el Teorema 1.3.1, si a un elemento de una base se multiplica por un coeficiente no nulo o se combina linealmente con otros elementos de la base, se obtiene una base diferente del mismo espacio o subespacio:

$$L[(q_1, q_2)] = L[(u_1, u_2)] \quad (5.188)$$

Tómese ahora el tercer elemento de la base \mathcal{B} , denotado u_3 . Restándole sus proyecciones ortogonales sobre los dos elementos de la nueva base previamente calculados, q_1 y q_2 , se obtiene:

$$\tilde{q}_3 = u_3 - \langle u_3, q_1 \rangle q_1 - \langle u_3, q_2 \rangle q_2 \quad (5.189)$$

\tilde{q}_3 es ortogonal a q_1 y q_2 :

$$\begin{aligned} \langle \tilde{q}_3, q_1 \rangle &= \langle u_3 - \langle u_3, q_1 \rangle q_1 - \langle u_3, q_2 \rangle q_2, q_1 \rangle = \\ &= \langle u_3, q_1 \rangle - \langle u_3, q_1 \rangle \langle q_1, q_1 \rangle - \langle u_3, q_2 \rangle \langle q_2, q_1 \rangle = \\ &= \langle u_3, q_1 \rangle - \langle u_3, q_1 \rangle - 0 = \\ &= 0 \end{aligned} \quad (5.190)$$

$$\begin{aligned} \langle \tilde{q}_3, q_2 \rangle &= \langle u_3 - \langle u_3, q_1 \rangle q_1 - \langle u_3, q_2 \rangle q_2, q_2 \rangle = \\ &= \langle u_3, q_2 \rangle - \langle u_3, q_1 \rangle \langle q_1, q_2 \rangle - \langle u_3, q_2 \rangle \langle q_2, q_2 \rangle = \\ &= \langle u_3, q_2 \rangle - 0 - \langle u_3, q_2 \rangle = \\ &= 0 \end{aligned} \quad (5.191)$$

Normalizando \tilde{q}_3 se tiene un tercer elemento que es, además de ser unitario y ortogonal a q_1 y q_2 , genera junto con estos un subespacio en dimensión tres que coincide con la clausura lineal de los tres primeros elementos de \mathcal{B} :

$$L[(q_1, q_2, q_3)] = L[(u_1, u_2, u_3)] \quad (5.192)$$

Si se sigue este proceso iterativamente desde $i = 3$ hasta n se acabará por obtener una base ortonormal (q_1, \dots, q_n) del espacio o subespacio V , justo como se deseaba:

$$L[(q_1, \dots, q_n)] = L[(u_1, \dots, u_n)] \quad (5.193)$$

A este algoritmo se le conoce como el “proceso de *Gram-Schmidt*”:

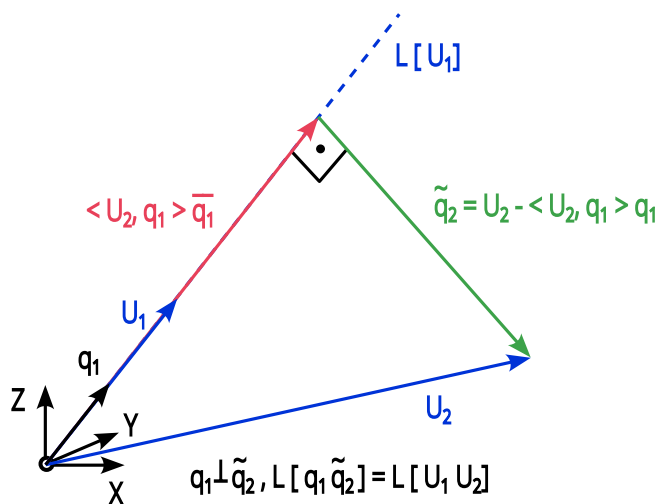
Teorema 5.4.1 – Proceso de *Gram-Schmidt*

Dada una base $\mathcal{B} = (u_1, u_2, \dots, u_n)$ de un espacio o subespacio V de dimensión n dotado de un producto interno $\langle \cdot, \cdot \rangle$ y de la norma canónica $\|\cdot\| = \sqrt{\langle \cdot, \cdot \rangle}$, la familia ortonormal ordenada (q_1, q_2, \dots, q_n) cuyos elementos se calculan aplicando el siguiente proceso iterativo:

$$\begin{aligned} q_1 &= u_1 / \|u_1\| \\ \tilde{q}_2 &= u_2 - \langle u_2, q_1 \rangle q_1, \quad q_2 = \tilde{q}_2 / \|\tilde{q}_2\| \\ &\vdots \\ \tilde{q}_n &= u_n - \sum_{i=1}^{n-1} \langle u_n, q_i \rangle q_i, \quad q_n = \tilde{q}_n / \|\tilde{q}_n\| \end{aligned} \quad (5.194)$$

Tiene la misma clausura lineal que \mathcal{B} , por lo que constituye una base ortonormal de V .

Figura 5.7: Al restarle a \bar{u}_2 su proyección ortogonal sobre \bar{u}_1 se obtiene un vector denotado \tilde{q}_2 que es ortogonal a \bar{u}_1 y que está contenido en el plano que definen \bar{u}_1 y \bar{u}_2 (y por tanto, en el subespacio $L[(\bar{u}_1, \bar{u}_2)]$). Normalizando \bar{u}_1 y \tilde{q}_2 se obtienen dos vectores ortogonales y unitarios cuya clausura lineal coincide con $L[(\bar{u}_1, \bar{u}_2)]$.



Ejemplo 5.4.1 La base de \mathbb{R}^3 formada por los vectores:

$$\bar{u}_1 = (2, -1, 0)^t \tag{5.195}$$

$$\bar{u}_2 = (3, 1, -1)^t \tag{5.196}$$

$$\bar{u}_3 = (25, 0, 1)^t \tag{5.197}$$

Puede transformarse en una base ortonormal de \mathbb{R}^3 aplicando el proceso de *Gram-Schmidt*, Teorema 5.4.1. En primer lugar, se normaliza \bar{u}_1 , obteniendo el primer vector de la nueva base:

$$\bar{q}_1 = \bar{u}_1 / \|\bar{u}_1\| = 1/\sqrt{5} \cdot (2, -1, 0)^t \tag{5.198}$$

A \bar{u}_2 se le sustrae su proyección ortogonal sobre \bar{q}_1 :

$$\begin{aligned} \tilde{q}_2 &= \bar{u}_2 - \langle \bar{u}_2, \bar{q}_1 \rangle \bar{q}_1 = \\ &= (3, 1, -1)^t - \frac{1}{5} \cdot (2, -1, 0)^t = (1, 2, -1)^t \end{aligned} \tag{5.199}$$

Y el vector resultante se normaliza, dando lugar a \bar{q}_2 :

$$\bar{q}_2 = \tilde{q}_2 / \|\tilde{q}_2\| = 1/\sqrt{6} \cdot (1, 2, -1)^t \tag{5.200}$$

Por último, se le resta a \bar{u}_3 sus proyecciones ortogonales sobre \bar{q}_1 y \bar{q}_2 :

$$\begin{aligned} \tilde{q}_3 &= \bar{u}_3 - \langle \bar{u}_3, \bar{q}_1 \rangle \bar{q}_1 - \langle \bar{u}_3, \bar{q}_2 \rangle \bar{q}_2 = \\ &= (25, 0, 1)^t - \frac{50}{5} \cdot (2, -1, 0)^t - \frac{24}{6} \cdot (1, 2, -1)^t = \\ &= (1, 2, 5)^t \end{aligned} \tag{5.201}$$

Y se normaliza el resultado, obteniendo así el tercer vector de la nueva base:

$$\bar{q}_3 = \tilde{q}_3 / \|\tilde{q}_3\| = 1/\sqrt{30} \cdot (1, 2, 5)^t \tag{5.202}$$

La familia ordenada $(\bar{q}_1, \bar{q}_2, \bar{q}_3) \subset \mathbb{R}^3$ es la base ortonormal buscada.

5.4.3 Una aplicación del método de *Gram-Schmidt*: la factorización QR

A continuación se utilizará el proceso de *Gram-Schmidt* para demostrar que toda matriz cuadrada de rango completo puede factorizarse en el producto de una matriz unitaria, u ortogonal si es real, por una matriz triangular superior.

Si $A \in \mathbb{K}^{n \times n}$ tiene rango completo, sus columnas forman una base de \mathbb{K}^n , $\mathcal{B} = (\bar{u}_1, \bar{u}_2, \dots, \bar{u}_n)$. Dicha base puede ortogonalizarse por el proceso de *Gram-Schmidt*, obteniéndose una base ortonormal de \mathbb{K}^n denominada $\mathcal{B}_o = (\bar{q}_1, \bar{q}_2, \dots, \bar{q}_n)$. Las ecuaciones de (5.194) muestran que los vectores de \mathcal{B} pueden ponerse en función de los de \mathcal{B}_o :

$$\begin{aligned} & \|\bar{u}_1\| \bar{q}_1 = \bar{u}_1 \\ \langle \bar{u}_2, \bar{q}_1 \rangle \bar{q}_1 + \|\bar{u}_2 - \langle \bar{u}_2, \bar{q}_1 \rangle \bar{q}_1\| \bar{q}_2 &= \bar{u}_2 \\ & \vdots \\ \sum_{i=1}^{n-1} \langle \bar{u}_n, \bar{q}_i \rangle \bar{q}_i + \|\bar{u}_n - \sum_{i=1}^{n-1} \langle \bar{u}_n, \bar{q}_i \rangle \bar{q}_i\| \bar{q}_n &= \bar{u}_n \end{aligned} \tag{5.203}$$

La propiedad 1 de la Proposición 2.2.5 nos dice que cada columna del producto matricial entre dos matrices equivale a una combinación lineal de las columnas de la primera matriz. Aplicándola, las n ecuaciones de (5.203) pueden escribirse como una única ecuación matricial:

$$QR = (\bar{q}_1 | \bar{q}_2 | \dots | \bar{q}_n)R = (\bar{u}_1 | \bar{u}_2 | \dots | \bar{u}_n) = A \tag{5.204}$$

Donde:

$$R = \begin{pmatrix} \|\bar{u}_1\| & \langle \bar{u}_2, \bar{q}_1 \rangle & \dots & \langle \bar{u}_1, \bar{q}_n \rangle \\ 0 & \|\bar{u}_2 - \langle \bar{u}_2, \bar{q}_1 \rangle \bar{q}_1\| & \dots & \langle \bar{u}_2, \bar{q}_n \rangle \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \|\bar{u}_n - \sum_{i=1}^{n-1} \langle \bar{u}_n, \bar{q}_i \rangle \bar{q}_i\| \end{pmatrix} \tag{5.205}$$

Se demuestra así el teorema que rige la factorización QR para matrices cuadradas de rango completo:

Teorema 5.4.2 – Factorización QR de matrices cuadradas de rango completo

Toda matriz $A \in \mathbb{K}^{n \times n}$ de rango completo puede descomponerse en el producto de una matriz unitaria (u ortogonal, si A es real) por una matriz triangular superior cuyos coeficientes en la diagonal principal son escalares reales positivos:

$$A = QR \tag{5.206}$$

Ambas matrices son de orden n .

Observación 5.4.1 La factorización QR de una matriz no tiene ninguna relación con los códigos QR o *Quick-Response*.

La demostración anterior muestra además cómo llevarla a cabo. Otro método para calcular R a partir de A y Q es despejarla del producto QR: $A = QR \Rightarrow R = Q^t A$.

Ejemplo 5.4.2 La matriz:

$$A = \begin{pmatrix} 2 & 3 & 25 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{pmatrix} \in \mathbb{R}^{3 \times 3} \quad (5.207)$$

Es de rango completo, por lo que se puede factorizar como el producto de una matriz ortogonal por una matriz triangular superior. Las tres columnas de A forman la base $(\bar{u}_1, \bar{u}_2, \bar{u}_3)$ de \mathbb{R}^3 . Esta base se había ortogonalizado en el Ejemplo 5.4.1, obteniendo una base ortonormal de \mathbb{R}^3 formada por los vectores:

$$\bar{q}_1 = \frac{1}{\sqrt{5}} \cdot (2, -1, 0)^t \quad (5.208)$$

$$\bar{q}_2 = \frac{1}{\sqrt{6}} \cdot (1, 2, -1)^t \quad (5.209)$$

$$\bar{q}_3 = \frac{1}{\sqrt{30}} \cdot (1, 2, 5)^t \quad (5.210)$$

Estos vectores conforman cada una de las columnas de Q :

$$Q = \begin{pmatrix} 2/\sqrt{5} & 1/\sqrt{6} & 1/\sqrt{30} \\ -1/\sqrt{5} & 2/\sqrt{6} & 2/\sqrt{30} \\ 0 & -1/\sqrt{6} & 5/\sqrt{30} \end{pmatrix} \in \mathbb{R}^{3 \times 3} \quad (5.211)$$

Los coeficientes de la matriz triangular superior R pueden despejarse de las fórmulas empleadas en el proceso de *Gram-Schmidt*, ecuaciones [(5.198) ... (5.202)], o bien puede calcularse la matriz R despejándola de la identidad $A = QR$, teniendo en cuenta que Q es ortogonal:

$$A = QR \Rightarrow R = Q^t A = \begin{pmatrix} 5/\sqrt{5} & 5/\sqrt{5} & 50/\sqrt{5} \\ 0 & 6/\sqrt{6} & 24/\sqrt{6} \\ 0 & 0 & 30/\sqrt{30} \end{pmatrix} \quad (5.212)$$

La descomposición QR puede aplicarse sobre cualquier matriz, sea esta cuadrada o rectangular, con la única condición de que su número de columnas sea menor o igual que su número de filas:

Teorema 5.4.3 – Factorización QR de matrices en $\mathbb{K}^{m \times n}$ con $n \leq m$

Toda matriz $A \in \mathbb{K}^{m \times n}$ de rango $r \leq n$ y con $n \leq m$ puede descomponerse en el producto de una matriz $Q \in \mathbb{K}^{m \times n}$ de columnas ortogonales y unitarias por una matriz triangular superior R de orden n con coeficientes reales positivos o nulos en la diagonal principal: $A = QR$.

Dicha descomposición puede también llevarse a cabo aplicando el método basado en *Gram-Schmidt* descrito arriba sobre r columnas de la matriz que formen una base de $\text{Im } A$, y después, buscando $n - r$ vectores columna unitarios cualesquiera que formen junto con los anteriores una base ortonormal $(\bar{q}_1, \dots, \bar{q}_m)$ de \mathbb{K}^n . Los $r \leq m$ primeros vectores de esta base formarían una base ortonormal de $\text{Im } A$.

5.5 Teorema de la proyección ortogonal.

Matrices de proyección ortogonal

5.5.1 Introducción

La propiedad $V = H \oplus H^\perp$ que cumple todo subespacio $H \subseteq V$ y su suplementario ortogonal H^\perp implica que todo elemento $v \in V$ se puede descomponer en la suma de un elemento en H más otro en H^\perp : $v = x + y$, $x \in H$, $y \in H^\perp$. Además, por la Proposición 1.4.4, esta descomposición es única: existe una única pareja de elementos $x \in H$ y $y \in H^\perp$ cuya suma es v . En esta sección se demostrará que el elemento $x \in H$ de la descomposición $v = x + y$ es, de todos los elementos de H , el más cercano en distancia *euclidiana* a $v \in V$.

El elemento $x \in H$ no es otro que la proyección ortogonal de v sobre H . En este apartado se definirá una matriz que se llamará “de proyección ortogonal sobre H ” y que permitirá calcular la proyección ortogonal de cualquier vector del espacio *euclidiano* de vectores de n componentes reales o complejas, sobre un subespacio H cualquiera de \mathbb{K}^n .

5.5.2 Teorema de la proyección ortogonal

Sea V un espacio vectorial dotado de un producto interno $\langle \cdot, \cdot \rangle$ y de la norma *euclidiana* o canónica $\| \cdot \| = \sqrt{\langle \cdot, \cdot \rangle}$, por la caracterización de la suma directa, Proposición 1.4.4, para todo $H \subseteq V$ se cumple $V = H \oplus H^\perp \Leftrightarrow v = x + y$, $v \in V$, $x \in H$, $y \in H^\perp$. Además, esta descomposición es única. En cuanto a la distancia, si se escoge cualquier elemento z de H , se cumple la siguiente desigualdad:

$$\begin{aligned} \|v - z\|^2 &= \|(v - x) + (x - z)\|^2 \stackrel{(*)}{=} \\ &\stackrel{(*)}{=} \|v - x\|^2 + \|x - z\|^2 \geq \|v - x\|^2 \end{aligned} \quad (5.213)$$

La igualdad en el paso $(*)$ se da en virtud del teorema de *Pitágoras*, ya que los elementos $v - x$ y $x - z$ son ortogonales entre sí, pues $v - x \in H^\perp$ y $x - z \in H$.

Así, se demuestra que la descomposición $v = x + y$ produce un elemento $x \in H$ denominado proyección ortogonal de v sobre H cuya distancia *euclidiana* a v es mínima respecto a cualquier otro elemento contenido en H (Figura 5.8).

Teorema 5.5.1 – Teorema de la proyección ortogonal

Dado un subespacio H de un espacio vectorial V dotado de un producto interno $\langle \cdot, \cdot \rangle$ y de la norma *euclidiana* o canónica $\| \cdot \| = \sqrt{\langle \cdot, \cdot \rangle}$, cada elemento $v \in V$ tiene unívocamente asociada una pareja de elementos $x \in H$, $y \in H^\perp$ tal que $v = x + y$. Además, para todo $z \in H$ se cumple:

$$d(v, z) = \|v - z\| \geq \|v - x\| = d(v, x) \quad (5.214)$$

Definición 5.5.1 – Proyección ortogonal de un elemento de V sobre un subespacio

El elemento $x \in H$ de la descomposición $v = x + y$, $v \in V$, $x \in H$, $y \in H^\perp$ recibe el nombre de proyección ortogonal de v sobre H . A este se le denotará $P_H v$.

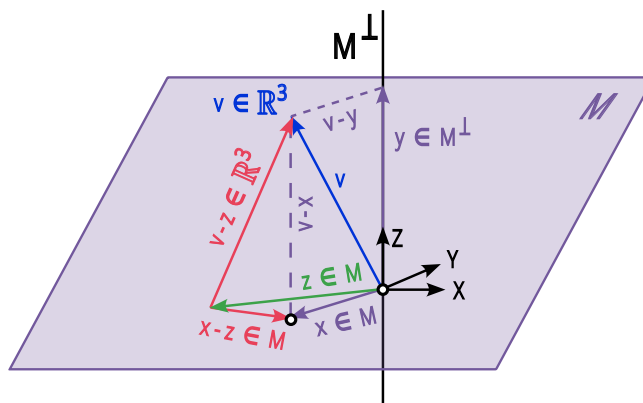
De acuerdo con el Teorema 5.5.1, la distancia *euclidiana* entre $v \in V$ y su proyección ortogonal sobre $H \subseteq V$ es mínima de entre todos los vectores de H . Esta propiedad motiva definir la distancia entre un elemento $v \in V$ y un subespacio $H \subseteq V$ como la distancia entre v y su proyección ortogonal sobre H :

Definición 5.5.2 – Distancia *euclidiana* entre $v \in V$ y un subespacio $H \subseteq V$

Dado un subespacio H de un espacio *euclidiano* V , se define la distancia *euclidiana* entre H y $v \in V$ como

$$d(v, H) = \min_{x \in H} d(v, x) = \|v - P_H v\|$$

Figura 5.8: Representación en el espacio *euclidiano* del Teorema 5.5.1 de la proyección ortogonal y de su demostración, ecuación (5.213). Dados $\bar{v} \in \mathbb{R}^3$ y un subespacio $M \subseteq \mathbb{R}^3$, la descomposición de \bar{v} en suma de un vector en M denotado \bar{x} más otro vector en M^\perp denotado \bar{y} es única. Al vector \bar{x} se le llama "proyección ortogonal de \bar{v} sobre M " y es, de todos los vectores contenidos en M , el de menor distancia *euclidiana* respecto a \bar{v} .



5.5.3 Cálculo de proyecciones ortogonales

Considérense un espacio vectorial *euclidiano* V de dimensión n y un subespacio H de V de dimensión $m \leq n$. Mediante el proceso de *Gram-Schmidt*, Teorema 5.4.1, toda base de H puede transformarse en una base ortogonal de este mismo subespacio que se denotará $\mathcal{B}_M = (u_1, \dots, u_m)$.

La base \mathcal{B}_M obtenida puede completarse añadiéndole, en virtud del Teorema 1.3.4 de completación de bases y el proceso de *Gram-Schmidt*, Teorema 5.4.1, $n - m$ elementos ortogonales entre sí y respecto de los elementos contenidos en \mathcal{B}_M , obteniéndose así la base ortogonal de V siguiente:

$$\mathcal{B} = (u_1, \dots, u_m, u_{m+1}, \dots, u_n) \tag{5.215}$$

Los $n - m$ elementos añadidos formarían una base ortogonal de H^\perp , $\mathcal{B}_{H^\perp} = (u_{m+1}, \dots, u_n)$. Así, por el Teorema 5.5.1 de la proyección ortogonal, todo $v \in V$ puede representarse como combinación lineal de los

elementos de la base ortogonal \mathcal{B} y al hacerlo, se desprende la descomposición de v en suma de un elemento en H más otro elemento en H^\perp :

$$v = x + y = \sum_{i=1}^m \frac{\langle v, u_i \rangle}{\langle u_i, u_i \rangle} u_i + \sum_{i=m+1}^n \frac{\langle v, u_i \rangle}{\langle u_i, u_i \rangle} u_i = P_H v + P_{H^\perp} v \quad (5.216)$$

En consecuencia, la proyección ortogonal de $v \in V$ sobre el subespacio $H \subseteq V$ puede calcularse como la suma de las proyecciones ortogonales de v sobre los elementos de una base ortogonal de H .

Ejemplo 5.5.1 Dada la base ortogonal del espacio euclidiano \mathbb{R}^3 :

$$\mathcal{B} = (\bar{u}_1, \bar{u}_2, \bar{u}_3) = \left(\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} \right) \quad (5.217)$$

Y el vector $\bar{v} = (1, 1, -1)^t \in \mathbb{R}^3$, defínase H como la clausura lineal de los dos primeros vectores de \mathcal{B} , $M = L[(\bar{u}_1, \bar{u}_2)] \subset \mathbb{R}^3$. El vector \bar{u}_3 es ortogonal tanto a \bar{u}_1 como a \bar{u}_2 , por lo que cualquier vector de \mathbb{R}^3 que sea proporcional a \bar{u}_3 , $\lambda \bar{u}_3 \forall \lambda \in \mathbb{R}$, es ortogonal al conjunto de los vectores contenidos en la clausura lineal de \bar{u}_1 y \bar{u}_2 , por lo que $L[\bar{u}_3] = H^\perp$.

Aplicando el teorema de la proyección ortogonal:

$$\bar{v} = \bar{x} + \bar{y}, \bar{x} \in H, \bar{y} \in H^\perp \Leftrightarrow \bar{x} = P_H \bar{v}, \bar{y} = P_{H^\perp} \bar{v} \quad (5.218)$$

La proyección ortogonal de \bar{v} sobre H puede calcularse como:

$$P_H \bar{v} = \frac{\langle \bar{v}, \bar{u}_1 \rangle}{\langle \bar{u}_1, \bar{u}_1 \rangle} \bar{u}_1 + \frac{\langle \bar{v}, \bar{u}_2 \rangle}{\langle \bar{u}_2, \bar{u}_2 \rangle} \bar{u}_2 = (1, 0, 0)^t \quad (5.219)$$

Mientras que la proyección ortogonal de \bar{v} sobre H^\perp queda:

$$P_{H^\perp} \bar{v} = \frac{\langle \bar{v}, \bar{u}_3 \rangle}{\langle \bar{u}_3, \bar{u}_3 \rangle} \bar{u}_3 = (0, 1, -1)^t = \bar{v} - P_H \bar{v} \quad (5.220)$$

5.5.4 Matrices de proyección ortogonal: definición y propiedades

Sea V un espacio de vectores de n componentes en el cuerpo \mathbb{K} dotado del producto escalar usual, Definición 5.1.7, y de la norma euclidiana, Definición 5.1.8, se demostrará que es posible calcular la proyección ortogonal de un vector columna $\bar{v} \in \mathbb{K}^n$ cualquiera sobre un subespacio $H \subseteq \mathbb{K}^n$ pre-multiplicándolo por una matriz cuadrada a la que se denominará matriz de proyección ortogonal sobre H .

Se parte de la fórmula que permite calcular la proyección ortogonal de \bar{v} sobre H , $P_H \bar{v}$, dada una base ortogonal de H :

$$P_H \bar{v} = \sum_{i=1}^m \frac{\langle \bar{v}, \bar{u}_i \rangle}{\langle \bar{u}_i, \bar{u}_i \rangle} \bar{u}_i = \sum_{i=1}^m \frac{\langle \bar{v}, \bar{u}_i \rangle}{\|\bar{u}_i\|^2} \bar{u}_i \quad (5.221)$$

Aplicando las Definiciones 5.1.7 y 5.1.8 del producto escalar usual y de la norma *euclidiana* y reordenando los términos de la ecuación anterior, (5.221), para expresarla como el producto de una matriz cuadrada por un vector columna, se obtiene:

$$\begin{aligned} P_H \bar{v} &= \sum_{i=1}^m \frac{\langle \bar{v}, \bar{u}_i \rangle}{\|\bar{u}_i\|^2} \bar{u}_i = \sum_{i=1}^m \frac{(\bar{v}^t \bar{u}_i^*) \bar{u}_i}{\|\bar{u}_i\|^2} = \\ &= \sum_{i=1}^m \frac{\bar{u}_i (\bar{u}_i^t \bar{v}^*)^*}{\|\bar{u}_i\|^2} = \left(\sum_{i=1}^m \frac{\bar{u}_i \bar{u}_i^h}{\|\bar{u}_i\|^2} \right) \bar{v} = P_M \bar{v} \end{aligned} \quad (5.222)$$

Los productos $\bar{u}_i \bar{u}_i^h$ son todos reales, por lo que P_H es una matriz real conocida como la “matriz de proyección ortogonal sobre H ”.

Definición 5.5.3 – Matriz de proyección ortogonal sobre un subespacio en \mathbb{K}^n

Sea H un subespacio en \mathbb{K}^n de dimensión $m \leq n$, se dice que $P_H \in \mathbb{R}^{n \times n}$ es la matriz de proyección ortogonal sobre H si para todo $\bar{v} \in \mathbb{K}^n$, $P_H \bar{v}$ es la proyección ortogonal de \bar{v} sobre H . Dicha matriz tiene la siguiente fórmula:

$$P_H = \sum_{i=1}^m \frac{\bar{u}_i \bar{u}_i^h}{\|\bar{u}_i\|^2} = \sum_{i=1}^m \frac{\bar{u}_i \bar{u}_i^h}{\bar{u}_i^h \bar{u}_i}$$

Donde $(\bar{u}_1, \dots, \bar{u}_m)$ es una base ortogonal de H .

Ejemplo 5.5.2 Dada la base $\mathcal{B} = L[(\bar{u}_1, \bar{u}_2, \bar{u}_3)]$ de \mathbb{R}^3 del Ejemplo 5.5.1, definase el subespacio $H = L[(\bar{u}_1, \bar{u}_2)] \subset \mathbb{R}^3$. La matriz de proyección ortogonal sobre H puede calcularse aplicando la ecuación (5.5.3) y con esta, puede obtenerse la proyección ortogonal del vector $\bar{v} = (1, 1, -1)^t \in \mathbb{R}^3$ sobre H :

$$\begin{aligned} P_H &= \frac{\bar{u}_1 \bar{u}_1^t}{\|\bar{u}_1\|^2} + \frac{\bar{u}_2 \bar{u}_2^t}{\|\bar{u}_2\|^2} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/2 & 1/2 \\ 0 & 1/2 & 1/2 \end{pmatrix} \Rightarrow \\ &\Rightarrow P_H \bar{v} = (1, 0, 0)^t \end{aligned} \quad (5.223)$$

Por otra parte, la matriz de proyección ortogonal sobre H^\perp y la proyección ortogonal de \bar{v} sobre H^\perp son:

$$\begin{aligned} P_{H^\perp} &= \frac{\bar{u}_3 \bar{u}_3^t}{\|\bar{u}_3\|^2} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1/2 & -1/2 \\ 0 & -1/2 & 1/2 \end{pmatrix} \Rightarrow \\ &\Rightarrow P_{H^\perp} \bar{v} = (0, 1, -1)^t \end{aligned} \quad (5.224)$$

Es fácil comprobar que:

$$\bar{v} = P_H \bar{v} + P_{H^\perp} \bar{v} = (1, 0, 0)^t + (0, 1, -1)^t = (1, 1, -1)^t \quad (5.225)$$

En función de las características del subespacio $H \subseteq \mathbb{K}^n$, pueden proponerse formulaciones alternativas a la de la Definición 5.5.3 para calcular la matriz de proyección ortogonal sobre H . A continuación se enumeran algunas de ellas.

1. Si el subespacio H es de dimensión uno engendrado por un vector $\bar{u} \in \mathbb{K}^n$ no nulo, su matriz de proyección ortogonal asociada queda:

$$P_{L[\bar{u}]} = \frac{\bar{u}\bar{u}^h}{\bar{u}^h\bar{u}} \quad (5.226)$$

2. Si se normalizan los vectores de una base ortogonal de H aplicando $\bar{q}_i = \bar{u}_i/\|\bar{u}_i\|$ para todo $i = 1, \dots, m$, la matriz de proyección ortogonal sobre H queda $\sum_{i=1}^m \bar{q}_i\bar{q}_i^h$. Así, por la propiedad 3 de la Proposición 2.2.5, dada una base ortonormal $(\bar{q}_1, \dots, \bar{q}_m)$ de H , la matriz de proyección ortogonal sobre H puede escribirse como el producto de la matriz de columnas ortogonales y unitarias $(\bar{q}_1|\bar{q}_2|\dots|\bar{q}_m)$ con su adjunta:

$$P_H = (\bar{q}_1|\bar{q}_2|\dots|\bar{q}_m) \begin{pmatrix} \bar{q}_1^h \\ \bar{q}_2^h \\ \vdots \\ \bar{q}_m^h \end{pmatrix} = QQ^h \quad (5.227)$$

Ejemplo 5.5.3 Sea la base ortonormal de \mathbb{R}^3 :

$$\mathcal{B} = (\bar{q}_1, \bar{q}_2, \bar{q}_3) = \left(\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}, \begin{pmatrix} 0 \\ 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix} \right) \quad (5.228)$$

Defínase el subespacio $M \subset \mathbb{R}^3$ y su suplementario ortogonal como:

$$H = L[(\bar{q}_1, \bar{q}_2)] \quad (5.229)$$

$$H^\perp = L[(\bar{q}_3)] \quad (5.230)$$

Aplicando la expresión (5.227) para calcular la matriz de proyección ortogonal sobre cada subespacio se obtiene:

$$\begin{aligned} P_H &= \bar{q}_1\bar{q}_1^t + \bar{q}_2\bar{q}_2^t = \begin{pmatrix} 1 & 0 \\ 0 & 1/\sqrt{2} \\ 0 & 1/\sqrt{2} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/2 & 1/2 \\ 0 & 1/2 & 1/2 \end{pmatrix} \end{aligned} \quad (5.231)$$

$$\begin{aligned} P_{H^\perp} &= \bar{q}_3\bar{q}_3^t = \begin{pmatrix} 0 \\ 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix} (0, 1/\sqrt{2}, -1/\sqrt{2}) \\ &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1/2 & -1/2 \\ 0 & -1/2 & 1/2 \end{pmatrix} \end{aligned} \quad (5.232)$$

Observación 5.5.1 † Se requiere que $A \in \mathbb{K}^{m \times n}$ tenga rango completo por columnas, $r(A) = n$, porque de lo contrario, de acuerdo con la propiedad 6 de la Proposición 3.4.4, la matriz $A^h A$ no sería invertible.

3. Supóngase que el subespacio $H \subseteq \mathbb{K}^m$ de dimensión n es la imagen de una matriz $A \in \mathbb{K}^{m \times n}$, $m \leq n$ cuyo rango coincide con n †. Sus columnas, denotadas $\bar{u}_1, \dots, \bar{u}_n$, formarían una base del subespacio H . A continuación, se demostrará que la matriz de proyección ortogonal sobre H puede obtenerse a partir de la matriz A mediante la fórmula:

$$P_H = A (A^h A)^{-1} A^h \quad (5.233)$$

La base $\mathcal{B} = (\bar{u}_1, \dots, \bar{u}_n)$ que constituyen las columnas de A puede transformarse mediante el proceso de *Gram-Schmidt*, Teorema 5.4.1, en la base ortonormal $(\bar{q}_1, \dots, \bar{q}_n)$ de H . Constrúyase una matriz Q cuyas columnas sean los vectores ortogonales y unitarios $(\bar{q}_1, \dots, \bar{q}_n)$. Por el Teorema 5.4.3 de la descomposición QR, las matrices A y Q están relacionadas entre sí a través de la identidad $A = QR$, donde $R \in \mathbb{K}^{n \times n}$ es una matriz triangular superior de orden n . Despejando Q :

$$A = QR \Rightarrow Q = AR^{-1} \quad (5.234)$$

Sustituyendo la identidad (5.234) en (5.227) y recordando que $Q^h Q = I$, se obtiene:

$$\begin{aligned} P_H &= QQ^h = (AR^{-1})(AR^{-1})^h = AR^{-1}(R^{-1})^h A^h = \\ &= A(R^h R)^{-1} A^h = A(R^h Q^h QR)^{-1} A^h = \\ &= A((QR)^h (QR))^{-1} A^h = A(A^h A)^{-1} A^h \end{aligned} \quad (5.235)$$

Demostrándose así la expresión (5.233). Esta permite calcular la matriz de proyección ortogonal sobre la imagen de una matriz A de rango completo cuyo número de columnas sea menor o igual que su número de filas.

Ejemplo 5.5.4 La matriz:

$$A = \begin{pmatrix} 1 & 1 \\ 1 & -1 \\ 1 & -1 \end{pmatrix} \in \mathbb{R}^{3 \times 2} \quad (5.236)$$

Tiene rango completo por columnas, pues $r(A) = 2 = n$. Por tanto, sus columnas forman una base de su imagen. Denótese $M = \text{Im } A$. La matriz de proyección ortogonal sobre H puede calcularse a partir de la matriz A aplicando la fórmula (5.233):

$$\begin{aligned} P_H &= A (A^t A)^{-1} A^t = A \cdot \frac{1}{8} \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix} \cdot A^t = \\ &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/2 & 1/2 \\ 0 & 1/2 & 1/2 \end{pmatrix} \end{aligned} \quad (5.237)$$

Las matrices de proyección ortogonal cumplen las siguientes propiedades:

Proposición 5.5.2 – Propiedades de las matrices de proyección ortogonal

Sea P_H la matriz de proyección ortogonal sobre un subespacio H de \mathbb{K}^n .

1. P_H es hermítica e idempotente.

$$P_H^h = P_H = P_H^2 \quad (5.238)$$

Y toda matriz $P \in \mathbb{R}^{n \times n}$ que sea hermítica e idempotente es matriz de proyección ortogonal sobre su imagen.

2. La imagen y el núcleo de P_H son suplementarios ortogonales el uno respecto al otro.

$$\text{Im } P_H = (\text{ker } P_H)^\perp \quad (5.239)$$

y en consecuencia, $H^\perp = \text{ker } P_H$.

3. La matriz de proyección ortogonal sobre H^\perp puede calcularse como:

$$P_{H^\perp} = I - P_H \quad (5.240)$$

Ejercicio 5.5.1 † Demostrar las propiedades de la matriz de proyección ortogonal enumeradas en la Proposición 5.5.2.

Su demostración se deja † al lector.

Ejemplo 5.5.5 La matriz:

$$P_H = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/2 & 1/2 \\ 0 & 1/2 & 1/2 \end{pmatrix} \quad (5.241)$$

Es una matriz de proyección ortogonal sobre cierto subespacio $H \subseteq \mathbb{R}^3$, y como tal, cumple las siguientes propiedades:

1. Simetría e idempotencia, propiedad 1 de la Proposición 5.5.2:

$$P_H^t = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/2 & 1/2 \\ 0 & 1/2 & 1/2 \end{pmatrix}^t = P_H \quad (5.242)$$

$$P_H^2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/2 & 1/2 \\ 0 & 1/2 & 1/2 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/2 & 1/2 \\ 0 & 1/2 & 1/2 \end{pmatrix} = P_H \quad (5.243)$$

2. Relación entre su imagen y su núcleo, propiedad 2 de la misma proposición:

$$\text{Im } P_H = L[(1, 0, 0)^t, (0, 1/2, 1/2)^t] = M \quad (5.244)$$

$$\text{ker } P_H = L[(0, -1, 1)^t] = H^\perp \quad (5.245)$$

3. Relación entre P_H y P_{H^\perp} , propiedad 3 de la misma proposición:

$$P_{H^\perp} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1/2 & -1/2 \\ 0 & -1/2 & 1/2 \end{pmatrix} = I - P_H \quad (5.246)$$

5.5.5 Resumen

1. La proyección ortogonal de $v \in V$ sobre un subespacio H de V es aquel elemento cuya distancia *euclidiana* a v es la menor de entre todos los elementos contenidos en H .
2. La descomposición de un elemento de V en la suma de un elemento perteneciente a un subespacio H más otro perteneciente a su suplementario ortogonal H^\perp es única.
3. Si de la descomposición anterior se descarta el sumando perteneciente a H^\perp se obtiene un operador lineal que asocia a cada elemento de V su proyección ortogonal sobre H .
4. Si $V = \mathbb{K}^n$, el operador anterior admite una representación matricial denominada matriz de proyección ortogonal sobre H .
5. La imagen de la matriz de proyección ortogonal sobre H coincide con el subespacio sobre el que proyecta. Es decir, con H ; y su núcleo es el suplementario ortogonal de H .
6. La matriz de proyección ortogonal sobre $H \subseteq \mathbb{K}^n$ es simétrica e idempotente, y toda matriz simétrica e idempotente es matriz de proyección ortogonal sobre su subespacio imagen.

5.6 El problema de mínimos cuadrados

5.6.1 Introducción

Muchos problemas que surgen en Ingeniería se formulan como un sistema lineal cuyas soluciones constituyen la solución al problema. Si el sistema es compatible, su solución o soluciones se pueden calcular mediante métodos como el mostrado en el apartado 4.4.2; pero puede acontecer que el sistema sea incompatible, en cuyo caso no admitiría solución. Esto suele suceder por exceso de restricciones, o bien por errores de medida o de redondeo en los coeficientes del sistema.

En el caso de que el sistema sea incompatible, es natural buscar un subconjunto de vectores de \mathbb{K}^n que minimicen el error $\|\bar{\mathbf{b}} - A\bar{\mathbf{x}}\|$. A estos vectores se les denomina soluciones de mínimos cuadrados, o pseudosoluciones, del sistema incompatible $A\bar{\mathbf{x}} = \bar{\mathbf{b}}$.

5.6.2 Soluciones de mínimos cuadrados de un sistema

Si el sistema lineal $A\bar{\mathbf{x}} = \bar{\mathbf{b}}$ con matriz de coeficientes $A \in \mathbb{K}^{m \times n}$ y vector de coeficientes independientes $\bar{\mathbf{b}} \in \mathbb{K}^m$ es incompatible, no existe ningún $\bar{\mathbf{u}} \in \mathbb{K}^n$ tal que $A\bar{\mathbf{u}} = \bar{\mathbf{b}}$, por lo que la distancia *euclidiana* entre $\bar{\mathbf{b}}$ y la imagen de $\bar{\mathbf{x}}$ por A cumple $d(\bar{\mathbf{b}}, A\bar{\mathbf{x}}) = \|\bar{\mathbf{b}} - A\bar{\mathbf{x}}\| > 0 \forall \bar{\mathbf{x}} \in \mathbb{K}^n$. El problema a resolver será el de encontrar un conjunto de vectores $\bar{\mathbf{u}} \in \mathbb{K}^n$ tales que la distancia $d(\bar{\mathbf{b}}, A\bar{\mathbf{u}}) = \|\bar{\mathbf{b}} - A\bar{\mathbf{u}}\|$ alcance un mínimo global.

El subconjunto $\{A\bar{\mathbf{x}} : \bar{\mathbf{x}} \in \mathbb{K}^n\} \subseteq \mathbb{K}^m$ constituye, de acuerdo con el punto 1 de la Definición 3.4.1, la imagen de A . Por el Teorema 5.5.1 de la proyección ortogonal, el vector de $\text{Im } A$ cuya distancia *euclidiana* a $\bar{\mathbf{b}}$ es mínima es justamente la proyección ortogonal de $\bar{\mathbf{b}}$ sobre $\text{Im } A$. Por tanto, el vector $\bar{\mathbf{u}}$ que minimiza la norma *euclidiana* $\|\bar{\mathbf{b}} - A\bar{\mathbf{x}}\|$ es aquel cuya imagen por A es proyección ortogonal de $\bar{\mathbf{b}}$ sobre $\text{Im } A$ (Figura 5.9),

$$\min_{\bar{\mathbf{x}} \in \mathbb{K}^n} \|\bar{\mathbf{b}} - A\bar{\mathbf{x}}\| = \min_{\bar{\mathbf{x}} \in \mathbb{K}^n} d(\bar{\mathbf{b}}, A\bar{\mathbf{x}}) = \|\bar{\mathbf{b}} - P_{\text{Im } A} \bar{\mathbf{b}}\| \quad (5.247)$$

Con este razonamiento se concluye que $\bar{\mathbf{u}} \in \mathbb{K}^n$ minimiza la distancia $\|\bar{\mathbf{b}} - A\bar{\mathbf{x}}\|$, si y sólo si, $A\bar{\mathbf{u}} = P_{\text{Im } A} \bar{\mathbf{b}}$. Por tanto, encontrar los vectores donde $\|\bar{\mathbf{b}} - A\bar{\mathbf{x}}\|$ alcanza el mínimo global equivale a resolver el sistema $A\bar{\mathbf{x}} = P_{\text{Im } A} \bar{\mathbf{b}}$, sistema siempre compatible y que tiene, de acuerdo con el Teorema 4.3.1 de *Rouché-Frobenius*, una o infinitas soluciones en función del rango de la matriz de coeficientes A . A los $\bar{\mathbf{x}} \in \mathbb{K}^n$ tales que la distancia $\|\bar{\mathbf{b}} - A\bar{\mathbf{x}}\|$ alcanza un mínimo global se les llama “soluciones de mínimos cuadrados” o “pseudosoluciones” del sistema.

Definición 5.6.1 — Solución de mínimos cuadrados de un sistema lineal

Se denomina solución de mínimos cuadrados o pseudosolución del sistema de ecuaciones lineal $A\bar{\mathbf{x}} = \bar{\mathbf{b}}$ a cualquier solución del sistema $A\bar{\mathbf{x}} = P_{\text{Im } A} \bar{\mathbf{b}}$.

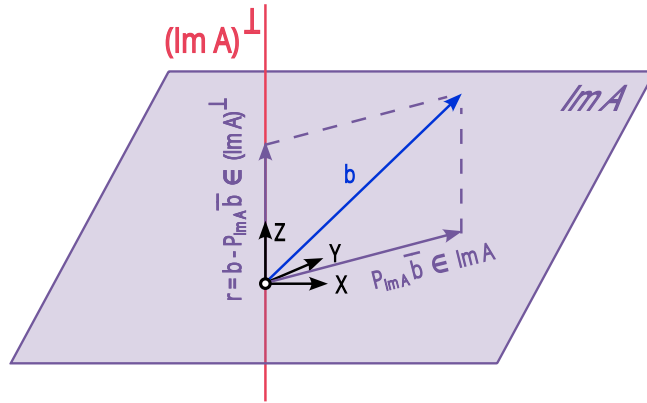


Figura 5.9: Si $\bar{b} \notin \text{Im } A$, el sistema $A\bar{x} = \bar{b}$ es incompatible, Teorema 4.3.1. Pero si \bar{b} se sustituye por su proyección ortogonal sobre $\text{Im } A$, el sistema se vuelve compatible, siendo las soluciones de ese nuevo sistema vectores de \mathbb{K}^n cuya imagen por A se encuentra a una distancia euclidiana mínima de \bar{b} .

Ejemplo 5.6.1 El sistema $A\bar{x} = \bar{b}$ definido por la matriz de coeficientes:

$$A = \begin{pmatrix} 1 & 1 \\ 1 & -1 \\ 1 & -1 \end{pmatrix} \tag{5.248}$$

Y el vector de coeficientes independientes $\bar{b} = (0, 1, 0)^t$ es incompatible, pues $\bar{b} \notin \text{Im } A$.

Sustituyendo \bar{b} por su proyección ortogonal sobre $\text{Im } A$ se obtiene un sistema compatible cuyas soluciones son las soluciones de mínimos cuadrados del sistema $A\bar{x} = \bar{b}$. Este sistema es:

$$A\bar{x} = P_H \bar{b} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/2 & 1/2 \\ 0 & 1/2 & 1/2 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 1/2 \\ 1/2 \end{pmatrix} \tag{5.249}$$

Este sistema es compatible determinado, pues $r(A) = 2 = n$ y $P_{\text{Im } A} \bar{b} \in \text{Im } A$. Su única solución, que es la solución de mínimos cuadrados del sistema $A\bar{x} = \bar{b}$, es el vector $\bar{u} = (1/4, -1/4)^t \in \mathbb{R}^2$.

5.6.3 Vector residuo de un sistema lineal incompatible

Se define el vector residuo de un sistema incompatible como el vector que resulta de sustraer a \bar{b} su proyección ortogonal sobre $\text{Im } A$. De acuerdo con la ecuación (5.247), este vector coincide con la diferencia entre \bar{b} y la imagen por A de una cualquiera de sus soluciones de mínimos cuadrados.

Definición 5.6.2 – Vector error de un sistema incompatible

Se define el vector residuo de un sistema $A\bar{x} = \bar{b}$ incompatible ($\bar{b} \notin \text{Im } A$) como:

$$r = \bar{b} - P_{\text{Im } A} \bar{b} = \bar{b} - A\bar{u} \in \mathbb{K}^m$$

Donde $\bar{u} \in \mathbb{K}^n$ es cualquier solución de mínimos cuadrados del sistema.

El módulo del vector residuo del sistema es precisamente la distancia entre $\bar{b} \in \mathbb{K}^n$ y la imagen de $\text{Im } A$, $d(\bar{b}, \text{Im } A)$.

Ejemplo 5.6.2 El sistema del Ejemplo 5.6.1 está caracterizado por un vector residuo igual a:

$$r = \bar{b} - P_{\text{Im}A} \bar{b} = (0, 1, 0)^t - (0, 1/2, 1/2)^t = (0, 1/2, -1/2)^t$$

5.6.4 Sistema de ecuaciones normales asociado a un sistema lineal

Un modo alternativo † de obtener las soluciones de mínimos cuadrados de un sistema incompatible, equivalente al mostrado en el apartado 5.6.2 anterior, es encontrar el vector o vectores $\bar{x} \in \mathbb{K}^m$ cuya imagen por A sea ortogonal al vector residuo $\bar{b} - A\bar{x}$ (Figura 5.9):

$$\langle \bar{b} - A\bar{x}, A\bar{x} \rangle = 0 \quad (5.250)$$

Aplicando la Definición 5.1.7 del producto escalar usual:

$$\begin{aligned} \langle \bar{b} - A\bar{x}, A\bar{x} \rangle &= \bar{x}^h A^h (\bar{b} - A\bar{x}) = \bar{x}^h A^h \bar{b} - \bar{x}^h A^h A\bar{x} = \\ &= \bar{x}^h (A^h \bar{b} - A^h A\bar{x}) = 0 \end{aligned} \quad (5.251)$$

El vector nulo $\bar{0}$ es solución de un sistema de ecuaciones, si y sólo si, $\bar{b} = \bar{0}$, pues el producto de una matriz no nula por un vector, o matriz, no nulo es siempre no nulo. Además, es necesario que $\bar{b} \neq \bar{0}$ para que un sistema lineal sea incompatible, por lo que puede asumirse que $\bar{x} \neq \bar{0}$ en la identidad (5.250) y queda que \bar{x} cumple la condición (5.250) de ortogonalidad entre $\bar{b} - A\bar{x}$ y $A\bar{x}$, si y sólo si:

$$A^h \bar{b} - A^h A\bar{x} = \bar{0} \quad (5.252)$$

Es decir, si y sólo si es solución del sistema $A^h A\bar{x} = A^h \bar{b}$. Como resultado, las soluciones de mínimos cuadrados de un sistema de ecuaciones lineales $A\bar{x} = \bar{b}$ coinciden con las soluciones del sistema lineal $A^h A\bar{x} = A^h \bar{b}$, lo que motiva la siguiente definición:

Definición 5.6.3 – Sistema de ecuaciones normales asociado al sistema $A\bar{x} = \bar{b}$

Dado un sistema lineal $A\bar{x} = \bar{b}$, su sistema de ecuaciones normales asociado es el siguiente:

$$A^h A\bar{x} = A^h \bar{b}$$

Este sistema es siempre compatible, pues $\text{Im} A^h = \text{Im}(A^h A) \Rightarrow A^h \bar{b} \in \text{Im} A^h A$, y sus soluciones son, como se ha demostrado arriba, las soluciones de mínimos cuadrados del sistema $A\bar{x} = \bar{b}$.

Observación 5.6.1 † En el campo de la simulación numérica, a esta metodología para resolver el problema de mínimos cuadrados ortogonalizando el residuo $\bar{b} - A\bar{x}$ con la imagen de A se la conoce como el “método de Galerkin”. Es muy útil para encontrar soluciones aproximadas de ecuaciones diferenciales no lineales como las que rigen la mecánica del sólido elástico o la mecánica de fluidos.

Proposición 5.6.1 – Soluciones de un sistema de ecuaciones normales

$\bar{u} \in \mathbb{K}^n$ es solución de mínimos cuadrados del sistema $A\bar{x} = \bar{b}$, si y sólo si, \bar{u} es solución de su sistema de ecuaciones normales asociado.

Si A tiene rango completo por columnas. Es decir, si $r(A) = n$, la propiedad 3 de la Proposición 2.2.5 garantiza que la matriz $A^h A \in \mathbb{K}^{n \times n}$ sea invertible, y la solución de mínimos cuadrados \bar{x} se podría despejar de la ecuación (5.6.3):

$$\bar{x} = (A^h A)^{-1} A^h \bar{b} \tag{5.253}$$

A la matriz $(A^h A)^{-1} A^h$ se le suele conocer con el nombre de “pseudoinversa de A ” o “inversa de *Moose-Penrose*” †.

Observación 5.6.2 † Más concretamente, la matriz A^+ es la “inversa por la izquierda” de A , pues $A^+ A$ es claramente igual a la matriz identidad pero AA^+ no tiene por qué serlo (lo sería, si y sólo si, A fuese cuadrada). Si $A \in \mathbb{K}^{m \times n}$ es una matriz de rango completo por filas, $r(A) = m$, su “inversa por la derecha” es la matriz $A^+ = A^h (AA^h)^{-1}$, que cumple $AA^+ = I$.

Definición 5.6.4 – Pseudoinversa o inversa de *Moose-Penrose*

Dada una matriz $A \in \mathbb{K}^{m \times n}$ de rango completo por columnas, a la matriz:

$$A^+ = (A^h A)^{-1} A^h$$

Se la suele conocer como pseudoinversa, o inversa de *Moose-Penrose*, de A .

Ejemplo 5.6.3 El sistema $A\bar{x} = \bar{b}$ definido por la matriz de coeficientes:

$$A = \begin{pmatrix} 1 & 1 \\ 1 & -1 \\ 1 & -1 \end{pmatrix} \tag{5.254}$$

Y el vector de coeficientes independientes $\bar{b} = (0, 1, 0)^t$ tiene el siguiente sistema de ecuaciones normales asociado:

$$A^t A \bar{x} = A^t \bar{b} \Rightarrow \begin{pmatrix} 3 & -1 \\ -1 & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \tag{5.255}$$

Su única solución es el vector $\bar{u} = (1/4, -1/4)^t$, la solución de mínimos cuadrados del sistema $A\bar{x} = \bar{b}$.

5.6.5 Una aplicación de los mínimos cuadrados: el problema de la regresión lineal

Imagínese que en un experimento se realizan m mediciones de una variable “dependiente” y en función de n variables denotadas $x^1, x^2 \dots x^n$ ‡ a las que se denominarán “variables independientes” del sistema. El problema de los mínimos cuadrados puede aplicarse para estimar los parámetros de un modelo matemático que represente el sistema físico en el que se realiza el experimento, suponiendo como hipótesis que la dependencia de la variable y en función de las n variables independientes es lineal:

$$y = \theta_1 x^1 + \theta_2 x^2 + \dots + \theta_n x^n + b \tag{5.256}$$

Observación 5.6.3 ‡ Aquí, los superíndices no representan potencias, sino el número asignado a cada variable independiente medida durante un experimento.

En cada una de las m mediciones de la variable dependiente y se fijan distintos valores de las n variables independientes, dando lugar a las $n+1$ -*uplas* $(x_1^1, x_1^2, \dots, x_1^n, \hat{y}_1)$, $(x_2^1, x_2^2, \dots, x_2^n, \hat{y}_2)$, \dots , $(x_m^1, x_m^2, \dots, x_m^n, \hat{y}_m)$. Debido, por un lado, a errores en la medición de y , y por otro lado a la inexactitud de la hipótesis de linealidad, así como a la presencia de factores incontrolables que influyen en el experimento, cada medición viene perturbada por un error ϵ que hace que el modelo matemático (5.256) no se cumpla de manera exacta, de modo que:

$$\hat{y} = \theta_1 x^1 + \theta_2 x^2 + \dots + \theta_n x^n + b + \epsilon \quad (5.257)$$

Siendo el error una variable aleatoria que suele distribuirse de manera normal (o *gaussiana*) con media μ y varianza σ^2 .

Sustituir los valores de las variables dependientes e independientes de cada una de las mediciones en la expresión (5.256) da lugar a las m ecuaciones siguientes:

$$\begin{aligned} \hat{y}_1 &= \theta_1 x_1^1 + \theta_2 x_1^2 + \dots + \theta_n x_1^n + b + \epsilon_1 \\ \hat{y}_2 &= \theta_1 x_2^1 + \theta_2 x_2^2 + \dots + \theta_n x_2^n + b + \epsilon_2 \\ &\vdots \\ \hat{y}_m &= \theta_1 x_m^1 + \theta_2 x_m^2 + \dots + \theta_n x_m^n + b + \epsilon_m \end{aligned} \quad (5.258)$$

Los parámetros del modelo (5.256) pueden estimarse a partir de las m mediciones resolviendo el siguiente problema: encontrar los estimadores del modelo $\hat{\theta}_1, \dots, \hat{\theta}_n, \hat{b} \in \mathbb{R}$ tal que la suma de los cuadrados de los errores de medición durante el experimento sea mínima:

$$\min_{\hat{\theta}_1, \dots, \hat{\theta}_n, \hat{b} \in \mathbb{R}} \sum_{i=1}^m \epsilon_i^2 \quad (5.259)$$

donde $\epsilon_i = \hat{\theta}_1 x_i^1 + \hat{\theta}_2 x_i^2 + \dots + \hat{\theta}_n x_i^n + \hat{b} - \hat{y}_i \forall i = 1, \dots, m$.

Resolver el problema (5.259) equivale a minimizar la norma euclidiana del vector $\bar{b} - A\bar{x}$, donde A , \bar{b} y \bar{x} son, respectivamente, la matriz de coeficientes, el vector de coeficientes independientes y el vector solución $(\hat{\theta}_1, \dots, \hat{\theta}_n, \hat{b})^t$ del sistema lineal:

$$\begin{pmatrix} x_1^1 & x_1^2 & \dots & x_1^n & 1 \\ x_2^1 & x_2^2 & \dots & x_2^n & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_m^1 & x_m^2 & \dots & x_m^n & 1 \end{pmatrix} \begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \\ b \end{pmatrix} = \begin{pmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_m \end{pmatrix} \quad (5.260)$$

Este sistema de ecuaciones lineales es compatible, si y sólo si, $\epsilon_i = 0 \forall i = 1, \dots, m$, lo que en la práctica no sucede nunca. Su solución de mínimos cuadrados, que si no hay ensayos redundantes es única, contiene los estimadores $\hat{\theta}_1, \dots, \hat{\theta}_n, \hat{b}$ del modelo que representa el sistema físico considerado.

5.6.6 Resumen

1. Los sistemas de ecuaciones lineales incompatibles no admiten solución exacta, pero es siempre posible encontrar al menos un vector cuya imagen por A tenga una distancia *euclidiana* mínima respecto del vector de coeficientes independientes del sistema. A estos vectores se les llama “pseudosoluciones” o “soluciones de mínimos cuadrados” del sistema.
2. Por el Teorema 5.5.1 de la proyección ortogonal, el vector de $\text{Im } A$ cuya distancia al vector de coeficientes independientes \bar{b} del sistema es mínima es su proyección ortogonal sobre $\text{Im } A$. Por tanto, para encontrar las soluciones de mínimos cuadrados de un sistema puede sustituirse \bar{b} por su proyección ortogonal sobre $\text{Im } A$ y resolverse este nuevo sistema.
3. La solución o soluciones de mínimos cuadrados pueden obtenerse también resolviendo el sistema de ecuaciones normales asociado al sistema original.



Diagonalización de matrices cuadradas

6

6.1 Introducción a la diagonalización de matrices cuadradas

6.1.1 Introducción

Después de la resolución de sistemas de ecuaciones lineales, la otra aplicación estrella del álgebra lineal es la llamada diagonalización de matrices pues, como razonaremos más adelante, en ella se basan los métodos de resolución de muchos problemas de física, matemáticas e ingeniería. Aunque es posible diagonalizar matrices rectangulares mediante su descomposición en valores singulares, este texto introductorio se centrará en la diagonalización de matrices cuadradas, que se fundamenta en la búsqueda de una base \mathcal{B} del espacio E tal que la matriz asociada a un endomorfismo lineal $f : E \mapsto E$ respecto de \mathcal{B} sea diagonal.

Tras introducir y definir el problema de diagonalizar una matriz cuadrada, se definirán los conceptos de valor y vector propio, polinomio característico y subespacio propio, así como sus propiedades, se enunciará la condición necesaria y suficiente que ha de cumplir toda matriz cuadrada para que sea diagonalizable y se propondrá un procedimiento analítico para llevar a cabo dicha diagonalización. Por último, en la sección de diagonalización unitaria, se mostrará que las matrices normales y las matrices reales simétricas pueden diagonalizarse de una manera especial debido a la ortogonalidad entre sus subespacios propios asociados.

6.1.2 Matrices diagonalizables

Recuérdese que, por la Definición 3.3.2, dos matrices cuadradas A, B son semejantes si existe alguna matriz P invertible tal que:

$$B = P^{-1}AP \quad (6.1)$$

Dada una matriz $A \in \mathbb{K}^{n \times n}$, podría plantearse el problema de encontrar una matriz D semejante a A que sea diagonal, y una matriz P que relacione ambas matrices. A esto se llama diagonalizar una matriz A . Si existe dicha matriz diagonal, se dice que A es diagonalizable.

Definición 6.1.1 — Matriz cuadrada diagonalizable

Una matriz cuadrada A es diagonalizable si es semejante a alguna matriz diagonal. Es decir, si existe alguna matriz P invertible tal que $P^{-1}AP$ es diagonal.

6.1	Introducción a la diagonalización	193
6.2	Valores y vectores propios	196
6.3	Polinomio característico y subespacios propios de una matriz	199
6.4	Diagonalización de matrices y sus propiedades	205
6.5	Diagonalización unitaria	216

6.1.3 Interpretación geométrica de la diagonalización

Estos son algunos motivos por los que puede ser deseable diagonalizar una matriz:

1. Las matrices diagonales tienen todos sus elementos nulos salvo los de la diagonal, por lo que estas matrices ocupan menos memoria de ordenador y es mucho más sencillo operar con ellas. Por ejemplo, sistemas de ecuaciones con matrices de coeficientes diagonales tienen sus ecuaciones desacopladas.
2. Tanto los elementos de la diagonal de D como las columnas de la matriz P tienen un significado físico: su cálculo proporciona información sobre la aplicación lineal f que representa A . Este significado físico es evidente cuando se ve que:

$$P^{-1}AP = D \Rightarrow AP = PD \Rightarrow$$

$$\Rightarrow A(\bar{v}_1 | \dots | \bar{v}_n) = (\bar{v}_1 | \dots | \bar{v}_n) \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} \quad (6.2)$$

Los vectores columna $\bar{v}_1, \bar{v}_2, \dots, \bar{v}_n$ son vectores cuya dirección se mantiene invariante al ser transformados por la matriz A . A estos vectores se les conoce como “vectores propios de la matriz A ”.

Las imágenes de cada \bar{v}_i son proporcionales a los vectores \bar{v}_i con un factor de proporcionalidad $\lambda_i \in \mathbb{K}$. A cada uno de estos escalares se le conoce como el “valor propio asociado a cada vector propio de A ” (Figura 6.1).

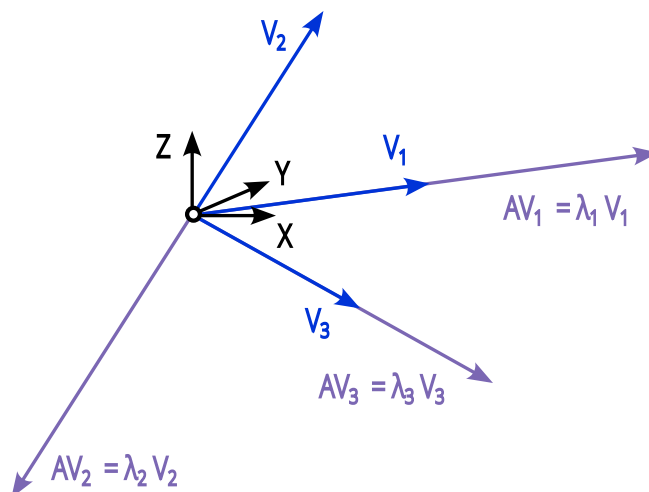


Figura 6.1: Los vectores $\bar{v}_1, \bar{v}_2, \bar{v}_3 \in \mathbb{R}^3$ son vectores propios de una cierta matriz real cuadrada A . Sus transformados por A son vectores proporcionales a \bar{v}_1, \bar{v}_2 y \bar{v}_3 con factores de proporcionalidad iguales, respectivamente, a λ_1, λ_2 y λ_3 . Estos coeficientes son los valores propios asociados a cada vector propio. En este ejemplo, el transformado de \bar{v}_3 por A tiene sentido opuesto respecto a \bar{v}_3 . Esto sucede cuando $\lambda_3 < 0$.

6.1.4 Aplicaciones de la diagonalización en la ingeniería

A veces, cuando se emplean matrices para modelizar sistemas físicos, sus valores y vectores propios proporcionan una información sobre los sistemas físicos modelados que permite resolver problemas de ingeniería como los siguientes:

1. El cálculo de modos de vibración de edificios, máquinas, etc. y sus frecuencias de resonancia. Es decir, aquellas en las que sus amplitudes de vibración se maximizan.
2. La estimación de las direcciones de fractura de componentes mecánicos bajo diferentes tipos de sollicitación mecánica: tracción, flexión, torsión, etc.
3. El estudio de la estabilidad de sistemas dinámicos de toda clase: sistemas de control y regulación automática, sistemas químicos, ecosistemas donde conviven varias especies, etc.
4. El diseño de motores de búsqueda en Internet como el algoritmo *PageRank* de *Google* ©.
5. La descomposición de señales de audio, vídeo, etc. en sus componentes fundamentales para su análisis o para su filtrado.
6. Análisis masivo de datos (*big data*). Un ejemplo es la elaboración de listas de recomendación personalizadas en sitios web, que se basa en diagonalizar una gran matriz.
7. La resolución de sistemas de ecuaciones diferenciales lineales para la simulación de sistemas dinámicos como los del punto 3.

6.1.5 Resumen

1. Una matriz cuadrada A es diagonalizable si entre sus matrices semejantes existe alguna que sea diagonal. Diagonalizar una matriz consiste en encontrar una matriz semejante a esta que sea diagonal.
2. Los elementos de la diagonal principal de una matriz diagonal semejante a A son los valores propios de A . Las columnas de una matriz de cambio de base que convierte la matriz A en una matriz diagonal semejante a esta son vectores propios asociados a cada valor propio de A .
3. Cualquier vector propio de una matriz A se convierte, al ser transformado por A , en otro vector con su misma dirección y con su longitud multiplicada por su valor propio asociado.
4. La diagonalización de matrices aparece en la resolución de numerosos problemas de ingeniería.

6.2 Valores y vectores propios

6.2.1 Introducción

Como se ha mostrado en el apartado 6.1.3, los valores propios de una matriz constituyen los elementos de una matriz diagonal cualquiera que sea semejante a ella, mientras que los los vectores propios asociados a cada valor propio constituyen las columnas de una matriz de cambio de base que transforma la matriz inicial en una matriz diagonal semejante. Por tanto, para llevar a cabo la diagonalización de una matriz será necesario calcular tanto sus valores propios como un conjunto de vectores propios asociados a cada valor propio.

En este apartado, se aportarán definiciones rigurosas de los conceptos de valor propio, vector propio, espectro y radio espectral de una matriz cuadrada, y se demostrará que vectores propios asociados a autovalores distintos son siempre linealmente independientes entre sí.

6.2.2 Definición de valor y vector propio de una matriz cuadrada

Definición 6.2.1 – Valores y vectores propios de una matriz cuadrada

Sea A una matriz cuadrada de orden n , se dice que:

1. El escalar $\lambda \in \mathbb{K}$ es un valor propio o autovalor de la matriz A si existe algún vector \bar{v} no nulo perteneciente a \mathbb{K}^n tal que $A\bar{v} = \lambda\bar{v}$.
2. El vector no nulo $\bar{v} \in \mathbb{K}^n$ es un vector propio o autovector de la matriz A si existe un escalar λ (su autovalor asociado) en \mathbb{K} tal que $A\bar{v} = \lambda\bar{v}$.

Ejemplo 6.2.1 Considérese la matriz:

$$A = \begin{pmatrix} 2 & -4 \\ 0 & 4 \end{pmatrix} \quad (6.3)$$

Y los vectores $\bar{v}_1 = (1, 0)^t$ y $\bar{v}_2 = (2, -1)^t$. Se tiene que:

$$A\bar{v}_1 = A \cdot (1, 0)^t = (2, 0)^t = 2 \cdot (1, 0)^t = 2 \cdot \bar{v}_1 \quad (6.4)$$

$$A\bar{v}_2 = A \cdot (2, -1)^t = (8, -4)^t = 4 \cdot (2, -1)^t = 4 \cdot \bar{v}_2 \quad (6.5)$$

Por lo que \bar{v}_1 y \bar{v}_2 son vectores propios asociados, respectivamente, a los valores propios $\lambda_1 = 2$ y $\lambda_2 = 4$.

A continuación, se demostrará que vectores propios asociados a autovalores distintos son siempre linealmente independientes. Sea \bar{v}_1 un vector propio de una matriz A cuadrada de orden n , en virtud del punto 2 de la Definición 6.2.1, este ha de ser distinto del vector nulo, por lo que el conjunto $\{\bar{v}_1\}$ es una familia libre.

Ahora supóngase que $\{\bar{v}_1, \bar{v}_2, \dots, \bar{v}_r\}$ es una familia libre formada por $r < n$ vectores propios de A . Tómesese otro vector propio de A denotado \bar{v}_{r+1} distinto de todos los contenidos en la familia. Por la Definición 1.2.3, para que los $r + 1$ autovectores fuesen linealmente independientes estos habrían de cumplir:

$$\alpha_1 \bar{v}_1 + \alpha_2 \bar{v}_2 + \dots + \alpha_{r+1} \bar{v}_{r+1} = \bar{0} \Leftrightarrow \alpha_1, \dots, \alpha_{r+1} \neq 0 \quad (6.6)$$

pre-multiplicando la expresión $\alpha_1 \bar{v}_1 + \alpha_2 \bar{v}_2 + \dots + \alpha_{r+1} \bar{v}_{r+1}$ por A :

$$\begin{aligned} & \alpha_1 A \bar{v}_1 + \alpha_2 A \bar{v}_2 + \dots + \alpha_{r+1} A \bar{v}_{r+1} = \\ & = \alpha_1 \lambda_1 \bar{v}_1 + \alpha_2 \lambda_2 \bar{v}_2 + \dots + \alpha_{r+1} \lambda_{r+1} \bar{v}_{r+1} = \bar{0} \end{aligned} \quad (6.7)$$

Sustrayendo la identidad $\alpha_1 \bar{v}_1 + \alpha_2 \bar{v}_2 + \dots + \alpha_{r+1} \bar{v}_{r+1} = \bar{0}$ multiplicada por λ_{r+1} a la expresión (6.7):

$$\begin{aligned} & \alpha_1 (\lambda_1 - \lambda_{r+1}) \bar{v}_1 + \alpha_2 (\lambda_2 - \lambda_{r+1}) \bar{v}_2 + \dots \\ & \dots + \alpha_r (\lambda_r - \lambda_{r+1}) \bar{v}_r = \bar{0} \end{aligned} \quad (6.8)$$

Si $\lambda_{r+1} \neq \lambda_i \forall i = 1, \dots, r$ y $\lambda_i - \lambda_{r+1} \neq 0 \forall i = 1, \dots, r$. Además, los vectores $\bar{v}_1, \dots, \bar{v}_r$ forman por hipótesis una familia libre. Por tanto, la identidad (6.8) se cumple, si y sólo si, $\alpha_1 = \alpha_2 = \dots = \alpha_r = 0$. Sustituyendo en (6.6) se tiene:

$$\alpha_{r+1} \bar{v}_{r+1} = \bar{0} \quad (6.9)$$

Por la Definición 6.2.1, punto 2, de vector propio, $\bar{v}_{r+1} \neq \bar{0}$, por lo que $\alpha_{r+1} = 0$ y se concluye que los vectores $\bar{v}_1, \dots, \bar{v}_r, \bar{v}_{r+1}$ forman una familia libre.

Se acaba de demostrar por inducción la siguiente proposición:

Proposición 6.2.1 – Independencia lineal de autovectores asociados a autovalores diferentes

Si $\bar{v}_1, \bar{v}_2, \dots, \bar{v}_r$ son r vectores propios de una matriz A cuadrada asociados a r valores propios $\lambda_1, \lambda_2, \dots, \lambda_r$ distintos dos a dos, dichos vectores son linealmente independientes.

Ejemplo 6.2.2 La matriz A del Ejemplo 6.2.1 tiene el vector propio $\bar{v}_1 = (1, 0)^t$ asociado a $\lambda_1 = 2$ y el vector propio $\bar{v}_2 = (2, -1)^t$ asociado a $\lambda_2 = 4$, entre otros. Cada uno de estos dos vectores propios está asociado a un valor propio distinto. La Proposición 6.2.1 establece que vectores propios asociados a autovalores distintos son siempre linealmente independientes entre sí. Como puede verse fácilmente, estos lo son.

Ejemplo 6.2.3 Considérense ahora la matriz:

$$B = \begin{pmatrix} -2 & -4 & 2 \\ -2 & 1 & 2 \\ 4 & 2 & 5 \end{pmatrix} \quad (6.10)$$

Y los vectores:

$$\bar{w}_1 = (2, 1, -1)^t \quad (6.11)$$

$$\bar{w}_2 = (2, -3, -1)^t \quad (6.12)$$

$$\bar{w}_3 = (1, 6, 16)^t \quad (6.13)$$

Puede comprobarse que \bar{w}_1 es vector propio asociado a $\lambda_1 = -5$, \bar{w}_2 es vector propio asociado a $\lambda_2 = 3$ y \bar{w}_3 es vector propio asociado a $\lambda_3 = 6$. Cada uno de estos dos vectores propios está asociado a un valor propio distinto. La Proposición 6.2.1 establece que vectores propios asociados a autovalores distintos son siempre linealmente independientes entre sí. Como puede verse fácilmente, estos lo son.

6.2.3 Espectro y radio espectral

Definición 6.2.2 Sea A una matriz cuadrada, se llama:

1. Espectro de A al conjunto de todos sus valores propios.
2. Radio espectral de A , y se denota $\rho(A)$, al máximo de los módulos de los valores propios.

Ejemplo 6.2.4 Los valores propios de la matriz:

$$C = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 1 & 1 \end{pmatrix} \quad (6.14)$$

Son $1 + i$, $1 - i$ y 1 †. Por tanto, su espectro es el conjunto:

$$\{1 + i, 1 - i, 1\} \subset \mathbb{C} \quad (6.15)$$

Y su radio espectral es:

$$\rho(A) = \max\{\sqrt{2}, \sqrt{2}, 1\} = \sqrt{2} \quad (6.16)$$

Observación 6.2.1 † Se describirá un método para su cálculo en el apartado 6.4.8.

6.3 Polinomio característico y subespacios propios de una matriz cuadrada

6.3.1 Introducción

Como se vio en el apartado 2.4.5, los valores y vectores propios definidos en el apartado anterior son los ingredientes necesarios para diagonalizar una matriz. En esta nueva sección, se mostrará una manera analítica de calcular los valores y vectores propios de una matriz A , así como y propiedades adicionales de estos.

Concretamente, los valores propios de una matriz cuadrada coinciden con las raíces de un polinomio definido en función de esta llamado “polinomio característico de la matriz”, mientras que los vectores propios asociados a cada autovalor de una matriz forman subespacios vectoriales conocidos como “subespacios característicos” o “propios” de la matriz.

6.3.2 Subespacio característico o propio

Según la Proposición 6.2.1, vectores propios asociados a autovalores diferentes son linealmente independientes entre sí. Si los autovectores están asociados al mismo valor propio, es muy fácil demostrar, aplicando la Definición 6.2.1, punto 2, de vector propio, que cualquier combinación lineal de ellos es también vector propio asociado al mismo autovalor (Figura 6.2). Sean \bar{v}_1, \bar{v}_2 autovectores de A asociados a λ , $\forall \alpha, \beta \in \mathbb{K}$

$$\begin{aligned} A \cdot (\alpha\bar{v}_1 + \beta\bar{v}_2) &= \alpha A\bar{v}_1 + \beta A\bar{v}_2 = \alpha\lambda\bar{v}_1 + \beta\lambda\bar{v}_2 = \\ &= \lambda \cdot (\alpha\bar{v}_1 + \beta\bar{v}_2) \end{aligned} \quad (6.17)$$

Lo que permite enunciar la siguiente proposición:

Proposición 6.3.1 – Relación entre autovectores asociados al mismo autovalor

Sean \bar{v}_1, \bar{v}_2 vectores propios de A asociados al mismo autovalor λ , cualquier vector combinación lineal de estos $\alpha\bar{v}_1 + \beta\bar{v}_2$, con $\alpha, \beta \in \mathbb{K}$ distintos de cero, es también vector propio de A asociado al mismo valor propio λ .

Ejemplo 6.3.1 $\lambda = 0$ es uno de los autovalores de la matriz:

$$A = \begin{pmatrix} 1 & 2 & -1 \\ -1 & -2 & 1 \\ 2 & 4 & -2 \end{pmatrix} \quad (6.18)$$

Dos vectores propios asociados a $\lambda = 0$ son $\bar{v}_1 = (2, -1, 0)^t$ y $\bar{v}_2 = (0, 1, 2)^t$:

$$A\bar{v}_1 = 0 \cdot \bar{v}_1 = \bar{0} \quad (6.19)$$

$$A\bar{v}_2 = 0 \cdot \bar{v}_2 = \bar{0} \quad (6.20)$$

Cualquier combinación lineal de \bar{v}_1 y \bar{v}_2 cumple:

$$A(\alpha\bar{v}_1 + \beta\bar{v}_2) = \alpha A\bar{v}_1 + \beta A\bar{v}_2 = \alpha \cdot \bar{0} + \beta \cdot \bar{0} = \bar{0} \quad (6.21)$$

Por lo que $\alpha\bar{v}_1 + \beta\bar{v}_2 \forall \alpha, \beta \in \mathbb{R}$ es también vector propio de A asociado a $\lambda = 0$ (excluido el vector nulo, que por definición no puede ser nunca vector propio.)

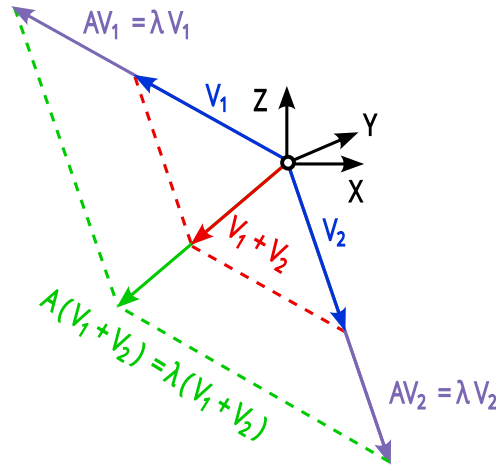


Figura 6.2: \bar{v}_1 y \bar{v}_2 son vectores propios de cierta matriz cuadrada asociados al mismo valor propio λ . Cualquier combinación lineal de estos, por ejemplo su suma $\bar{v}_1 + \bar{v}_2$, es también un vector propio de la misma matriz asociado al mismo autovalor.

Como consecuencia de la Proposición 6.3.1 y de la Definición 1.2.6, el conjunto de todos los vectores propios asociados a un determinado valor propio constituye, junto con el vector nulo, un subespacio vectorial al que se denominará “subespacio característico” o “propio” de la matriz asociado al susodicho valor.

Definición 6.3.1 – Subespacio característico o propio

Si λ es valor propio de una matriz A cuadrada, se denomina subespacio característico, o propio, de A asociado a λ al conjunto formado por todos los vectores propios de A asociados a λ , junto con el vector nulo.

Ejemplo 6.3.2 Los vectores $\bar{v}_1 = (2, -1, 0)^t$ y $\bar{v}_2 = (0, 1, 2)^t$ de \mathbb{R}^3 son, como se ha visto en el Ejemplo 6.3.1 vectores propios de la matriz A , ecuación (6.18) asociados al autovalor $\lambda = 0$. Por la Proposición 6.3.1, cualquier combinación lineal de estos dos vectores es también vector propio de A asociado a $\lambda = 0$.

Puede comprobarse † que el subespacio $L[(\bar{v}_1, \bar{v}_2)]$ es precisamente $\ker A$. De hecho, todo vector que cumple $A\bar{v} = \bar{0}$ (salvo el vector nulo) es autovector asociado al valor propio $\lambda = 0$.

Ejercicio 6.3.1 † Calcular el núcleo de la matriz A , ecuación (6.18).

Para calcular el subespacio propio de una matriz cuadrada asociado a alguno de sus autovalores, nótese que si \bar{v} es vector propio asociado a λ , se cumple:

$$A\bar{v} = \lambda\bar{v} \Rightarrow \lambda\bar{v} - A\bar{v} = \bar{0} \Rightarrow (\lambda I - A)\bar{v} = \bar{0} \quad (6.22)$$

Por lo que \bar{v} pertenecería al núcleo de la matriz $(\lambda I - A)$.

Se concluye que el subespacio propio de $A \in \mathbb{K}^{n \times n}$ asociado al valor propio λ coincide con $\ker(\lambda I - A)$, de forma que:

$$\bar{v} \in \ker(\lambda I - A) \Leftrightarrow A\bar{v} = \lambda\bar{v} \quad (6.23)$$

Por el Teorema 3.4.5 del rango-nulidad, $\dim \ker(\lambda I - A) > 0$, si y sólo si, existen vectores $\bar{v} \in \mathbb{K}^n$ no nulos tal que $(\lambda I - A)\bar{v} = \bar{0}$, por lo que $\lambda \in \mathbb{K}$ es valor propio de A , si y sólo si, $r(\lambda I - A) < n$. Esta última condición puede aplicarse para calcular los valores propios asociados a cualquier matriz cuadrada.

Ejemplo 6.3.3 La matriz B , ecuación (6.10) del Ejemplo 6.2.3, tiene entre sus valores propios el escalar real $\lambda = -5$, ya que:

$$r(-5I - B) = r \begin{pmatrix} -3 & 4 & -2 \\ 2 & -6 & -2 \\ -4 & -2 & -10 \end{pmatrix} = 2 < 3 = n \quad (6.24)$$

Su subespacio propio asociado coincide con el núcleo de la matriz $\ker(-5I - B)$:

$$\ker(-5I - B) = \ker \begin{pmatrix} -3 & 4 & -2 \\ 2 & -6 & -2 \\ -4 & -2 & -10 \end{pmatrix} = L \left[\left(\begin{pmatrix} 2 \\ 1 \\ -1 \end{pmatrix} \right) \right] \quad (6.25)$$

6.3.3 Polinomio característico

Por el Corolario 3.4.6, el determinante de la matriz $\lambda I - A$ es cero, si y sólo si, $r(\lambda I - A) < n$, por lo que el determinante de esa matriz puede emplearse para calcular los valores propios de A :

$$|\lambda I - A| = \begin{vmatrix} \lambda - a_{11} & -a_{12} & \dots & -a_{1n} \\ -a_{21} & \lambda - a_{22} & \dots & -a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ -a_{n1} & -a_{n2} & \dots & \lambda - a_{nn} \end{vmatrix} = 0 \quad (6.26)$$

Los ceros del polinomio resultante de evaluar este determinante coinciden con sus autovalores. A este polinomio se le conoce como “polinomio característico de A ”.

Definición 6.3.2 – Polinomio característico de una matriz cuadrada

Se llama polinomio característico de $A \in \mathbb{K}^{n \times n}$, y se denota por χ_A , al polinomio de grado n definido como:

$$\chi_A(\lambda) = |\lambda I - A|$$

Ejercicio 6.3.2 † Calcular los subespacios propios de B, ecuación (6.10), asociados a $\lambda_2 = 3$ y $\lambda_3 = 6$.

Ejemplo 6.3.4 El polinomio característico de la matriz B, ecuación (6.10) del Ejemplo 6.2.3, es:

$$\begin{aligned}\chi_B(\lambda) &= |\lambda I - B| = \det \begin{pmatrix} \lambda + 2 & 4 & -2 \\ 2 & \lambda - 1 & -2 \\ -4 & -2 & \lambda - 5 \end{pmatrix} = \\ &= \lambda^3 - 4\lambda^2 - 27\lambda + 90\end{aligned}\quad (6.27)$$

Las raíces de $\chi_B(\lambda)$, que coinciden con sus valores propios, son $\lambda_1 = -5$, $\lambda_2 = 3$ y $\lambda_3 = 6$ †.

El polinomio característico χ_A de una matriz cuadrada $A \in \mathbb{K}^{n \times n}$ cumple estas propiedades:

Proposición 6.3.2 – Propiedades del polinomio característico

1. El coeficiente de grado $n - 1$ de χ_A coincide con el opuesto de la traza de A, o sea, con $-\text{tr}A = -(a_{11} + a_{22} + \dots + a_{nn})$.

$$\begin{aligned}\chi_A(\lambda) &= (\lambda - a_{11})(\lambda - a_{22}) \dots (\lambda - a_{nn}) + \dots = \\ &= \lambda^n - \text{tr}A \cdot \lambda^{n-1} + (\text{términos de grado menor que } n - 1)\end{aligned}\quad (6.28)$$

2. El término independiente de χ_A es igual a:

$$\chi_A(0) = \det(-A) = (-1)^n \det(A) \quad (6.29)$$

3. Dos matrices semejantes tienen el mismo polinomio característico.

Ejercicio 6.3.3 ‡ Demostrar, aplicando las fórmulas del determinante correspondientes, que el polinomio característico de una matriz cuadrada de orden dos o tres cumple las propiedades 1 y 2 de la Proposición 6.3.2.

Las propiedades 1 y 2 son fáciles de comprobar para los casos $n = 2$ y $n = 3$ (el caso $n = 1$ es trivial) desarrollando el determinante de $A - \lambda I$ en función de $\lambda \in \mathbb{K}$ y reordenando términos ‡, siendo A una matriz genérica con coeficientes en \mathbb{K} . Demostrar que toda matriz A, sea cual sea su orden, cumple estas dos propiedades es más engorroso y se deja fuera del alcance de este libro. El lector interesado encontrará demostraciones completas en otros manuales de álgebra lineal.

Demostración: (sólo de la propiedad 3)

Si $A, B \in \mathbb{K}^{n \times n}$ son semejantes, existe alguna matriz $P \in \mathbb{K}^{n \times n}$ invertible tal que $B = P^{-1}AP$, Definición 3.3.2. Sustituyendo B y aplicando la Definición 6.3.2 de polinomio característico y las propiedades del determinante, Proposición 2.3.2, se tiene:

$$\begin{aligned}\chi_B(\lambda) &= |\lambda I - B| = |\lambda I - P^{-1}AP| \\ &= |P^{-1}(\lambda I - A)P| = \\ &= |P^{-1}| \cdot |\lambda I - A| \cdot |P| \\ &= |\lambda I - A| = \chi_A(\lambda)\end{aligned}\quad (6.30)$$

□

Ejemplo 6.3.5 Los valores propios de la matriz B, ecuación (6.10) del Ejemplo 6.2.3, son $\lambda_1 = -5$, $\lambda_2 = 3$ y $\lambda_3 = 6$. El determinante de B es $\det(B) = -90$, por lo que el término independiente de $\chi_B(\lambda)$ es:

$$(-1)^n \cdot \det(B) = (-1)^3 \cdot (-90) = 90 \quad (6.31)$$

La traza de B es $\text{tr}(B) = 4$, por lo que el coeficiente del término de grado $n - 1$ es igual a:

$$-\text{tr}(B) = -4 \quad (6.32)$$

La igualdad de los polinomios característicos de matrices semejantes, punto 3 de la Proposición 6.3.2 junto con el punto 1 de la Proposición 6.3.2 y el punto 2 de la Proposición 6.3.2, implican lo siguiente:

Corolario 6.3.3 – Traza y determinante de matrices cuadradas semejantes

Dos matrices cuadradas semejantes tienen igual traza y determinante. Además, toda matriz A cuadrada cumple:

$$\det(A) = \lambda_1 \lambda_2 \dots \lambda_n \quad (6.33)$$

$$\text{tr}(A) = \lambda_1 + \lambda_2 + \dots + \lambda_n \quad (6.34)$$

Donde $\lambda_1, \lambda_2, \dots, \lambda_n$ son las n raíces del polinomio $\chi_A(\lambda)$.

Ejemplo 6.3.6 Los valores propios de la matriz B, ecuación (6.10) del Ejemplo 6.2.3, coinciden con las raíces de $\chi_B(\lambda)$ y son $\lambda_1 = -5$, $\lambda_2 = 3$ y $\lambda_3 = 6$. El producto de estas coincide con el determinante de B:

$$\det(B) = -90 = \lambda_1 \lambda_2 \lambda_3 \quad (6.35)$$

Mientras que la traza de B es:

$$\text{tr}(B) = 4 = \lambda_1 + \lambda_2 + \lambda_3 \quad (6.36)$$

Ejemplo 6.3.7 Como se vio en el Ejemplo 3.3.3 del apartado 3.3.4, las matrices:

$$A = \begin{pmatrix} 0 & 1 & -1 \\ 1 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix} \quad (6.37)$$

$$B = \begin{pmatrix} -1 & 0 & 0 \\ 2 & 0 & -1 \\ -1 & -1 & 0 \end{pmatrix} \quad (6.38)$$

Son semejantes. Por lo que se cumple † que $\chi_A(\lambda) = \chi_B(\lambda)$, $\det(A) = \det(B)$ y $\text{tr}(A) = \text{tr}(B)$.

Ejercicio 6.3.4 † Comprobar estas tres igualdades calculando las trazas, los determinantes y los polinomios característicos de ambas matrices.

6.3.4 Resumen de los últimos apartados

1. Familias de vectores propios de una misma matriz asociados a autovalores distintos dos a dos son siempre familias libres.
2. El conjunto de los vectores propios asociados a un mismo autovalor, junto con el vector cero, forman un subespacio llamado “subespacio característico” o “propio”.
3. Todos los valores propios de una matriz son raíces de su polinomio característico asociado.
4. Dos matrices semejantes tienen polinomios característicos idénticos, y como consecuencia, sus valores propios, trazas y determinantes coinciden.

6.4 Diagonalización de una matriz cuadrada.

Propiedades de la diagonalización.

Polinomios de matrices. Teorema de Cayley-Hamilton

6.4.1 Introducción

Si una matriz cuadrada $A \in \mathbb{K}^{n \times n}$ es semejante a una matriz diagonal D , la identidad $AP = PD$ mostrada en la ecuación (6.2) significa que la i -ésima columna de P es proporcional a la i -ésima columna de AP . En consecuencia, el i -ésimo elemento de la diagonal principal de D es un valor propio de A que tiene entre sus vectores propios asociados la i -ésima columna de P . Se tiene que las n columnas de P forman una familia libre, pues de lo contrario P no sería invertible, y componen una base de \mathbb{K}^n formada por vectores propios de A .

En esta sección, se enunciará la condición necesaria y suficiente que debe cumplir una matriz cuadrada A para que sea diagonalizable, y en el caso de que A lo sea, el modo en el que se construyen las matrices D y P tras haber calculado sus valores propios y sus subespacios característicos.

6.4.2 Multiplicidad geométrica y algebraica

Para estudiar si una matriz cuadrada es o no diagonalizable, es necesario conocer la multiplicidad geométrica y la multiplicidad algebraica de cada uno de sus autovalores. En este apartado, se definen ambos conceptos y se demuestran sus propiedades fundamentales.

Definición 6.4.1 — Multiplicidades algebraica y geométrica de un autovalor

Se dice que el valor propio λ de $A \in \mathbb{K}^{n \times n}$ tiene:

1. Multiplicidad algebraica igual a m si tiene multiplicidad m como raíz de χ_A .
2. Multiplicidad algebraica igual a σ si $\dim \ker(\lambda I - A) = \sigma$.

La multiplicidad geométrica de cada valor propio de una matriz cuadrada cualquiera no puede exceder nunca su multiplicidad algebraica, como se demostrará a continuación.

Proposición 6.4.1 — Acotación de las multiplicidades algebraica y geométrica de un autovalor

La multiplicidad geométrica de cada uno de los autovalores de una matriz A cuadrada es menor o igual que su multiplicidad algebraica correspondiente.

Dicho de otra forma, supóngase que el polinomio característico de $A \in \mathbb{K}^{n \times n}$ se factoriza como:

$$\chi_A(\lambda) = (\lambda - \lambda_1)^{m_1} \dots (\lambda - \lambda_r)^{m_r} \quad (6.39)$$

Siendo λ_i , $i = 1, \dots, r$ los $r \leq n$ valores propios distintos dos a dos de A , y m_i , $i = 1, \dots, r$ la multiplicidad algebraica de cada uno de ellos. Entonces, se verifica:

$$1 \leq \sigma_i \leq m_i \quad \forall i = 1, \dots, r \quad (6.40)$$

Demostración:

Sea $A \in \mathbb{K}^{n \times n}$ una matriz cuyos autovalores son $\lambda_1, \dots, \lambda_r$, $r \leq n$ con $\lambda_i \neq \lambda_j$ si $i \neq j$. Escógase su i -ésimo valor propio, λ_i , y constrúyase una matriz semejante a A que tenga λ_i entre sus valores propios, con una multiplicidad algebraica m_i igual o mayor a la dimensión σ_i de su subespacio propio asociado. Sea \mathcal{B} una base de \mathbb{K}^n cuyos σ_i primeros vectores formen una base del subespacio propio $\ker(\lambda_i I - A)$. Si P es una matriz cuyas columnas son los vectores de la base \mathcal{B} en el mismo orden, se tiene:

$$P^{-1}AP = \begin{pmatrix} \lambda_i I & B \\ 0 & C \end{pmatrix} \quad (6.41)$$

Donde I denota la matriz identidad de orden σ_i . Por la propiedad 3 de la Proposición 6.3.2, el polinomio característico de la matriz $P^{-1}AP$, que es igual a $(\lambda - \lambda_i)^{\sigma_i} \chi_C(\lambda)$, ha de coincidir con el de A . Si C no tuviese λ_i entre sus valores propios, las multiplicidades algebraica y geométrica de λ_i como autovalor de A coincidirían, es decir, $\sigma_i = m_i$. Si λ_i fuese valor propio también de C , la multiplicidad geométrica de λ_i como autovalor de A sería mayor que la algebraica, es decir, $\sigma_i > m_i$. En todo caso, σ_i no puede exceder nunca m_i . Repitiendo este razonamiento para cada $i = 1, \dots, r$ se demuestra la Proposición 6.4.1. □

Ejemplo 6.4.1 El polinomio característico de la matriz B , ecuación (6.10), es $\chi_B(\lambda) = \lambda^3 - 4\lambda^2 - 27\lambda + 90$. Las raíces de este polinomio son sus valores propios, $\lambda_1 = -5$, $\lambda_2 = 3$ y $\lambda_3 = 6$, todos con multiplicidad algebraica igual a uno. Los subespacios propios asociados a cada uno de estos autovalores son:

$$\begin{aligned} \ker(-5I - B) &= \ker \begin{pmatrix} -3 & 4 & -2 \\ 2 & -6 & -2 \\ -4 & -2 & -10 \end{pmatrix} = L \left[\left(\begin{pmatrix} 2 \\ 1 \\ -1 \end{pmatrix} \right) \right] = \\ &= L[(\bar{v}_1)] \end{aligned} \quad (6.42)$$

$$\begin{aligned} \ker(3I - B) &= \ker \begin{pmatrix} 5 & 4 & -2 \\ 2 & 2 & -2 \\ -4 & -2 & -2 \end{pmatrix} = L \left[\left(\begin{pmatrix} 2 \\ -3 \\ -1 \end{pmatrix} \right) \right] = \\ &= L[(\bar{v}_2)] \end{aligned} \quad (6.43)$$

$$\begin{aligned} \ker(6I - B) &= \ker \begin{pmatrix} 8 & 4 & -2 \\ 2 & 5 & -2 \\ -4 & -2 & 1 \end{pmatrix} = L \left[\left(\begin{pmatrix} 1 \\ 6 \\ 16 \end{pmatrix} \right) \right] = \\ &= L[(\bar{v}_3)] \end{aligned} \quad (6.44)$$

Todos estos subespacios tienen dimensión unidad, por lo que las multiplicidades geométricas de los tres valores propios son también igual a uno. Ninguna de ellas excede la multiplicidad algebraica de su autovalor correspondiente.

6.4.3 Caracterización de una matriz diagonalizable

Si todos los autovalores de $A \in \mathbb{K}^{n \times n}$ se encuentran en el cuerpo \mathbb{K} y tienen multiplicidad algebraica unidad, el número de valores propios pertenecientes a \mathbb{K} distintos dos a dos sería de n por el teorema fundamental del álgebra, Teorema C.4.1. A cada uno de los n autovalores se le puede asociar un vector propio. Por la Proposición 6.2.1, autovectores asociados a valores propios distintos son linealmente independientes, por lo que los n autovectores formarían una base de \mathbb{K}^n . Como se había visto en el apartado 6.1.3, toda matriz P que diagonaliza $A \in \mathbb{K}^{n \times n}$ está formada por n vectores propios de A linealmente independientes que forman una base de \mathbb{K}^n . Con este razonamiento, se obtiene una condición suficiente, pero no necesaria, para que una matriz cuadrada sea diagonalizable en \mathbb{K} .

Proposición 6.4.2 — Condición suficiente para que una matriz cuadrada sea diagonalizable

Si todos los valores propios de A pertenecen a \mathbb{K} y tienen multiplicidad algebraica unidad, la matriz es diagonalizable en \mathbb{K} .

Ejemplo 6.4.2 Todos los valores propios de la matriz B , ecuación (6.10) del Ejemplo 6.2.3, son reales y tienen multiplicidad algebraica unidad, por lo que es diagonalizable en \mathbb{R} .

Ejemplo 6.4.3 Los valores propios de la matriz real C , ecuación (6.14) del Ejemplo 6.2.4, son $\lambda_1 = 1 + i$, $\lambda_2 = 1 - i$ y $\lambda_3 = 1$. Dos de sus autovalores pertenecen a \mathbb{C} , por lo que la matriz no es diagonalizable en \mathbb{R} .

Supóngase ahora que A tiene valores propios con multiplicidad algebraica mayor que uno. Si, a pesar de ello, todos los autovalores pertenecen a \mathbb{K} y la suma de las multiplicidades geométricas es igual a n , sería posible encontrar n vectores propios en \mathbb{K}^n linealmente independientes entre sí de entre los subespacios propios de A , y estos formarían una base de \mathbb{K}^n . Con esos n vectores se podría construir una matriz P de cambio de base que diagonalizase A .

Si alguno de los valores propios de A tuviese una multiplicidad geométrica menor que la algebraica o no perteneciese al cuerpo \mathbb{K} , la suma de las multiplicidades geométricas de los autovalores pertenecientes a \mathbb{K} sería menor que n , no existiría ninguna base de \mathbb{K}^n compuesta por vectores propios de A , y no se podría construir una matriz P que diagonalizase A en \mathbb{K} .

En conclusión, la condición necesaria y suficiente que toda matriz cuadrada debe cumplir para ser diagonalizable es la que enuncia la siguiente proposición:

Proposición 6.4.3 – Condición necesaria y suficiente para que una matriz cuadrada sea diagonalizable

Una matriz A cuadrada de orden n es diagonalizable en \mathbb{K} si y sólo si la suma de las multiplicidades geométricas de los valores propios de A pertenecientes al cuerpo \mathbb{K} es igual a n .

Ejemplo 6.4.4 Una matriz P que diagonaliza la matriz B , ecuación (6.10) del Ejemplo 6.2.3, es:

$$P = (\bar{v}_1 | \bar{v}_2 | \bar{v}_3) = \begin{pmatrix} 2 & 2 & 1 \\ 1 & -3 & 6 \\ -1 & -1 & 16 \end{pmatrix} \quad (6.45)$$

Donde $\bar{v}_1, \bar{v}_2, \bar{v}_3$ son, respectivamente, vectores propios asociados a los autovalores de la matriz, ya calculados en el Ejemplo 6.4.1: $-5, 3$ y 6 . Las columnas de la matriz P forman una base de \mathbb{R}^3 , así que $r(P) = 3$ y esta es invertible.

Ejemplo 6.4.5 La matriz:

$$D = \begin{pmatrix} -1 & 2 \\ 0 & -1 \end{pmatrix} \quad (6.46)$$

Tiene el polinomio característico:

$$\chi_C(\lambda) = \det(\lambda I - C) = \det \begin{pmatrix} \lambda + 1 & 2 \\ 0 & \lambda + 1 \end{pmatrix} = (\lambda + 1)^2 \quad (6.47)$$

La única raíz de $\chi_C(\lambda)$ es $\lambda = -1$. Su multiplicidad algebraica es dos. El subespacio propio asociado a este único autovalor de C es:

$$\ker(-I - C) = \ker \begin{pmatrix} 0 & -2 \\ 0 & 0 \end{pmatrix} = L \left[\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix} \right) \right] \quad (6.48)$$

$\sigma = 1 < 2 = m$, por lo que C no es diagonalizable.

6.4.4 Propiedades de la diagonalización

Toda matriz $A \in \mathbb{K}^{n \times n}$ cuadrada cumple las siguientes propiedades en relación a sus valores y vectores propios:

Proposición 6.4.4 – Propiedades de la diagonalización de matrices

1. Una matriz cuadrada real puede no ser diagonalizable en \mathbb{R} y serlo en \mathbb{C} . Esto sucede cuando el espectro de la matriz contiene autovalores con parte imaginaria no nula, pero todos ellos, tanto los reales como los complejos, cumplen la condición de la Proposición 6.4.3.

2. Si A es una matriz real cuadrada de orden n , y $\lambda \in \mathbb{K}$ es valor propio de A , su conjugado $\lambda^* \in \mathbb{K}$ es también valor propio de A , y si $\bar{v} \in \mathbb{K}^n$ es un vector propio de A asociado a λ , $\bar{v}^* \in \mathbb{K}^n$ lo es respecto del autovalor λ^* pues, por la propiedad 3 de la Proposición C.5.1:

$$A\bar{v} = \lambda\bar{v} \Rightarrow (A\bar{v})^* = (\lambda\bar{v})^* \Rightarrow A\bar{v}^* = \lambda^*\bar{v}^* \quad (6.49)$$

Además, si A es una matriz real de orden impar, al menos uno de sus autovalores debe ser real.

3. Si A es una matriz real cuadrada de orden n , $\lambda \in \mathbb{K}$ es un valor propio de A , y $\bar{v} \in \mathbb{K}^n$ es un vector propio de A asociado a λ :

- a) $\lambda \in \mathbb{K}$ es un valor propio de A^t , pues:

$$\det(\lambda I - A) = \det(\lambda I - A)^t = \det(\lambda I - A^t) \quad (6.50)$$

- b) $\lambda^* \in \mathbb{K}$ es un valor propio de A^* y $\bar{v}^* \in \mathbb{K}^n$ es un vector propio de A^* asociado a λ^* , pues:

$$Ax = \lambda x \Rightarrow (Ax)^* = (\lambda x)^* \Rightarrow A^*x^* = \lambda^*x^* \quad (6.51)$$

- c) $\lambda^* \in \mathbb{K}$ es un valor propio de A^h , pues:

$$\det(\lambda I - A) = \det(\lambda I - A)^h = \det(\lambda^* I - A^h) \quad (6.52)$$

- d) $\lambda^{-1} \in \mathbb{K}$ es un valor propio de A^{-1} y $\bar{v} \in \mathbb{K}^n$ es un vector propio de A^{-1} asociado a λ^{-1} , pues:

$$Ax = \lambda x \Rightarrow x = A^{-1}\lambda x \Rightarrow \lambda^{-1}x = A^{-1}x \quad (6.53)$$

- e) $\lambda^k \in \mathbb{K}$ es un valor propio de A^k y $\bar{v} \in \mathbb{K}^n$ es un vector propio de A^k asociado a λ^k , pues:

$$Ax = \lambda x \Rightarrow A^k x = A^{k-1}\lambda x = A^{k-2}\lambda^2 x = \dots = \lambda^k x \quad (6.54)$$

4. Si $A \in \mathbb{K}^{n \times n}$ es diagonalizable, A^t , A^* , A^h , A^{-1} y $A^k \forall k \in \mathbb{N}$, $k \neq 0$, también lo son. En efecto:

- a) Las matrices que diagonalizan A^t son las inversas de las transpuestas de las matrices que diagonalizan A .

$$D = (P^{-1}AP)^t = P^t A^t (P^{-1})^t = P^t A^t (P^t)^{-1} \quad (6.55)$$

Donde se aplica que $(A^{-1})^t = (A^t)^{-1}$, ya que $(A^{-1})^t A^t = (A^{-1}A)^t = I \Rightarrow (A^{-1})^t = (A^t)^{-1}$.

- b) Las matrices que diagonalizan A^* son los conjugados de las matrices que diagonalizan A .

$$D^* = (P^{-1}AP)^* = (P^{-1})^* A^* P^* = (P^*)^{-1} A^* P^* \quad (6.56)$$

Donde se aplica que $(A^{-1})^* = (A^*)^{-1}$, ya que $(A^{-1})^* A^* = (A^{-1}A)^* = I \Rightarrow (A^{-1})^* = (A^*)^{-1}$.

c) Las matrices que diagonalizan A^h son las inversas de las transpuestas conjugadas de las matrices que diagonalizan A .

$$D^* = D^h = (P^{-1}AP)^h = P^h A^h (P^{-1})^h = \quad (6.57)$$

$$= P^h A^h (P^h)^{-1} \quad (6.58)$$

Donde se aplica que $(A^{-1})^h = (A^h)^{-1}$, ya que $(A^{-1})^h A^h = (A^{-1}A)^h = I \Rightarrow (A^{-1})^h = (A^h)^{-1}$.

d) Las matrices que diagonalizan A^{-1} son las mismas que diagonalizan A .

$$\text{diag}(\lambda_1^{-1}, \dots, \lambda_n^{-1}) = D^{-1} = (P^{-1}AP)^{-1} = \quad (6.59)$$

$$= P^{-1}A^{-1}P \quad (6.60)$$

e) Las matrices que diagonalizan A^k son las mismas que diagonalizan A .

$$\begin{aligned} \text{diag}(\lambda_1^k, \dots, \lambda_n^k) = D^k &= (P^{-1}AP)^k = \\ &= (P^{-1}AP)(P^{-1}AP) \dots (P^{-1}AP) = \\ &= P^{-1}A^k P \end{aligned} \quad (6.61)$$

Ejercicio 6.4.1 † Calcular los valores propios y los subespacios característicos de la matriz C .

Ejemplo 6.4.6 Los valores propios de la matriz real C , ecuación (6.14) del Ejemplo 6.2.4, son $\lambda_1 = 1 + i$, $\lambda_2 = 1 - i$ y $\lambda_3 = 1$. Esta matriz no es diagonalizable en \mathbb{R} , pero sí es diagonalizable en \mathbb{C} , ya que todos sus valores propios tienen multiplicidad algebraica unidad. Una matriz que diagonaliza C † es:

$$P = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 1 & 0 \\ -i & i & 0 \end{pmatrix} \quad (6.62)$$

Ejemplo 6.4.7 Los dos valores propios de la matriz real C del Ejemplo 6.2.4 con parte imaginaria no nula son uno el conjugado del otro: $\lambda_1 = \lambda_2^*$, mientras que su otro autovalor es real: $\lambda_3 \in \mathbb{R}$. Un vector propio asociado a $\lambda_1 = 1 + i$ es el $\bar{v}_1 = (0, 1, -i)^t$, y un vector propio asociado a $\lambda_2 = 1 - i = \lambda_1^*$ es $\bar{v}_2 = (0, 1, i)^t = \bar{v}_1^*$. Todos los vectores propios asociados al autovalor real $\lambda_3 = 1 \in \mathbb{R}$, como por ejemplo $\bar{v}_3 = (1, 0, 0)^t$, son reales.

6.4.5 Polinomios de matrices cuadradas

Al resultado de evaluar un determinado polinomio con coeficientes escalares sobre una matriz cuadrada se le conoce como “polinomio de una matriz”. Esta es su definición:

Definición 6.4.2 – Polinomio de una matriz cuadrada

Dados una matriz A cuadrada y un polinomio:

$$P(\lambda) = \sum_{i=0}^m a_i \lambda^i \quad (6.63)$$

con $\lambda \in \mathbb{K}$ y coeficientes en el cuerpo \mathbb{K} , se define la matriz $P(A)$ como la obtenida sustituyendo en la expresión del polinomio las potencias de λ por potencias de A . Es decir:

$$P(A) = \sum_{i=0}^m a_i A^i = a_0 I + a_1 A + a_2 A^2 + \cdots + a_m A^m \quad (6.64)$$

Considerando por convenio que $A^0 = I$.

Ejemplo 6.4.8 Se tiene la matriz B , ecuación (6.10) del Ejemplo 6.2.3, y el polinomio $P(\lambda) = \lambda^2 + 3\lambda + 2$. El resultado de evaluar el polinomio P sobre B es, de acuerdo con la Definición 6.4.2, la matriz siguiente:

$$P(B) = B^2 + 3B^1 + 2B^0 = B^2 + 3B + 2I = \begin{pmatrix} 16 & -4 & 4 \\ 4 & 18 & 14 \\ 20 & 2 & 54 \end{pmatrix}$$

Los valores y vectores propios de A y $P(A)$ están relacionados entre sí, para todo polinomio P , del modo que dictan las propiedades que se enumeran a continuación:

Proposición 6.4.5 – Propiedades del polinomio de una matriz cuadrada

Para todo polinomio P y matriz cuadrada A se cumplen:

1. Si λ_0 es valor propio de A , $P(\lambda_0)$ es valor propio de $P(A)$.
2. Los vectores propios de A son también vectores propios de $P(A)$.
3. Toda matriz P que diagonaliza A diagonaliza también $P(A)$. Dada la identidad $P^{-1}AP = D$, con D diagonal:

$$P^{-1}P(A)P = P(D)$$

Siendo $P(D)$ también diagonal.

Ejercicio 6.4.2 † Demostrar estas propiedades aplicando la Definición 6.4.2 del polinomio de una matriz cuadrada.

Se deja al lector † la demostración de las tres propiedades de la Proposición 6.4.5.

Ejercicio 6.4.3 † Calcular los subespacios propios de $P(B)$.

Ejemplo 6.4.9 Los valores propios de la matriz B , ecuación (6.10) del Ejemplo 6.2.3, son $\lambda_1 = -5$, $\lambda_2 = 3$ y $\lambda_3 = 6$. Los de $P(B)$, ecuación (6.4.8), son:

$$P(\lambda_1) = (-5)^2 + 3 \cdot (-5) + 2 = 12 \quad (6.65)$$

$$P(\lambda_2) = 3^2 + 3 \cdot 3 + 2 = 20 \quad (6.66)$$

$$P(\lambda_3) = 6^2 + 3 \cdot 6 + 2 = 56 \quad (6.67)$$

Puede comprobarse † que los subespacios propios de $P(B)$ asociados a $P(\lambda_1)$, $P(\lambda_2)$ y $P(\lambda_3)$ coinciden, respectivamente, con los subespacios propios de B asociados a λ_1 , λ_2 y λ_3 , por lo que cualquier matriz que diagonaliza B también diagonaliza $P(B)$.

6.4.6 Teorema de Cayley-Hamilton

Una propiedad muy interesante de las matrices cuadradas es que, para toda $A \in \mathbb{K}^{n \times n}$, el resultado de evaluar su polinomio característico $\chi_A(\lambda)$ sobre esta es siempre la matriz nula.

Teorema 6.4.6 — Teorema de Cayley-Hamilton

Toda matriz A cuadrada cumple:

$$\chi_A(A) = O$$

Observación 6.4.1 ‡ En Matemáticas, a un operador que hace cero un elemento determinado se le suele llamar “aniquilador”, o “anulador”, de dicho elemento. Por lo que también puede decirse que el operador aniquilador que caracteriza a toda matriz cuadrada es su polinomio característico.

En otras palabras, toda matriz A cuadrada satisface su propio polinomio característico ‡.

Demostración:

Se demostrará el teorema únicamente para las matrices diagonalizables. Dada una matriz $A \in \mathbb{K}^{n \times n}$ diagonalizable cuyo polinomio característico factorizado es:

$$\chi_A(\lambda) = (\lambda - \lambda_1)^{m_1} \dots (\lambda - \lambda_r)^{m_r}, \quad r \leq n, \quad \lambda_i \neq \lambda_j, \quad i \neq j \quad (6.68)$$

Evaluando χ_A sobre A , se obtiene:

$$\chi_A(A) = (A - \lambda_1 I)^{m_1} \dots (A - \lambda_r I)^{m_r} \quad (6.69)$$

Según la propiedad 3 de la Proposición 6.3.2, dos matrices semejantes tienen el mismo polinomio característico. La matriz diagonal D es semejante a A , por lo que se cumple:

$$\chi_D(\lambda) = (\lambda - \lambda_1)^{m_1} \dots (\lambda - \lambda_r)^{m_r} \Rightarrow \quad (6.70)$$

$$\Rightarrow \chi_D(D) = (D - \lambda_1 I)^{m_1} \dots (D - \lambda_r I)^{m_r} = P^{-1} \chi_A(A) P \quad (6.71)$$

Cada una de las matrices $D - \lambda_1 I$, $D - \lambda_2 I$, ..., $D - \lambda_r I$, así como sus m_i -ésimas potencias, es diagonal. Además, los coeficientes de sus diagonales principales correspondientes a cada uno de los autovalores $\lambda_1, \dots, \lambda_r$ se hacen cero al restarles $\lambda_1, \dots, \lambda_r$, por lo que son matrices nulas por bloques.

Multiplicando sucesivamente estas r matrices se obtiene la matriz nula, por lo que $\chi_D(\lambda) = 0$. Sustituyendo este resultado en la ecuación (6.71), puede verse que esta implica que $\chi_A(\lambda) = 0$.

□

La demostración del teorema de *Cayley-Hamilton* para matrices no diagonalizables requiere de la llamada “forma canónica de Jordan” de la matriz, que se encuentra fuera del alcance de este libro. El lector interesado encontrará la demostración completa del teorema en numerosos libros de álgebra.

Ejemplo 6.4.10 Las matrices B, C, D de los Ejemplos 6.2.3, 6.4.5 y 6.2.4 satisfacen su propio polinomio característico:

$$\chi_B(\lambda) = B^3 - 4B^2 - 27B + 90I = 0 \quad (6.72)$$

$$\chi_C(\lambda) = D^3 - 3D^2 + 4D - 2I = 0 \quad (6.73)$$

$$\chi_D(\lambda) = (C + I)^2 = 0 \quad (6.74)$$

6.4.7 Valores propios de las matrices hermiticas, antihermiticas y unitarias

Recuérdese que, por el punto 3 de la Definición 2.1.16, una matriz A es hermitica si cumple $A^h = A$. Supóngase que \bar{v} es valor propio de A asociado al valor propio λ . Aplicando la Definición 5.1.7 del producto escalar usual:

$$\langle A\bar{v}, \bar{v} \rangle = \bar{v}^h A\bar{v} = \lambda \bar{v}^h \bar{v} = \lambda \|\bar{v}\|^2 \quad (6.75)$$

$$\langle \bar{v}, A\bar{v} \rangle = (A\bar{v})^h \bar{v} = \bar{v}^h A^h \bar{v} = \bar{v}^h A\bar{v} = \lambda \bar{v}^h \bar{v} = \lambda \|\bar{v}\|^2 \quad (6.76)$$

Comparando las ecuaciones (6.75) y (6.76) y aplicando la propiedad 2, Proposición 5.1.5, de simetría hermitica del producto escalar usual, la positividad de la norma, punto 1 de la Definición 5.1.10, y la propiedad 3 de la Proposición C.5.1 del conjugado del producto de números complejos, se tiene:

$$\lambda \|\bar{v}\|^2 = (\lambda \|\bar{v}\|^2)^* = \lambda^* (\|\bar{v}\|^2)^* = \lambda^* \|\bar{v}\|^2 \quad (6.77)$$

Por tanto:

$$\lambda = \lambda^* \quad (6.78)$$

Condición que se cumple, si y solo si, λ tiene parte imaginaria nula: $\lambda \in \mathbb{R}$. Se acaba de demostrar lo siguiente:

Proposición 6.4.7 – Autovalores de una matriz hermitica

Si A es hermitica, todos sus valores propios son reales.

Por el punto 4 de la Definición 2.1.16, las matrices antihermiticas cumplen $A^h = -A$. Como en la demostración anterior, se supone que \bar{v} es valor propio de A asociado al valor propio λ y se aplica la Definición 5.1.7 del

producto escalar usual a $\langle A\bar{v}, \bar{v} \rangle$ y $\langle \bar{v}, A\bar{v} \rangle$, obteniéndose:

$$\langle A\bar{v}, \bar{v} \rangle = \bar{v}^h A\bar{v} = \lambda \bar{v}^h \bar{v} = \lambda \|\bar{v}\|^2 \quad (6.79)$$

$$\langle \bar{v}, A\bar{v} \rangle = (A\bar{v})^h \bar{v} = \bar{v}^h A^h \bar{v} = -\bar{v}^h A\bar{v} = -\lambda \bar{v}^h \bar{v} = -\lambda \|\bar{v}\|^2 \quad (6.80)$$

Por la propiedad 2 de simetría hermítica del producto escalar usual, Proposición 5.1.5, se tiene:

$$\lambda \|\bar{v}\|^2 = (-\lambda \|\bar{v}\|^2)^* = -\lambda^* (\|\bar{v}\|^2)^* = -\lambda^* \|\bar{v}\|^2 \quad (6.81)$$

Por tanto:

$$\lambda = -\lambda^* \quad (6.82)$$

Identidad que se cumple, si y sólo si, la parte real de λ es nula, por lo que los autovalores de una matriz antihermítica son imaginarios puros.

Proposición 6.4.8 – Autovalores de una matriz antihermítica

Si A es antihermítica, todos sus valores propios son imaginarios puros.

Por último, supóngase que la matriz cuadrada A es unitaria. Por la Definición 5.3.1 esta cumpliría $AA^h = A^h A = I$. Por la Proposición 5.3.1, una manera equivalente de definir las matrices unitarias de orden n es como el conjunto de las matrices cuadradas que cumplen, $\forall \bar{x} \in \mathbb{K}^n$:

$$\|\bar{x}\| = \|A\bar{x}\| \quad (6.83)$$

Si \bar{v} es vector propio de una matriz unitaria A asociado al autovalor λ , sustituyendo en la ecuación (6.83) y aplicando la Definición 6.2.1, punto 2, de vector propio, se cumple:

$$\|\bar{v}\| = \|A\bar{v}\| = \|\lambda\bar{v}\| = |\lambda| \cdot \|\bar{v}\| \quad (6.84)$$

De lo que se deduce que $|\lambda| = 1$. Es decir, todo autovalor λ de una matriz unitaria tiene módulo unidad:

Proposición 6.4.9 – Autovalores de una matriz unitaria

Si A es unitaria, todos sus valores propios tienen módulo unidad.

6.4.8 Diagonalización analítica de una matriz cuadrada

De todo lo visto hasta ahora sobre la diagonalización, se deduce el siguiente procedimiento para diagonalizar matrices cuadradas de manera analítica:

1. Se calculan las raíces $\lambda_1, \dots, \lambda_r$ del polinomio característico de A , $\chi_A(\lambda) = \det(\lambda I - A)$, y sus multiplicidades algebraicas. Estos son los valores propios de la matriz.
2. Se calculan las multiplicidades geométricas de los r valores propios A , $\dim \ker(\lambda_i I - A) \forall i = 1, \dots, r$. A es diagonalizable si y sólo si ambas multiplicidades coinciden para cada λ_i .
3. Si A resulta diagonalizable, se obtiene una base del subespacio propio asociado a cada autovalor, $\ker(\lambda_i I - A) \forall i = 1, \dots, r$.

4. La matriz D es una matriz diagonal que se puede denotar como $\text{diag}(\lambda_1, \dots, \lambda_r)$, donde cada λ_i se repite en D un número de veces igual a su multiplicidad. La columna i -ésima de P es un vector propio asociado al autovalor λ_i , $\forall i = 1, \dots, n$, de los contenidos en las bases calculadas en el paso anterior.

6.4.9 Resumen

1. Una matriz cuadrada es diagonalizable, si y sólo si, las multiplicidades algebraicas de todos sus valores propios coinciden con sus multiplicidades geométricas.
2. Si una matriz cuadrada es diagonalizable, también lo son su transpuesta, su conjugada, su transpuesta conjugada, su inversa, cualquier potencia de esta y el resultado de evaluar un polinomio sobre esta matriz.
3. Toda matriz cuadrada satisface su propio polinomio característico.
4. Las matrices hermiticas tienen todos sus valores propios reales. Las matrices antihermiticas tienen todos sus valores propios imaginarios puros. Los valores propios de las matrices unitarias tienen todos módulo unidad.

6.5 Diagonalización unitaria y ortogonal.

Teorema espectral. Descomposición espectral de una matriz

6.5.1 Introducción

Por un lado, vectores propios asociados a un mismo valor propio de $A \in \mathbb{K}^{n \times n}$ forman, junto con el vector cero, un subespacio llamado “característico” o “propio”, Definición 6.3.1, de modo tal que, por la Proposición 6.2.1, autovectores asociados a valores propios distintos son linealmente independientes entre sí.

Por otro lado, de acuerdo con el punto 2 de la Definición 6.2.1, cada vector propio \bar{v} de una matriz A cuadrada está asociado a un único valor propio de la matriz A . En consecuencia, la intersección de subespacios propios asociados a valores propios distintos produce el subespacio nulo de \mathbb{K}^n : $\forall i, j, i \neq j$

$$\ker(\lambda_i I - A) \cap \ker(\lambda_j I - A) = \{\bar{0}\} \quad (6.85)$$

Si además la matriz A es diagonalizable, la unión de bases de todos sus subespacios propios produce una base de \mathbb{K}^n . Por el teorema de *Grassmann*, esto implica que los r subespacios propios de A forman suma directa y que su suma es igual a \mathbb{K}^n :

$$\begin{aligned} \mathbb{K}^n &= \ker(\lambda_1 I - A) \oplus \ker(\lambda_2 I - A) \oplus \dots \oplus \ker(\lambda_r I - A) = \\ &= \bigoplus_{i=1}^r \ker(\lambda_i I - A) \end{aligned} \quad (6.86)$$

En esta sección se estudiará bajo qué condiciones los r subespacios propios asociados a una matriz cuadrada A diagonalizable son ortogonales dos a dos: $\forall i, j, i \neq j$

$$\ker(\lambda_i I - A) \perp \ker(\lambda_j I - A) \quad (6.87)$$

O dicho de otro modo, las condiciones que han de cumplirse para que vectores propios de una misma matriz A asociados a valores propios distintos dos a dos de una misma matriz A sean ortogonales entre sí: $\forall i, j, i \neq j$

$$\langle \bar{v}_i, \bar{v}_j \rangle = 0 \quad (6.88)$$

Si esto se cumpliera, sería posible encontrar n vectores propios de A que formen una base ortonormal de \mathbb{K}^n , y de este modo, construir una matriz de cambio de base unitaria, Definición 5.3.1, que diagonalice A :

$$U^{-1}AU = U^hAU = D \quad (6.89)$$

6.5.2 Semejanza y diagonalización unitarias

Definición 6.5.1 — Matrices unitariamente semejantes

Se dice que dos matrices cuadradas A, B son unitariamente semejantes si existe una matriz unitaria U tal que $B = U^hAU$.

Definición 6.5.2 – Matriz unitariamente diagonalizable

Se dice que la matriz $A \in \mathbb{K}^{n \times n}$ es unitariamente diagonalizable si existe una matriz unitaria $U \in \mathbb{K}^{n \times n}$ tal que $U^h A U$ es diagonal.

6.5.3 Semejanza y diagonalización ortogonales**Definición 6.5.3 – Matrices ortogonalmente semejantes**

Se dice que dos matrices cuadradas reales A, B son ortogonalmente semejantes † si existe una matriz ortogonal Q tal que $B = Q^t A Q$.

Observación 6.5.1 † Dos matrices ortogonalmente semejantes cumplen una relación de congruencia, Definición 3.3.3.

Definición 6.5.4 – Matriz ortogonalmente diagonalizable

Se dice que $A \in \mathbb{R}^{n \times n}$ es diagonalizable ortogonalmente si existe una matriz ortogonal $Q \in \mathbb{R}^{n \times n}$ tal que $Q^t A Q$ es diagonal.

La diagonalización ortogonal puede interpretarse como un caso particular de diagonalización unitaria en la que A es una matriz real, unitariamente diagonalizable y cuyos valores propios son reales. Los autovectores asociados a estas matrices tienen sus coeficientes reales y pueden formar bases ortonormales de \mathbb{R}^n .

6.5.4 El teorema espectral

Una matriz $A \in \mathbb{K}^{n \times n}$ hermítica cumple, de acuerdo al punto 3 de la Definición 2.1.16, $A^h = A$. Tómense dos valores propios distintos de A denotados λ_1, λ_2 y dos vectores propios \bar{v}_1, \bar{v}_2 asociados a cada autovalor. Por la definición 6.2.1 de valor y vector propio:

$$A \bar{v}_1 = \lambda_1 \bar{v}_1 \quad (6.90)$$

$$A \bar{v}_2 = \lambda_2 \bar{v}_2 \quad (6.91)$$

Si se desarrolla el producto $\lambda_1 \langle \bar{v}_1, \bar{v}_2 \rangle$ aplicando la Definición 5.1.7 del producto escalar usual y la Proposición 5.1.5, se obtiene:

$$\begin{aligned} \lambda_1 \langle \bar{v}_1, \bar{v}_2 \rangle &= \langle \lambda_1 \bar{v}_1, \bar{v}_2 \rangle = \langle A \bar{v}_1, \bar{v}_2 \rangle = \bar{v}_2^h A \bar{v}_1 \stackrel{(*)}{=} \bar{v}_2^h A^h \bar{v}_1 = \\ &= (A \bar{v}_2)^h \bar{v}_1 = \langle \bar{v}_1, A \bar{v}_2 \rangle = \langle \bar{v}_1, \lambda_2 \bar{v}_2 \rangle = \\ &= \lambda_2^* \langle \bar{v}_1, \bar{v}_2 \rangle \stackrel{(**)}{=} \lambda_2 \langle \bar{v}_1, \bar{v}_2 \rangle \end{aligned} \quad (6.92)$$

Demostrándose que $\lambda_1 \langle \bar{v}_1, \bar{v}_2 \rangle = \lambda_2 \langle \bar{v}_1, \bar{v}_2 \rangle$. En el paso (*) se ha considerado la hipótesis de hermiticidad de A , y en (**) se ha aplicado la Proposición 6.4.7 según la cual los valores propios de las matrices hermíticas son siempre reales.

Si $\lambda_1 \neq \lambda_2$, la identidad $\lambda_1 \langle \bar{v}_1, \bar{v}_2 \rangle = \lambda_2 \langle \bar{v}_1, \bar{v}_2 \rangle$ se cumple, si y sólo si, $\langle \bar{v}_1, \bar{v}_2 \rangle = 0$, demostrándose que vectores propios de una matriz hermítica A asociados a autovalores distintos son siempre ortogonales y que, por tanto, es posible construir una base ortogonal de \mathbb{K}^n compuesta por n vectores propios de A y una matriz unitaria que la diagonalice.

Teorema 6.5.1 – Teorema espectral para matrices hermíticas

Toda matriz hermítica es unitariamente diagonalizable.

El teorema anterior proporciona una condición suficiente, pero no necesaria, para que una matriz cuadrada sea unitariamente diagonalizable. De hecho, existen matrices no hermíticas que son unitariamente diagonalizables. La condición necesaria que ha de cumplir una matriz cuadrada para ser unitariamente diagonalizable se estudiará tras definir las llamadas “matrices normales”.

Definición 6.5.5 – Matrices normales

Se dice que la matriz $A \in \mathbb{K}^{n \times n}$ es normal si cumple la condición

$$A^h A = A A^h$$

Ejercicio 6.5.1 † Demostrar, aplicando la Definición 6.5.5 de matriz normal, que las matrices unitarias, hermíticas y antihermíticas, Definiciones 5.3.1 y 2.1.16, son matrices normales.

Se propone al lector † demostrar que las matrices unitarias, hermíticas y antihermíticas son matrices normales.

La identidad $A^h A = A A^h$, ecuación (6.5.5) proporciona una condición no sólo suficiente, sino también necesaria, para que una matriz cuadrada sea unitariamente diagonalizable:

Teorema 6.5.2 – Teorema espectral

Una matriz cuadrada es unitariamente diagonalizable, si y solo si, es normal.

Como consecuencia de este teorema, todas las matrices unitarias, hermíticas y antihermíticas son unitariamente diagonalizables, así como muchas otras que, sin ser unitarias, hermíticas o antihermíticas, cumplen la identidad (6.5.5).

Demostración:

Para demostrar este teorema se emplea la llamada “descomposición de Schur” de una matriz cuadrada. Aunque está fuera del alcance de este libro, se puede comprobar que toda matriz A cuadrada se puede factorizar de la forma:

$$A = U^h T U$$

Donde U es una matriz unitaria y T es una matriz triangular superior.

Multiplicando A por su adjunta y aplicando (6.5.4):

$$A A^h = U^h T U (U^h T U)^h = U^h T U U^h T^h U = U^h T T^h U \quad (6.93)$$

$$A^h A = (U^h T U)^h U^h T U = U^h T^t U U^h T U = U^h T^h T U \quad (6.94)$$

Si T es diagonal, $T T^h$ y $T^h T$ también lo son, y además, se cumple $T T^h = T^h T$, por lo que se cumple:

$$A A^h = U^h T T^h U = U^h T^h T U = A^h A \quad (6.95)$$

Demostrándose que toda matriz unitariamente diagonalizable es normal. Para demostrar el converso de esta proposición se razona así: supóngase

que $AA^h = A^hA$. Comparando la expresión (6.93) con la (6.94) se obtiene la identidad:

$$U^hTT^hU = U^hT^hTU \Rightarrow TT^h = T^hT \quad (6.96)$$

El cumplimiento de la identidad $TT^h = T^hT$ implica que T es diagonal \dagger de modo que toda matriz normal es unitariamente diagonalizable. Queda así demostrado el teorema espectral.

□

Ejercicio 6.5.2 \dagger Demostrar, aplicando las Definiciones 2.2.3 y 2.1.13 del producto matricial y de la transpuesta conjugada, que las únicas matrices triangulares que cumplen la identidad $AA^h = A^hA$ son las matrices diagonales.

Ejemplo 6.5.1 La matriz cuadrada real:

$$A = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix} \quad (6.97)$$

Es normal, pues cumple la identidad $A^hA = AA^h$. Sus autovalores son:

$$\lambda_1 = \frac{1}{2} + \frac{1}{2\sqrt{3}}i \quad (6.98)$$

$$\lambda_2 = \frac{1}{2} - \frac{1}{2\sqrt{3}}i \quad (6.99)$$

$$\lambda_3 = 2 \quad (6.100)$$

Todos tienen multiplicidad algebraica unidad, por lo que la matriz es diagonalizable en \mathbb{C} . Una matriz que la diagonaliza es la siguiente:

$$U = (\bar{u}_1 | \bar{u}_2 | \bar{u}_3) = \begin{pmatrix} 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \\ 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \\ -1/\sqrt{3} & -1/\sqrt{3} & 1/\sqrt{3} \end{pmatrix} \quad (6.101)$$

Donde \bar{u}_1 , \bar{u}_2 y \bar{u}_3 son vectores propios asociados a λ_1 , λ_2 y λ_3 respectivamente. Estos forman una familia ortonormal respecto al producto escalar usual en \mathbb{C}^3 y la norma canónica en \mathbb{C}^3 , por lo que U es una matriz unitaria, U^hAU es diagonal, y la matriz A es, en base a la Definición 6.5.2, una matriz unitariamente diagonalizable.

Por lo que respecta a las matrices cuadradas reales, como consecuencia del Teorema 6.5.2, las matrices ortogonales, simétricas y antisimétricas son todas ellas unitariamente ortogonales. A continuación, se estudiará qué condición o condiciones ha de cumplir una matriz cuadrada real para que sea diagonalizable ortogonalmente en el sentido de la Definición 6.5.4. Para ello, es necesaria la versión real de la descomposición de *Schur*, que asegura que toda matriz real cuadrada A puede factorizarse como:

$$A = Q^tTQ \quad (6.102)$$

Donde T es una matriz real triangular superior y Q es una matriz ortogonal. Supóngase que T es una matriz real diagonal. En ese caso, $T = T^t$ y se tiene:

$$A = Q^tTQ = Q^tT^tQ = (Q^tTQ)^t = A^t \quad (6.103)$$

Por lo que A sería una matriz real simétrica. Esto implica que toda matriz ortogonalmente diagonalizable es real simétrica. Para demostrar el

reverso de esta proposición, supóngase que A es simétrica. Entonces:

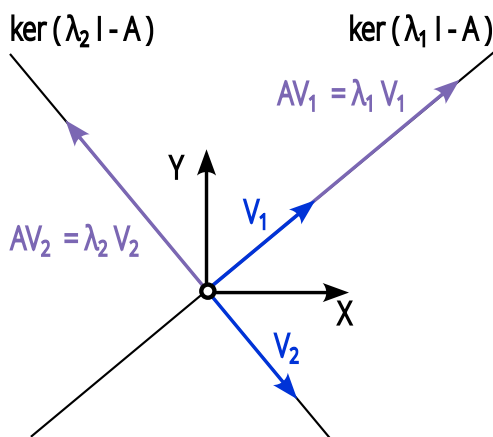
$$A = Q^t T Q \Rightarrow A^t = (Q^t T Q)^t = Q^t T^t Q = A \quad (6.104)$$

Lo que implica que $T = T^t$. Al ser T una matriz triangular, esta identidad sólo puede cumplirse si T es diagonal. Por tanto, toda matriz real simétrica es ortogonalmente diagonalizable. Se concluye que la simetría es una condición necesaria y suficiente para que una matriz real sea ortogonalmente diagonalizable.

Teorema 6.5.3 – Teorema espectral para matrices reales simétricas

Una matriz real cuadrada es unitariamente diagonalizable, si y sólo si, es simétrica.

Figura 6.3: Vectores pertenecientes a subespacios propios distintos de una misma matriz real simétrica A de orden dos, $\bar{v}_1 \in \ker(\lambda_1 I - A)$, $\bar{v}_2 \in \ker(\lambda_2 I - A)$, $\lambda_1 \neq \lambda_2$, son ortogonales entre sí. Además, la suma directa de ambos subespacios llena todo el espacio \mathbb{R}^2 , lo que permite construir una base ortonormal de \mathbb{R}^2 formada por vectores propios de A y una matriz ortogonal que la diagonalice.



Ejemplo 6.5.2 Los valores propios de la matriz real simétrica:

$$B = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \quad (6.105)$$

Son $\lambda_1 = -1$ con multiplicidad $m_1 = \sigma_1 = 2$, y $\lambda_2 = 2$ con multiplicidad $m_2 = \sigma_2 = 1$. Una matriz que la diagonaliza es:

$$Q = (\bar{q}_1 | \bar{q}_2 | \bar{q}_3) = \begin{pmatrix} -1/\sqrt{2} & -1/\sqrt{6} & 1/\sqrt{3} \\ 1/\sqrt{3} & -1/\sqrt{6} & 1/\sqrt{3} \\ 0 & 2/\sqrt{6} & 1/\sqrt{3} \end{pmatrix} \quad (6.106)$$

Donde $\bar{q}_1, \bar{q}_2, \bar{q}_3$ son autovectores asociados, respectivamente, a $\lambda_1, \lambda_1, \lambda_2$. Estos forman una familia ortonormal respecto al producto escalar en \mathbb{R}^3 y la norma euclidiana, por lo que Q es una matriz ortogonal, $Q^t A Q$ es diagonal, y A es, en base a la Definición 6.5.4, una matriz ortogonalmente diagonalizable.

6.5.5 Descomposición espectral de una matriz real simétrica

Acaba de demostrarse que a toda matriz $A \in \mathbb{R}^{n \times n}$ simétrica se le puede asociar una matriz ortogonal $Q = (\bar{q}_1 | \dots | \bar{q}_n)$ que la diagonalice, donde los \bar{q}_i son vectores propios unitarios de A que forman una base ortonormal.

mal de \mathbb{R}^n . Despejando A de la expresión $D = Q^t A Q$ y aplicando dos veces la propiedad 3 de la Proposición 2.2.5, se obtiene una expresión de la matriz A como combinación lineal de matrices de proyección ortogonal sobre los subespacios de dimensión uno que cada columna de Q genera:

$$D = Q^t A Q \Rightarrow A = Q \cdot \text{diag}(\lambda_1, \dots, \lambda_n) \cdot Q^t = \sum_{i=1}^n \lambda_i \bar{q}_i \bar{q}_i^t \quad (6.107)$$

El coeficiente escalar en cada término de la combinación lineal es el valor propio λ_i asociado a cada vector propio \bar{q}_i . Si se agrupan aquellos términos que comparten el mismo factor común λ_i , $i = 1 \dots, r$, se obtienen matrices de proyección ortogonal sobre cada uno de los subespacios propios de A :

$$A = \sum_{i=1}^r \left(\lambda_i \sum_{j=1}^{\sigma_i} \bar{q}_j \bar{q}_j^t \right) = \sum_{i=1}^r \lambda_i P_{\ker(\lambda_i I - A)} \quad (6.108)$$

Donde $P_{\ker(\lambda_i I - A)}$ denota la matriz de proyección ortogonal sobre el subespacio propio asociado a un autovalor λ_i cuya multiplicidad es σ_i , mientras que r representa el número de valores propios distintos de A .

Se concluye que toda matriz real simétrica puede descomponerse en una combinación lineal de las matrices de proyección ortogonal sobre sus subespacios propios.

Teorema 6.5.4 – Descomposición espectral de una matriz real simétrica

Toda matriz real simétrica A puede expresarse como:

$$A = \sum_{i=1}^r \lambda_i P_{\ker(\lambda_i I - A)}$$

Ejemplo 6.5.3 Los subespacios propios de la matriz:

$$B = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \quad (6.109)$$

Asociados a sus valores propios, $\lambda_1 = -1$ y $\lambda_2 = 2$, son:

$$\ker(-I - B) = L[(\bar{q}_1, \bar{q}_2)] = L \left[\left(\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ -1 \\ 2 \end{pmatrix} \right) \right] \quad (6.110)$$

$$\ker(2I - B) = L[(\bar{q}_3)] = L \left[\left(\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right) \right] \quad (6.111)$$

Las matrices de proyección ortogonal sobre cada subespacio son:

$$P_{\ker(-I-B)} = \bar{q}_1 \bar{q}_1^t + \bar{q}_2 \bar{q}_2^t = \frac{1}{3} \begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix} \quad (6.112)$$

$$P_{\ker(2I-B)} = \bar{q}_3 \bar{q}_3^t = \frac{1}{3} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \quad (6.113)$$

De modo que, en virtud del Teorema 6.5.4 de la descomposición espectral:

$$B = -P_{\ker(-I-B)} + 2 \cdot P_{\ker(2I-B)} = -\bar{q}_1 \bar{q}_1^t - \bar{q}_2 \bar{q}_2^t + 2 \cdot \bar{q}_3 \bar{q}_3^t \quad (6.114)$$

Multiplicando ambos miembros de la identidad (6.5.4) por $\bar{x} \in \mathbb{R}^n$, se obtiene la expresión:

$$A\bar{x} = \sum_{j=1}^r \lambda_j P_{\ker(\lambda_j I - A)} \bar{x} = \lambda_1 P_{\ker(\lambda_1 I - A)} \bar{x} + \dots + \lambda_r P_{\ker(\lambda_r I - A)} \bar{x} \quad (6.115)$$

Que implica que la imagen de un vector cualquiera de \mathbb{R}^n por una matriz A real simétrica de orden n puede descomponerse en combinación lineal de las proyecciones ortogonales de \bar{x} sobre los subespacios propios de A , con los valores propios de A como coeficientes de la combinación lineal.

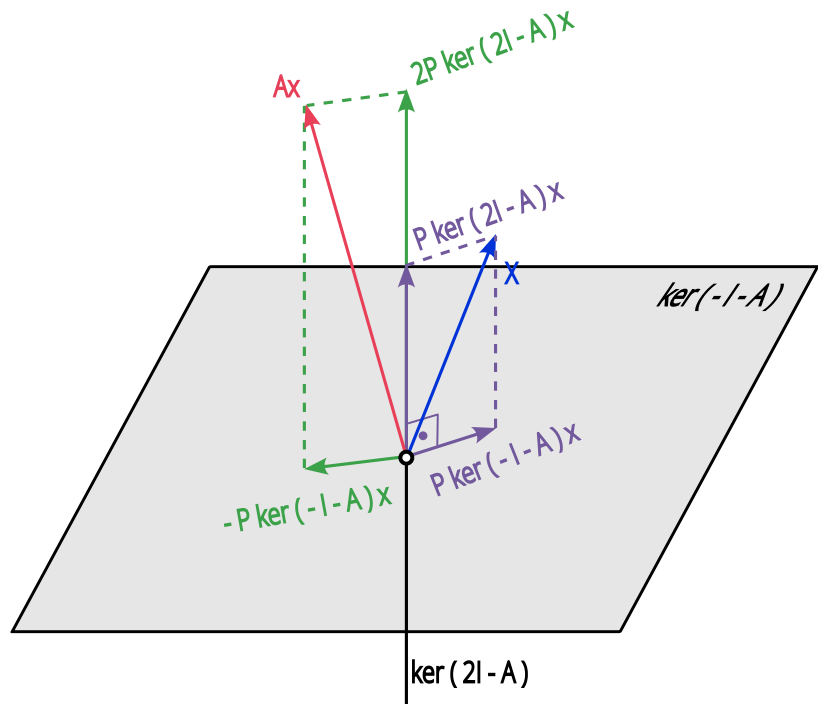


Figura 6.4: La imagen de $\bar{x} \in \mathbb{R}^3$ por la matriz real simétrica A del Ejemplo 6.5.4 puede escribirse como combinación lineal de las proyecciones ortogonales de \bar{x} sobre los subespacios propios de A , con coeficientes escalares iguales a los valores propios asociados a cada subespacio: $\lambda_1 = -1$ y $\lambda_2 = 2$.

Ejemplo 6.5.4 Como se ha visto en el Ejemplo 6.5.3:

$$\begin{aligned} B &= \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} = -P_{\ker(-I-B)} + 2 \cdot P_{\ker(2I-B)} = \\ &= -\bar{q}_1 \bar{q}_1^t - \bar{q}_2 \bar{q}_2^t + 2 \cdot \bar{q}_3 \bar{q}_3^t \end{aligned} \quad (6.116)$$

Para todo $\bar{x} \in \mathbb{R}^3$ se cumple:

$$B\bar{x} = -P_{\ker(-I-B)}\bar{x} + 2 \cdot P_{\ker(2I-B)}\bar{x} \quad (6.117)$$

Como ejemplo, se calcula la imagen por A de $\bar{x} = (1, 2, 1)^t$:

$$B \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} = -\frac{1}{3} \begin{pmatrix} -1 \\ 2 \\ -1 \end{pmatrix} + 2 \cdot \frac{1}{3} \begin{pmatrix} 4 \\ 4 \\ 4 \end{pmatrix} = \begin{pmatrix} 3 \\ 2 \\ 3 \end{pmatrix} \quad (6.118)$$

6.5.6 Descomposición espectral de una matriz normal

El Teorema 6.5.5 de descomposición espectral puede extenderse al conjunto de las matrices normales, Definición 6.5.5, sean estas reales o complejas. De acuerdo con teorema espectral, Teorema 6.5.2, una matriz cuadrada es normal, si y sólo si, es diagonalizable unitariamente. Es decir, si y sólo si, esta puede descomponerse de la forma $A = UDU^h$, con D diagonal y U unitaria. Las columnas de $U = (q_1 | \dots | q_n)$ forman una base ortonormal de \mathbb{K}^n compuesta por vectores propios de A. Esto último implica, por un razonamiento análogo al del apartado 6.5.5 anterior, que si A es normal, se puede descomponer en la suma de las matrices de proyección ortogonal sobre cada uno de sus subespacios propios, ponderada por los valores propios asociados a cada subespacio:

$$A = \sum_{i=1}^r \left(\lambda_i \sum_{j=1}^{\sigma_i} \bar{q}_j \bar{q}_j^h \right) = \sum_{i=1}^r \lambda_i P_{\ker(\lambda_i I - A)} \quad (6.119)$$

Lo que permite enunciar una versión más general del Teorema 6.5.2.

Teorema 6.5.5 – Descomposición espectral de una matriz normal

Toda matriz normal A puede expresarse como:

$$A = \sum_{i=1}^r \lambda_i P_{\ker(\lambda_i I - A)}$$

Por último, como consecuencia del teorema de descomposición espectral, al igual que sucedía con las matrices reales simétricas, la imagen por una matriz $A \in \mathbb{K}^{n \times n}$ normal de un vector $\bar{x} \in \mathbb{K}^n$ puede descomponerse en combinación lineal de las proyecciones ortogonales sobre los subespacios propios de la matriz, con coeficientes iguales a los autovalores asociados a cada subespacio, ecuación (6.115).

Observación 6.5.2 Es muy importante recordar que si un subespacio está contenido en \mathbb{C}^n , su matriz de proyección ortogonal asociada se formula mediante la transpuesta conjugada, denotada con el superíndice "h". No basta con sólo transponer, como sucede en el caso real.

6.5.7 Resumen

1. Una matriz cuadrada es unitariamente diagonalizable si existen matrices unitarias que la diagonalicen.
2. Una matriz cuadrada real es ortogonalmente diagonalizable si existen matrices ortogonales que la diagonalicen.
3. Una matriz cuadrada es unitariamente diagonalizable, si y sólo si, es una matriz normal.
4. Una matriz cuadrada real es ortogonalmente diagonalizable, si y sólo si, es simétrica.
5. Toda matriz normal puede descomponerse en una combinación lineal de matrices de proyección ortogonal sobre sus subespacios propios, con coeficientes iguales a sus valores propios correspondientes.



Formas bilineales y cuadráticas

7.1 Definiciones y propiedades de las formas bilineales y cuadráticas. Matriz de *Gram* de una forma bilineal o cuadrática

7.1.1 Introducción

Los polinomios con coeficientes y variables independientes reales P , Q y R definidos como:

$$P(x, y) = P(\bar{x}) = x^2 + 2y^2 \quad (7.1)$$

$$Q(x, y) = Q(\bar{x}) = x^2 + 2y^2 - 4xy \quad (7.2)$$

$$R(x, y, z) = R(\bar{x}) = x^2 - y^2 + z^2 + 2xy - 2xz - yz \quad (7.3)$$

Tienen todos sus términos de segundo grado. Polinomios como estos son expresables matricialmente mediante la forma $\bar{x}^t A \bar{x}$, donde $\bar{x} \in \mathbb{R}^n$ es un vector columna que contiene las variables independientes del polinomio, y $A \in \mathbb{R}^{n \times n}$ es una matriz real simétrica que contiene sus coeficientes escalares y que caracteriza unívocamente el polinomio.

A la función que asigna a cada n -upla de variables independientes el valor $\bar{x}^t A \bar{x}$ se le conoce como “forma cuadrática asociada a la matriz A ”. Las formas cuadráticas son casos particulares de una familia mayor de operadores llamados “formas bilineales” que se definirán al inicio de esta sección.

Más adelante, tras presentar la definición de forma cuadrática, se demostrará la relación existente entre las cotas máxima y mínima de una forma cuadrática, los vectores en los que se alcanzan, y los valores y vectores propios de la matriz A que caracteriza la forma cuadrática.

7.1.2 Definición y propiedades de las formas bilineales. Matrices de *Gram*

Definición 7.1.1 — Formas bilineales

Dado un espacio vectorial V sobre el cuerpo \mathbb{K} , se llama forma bilineal a una función $B : V \times V \mapsto \mathbb{K}$ que es lineal en cada uno de sus argumentos tomados por separado:

$$B(\lambda_1 x_1 + \lambda_2 x_2, y) = \lambda_1 B(x_1, y) + \lambda_2 B(x_2, y) \quad (7.4)$$

$$B(x, \lambda_1 y_1 + \lambda_2 y_2) = \lambda_1 B(x, y_1) + \lambda_2 B(x, y_2) \quad (7.5)$$

Donde x, y, x_1, x_2, y_1, y_2 son elementos en V , y $\lambda_1, \lambda_2 \in \mathbb{K}$.

7.1	Definición de forma bilineal y cuadrática . . .	225
7.2	Cocientes de <i>Rayleigh</i> . Clasificación de matrices reales simétricas	233
7.3	Clasificación y estudio de las formas cónicas . . .	240

Definición 7.1.2 – Formas bilineales simétricas

Se dice que una forma bilineal B definida en el espacio V es simétrica si esta cumple:

$$B(x, y) = B(y, x)$$

Para toda pareja de elementos $x, y \in V$.

Dados un espacio vectorial V de dimensión n y una forma bilineal $B(x, y)$ definida en V , tómesese una base del espacio V denotada $\mathcal{B} = (u_1, \dots, u_n)$ y evalúese B en todas las parejas posibles de elementos de la base \mathcal{B} obteniendo, $\forall i, j = 1, \dots, n$

$$a_{ij} = B(u_i, u_j) \in \mathbb{K} \quad (7.6)$$

A la matriz $A \in \mathbb{K}^{n \times n}$ formada por todos los coeficientes escalares a_{ij} obtenidos de esta manera se le llama “matriz de Gram de la forma bilineal B respecto a la base \mathcal{B} ”.

Definición 7.1.3 – Matriz de Gram de una forma bilineal

A la matriz $A \in \mathbb{K}^{n \times n}$ cuyos coeficientes, denotados a_{ij} , son las imágenes por la forma bilineal $B(x, y) : (x, y) \mapsto \mathbb{K}$ definida en V de todas las parejas posibles de elementos de una base $\mathcal{B} = (u_1, \dots, u_n)$ de V :

$$\forall i, j = 1, \dots, n, a_{ij} = B(u_i, u_j) \in \mathbb{K}$$

Sean $[x]^{\mathcal{B}} = (x_1, \dots, x_n)^t$ e $[y]^{\mathcal{B}} = (y_1, \dots, y_n)^t$ los vectores de coordenadas de una pareja de elementos cualesquiera de V respecto de la base \mathcal{B} , aplicando las Definiciones 1.3.6 y 1.3.7 se obtienen las descomposiciones siguientes:

$$x = x_1 u_1 + \dots + x_n u_n \quad (7.7)$$

$$y = y_1 u_1 + \dots + y_n u_n \quad (7.8)$$

Sustituyendo en $B(x, y)$ y aplicando la bilinealidad del operador B , ecuaciones (7.4) y (7.5), y la Definición 7.1.3 de matriz de Gram:

$$\begin{aligned} B(x, y) &= B(x_1 u_1 + \dots + x_n u_n, y_1 u_1 + \dots + y_n u_n) = \\ &= x_1 y_1 \cdot B(u_1, u_1) + \dots + x_1 y_n \cdot B(u_1, u_n) + \dots \\ &\dots + x_n y_1 \cdot B(u_n, u_1) + \dots + x_n y_n \cdot B(u_n, u_n) = \quad (7.9) \\ &= x_1 y_1 \cdot a_{11} + \dots + x_n y_n \cdot a_{nn} = \\ &= ([x]^{\mathcal{B}})^t A [y]^{\mathcal{B}} \end{aligned}$$

Se concluye que, dados un espacio V y una base \mathcal{B} de V , toda forma bilineal B en V puede definirse unívocamente mediante su matriz de Gram A asociada, y que la imagen por B de cualquier pareja de elementos $x, y \in V$ puede calcularse mediante la expresión:

$$B(x, y) = ([x]^{\mathcal{B}})^t A [y]^{\mathcal{B}} \quad (7.10)$$

Proposición 7.1.1 – Expresión de una forma bilineal mediante una matriz de Gram

Sean la forma bilineal B en el espacio vectorial V , una base \mathcal{B} de V y la matriz de Gram A asociada a B respecto de la base \mathcal{B} , si cumple:

$$B(x, y) = ([x]^{\mathcal{B}})^t A [y]^{\mathcal{B}}$$

Para toda pareja de elementos $x, y \in V$.

Ejemplo 7.1.1 Defínase la forma bilineal real $B_1(\bar{x}, \bar{y})$, $\bar{x}, \bar{y} \in \mathbb{R}^3$, como aquella cuya matriz de Gram asociada respecto de la base canónica de \mathbb{R}^3 es:

$$A = \begin{pmatrix} 0 & 1 & 2 \\ -2 & 1 & -1 \\ 3 & 5 & 0 \end{pmatrix} \quad (7.11)$$

B_1 puede expresarse como:

$$B_1(\bar{x}, \bar{y}) = \bar{x}^t A \bar{y} = (x_1, x_2, x_3) \begin{pmatrix} 0 & 1 & 2 \\ -2 & 1 & -1 \\ 3 & 5 & 0 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} \quad (7.12)$$

Ejemplo 7.1.2 Defínase la forma bilineal compleja $B_2(\bar{x}, \bar{y})$, $\bar{x}, \bar{y} \in \mathbb{C}^2$, como aquella cuya matriz de Gram asociada respecto de la base canónica de \mathbb{C}^2 es:

$$B = \begin{pmatrix} 2 & i \\ -i & -2 \end{pmatrix} \quad (7.13)$$

B_2 puede expresarse como:

$$B_2(\bar{x}, \bar{y}) = \bar{x}^t B \bar{y} = (x_1, x_2) \begin{pmatrix} 2 & i \\ -i & -2 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \quad (7.14)$$

Ejemplo 7.1.3 Defínase la forma bilineal real $B_3(\bar{x}, \bar{y})$, $\bar{x}, \bar{y} \in \mathbb{R}^3$, como aquella cuya matriz de Gram asociada respecto de la base canónica de \mathbb{R}^3 es:

$$C = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \quad (7.15)$$

C es una matriz real simétrica, por lo que B_3 puede expresarse como:

$$\begin{aligned} B_3(\bar{x}, \bar{y}) &= \bar{x}^t C \bar{y} = (x_1 \ x_2 \ x_3) \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \\ &= \bar{y}^t C \bar{x} = B_3(\bar{y}, \bar{x}) \end{aligned} \quad (7.16)$$

Por lo que $B_3(\bar{x}, \bar{y})$ es una forma bilineal simétrica.

En la siguiente proposición se enumeran dos propiedades fundamentales que cumplen las formas bilineales:

Proposición 7.1.2 — Propiedades de las formas bilineales

Toda forma bilineal $B : V \times V \mapsto \mathbb{K}$ cumple:

1. $B(x, 0) = B(0, y) = 0$.
2. $B(-x, y) = B(x, -y) = -B(x, y)$.

Para toda pareja de elementos x, y del espacio vectorial V ,

Demostración:

1. Por la propiedad 3 de la Proposición 1.1.3, $0 = 0 \cdot x_1 + 0 \cdot x_2 \forall x_1, x_2 \in V$, por lo que:

$$B(0, y) = 0 \cdot B(x_1, y) + 0 \cdot B(x_2, y) = 0 \quad (7.17)$$

$$B(x, 0) = 0 \cdot B(x, y_1) + 0 \cdot B(x, y_2) = 0 \quad (7.18)$$

2. Por la propiedad 4 de la Proposición 1.1.3, $-x = (-1) \cdot x \forall x \in V$, por lo que:

$$B(-x, y) = (-1) \cdot B(x, y) = -B(x, y) \quad (7.19)$$

$$B(x, -y) = (-1) \cdot B(x, y) = -B(x, y) \quad (7.20)$$

□

7.1.3 Cambios de base de una forma bilineal

Sea B una forma bilineal en un espacio vectorial V , por cada base de V se obtiene una matriz de *Gram* distinta que caracteriza B . Denótese $A^{\mathcal{B}}$ y $A^{\mathcal{B}'}$ a las dos matrices de *Gram* asociadas a una misma forma bilineal B respecto de las bases \mathcal{B} y \mathcal{B}' . Por la Proposición 7.1.1:

$$B(x, y) = ([x]^{\mathcal{B}})^t A^{\mathcal{B}} [y]^{\mathcal{B}} = ([x]^{\mathcal{B}'})^t A^{\mathcal{B}'} [y]^{\mathcal{B}'} \quad (7.21)$$

Llamando $P_{\mathcal{B}'}^{\mathcal{B}}$ a la matriz de cambio de base de \mathcal{B} a \mathcal{B}' , las coordenadas de dos elementos $x, y \in V$ respecto de cada base están relacionadas por las expresiones:

$$[x]^{\mathcal{B}} = P_{\mathcal{B}'}^{\mathcal{B}} [x]^{\mathcal{B}'} \quad (7.22)$$

$$[y]^{\mathcal{B}} = P_{\mathcal{B}'}^{\mathcal{B}} [y]^{\mathcal{B}'} \quad (7.23)$$

Sustituyendo (7.22) y (7.23) en la expresión (7.21) se obtiene:

$$\begin{aligned} B(x, y) &= (P_{\mathcal{B}'}^{\mathcal{B}} [x]^{\mathcal{B}'})^t A^{\mathcal{B}} P_{\mathcal{B}'}^{\mathcal{B}} [y]^{\mathcal{B}'} = \\ &= ([x]^{\mathcal{B}'})^t (P_{\mathcal{B}'}^{\mathcal{B}})^t A^{\mathcal{B}} P_{\mathcal{B}'}^{\mathcal{B}} [y]^{\mathcal{B}'} = ([x]^{\mathcal{B}'})^t A^{\mathcal{B}'} [y]^{\mathcal{B}'} \end{aligned} \quad (7.24)$$

De lo que se deduce que dos matrices de *Gram* asociadas a la misma forma bilineal respecto de bases distintas están asociadas entre si por la siguiente relación de congruencia, Definición 3.3.3:

$$A^{\mathcal{B}'} = (P_{\mathcal{B}'}^{\mathcal{B}})^t A^{\mathcal{B}} P_{\mathcal{B}'}^{\mathcal{B}} \quad (7.25)$$

Proposición 7.1.3 – Cambio de base de una forma bilineal

Dos matrices de *Gram* $A^{\mathcal{B}}$ y $A^{\mathcal{B}'}$ asociadas a la misma forma bilineal B respecto de las bases \mathcal{B} y \mathcal{B}' están relacionadas entre si mediante la siguiente relación de congruencia:

$$A^{\mathcal{B}'} = (P_{\mathcal{B}'}^{\mathcal{B}})^t A^{\mathcal{B}} P_{\mathcal{B}'}^{\mathcal{B}}$$

Donde $P_{\mathcal{B}'}^{\mathcal{B}}$ es la matriz de cambio de base de \mathcal{B} a \mathcal{B}' .

Ejemplo 7.1.4 Sean la forma bilineal real B_1 del Ejemplo 7.1.1 y la matriz de cambio de base de la base canónica a cierta base \mathcal{B}' :

$$P = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 1 \\ 2 & 0 & -1 \end{pmatrix} \quad (7.26)$$

La matriz de *Gram* de $B_1(\bar{x}, \bar{y})$ respecto de la base \mathcal{B}' es:

$$B = P^t B P = \begin{pmatrix} 10 & 11 & 3 \\ -4 & 1 & 4 \\ -11 & -5 & 3 \end{pmatrix} \quad (7.27)$$

Con esta nueva matriz de *Gram*, B_1 puede expresarse como:

$$B_1(\bar{x}, \bar{y}) = ([\bar{x}]^{\mathcal{B}'})^t B [\bar{y}]^{\mathcal{B}'} \quad (7.28)$$

7.1.4 Las formas bilineales en los espacios \mathbb{R}^n

De ahora en adelante, se supondrá que V es un espacio de vectores reales de n componentes \mathbb{R}^n y que $\bar{x}, \bar{y} \in \mathbb{R}^n$. Sea B una forma bilineal que asocia a cada pareja de vectores en \mathbb{R}^n un escalar real:

$$B(\bar{x}, \bar{y}) : \mathbb{R}^3 \times \mathbb{R}^3 \mapsto \mathbb{R} \quad (7.29)$$

Aplicando la Proposición 7.1.1, pueden calcularse los coeficientes $a_{ij} = B(\bar{e}_i, \bar{e}_j)$ de la matriz de *Gram* $A \in \mathbb{R}^{n \times n}$ de B respecto de la base canónica de \mathbb{R}^n , denotada $(\bar{e}_1, \dots, \bar{e}_n)$, obteniéndose así la siguiente expresión de la forma bilineal B :

$$B(\bar{x}, \bar{y}) = \bar{x}^t A \bar{y} \quad (7.30)$$

Si B fuese una forma bilineal simétrica, esta cumpliría, de acuerdo con la Definición 7.1.2, la propiedad $B(\bar{x}, \bar{y}) = B(\bar{y}, \bar{x})$, y en tal caso la matriz A sería una matriz real simétrica.

Ejemplo 7.1.5 Las formas bilineales B_1 y B_3 de los Ejemplos 7.1.1 y 7.1.3 son ambas formas bilineales en \mathbb{R}^3 . Estas asocian a cada pareja de vectores de tres componentes reales un escalar en el cuerpo de los números reales. Sus expresiones en función de sus matrices de *Gram* respecto de la base canónica de \mathbb{R}^3 vienen mostradas en los ejemplos citados.

7.1.5 Definición de forma cuadrática

Las formas bilineales simétricas en \mathbb{R}^n son funciones que asocian un escalar real a cada pareja de vectores de n componentes reales mediante el producto $\bar{x}^t A \bar{y}$, siendo A una matriz real simétrica. Si se hacen coincidir los vectores \bar{x}, \bar{y} evaluados de manera que $\bar{x} = \bar{y} \forall \bar{x} \in \mathbb{R}^n$, la forma bilineal se transforma en una función de una sola variable en \mathbb{R}^n denominada “forma cuadrática”:

Definición 7.1.4 — Formas cuadráticas

Se define la forma cuadrática asociada a la matriz real simétrica $A \in \mathbb{R}^{n \times n}$ como la función $Q(\bar{x}) : \mathbb{R}^n \mapsto \mathbb{R}$ definida como

$$Q(\bar{x}) = \bar{x}^t A \bar{x}$$

Ejemplo 7.1.6 La matriz real simétrica que define la forma bilineal $B_3(\bar{x}, \bar{y})$ del Ejemplo 7.1.3 respecto de la base canónica de \mathbb{R}^3 tiene asociada la forma cuadrática $Q(\bar{x})$ siguiente:

$$Q(\bar{x}) = \bar{x}^t C \bar{x} = (x_1, x_2, x_3) \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

Esta es una función que asocia un escalar real a cada vector de \mathbb{R}^3 .

Al contrario que las formas bilineales que, por la Definición 7.1.1, son lineales en cada una de sus componentes de entrada, las formas cuadráticas no se consideran como tales:

Proposición 7.1.4 Sea $Q : \mathbb{R}^n \mapsto \mathbb{R}$ una forma cuadrática, $\forall \bar{x} \in \mathbb{R}^n$, $\forall \alpha \in \mathbb{R}$, se cumple:

$$Q(\alpha \bar{x}) = \alpha^2 \cdot Q(\bar{x})$$

Demostración:

Por aplicación directa de la Definición 7.1.4, sustituyendo \bar{x} por $\alpha \bar{x} \forall \bar{x} \in \mathbb{R}^n$, $\forall \alpha \in \mathbb{R}$ y para toda matriz real simétrica $A \in \mathbb{R}^{n \times n}$.

□

7.1.6 Acotación de las formas cuadráticas

Dada una forma cuadrática $Q(\bar{x})$ asociada a la matriz real simétrica A , si $\lambda \in \mathbb{R}$ es un valor propio de A , y $\bar{v} \in \mathbb{R}^n$ es un vector propio de A asociado a λ , $Q(\bar{v})$ adopta el siguiente valor:

$$Q(\bar{v}) = \bar{v}^t A \bar{v} = \bar{v}^t \lambda \bar{v} = \lambda \bar{v}^t \bar{v} = \lambda \|\bar{v}\|^2 \quad (7.31)$$

Lo que permite enunciar esta proposición:

Proposición 7.1.5 — Evaluación de una forma cuadrática en un vector propio

Sea $Q : \mathbb{R}^n \mapsto \mathbb{R}$ una forma cuadrática caracterizada por la matriz real simétrica A , si $\bar{v} \in \mathbb{R}^n$ es un vector propio de A :

$$Q(\bar{v}) = \lambda \|\bar{v}\|^2$$

Siendo λ su valor propio asociado.

En lo siguiente, se utilizará este resultado para acotar superior e inferiormente las imágenes por la forma cuadrática $Q(\bar{x})$ de los vectores de \mathbb{R}^n en función de los valores propios de la matriz real simétrica A que lo caracteriza. Por el teorema espectral para matrices reales simétricas, Teorema 6.5.3, es posible encontrar una matriz ortogonal $Q \in \mathbb{R}^{n \times n}$ tal que:

$$Q^t A Q = D \quad (7.32)$$

Donde D es una matriz diagonal que contiene los valores propios de A ordenados de mayor a menor a mayor a lo largo de su diagonal principal:

$$D = \text{diag}(\lambda_{\min}, \dots, \lambda_{\max}) \quad (7.33)$$

Sustituyendo \bar{x} por $Q\bar{y}$, con $\bar{y} \in \mathbb{R}^n$, se obtiene:

$$\begin{aligned} \bar{x}^t A \bar{x} &= \bar{y}^t Q^t A Q \bar{y} = \bar{y}^t \cdot \text{diag}(\lambda_{\min}, \dots, \lambda_{\max}) \cdot \bar{y} = \\ &= \lambda_{\min} |y_1|^2 + \dots + \lambda_{\max} |y_n|^2 \end{aligned} \quad (7.34)$$

La suma $\lambda_{\min} |y_1|^2 + \dots + \lambda_{\max} |y_n|^2$ está acotada superiormente por esta otra: $\lambda_{\max} (|y_1|^2 + \dots + |y_n|^2)$ para todo $\bar{y} \in \mathbb{R}^n$, por lo que:

$$\bar{x}^t A \bar{x} \leq \lambda_{\max} (|y_1|^2 + \dots + |y_n|^2) \quad (7.35)$$

La matriz Q , al ser ortogonal, cumple la propiedad 3 de conservación de la norma euclidiana de los vectores transformados, Proposición 5.3.3, es decir, $\|\bar{y}\| = \|Q\bar{y}\| = \|\bar{x}\| \forall \bar{x} \in \mathbb{R}^n$. Por lo que la identidad (7.35) puede reescribirse como:

$$\bar{x}^t A \bar{x} \leq \lambda_{\max} \|\bar{y}\|^2 = \lambda_{\max} \|\bar{x}\|^2 \quad (7.36)$$

Así pues, $\bar{x}^t A \bar{x} \leq \lambda_{\max} \|\bar{x}\|^2 \forall \bar{x} \in \mathbb{R}^n$. En virtud de la Proposición 7.1.5, la cota superior $\lambda_{\max} \|\bar{x}\|^2$ se alcanza, si y sólo si, \bar{x} es vector propio de A asociado a su máximo autovalor. De manera análoga, a suma $\lambda_{\min} |y_1|^2 + \dots + \lambda_{\max} |y_n|^2$ está acotada inferiormente por esta otra: $\lambda_{\min} (|y_1|^2 + \dots + |y_n|^2)$, lo que lleva a la identidad:

$$\bar{x}^t A \bar{x} \geq \lambda_{\min} \|\bar{x}\|^2 \forall \bar{x} \in \mathbb{R}^n \quad (7.37)$$

Alcanzándose la cota inferior $\lambda_{\min} \|\bar{x}\|^2$, si y sólo si, \bar{x} es vector propio de A asociado a su mínimo autovalor.

Se acaba de demostrar el siguiente teorema:

Teorema 7.1.6 — Acotación de las formas cuadráticas

Sean λ_{\min} y λ_{\max} los autovalores mínimo y máximo de la matriz real simétrica $A \in \mathbb{R}^{n \times n}$, para todo $\bar{x} \in \mathbb{R}^n$ se verifica:

$$\lambda_{\min} \|\bar{x}\|^2 \leq \bar{x}^t A \bar{x} \leq \lambda_{\max} \|\bar{x}\|^2$$

Alcanzándose las cotas inferior y superior, si y sólo si, \bar{x} es vector propio asociado a λ_{\min} y λ_{\max} respectivamente.

Ejemplo 7.1.7 Los valores propios de la matriz C del Ejemplo 7.1.3 son $\lambda_1 = -1$ y $\lambda_2 = 2$. Un vector propio asociado a λ_1 es $\bar{v}_1 = 1/\sqrt{2} \cdot (-1, 1, 0)^t$. Un vector propio asociado a λ_2 es $\bar{v}_2 = 1/\sqrt{3} \cdot (1, 1, 1)^t$. Ambos son unitarios respecto de la norma euclidiana. Véase que:

$$\bar{v}_1^t C \bar{v}_1 = \lambda_1 \|\bar{v}_1\|^2 = -1 \cdot 1 = -1 \quad (7.38)$$

$$\bar{v}_2^t C \bar{v}_2 = \lambda_2 \|\bar{v}_2\|^2 = 2 \cdot 1 = 2 \quad (7.39)$$

Además, $\forall \bar{x} \in \mathbb{R}^3$:

$$-\|\bar{x}\|^2 \leq \bar{x}^t C \bar{x} \leq 2 \cdot \|\bar{x}\|^2 \quad (7.40)$$

Y si \bar{x} tiene norma euclidiana unidad:

$$-1 \leq \bar{x}^t C \bar{x} \leq 2 \quad (7.41)$$

7.2 Cocientes de Rayleigh. Clasificación de matrices reales simétricas. Criterio de Sylvester

7.2.1 Introducción

A toda matriz real simétrica se le puede asociar un operador llamado “cociente de Rayleigh” que se diferencia de las formas cuadrática únicamente en la presencia de un factor que normaliza el vector de entrada. Tras definirlo formalmente, se mostrará que el cociente de Rayleigh asociado a una matriz real simétrica cualquiera se encuentra acotado superior e inferiormente por los autovalores máximo y mínimo de la matriz que lo define. Finalmente, se clasificarán las matrices reales simétricas en función de los valores que acotan sus formas cuadráticas asociadas y se propondrá un criterio para comprobar si una matriz real simétrica es definida positiva.

7.2.2 Definición de cociente de Rayleigh

Definición 7.2.1 – Cocientes de Rayleigh

Se llama cociente de Rayleigh asociado a la matriz real simétrica $A \in \mathbb{R}^{n \times n}$ a la función $R : \mathbb{R}^n \setminus \{0\} \mapsto \mathbb{R}$ † definida, para todo $\bar{x} \neq \bar{0}$, como:

$$R(\bar{x}) = \frac{\bar{x}^t A \bar{x}}{\|\bar{x}\|^2}$$

Claramente, al igual que las formas cuadráticas, Proposición 7.1.4, los cocientes de Rayleigh no son funciones lineales en $\bar{x} \in \mathbb{R}^n$.

Ejemplo 7.2.1 El cociente de Rayleigh $R(\bar{x})$ cuya matriz característica es la matriz C del Ejemplo 7.1.3 puede expresarse como:

$$R(\bar{x}) = \frac{1}{\|\bar{x}\|^2} \cdot (x_1, x_2, x_3) \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

Donde $\|\bar{x}\|^2 = x_1^2 + x_2^2 + x_3^2$.

Observación 7.2.1 † La notación $\mathbb{R}^n \setminus \{0\}$ puede leerse como “el conjunto formado por todos los vectores de \mathbb{R}^n a excepción del vector nulo”, donde “\” representa la diferencia de conjuntos, Definición A.2.6. El vector nulo se excluye del dominio de los cocientes de Rayleigh porque el cociente de un escalar por cero es una operación no definida.

7.2.3 Acotación de los cocientes de Rayleigh

Obsérvese que si \bar{v} es un vector propio de A asociado al autovalor λ se tiene:

$$R(\bar{v}) = \frac{\bar{v}^t A \bar{v}}{\|\bar{v}\|^2} = \frac{\bar{v}^t \lambda \bar{v}}{\|\bar{v}\|^2} = \frac{\lambda \|\bar{v}\|^2}{\|\bar{v}\|^2} = \lambda \tag{7.42}$$

Lo que permite enunciar la siguiente proposición:

Proposición 7.2.1 — Evaluación de una forma cuadrática en un vector propio

Sea $R : \mathbb{R}^n \mapsto \mathbb{R}$ un cociente de *Rayleigh* caracterizada por la matriz real simétrica A , si $\bar{v} \in \mathbb{R}^n$ es un vector propio de A :

$$R(\bar{v}) = \lambda$$

Siendo λ su valor propio asociado.

Ejercicio 7.2.1 † Calcular los valores propios y los subespacios característicos de la matriz C que caracteriza el cociente de *Rayleigh* definido por la ecuación 7.2.1.

Ejemplo 7.2.2 Uno de los valores propios de la matriz que caracteriza el cociente de *Rayleigh* $R(\bar{x})$ definido en el Ejemplo 7.2.1 es $\lambda = 2$ †. Un vector propio de A asociado a este autovalor es $\bar{v} = (1, 1, 1)^t$. Evaluando $R(\bar{x})$ en \bar{v} se obtiene:

$$R(\bar{v}) = \frac{\bar{v}^t A \bar{v}}{\|\bar{v}\|^2} = \frac{6}{3} = 2 = \lambda$$

El Teorema 7.1.6 de acotación de las formas cuadráticas garantiza que $\forall \bar{x} \neq \bar{0}$ se cumple:

$$\lambda_{\min} \leq \frac{\bar{x}^t A \bar{x}}{\|\bar{x}\|^2} \leq \lambda_{\max} \quad (7.43)$$

Y la continuidad de la función R en su dominio garantiza que R alcanza todos los valores del intervalo cerrado $[\lambda_{\min}, \lambda_{\max}] \subset \mathbb{R}$, o dicho en otras palabras, que para todo $y \in [\lambda_{\min}, \lambda_{\max}]$ existe al menos un $\bar{x} \in \mathbb{R}^n$ tal que $R(\bar{x}) = y$.

Teorema 7.2.2 — Acotación de los cocientes de *Rayleigh*

Sean λ_{\min} y λ_{\max} los autovalores mínimo y máximo de la matriz real simétrica $A \in \mathbb{R}^{n \times n}$, para todo $\bar{x} \in \mathbb{R}^n$ se verifica:

$$\lambda_{\min} \leq \frac{\bar{x}^t A \bar{x}}{\|\bar{x}\|^2} \leq \lambda_{\max}$$

Alcanzándose las cotas inferior y superior si \bar{x} es vector propio asociado a λ_{\min} y λ_{\max} respectivamente.

Ejemplo 7.2.3 Los valores propios de la C del Ejemplo 7.1.3 son -1 y 2 , por lo que las imágenes de su cociente de *Rayleigh* asociado están acotadas por la identidad siguiente para todo $\bar{x} \neq \bar{0}$:

$$-1 \leq R(\bar{x}) \leq 2$$

7.2.4 Clasificación de matrices reales simétricas

Los signos y valores de las imágenes de una forma cuadrática (o de un cociente de *Rayleigh*) pueden adoptar en función de $\bar{x} \in \mathbb{R}^n$ vienen deter-

minados, en base al Teorema 7.1.6, por los de los autovalores de la matriz simétrica que la define. Debido a la presencia de las formas cuadráticas y los cocientes de Rayleigh en tantos modelos de la física y en algunos algoritmos numéricos y matemáticos, resulta de interés clasificar las matrices reales simétricas, y por extensión al caso complejo, las matrices hermiticas (cuyos valores propios son, por la Proposición 6.4.7, siempre reales) en base a los signos de sus valores propios, así como a la existencia o no de autovalores nulos.

Definición 7.2.2 – Clasificación de matrices reales simétricas

Dada una matriz A real simétrica, se dice que:

1. A es semidefinida positiva si $\bar{x}^t A \bar{x} \geq 0 \forall \bar{x} \in \mathbb{R}^n$. Es decir, si sus autovalores sean todos positivos o nulos ($\lambda_k \geq 0$).
 - a) En particular, A es definida positiva si $\bar{x}^t A \bar{x} > 0 \forall \bar{x} \in \mathbb{R}^n, \bar{x} \neq \bar{0}$. Es decir, si sus autovalores son todos estrictamente positivos ($\lambda_k > 0$).
2. A es semidefinida negativa si $\bar{x}^t A \bar{x} \leq 0 \forall \bar{x} \in \mathbb{R}^n$. Es decir, si sus autovalores sean todos negativos o nulos ($\lambda_k \leq 0$).
 - a) En particular, A es definida negativa si $\bar{x}^t A \bar{x} < 0 \forall \bar{x} \in \mathbb{R}^n, \bar{x} \neq \bar{0}$. Es decir, si sus autovalores sean todos estrictamente negativos ($\lambda_k < 0$).
3. A es indefinida si existen vectores $\bar{x}, \bar{y} \in \mathbb{R}^n$ tales que $\bar{x}^t A \bar{x} > 0$ e $\bar{y}^t A \bar{y} < 0$. Es decir, si A posee autovalores tanto positivos como negativos.

Ejemplo 7.2.4 Los valores propios de la matriz C del Ejemplo 7.1.3 son -1 y 2 . A tiene autovalores positivos y negativos. Por el punto 3 de la Definición 7.2.2, es indefinida.

7.2.5 El criterio de Sylvester

A continuación, se demostrará una condición necesaria y suficiente para que una matriz real simétrica sea definida positiva. Esta condición puede aplicarse para reconocer aquellas matrices reales simétricas que son definidas positivas o definidas negativas sin necesidad de calcular explícitamente sus valores propios. Con un pequeño cambio en la condición que más adelante se justificará, el criterio de Sylvester podrá ser también aplicado para identificar matrices reales simétricas definidas negativas.

Sea A_i la matriz cuadrada de $\mathbb{R}^{i \times i}$ cuyas filas y columnas coinciden con las i primeras filas y columnas de A . Si A es definida positiva, también lo es cada una de las i submatrices de A denotadas como $A_i, 1 \leq i \leq n$, puesto que, para todo vector $\bar{x}_i = (x_1, \dots, x_i) \in \mathbb{R}^i$ no nulo, aplicando la Definición 7.2.2 de matriz definida positiva, punto 1a, se tiene que:

$$\bar{x}_i^t A_i \bar{x}_i = (x_1, \dots, x_i, 0 \dots 0) A (x_1, \dots, x_i, 0, \dots 0)^t > 0 \quad (7.44)$$

Pues $\bar{x}^t A \bar{x} > 0 \forall \bar{x} \neq \bar{0}$.

Como indica el Corolario 6.3.3, el producto de los autovalores de una matriz cuadrada coincide con su determinante. Esta propiedad implica

que si se cumple (7.44) $\forall i = 1, \dots, n$, todos sus valores propios serían positivos, por lo que A sería definida positiva.

El reverso de esta proposición puede demostrarse por inducción: el caso $i = 1$ es trivial. Ahora, supóngase que A_i es definida positiva, en cuyo caso $\det(A_i) > 0$ y $\det(A_j) > 0 \forall j = 1, \dots, i$; y supóngase también que $\det(A_{i+1}) > 0$. La matriz A_{i+1} puede escribirse en función de A_i como una matriz por bloques:

$$A_{i+1} = \begin{pmatrix} A_i & \bar{v}_i \\ \bar{v}_i^t & d \end{pmatrix} \quad (7.45)$$

Donde d es una constante real, y \bar{v}_i es un vector de i componentes reales. Sea \bar{x}_{i+1} un vector de $i+1$ componentes reales cuyos i primeros coeficientes forman el subvector \bar{x}_i de i vectores reales y cuyo último componente se denota como $x_{i+1} \in \mathbb{R}$. Es decir:

$$\bar{x}_{i+1} = (\bar{x}_i, x_{i+1})^t \quad (7.46)$$

La forma cuadrática $\bar{x}_{i+1}^t A_{i+1} \bar{x}_{i+1}$ puede desarrollarse como:

$$\begin{aligned} \bar{x}_{i+1}^t A_{i+1} \bar{x}_{i+1} &= \bar{x}_i^t A_i \bar{x}_i + x_{i+1} \bar{x}_i^t \bar{v}_i + x_{i+1} \bar{v}_i^t \bar{x}_i + d \cdot |x_{i+1}|^2 = \\ &= (\bar{x}_i + c)^t A_i (\bar{x}_i + c) + |x_{i+1}|^2 \cdot (d - \bar{v}_i^t A_i^{-1} \bar{v}_i) \end{aligned} \quad (7.47)$$

Donde $c = \bar{x}_{i+1} A_i^{-1} \bar{v}_i$.

El determinante de una matriz cuadrada por bloques puede calcularse mediante la siguiente fórmula:

$$\det \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \det(A) \cdot \det(D - CA^{-1}B) \quad (7.48)$$

Aplicando (7.48) a la matriz A_{i+1} , ecuación (7.45), se obtiene lo siguiente:

$$\det(A_{i+1}) = \det(A_i) \cdot \det(d - \bar{v}_i^t A_i^{-1} \bar{v}_i) \quad (7.49)$$

La positividad de $\det(A_i)$ y la hipótesis de que $\det(A_{i+1}) > 0$ implican:

$$d - \bar{v}_i^t A_i^{-1} \bar{v}_i > 0 \quad (7.50)$$

Que a su vez implica la positividad de la forma cuadrática desarrollada en la ecuación (7.47):

$$\bar{x}_{i+1}^t A_{i+1} \bar{x}_{i+1} > 0 \quad (7.51)$$

Y así, en virtud del punto 1a de la Definición 7.2.2, se concluye que A_{i+1} es una matriz definida positiva, demostrándose que toda matriz real simétrica que cumpla $\det(A_i) > 0 \forall i = 1, 2, \dots, n$ es definida positiva.

Teorema 7.2.3 – Criterio de Sylvester

Una matriz real simétrica $A \in \mathbb{R}^{n \times n}$ es definida positiva, si y sólo si, $\det(A_i) > 0$ para todo $i = 1, \dots, n$, donde:

$$A_i = \begin{pmatrix} a_{11} & \cdots & a_{1j} \\ \vdots & \ddots & \vdots \\ a_{1j} & \cdots & a_{jj} \end{pmatrix}$$

Si se aplica la propiedad 1 de la Proposición 6.4.5 a una matriz A cuadrada, se tiene que si λ es valor propio de A , $-\lambda$ es valor propio de $-A$, por lo que si la matriz real simétrica A es definida positiva, $-A$ sería definida negativa. Esto permite enunciar una versión del criterio de *Sylvester* para matrices definidas negativas:

Corolario 7.2.4 – Criterio de *Sylvester* para matrices definidas negativas

Una matriz real simétrica $A \in \mathbb{R}^{n \times n}$ es definida negativa, si y sólo si, $\det(A_i) < 0$ para todo $i = 1, \dots, n$.

Ejemplo 7.2.5 Los menores de la matriz:

$$B = \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (7.52)$$

Son:

$$\begin{aligned} |B_1| &= \det(2) = 2 > 0 \\ |B_2| &= \det \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} = 3 > 0 \\ |B_3| &= \det(B) = 3 > 0 \end{aligned} \quad (7.53)$$

Por lo que B es definida positiva. De hecho, sus dos valores propios son números reales positivos: $\lambda_1 = 1$ y $\lambda_2 = 5$.

7.2.6 Una aplicación práctica del cociente de *Rayleigh*: el método de la potencia

En este apartado, se describirá un algoritmo iterativo para el cálculo de valores y vectores propios muy sencillo de implementar numéricamente en un ordenador y que está basado en el cociente de *Rayleigh* de una matriz cuadrada no necesariamente simétrica ni real. Por lo razonado en el apartado 6.4.3, si una matriz A cuadrada de orden n es diagonalizable, es posible construir una base \mathcal{B} de \mathbb{K}^n formada por n vectores propios de A denotados $\bar{u}_1, \bar{u}_2, \dots, \bar{u}_n$. Los vectores de la base se ordenan de modo que $|\lambda_1| \geq |\lambda_2| \geq \dots \geq |\lambda_n|$, donde λ_i es el autovalor correspondiente al elemento i -ésimo de \mathcal{B} . Tómese un vector (columna) $\bar{x}_0 \in \mathbb{K}^n$. Su expresión como combinación lineal de los vectores de \mathcal{B} es:

$$\bar{x}_0 = x_1 \bar{u}_1 + x_2 \bar{u}_2 + \dots + x_n \bar{u}_n \quad (7.54)$$

Con coeficientes escalares $x_1, \dots, x_n \in \mathbb{K}$.

Pre-multiplicando \bar{x}_0 por la matriz A se obtiene:

$$\begin{aligned} A\bar{x}_0 &= x_1 A\bar{u}_1 + x_2 A\bar{u}_2 + \dots + x_n A\bar{u}_n = \\ &= x_1 \lambda_1 \bar{u}_1 + x_2 \lambda_2 \bar{u}_2 + \dots + x_n \lambda_n \bar{u}_n = \\ &= \bar{x}_1 \end{aligned} \quad (7.55)$$

Y si se vuelve a pre-multiplicar por A:

$$\begin{aligned} A^2 \bar{x}_0 &= A \bar{x}_1 = x_1 \lambda_1 A \bar{u}_1 + x_2 \lambda_2 A \bar{u}_2 + \cdots + x_n \lambda_n A \bar{u}_n = \\ &= x_1 \lambda_1^2 \bar{u}_1 + x_2 \lambda_2^2 \bar{u}_2 + \cdots + x_n \lambda_n^2 \bar{u}_n = \\ &= \bar{x}_2 \end{aligned} \quad (7.56)$$

Pre-multiplicando el vector \bar{x}_0 por A k veces:

$$\begin{aligned} A^k \bar{x}_0 &= A \bar{x}_{k-1} = x_1 \lambda_1^k \bar{u}_1 + x_2 \lambda_2^k \bar{u}_2 + \cdots + x_n \lambda_n^k \bar{u}_n = \\ &= \bar{x}_k \end{aligned} \quad (7.57)$$

Y dividiendo por la k -ésima potencia del autovalor con mayor módulo de la matriz, conocido también como el autovalor “dominante”:

$$\begin{aligned} \frac{A^k \bar{x}_0}{\lambda_1^k} &= x_1 \left(\frac{\lambda_1}{\lambda_1} \right)^k \bar{u}_1 + \cdots + x_n \left(\frac{\lambda_n}{\lambda_1} \right)^k \bar{u}_n = \\ &= x_1 \bar{u}_1 + x_2 \left(\frac{\lambda_2}{\lambda_1} \right)^k \bar{u}_2 + \cdots + x_n \left(\frac{\lambda_n}{\lambda_1} \right)^k \bar{u}_n \end{aligned} \quad (7.58)$$

Las Proposiciones C.8.2 y C.9.2 establecen que el módulo del producto de números complejos es el producto de sus módulos, y que el módulo de su cociente es el cociente de sus módulos, por lo que $\lim_{k \rightarrow \infty} (\lambda_i / \lambda_1)^k = 0 \forall i > 1$, mientras que $(\lambda_1 / \lambda_1)^k = 1 \forall k \geq 0$. En consecuencia:

$$\begin{aligned} \frac{A^k \bar{x}_0}{\lambda_1^k} &= x_1 \bar{u}_1 + x_2 \left(\frac{\lambda_2}{\lambda_1} \right)^k \bar{u}_2 + \cdots \\ &\dots + x_n \left(\frac{\lambda_n}{\lambda_1} \right)^k \bar{u}_n \xrightarrow{k \rightarrow \infty} x_1 \bar{u}_1 \end{aligned} \quad (7.59)$$

Se concluye que, dadas una matriz cuadrada de orden n diagonalizable y un vector columna \bar{x}_0 en \mathbb{K}^n , pre-multiplicando el vector por A repetidas veces se obtiene una secuencia de vectores en \mathbb{K}^n , $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_k$, que converge a un vector propio de A asociado a su autovalor dominante:

$$A^k \bar{x}_0 = \bar{x}_k \quad (7.60)$$

$$A \bar{x}_k \approx \lambda_1 \bar{x}_k \quad (7.61)$$

Siendo el valor de λ_1 estimable mediante el cociente:

$$\frac{\bar{x}_k^h A \bar{x}_k}{\bar{x}_k^h \bar{x}_k} = \frac{\lambda_1 \bar{x}_k^h \bar{x}_k}{\bar{x}_k^h \bar{x}_k} \approx \lambda_1 \quad (7.62)$$

A este algoritmo se le conoce con el nombre de “método de la potencia”.

Para obtener los autovalores restantes $\lambda_2, \dots, \lambda_n$ de la matriz, así como vectores propios asociados a estos, puede aplicarse sobre A un procedimiento conocido como “deflación”, descrito en libros especializados en computación numérica. Otra posibilidad es la siguiente: sean un vector columna inicial $\bar{x}_0 \in \mathbb{K}^n$ y un escalar $\mu \in \mathbb{K}$, si el parámetro μ se encuentra lo suficientemente cercano a alguno de los valores propios de A, supóngase que a λ_j , $1 \leq j \leq n$, pre-multiplicar \bar{x}_0 por la matriz $(A - \mu I)^{-1}$ repetidas veces produce una secuencia de vectores que converge a un vector propio asociado al autovalor λ_j pues, por aplicación de la propiedad

4 de la Proposición 6.4.4 y de la propiedad 3 de la Proposición 6.4.5:

$$\begin{aligned} \lambda \in \mathbb{K} \text{ es valor propio de } A \in \mathbb{K}^{n \times n} &\Rightarrow \\ \Rightarrow \frac{1}{\lambda - \mu} \text{ es valor propio de } (A - \mu I)^{-1} &\quad (7.63) \end{aligned}$$

A este algoritmo para calcular de valores y vectores propios se le suele llamar “método de la potencia inversa”, o también “iteración inversa”.

7.2.7 Resumen de los últimos apartados

1. Una forma cuadrática es una aplicación no lineal que asigna a un vector de componentes reales un escalar real a través de una matriz real simétrica.
2. Cada forma cuadrática, al igual que su cociente de *Rayleigh* correspondiente, está acotada superior e inferiormente en función de los autovalores máximo y mínimo de su matriz simétrica asociada.
3. De acuerdo a lo anterior, las matrices reales simétricas se pueden clasificar según los signos de los autovalores que forman su espectro. Esta clasificación determina los signos que pueden tener las formas cuadráticas construidas con estas matrices.
4. Una matriz real simétrica es definida positiva, si y sólo si, todos los menores construidos desde el primer elemento de su diagonal principal tienen un determinante mayor que cero.
5. Una matriz real simétrica es definida negativa, si y sólo si, todos los menores construidos desde el primer elemento de su diagonal principal tienen un determinante menor que cero.

7.3 Aplicación de las formas cuadráticas a la clasificación y el estudio de las curvas cónicas

7.3.1 Introducción

Una aplicación interesante de las formas cuadráticas es la representación matemática y el estudio de las llamadas “formas cónicas”: la elipse, la hipérbola, la parábola, y todos sus casos degenerados. Como se verá en esta sección, el cálculo de los valores y vectores propios de la matriz simétrica presente en la ecuación de una cónica permiten determinar tanto el tipo de cónica del que se trata como sus propiedades geométricas.

7.3.2 Las curvas cónicas

Las curvas cónicas, o secciones cónicas, son curvas que resultan de intersectar un cono con un plano. Los tipos más conocidos de cónicas son:

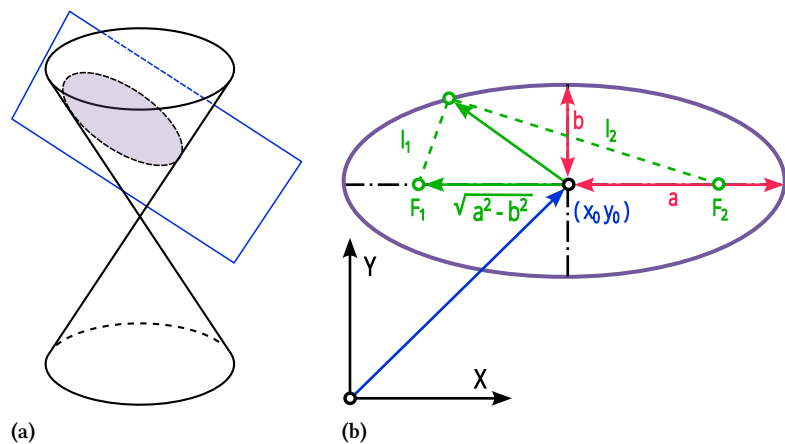
1. La elipse, cuya ecuación implícita reducida es:

$$\frac{(x - x_0)^2}{a^2} + \frac{(y - y_0)^2}{b^2} = 1 \tag{7.64}$$

x_0 y y_0 son las coordenadas del centro de la elipse, mientras que a y b son las longitudes de sus semiejes.

Como caso particular, si $a = b$ se tiene una circunferencia de radio a cuya ecuación reducida es $(x - x_0)^2 + (y - y_0)^2 = a^2$.

Figura 7.1: Centro, ejes, longitudes de los semiejes y focos (F_1 y F_2) de una elipse. La elipse puede definirse, o bien como la intersección entre un cono y un plano dispuesto con una inclinación menor a la de la generatriz (Figura 7.1a), o bien como el conjunto de todos los puntos del plano euclidiano que cumplen la identidad $l_1 + l_2 = 2a$, donde l_1 y l_2 son las distancias a cada foco de la elipse (Figura 7.1b). La distancia entre cada foco y el centro de la elipse es igual a $\sqrt{a^2 - b^2}$.



2. La hipérbola, cuya ecuación implícita reducida es una de las dos siguientes:

$$\begin{aligned} \frac{(x - x_0)^2}{a^2} - \frac{(y - y_0)^2}{b^2} &= 1 \\ \frac{(y - y_0)^2}{a^2} - \frac{(x - x_0)^2}{b^2} &= 1 \end{aligned} \tag{7.65}$$

Dependiendo de su orientación. En esta ecuación, x_0 y y_0 son las coordenadas de su centro, mientras que a y b son las longitudes de sus semiejes.

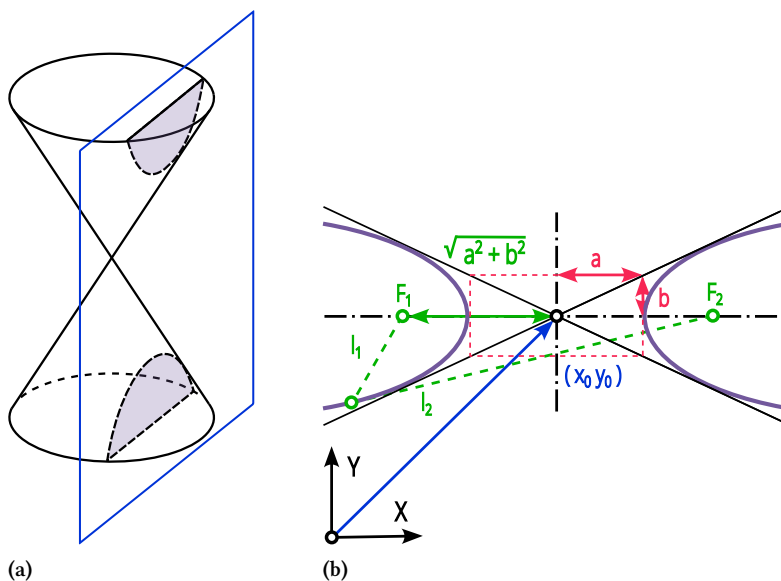


Figura 7.2: Centro, ejes, longitudes de los semiejes y focos (F_1 y F_2) de una hipérbola. La hipérbola puede definirse, o bien como la intersección entre un cono y un plano dispuesto verticalmente (Figura 7.2a), o bien como el conjunto de todos los puntos del plano euclidiano que cumplen la identidad $l_1 - l_2 = 2a$, donde l_1 y l_2 son las distancias a cada foco de la hipérbola (Figura 7.2b). La distancia entre cada foco y el centro de la hipérbola es igual a $\sqrt{a^2 + b^2}$.

3. La parábola, cuya ecuación implícita reducida es una de las dos siguientes:

$$\begin{aligned} y - y_0 &= a(x - x_0)^2 \\ x - x_0 &= a(y - y_0)^2 \end{aligned} \tag{7.66}$$

Dependiendo de su orientación. x_0 y y_0 son las coordenadas de su centro y a es un parámetro de forma que determina la derivada segunda de la curva.

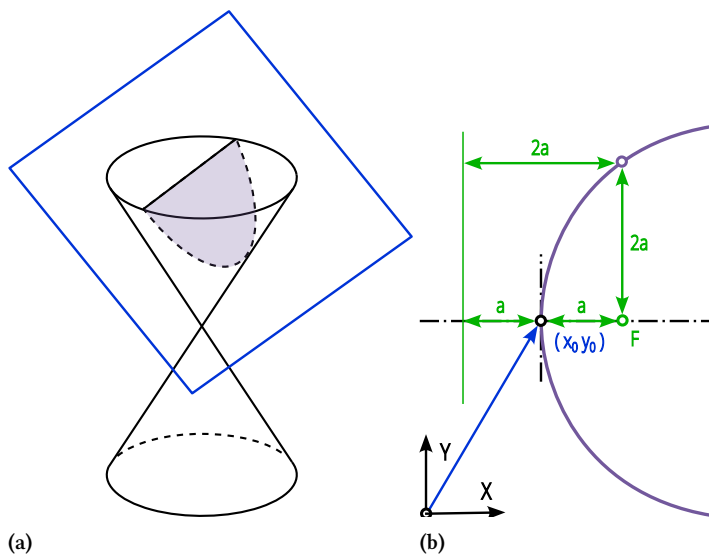


Figura 7.3: Centro, directriz y foco F de una parábola. La parábola puede definirse, o bien como la intersección entre un cono y un plano dispuesto con una inclinación igual a la de la generatriz (Figura 7.3a) o bien, como el conjunto de todos los puntos del plano euclidiano cuya distancia a una recta llamada "directriz" coincide con su distancia a un punto denominado "foco" (Figura 7.3b). La constante a de la ecuación implícita de una parábola es la mitad de la distancia entre el foco y la directriz.

Las tres ecuaciones características [(7.64) ... 7.66]] de cada uno de los tres tipos fundamentales de cónicas son fáciles de identificar. Pero si la cónica está girada y/o su centro no coincide con el origen de coordenadas, los ejes x e y ya no coinciden con sus ejes de simetría y las ecuaciones características se vuelven más complicadas de identificar.

En general, la ecuación general de la cónica en (x, y) tiene forma de polinomio de segundo grado en dos variables:

$$a_{11}x^2 + 2a_{12}xy + a_{22}y^2 + b_1x + b_2y + c = 0 \quad (7.67)$$

El objetivo de esta sección es, dada la ecuación de una cónica, aprender a identificar de qué tipo de cónica se trata y sus parámetros geométricos fundamentales.

7.3.3 Clasificación y estudio de cónicas

Comiencese escribiendo la ecuación (7.67) de forma matricial, observando que sus tres primeros términos constituyen una forma cuadrática asociada a una matriz real simétrica A :

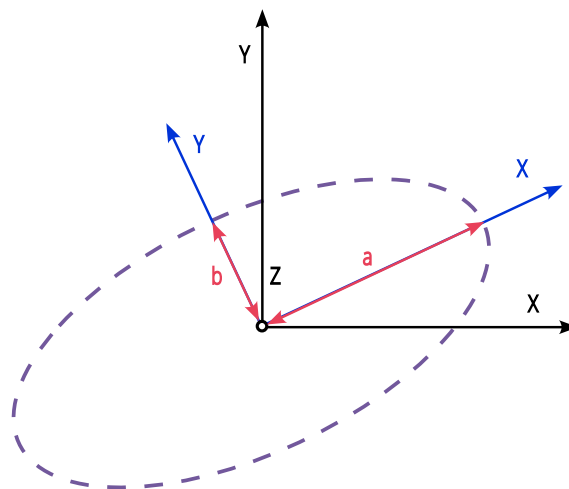
$$(x, y) \begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + (b_1, b_2) \begin{pmatrix} x \\ y \end{pmatrix} + c = 0 \quad (7.68)$$

Con el cambio de variable $(x, y)^t = Q(X, Y)^t$, donde Q es la matriz ortogonal que diagonaliza A , se sustituyen los ejes x y y por otros nuevos paralelos a los ejes de simetría de la cónica, de modo que la ecuación (7.69) se transforma en la siguiente:

$$(X, Y)Q^tAQ \begin{pmatrix} X \\ Y \end{pmatrix} + (b_1, b_2)Q \begin{pmatrix} X \\ Y \end{pmatrix} + c = 0 \Rightarrow \quad (7.69)$$

$$\Rightarrow (X, Y) \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix} + (c_1, c_2) \begin{pmatrix} X \\ Y \end{pmatrix} + c = 0 \quad (7.70)$$

Figura 7.4: Al realizar el cambio de variable $(x, y)^t = Q(X, Y)^t$, se sustituyen las coordenadas (x, y) de la cónica (en negro en la figura) por un nuevo sistema de coordenadas (X, Y) (en azul en la figura) que tiene sus ejes de abscisas y ordenadas paralelos a los ejes de simetría de una cónica que, en el ejemplo ilustrado, es una elipse. La matriz A de la cónica en el nuevo sistema de coordenadas será una matriz diagonal.



Esta última expresión puede escribirse como:

$$\lambda_1 X^2 + \lambda_2 Y^2 + c_1 X + c_2 Y + c = 0 \quad (7.71)$$

Donde λ_1 y λ_2 son los valores propios de $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix}$.

Completando cuadrados en la expresión (7.71), se obtiene una ecuación de la cónica en la forma reducida:

$$\lambda_1 \left(X - \frac{c_1}{2\lambda_1} \right)^2 + \lambda_2 \left(Y - \frac{c_2}{2\lambda_2} \right)^2 = d \Rightarrow \quad (7.72)$$

$$\frac{\left(X - \frac{c_1}{2\lambda_1} \right)^2}{d/\lambda_1} + \frac{\left(Y - \frac{c_2}{2\lambda_2} \right)^2}{d/\lambda_2} = 1 \quad (7.73)$$

Pudiéndose extraer de ella las longitudes de los semiejes de la cónica, $a = \sqrt{d/\lambda_1}$ y $b = \sqrt{d/\lambda_2}$, y obtener el centro de la cónica simplemente deshaciendo el cambio de coordenadas:

$$(x_0, y_0)^t = Q \left(\frac{c_1}{2\lambda_1}, \frac{c_2}{2\lambda_2} \right)^t \quad (7.74)$$

En función de los valores de λ_1 y λ_2 , se puede clasificar la cónica y extraer los parámetros geométricos que las caracterizan:

1. Si los autovalores λ_1 y λ_2 son no nulos y del mismo signo, A es definida positiva o definida negativa, y la cónica es de tipo elíptico: una elipse, un punto o el conjunto vacío. Si λ_1 y d tuviesen igual signo, la curva sería una elipse; si λ_1 y d tuviesen distinto signo, ningún punto del plano euclidiano \mathbb{R}^2 cumpliría la ecuación característica, o en otras palabras, el conjunto de puntos del plano que conformarían la curva se trataría del conjunto vacío \emptyset . Si $d = 0$, uno y sólo un punto del plano euclidiano cumpliría la ecuación característica: el centro de la cónica (x_0, y_0) .
2. Si los autovalores λ_1 y λ_2 es no nulos y de distinto signo, A es indefinida, y la cónica es de tipo hiperbólica: una hipérbola o dos rectas secantes. Si $d \neq 0$, la curva se trataría de una hipérbola, mientras que si $d = 0$ esta consistiría en dos rectas que se cortarían en el centro (x_0, y_0) .
3. Si alguno de los valores propios λ_1 y λ_2 es nulo, A es semidefinida positiva o negativa, y la cónica es de tipo parabólico. Por ejemplo, si $\lambda_2 = 0$ la ecuación (7.73) se reduciría a:

$$\lambda_1 X^2 + c_1 X + c_2 Y + c = 0 \quad (7.75)$$

Completando cuadrados en X se obtendría una ecuación de la forma:

$$\lambda_1 X^2 + c_2 Y + d = 0 \quad (7.76)$$

Si $c_2 \neq 0$, la curva representada por la ecuación consistiría en una parábola. Si $c_2 = 0$, los puntos de \mathbb{R}^2 que satisfarían la ecuación formarían dos rectas paralelas de ecuaciones $X = \pm e$ las cuales a su vez podrían ser dos rectas reales, dos rectas complejas (en \mathbb{C}^2) o dos rectas coincidentes en una sola.

Observación 7.3.1 Las curvas cónicas se pueden extender a los espacios \mathbb{R}^n , con $n > 2$, mediante las llamadas "superficies cuadráticas". Estas se caracterizan por ecuaciones implícitas que tienen la misma forma que la expresión (7.68), con una matriz $A \in \mathbb{R}^{n \times n}$ real simétrica y el vector de coordenadas $\bar{x} \in \mathbb{R}^n$. En función de los valores propios de A, la superficie puede ser:

1. Un elipsoide: si A es definida positiva o negativa.
2. Un hiperboloide: si A es indefinida.
3. Un paraboloides: si A es semidefinida positiva o negativa.

Algunas leyes físicas, como el principio de mínima acción, se expresan en función de los puntos estacionarios de una superficie cuadrática que representa propiedades del sistema como la energía y el momento que contiene.

Ejemplo 7.3.1 Considérese la curva cónica cuya ecuación implícita es el polinomio de segundo grado en dos variables:

$$6x^2 + 9y^2 - 4xy + 4x - 18y + 4 = 0 \quad (7.77)$$

En adelante, se mostrará cómo clasificar dicha cónica y estudiarla, calculando su centro, sus ejes de simetría y sus radios. Expresando la identidad (7.77) en forma matricial:

$$\begin{aligned} & \bar{x}^t A \bar{x} + \mathbf{b}^t \bar{x} + c = \\ & = (x, y) \begin{pmatrix} 6 & -2 \\ -2 & 9 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + (4, -18) \begin{pmatrix} x \\ y \end{pmatrix} + 4 = 0 \end{aligned} \quad (7.78)$$

Los valores propios de A son $\lambda_1 = 5$ y $\lambda_2 = 10$. Al ser positivos, la curva es elíptica. Dos vectores propios unitarios asociados a cada autovalor son, respectivamente, $\bar{q}_1 = 1/\sqrt{5} \cdot (-1, 2)^t$ y $\bar{q}_2 = 1/\sqrt{5} \cdot (2, 1)^t$. Estos tienen las direcciones de los ejes de simetría de la elipse. Una matriz ortogonal que diagonaliza A es:

$$Q = (\bar{q}_1 | \bar{q}_2) = \frac{1}{\sqrt{5}} \begin{pmatrix} -1 & 2 \\ 2 & 1 \end{pmatrix} \quad (7.79)$$

Aplicando el cambio de variable $(x, y)^t = Q(X, Y)^t$, se obtiene la forma matricial de la elipse en el nuevo sistema de coordenadas:

$$(X, Y) \begin{pmatrix} X \\ Y \end{pmatrix} + \frac{1}{\sqrt{5}}(-40, -10) \begin{pmatrix} X \\ Y \end{pmatrix} + 4 = 0 \quad (7.80)$$

Completando cuadrados en X y en Y , la ecuación (7.80) se reduce a:

$$\begin{aligned} & 10 \left(X - \frac{2}{\sqrt{5}} \right)^2 + 5 \left(Y - \frac{1}{\sqrt{5}} \right)^2 = 5 \Rightarrow \\ & \Rightarrow \frac{(X - 2/\sqrt{5})^2}{(1/\sqrt{2})^2} + \frac{(Y - 1/\sqrt{5})^2}{1^2} = 1 \end{aligned} \quad (7.81)$$

De la forma reducida de la elipse en el sistema X, Y , ecuación (7.81), se deduce que los radios mayor y menor de la elipse son 1 y $1/\sqrt{2}$.

Por último, para obtener el centro de la elipse, se transforma el vector $(2/\sqrt{5}, 1/\sqrt{5})$, que es el centro de la elipse en el sistema de coordenadas (X, Y) , al sistema inicial de coordenadas (x, y) :

$$\begin{pmatrix} x \\ y \end{pmatrix} = \frac{1}{\sqrt{5}} \begin{pmatrix} -1 & 2 \\ 2 & 1 \end{pmatrix} \cdot \frac{1}{\sqrt{5}} \begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad (7.82)$$

7.3.4 Resumen

1. Hay tres clases fundamentales de cónicas: la elipse, la parábola y la hipérbola. Las tres son fácilmente identificables por sus ecuaciones cuando las curvas están centradas en el origen de coordenadas, pero la identificación se complica si se encuentran giradas y trasladadas sobre el plano.
2. Para identificar la cónica a partir de la ecuación cuadrática que la caracteriza, así como sus propiedades, habrá que efectuar un cambio de variable, o de coordenadas, de modo tal que la ecuación se transforme en otra más simple que permita identificar inmediatamente el tipo de cónica.
3. Una forma de realizar este cambio de coordenadas es diagonalizar ortogonalmente la matriz real simétrica presente en la forma cuadrática que define la cónica.
4. Si esa matriz es definida positiva o negativa, se tratará de una elipse. Si es indefinida, de una hipérbola. Si uno de los valores propios es nulo, de una parábola.
5. Una vez obtenido el centro de la cónica en el nuevo sistema de coordenadas, se pueden calcular sus coordenadas respecto del sistema de referencia inicial deshaciendo la transformación geométrica.



APÉNDICES



A

Fundamentos de lógica y conjuntos

El objetivo de este apéndice es dotar al lector de los fundamentos en lógica, teoría de conjuntos y razonamiento matemático necesarios (y suficientes) para la comprensión de todo el contenido de este libro. Para distinguir los conjuntos de las proposiciones lógicas, los primeros se denotarán mediante letras mayúsculas cursivas como P , Q o R †; mientras que las segundas se representarán mediante letras mayúsculas caligráficas como \mathcal{P} , \mathcal{Q} o \mathcal{R} .

A.1	Elementos de lógica . . .	247
A.2	Teoría básica de conjuntos	253
A.3	Técnicas fundamentales de demostración	259

Observación A.0.1 † Salvo los conjuntos de los números naturales (\mathbb{N}), enteros (\mathbb{Z}), racionales (\mathbb{Q}), reales (\mathbb{R}) y complejos (\mathbb{C}).

A.1 Elementos de lógica

Las proposiciones se definen como aquellas afirmaciones que pueden ser verdaderas o falsas, pero no ni verdaderas ni falsas, ni verdaderas y falsas al mismo tiempo. Por ejemplo, la afirmación “Pepe tiene un coche” puede ser verdadera o falsa, pero no puede ser verdadera y falsa a la vez, pues no se puede poseer un coche y no poseerlo al mismo tiempo, ni puede no ser ni verdadera ni falsa, o bien se tiene un coche, o bien no se tiene.

Las palabras afirmación y enunciado se suelen emplear como sinónimos de proposición, aunque esos dos términos tienen un significado más general que puede englobar sintagmas cuya veracidad o falsedad no puede ser probada de manera empírica y lógica por tratarse de juicios de valor. Por ejemplo, la afirmación “Juana es lista” es una apreciación subjetiva que no es ni verdadera ni falsa desde un plano objetivo, por lo que no sería una proposición.

Las proposiciones, al igual que los números, pueden combinarse mediante operaciones para obtener nuevas proposiciones cuya veracidad o falsedad dependerá de la veracidad y falsedad de las proposiciones que las componen. Tres operaciones lógicas elementales que se definen a continuación son la suma, el producto y la negación:

Definición A.1.1 — Suma, producto y negación lógica

Sean \mathcal{P} y \mathcal{Q} dos proposiciones lógicas, se definen:

1. Su suma, como una nueva proposición denotada:

$$\mathcal{P} \wedge \mathcal{Q} \tag{A.1}$$

Que es verdadera si ambas son verdaderas, y falsa si al menos una de ellas es falsa. El símbolo “ \wedge ” se lee como la conjunción “y”.

2. Su producto, como una nueva proposición denotada:

$$\mathcal{P} \vee \mathcal{Q} \quad (\text{A.2})$$

Que es verdadera si al menos una de ellas es verdadera, y falsa si ambas son falsas. El símbolo “ \vee ” se lee como la conjunción “o”.

Sea \mathcal{P} una proposición lógica, se define:

3. Su negación, como una nueva proposición denotada:

$$\neg \mathcal{P} \quad (\text{A.3})$$

Que es verdadera si \mathcal{P} es falsa, y falsa si \mathcal{P} es verdadera. El símbolo “ \neg ” se lee como la conjunción “no”.

Ejemplo A.1.1 Dadas las proposiciones:

1. x es un número racional.
2. x es un número real.

Supóngase que $x = 1/2$. La proposición:

$$x \text{ es un número racional} \wedge x \text{ es un número real} \quad (\text{A.4})$$

Sería verdadera, pues el número $1/2$ es tanto real como racional.

Supóngase ahora que $x = \sqrt{2}$. La proposición:

$$x \text{ es un número racional} \wedge x \text{ es un número real} \quad (\text{A.5})$$

Sería falsa, pues el número $\sqrt{2}$ no es racional. Pero las proposiciones:

$$x \text{ no es un número racional} \wedge x \text{ es un número real} \quad (\text{A.6})$$

$$x \text{ es un número racional} \vee x \text{ es un número real} \quad (\text{A.7})$$

sí que serían verdaderas, pues $\sqrt{2}$ que es un número real.

Una proposición como las del Ejemplo A.1.1, que dependa de una variable o índice x , puede ser:

1. Verdadera para cualquier valor de x .
2. Verdadera sólo para algunos valores de x .
3. Verdadera para un único valor de x .
4. Falsa para todo valor de x .

Para enunciar la veracidad o falsedad de una proposición en función de una variable de acuerdo con la clasificación anterior, se emplean los llamados “cuantificadores”:

Definición A.1.2 – Cuantificador universal y cuantificador existencial

Sea $\mathcal{P}(x)$ una proposición que depende de un índice o variable x , se definen:

1. El cuantificador universal \forall (“para todo”), que indica que $\mathcal{P}(x)$ es verdadera en cada uno de los posibles valores que x puede adoptar:

$$\mathcal{P}(x) \forall x \quad (\text{A.8})$$

2. El cuantificador existencial \exists (“existe”), que indica la existencia de al menos un valor de x tal que $\dagger \mathcal{P}(x)$ es verdadera:

$$\exists x : \mathcal{P}(x) \quad (\text{A.9})$$

Si $\mathcal{P}(x)$ es verdadera para un único valor de x , se escribe $\exists ! x : \mathcal{P}(x)$, leyéndose “ $\exists !$ ” como “existe un único”.

Si $\mathcal{P}(x)$ no es verdadera para ningún valor de x , se denota $\nexists x : \mathcal{P}(x)$, leyéndose “ \nexists ” como “no existe”.

Observación A.1.1 \dagger En Matemáticas, el sintagma “tal que” suele representarse mediante los dos puntos “:” (la notación empleada en este texto), mediante el símbolo “\” (barra invertida) o con el símbolo “|” (barra vertical).

Ejemplo A.1.2 Denótese $\mathcal{P}(x)$ a la proposición:

$$2x \text{ es un número par.} \quad (\text{A.10})$$

Sea \mathbb{N} el conjunto de todos los números naturales, la proposición $\mathcal{P}(x)$ es verdadera para todo valor de x en el conjunto \mathbb{N} . Esto se enuncia matemáticamente así:

$$\mathcal{P}(x) \forall x \in \mathbb{N} \quad (\text{A.11})$$

Considérese ahora la proposición:

$$\mathcal{Q}(x) = x \text{ es un número par.} \quad (\text{A.12})$$

Sólo son pares los números naturales divisibles por dos, por lo que $\mathcal{Q}(x)$ es verdadera sólo para algunos valores de \mathbb{N} , escribiéndose:

$$\exists x \in \mathbb{N} : \mathcal{Q}(x) \quad (\text{A.13})$$

Por último, es fácil comprobar que la proposición:

$$\mathcal{R}(x) = 2x + 1 \text{ es un número par.} \quad (\text{A.14})$$

Es falsa para todo valor de \mathbb{N} , por lo que puede escribirse:

$$\nexists x \in \mathbb{N} : \mathcal{R}(x) \quad (\text{A.15})$$

Además de la suma y el producto, otra relación que se puede establecer entre dos proposiciones es la implicación simple. Esta se da cuando la veracidad de una proposición conlleva automáticamente la veracidad de otra proposición.

Definición A.1.3 – Implicación simple

Si una proposición Q se cumple siempre que se cumple otra proposición denotada P , se escribe:

$$P \Rightarrow Q$$

Y se dice que:

1. Si P entonces Q .
2. P implica Q .
3. P es condición suficiente para que se cumpla Q .
4. Q es condición necesaria para que se cumpla P .

Ejemplo A.1.3 Si x es un número natural, x es un número entero, pues el conjunto de los números enteros incluye a todos los naturales, por tanto:

$$x \text{ es un número natural} \Rightarrow x \text{ es un número entero.}$$

Ejemplo A.1.4 Consideremos las proposiciones:

1. Pepita saca un cinco en álgebra.
2. Pepita aprueba álgebra.

La implicación:

$$\text{Pepita saca un cinco en álgebra} \Rightarrow \text{Pepita aprueba álgebra} \quad (\text{A.16})$$

Es claramente verdadera.

Definición A.1.4 – Converso, inverso y contrapositivo de una implicación simple

Dadas dos proposiciones P y Q que cumplen la implicación:

$$P \Rightarrow Q \quad (\text{A.17})$$

Se llama:

1. Converso (o recíproco) de $P \Rightarrow Q$, a la implicación:

$$Q \Rightarrow P \quad (\text{A.18})$$

2. Inverso (o reverso) de $P \Rightarrow Q$, a la implicación:

$$\neg P \Rightarrow \neg Q \quad (\text{A.19})$$

3. Contrapositivo de $\mathcal{P} \Rightarrow \mathcal{Q}$, a la implicación:

$$\neg \mathcal{Q} \Rightarrow \neg \mathcal{P} \quad (\text{A.20})$$

En general, la veracidad o falsedad del converso y el inverso de una implicación entre dos proposiciones depende no sólo de la veracidad o falsedad de las proposiciones presentes en la implicación, sino también de su naturaleza, mientras que la veracidad o falsedad del contrapositivo de una implicación tiene una relación unívoca † con la veracidad o falsedad de las proposiciones presentes en la aplicación. Todo esto puede verse en los siguientes ejemplos:

Observación A.1.2 † El vocablo “unívoco” significa en este contexto “que depende de una única cosa, y no de otras”.

Ejemplo A.1.5 Se había razonado en el Ejemplo A.1.3 que:

$$x \text{ es un número natural} \Rightarrow x \text{ es un número entero.} \quad (\text{A.21})$$

El converso de esta implicación es la proposición:

$$x \text{ es un número entero} \Rightarrow x \text{ es un número natural.} \quad (\text{A.22})$$

Pero no todos los números enteros son naturales (concretamente, los enteros negativos) por lo que esta última proposición es falsa. El inverso (o reverso) de la implicación (A.21) es:

$$x \text{ no es un número natural} \Rightarrow x \text{ no es un número entero.} \quad (\text{A.23})$$

Pero los números enteros negativos no pertenecen al conjunto de los números naturales, por lo que esta proposición también es falsa.

Sin embargo, el contrapositivo de la implicación (A.21):

$$x \text{ no es un número entero} \Rightarrow x \text{ no es un número natural} \quad (\text{A.24})$$

Sí que es una proposición verdadera. Por ejemplo, los números $1/2$, π y $1 + i$ no son enteros, por lo que tampoco son naturales.

Ejemplo A.1.6 El converso de la implicación (A.16):

$$\text{Pepita aprueba álgebra} \Rightarrow \text{Pepita saca un cinco en álgebra} \quad (\text{A.25})$$

Es falso, pues es posible aprobar álgebra con, por ejemplo, un seis. Esto último es un “contraejemplo” que permite demostrar la falsedad de la implicación (A.25). El inverso de la implicación (A.16):

$$\begin{aligned} \text{Pepita no saca un cinco en álgebra} &\Rightarrow \\ &\Rightarrow \text{Pepita no aprueba álgebra} \end{aligned} \quad (\text{A.26})$$

Tampoco es verdadero por la misma razón por la que era falso el converso, porque con una nota mayor que cinco se hubiese aprobado álgebra igualmente.

Sin embargo, el contrapositivo de la implicación (A.16):

$$\begin{aligned} \text{Pepita no aprueba álgebra} &\Rightarrow \\ \Rightarrow \text{Pepita no saca un cinco en álgebra} &\quad (A.27) \end{aligned}$$

Sí es verdadero, pues si Pepita hubiese sacado un cinco habría aprobado.

En los Ejemplos A.1.5 y A.1.6, se ha visto que, tanto las implicaciones enunciadas, (A.21) y (A.25) como sus contrapositivos, (A.24) y (A.27), eran proposiciones verdaderas. Esto no es una coincidencia. Mediante el uso de “tablas de verdad” es posible demostrar que, en general, toda implicación es equivalente a su contrapositivo. Es decir, que si es verdadera una relación de implicación entre dos proposiciones, también lo es el contrapositivo de esa implicación.

Proposición A.1.1 — Equivalencia entre una implicación simple y su contrapositivo

Si \mathcal{P} y \mathcal{Q} son dos proposiciones tales que la implicación $\mathcal{P} \Rightarrow \mathcal{Q}$ es verdadera, su contrapositivo. Es decir, la implicación:

$$\neg \mathcal{Q} \Rightarrow \neg \mathcal{P}$$

Es también verdadera, y viceversa; si $\neg \mathcal{Q} \Rightarrow \neg \mathcal{P}$ es verdadera, también lo es $\mathcal{P} \Rightarrow \mathcal{Q}$. Por tanto, toda implicación es equivalente a su contrapositivo.

Sin embargo, una implicación simple y su converso no son siempre equivalentes. Si lo fueran, se tendría una relación de implicación doble en la que \mathcal{P} implicaría \mathcal{Q} , y viceversa.

Definición A.1.5 — Implicación doble

Si $\mathcal{P} \Rightarrow \mathcal{Q}$ y $\mathcal{Q} \Rightarrow \mathcal{P}$, puede escribirse:

$$\mathcal{P} \Leftrightarrow \mathcal{Q}$$

Y decirse que:

1. \mathcal{P} , si y sólo si, \mathcal{Q} .
2. \mathcal{Q} , si y sólo si, \mathcal{P} .
3. \mathcal{P} es condición necesaria y suficiente para que se cumpla \mathcal{Q} .
4. \mathcal{Q} es condición necesaria y suficiente para que se cumpla \mathcal{P} .
5. \mathcal{P} y \mathcal{Q} son equivalentes.

Ejemplo A.1.7 Consideremos las proposiciones:

1. Pepita saca al menos un cinco.
2. Pepita aprueba álgebra.

La implicación:

$$\text{Pepita saca al menos un cinco} \Rightarrow \text{Pepita aprueba álgebra} \quad (\text{A.28})$$

Es verdadera, así como la implicación:

$$\text{Pepita aprueba álgebra} \Rightarrow \text{Pepita saca al menos un cinco}, \quad (\text{A.29})$$

Pues con una nota menor que cinco se suspendería, por tanto, ambas proposiciones son equivalentes y puede escribirse:

$$\text{Pepita aprueba álgebra} \Leftrightarrow \text{Pepita saca al menos un cinco}. \quad (\text{A.30})$$

La distinción entre implicación simple, Definición A.1.3, e implicación doble, Definición A.1.5, es vital para una adecuada comprensión y aplicación de numerosos conceptos matemáticos, como muestra el siguiente ejemplo:

Ejemplo A.1.8 Se había razonado en el Ejemplo A.1.3 que la implicación:

$$x \text{ es un número natural} \Rightarrow x \text{ es un número entero}. \quad (\text{A.31})$$

Es verdadera, mientras que su converso es falso, Ejemplo A.1.5. Considérese ahora la implicación:

$$\begin{aligned} x \text{ es un número natural} &\Rightarrow \\ \Rightarrow x \text{ es un número entero no negativo.} &\quad (\text{A.32}) \end{aligned}$$

El converso de esta última implicación sí que es verdadero, ya que el conjunto de todos los números naturales (incluyendo el cero) coincide con el conjunto de todos los números enteros no negativos. Por tanto, puede escribirse:

$$\begin{aligned} x \text{ es un número natural} &\Leftrightarrow \\ \Leftrightarrow x \text{ es un número entero no negativo.} &\quad (\text{A.33}) \end{aligned}$$

A.2 Teoría básica de conjuntos

Definición A.2.1 — Conjuntos finitos e infinitos. Conjunto vacío

Se llama conjunto a una colección de objetos distintos dos a dos y sin ordenación preestablecida denominados elementos del conjunto.

Un conjunto se dice finito si contiene un número finito de elementos, y se dice infinito si contiene infinitos elementos.

Se llama conjunto vacío, y se denota por \emptyset , al conjunto que no contiene ningún elemento.

Un conjunto de objetos matemáticos puede definirse de dos maneras distintas: especificando explícitamente todos los elementos que lo contienen, o bien enunciando una condición lógica que un elemento ha de cumplir para que pertenezca al conjunto. Por ejemplo, el conjunto que contiene todos los números pares menores o iguales que 10 puede definirse explícitamente así:

$$A = \{0, 2, 4, 8, 10\} \quad (\text{A.34})$$

O implícitamente, mediante la condición:

$$A = \{n \in \mathbb{N} : n \leq 10 \wedge n = 2c \text{ para algún } c \in \mathbb{N}\} \quad (\text{A.35})$$

Observación A.2.1 En una definición intensiva de un conjunto, es común separar por comas todas las condiciones lógicas que han de cumplirse para que un elemento pertenezca a este, en lugar de utilizar el operador “ \wedge ”.

Es decir, como el conjunto formado por todos los números naturales “tal que”, Observación A.1.1, estos son menores o iguales que 10 y además se pueden descomponer de la forma $n = 2c$ para algún c natural. Al primer modo de definir un conjunto (A.34) se le denomina “definición extensiva”, mientras que al segundo modo (A.35) se le denomina “definición intensiva” o “por comprensión”, Observación A.2.1.

En cualquier caso, los conjuntos no ordenados se definen siempre mediante corchetes, $\{\cdot\}$. Al no preestablecerse ningún orden entre sus elementos, el conjunto A definido en la ecuación (A.34) coincide con el conjunto:

$$A = \{2, 0, 10, 6, 8\} \quad (\text{A.36})$$

A.2.1 Relaciones entre conjuntos y elementos

Un elemento matemático respecto de un conjunto de objetos puede cumplir una relación de pertenencia si este se encuentra entre los objetos del conjunto, o de no pertenencia si no se encuentra entre los elementos de dicho conjunto. De forma similar, si se tienen dos conjuntos de objetos, estos pueden ser disjuntos (si no tienen ningún elemento en común), anidados (si uno de ellos está contenido en el otro) o intersecantes (si comparten al menos una parte de todos los objetos que los componen).

Definición A.2.2 — Relaciones entre conjuntos y elementos

Sea E un conjunto y a un elemento del mismo, se dice que “ a pertenece a E ” y se escribe:

$$a \in E$$

En caso contrario, se escribe $a \notin E$.

Ejemplo A.2.1 Sea E el conjunto de todos los gatos, se tiene que:

$$\text{Garfield} \in E \quad (\text{A.37})$$

$$\text{Pequeño ayudante de Santa Claus} \notin E \quad (\text{A.38})$$

Ejemplo A.2.2 El número $x = -2$ pertenece al conjunto de los números enteros, pero no al de los números naturales, por lo que puede escribirse:

$$x \in \mathbb{Z} \quad (\text{A.39})$$

$$x \notin \mathbb{N} \quad (\text{A.40})$$

Definición A.2.3 — Relaciones entre conjuntos

Sean E, F dos conjuntos, se dice que:

1. “ E está contenido en F ”, denotándose:

$$E \subset F \quad (\text{A.41})$$

Si todo elemento de E pertenece a F . En este caso, se dice que E es un subconjunto de F .

2. “ E y F son iguales”, denotándose:

$$E = F \quad (\text{A.42})$$

Si todo elemento de E pertenece a F y viceversa, si todo elemento de F pertenece a E . Es decir, si:

$$(E \subset F) \wedge (F \subset E) \quad (\text{A.43})$$

En caso contrario, se escribe $E \neq F$.

3. “ E está contenido o es igual a F ”, denotándose:

$$E \subseteq F \quad (\text{A.44})$$

Si, o bien $E \subset F$, o bien $E = F$:

$$(E \subset F) \vee (E = F) \quad (\text{A.45})$$

4. “ E no está contenido en F ”, denotándose:

$$E \not\subseteq F \quad (\text{A.46})$$

Si E contiene elementos que no pertenecen a F .

Ejemplo A.2.3 El conjunto de todos los gatos blancos F es un subconjunto del conjunto de todos los gatos E . Es decir:

$$F \subset E$$

Pues todos los gatos blancos son, obviamente, gatos.

Ejemplo A.2.4 Todos los números enteros son números naturales, pero no todos los números naturales son números enteros, por lo que puede decirse que el conjunto de los números naturales está “contenido” en el de los números enteros:

$$\mathbb{N} \subset \mathbb{Z} \quad (\text{A.47})$$

No todos los números enteros son naturales, por lo que \mathbb{Z} no se encuentra contenido a \mathbb{N} :

$$\mathbb{Z} \not\subset \mathbb{N} \quad (\text{A.48})$$

Razonando de un modo análogo en cada uno de los cinco conjuntos de números mencionados en la Observación A.0.1, se deduce la siguiente relación de anidación entre estos:

$$\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C} \quad (\text{A.49})$$

A.2.2 Operaciones entre conjuntos

Al igual que las proposiciones lógicas, sección A.1, los conjuntos también pueden operarse los unos con los otros para construir nuevos conjuntos. A continuación se definen la unión, la intersección y la diferencia entre conjuntos, y el complementario de un conjunto. Las propiedades que cumplen estas operaciones se dejan fuera del alcance de este libro, pero pueden consultarse en manuales de matemáticas más avanzados o especializados en teoría de conjuntos.

Definición A.2.4 – Unión de conjuntos

Se define la unión de dos conjuntos E y F (Figura A.1) como:

$$E \cup F = \{a : a \in E \vee a \in F\}$$

La unión entre E y F es aquel conjunto cuyos elementos se encuentran o en E , o en F , o en ambos.

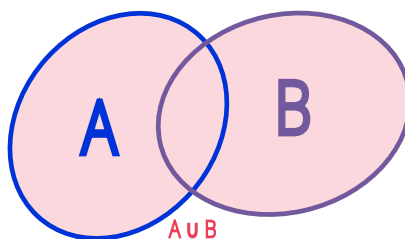


Figura A.1: Diagrama de Venn de los conjuntos A y B y de su unión, $A \cup B$.

Definición A.2.5 – Intersección de conjuntos. Conjuntos disjuntos

Se define la intersección de dos conjuntos E y F (Figura A.2) como:

$$E \cap F = \{a : a \in E \wedge a \in F\}$$

Si $E \cap F = \emptyset$ se dice que E y F son disjuntos.

La intersección entre E y F es aquel conjunto cuyos elementos se encuentran tanto en E como en F .

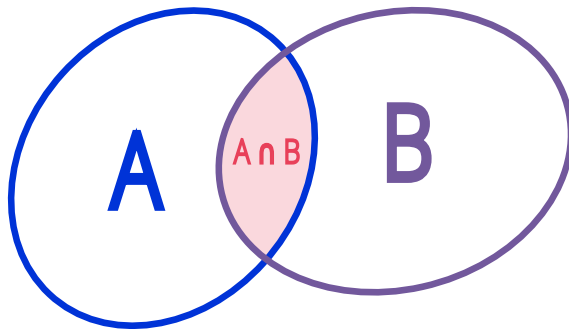


Figura A.2: Diagrama de Venn de los conjuntos A y B y de su intersección, $A \cap B$.

Ejemplo A.2.5 Sea E el conjunto de todos los gatos, y sea F el conjunto de todos los perros, su unión, $E \cup F$, es el conjunto de todos los animales que son perros o gatos, mientras que, por lo que respecta a su intersección:

$$E \cap F = \emptyset$$

Pues un animal no puede ser perro y gato a la vez.

Ejemplo A.2.6 La unión de los conjuntos $E = \{\text{perro, gato, pez}\}$ y $F = \{\text{gato, vaca}\}$, denotada $E \cup F$, es el conjunto formado por los elementos pertenecientes o bien a E , o bien a F , o bien a ambos conjuntos a la vez:

$$E \cup F = \{\text{perro, gato, pez, vaca}\} \quad (\text{A.50})$$

Su intersección, que se denota $E \cap F$, es el conjunto formado por aquellos elementos que E y F tienen en común, o sea, el elemento “gato” en este ejemplo en particular:

$$E \cap F = \{\text{gato}\} \quad (\text{A.51})$$

Ejemplo A.2.7 El conjunto de los números enteros puede descomponerse en la unión de tres conjuntos: el de los números enteros positivos, denotado \mathbb{Z}^+ ; el de los enteros negativos, denotado \mathbb{Z}^- , y el conjunto $\{0\}$ que contiene sólo el número cero:

$$\mathbb{Z} = \mathbb{Z}^+ \cup \mathbb{Z}^- \cup \{0\} \quad (\text{A.52})$$

Los conjuntos \mathbb{Z}^+ y \mathbb{Z}^- son disjuntos, mientras que la intersección de \mathbb{N} y \mathbb{Z} coincide con \mathbb{N} al pertenecer todos sus elementos a \mathbb{Z} :

$$\mathbb{Z}^+ \cap \mathbb{Z}^- = \emptyset \tag{A.53}$$

$$\mathbb{N} \cap \mathbb{Z} = \mathbb{N} \tag{A.54}$$

Definición A.2.6 – Diferencia de conjuntos

Se define la diferencia entre los conjuntos E y F (Figura A.3) como:

$$E \setminus F = \{a \in E : a \notin F\}$$

La diferencia entre E y F es el conjunto que contiene aquellos elementos de E que no se encuentran en F .

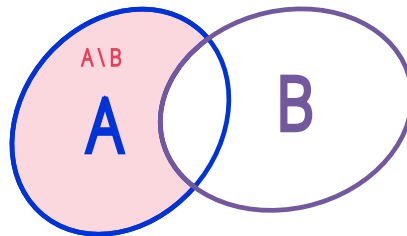


Figura A.3: Diagrama de Venn de los conjuntos A y B y de su diferencia, $A \setminus B$.

Definición A.2.7 – Complementario de un conjunto respecto de otro

Dado un conjunto E y otro conjunto F contenido o igual a E , $F \subseteq E$, se define el complementario de F respecto a E (Figura A.4), denotado F^c , F' , o $c(F)$, como:

$$F^c = E \setminus F = \{a \in E : a \notin F\}$$

El complementario de F respecto de E es el conjunto que contiene aquellos elementos de E que no se encuentran en F .

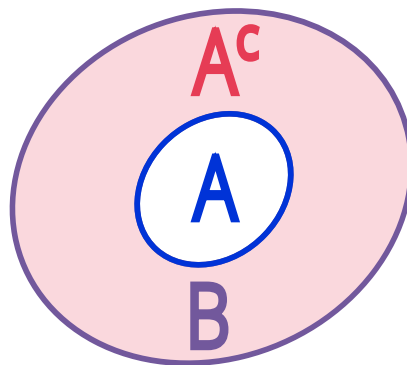


Figura A.4: Diagrama de Venn del complementario del conjunto A respecto de B , A^c .

Ejemplo A.2.8 Sea G el conjunto de todos los perros negros, su complementario respecto al conjunto F de todos los perros es el conjunto:

$$G^c = F \setminus G$$

Es decir, el conjunto de todos los perros que no son negros.

Ejemplo A.2.9 El complementario del conjunto de los números naturales respecto del conjunto de los números enteros es igual al conjunto de los enteros negativos:

$$\mathbb{N}^c = \mathbb{Z} \setminus \mathbb{N} = \mathbb{Z}^-$$

Definición A.2.8 — Producto cartesiano de conjuntos

Dados dos conjuntos E, F se define su producto cartesiano $E \times F$ como el conjunto de todas las tuplas o pares ordenados de elementos pertenecientes, respectivamente, a E y F :

$$E \times F = \{(x, y) : x \in E, y \in F\} \quad (\text{A.55})$$

Dados n conjuntos E_1, E_2, \dots, E_n , se define su producto cartesiano $E_1 \times E_2 \times \dots \times E_n$ como el conjunto de todas n -uplas o tuplas formadas por una secuencia de n elementos pertenecientes, respectivamente, a E_1, E_2, \dots, E_n :

$$E_1 \times \dots \times E_n = \{(x_1, \dots, x_n) : x_i \in E_i, \forall i = 1, \dots, n\} \quad (\text{A.56})$$

Para diferenciar las tuplas de los conjuntos, que de acuerdo con la Definición A.2.1 no son ordenados ni admiten repeticiones de sus elementos, las tuplas se denotan mediante paréntesis en lugar de corchetes: (\cdot) .

Ejemplo A.2.10 El conjunto $\mathbb{R} \times \mathbb{R} = \mathbb{R}^2$ contiene todos los pares ordenados de números reales, por ejemplo $(1, -2)$ y $(-\pi, \sqrt{2})$. Al tratarse de tuplas ordenadas, $\forall a, b \in \mathbb{R}$

$$(a, b) \neq (b, a)$$

A.3 Técnicas fundamentales de demostración

La base del pensamiento matemático es el pensamiento deductivo: a un objeto o colección de objetos se le asignan unas propiedades axiomáticas que se consideran verdaderas, y de esos axiomas \dagger se deducen de manera lógica otras proposiciones y relaciones entre los objetos. Un ejemplo muy sencillo viene dado por el siguiente silogismo:

Observación A.3.1 \dagger A los “axiomas” se les puede denominar también “postulados” o “premisas”.

1. Axioma mayor: todos los hombres son mortales.
2. Axioma menor: Sócrates es un hombre.
3. Conclusión: Sócrates es mortal.

De las dos premisas 1 y 2 se deduce que Sócrates, por ser un hombre, es un ser mortal.

Otro silogismo es el siguiente:

1. Axioma mayor: ningún héroe es un cobarde.
2. Axioma menor: algunos soldados son cobardes.
3. Conclusión: algunos soldados no son héroes.

Porque, por la premisa mayor, una condición necesaria para ser héroe es no ser cobarde, pero algunos soldados lo son.

En general, el razonamiento matemático cumple la siguiente estructura deductiva:

1. Se define formalmente un objeto o colección de objetos asignándole unos axiomas.
2. De esos axiomas se derivan mediante el razonamiento lógico nuevas propiedades de esos objetos o relaciones entre ellos:
 - a) Proposiciones: afirmaciones que se deducen directamente de un conjunto de axiomas.
 - b) Teoremas: afirmaciones más importantes que las proposiciones y habitualmente más difíciles de demostrar.
 - c) Corolarios: una consecuencia directa de un teorema para un caso específico.
 - d) Lemas: proposiciones secundarias útiles para demostrar un determinado teorema.

En lo que sigue, se describirán las técnicas de razonamiento lógico más extendidas para demostrar lógicamente la veracidad de proposiciones, lemas, teoremas y corolarios matemáticos a partir de un conjunto de axiomas, premisas o proposiciones previamente enunciadas y demostradas.

A.3.1 Demostración directa

Se trata del método más simple para demostrar el cumplimiento de una proposición a partir de unas premisas iniciales o de otras proposiciones previamente demostradas. Consiste en aplicar estas de manera secuencial hasta probar la veracidad de lo que se desea demostrar.

Como ejemplo, se mostrará cómo a partir de las cuatro propiedades axiomáticas de los grupos, puntos 1 a 4 de la Definición B.1.1, puede deducirse que, dado un par $(E, *)$ con estructura de grupo, a cada elemento de E le corresponde una única inversa respecto del operador “*”. Supóngase que

un elemento $a \in E$ tiene dos inversas que se denotarán a^{-1} y \hat{a}^{-1} . Aplicando sucesivamente las diversas propiedades axiomáticas de los grupos sobre a^{-1} :

$$\begin{aligned} a^{-1} &\stackrel{(1)}{=} a^{-1} * i \stackrel{(2)}{=} a^{-1} * (a * \hat{a}^{-1}) \stackrel{(3)}{=} \\ &\stackrel{(3)}{=} (a^{-1} * a) * \hat{a}^{-1} \stackrel{(4)}{=} i * \hat{a}^{-1} \stackrel{(5)}{=} \hat{a}^{-1} \end{aligned} \tag{A.57}$$

En el paso (1) se aplica la propiedad axiomática de la presencia del elemento identidad, punto 3 de la Definición B.1.1. En (2) se aplica la propiedad de existencia de inversa, punto 4 de la misma definición. En (3) se aplica la propiedad asociativa, punto 2. En (4) se aplica de nuevo la propiedad de existencia de inversa, punto 4. Finalmente, en (5) se aplica la propiedad de existencia del elemento identidad, punto 3.

La conclusión a la que se llega es que $a^{-1} = \hat{a}^{-1}$ para todo $a \in E$, demostrándose que cada elemento de E le corresponde una y sólo una inversa respecto del operador “*”.

A.3.2 Demostración por inducción

Este método puede aplicarse para demostrar la veracidad de proposiciones que dependen de un índice o variable k para cada uno de sus posibles valores, siendo este un número natural que puede o no incluir el cero. En el primer caso, k pertenecería al conjunto $\mathbb{N} = \{0, 1, 2, \dots\}$ de los números naturales, mientras que en el segundo caso, k pertenecería al conjunto de los números naturales positivos $\mathbb{N}^+ = \{1, 2, 3, \dots\}$.

Asumiendo que k es una variable que toma valores de entre el conjunto \mathbb{N} , el método de la inducción tiene su base en el siguiente teorema:

Teorema A.3.1 — Principio de la inducción matemática

Sea $\mathcal{P}(k)$ una proposición que es función de un índice natural $k \in \mathbb{N}$, $\mathbb{N} = \{0, 1, 2, \dots\}$. Si se cumplen estas dos afirmaciones:

1. $\mathcal{P}(0)$ es una proposición verdadera.
2. Para todo $k > 0$, la veracidad de $\mathcal{P}(k)$ implica la veracidad de $\mathcal{P}(k + 1)$.

Se tiene que la proposición $\mathcal{P}(k)$ es verdadera para todo $k \in \mathbb{N}$.

Si la variable k tomara valores de entre el conjunto \mathbb{N}^+ , el teorema anterior seguiría siendo válido. Solamente habría que cambiar $\mathcal{P}(0)$ por $\mathcal{P}(1)$ y la condición $k > 0$ por $k > 1$.

En resumen, el método consiste en lo siguiente: se comienza demostrando que el “caso base” $\mathcal{P}(0)$ se cumple, y a continuación se prueba que, para cualquier valor de k mayor que cero, el cumplimiento de $\mathcal{P}(k)$, que se asume por hipótesis, implica el cumplimiento de la proposición $\mathcal{P}(k + 1)$. A esto último se le llama “paso inductivo”.

El principio de inducción matemática puede aplicarse para demostrar la veracidad de la siguiente fórmula para la suma finita de k términos que

Observación A.3.2 El símbolo " \sum ", llamado "sumatorio", sirve para representar de forma abreviada sumas de términos que dependen de una variable llamada "índice" desde un valor inicial de este hasta un valor final. Por ejemplo, la ecuación (A.58) representa una suma compuesta por k términos que son función de la variable $n \in \mathbb{N}$. En cada sumando, n adopta un valor diferente desde 1 hasta k . Si k fuese infinito, el sumatorio representaría una serie infinita. Otro símbolo similar a este es el llamado "productorio", denotado como " \prod ", que representa abreviadamente productos en lugar de sumas.

varían desde 1 hasta k :

$$\sum_{n=1}^k n = 1 + 2 + \dots + k = \frac{k(k+1)}{2} \quad (\text{A.58})$$

para todo $k \in \mathbb{N}$, $k \geq 1$.

1. Caso base, $k = 1$: simplemente sustituyendo $k = 1$ en la expresión (A.58) se demuestra que el caso $\mathcal{P}(1)$ se cumple. Es decir que la fórmula (A.58) es verdadera para $k = 1$:

$$\sum_{n=1}^1 n = \frac{1(1+1)}{2} = \frac{1 \cdot 2}{2} = \frac{2}{2} = 1 \Rightarrow \mathcal{P}(1) \text{ es verdadera.} \quad (\text{A.59})$$

2. Paso inductivo, $k > 1$: se comprueba que la veracidad de $\mathcal{P}(k)$ implica la veracidad de $\mathcal{P}(k+1)$ sea cual sea el valor de k . Asumiendo que la fórmula (A.58) se cumple en el caso k , se sustituye k por $k+1$ y se reordenan términos, obteniendo:

$$\begin{aligned} \sum_{n=1}^{k+1} n &= (1 + 2 + \dots + k) + (k+1) = \\ &= \frac{k(k+1)}{2} + (k+1) = \frac{k(k+1)}{2} + \frac{2(k+1)}{2} = \\ &= \frac{(2+k)(k+1)}{2} = \frac{(k+1)((k+1)+1)}{2} \end{aligned} \quad (\text{A.60})$$

Así, se comprueba que la veracidad del caso k implica que la fórmula (A.58) se cumple también en el caso $k+1$.

Con este razonamiento, puede darse por demostrada la identidad (A.58).

A.3.3 Demostración por contraposición

Esta técnica permite demostrar implicaciones simples del tipo $\mathcal{P} \Rightarrow \mathcal{Q}$ aplicando la equivalencia entre toda implicación simple y su contrapositivo, Proposición A.1.1:

$$(\mathcal{P} \Rightarrow \mathcal{Q}) \Leftrightarrow (\neg \mathcal{Q} \Rightarrow \neg \mathcal{P}) \quad (\text{A.61})$$

Es un método útil en casos en los que el contrapositivo resulte más fácil, o incluso trivial de demostrar en relación con la implicación original. Puede aplicarse, por ejemplo, para demostrar que la proposición:

$$x \cdot y \text{ no es divisible por } 5 \Rightarrow \text{ni } x \text{ ni } y \text{ son divisibles por } 5 \quad (\text{A.62})$$

Es verdadera $\forall x, y \in \mathbb{N}$. Su contrapositivo es la implicación:

$$x \text{ ó } y \text{ es divisible por } 5 \Rightarrow x \cdot y \text{ es divisible por } 5 \quad (\text{A.63})$$

Supóngase que x es divisible por 5. En tal caso, x podría factorizarse como $x = 5c$ para algún $c \in \mathbb{N}$. Multiplicando ambos miembros de la identidad $x = 5c$ por el número natural y se obtiene $xy = 5cy$, de lo que se concluye que el producto xy también es divisible por 5. Invirtiendo los roles

de x, y y repitiendo el razonamiento anterior se termina de demostrar por deducción directa la implicación (A.63) para toda pareja de valores $x, y \in \mathbb{N}$, proposición que equivale a la implicación (A.62) que se deseaba probar.

A.3.4 Demostración por contradicción: reducción al absurdo

Este método consiste en probar la veracidad de una proposición \mathcal{P} demostrando que el no cumplimiento de \mathcal{P} conlleva consecuencias absurdas o lógicamente contradictorias. Esto puede esquematizarse de manera lógica mediante el siguiente esquema:

$$\neg \mathcal{P} \Rightarrow (\mathcal{Q} = \neg \mathcal{Q}) \quad (\text{A.64})$$

La reducción al absurdo puede emplearse para demostrar que el número $\sqrt{2}$ no es un número racional. Es decir, $\sqrt{2} \notin \mathbb{Q}$. Supóngase que $\sqrt{2}$ es un número racional. En ese caso, podría escribirse:

$$\sqrt{2} = \frac{a}{b} \quad (\text{A.65})$$

Para alguna pareja de valores $a, b \in \mathbb{Z}$ mutuamente primos, o sea, sin factores comunes mayores que uno. Reordenando y elevando al cuadrado:

$$b\sqrt{2} = a \Rightarrow (b\sqrt{2})^2 = a^2 \Rightarrow 2b^2 = a^2 \quad (\text{A.66})$$

El término $2b^2$ es evidentemente par, por lo que a^2 es también par. Si a^2 es par, también lo es a , y este podría factorizarse como $a = 2c$ para algún $c \in \mathbb{Z}$. Sustituyendo en la expresión (A.66) y elevando al cuadrado:

$$2b^2 = (2c)^2 \Rightarrow 2b^2 = 4c^2 \Rightarrow b^2 = 2c^2 \quad (\text{A.67})$$

Por un razonamiento análogo al aplicado sobre la identidad $2b^2 = a^2$, se tiene que b es par. Así, se concluye que si fuese posible escribir $\sqrt{2}$ como un cociente a/b , a y b deberían ser ambos números pares, por lo que tendrían 2 como factor común, de tal forma que sería imposible encontrar dos números enteros mutuamente primos cuyo cuadrado de su cociente sea igual a dos. Esta contradicción implica que $\sqrt{2}$ no es un número racional.

A.3.5 Demostración constructiva

Se trata de un método consistente en demostrar la existencia de un objeto matemático con unas propiedades determinadas proponiendo un método para crear ese objeto con las propiedades deseadas. El proceso de *Gram-Schmidt*, Teorema 5.4.1, y la factorización QR, Teorema 5.4.2, son ejemplos de demostraciones constructivas contenidos en este libro.

A esta clase de demostraciones pertenecen las llamadas “demostraciones por contraejemplo”, que como su nombre indica consisten en demostrar la falsedad de una proposición o de una implicación aportando un caso en el que esta no se cumpla. Un ejemplo trivial es el siguiente: para demostrar que la proposición “todos los números primos son impares” es falsa, puede hacerse notar que el número dos es par y al mismo tiempo primo, pues sus únicos divisores son los números uno y dos.



B

Grupos, anillos y cuerpos

Cuando se definen conjuntos de objetos matemáticos y operaciones entre sus elementos, es común encontrar que operaciones que no tienen ninguna relación entre si comparten propiedades como, por ejemplo, la propiedad conmutativa, la asociativa y la distributiva. Esto motiva que se definan conjuntos y operaciones matemáticas de manera abstracta y axiomática, sin necesidad de describir los elementos que los conforman ni de formular las operaciones que los relacionan, limitándose a enunciar una serie de premisas de las que se puedan derivar por deducción lógica todas sus propiedades y características.

Tres de estas generalizaciones son los grupos, los anillos y los cuerpos. En cada una de ellas, se establece una lista de axiomas que se han de cumplir para que un conjunto de elementos dado, junto con una o varias operaciones matemáticas, sea considerado un grupo, un anillo o un cuerpo. Partiendo de esos axiomas, se pueden deducir propiedades que objetos matemáticos y operaciones muy diferentes entre si comparten sin necesidad de considerar las características individuales de los objetos estudiados ni las fórmulas de cada operación.

B.1 Grupos	265
B.2 Anillos	267
B.3 Cuerpos	268

B.1 Grupos

Definición B.1.1 – Grupo. Grupo conmutativo o abeliano

Dado un conjunto E y una operación “ $*$ ” entre elementos de E , se dice que el par $(E, *)$ tiene estructura de grupo si el operador “ $*$ ” cumple al menos las cuatro propiedades siguientes:

1. Propiedad de clausura: $\forall a, b \in E$

$$a * b \in E \tag{B.1}$$

2. Propiedad asociativa: $\forall a, b, c \in E$

$$(a * b) * c = a * (b * c) \tag{B.2}$$

3. Existencia del elemento identidad: existe un elemento en E llamado elemento identidad y denotado i , tal que, $\forall a \in E$

$$a * i = i * a = a \tag{B.3}$$

4. Existencia de inversa: a cada $a \in E$ le corresponde un elemento inverso, denotado a^{-1} y perteneciente también a E , tal que

$$a * a^{-1} = a^{-1} * a = i \quad (\text{B.4})$$

Si además el operador “ $*$ ” cumple la siguiente propiedad:

5. Propiedad conmutativa: $\forall a, b \in E$

$$a * b = b * a \quad (\text{B.5})$$

Se dice que el grupo $(E, *)$ es conmutativo o abeliano.

Estos son algunos ejemplos:

1. El conjunto \mathbb{Z} de los números enteros, junto con su suma, forman el grupo $(\mathbb{Z}, +)$, que además es abeliano, pues por el punto 5 de la Definición B.1.1 la suma de números enteros es conmutativa. El elemento identidad de la suma de enteros, punto 3 de la misma definición, es el cero, $0 \in \mathbb{Z}$, mientras que el inverso de $x \in \mathbb{Z}$ respecto de la suma, punto 4, es $-x \forall x \in \mathbb{Z}$. Las tuplas $(\mathbb{Q}, +)$, $(\mathbb{R}, +)$ y $(\mathbb{C}, +)$ también constituyen grupos abelianos.
2. El producto de números enteros no cumple el cuarto axioma, punto 4 de la Definición B.1.1, de existencia de la inversa, pues ningún elemento de \mathbb{Z} , a excepción de $x = 1$, admite inversa respecto del producto: no existe ningún $x^{-1} \in \mathbb{Z}$, $x \neq 1$ tal que $x \cdot x^{-1} = x^{-1} \cdot x = 1$, donde 1 es el elemento identidad del producto de enteros, punto 3 de la misma definición. Por tanto, el par (\mathbb{Z}, \cdot) no constituye un grupo.
3. Los pares (\mathbb{Q}, \cdot) , (\mathbb{R}, \cdot) y (\mathbb{C}, \cdot) tampoco constituyen grupos, pues no todos los elementos de \mathbb{Q} , \mathbb{R} y \mathbb{C} admiten inversa respecto del producto. En concreto, no existe ningún x tal que $0 \cdot x = 1$, por lo que no se cumple el punto 3 de la Definición B.1.1. Sin embargo, los pares $(\mathbb{Q} \setminus \{0\}, \cdot)$, $(\mathbb{R} \setminus \{0\}, \cdot)$ y $(\mathbb{C} \setminus \{0\}, \cdot)$ sí que forman grupos porque, excluido el cero, todos sus elementos admiten inversa respecto del producto.
4. Conjuntos de matrices con m filas y n columnas junto con la suma matricial, Definición 2.2.1, forman grupos abelianos denotados $(\mathbb{K}^{m \times n}, +)$. El elemento identidad es en este caso la matriz nula con m filas y n columnas, Definición 2.1.8, mientras que el elemento inverso asociado a cada matriz $A \in \mathbb{K}^{m \times n}$ es su opuesto $-A$, Definición 2.1.11.
5. Las propiedades axiomáticas 1 a 5 de la suma de elementos de un espacio vectorial, Definición 1.1.1, coinciden con las de los grupos abelianos, puntos 1 a 5 de la Definición B.1.1, por lo que puede decirse que el par $(V, +)$, donde V es un espacio vectorial y “ $+$ ” representa la suma de elementos en V , forman un grupo abeliano.
6. Conjuntos de matrices cuadradas de igual orden y con rango máximo, Definición 3.4.6, junto con el producto matricial, Definición 2.2.3, forman grupos no abelianos debido a la propiedad 1 de no con-

mutatividad del producto matricial, Proposición 2.2.3. El elemento identidad del producto matricial es la matriz identidad, Definición 2.1.7, mientras que el elemento inverso asociado a cada matriz cuadrada de rango máximo respecto del producto matricial es su matriz inversa, Definición 2.5.1.

7. Otro ejemplo de grupo muy curioso es el conjunto de todos los movimientos que se pueden realizar sobre las caras de un cubo de Rubik junto con el operador de concatenación, o combinación, de movimientos. Gracias a la teoría de grupos, que es el campo de la matemática que estudia las propiedades que se derivan por deducción lógica de los axiomas 1 a 5 de grupo, Definición B.1.1, se sabe que es teóricamente posible resolver cualquier cubo de Rubik, sea cual sea su configuración de partida, con un número igual o menor de 20 movimientos.

B.2 Anillos

Definición B.2.1 – Anillo. Anillos asociativo y conmutativo. Anillo con identidad

Dado un conjunto E y dos operaciones “+” (comúnmente llamada suma o adición) y “·” (comúnmente llamada producto o multiplicación) entre elementos de E , se dice que la terna $(E, +, \cdot)$ tiene estructura de anillo si los operadores “+” y “·” cumplen al menos las tres propiedades siguientes:

1. $(E, +)$ tiene estructura de grupo abeliano, Definición B.1.1, con un elemento identidad (comúnmente llamado neutro o cero) denotado 0 y un inverso aditivo para cada $a \in E$ (habitualmente llamado opuesto) denotado $-a$.

2. Propiedad de clausura del operador “·”: $\forall a, b \in E$

$$a \cdot b \in E \quad (\text{B.6})$$

3. Propiedad distributiva del operador “·” respecto del operador “+” por la derecha y por la izquierda: $\forall a, b, c \in E$

$$a \cdot (b + c) = a \cdot b + a \cdot c \quad (\text{B.7})$$

$$(a + b) \cdot c = a \cdot c + b \cdot c \quad (\text{B.8})$$

Si además el operador “·” cumple la siguiente propiedad:

4. Propiedad asociativa del operador “·”: $\forall a, b, c \in E$

$$(a \cdot b) \cdot c = a \cdot (b \cdot c) \quad (\text{B.9})$$

Se dice que el anillo $(E, +, \cdot)$ es asociativo.

Si el operador “ \cdot ” cumple la siguiente propiedad:

5. Propiedad conmutativa del operador “ \cdot ”: $\forall a, b \in E$

$$a \cdot b = b \cdot a \quad (\text{B.10})$$

Se dice que el anillo $(E, +, \cdot)$ es conmutativo.

Y si “ \cdot ” cumple la siguiente propiedad:

6. Existencia del elemento identidad respecto de la operación “ \cdot ”: existe un elemento en E llamado elemento identidad, o unidad, y denotado 1 (no necesariamente coincidente con el elemento neutro de E asociado a la operación “ $+$ ”) tal que, $\forall a \in E$

$$a \cdot 1 = 1 \cdot a = a \quad (\text{B.11})$$

Se dice que $(E, +, \cdot)$ es un anillo con identidad.

Algunos ejemplos son:

1. Los conjuntos \mathbb{Z} , \mathbb{Q} , \mathbb{R} y \mathbb{C} , junto con la suma y el producto habituales en estos conjuntos, constituyen anillos asociativos, conmutativos y con identidad.
2. La terna $(\mathbb{K}^{n \times n}, +, \cdot)$ formada por el conjunto de todas las matrices cuadradas de orden n en el cuerpo \mathbb{K} con las operaciones de suma de matrices, Definición 2.2.1, y producto matricial, Definición 2.2.3, constituye un anillo asociativo y con identidad pero no conmutativo, pues el producto matricial no lo es por la propiedad 1 de la Proposición 2.2.3. El elemento identidad de este anillo es la matriz identidad de orden n , Definición 2.1.7.
3. Debido a la correspondencia entre matrices cuadradas e isomorfismos, punto 1 de la Definición 3.1.7, dado un espacio vectorial V cualquiera, la terna $(f : V \mapsto V, +, \circ)$ también constituye un anillo asociativo, con identidad y no conmutativo, siendo “ $+$ ” la suma de aplicaciones, ecuación (1.23), y siendo \circ la composición de aplicaciones, Definición 3.5.1.

B.3 Cuerpos

Definición B.3.1 – Cuerpo

Dado un conjunto E y dos operaciones “ $+$ ” (comúnmente llamada suma o adición) y “ \cdot ” (comúnmente llamada producto o multiplicación) entre elementos de E , se dice que la terna $(E, +, \cdot)$ tiene estructura de cuerpo si $(E, +, \cdot)$ es un anillo con identidad tanto asociativo como conmutativo, puntos 1 a 6 de la Definición B.2.1, y si el operador “ \cdot ” también admite inversa. Es decir, si se cumple la siguiente propiedad:

7. Existencia de inversa multiplicativa: a cada $a \in E$, $a \neq 0$, donde 0 denota el elemento neutro asociado a la operación “ $+$ ”, punto 1 de la Definición B.2.1, le corresponde un elemento inverso,

denotado a^{-1} y perteneciente también a E , tal que:

$$a \cdot a^{-1} = a^{-1} \cdot a = 1 \quad (\text{B.12})$$

Donde 1 representa el elemento identidad, o unidad, asociado al operador “ \cdot ”, punto 6 de la Definición B.2.1.

El axioma 4 de existencia de inversa aditiva, y el axioma 1 de existencia de inversa multiplicativa, B.1.1, permiten definir dos nuevas operaciones comúnmente denominadas “sustracción” y “división”:

Definición B.3.2 – Sustracción y división de elementos de un cuerpo

Dados un cuerpo $(E, +, \cdot)$ y dos elementos $a, b \in E$, se definen:

1. La sustracción entre a y b como aquel elemento $x \in E$ que cumple $x + b = b + x = a$. Es decir, como la suma entre a y el opuesto de b .

$$x = a - b = a + (-b) \quad (\text{B.13})$$

2. La división entre a y b , con $b \neq 0$, como aquel elemento $x \in E$ que cumple $x \cdot b = b \cdot x = a$. Es decir, como el producto entre a y la inversa multiplicativa de b .

$$x = \frac{a}{b} = a \cdot b^{-1} \quad (\text{B.14})$$

Los anillos $(\mathbb{Q}, +, \cdot)$, $(\mathbb{R}, +, \cdot)$ y $(\mathbb{C}, +, \cdot)$ cumplen la propiedad 7 de existencia de inversa respecto del producto, Definición B.2.1, por lo que constituyen cuerpos a los que se denota, respectivamente, con los símbolos \mathbb{Q} , \mathbb{R} y \mathbb{C} . Respecto a los números que contienen cada uno de estos conjuntos, recuérdese que:

1. El conjunto \mathbb{Q} de los números racionales está formado por todos los cocientes posibles entre números enteros no necesariamente divisibles entre sí.
2. El conjunto \mathbb{R} de los números reales está formado por la unión del conjunto \mathbb{Q} con el conjunto formado por aquellos números que, a pesar de no poderse expresar como el cociente de dos números enteros, son límites de sucesiones de números racionales. Dos ejemplos son los números π y e , que pueden expresarse como las siguientes series infinitas:

$$\pi = 4 \cdot \sum_{n=0}^{\infty} \frac{(-1)^n}{2n+1} \quad (\text{B.15})$$

$$e = \sum_{n=0}^{\infty} \frac{1}{n!} \quad (\text{B.16})$$

A números como estos se les conoce como “números irracionales”.

El conjunto de los números complejos, así como sus operaciones asociadas de suma y producto, se definen en el Apéndice C.

Sin embargo, el anillo $(\mathbb{Z}, +, \cdot)$ no puede constituir un cuerpo porque no cumple la propiedad de existencia de inversa respecto del producto, como ya se había razonado en el Ejemplo 2 del apartado B.1.



C

El cuerpo de los números complejos

C.1 Introducción

Los llamados “números complejos” son un artificio propuesto por primera vez en el siglo XIV por pioneros como el matemático italiano Gerolamo Cardano para garantizar que todos los polinomios tuviesen raíces. Como ejemplo, considérese el siguiente polinomio:

$$P(x) = x^2 - 1 \quad (\text{C.1})$$

Claramente, si x es igual a 1 ó -1 , el polinomio vale cero: $P(x) = 0$, por lo que estas son las dos raíces del polinomio, que por supuesto son números reales.

Considérese ahora el siguiente polinomio:

$$Q(x) = x^2 + 1 \quad (\text{C.2})$$

Igualando $Q(x)$ a cero y espejando la incógnita x , se tiene que x es raíz del polinomio si cumple:

$$x^2 + 1 = 0 \Rightarrow x^2 = -1 \Rightarrow x = \sqrt{-1} \notin \mathbb{R} \quad (\text{C.3})$$

No existe ningún número real cuyo cuadrado sea igual a -1 , por lo que $Q(x)$ no tiene raíces en \mathbb{R} . Si existiese un cuerpo de escalares que incluyese todos los números reales y que contuviese al menos un elemento cuyo cuadrado fuese igual al número real -1 , la ecuación $Q(x) = 0$ tendría soluciones en ese cuerpo, y el polinomio $Q(x)$ sí que tendría raíces. Aquí es donde entran en juego los números complejos.

Pero estos números son mucho más que un simple artificio para expresar raíces de polinomios. Sus propiedades, especialmente su relación con las funciones ondulatorias, han hecho de ellos una herramienta de enorme utilidad para la modelización y la resolución de problemas en matemática aplicada, física e ingeniería. Se explicará por qué al finalizar la sección. Antes, se definirán el conjunto de los números complejos, unas operaciones de suma y producto que dotan al conjunto de los números complejos de una estructura de cuerpo, Definición B.3.1, los conceptos de módulo, conjugación y argumento de un número complejo y las fórmulas de *Euler* y de *Moivre*, así como el cálculo de raíces de números complejos.

C.1	Introducción	271
C.2	Definición de número complejo	272
C.3	Suma y producto de números complejos . . .	273
C.4	El cuerpo de los números complejos	275
C.5	Conjugación de números complejos	276
C.6	Módulo de un número complejo	276
C.7	Inversa y cociente de números complejos . . .	277
C.8	Argumento y forma trigonométrica de un número complejo	278
C.9	Identidad de <i>Euler</i> . . .	280
C.10	Fórmula de <i>De Moivre</i> .	281
C.11	Raíces cuadradas y enésimas de un número complejo	282
C.12	Los números complejos en la ingeniería	283
C.13	Resumen	285

C.2 Definición de número complejo

A continuación, se definirán los números complejos como elementos de un conjunto al que se denotará \mathbb{C} .

Definición C.2.1 — El conjunto de los números complejos

El conjunto de los números complejos se define como:

$$\mathbb{C} = \{z : z = a + bi, a, b \in \mathbb{R}, i^2 = i \cdot i = -1\}$$

Es decir, es el conjunto de todos los elementos que se pueden representar de la forma $a + bi$, donde a, b son parámetros escogidos de entre el conjunto de los números reales, mientras que i es un elemento denominado “unidad imaginaria” que cumple la propiedad axiomática $i^2 = i \cdot i = -1$.

A esta forma de representar un número complejo se la llama “forma binómica”:

Definición C.2.2 — Forma binómica de un número complejo

A la expresión del número complejo $z \in \mathbb{C}$ como $a + bi$, donde $a, b \in \mathbb{R}$ e i denota la unidad imaginaria, Definición C.2.1, se le conoce como forma binómica de z .

A los parámetros reales a, b que caracterizan un número complejo z se les llama, respectivamente, parte real e imaginaria de z :

Definición C.2.3 — Parte real e imaginaria de un número complejo

Sea z un número complejo cuya forma binómica es $a + bi$, $a, b \in \mathbb{R}$, se llama:

1. Parte real de z , y se denota $\operatorname{Re} z$, al coeficiente $a \in \mathbb{R}$:

$$\operatorname{Re} z = a \tag{C.4}$$

2. Parte imaginaria de z , y se denota $\operatorname{Im} z$, al coeficiente $b \in \mathbb{R}$:

$$\operatorname{Im} z = b \tag{C.5}$$

Todo número complejo $z = a + bi$ queda determinado unívocamente por los dos números reales (a, b) que lo definen. Así, puede identificarse geoméricamente el conjunto de los números complejos con el plano \mathbb{R}^2 , de modo tal que a cada número complejo se le asocia un único punto en \mathbb{R}^2 ; y viceversa, a cada punto en \mathbb{R}^2 le corresponde un único número complejo z .

Proposición C.2.1 — Relación biunívoca entre \mathbb{C} y \mathbb{R}^2

Un número complejo $z \in \mathbb{C}$ queda determinado unívocamente por su parte real $a = \operatorname{Re} z$ y por su parte imaginaria $b = \operatorname{Im} z$, por lo que es posible establecer una relación biunívoca entre cada elemento $z \in \mathbb{C}$ y el punto $(\operatorname{Re} z, \operatorname{Im} z) = (a, b)$ del plano euclidiano \mathbb{R}^2 .

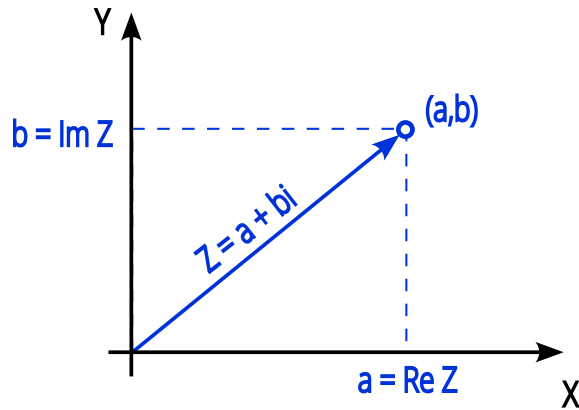


Figura C.1: Representación del número complejo $z = a + bi$ como un punto del plano euclidiano \mathbb{R}^2 con coordenadas $a = \operatorname{Re} z$, $b = \operatorname{Im} z$. A la representación gráfica de un número complejo como un vector fijo en \mathbb{R}^2 se la suele conocer en física e ingeniería como el fador de z .

En virtud de esta proposición, puede decirse que \mathbb{C} y \mathbb{R}^2 son conjuntos equivalentes (Figura C.1) o, utilizando la nomenclatura de la Definición 3.1.7, isomorfos.

C.3 Suma y producto de números complejos

Como se verá a continuación, la suma y el producto de números complejos se asemeja a la suma y producto de binomios.

Definición C.3.1 – Suma de números complejos

Se define la suma $z + w$ de dos números complejos $z = a + bi$, $w = c + di$ como el número complejo dado por:

$$\begin{aligned} z + w &= (a + bi) + (c + di) = a + c + bi + di \\ &= (a + c) + (b + d)i \end{aligned}$$

La suma de números complejos se efectúa agrupando sus partes reales e imaginarias. Por la equivalencia entre \mathbb{C} y \mathbb{R}^2 que establece la Proposición C.2.1, sumar números complejos puede visualizarse como la suma de vectores en el plano euclidiano mediante la regla del paralelogramo.

Definición C.3.2 – Producto de números complejos

Se define el producto $z \cdot w$ de dos números complejos $z = a + bi$, $w = c + di$ como el número complejo dado por:

$$\begin{aligned} z \cdot w &= (a + bi)(c + di) = ac + adi + cbi + bdi^2 = \\ &= (ac - bd) + (ad + bc)i \end{aligned}$$

Observación C.3.1 El símbolo producto “ \cdot ” se suele omitir por simplicidad.

La multiplicación de números complejos tiene, al igual que la suma, una interpretación geométrica que se estudiará más adelante, en el apartado C.8.

Proposición C.3.1 – Propiedades de la suma y del producto de números complejos

Tanto la suma como el producto de números complejos cumplen las siguientes propiedades $\forall z, w, u \in \mathbb{C}$:

1. Propiedad de clausura de la suma y del producto.

$$z + w \in \mathbb{C} \quad (\text{C.6})$$

$$zw \in \mathbb{C} \quad (\text{C.7})$$

2. Propiedad conmutativa de la suma y del producto.

$$z + w = w + z \quad (\text{C.8})$$

$$zw = wz \quad (\text{C.9})$$

3. Propiedad asociativa de la suma y del producto.

$$(z + w) + u = z + (w + u) \quad (\text{C.10})$$

$$(zw)u = z(wu) \quad (\text{C.11})$$

4. Propiedad distributiva del producto respecto a la suma.

$$z(w + u) = zw + zu \quad (\text{C.12})$$

$$(w + u)z = wz + uz \quad (\text{C.13})$$

5. Existencia de elementos identidad respecto de la suma y del producto: $\forall z \in \mathbb{C}$

$$z = z + 0 = 0 + z \quad (\text{C.14})$$

$$z = z \cdot 1 = 1 \cdot z \quad (\text{C.15})$$

Donde $0 = 0 + 0 \cdot i$, $1 = 1 + 0 \cdot i$ son, respectivamente, el elemento identidad de la suma y del producto.

6. Existencia del elemento opuesto respecto de la suma \dagger denotado $-z$: $\forall z \in \mathbb{C}$, $z = a + bi$

$$\exists -z : z + (-z) = -z + z = 0 \quad (\text{C.16})$$

Siendo $-z = -a + (-b)i = -a - bi$.

7. Existencia del elemento inverso respecto del producto \ddagger denotado z^{-1} : $\forall z \in \mathbb{C}$, $z \neq 0$

$$\exists z^{-1} \in \mathbb{C} : zz^{-1} = z^{-1}z = 1 \quad (\text{C.17})$$

Observación C.3.2 \dagger Gracias a esta propiedad, se puede definir la resta de dos números complejos z y w como $z - w = z + (-w) = z + (-1) \cdot w$.

Observación C.3.3 \ddagger De esta propiedad se deriva la operación cociente entre dos números complejos z y w . Se presentará en el apartado C.7, tras haber definido con mayor detenimiento la inversa de un número complejo.

Se deja como ejercicio † para el lector la demostración de estas propiedades.

Ejercicio C.3.1 † Demostrar que la suma, Definición C.8.1, y el producto, Definición C.3.2, de números complejos cumplen estas propiedades.

C.4 El cuerpo de los números complejos

Las propiedades 1 a 7 de la suma y el producto de números complejos, Proposición C.3.1, dotan al conjunto de los números complejos de una estructura de cuerpo, Definición B.3.1. Por ello, de ahora en adelante se hablará de \mathbb{C} como del cuerpo de los números complejos:

Definición C.4.1 – El cuerpo de los números complejos

Se llama cuerpo de los números complejos a la terna $(\mathbb{C}, +, \cdot)$, donde \mathbb{C} es el conjunto de los números complejos, Definición C.2.1, y los símbolos “+” y “ \cdot ” representan las operaciones suma y producto de números complejos, Definiciones C.3.1 y C.3.2.

Nótese que el conjunto de los números reales \mathbb{R} está contenido en \mathbb{C} : es el subconjunto formado por todos los números complejos cuya parte imaginaria es nula. Es decir, que cumplen $\text{Im } z = 0$, con las mismas operaciones de suma y de producto que los números complejos, que son equivalentes a la suma y el producto de números reales. Por ello, puede decirse que el cuerpo de los números complejos \mathbb{C} , con las operaciones de suma y producto de números complejos, Definiciones C.3.1 y C.3.2, es una “extensión” del cuerpo de los números reales \mathbb{R} .

Por otro lado, si $z \in \mathbb{C}$ tiene parte real nula y parte imaginaria no nula, se dice que z es un número imaginario puro.

Volviendo al polinomio $Q(x) = x^2 + 1$, despejando x de la ecuación cuadrática $Q(x) = 0$, ecuación (C.3), se tiene que x es raíz de $Q(x)$, si y sólo si, su cuadrado es igual a -1 . Por definición, el cuadrado de la unidad imaginaria, Definición C.2.1, es precisamente -1 , por lo que las dos raíces (complejas) de $Q(x)$ son:

$$x^2 + 1 = 0 \Rightarrow x^2 = -1 \Rightarrow x = \pm i \in \mathbb{C} \quad (\text{C.18})$$

En general, todos los polinomios con coeficientes reales o complejos tienen raíces en \mathbb{C} , y además, el número de raíces complejas de un polinomio coincide siempre con su grado.

Teorema C.4.1 – Teorema fundamental del álgebra

Sea $P(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0$ un polinomio con coeficientes en \mathbb{C} , $P(x)$ tiene exactamente n raíces.

Se omite la demostración de este teorema aparentemente tan simple por requerir para su comprensión de conocimientos profundos en análisis de funciones complejas.

C.5 Conjugación de números complejos

Definición C.5.1 – Conjugado de un número complejo

A cada número complejo $z = a + bi$ se le asocia el número z^* dado por:

$$z^* = a + (-1)i = a - bi$$

A $z^* \in \mathbb{C}$ se le denomina conjugado de z .

De acuerdo con esta definición, conjugar un número complejo consiste en cambiar de signo su parte imaginaria. Por la equivalencia entre el conjunto de los números complejos y el plano euclidiano, conjugar un número $z \in \mathbb{C}$ equivale a calcular el simétrico del fador de z respecto del eje de abscisas (Figura C.1).

Proposición C.5.1 – Propiedades del conjugado de un número complejo

La conjugación de números complejos cumple las siguientes propiedades $\forall z, w \in \mathbb{C}$:

1. El operador conjugación es su propio inverso.

$$(z^*)^* = z \quad (\text{C.19})$$

2. El conjugado de la suma coincide con la suma de los conjugados.

$$(z + w)^* = z^* + w^* \quad (\text{C.20})$$

3. El conjugado de la producto coincide con la producto de los conjugados.

$$(zw)^* = z^*w^* \quad (\text{C.21})$$

4. La parte real de $z \in \mathbb{C}$ puede expresarse en función de su conjugado como:

$$\operatorname{Re} z = \frac{1}{2}(z + z^*) \quad (\text{C.22})$$

5. La parte imaginaria de $z \in \mathbb{C}$ puede expresarse en función de su conjugado como:

$$\operatorname{Im} z = \frac{-i}{2}(z - z^*) = \frac{z - z^*}{2i} \quad (\text{C.23})$$

Ejercicio C.5.1 † Demostrar que el conjugado de un número complejo, Definición C.5.1, cumple estas propiedades.

La demostración de estas propiedades se deja como ejercicio † al lector.

C.6 Módulo de un número complejo

Definición C.6.1 – Módulo de un número complejo

Se define el módulo de un número complejo $z = a + bi$ como el escalar real no nulo:

$$|z| = \sqrt{z \cdot z^*} = \sqrt{a^2 + b^2}$$

El módulo de $z = a + bi \in \mathbb{C}$ coincide con la longitud de la representación gráfica de z en el plano euclidiano, el fasor de z (Figura C.1). Es decir, con la norma euclidiana del vector (a, b) , Definición 5.1.4. Por ello, las propiedades del módulo que se enumeran a continuación son deducibles tanto de su definición como de las propiedades de la norma, Definición 5.1.10:

Proposición C.6.1 – Propiedades del módulo de un número complejo

El módulo cumple las siguientes propiedades $\forall z, w \in \mathbb{C}$:

1. El módulo de un número complejo coincide con el de su conjugado.

$$|z| = |z^*| \quad (\text{C.24})$$

2. El módulo del producto coincide con el producto de los módulos.

$$|zw| = |z||w| \quad (\text{C.25})$$

3. El cuadrado del módulo de un número complejo coincide con el producto de este por su conjugado.

$$z(z^*) = |z|^2 \quad (\text{C.26})$$

4. El módulo cumple las desigualdades triangulares:

$$|\operatorname{Re} z| \leq |z| \leq |\operatorname{Re} z| + |\operatorname{Im} z| \quad (\text{C.27})$$

$$|\operatorname{Im} z| \leq |z| \leq |\operatorname{Re} z| + |\operatorname{Im} z| \quad (\text{C.28})$$

$$|z + w| \leq |z| + |w| \quad (\text{C.29})$$

Además, si $w \neq 0$:

$$\begin{aligned} |z + w| = |z| + |w| &\Leftrightarrow \\ \Leftrightarrow z = \alpha w &\text{ para algún } \alpha \in \mathbb{R}, \alpha \geq 0 \end{aligned} \quad (\text{C.30})$$

Ejercicio C.6.1 † Demostrar que el módulo de un número complejo, Definición C.6.1, cumple estas propiedades.

Su demostración se deja como ejercicio † al lector.

C.7 Inversa y cociente de números complejos

Definición C.7.1 – Inversa de un número complejo

Se dice que $z^{-1} \in \mathbb{C}$ es la inversa del número complejo z si:

$$zz^{-1} = z^{-1}z = 1$$

La propiedad 7 de la Proposición C.3.1 garantiza que cada $z \in \mathbb{C}$ distinto de cero tenga una inversa que además es única para cada z .

Las propiedades del módulo y de la conjugación, Propositiones C.6.1 y C.5.1, pueden aplicarse para calcular la inversa de un número complejo cualquiera $z = a + bi$ distinto de cero:

$$\begin{aligned} \frac{1}{z} &= \frac{z^*}{z(z^*)} = \frac{1}{|z|^2} \cdot z^* \Rightarrow \\ \Rightarrow \frac{1}{a + bi} &= \frac{(a - bi)}{(a + bi)(a - bi)} = \frac{1}{a^2 + b^2} \cdot (abi) \end{aligned} \quad (\text{C.31})$$

La propiedad 7 de existencia de inversa multiplicativa, Proposition C.3.1, permite definir la división de números complejos de la siguiente forma:

Definición C.7.2 – División de números complejos

Se dice que $z/w \in \mathbb{C}$ es el cociente de los números complejos z y w (con $w \neq 0$) si cumple:

$$w \cdot \frac{z}{w} = \frac{z}{w} \cdot w = z$$

A z se le llama dividendo, o numerador, del cociente; mientras que a w se le denomina divisor, o denominador, del cociente.

El cociente entre z y w , con $w \neq 0$, puede calcularse de modo sencillo mediante una estrategia similar a la propuesta en la ecuación (C.31) para el cálculo de inversas:

$$\begin{aligned} \frac{z}{w} &= \frac{z(w^*)}{w(w^*)} = \frac{1}{|w|^2} \cdot z(w^*) \Rightarrow \\ \Rightarrow \frac{a + bi}{c + di} &= \frac{(a + bi)(c - di)}{(c + di)(c - di)} = \frac{1}{c^2 + d^2} \cdot (a + bi)(c - di) \end{aligned} \quad (\text{C.32})$$

C.8 Argumento y forma trigonométrica de un número complejo

Debido a la correspondencia biunívoca entre números complejos y vectores en \mathbb{R}^2 (Figura C.1), todo número complejo $z = a + bi$ distinto de cero se puede caracterizar por un valor real θ que representa el ángulo del fasor de z respecto del eje de abscisas del plano euclidiano. De este modo, z se podría escribir de la siguiente forma:

$$\begin{aligned} z = a + bi &= |z| \cos \theta + |z| \operatorname{sen} \theta \cdot i = \\ &= |z| \cdot (\cos \theta + i \operatorname{sen} \theta) \end{aligned} \quad (\text{C.34})$$

A $\theta \in \mathbb{R}$ se le conoce como “argumento de z ”.

Definición C.8.1 – Argumento de un número complejo

Se dice que el número real θ es argumento, o fase †, de $z \in \mathbb{C}$, $z \neq 0$, si este cumple la identidad:

$$z = a + bi = |z| \cdot (\cos \theta + i \operatorname{sen} \theta)$$

Donde $|z|$ es el módulo de z , Definición C.6.1, y $\theta \in \mathbb{R}$ es un argumento de z .

Observación C.8.1 † Al argumento de un número complejo se le suele llamar también “fase”.

La siguiente proposición muestra dos propiedades fundamentales del argumento de un número complejo. Su demostración † se deja al lector.

Proposición C.8.1 — Propiedades del argumento

Dado $z \in \mathbb{C}$:

1. Si θ es un argumento de z , $-\theta$ lo es de z^* .
2. Si θ es un argumento de z , también lo es $\theta + 2\pi k \forall k \in \mathbb{Z}$.

Ejercicio C.8.1 † Demostrar estas dos propiedades del argumento a partir de la Definición C.8.1, ayudándose de la forma trigonométrica (C.8.1) de un número complejo si fuese necesario.

Véase que, debido a la ciclicidad de los ángulos, propiedad 2 de la Proposición C.8.1, a cada $z \neq 0$ se le pueden asignar infinitos argumentos. Para evitar esta redundancia, se caracteriza el argumento de los números complejos mediante valores contenidos en ciertos intervalos de \mathbb{R} .

Definición C.8.2 — Argumento principal y polar

Si $\theta \in \mathbb{R}$ es un argumento de $z \in \mathbb{C}$ y:

1. $\theta \in (-\pi, \pi]$, se dice que θ es el argumento principal de z .
2. $\theta \in [0, 2\pi)$, se dice que θ es el argumento polar de z .

Conocido el módulo y un argumento de $z \in \mathbb{C}$, es posible expresarlo mediante la llamada “forma trigonométrica”:

Definición C.8.3 — Forma trigonométrica de un número complejo

Se llama forma trigonométrica de $z \in \mathbb{C}$ a la expresión:

$$|z| \cdot (\cos \theta + i \operatorname{sen} \theta)$$

Donde $|z|$ es el módulo de z , y $\theta \in \mathbb{R}$ es un argumento de z , Definiciones C.6.1 y C.8.1.

Esta forma de representar un número complejo enfatiza aún más la correspondencia biunívoca entre los conjuntos \mathbb{C} y \mathbb{R}^2 (Figura C.2).

Ahora, aplíquense las fórmulas del seno y del coseno de la suma de ángulos:

$$\operatorname{sen}(\theta + \varphi) = \cos \theta \operatorname{sen} \varphi + \operatorname{sen} \theta \cos \varphi \tag{C.35}$$

$$\cos(\theta + \varphi) = \cos \theta \cos \varphi - \operatorname{sen} \theta \operatorname{sen} \varphi \tag{C.36}$$

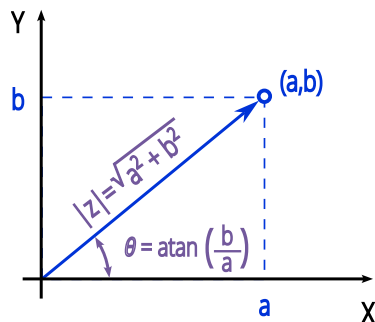


Figura C.2: Equivalencia entre el módulo y el argumento del número complejo $z = a + bi$, y la longitud y el ángulo respecto del eje de abscisas de su representación fasorial en \mathbb{R}^2 .

Al producto de dos números complejos escritos en forma trigonométrica. Se obtiene:

$$\begin{aligned} zw &= |z||w| \cdot (\cos \theta + i \operatorname{sen} \theta)(\cos \varphi + i \operatorname{sen} \varphi) = \\ &= |z||w| \cdot [\cos \theta \cos \varphi - \operatorname{sen} \theta \operatorname{sen} \varphi + \\ &\quad + i \cdot (\cos \theta \operatorname{sen} \varphi + \operatorname{sen} \theta \cos \varphi)] = \\ &= |z||w| \cdot (\cos(\theta + \varphi) + i \operatorname{sen}(\theta + \varphi)) \end{aligned} \quad (\text{C.37})$$

De este resultado se concluye que el producto de números complejos tiene la siguiente interpretación geométrica: el módulo del producto de números complejos es el producto de sus módulos, y el argumento del producto es la suma de sus argumentos.

Proposición C.8.2 – Producto de dos números complejos en forma trigonométrica

El producto de dos números complejos:

$$z = |z| \cdot (\cos \theta + i \operatorname{sen} \theta) \quad (\text{C.38})$$

$$w = |w| \cdot (\cos \varphi + i \operatorname{sen} \varphi) \quad (\text{C.39})$$

Tiene la siguiente forma trigonométrica:

$$zw = |z||w| \cdot (\cos(\theta + \varphi) + i \operatorname{sen}(\theta + \varphi)) \quad (\text{C.40})$$

C.9 Identidad de Euler

La propiedad enunciada en la Proposición C.8.2 sugiere que tal vez sea posible expresar un número complejo cualquiera $z = |z| \cdot (\cos \theta + i \operatorname{sen} \theta)$ de la forma $z = |z|e^{i\theta}$, pues el producto de funciones exponenciales cumple esta misma propiedad:

$$zw = |z|e^{i\theta} \cdot |w|e^{i\varphi} = |z||w|e^{i(\theta+\varphi)} \quad (\text{C.41})$$

Para demostrar que, efectivamente, z se puede expresar de esta forma, supóngase que $|z| \cdot (\cos \theta + i \operatorname{sen} \theta) = |z|e^{i\theta}$. Es decir, que:

$$\cos \theta + i \operatorname{sen} \theta = e^{i\theta} \quad (\text{C.42})$$

Tanto el seno como el coseno pueden expresarse mediante unas series infinitas conocidas en el análisis infinitesimal real como “polinomios de Taylor”:

$$\cos x = 1 - \frac{x^2}{2} + \frac{x^4}{24} - \dots = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} x^{2n} \quad (\text{C.43})$$

$$\operatorname{sen} x = x - \frac{x^3}{6} + \frac{x^5}{120} - \dots = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} x^{2n+1} \quad (\text{C.44})$$

$$e^x = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \dots = \sum_{n=0}^{\infty} \frac{x^n}{n!} \quad (\text{C.45})$$

Sustituyendo $x = \theta$ en las expresiones (C.43) y (C.44), $x = i\theta$ en la ecuación (C.45), y sumando, se demuestra la igualdad (C.42). A esta se la suele conocer con el nombre de “identidad de Euler” en honor del matemático que la descubrió.

Teorema C.9.1 – Identidad de Euler

Para todo $\theta \in \mathbb{R}$ se cumple:

$$\forall \theta \in \mathbb{R}, e^{i\theta} = \cos \theta + i \operatorname{sen} \theta$$

Donde i es la unidad imaginaria, Definición C.2.1.

Así pues, cualquier número complejo no nulo puede expresarse de la forma:

$$z = |z|e^{i\theta} \quad \forall \theta \in \mathbb{R} \quad (\text{C.46})$$

Expresión que se conoce como “forma exponencial de z ”.

Definición C.9.1 – Forma exponencial de un número complejo

Se llama forma exponencial de $z \in \mathbb{C}$ a la expresión:

$$|z|e^{i\theta}$$

Donde $|z|$ es el módulo de z , Definición C.6.1, y $\theta \in \mathbb{R}$ es un argumento de z , Definición C.8.1.

La posibilidad de representar números complejos en forma exponencial puede ser empleada para demostrar fácilmente † que el cociente z/w de dos números complejos z y w (con $w \neq 0$) tiene módulo igual al cociente de los módulos de z y w , y argumento igual a la diferencia entre los argumentos de z y w .

Ejercicio C.9.1 † Demostrar la Proposición C.9.2 con la ayuda de la identidad de Euler, representando z y w en forma exponencial, y de las propiedades de las potencias.

Proposición C.9.2 – Cociente de dos números complejos en forma trigonométrica

El cociente de dos números complejos:

$$z = |z| \cdot (\cos \theta + i \operatorname{sen} \theta) \quad (\text{C.47})$$

$$w = |w| \cdot (\cos \varphi + i \operatorname{sen} \varphi) \quad (\text{C.48})$$

Tiene la siguiente forma trigonométrica:

$$\frac{z}{w} = \frac{|z|}{|w|} \cdot (\cos(\theta - \varphi) + i \operatorname{sen}(\theta - \varphi)) \quad (\text{C.49})$$

C.10 Fórmula de De Moivre

Sea z un número complejo expresado en forma exponencial, si se eleva a la n -ésima potencia (con $n \in \mathbb{Z}$), se obtiene:

$$z^n = (|z|e^{i\theta})^n = |z|^n e^{in\theta} \quad (\text{C.50})$$

Resultado que puede ser expresado en forma trigonométrica como:

$$z^n = |z|^n \cdot (\cos \theta + i \operatorname{sen} \theta)^n = |z|^n \cdot (\cos n\theta + i \operatorname{sen} n\theta) \quad (\text{C.51})$$

A esta última expresión se la conoce como “fórmula de *De Moivre*”.

Proposición C.10.1 – Fórmula de *De Moivre*

La enésima potencia de un número complejo z cualquiera puede expresarse en forma trigonométrica como:

$$z^n = |z|^n \cdot (\cos n\theta + i \operatorname{sen} n\theta)$$

Donde $|z|$ es el módulo de z , Definición C.6.1, y $\theta \in \mathbb{R}$ es un argumento de este, Definición C.8.1.

C.11 Raíces cuadradas y enésimas de un número complejo

Definición C.11.1 – Raíz cuadrada de un número complejo

Se dice que $w \in \mathbb{C}$ es raíz cuadrada del número complejo z si $z = w^2$.

Si $w = |w|e^{i\beta}$ es una raíz cuadrada de $z = |z|e^{i\alpha}$, debe cumplir:

$$|z|e^{i\alpha} = (|w|e^{i\beta})^2 = |w|^2 e^{i2\beta} \quad (\text{C.52})$$

De este resultado se deduce que $|z| = |w|^2$ y que $\beta = \frac{\alpha}{2}$. Por tanto, todos los números complejos que son raíces cuadradas de z tienen modulo igual a $\sqrt{|z|}$, y por lo que respecta a los argumentos de las raíces cuadradas de z , se tiene que si α es un argumento cualquiera de z , sus raíces cuadradas han de tener como argumento cualquiera de los valores contenidos en el conjunto:

$$\left\{ \beta : \beta = \frac{\alpha}{2} + k\pi \quad \forall k \in \mathbb{Z} \right\} \quad (\text{C.53})$$

En total, cada $z \in \mathbb{C}$, $z \neq 0$, tiene dos raíces cuadradas distintas.

El resultado anterior puede generalizarse para todo $n \in \mathbb{Z}$, $n \leq 1$:

Definición C.11.2 – Raíz enésimas de un número complejo

Se dice que $w \in \mathbb{C}$ es la raíz enésima del número complejo z si $z = w^n$, siendo $n \in \mathbb{Z}$, $n \leq 1$.

w es la raíz enésima de $z \in \mathbb{C}$ si cumple:

$$|z| = |w|^n \quad (\text{C.54})$$

$$\beta = \frac{\alpha}{n} + k \frac{2\pi}{n} \quad \text{para algún } k \in \mathbb{Z} \quad (\text{C.55})$$

En total, cada $z \in \mathbb{C}$, $z \neq 0$, tiene n raíces enésimas distintas.

En la Figura C.3, se representan gráficamente las tres raíces cúbicas distintas dos a dos de $z = -1$. El resultado de elevar al cubo cualquiera de ellas es el número -1 .

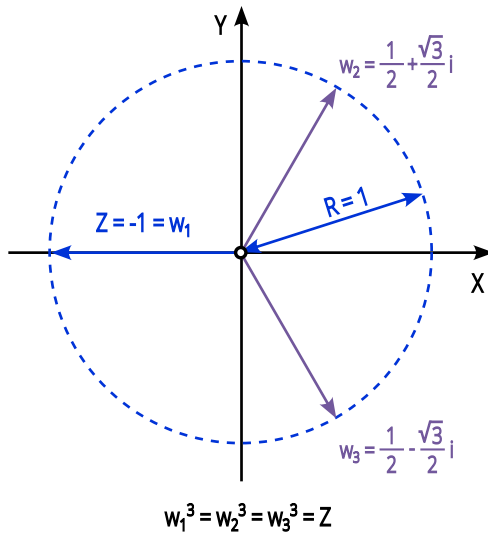


Figura C.3: Las representaciones fasoriales de las tres raíces cúbicas complejas de $z = -1$, denotadas w_1 , w_2 y w_3 . La línea discontinua representa una circunferencia de radio unidad con centro el origen del plano euclidiano. Todos los $z = a + bi \in \mathbb{C}$ que cumplen $|z| = 1$ tienen su representación en \mathbb{R}^2 contenida en esta circunferencia.

C.12 Los números complejos en la ingeniería

Los fenómenos físicos ondulatorios en una dimensión se suelen expresar matemáticamente como una suma de sinusoides:

$$x(t) = \sum_{k=1}^{\infty} A_k \cos\left(\frac{kt}{T} + \varphi\right) = \tag{C.56}$$

$$= \sum_{k=1}^{\infty} A_{1,k} \cos\left(\frac{kt}{T}\right) + \sum_{k=1}^{\infty} A_{2,k} \text{sen}\left(\frac{kt}{T}\right) \tag{C.57}$$

Donde t es una variable independiente (generalmente real), A_k , $A_{1,k}$ y $A_{2,k}$ son constantes, T es el periodo de las oscilaciones, y φ es un ángulo de desfase.

Si una oscilación es puramente sinusoidal, los términos de la ecuación (C.57) con periodo menor que T son nulos para todo instante t . Si se descartan, queda la siguiente expresión:

$$x(t) = A \cos\left(\frac{t}{T} + \varphi\right) = A_1 \cos\left(\frac{t}{T}\right) + A_2 \text{sen}\left(\frac{t}{T}\right) \tag{C.58}$$

Operar con funciones trigonométricas es, en general, muy engorroso y requiere recordar y aplicar relaciones trigonométricas como las fórmulas de los cosenos y senos del ángulo doble y mitad, o de la suma, ecuaciones (C.35) y (C.36), o diferencia de ángulos, entre muchas otras.

Afortunadamente, la identidad de *Euler*, Teorema C.9.1, permite representar fenómenos ondulatorios de forma más simple y compacta mediante fórmulas que contienen exponenciales complejas en lugar de funciones trigonométricas, pues:

$$\frac{Ae^{(\frac{t}{T}+\varphi)i} + Ae^{-(\frac{t}{T}+\varphi)i}}{2} = \operatorname{Re} Ae^{(\frac{t}{T}+\varphi)i} = A \cos\left(\frac{t}{T} + \varphi\right) \quad (\text{C.59})$$

Al ser más sencillo operar con exponenciales que con funciones trigonométricas, muchos modelos físicos y matemáticos que rigen sistemas oscilatorios contienen números complejos en sus formulaciones y descripciones. Por poner un ejemplo, el electrotecnia suelen emplearse números complejos para caracterizar las resistencias eléctricas de circuitos y máquinas de corriente alterna. A esos valores se les conoce como “impedancias”. Las partes reales de las impedancias representan resistencia térmica de efecto Joule, mientras que las partes imaginarias representan la presencia de efectos parasitarios, o bien de inductancia si son positivas, o bien de capacitancia si son negativas, que alteran el flujo de la corriente. También es común el uso de la llamada “notación de fasor” (Figura C.1) para caracterizar la amplitud y la fase de una tensión o corriente alterna.

Otras aplicaciones de los números complejos en ingeniería son:

1. En mecánica: para representar fuerzas variables en el tiempo y estudiar su efecto sobre las vibraciones y el ruido de las máquinas y las estructuras.
2. En electrónica: para diseñar filtros y amplificadores de señal.
3. En física cuántica: para representar funciones de onda de partículas subatómicas y en la ecuación de *Schrödinger*.
4. En ecuaciones diferenciales: para saber si la dinámica de un sistema físico es estable o inestable, así como los periodos y las frecuencias de las oscilaciones de dicho sistema en el tiempo.
5. En automatización: para el diseño y ajuste de los sistemas de control, el análisis de su estabilidad y la determinación de sus características.

C.13 Resumen

1. El cuerpo de los números complejos es una extensión del cuerpo de los números reales que permite que todo polinomio pueda tener al menos una raíz.
2. Todo polinomio de grado n con coeficientes reales o complejos tiene n raíces complejas (teorema fundamental del álgebra).
3. Cada número complejo es unívocamente representable por un punto en el plano \mathbb{R}^2 .
4. A cada número complejo se le puede asociar su conjugado y viene caracterizado por su parte real, su parte imaginaria, un módulo e infinitos argumentos.
5. Conocidos el módulo y alguno de los argumentos de un número complejo, este se puede representar en forma trigonométrica y exponencial.
6. Multiplicar dos números complejos equivale a multiplicar sus módulos y sumar sus argumentos. Dividirlos equivale a dividir sus módulos y restar sus argumentos.
7. Cada número complejo tiene n raíces enésimas distintas dos a dos. Los argumentos de las raíces de $z \in \mathbb{C}$ son divisores de los argumentos de z y forman polígonos regulares sobre el plano \mathbb{R}^2 .
8. A pesar de llamarse complejos, el propósito de estos números es facilitar la resolución de problemas que se presentan en física y en ingeniería.



Esta página está en blanco intencionalmente.

Bibliografía

A continuación, se indica una lista de manuales aconsejables de álgebra lineal tanto en español como en inglés que pueden complementar la lectura y el estudio de este libro.

- [1] Alberto Lastra Sedano et al. *Álgebra Lineal: Ideas, resultados, métodos y aplicaciones*. Madrid, España: Editorial Sanz y Torres S.L., 2024. ISBN: 978-8419947185.
- [2] Gilbert Strang. *Introduction to Linear Algebra*. 5th. Wellesley-Cambridge Press, 2016. ISBN: 978-1733146654.
- [3] David Poole. *Linear Algebra: A Modern Introduction*. 4th. Cengage Learning, 2014. ISBN: 978-6075263113.
- [4] David C. Lay, Steven R. Lay y Judi J. McDonald. *Linear Algebra and Its Applications*. 5th. Pearson, 2015. ISBN: 978-0321982384.
- [5] Otto Bretscher. *Linear Algebra with Applications*. 5th. Pearson, 2012. ISBN: 978-0321796974.
- [6] Jim Hefferon. *Linear Algebra*. 3rd. Orthogonal Publishing, 2017. ISBN: 978-1944325039.
- [7] Bernard Kolman y David R. Hill. *Álgebra Lineal con Aplicaciones*. 9na. Pearson Educación, 2013. ISBN: 978-0132296540.
- [8] Howard Anton y Chris Rorres. *Elementary Linear Algebra*. 11th. Wiley, 2013. ISBN: 978-1118473504.
- [9] José Javier García. *Álgebra Lineal*. 1ra. Editorial Reverté, 2017. ISBN: 978-8497692052.
- [10] Stephen H. Friedberg, Arnold J. Insel y Lawrence E. Spence. *Linear Algebra*. 5th. Pearson, 2018. ISBN: 978-0134860244.
- [11] Sheldon Axler. *Linear Algebra Done Right*. 3rd. Springer, 2015. ISBN: 978-3319110790.
- [12] Kenneth Hoffman y Ray Kunze. *Linear Algebra*. 2nd. Pearson, 2015. ISBN: 978-9332550070.
- [13] Carl D. Meyer. *Matrix Analysis and Applied Linear Algebra*. 1st. SIAM, 2000. ISBN: 978-0898714548.
- [14] Seymour Lipschutz y Marc Lipson. *Schaum's Outline of Linear Algebra*. 6th. McGraw-Hill Education, 2017. ISBN: 978-1260011449.
- [15] Lloyd N. Trefethen y David Bau III. *Numerical Linear Algebra*. 1st. SIAM, 1997. ISBN: 978-0898714876.

Esta página está en blanco intencionalmente.



POLITÉCNICA

UNIVERSIDAD
POLITÉCNICA
DE MADRID



ESCUELA TÉCNICA SUPERIOR
DE INGENIEROS DE MINAS
Y ENERGÍA