MA–360: ÁLGEBRA LINEAL I

Joseph C. Várilly

Escuela de Matemática, Universidad de Costa Rica

I Ciclo Lectivo del 2021

Introducción

Este es un curso de *álgebra lineal*. Por “álgebra” se puede entender la manipulación de cantidades abstractas (numéricas o no) por reglas formales análogas a las conocidas operaciones aritméticas: $+$, $-$, $\times$, $\div$. Las variables $\boldsymbol{x}$, $\boldsymbol{y}$, $\boldsymbol{z}$, etc. que aparecen no necesariamente significan números incógnitos: más comunmente, representan arreglos o *vectores*, es decir, hileras o cadenas de números. La palabra *lineal* indica que las operaciones permitidas sobre estos arreglos de números se limitan a aquellas que no distorsionan la geometría rectilínea de la totalidad de arreglos numéricos. Aunque se omiten algunas operaciones importantes (por ejemplo, la de eligir el mayor elemento de un arreglo), el álgebra lineal es tan amplio y general que merece un curso – o dos cursos – dedicado exclusivamente a este tema.

El estudio del álgebra lineal comprende aspectos *estructurales* y *algorítmicos*. Para entender bien la teoría de espacios vectoriales y matrices, se debe poner atención también a su presentación en algoritmos. El algoritmo primordial del álgebra lineal es el método de eliminación para la resolución de ecuaciones lineales simultáneas: en este curso, se enfocará tanto su significado abstracto como sus detalles computacionales.

Como tema preparatorio, en el primer capítulo se consideran vectores con dos o más coordenadas reales y sus productos escalar y vectorial (este último sólo aparece cuando hay tres coordenadas). En el Capítulo 2, se introduce el concepto general de espacio vectorial y la noción fundamental de independencia lineal de vectores. En el tercer capítulo, etran las matrices y su relación con sistemas de ecuaciones lineales. Se estudia en detalle el algoritmo de eliminación gaussiana para la resolución de dichos sistemas. El Capítulo 4 trata de las aplicaciones lineales entre espacios vectoriales y las matrices asociadas a tales aplicaciones. El Capítulo 5 considera brevemente la noción de ortogonalidad de vectores. En el último capítulo, el tópico es el cálculo y el empleo de los determinantes de matrices.

Programa de materias

1 Vectores en el espacio euclidiano $\mathbb{R}^n$

El espacio $\mathbb{R}^n$ de n coordenadas reales. Representación geométrica de puntos por vectores. El producto escalar de dos vectores, la longitud de un vector. El producto vectorial en $\mathbb{R}^3$ y sus propiedades geométricos. Rectas y planos en $\mathbb{R}^3$. Conjuntos convexos y afines en $\mathbb{R}^n$.

2 Espacios vectoriales

El concepto de espacios vectorial; cuerpos de escalares. Dependencia e independencia lineal de conjuntos de vectores. Subespacios vectoriales, bases y dimensión de un espacio vectorial. Compleción de una base parcial. Suma directa de dos o más espacios vectoriales.

3 Matrices y ecuaciones lineales

Matrices rectangulares y cuadradas, producto de matrices. Sistemas de ecuaciones lineales y matrices. Operaciones de fila, el algoritmo de eliminación gaussiana simple. Factorización LDU de matrices. Eliminación gaussiana con pivoteo parcial. Inversión de matrices.

4 Aplicaciones lineales y sus matrices

Aplicaciones lineales. Sus matrices con respecto a un par de bases. Formas lineales y el espacio vectorial dual. Núcleo e imagen de una aplicación lineal; su rango y nulidad. Transpuesta de una aplicación lineal; espacio de columnas y espacio de filas de una matriz. El teorema de rango y nulidad. Forma escalonada de una matriz, cálculo de rangos. Cambios de bases, matrices equivalentes y matrices semejantes.

5 Espacios vectoriales euclidianos

Espacios euclidianos con productos escalares reales. Ortogonalidad y bases ortonormales, el algoritmo de Gram y Schmidt. Productos escalares complejos, matrices adjuntas.

6 Trazas y determinantes

Traza y determinante de una matriz cuadrada. Propiedades de determinantes, independencia de cambio de base. Cálculo de determinantes por expansiones de Laplace o por eliminación gaussiana. La regla de Cramer para sistemas de ecuaciones lineales.

Bibliografía

En la página del curso, se colocará semanalmente unos apuntes detallados. En la biblioteca se encuentra un gran surtido de libros sobre álgebra lineal; algunos son textos básicos con muchos ejercicios de rutina. A continuación se ofrece una lista de libros recomendables, que tratan la materia con mayor profundidad.

1 Sheldon Axler, *Linear Algebra Done Right*, 3ª edición, Springer, Cham, Suiza, 2015.

2 Feliks R. Gantmacher, *The Theory of Matrices I*, Chelsea, New York, 1959.

3 Lidia I. Goloviná, *Algebra lineal y algunas de sus aplicaciones*, Mir, Moscú, 1974.

4 Roger A. Horn y Charles R. Johnson, *Matrix Analysis*, Cambridge Univ. Press, 2013.

5 Yitzhak Katznelson y Yonatan R. Katznelson, *A (Terse) Introduction to Linear Algebra*, Americal Mathematical Society, Providence, RI, 2008.

6 Anatoly I. Maltsev, *Fundamentos de álgebra lineal*, Mir, Moscú, 1972.

7 Peter Petersen, *Linear Algebra*, Springer, New York, 2012.

8 Belkacem Said-Houari, *Linear Algebra*, Birkhäuser, Cham, Suiza, 2017.

9 Jesús Sánchez Guevara, *Álgebra lineal fundamental*, Editorial de la UCR, San José, 2020.

10 Georgi E. Shilov, *Linear Algebra*, Dover Books, Mineola, NY, 1977.

11 Yisong Yang, *A Concise Text on Advanced Linear Algebra*, Cambridge University Press, Cambridge, 2015.

12 Xingshi Zhan, *Matrix Theory*, Americal Mathematical Society, Providence, RI, 2013.

Índice general

1 Vectores en el espacio euclidiano $\mathbb{R}^n$

In mathematics and especially in physics two very different kinds of quantity present themselves. Consider, for example, mass, time, density, temperature, force, displacement of a point, velocity, and acceleration. Of these quantities some can be represented adequately by a single number [. . .] A vector *is a quantity which is considered as possessing* direction *as well as* magnitude. *A* scalar *is a quantity which is considered as possessing* magnitude *but no direction.*

— Josiah Willard Gibbs[1]

El epígrafe que introduce este capítulo está tomado del primer libro de texto que se dedicaba al tratamiento de los *vectores* en el espacio tridimensional como entes que representan cantidades físicas que tienen magnitud y direccionalidad. (Su autor principal, el físico norteamericano Gibbs, quiso enfatizar el uso del concepto de vector en mecánica.) Los dos aspectos – magnitud y direccionalidad – están captados en un juego de tres números $\boldsymbol{r} = (x, y, z)$ referidos a alguna unidad de medición.

Si se magnifica o dilata un vector por un factor de escala $c > 0$, el vector dilatada $c\boldsymbol{r} = (cx, cy, cz)$ conserva la dirección original del vector $\boldsymbol{r}$. En cambio, si se toma $c = -1$, se recibe un vector $-\boldsymbol{r} = (-x, -y, -z)$ en la dirección opuesta de $\boldsymbol{r}$ pero con la misma magnitud.

Es menos evidente como *sumar* dos vectores. Por ejemplo, si dos vectores $\boldsymbol{r} = (x, y, z)$ y $\boldsymbol{s} = (u, v, w)$ representan fuerzas ejercidas simultáneamente sobre un objeto pequeño, ¿cuál será la dirección y magnitud del movimiento resultante? Un principio mecánico, concebido por Simon Stevinus (1586) y luego por Galileo, después refinado por Isaac Newton (1684), es el llamado *paralelogramo de fuerzas*: la suma de dos vectores salientes de un origen común es la diagonal del paralelogramo determinado por los dos vectores iniciales (Figura 1.1).

También es posible considerar vectores en el plano $z = 0$ como puntos en $\mathbb{R}^2$, esto es, pares ordenadas de números reales $\boldsymbol{r} = (x, y)$ y $\boldsymbol{s} = (u, v)$. En tal caso, denotamos el origen del plano por $\mathbf{0} = (0, 0)$. Un pequeño ejercicio con triángulos congruentes muestra que si $\boldsymbol{r}$, $\mathbf{0}$, $\boldsymbol{s}$ son tres vértices de un paralelogramo, el cuarto vértice – opuesto a $\mathbf{0}$ – es

$$\boldsymbol{r} + \boldsymbol{s} = (x, y) + (u, v) := (x + u, y + v).$$

Esta ley de suma se ejemplifica en la Figura 1.1.

[1]En el libro *Vector Analysis* de J. W. Gibbs y E. B. Wilson: Yale University Press, New Haven, CT, 1901. Citado en: A. Ostermann y G. Wanner, *Geometry by Its History*, Springer, Berlin, 2012; p. 261.

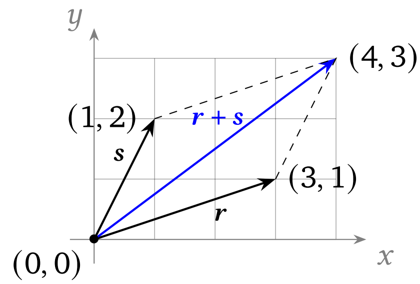

Figura 1.1: El paralelogramo de fuerzas en el plano $\mathbb{R}^2$

También es posible verificar que en el espacio tridimensional $\mathbb{R}^3$, un paralelogramo con origen $(0, 0, 0)$ y vértices (x, y, z) y (u, v, w), tiene como cuarto vértice – opuesto a $(0, 0, 0)$ – el punto

$$(x, y, z) + (u, v, w) := (x + u, y + v, z + w).$$

En resumen: en $\mathbb{R}^2$ o $\mathbb{R}^3$, se dispone de dos operaciones sobre vectores definidos "coordenada por coordenada": la ley de suma antedicha y la *dilatación* $c(x, y) := (cx, cy)$, respectivamente $c(x, y, z) := (cx, cy, cz)$, donde el factor de magnificación $c \in \mathbb{R}$ puede ser positiva o negativa (o cero). Por analogía,[2] estas operaciones son también aplicables a juegos ordenadas de n coordenadas reales, donde $n = 1, 2, 3, 4, \dots$, dando lugar a la siguiente definición.

Definición 1.1. Denótese por $\mathbb{R}$ los números reales. El plano $\mathbb{R}^2$ y el espacio $\mathbb{R}^3$ son los conjuntos de pares ordenados, respectivamente triples ordenadas de números reales:

$$\underline{\mathbb{R}^2} := \{\, (x_1, x_2) : x_1, x_2 \in \mathbb{R} \,\}, \qquad \underline{\mathbb{R}^3} := \{\, (x_1, x_2, x_3) : x_1, x_2, x_3 \in \mathbb{R} \,\}.$$

Más generalmente, si $n \in \mathbb{N}^*$, la totalidad de *n-tuplas ordenadas* es[3]

$$\underline{\mathbb{R}^n} := \{\, \boldsymbol{x} = (x_1, \dots, x_n) : x_1, \dots, x_n \in \mathbb{R} \,\}.$$

El elemento $\boldsymbol{x} \in \mathbb{R}^n$ es un **vector**; los números $x_1, x_2, \dots, x_n$ son las *coordenadas* del vector $\boldsymbol{x}$. Un número real $c \in \mathbb{R}$ se llamará un **escalar**.

[2]La idea de manipular n-tuplas de números de esta manera se remonta a Grassmann. Citando de nuevo a Ostermann y Wanner (p. 260): "Hermann Grassmann [...] published in 1844 his work *Die lineale Ausdehnunglehre* [La teoría de extensión lineal], an unreadable book, interspersed with mystic and abstract considerations."

[3]De ahora en adelante, los **números naturales** se denota por $\underline{\mathbb{N}} := \{0, 1, 2, 3, \dots\}$. Aquí se usa el "convenio francés": 0 se considera como número natural. Cuando se excluye 0, se denota los **enteros positivos** por $\underline{\mathbb{N}^*} := \{1, 2, 3, \dots\}$.

Se definen dos operaciones algebraicas fundamentales en $\mathbb{R}^n$: la *suma de dos vectores*, y la *multiplicación* de un vector por un escalar. Si $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ y si $c \in \mathbb{R}$, se definen

$$\begin{aligned} \underline{\boldsymbol{x} + \boldsymbol{y}} &= (x_1, x_2, \dots, x_n) + (y_1, y_2, \dots, y_n) := (x_1 + y_1, x_2 + y_2, \dots, x_n + y_n), \\ \underline{c\,\boldsymbol{x}} &= c(x_1, x_2, \dots, x_n) := (cx_1, cx_2, \dots, cx_n). \end{aligned} \tag{1.1}$$

El conjunto $\mathbb{R}^n$ dotado con estas dos operaciones es un *espacio vectorial*. ◊

Obsérvese que la suma de vectores cumple las propiedades usuales de una suma numérica. Es asociativa: $\boldsymbol{x} + (\boldsymbol{y} + \boldsymbol{z}) = (\boldsymbol{x} + \boldsymbol{y}) + \boldsymbol{z}$. Es conmutativa: $\boldsymbol{x} + \boldsymbol{y} = \boldsymbol{y} + \boldsymbol{x}$. Hay un vector "cero" $\mathbf{0}$ tal que $\boldsymbol{x} + \mathbf{0} = \mathbf{0} + \boldsymbol{x} = \boldsymbol{x}$. Y hay un "negativo" $-\boldsymbol{x}$ de cada vector $\boldsymbol{x}$ tal que $(-\boldsymbol{x}) + \boldsymbol{x} = \boldsymbol{x} + (-\boldsymbol{x}) = \mathbf{0}$. Por supuesto, el cero es $\mathbf{0} = (0, 0, \dots, 0)$, y además $-\boldsymbol{x} = (-x_1, -x_2, \dots, -x_n)$.

1.1. El producto escalar de dos vectores en $\mathbb{R}^n$

Sería natural ensayar una definición de "producto" de dos vectores en $\mathbb{R}^n$ por esta receta "coordenada por coordenada", al poner $\boldsymbol{x}\,\boldsymbol{y} := (x_1y_1, x_2y_2, \dots, x_ny_n)$. Una tal definición no resulta conveniente, porque el producto de dos vectores no ceros puede dar cero – por ejemplo, en $\mathbb{R}^2$, el producto de $(1, 0)$ y $(0, 1)$ sería $(0, 0)$. Resulta más apropiado definir un producto de dos vectores con un resultado *escalar*.

Definición 1.2. Si $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, su **producto escalar** es el número $\underline{\boldsymbol{x} \cdot \boldsymbol{y}} \in \mathbb{R}$ definido por

$$\boxed{\boldsymbol{x} \cdot \boldsymbol{y} := x_1y_1 + x_2y_2 + \cdots + x_ny_n.} \tag{1.2}$$

Como sinónimos de "producto escalar", se dice *producto interno* o *producto punto*. ◊

Proposición 1.3. *El producto escalar tiene las siguientes propiedades. Si $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \mathbb{R}^n$ y si $c \in \mathbb{R}$, entonces*

(a) $\boldsymbol{x} \cdot \boldsymbol{y} = \boldsymbol{y} \cdot \boldsymbol{x}$,

(b) $\boldsymbol{x} \cdot (\boldsymbol{y} + \boldsymbol{z}) = \boldsymbol{x} \cdot \boldsymbol{y} + \boldsymbol{x} \cdot \boldsymbol{z}$,

(c) $(\boldsymbol{x} + \boldsymbol{z}) \cdot \boldsymbol{y} = \boldsymbol{x} \cdot \boldsymbol{y} + \boldsymbol{z} \cdot \boldsymbol{y}$,

(d) $\boldsymbol{x} \cdot (c\boldsymbol{y}) = (c\boldsymbol{x}) \cdot \boldsymbol{y} = c(\boldsymbol{x} \cdot \boldsymbol{y})$, *y además:*

(e) $\boldsymbol{x} \cdot \boldsymbol{x} \geqslant 0$, *con igualdad solo si $\boldsymbol{x} = \mathbf{0}$ en $\mathbb{R}^n$.*

Demostración. Todos son cálculos sencillos usando la fórmula (1.2). Nótese en particular que

$$\boldsymbol{x} \cdot \boldsymbol{x} = x_1^2 + x_2^2 + \cdots + x_n^2 \geqslant 0. \qquad \square$$

Definición 1.4. Si $\boldsymbol{x} \in \mathbb{R}^n$, su **norma** (o *longitud*) es el número no negativo $\|\boldsymbol{x}\|$ dado por

$$\boxed{\|\boldsymbol{x}\| := \sqrt{\boldsymbol{x} \cdot \boldsymbol{x}}} = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}. \tag{1.3}$$

En $\mathbb{R}^2$ o $\mathbb{R}^3$, la fórmula de Pitágoras muestra que esta es la longitud del segmento de recta cuyos extremos son los puntos $\mathbf{0}$ y $\boldsymbol{x}$. ◊

Proposición 1.5. *Si $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, se verifica la* ***desigualdad de Cauchy****:*

$$\boxed{|\boldsymbol{x} \cdot \boldsymbol{y}| \leqslant \|\boldsymbol{x}\| \, \|\boldsymbol{y}\|} \tag{1.4}$$

con igualdad solo si $\boldsymbol{x}$, $\boldsymbol{y}$ son proporcionales.[4]

Demostración. Está claro que (1.4) se cumple con igualdad si $\boldsymbol{x} = \mathbf{0}$ o si $\boldsymbol{y} = \mathbf{0}$ (el vector $\mathbf{0}$ es proporcional a cada vector $\boldsymbol{x}$ porque $\mathbf{0} = 0\boldsymbol{x}$). Supóngase entonces que $\boldsymbol{x} \neq \mathbf{0}$, $\boldsymbol{y} \neq \mathbf{0}$.

Para $t \in \mathbb{R}$, considérese la función $f(t) := \|\boldsymbol{x} + t\boldsymbol{y}\|^2$. Resulta que

$$\begin{aligned} f(t) &= (\boldsymbol{x} + t\boldsymbol{y}) \cdot (\boldsymbol{x} + t\boldsymbol{y}) = (\boldsymbol{x} \cdot \boldsymbol{x}) + 2t(\boldsymbol{x} \cdot \boldsymbol{y}) + t^2(\boldsymbol{y} \cdot \boldsymbol{y}) \\ &= \|\boldsymbol{y}\|^2 t^2 + 2(\boldsymbol{x} \cdot \boldsymbol{y})t + \|\boldsymbol{x}\|^2 \\ &\equiv at^2 + bt + c, \end{aligned}$$

una función cuadrática de t, con $a > 0$. De la Proposición 1.3(e), el discriminante de la *ecuación* cuadrática $at^2 + bt + c = 0$ no puede ser positivo: si t_1, t_2 fueran dos raíces distintas de esta ecuación, sería $f(t) < 0$ para $t_1 < t < t_2$; pero $f(t) = \|\boldsymbol{x} + t\boldsymbol{y}\|^2 \geqslant 0$ para todo t. Se concluye que $\underline{b^2 - 4ac \leqslant 0}$, esto es,

$$4(\boldsymbol{x} \cdot \boldsymbol{y})^2 - 4\|\boldsymbol{x}\|^2\|\boldsymbol{y}\|^2 \leqslant 0,$$

o bien

$$(\boldsymbol{x} \cdot \boldsymbol{y})^2 \leqslant \|\boldsymbol{x}\|^2\|\boldsymbol{y}\|^2.$$

Al tomar raíces cuadradas en ambos lados, se obtiene la desigualdad deseada:

$$|\boldsymbol{x} \cdot \boldsymbol{y}| \leqslant \|\boldsymbol{x}\| \, \|\boldsymbol{y}\|.$$

Hay igualdad en estas relaciones solo si $b^2 - 4ac = 0$, es decir, solo si la ecuación cuadrática $at^2+bt+c = 0$ posee una raíz real $t = t_0$. Pero entonces $f(t_0) = \|\boldsymbol{x}+t_0\boldsymbol{y}\|^2 = 0$, y por ende $\boldsymbol{x} + t_0\boldsymbol{y} = \mathbf{0}$ en $\mathbb{R}^n$. Luego $\boldsymbol{x} = -t_0\boldsymbol{y}$ para algún $t_0 \in \mathbb{R}$, lo cual dice que $\boldsymbol{x}$, $\boldsymbol{y}$ son proporcionales. □

[4] Al lado izquierdo de la fórmula (1.4), las barras verticales denotan el *valor absoluto* del escalar $\boldsymbol{x} \cdot \boldsymbol{y}$.

En consecuencia, para $\boldsymbol{x} \neq \mathbf{0}$, $\boldsymbol{y} \neq \mathbf{0}$ se cumple

$$-1 \leqslant \frac{\boldsymbol{x} \cdot \boldsymbol{y}}{\|\boldsymbol{x}\|\,\|\boldsymbol{y}\|} \leqslant 1,$$

así que hay un único *ángulo* θ con $0 \leqslant \theta \leqslant \pi$ tal que

$$\boxed{\frac{\boldsymbol{x} \cdot \boldsymbol{y}}{\|\boldsymbol{x}\|\,\|\boldsymbol{y}\|} = \cos\theta.} \tag{1.5}$$

Se dice que θ es *el ángulo entre los vectores* no ceros $\boldsymbol{x}, \boldsymbol{y}$.

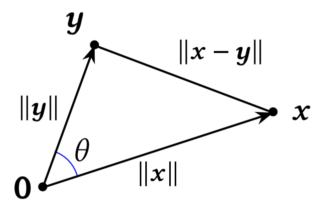

Figura 1.2: La ley de cosenos en un triángulo

Se mide la **distancia** entre dos vectores $\boldsymbol{x}$, $\boldsymbol{y}$ por la cantidad $\|\boldsymbol{x} - \boldsymbol{y}\|$. Al usar la fórmula (1.3), esto corresponde con la fórmula pitagórica para la distancia entre dos puntos con coordenadas $(x_1, x_2, \dots, x_n)$ y $(y_1, y_2, \dots, y_n)$ cuando $n = 2$ o 3.

De las Definiciones 1.1 y 1.2, se puede comprobar que

$$\|\boldsymbol{x} - \boldsymbol{y}\|^2 = \|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2 - 2\|\boldsymbol{x}\|\,\|\boldsymbol{y}\| \cos\theta.$$

Ahora $\|\boldsymbol{x}\|$, $\|\boldsymbol{y}\|$ y $\|\boldsymbol{x} - \boldsymbol{y}\|$ son las longitudes de los lados del triángulo con vértices $\mathbf{0}, \boldsymbol{x}, \boldsymbol{y}$ (Figura 1.2). Esta última fórmula es entonces la *ley de cosenos* que relaciona las longitudes de los lados y un ángulo en un triángulo Esto confirma que θ es el ángulo entre los dos segmentos $\overline{\mathbf{0}\boldsymbol{x}}$ y $\overline{\mathbf{0}\boldsymbol{y}}$.[5]

Proposición 1.6. *La norma tiene las siguientes propiedades. Si $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ y si $c \in \mathbb{R}$, entonces*

(a) $\|c\boldsymbol{x}\| = |c|\,\|\boldsymbol{x}\|$,

(b) $\|\boldsymbol{x} + \boldsymbol{y}\| \leqslant \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$, *y además*

(c) $\|\boldsymbol{x}\| \geqslant 0$, *con igualdad solo si $\boldsymbol{x} = \mathbf{0}$ en $\mathbb{R}^n$.*

[5]Si $\boldsymbol{a}, \boldsymbol{b}$ son dos puntos de $\mathbb{R}^n$, la notación $\overline{\boldsymbol{a}\,\boldsymbol{b}}$ indica el *segmento de recta* con extremos $\boldsymbol{a}$ y $\boldsymbol{b}$.

Demostración. La propiedad (a) sigue directamente de (1.3) porque $\sqrt{c^2} = |c|$. La propiedad (c) es una reinterpretación de la Proposición 1.3(e).

Para la desigualdad (b), fíjese que

$$\begin{aligned}\|\boldsymbol{x}+\boldsymbol{y}\|^2 &= (\boldsymbol{x}+\boldsymbol{y})\cdot(\boldsymbol{x}+\boldsymbol{y}) = \boldsymbol{x}\cdot\boldsymbol{x} + 2\,\boldsymbol{x}\cdot\boldsymbol{y} + \boldsymbol{y}\cdot\boldsymbol{y} \\ &= \|\boldsymbol{x}\|^2 + 2\,\boldsymbol{x}\cdot\boldsymbol{y} + \|\boldsymbol{y}\|^2 \\ &\leqslant \|\boldsymbol{x}\|^2 + 2\|\boldsymbol{x}\|\,\|\boldsymbol{y}\| + \|\boldsymbol{y}\|^2 \quad \text{por la desigualdad de Cauchy} \\ &= \big(\|\boldsymbol{x}\| + \|\boldsymbol{y}\|\big)^2. \end{aligned}$$ □

Definición 1.7. Dos vectores $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ son **ortogonales** (o *perpendiculares*) si $\boldsymbol{x}\cdot\boldsymbol{y} = 0$. ◊

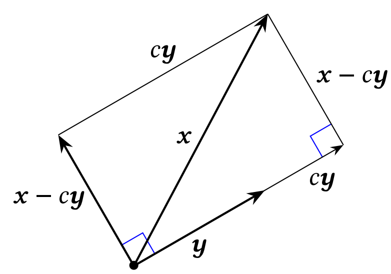

Figura 1.3: La proyección ortogonal del vector $\boldsymbol{x}$ a lo largo de $\boldsymbol{y}$

Si $\boldsymbol{x}$, $\boldsymbol{y}$ son dos vectores con $\boldsymbol{y} \neq \boldsymbol{0}$, se puede escribir $\boldsymbol{x}$ como la suma de un vector $c\boldsymbol{y}$ proporcional a $\boldsymbol{y}$ más un vector ortogonal a $\boldsymbol{y}$. Para determinar el valor de c, se resuelve la ecuación:

$$(\boldsymbol{x} - c\boldsymbol{y})\cdot\boldsymbol{y} = 0, \quad \text{de donde} \quad c = \frac{\boldsymbol{x}\cdot\boldsymbol{y}}{\boldsymbol{y}\cdot\boldsymbol{y}}.$$

El vector $c\boldsymbol{y}$, escrito como

$$\operatorname{proy}_{\boldsymbol{y}}(\boldsymbol{x}) := \frac{\boldsymbol{x}\cdot\boldsymbol{y}}{\boldsymbol{y}\cdot\boldsymbol{y}}\,\boldsymbol{y} = \left(\frac{\boldsymbol{x}\cdot\boldsymbol{y}}{\|\boldsymbol{y}\|}\right)\frac{\boldsymbol{y}}{\|\boldsymbol{y}\|}$$

se llama la **proyección** de $\boldsymbol{x}$ a lo largo de $\boldsymbol{y}$ (véase la Figura 1.3).

Considérese nuevamente el paralelogramo en $\mathbb{R}^n$ determinado por los tres vértices consecutivos $\boldsymbol{x}$, $\boldsymbol{0}$, $\boldsymbol{y}$: los segmentos $\overline{\boldsymbol{0}\boldsymbol{x}}$ y $\overline{\boldsymbol{0}\boldsymbol{y}}$ son lados adyacentes del paralelogramo y el cuarto vértice es la suma $\boldsymbol{x}+\boldsymbol{y}$. Cuando $\boldsymbol{x}$, $\boldsymbol{y}$ son ortogonales, este paralelogramo es un *rectángulo*.

1.2. El producto vectorial en $\mathbb{R}^3$

Entre todos los espacios vectoriales $\mathbb{R}^n$, el caso $n = 3$ se distingue por la posibilidad de definir un segundo producto de vectores en $\mathbb{R}^3$, cuyo resultado es un *vector*, no un escalar. Para manejarlo, resulta cómodo usar las notaciones

$$\begin{aligned} \boldsymbol{x} &= (x_1, x_2, x_3) \equiv (a, b, c), \\ \boldsymbol{y} &= (y_1, y_2, y_3) \equiv (p, q, r), \\ \boldsymbol{z} &= (z_1, z_2, z_3) \equiv (u, v, w). \end{aligned}$$

Definición 1.8. El **producto vectorial** de $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^3$ es el vector $\boldsymbol{x} \times \boldsymbol{y} \in \mathbb{R}^3$ definido por

$$\begin{aligned} &\boxed{\boldsymbol{x} \times \boldsymbol{y} := (x_2y_3 - x_3y_2, x_3y_1 - x_1y_3, x_1y_2 - x_2y_1)} \\ &= (br - cq, cp - ar, aq - bp). \end{aligned} \tag{1.6}$$

El producto vectorial también se llama el *producto cruz* en $\mathbb{R}^3$. ◊

Lema 1.9. *El producto vectorial tiene las siguientes propiedades, para $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \mathbb{R}^3$, $t \in \mathbb{R}$:*

(a) $\boldsymbol{x} \times \boldsymbol{y} = -\boldsymbol{y} \times \boldsymbol{x}$,

(b) $\boldsymbol{x} \times (\boldsymbol{y} + \boldsymbol{z}) = \boldsymbol{x} \times \boldsymbol{y} + \boldsymbol{x} \times \boldsymbol{z}$; $\quad (\boldsymbol{x} + \boldsymbol{z}) \times \boldsymbol{y} = \boldsymbol{x} \times \boldsymbol{y} + \boldsymbol{z} \times \boldsymbol{y}$,

(c) $\boldsymbol{x} \times (t\boldsymbol{y}) = (t\boldsymbol{x}) \times \boldsymbol{y} = t(\boldsymbol{x} \times \boldsymbol{y})$.

Demostración. El intercambio $\boldsymbol{x} \leftrightarrow \boldsymbol{y}$ produce un cambio de signo al lado derecho de (1.6), de donde la propiedad (a) es evidente. Esto significa que el producto cruz es **anticonmutativo**.

Se verifican las propiedades (b) y (c) por cálculos directos. □

Nótese que la anticonmutatividad conlleva la relación $\boldsymbol{x} \times \boldsymbol{x} = \mathbf{0}$, que también es evidente de (1.6).

Lema 1.10. *Si $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \mathbb{R}^3$, entonces*

$$\boldsymbol{x} \cdot (\boldsymbol{y} \times \boldsymbol{z}) = (\boldsymbol{x} \times \boldsymbol{y}) \cdot \boldsymbol{z}.$$

En consecuencia, $\boldsymbol{x} \times \boldsymbol{y}$ es ortogonal a los dos vectores $\boldsymbol{x}$, $\boldsymbol{y}$.

Demostración. Es cuestión de notar que

$$\begin{aligned} (\boldsymbol{x} \times \boldsymbol{y}) \cdot \boldsymbol{z} = (\boldsymbol{x} \times \boldsymbol{y}) \cdot (u, v.w) &= (br - cq)u + (cp - ar)v + (aq - bp)w \\ &= aqw + bru + cpv - arv - bpw - cqu \\ &= a(qw - rv) + b(ru - pw) + c(pv - qu) \\ &= \boldsymbol{x} \cdot (\boldsymbol{y} \times \boldsymbol{z}). \end{aligned}$$

En consecuencia, se obtiene

$$\begin{aligned}(\boldsymbol{x}\times\boldsymbol{y})\cdot\boldsymbol{y} &= \boldsymbol{x}\cdot(\boldsymbol{y}\times\boldsymbol{y}) = \boldsymbol{x}\cdot\mathbf{0} = 0,\\ (\boldsymbol{x}\times\boldsymbol{y})\cdot\boldsymbol{x} &= -(\boldsymbol{y}\times\boldsymbol{x})\cdot\boldsymbol{x} = -\boldsymbol{y}\cdot(\boldsymbol{x}\times\boldsymbol{x}) = -\boldsymbol{y}\cdot\mathbf{0} = 0.\end{aligned}$$

Esto comprueba la ortogonalidad de $\boldsymbol{x}\times\boldsymbol{y}$ con $\boldsymbol{x}$ y con $\boldsymbol{y}$. ◻

Proposición 1.11. *El producto vectorial* no es asociativo*; se verifican las fórmulas*

$$\begin{aligned}\boldsymbol{x}\times(\boldsymbol{y}\times\boldsymbol{z}) &= (\boldsymbol{x}\cdot\boldsymbol{z})\boldsymbol{y} - (\boldsymbol{x}\cdot\boldsymbol{y})\boldsymbol{z},\\ (\boldsymbol{x}\times\boldsymbol{y})\times\boldsymbol{z} &= (\boldsymbol{x}\cdot\boldsymbol{z})\boldsymbol{y} - (\boldsymbol{y}\cdot\boldsymbol{z})\boldsymbol{x},\end{aligned}\tag{1.7}$$

cuyos lados derechos son distintos, en general.

Demostración. La primera fórmula en (1.7) se verifica por un cálculo directo. La segunda entonces sigue de inmediato:

$$(\boldsymbol{x}\times\boldsymbol{y})\times\boldsymbol{z} = -\boldsymbol{z}\times(\boldsymbol{x}\times\boldsymbol{y}) = -(\boldsymbol{z}\cdot\boldsymbol{y})\boldsymbol{x} + (\boldsymbol{z}\cdot\boldsymbol{x})\boldsymbol{y} = (\boldsymbol{x}\cdot\boldsymbol{z})\boldsymbol{y} - (\boldsymbol{y}\cdot\boldsymbol{z})\boldsymbol{x}.$$

Como $\boldsymbol{x}\times(\boldsymbol{y}\times\boldsymbol{z}) \neq (\boldsymbol{x}\times\boldsymbol{y})\times\boldsymbol{z}$, *no se debe escribir* "$\boldsymbol{x}\times\boldsymbol{y}\times\boldsymbol{z}$", ya que esa expresión resulta ambigua cuando $\boldsymbol{x}$ y $\boldsymbol{z}$ no son proporcionales. ◻

El **producto triple** de tres vectores en $\mathbb{R}^3$ se define como

$$\boxed{[\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}] := \boldsymbol{x}\cdot(\boldsymbol{y}\times\boldsymbol{z}) = (\boldsymbol{x}\times\boldsymbol{y})\cdot\boldsymbol{z},}\tag{1.8}$$

en vista del Lema 1.9. Explícitamente:

$$[\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}] = x_1y_2z_3 + x_2y_3z_1 + x_3y_1z_2 - x_1y_3z_2 - x_2y_1z_3 - x_3y_2z_1.$$

De esta fórmula se ve que $[\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}] = 0$ si dos de los tres vectores $\boldsymbol{x}$, $\boldsymbol{y}$, $\boldsymbol{z}$ coinciden.

También es posible comprobar la fórmula:

$$\|\boldsymbol{x}\times\boldsymbol{y}\|^2 = \|\boldsymbol{x}\|^2\|\boldsymbol{y}\|^2 - (\boldsymbol{x}\cdot\boldsymbol{y})^2.$$

Habida cuenta de (1.5), esto da

$$\|\boldsymbol{x}\times\boldsymbol{y}\|^2 = \|\boldsymbol{x}\|^2\|\boldsymbol{y}\|^2 - \|\boldsymbol{x}\|^2\|\boldsymbol{y}\|^2\cos^2\theta = \|\boldsymbol{x}\|^2\|\boldsymbol{y}\|^2\operatorname{sen}^2\theta,$$

de donde sigue

$$\boxed{\|\boldsymbol{x}\times\boldsymbol{y}\| = \|\boldsymbol{x}\|\,\|\boldsymbol{y}\|\,|\operatorname{sen}\theta|.}\tag{1.9}$$

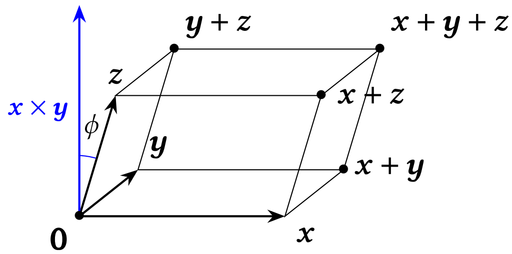

Figura 1.4: El paralelepípedo con lados adyacentes $\boldsymbol{x}$, $\boldsymbol{y}$, $\boldsymbol{z}$

Esta cantidad $\|\boldsymbol{x}\times\boldsymbol{y}\|$ es entonces *el área de un paralelogramo* cuyos lados adyacentes miden $\|\boldsymbol{x}\|$ y $\|\boldsymbol{y}\|$ e incluyen un ángulo θ; en particular, es el área del paralelogramo cuyos cuatro vértices son $\boldsymbol{x}$, $\mathbf{0}$, $\boldsymbol{y}$, $\boldsymbol{x}+\boldsymbol{y}$ (Figura 1.1).

⟦ Nótese que esta área es nula cuando los vectores $\boldsymbol{x}$, $\boldsymbol{y}$ son proporcionales, pues entonces el paralelogramo colapsa en un segmento de recta; ya se sabe que $\boldsymbol{x}\times(c\boldsymbol{x})=\mathbf{0}$. A la inversa, si $\boldsymbol{x}\times\boldsymbol{y}\neq\mathbf{0}$, o lo que es lo mismo, si $\|\boldsymbol{x}\times\boldsymbol{y}\|\neq 0$, la fórmula (1.9) indica que $\operatorname{sen}\theta\neq 0$, así que $\boldsymbol{x}$, $\boldsymbol{y}$ *no* son proporcionales. ⟧

El producto triple $[\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}]$ también tiene una interpretación geométrica. Considérese el *paralelepípedo* con vértice $\mathbf{0}$ cuyos tres vértices adyacentes son $\boldsymbol{x}$, $\boldsymbol{y}$, $\boldsymbol{z}$ (véase la Figura 1.4). Al tomar el paralelogramo con vértices $\boldsymbol{x}$, $\mathbf{0}$, $\boldsymbol{y}$, $\boldsymbol{x}+\boldsymbol{y}$ como una base de este paralelepípedo, la altura correspondiente es $\|\boldsymbol{z}\|\cos\phi$, donde ϕ es el ángulo agudo entre $\boldsymbol{z}$ y un vector perpendicular a la base. Un vector perpendicular a la base será $\pm(\boldsymbol{x}\times\boldsymbol{y})$; se escoge el signo para que ϕ sea agudo. Entonces la conocida fórmula "volumen = base por altura" dice que el volumen del paralelepípedo es[6]

$$\mathrm{Vol}=\|\boldsymbol{x}\times\boldsymbol{y}\|\,\|\boldsymbol{z}\|\cos\phi=\pm(\boldsymbol{x}\times\boldsymbol{y})\cdot\boldsymbol{z}=\pm[\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}].$$

1.3. Rectas y planos en $\mathbb{R}^2$ y $\mathbb{R}^3$

Para describir la geometría analítica de $\mathbb{R}^3$, es oportuno emplear las coordenadas $\boldsymbol{r}=(x,y,z)$ en lugar de $\boldsymbol{x}=(x_1,x_2,x_3)$. El conjunto $\mathbb{R}^3$ incluye una copia de $\mathbb{R}^2$ que corresponde a tomar $z=0$, en cuyo caso se escribe $\boldsymbol{r}=(x,y)$ simplemente.

► Conviene hacer un rápido recordatorio de las rectas en $\mathbb{R}^2$. Se sabe que una recta obedece una *ecuación de primer grado*:

$$ax+by+c=0,$$

donde $a,b,c\in\mathbb{R}$ son constantes sujetos a la condición con $a^2+b^2>0$; porque si fuera $a=b=0$, la ecuación trivial $c=0$ no tendría información acerca del punto (x,y).

[6]Nótese que el escalar $[\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}]$ puede ser positivo o negativo – o bien cero si $\boldsymbol{x}$, $\boldsymbol{y}$, $\boldsymbol{z}$ son coplanarios.

La recta que pasa por el punto específico (x_0, y_0) con pendiente m tiene la ecuación

$$y - y_0 = m(x - x_0).$$

La **recta que pasa por dos puntos** distintos (x_1, y_1) y (x_2, y_2) de $\mathbb{R}^2$ tiene la ecuación:

$$\boxed{(y_2 - y_1)(x - x_1) - (x_2 - x_1)(y - y_1) = 0.} \tag{1.10}$$

que (a) es de primer grado; y (b) pasa por los dos puntos indicados.

La ecuación se podría escribir alternativamente como una igualdad entre dos fracciones, que a su vez determinan una nueva variable auxiliar o **parámetro** t, al poner[7]

$$\frac{y - y_1}{y_2 - y_1} = \frac{x - x_1}{x_2 - x_1} =: t.$$

Al escribir $a := x_2 - x_1$ y $b := y_2 - y_1$, la sola ecuación (1.10) se convierte en *dos ecuaciones con un parámetro libre t*, así:

$$\begin{aligned} x &= x_1 + at, \\ y &= y_1 + bt, \end{aligned}$$

que *representa la misma recta en forma paramétrica*. Para volver al formato original (1.10), se debe *eliminar* el parámetro t entre estos dos ecuaciones.

► En $\mathbb{R}^3$, una recta queda determinado por su *dirección* y por la especificación de uno de sus puntos. Sea $\boldsymbol{r}_0 = (x_0, y_0, z_0)$ un punto conocido en la recta. La dirección de la recta está determinada por un vector $\boldsymbol{v} = (l, m, n)$ tal que el segmento $\overline{\boldsymbol{0v}}$ es paralelo a la recta en cuestión. Esto significa que el vector $\boldsymbol{r} - \boldsymbol{r}_0$ es *proporcional* a $\boldsymbol{v}$; en otras palabras,

$$\boldsymbol{r} - \boldsymbol{r}_0 = t\boldsymbol{v}, \tag{1.11}$$

donde la constante de proporcionalidad $t \in \mathbb{R}$ es un parámetro libre: los diversos valores de t corresponden a distintos puntos $\boldsymbol{r}$ de la recta. Luego (1.11) es la ecuación deseada de la recta *en una forma paramétrica*. En coordenadas cartesianas, se obtiene

$$\begin{aligned} x &= x_0 + lt, \\ y &= y_0 + mt, \\ z &= z_0 + nt. \end{aligned} \tag{1.12}$$

[7]Este formato de la ecuación de recta, en contraste con la fórmula (1.10), parece admitir dos excepciones: si $y_2 = y_1$ o si $x_2 = x_1$, una de la fracciones tendrá denominador 0. En tales casos se decide que el *numerador* también debe ser 0; por ejemplo, si fuera $y_2 = y_1$ la ecuación de la recta sería $y = y_1$, sin necesidad de mencionar la variable x.

Al eliminar el parámetro t, estas tres ecuaciones se reducen a dos:

$$\frac{x - x_0}{l} = \frac{y - y_0}{m} = \frac{z - z_0}{n}. \tag{1.13}$$

⟦ Si uno de los denominadores es cero, por ejemplo si $l = 0$, $m \neq 0$, $n \neq 0$, el par de ecuaciones debe interpretarse como $x = x_0$, $(y - y_0)/m = (z - z_0)/n$. ⟧

Una recta también queda determinado por dos puntos, digamos $\boldsymbol{r}_0$ y $\boldsymbol{r}_1$. En este caso basta tomar $\boldsymbol{v} = \boldsymbol{r}_1 - \boldsymbol{r}_0$, en cuyo caso (1.11) se convierte en:

$$\boxed{\boldsymbol{r} = (1 - t)\boldsymbol{r}_0 + t\boldsymbol{r}_1.} \tag{1.14}$$

Obsérvese que $t = 0$ corresponde a $\boldsymbol{r} = \boldsymbol{r}_0$ y que $t = 1$ corresponde a $\boldsymbol{r} = \boldsymbol{r}_1$.

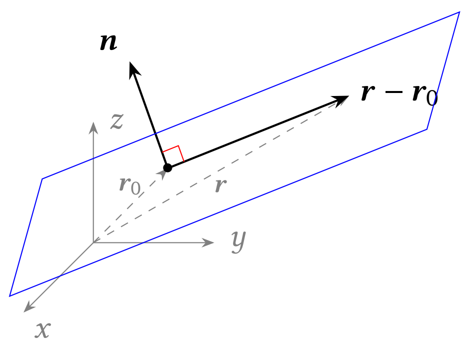

Figura 1.5: El plano que pasa por $\boldsymbol{r}_0$ con vector normal $\boldsymbol{n}$

► Un **plano** en $\mathbb{R}^3$ queda determinado por su *inclinación* y por uno de sus puntos $\boldsymbol{r}_0$; dado el punto $\boldsymbol{r}_0$, se selecciona el plano deseado de entre una familia de planos paralelos con la misma inclinación. Entonces, sea $\boldsymbol{r}_0$ un punto en el plano. Se determina la inclinación al especificar un *vector normal* $\boldsymbol{n} = (a, b, c)$ tal que el segmento $\overline{\boldsymbol{0}\boldsymbol{n}}$ sea perpendicular a ese plano. Esto significa que $\boldsymbol{n}$ es ortogonal al vector $\boldsymbol{r} - \boldsymbol{r}_0$,

$$\boldsymbol{n} \cdot (\boldsymbol{r} - \boldsymbol{r}_0) = 0. \tag{1.15}$$

Esta es la ecuación deseada del plano (véase la Figura 1.5). En coordenadas, se obtiene

$$a(x - x_0) + b(y - y_0) + c(z - z_0) = 0,$$

o bien:

$$\boxed{ax + by + cz + d = 0,} \tag{1.16}$$

donde $d = -(ax_0 + by_0 + cz_0)$ es una constante.

Un plano también *queda determinado por tres de sus puntos* $\boldsymbol{r}_0$, $\boldsymbol{r}_1$, $\boldsymbol{r}_2$. Las diferencias $\boldsymbol{r}_1 - \boldsymbol{r}_0$ y $\boldsymbol{r}_2 - \boldsymbol{r}_0$ son vectores paralelos al plano; su producto vectorial será un vector perpendicular a ambos y por ende perpendicular al plano. Luego podemos tomar como vector normal $\boldsymbol{n} = (\boldsymbol{r}_1 - \boldsymbol{r}_0) \times (\boldsymbol{r}_2 - \boldsymbol{r}_0)$, en cuyo caso la ecuación (1.15) se convierte en

$$\big[\boldsymbol{r} - \boldsymbol{r}_0, \boldsymbol{r}_1 - \boldsymbol{r}_0, \boldsymbol{r}_2 - \boldsymbol{r}_0\big] = 0.$$

Una observación geométrica que salta a la vista en $\mathbb{R}^3$ es que *una recta es la intersección de dos planos* (no paralelos). Esto es: si dos planos tienen al menos un punto en común, tiene muchos puntos en común y el conjunto de intersección es efectivamente una *recta*.

Es fácil entender esa observación si se observe que un plano en $\mathbb{R}^3$ *obedece una ecuación de primer grado* (1.16). En cambio, una *recta* requiere ***dos*** *ecuaciones de primer grado*, véase la fórmula (1.13) de nuevo. Nótese que la recta (1.13) es la intersección de estos dos planos, por ejemplo:[8]

$$m(x - x_0) = l(y - y_0) \quad \text{y} \quad n(y - y_0) = m(z - z_0).$$

► En $\mathbb{R}^n$, donde n no es necesariamente 3, se pueden repetir muchos de estos argumentos (excepto los que involucran el producto vectorial). Por ejemplo, (1.11) o bien (1.14) representa una recta en $\mathbb{R}^n$; el análogo de (1.13) es una colección de $(n - 1)$ ecuaciones. La ecuación (1.15) ahora representa un *hiperplano* en $\mathbb{R}^n$ perpendicular al vector $\boldsymbol{n}$.

En $\mathbb{R}^4$, una ecuación de primer grado $ax + by + cz + dw + e = 0$ representa un hiperplano. Un *plano* requiere dos ecuaciones de primer grado; y una *recta* requiere tres. En $\mathbb{R}^4$, no es cierto que la intersección de dos planos es una recta (o vacía); podría ser un solo punto. Por ejemplo, las dos pares de ecuaciones $x = y = 0$ y $z = w = 0$ corresponden a dos planos en $\mathbb{R}^4$ cuya intersección es solamente el origen $(x, y, z, w) = (0, 0, 0, 0)$.

► La fórmula (1.14) tiene una importante extensión, válida en $\mathbb{R}^n$ para cualquier n. En primer lugar, es posible escribirla con *dos parámetros ligados* s, t en vez de un solo parámetro libre t, así:

$$\boldsymbol{r} = s\,\boldsymbol{r}_0 + t\,\boldsymbol{r}_1 \quad \text{con} \quad s + t = 1. \tag{1.17a}$$

De igual modo, el plano en $\mathbb{R}^3$ que pasa por los tres puntos (no colineales) $\boldsymbol{r}_0$, $\boldsymbol{r}_1$, $\boldsymbol{r}_2$ tiene una descripción con tres parámetros ligados:

$$\boldsymbol{r} = s\,\boldsymbol{r}_0 + t\,\boldsymbol{r}_1 + u\,\boldsymbol{r}_2 \quad \text{con} \quad s + t + u = 1. \tag{1.17b}$$

Es evidente que el conjunto de los $\boldsymbol{r}$ que cumplen (1.17b) contiene $\boldsymbol{r}_0$, $\boldsymbol{r}_1$, $\boldsymbol{r}_2$ al escoger los parámetros $(s, t, u) = (1, 0, 0)$, $(0, 1, 0)$, $(0, 0, 1)$ respectivamente. Por otro lado, esta

[8] Estos dos planos no son los únicos que pasan por la recta dada. En efecto, $n(x - x_0) = l(z - z_0)$ es un tercer plano que incluye la recta (1.13).

fórmula puede ser considerado como un sistema de ecuaciones de primer grado para las variables s, t, u:

$$\begin{aligned} x_0\, s + x_1\, t + x_2\, u &= x, \\ y_0\, s + y_1\, t + y_2\, u &= y, \\ z_0\, s + z_1\, t + z_2\, u &= z, \\ s + t + u &= 1. \end{aligned} \tag{1.18}$$

Al haber cuatro ecuaciones para las tres variables s, t, u, solo hay una solución cuando las cantidades $x, y, z, 1$ al lado derecho obedecen cierta restricción. Como se verá más adelante, esta restricción tiene la forma $ax + by + cz + d = 0$, una ecuación de primer grado, de modo que la versión parametrizada (1.17b) representa un *plano* (1.16).

Es fácil generalizar las fórmulas (1.17) a vectores en $\mathbb{R}^n$, como sigue.

Definición 1.12. Sea $\{\boldsymbol{r}_0, \boldsymbol{r}_1, \dots, \boldsymbol{r}_k\}$ un juego de $(k + 1)$ vectores en $\mathbb{R}^n$. La expresión

$$\boldsymbol{r} = t_0\, \boldsymbol{r}_0 + t_1\, \boldsymbol{r}_1 + \cdots + t_k\, \boldsymbol{r}_k \quad \text{con} \quad t_0 + t_1 + \cdots + t_k = 1 \tag{1.19}$$

se llama una **combinación afín** de los vectores dados.

Los parámetros $t_0, t_1, \dots, t_k \in \mathbb{R}$ de esta combinación pueden ser positivos, negativos o ceros (toda vez que su suma sea 1).

Si además se pide *no negatividad*: $t_0 \geqslant 0$, $t_1 \geqslant 0, \dots, t_k \geqslant 0$, se dice que (1.19) es una **combinación convexa** de los vectores dados. ◊

1.4. Ejercicios sobre vectores en $\mathbb{R}^n$

Ejercicio 1.1. Sean $\boldsymbol{x} = (3, 4)$, $\boldsymbol{y} = (1, t)$ dos vectores en $\mathbb{R}^2$. Determinar los valores de t para que:

(a) $\boldsymbol{x}$, $\boldsymbol{y}$ sean ortogonales;

(b) $\boldsymbol{x}$, $\boldsymbol{y}$ sean paralelos (es decir, proporcionales);

(c) el ángulo entre $\boldsymbol{x}$, $\boldsymbol{y}$ sea $\pi/4$;

(d) el ángulo entre $\boldsymbol{x}$, $\boldsymbol{y}$ sea $\pi/3$.

Ejercicio 1.2. Calcular los vectores $\boldsymbol{x} + \boldsymbol{y}$, $\boldsymbol{x} - \boldsymbol{y}$, $\boldsymbol{x} \times \boldsymbol{y}$; los escalares $\boldsymbol{x} \cdot \boldsymbol{y}$, $\|\boldsymbol{x}\|$, $\|\boldsymbol{y}\|$ y la proyección $\operatorname{proy}_{\boldsymbol{y}}(\boldsymbol{x}) := \frac{\boldsymbol{x}\cdot\boldsymbol{y}}{\boldsymbol{y}\cdot\boldsymbol{y}}\, \boldsymbol{y}$ de $\boldsymbol{x}$ a lo largo de $\boldsymbol{y}$, en los siguientes casos:

$$\begin{aligned} &\text{(a)} \quad \boldsymbol{x} = (2, -1, 5), \quad \boldsymbol{y} = (-1, 1, 1); \\ &\text{(b)} \quad \boldsymbol{x} = (6, 3, -1), \quad \boldsymbol{y} = (12, -3, 7). \end{aligned}$$

Ejercicio 1.3. Si $\triangle ABC$ es un triángulo en $\mathbb{R}^3$, se denotan sus *vértices* por los tres vectores $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$; sus *lados* son los segmentos $\overline{\boldsymbol{ab}}, \overline{\boldsymbol{bc}}, \overline{\boldsymbol{ca}}$, representados por los tres vectores $\boldsymbol{b}-\boldsymbol{a}$, $\boldsymbol{c}-\boldsymbol{b}$, $\boldsymbol{a}-\boldsymbol{c}$, respectivamente.

Determinar las longitudes de los lados del triángulo cuyos vértices son $\boldsymbol{a} = (3,-1,-1)$, $\boldsymbol{b} = (1,2,1)$, $\boldsymbol{c} = (6,-1,2)$ y comprobar que ésta es un triángulo rectángulo.

⟦ Indicación: calcular las normas de los vectores que representan los lados. ⟧

Ejercicio 1.4. Verificar que el vector $\frac{1}{2}(\boldsymbol{a}+\boldsymbol{b})$ es el punto medio del segmento $\overline{\boldsymbol{ab}}$. Si $\boldsymbol{a}$, $\boldsymbol{b}$ y $\boldsymbol{c}$ son vértices de un triángulo, sean $\boldsymbol{l}$ y $\boldsymbol{m}$ los puntos medios de lados $\overline{\boldsymbol{ca}}$ y $\overline{\boldsymbol{ab}}$ respectivamente; comprobar que el segmento $\overline{\boldsymbol{lm}}$ es paralelo[9] al lado $\overline{\boldsymbol{bc}}$ y que mide la mitad de la longitud del lado $\overline{\boldsymbol{bc}}$.

Ejercicio 1.5. Un polígono $ABCD$ en el espacio $\mathbb{R}^3$ tiene vértices consecutivos $\boldsymbol{a} = (0,0,0)$, $\boldsymbol{b} = (1,0,0)$, $\boldsymbol{c} = (\frac{1}{2}, \frac{1}{2}, \frac{\sqrt{2}}{2})$, $\boldsymbol{d} = (0,1,0)$. Comprobar que los cuatro lados tienen igual longitud, pero que $ABCD$ no es un rombo.

⟦ Indicación: hacer un dibujo de este polígono en $\mathbb{R}^3$. ⟧

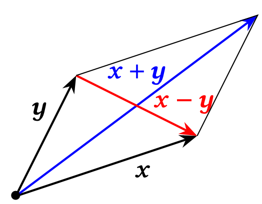

Figura 1.6: Ley del paralelogramo

Ejercicio 1.6. Demostrar la *ley del paralelogramo* para vectores en $\mathbb{R}^n$:

$$\|\boldsymbol{x}+\boldsymbol{y}\|^2 + \|\boldsymbol{x}-\boldsymbol{y}\|^2 = 2\|\boldsymbol{x}\|^2 + 2\|\boldsymbol{y}\|^2.$$

Explicar el significado geométrico de esta identidad en $\mathbb{R}^2$.

⟦ Indicación: considerar el cuadrilátero con vértices $\boldsymbol{0}$, $\boldsymbol{x}$, $\boldsymbol{x}+\boldsymbol{y}$, $\boldsymbol{y}$: Figura 1.6. ⟧

Ejercicio 1.7. Comprobar la siguiente relación entre tres vectores $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \mathbb{R}^3$:

$$\|\boldsymbol{x}+\boldsymbol{y}+\boldsymbol{z}\|^2 + \|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2 + \|\boldsymbol{z}\|^2 = \|\boldsymbol{x}+\boldsymbol{y}\|^2 + \|\boldsymbol{y}+\boldsymbol{z}\|^2 + \|\boldsymbol{z}+\boldsymbol{x}\|^2.$$

⟦ Esta identidad está satisfecha por los lados y ciertos diagonales del paralelepípedo con vértices $\boldsymbol{0}, \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$. Resulta posible deducirla de la ley del paralelogramo en el Ejercicio 1.6. ⟧

Ejercicio 1.8. Sean $\boldsymbol{x}, \boldsymbol{y}$ dos vectores en $\mathbb{R}^n$. Demostrar que los vectores $\boldsymbol{a} := \|\boldsymbol{y}\|\,\boldsymbol{x} + \|\boldsymbol{x}\|\,\boldsymbol{y}$ y $\boldsymbol{b} := \|\boldsymbol{y}\|\,\boldsymbol{x} - \|\boldsymbol{x}\|\,\boldsymbol{y}$ son ortogonales.

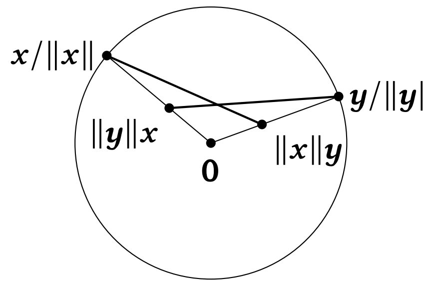

Figura 1.7: La fórmula de simetría

Ejercicio 1.9. Si los vectores $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ tienen normas $\|\boldsymbol{x}\| > 0$, $\|\boldsymbol{y}\| > 0$, comprobar esta *fórmula de simetría*:

$$\left\| \frac{\boldsymbol{y}}{\|\boldsymbol{y}\|} - \|\boldsymbol{y}\|\boldsymbol{x} \right\| = \left\| \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|} - \|\boldsymbol{x}\|\boldsymbol{y} \right\|.$$

⟦ Indicación: la Figura 1.7 ilustra esta fórmula en el círculo unitario de $\mathbb{R}^2$. ⟧

Ejercicio 1.10. Verificar las siguientes fórmulas para el producto vectorial y el producto triple en $\mathbb{R}^3$:

(I) $\boldsymbol{x} \times (\boldsymbol{y} \times \boldsymbol{z}) + \boldsymbol{y} \times (\boldsymbol{z} \times \boldsymbol{x}) + \boldsymbol{z} \times (\boldsymbol{x} \times \boldsymbol{y}) = 0$;

(II) $\|\boldsymbol{x} \times \boldsymbol{y}\|^2 = \|\boldsymbol{x}\|^2 \|\boldsymbol{y}\|^2 - (\boldsymbol{x} \cdot \boldsymbol{y})^2$;

(III) $(\boldsymbol{x} \times \boldsymbol{y}) \cdot (\boldsymbol{z} \times \boldsymbol{w}) = (\boldsymbol{x} \cdot \boldsymbol{z})(\boldsymbol{y} \cdot \boldsymbol{w}) - (\boldsymbol{x} \cdot \boldsymbol{w})(\boldsymbol{y} \cdot \boldsymbol{z})$.

Ejercicio 1.11. Demostrar que el área del triángulo $\triangle ABC$ del Ejercicio 1.3 está dada por $\frac{1}{2}\|(\boldsymbol{b} - \boldsymbol{a}) \times (\boldsymbol{c} - \boldsymbol{a})\|$.
⟦ Indicación: una diagonal divide un paralelogramo en dos triángulos de igual área. ⟧
Después, calcular el área del triángulo con vértices $(-2, 3, 1)$, $(4, 2, -2)$, $(2, 0, 1)$.

Ejercicio 1.12. Si $\boldsymbol{r}_0$, $\boldsymbol{r}_1$ son vectores en $\mathbb{R}^n$, el *segmento* $\overline{\boldsymbol{r}_0\boldsymbol{r}_1}$ consiste de los puntos $\boldsymbol{r}$ de la recta (1.14) que pasa por $\boldsymbol{r}_0$ y $\boldsymbol{r}_1$ tales que

$$\|\boldsymbol{r}_1 - \boldsymbol{r}\| + \|\boldsymbol{r} - \boldsymbol{r}_0\| = \|\boldsymbol{r}_1 - \boldsymbol{r}_0\|.$$

Comprobar que $\overline{\boldsymbol{r}_0\boldsymbol{r}_1} = \{\, (1 - t)\boldsymbol{r}_0 + t\boldsymbol{r}_1 : \underline{0 \leqslant t \leqslant 1} \,\}$.

Ejercicio 1.13. Sean $\boldsymbol{r} = \boldsymbol{r}_0 + t\boldsymbol{a}$ $(t \in \mathbb{R})$; y $\boldsymbol{r} = \boldsymbol{r}_1 + s\boldsymbol{b}$ $(s \in \mathbb{R})$ dos rectas en $\mathbb{R}^3$ dadas en forma paramétrica. Demostrar que estas rectas se cortan (tienen un punto de intersección) si y solo si $[\boldsymbol{r}_0, \boldsymbol{a}, \boldsymbol{b}] = [\boldsymbol{r}_1, \boldsymbol{a}, \boldsymbol{b}]$.

[9] Dos vectores son *paralelos* si son proporcionales (esto es, si uno es un múltiplo del otro).

Ejercicio 1.14. Encontrar las ecuaciones paramétricas de la recta que pasa por $(3, 1, 5)$ y es paralela a la recta:

$$x = 4 - t, \qquad y = 2 + 3t, \qquad z = -4 + t.$$

Ejercicio 1.15. Demostrar que las dos rectas

$$\frac{x-1}{2} = \frac{y+1}{-3} = \frac{z}{4} \qquad \text{y} \qquad \frac{x-5}{2} = \frac{y+7}{-3} = \frac{z-8}{4}$$

coinciden; es decir, estos dos pares de ecuaciones representan los mismos puntos de $\mathbb{R}^3$.

Ejercicio 1.16. (a) Demostrar que estas dos rectas, dadas en forma paramétrica:

$$\begin{aligned} &x = 1 + t, &&y = -3 + 2t, &&z = -2 - t, \\ &x = 17 + 3s, &&y = 4 + s, &&z = -8 - s, \end{aligned}$$

se cortan; y hallar su (único) punto de intersección.

(b) Demostrar que estas otras dos rectas en $\mathbb{R}^3$, dadas en forma paramétrica:

$$\begin{aligned} &x = 2 - t, &&y = 1 + t, &&z = -2 - t, \\ &x = 1 + s, &&y = -2s, &&z = 3 + 2s, \end{aligned}$$

son *rectas sesgadas*: no tienen un punto de intersección pero no son paralelas.

Ejercicio 1.17. Obtener la ecuación (1.16) del plano $\mathcal{P}$ en los siguientes casos:

(a) $\mathcal{P}$ pasa por $\boldsymbol{r}_0 = (4, 2, -1)$; $\boldsymbol{n} = (1, -1, 3)$ es un vector normal a $\mathcal{P}$.

(b) $\mathcal{P}$ pasa por $\boldsymbol{r}_0 = (2, -1, 3)$ y es paralelo al plano $x - 2y - 3z + 6 = 0$.

(c) $\mathcal{P}$ incluye las dos rectas (que sí se cortan):

$$\frac{x+1}{2} = \frac{y-2}{3} = \frac{z-1}{1} \quad \text{y} \quad \frac{x+1}{1} = \frac{y-2}{-1} = \frac{z-1}{2}.$$

Ejercicio 1.18. Los dos planos

$$2x - y + z = 1 \qquad \text{y} \qquad 3x + y + z = 2$$

pasan por el punto $(1, 0, -1)$. Encontrar la ecuación paramétrica de su recta de intersección. ⟦ Indicación: calcular el producto vectorial de dos vectores normales. ⟧

Ejercicio 1.19. Si $\mathcal{P}$ es el plano que pasa por el punto $\boldsymbol{r}_0$ con vector normal $\boldsymbol{n}$, y si $\boldsymbol{r}_1$ es un punto *fuera* del plano $\mathcal{P}$, demostrar que la distancia d desde $\boldsymbol{r}_1$ al plano $\mathcal{P}$ está dada por la fórmula:

$$d = \frac{|(\boldsymbol{r}_1 - \boldsymbol{r}_0) \cdot \boldsymbol{n}|}{\|\boldsymbol{n}\|}.$$

⟦ Indicación: en un dibujo similar a la Figura 1.5, cuál sería la interpretación geométrica de la proyección de $\boldsymbol{r}_1 - \boldsymbol{r}_0$ a lo largo de $\boldsymbol{n}$? ⟧

Hallar la distancia de $(1, 1, 1)$ al plano $x + y + z = 0$.

Ejercicio 1.20. Demostrar que la ecuación

$$[\boldsymbol{r}_0, \boldsymbol{r}_1, \boldsymbol{r}] + [\boldsymbol{r}_0, \boldsymbol{r}, \boldsymbol{r}_2] + [\boldsymbol{r}, \boldsymbol{r}_1, \boldsymbol{r}_2] = [\boldsymbol{r}_0, \boldsymbol{r}_1, \boldsymbol{r}_2]$$

representa el plano que pasa por los tres puntos no colineales $\boldsymbol{r}_0$, $\boldsymbol{r}_1$, $\boldsymbol{r}_2$. Obtener la ecuación del plano que pasa por los tres puntos $(-2, 3, -1)$, $(2, 2, 3)$, $(-4, -1, 1)$.

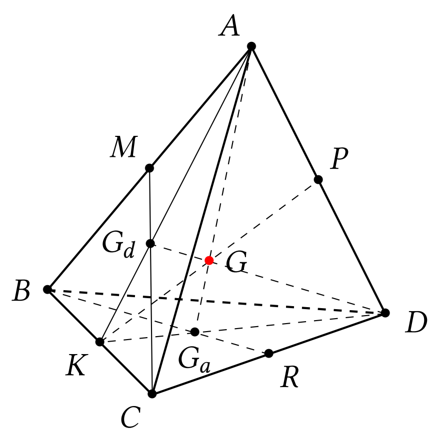

Figura 1.8: Las medianas y bimedianas de un tetraedro son concurrentes

Ejercicio 1.21. Un *tetraedro ABCD* tiene vértices $\boldsymbol{a} = (1, 1, 1)$, $\boldsymbol{b} = (1, -1, -1)$, $\boldsymbol{c} = (-1, 1, -1)$, $\boldsymbol{d} = (-1, -1, 1)$; mostrar que *ABCD* es regular (es decir, que todas sus facetas son triángulos equiláteros). Hallar las ecuaciones de los cuatro planos que incluyen las facetas $\triangle\, BCD$, $\triangle\, ACD$, $\triangle\, ABD$ y $\triangle\, ABC$.

Ejercicio 1.22. (a) Sea $\overline{\boldsymbol{ab}}$ un segmento en $\mathbb{R}^n$, que es parte de la recta parametrizada $\boldsymbol{r} = (1 - t)\boldsymbol{a} + t\boldsymbol{b}$, con $t \in \mathbb{R}$. Demostrar que el punto $\boldsymbol{r}$ divide este segmento[10] en la proporción $t : (1 - t) = \pm\|\boldsymbol{r} - \boldsymbol{a}\| : \|\boldsymbol{r} - \boldsymbol{b}\|$.

[10] Si el punto $\boldsymbol{r}$ de la recta está fuera del segmento referido, esta proporción es negativa.

(b) En el triángulo $\triangle ABC$ (Ejercicio 1.3) las *medianas* son los segmentos que unen cada vértice con el punto medio del lado opuesto (Ejercicio 1.4). Demostrar que el punto $\frac{1}{3}(\boldsymbol{a} + \boldsymbol{b} + \boldsymbol{c})$ queda en las tres medianas y las divide en la proporción $\frac{2}{3} : \frac{1}{3}$.

⟦ El punto de concurrencia de las medianas se llama el *centroide* (o baricentro) del triángulo $\triangle ABC$. ⟧

Ejercicio 1.23. El *teorema de Commandino* dice que en un tetraedro $ABCD$ cualquiera, las cuatro *medianas* y las tres *bimedianas* son concurrentes. (Véase la Figura 1.8.) Una mediana es el segmento, como AG_a, que une un vértice al centroide de la faceta opuesta (Ejercicio 1.22). Una bimediana es el segmento, como KP, que une los puntos medios de dos aristas opuestas (Ejercicio 1.4).

Demostrar este teorema, al comprobar que el punto $G = \frac{1}{4}(\boldsymbol{a} + \boldsymbol{b} + \boldsymbol{c} + \boldsymbol{d})$ es el punto medio de la bimediana KP y además divide la mediana AG_a en la proporción $\frac{3}{4} : \frac{1}{4}$.

2 Espacios vectoriales

Es sumamente difícil comprender la Matemática moderna sin una noción suficientemente clara de su estructura y del método con que procede, en sí mismos y por contraste con los de otras disciplinas científicas.

— José Joaquín Trejos Fernández[1]

De ahora en adelante, es conveniente adoptar un punto de vista más general (y por ende más abstracta) sobre los vectores. En este capítulo los vectores aparecen como símbolos que pertenecen a ciertos conjuntos; la suma y la multiplicación escalar se definen como reglas de combinación formales. Este punto de vista tiene la ventaja de resaltar la propiedad característica de los vectores, que es su *dependencia o independencia lineal*.

El primer paso es hacer una lista de las propiedades algebraicas de los escalares.

Definición 2.1. Un **cuerpo**[2] es un conjunto no vacío $\mathbb{F}$, cuyos elementos serán llamados **escalares**, en donde se definen dos operaciones, *suma* y *producto*; si $a, b \in \mathbb{F}$, su suma se escribe $\underline{a+b}$ y su producto se denota por $\underline{ab}$. Las formas concretas de la suma y del producto deben cumplir los siguientes requisitos, para $a, b, c \in \mathbb{F}$:

(a) $a + (b + c) = (a + b) + c$ (*asociatividad* de la suma);

(b) $a + b = b + a$ (*conmutatividad* de la suma);

(c) hay un único *cero* $\underline{0 \in \mathbb{F}}$ tal que $0 + a = a$;

(d) a cada $a \in \mathbb{F}$ le corresponde un único *negativo* $-a$, tal que $a + (-a) = 0$;

(e) $a(bc) = (ab)c$ (asociatividad del producto);

(f) $ab = ba$ (conmutatividad del producto);

(g) hay una única *identidad* $\underline{1 \in \mathbb{F}}$, con $1 \neq 0$, tal que $1a = a$;

(h) a cada $a \in \mathbb{F}$ con $a \neq 0$ le corresponde un único *recíproco* a^{-1}, tal que $aa^{-1} = 1$;

(i) $a(b + c) = ab + ac$ (*distributividad* del producto sobre la suma).

Nótese que si $a \in \mathbb{F}$, entonces $0a = (0 + 0)a = 0a + 0a$, así que $\underline{0a = 0}$. ◊

[1] En la introducción de su libro *Conceptos fundamentales de álgebra moderna*, Editorial Trejos Hermanos, San José, Costa Rica, 1968.

[2] El término cuerpo viene del alemán *Körper*, introducido por Dedekind en 1871; se llama *corps* en francés, pero en inglés se usa la palabra *field*. En español, no debe usarse la traducción secundaria *campo*, que denota campos vectoriales, campos magnéticos, etc.

Los **números naturales** $\underline{\mathbb{N}} := \{0, 1, 2, 3, \dots\}$ no forman un cuerpo, porque $1, 2, 3, \dots$ no tienen negativos en $\mathbb{N}$ [la propiedad (d)]. Los **números enteros**

$$\underline{\mathbb{Z}} := \{\dots, -3, -2, -1, 0, 1, 2, 3, \dots\}$$

tienen negativos y 1, pero $\pm 2, \pm 3$, etc. no poseen recíprocos en $\mathbb{Z}$ [la propiedad (h)].

Ejemplo 2.2. Los **números racionales** $\mathbb{F} = \underline{\mathbb{Q}}$ forman un cuerpo. Sus elementos son las fracciones p/q con $p \in \mathbb{Z}$, $q \in \mathbb{Z} \setminus \{0\}$. ◊

Ejemplo 2.3. Los **números reales** $\mathbb{F} = \underline{\mathbb{R}}$ forman otro cuerpo con $\mathbb{Q} \subset \mathbb{R}$. De hecho, la lista de propiedades (a–i) de la Definición 2.1 no es más que un catálogo de las propiedades conocidas de los números racionales y reales, Un cuerpo, entonces, es un conjunto cuyos elementos se pueden sumar, restar, multiplicar y dividir como si fueran números racionales o reales. ◊

Ejemplo 2.4. Un cuerpo tiene al menos dos elementos distintos, 0 y 1. Existe un cuerpo con solo esos dos elementos, $\underline{\mathbb{F}_2 := \{0, 1\}}$. Sus reglas para suma y producto son:

$$\begin{aligned} 0 + 0 = 1 + 1 = 0, &\quad 0 + 1 = 1 + 0 = 1, \\ 0 \cdot 0 = 0 \cdot 1 = 1 \cdot 0 = 0, &\quad 1 \cdot 1 = 1. \end{aligned}$$

(Para entender mejor estas operaciones, se puede reemplazar los "números" 0 y 1 por las palabras **par** e **impar**.) La aritmética de $\mathbb{F}_2$ se llama *aritmética booleana* o *binaria*. ◊

Ejemplo 2.5. Los **números complejos** $\mathbb{F} = \mathbb{C}$ forman un cuerpo, que incluye los números reales $\mathbb{R}$ como una parte. Al conjunto $\mathbb{R}$ se le agrega un elemento extra i, con la propiedad $\underline{i^2 = -1}$. (Los cuadrados de números reales son no negativos.) Entonces se define

$$\underline{\mathbb{C}} := \{\, a + bi : a, b \in \mathbb{R} \,\}, \tag{2.1}$$

esto es, el conjunto de polinomios de primer grado en la "incógnita" i, con las operaciones usuales sobre polinomios seguido por la regla de cancelación $i^2 = -1$. De esta manera,[3] la suma y el producto están dados por

$$\begin{aligned} (a + bi) + (c + di) &= (a + c) + (b + d)i, \\ (a + bi)(c + di) &= (ac - bd) + (ad + bc)i, \end{aligned}$$

para $a, b, c, d \in \mathbb{R}$. Así, por ejemplo,

$$(5 + 3i) + (7 - 2i) = 12 + i, \qquad (5 + 3i)(7 - 2i) = 35 + 11i - 6i^2 = 41 + 11i.$$

[3] También se denota un número complejo por $a + b\sqrt{-1}$. Leonhard Euler introdujo la letra i como sinónimo de $\sqrt{-1}$ en 1777 y hoy en día esta costumbre es casi universal. Una excepción se ve en textos de ingeniería eléctrica, donde se usa $j = \sqrt{-1}$ porque la letra i está reservada para la corriente eléctrica.

Si $a + bi \neq 0$, esto es, si $(a, b) \neq (0, 0)$ en $\mathbb{R}^2$, entonces $a^2 + b^2 > 0$. Nótese además que $(a + bi)(a - bi) = a^2 + b^2$. Entonces el *recíproco* de $a + bi$ es

$$\frac{1}{a + bi} = \frac{a - bi}{a^2 + b^2} = \frac{a}{a^2 + b^2} - \frac{b}{a^2 + b^2}\, i.$$

Si $z = a + bi \in \mathbb{C}$, su **conjugado complejo** es $\bar{z} := a - bi$, de modo que $z\bar{z} = a^2 + b^2$. El **valor absoluto** de z es el número real no negativo $|z| := \sqrt{a^2 + b^2}$. ◊

2.1. El concepto de espacio vectorial

Definición 2.6. Un **espacio vectorial** sobre un cuerpo $\mathbb{F}$ es un conjunto no vacío V cuyos elementos se llaman **vectores**, en donde se definen dos operaciones algebraicas. A cada par ordenado de vectores $(\boldsymbol{x}, \boldsymbol{y})$ le corresponde una **suma** $\boldsymbol{x} + \boldsymbol{y} \in V$; a cada escalar $c \in \mathbb{F}$ y a cada vecrtor $\boldsymbol{x} \in V$ le corresponde un **múltiplo** $c\boldsymbol{x} \in V$. Estas operaciones deben satisfacer las siguientes condiciones: si $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in V$ y $a, b \in \mathbb{F}$, entonces[4]

(a) $\boldsymbol{x} + \boldsymbol{y} = \boldsymbol{y} + \boldsymbol{x}$,

(b) $\boldsymbol{x} + (\boldsymbol{y} + \boldsymbol{z}) = (\boldsymbol{x} + \boldsymbol{y}) + \boldsymbol{z}$,

(c) hay un único *cero* $\mathbf{0} \in V$ tal que $\mathbf{0} + \boldsymbol{x} = \boldsymbol{x}$,

(d) a cada $\boldsymbol{x} \in V$ le corresponde un único *negativo* $-\boldsymbol{x} \in V$ tal que $\boldsymbol{x} + (-\boldsymbol{x}) = \mathbf{0}$,

(e) $a(\boldsymbol{x} + \boldsymbol{y}) = a\boldsymbol{x} + a\boldsymbol{y}$,

(f) $(a + b)\boldsymbol{x} = a\boldsymbol{x} + b\boldsymbol{x}$,

(g) $a(b\boldsymbol{x}) = (ab)\boldsymbol{x}$,

(h) $1\boldsymbol{x} = \boldsymbol{x}$. ◊

Fíjese que si $\boldsymbol{x} \in V$, entonces $0\boldsymbol{x} = (0 + 0)\boldsymbol{x} = 0\boldsymbol{x} + 0\boldsymbol{x}$, así que $0\boldsymbol{x} = \mathbf{0}$. (En estos apuntes, se usará la letra negrilla $\mathbf{0}$ para el cero vectorial y la letra clara 0 para el cero escalar.) Del mismo modo, es fácil comprobar que $(-1)\boldsymbol{x} = -\boldsymbol{x}$. Obviamente, se escribirá $\boldsymbol{x} - \boldsymbol{y} := \boldsymbol{x} + (-\boldsymbol{y})$.

Ejemplo 2.7. $\mathbb{R}^n$ es un espacio vectorial sobre el cuerpo $\mathbb{R}$. ◊

[4] Esta es esencialmente la definición original de espacio vectorial dada por Stefan Banach en su libro: *Theorie des opérations linéaires*, Monografie Matematyczne **1**, Warszawa, 1932. En vez de las condiciones (c) y (d), Banach puso: "$\boldsymbol{x} + \boldsymbol{y} = \boldsymbol{x} + \boldsymbol{z}$ implica $\boldsymbol{y} = \boldsymbol{z}$" y de ahí dedujo la existencia única del vector nulo $\mathbf{0}$.

Ejemplo 2.8. Más generalmente, para cualquier cuerpo $\mathbb{F}$ y cualquier $n \in \mathbb{N}$, $\mathbb{F}^n$ denotará la totalidad de las "n-tuplas" $\boldsymbol{x} = (x_1, x_2, \dots, x_n)$ con cada $x_k \in \mathbb{F}$. Al definir sumas y múltiplos de n-tuplas "entrada por entrada":

$$(x_1, \dots, x_n) + (y_1, \dots, y_n) := (x_1 + y_1, \dots, x_n + y_n), \qquad c(x_1, \dots, x_n) := (cx_1, \dots, cx_n),$$

se ve que $\mathbb{F}^n$ es un espacio vectorial sobre $\mathbb{F}$.

Fíjese que $\mathbb{F}^1 = \mathbb{F}$. Tampoco se excluye el caso trivial $n = 0$. Si se denota por $\mathbf{0}$ la "tupla vacía" ( ), se obtiene $\mathbb{F}^0 = \{\mathbf{0}\}$ con las operaciones triviales $\mathbf{0}+\mathbf{0} := \mathbf{0}$ y $c\,\mathbf{0} := \mathbf{0}$. ◊

Ejemplo 2.9. El espacio vectorial $\mathbb{F}_2^8$,sobre el cuerpo $\mathbb{F}_2$, es el espacio de los *bytes* u *octetos* usados en informática; los escalares en $\mathbb{F}_2$ son los *bits*. A cada octeto le corresponde un número entero entre 0 y 255; pero la operación de suma en $\mathbb{F}_2^8$ *difiere* de la suma de enteros. Por ejemplo, dadas las representaciones binarias $114 \leftrightarrow (0\,1\,1\,1\,0\,0\,1\,0)$ y $213 \leftrightarrow (1\,1\,0\,1\,0\,1\,0\,1)$, se ve que

$$(0\,1\,1\,1\,0\,0\,1\,0) + (1\,1\,0\,1\,0\,1\,0\,1) = (1\,0\,1\,0\,0\,1\,1\,1) \leftrightarrow 167 \in \mathbb{N},$$

pero $114 + 213 \neq 167$ en $\mathbb{N}$. ◊

Ejemplo 2.10. La totalidad de *polinomios con coeficientes reales* es un espacio vectorial (sobre $\mathbb{R}$) comúnmente denotado por $\mathbb{R}[t]$: el símbolo t es la *incógnita* y un elemento típico de $\mathbb{R}[t]$ es de la forma

$$p(t) = a_0 + a_1 t + a_2 t^2 + \cdots + a_n t^n \quad \text{con cada} \quad a_k \in \mathbb{R},\ a_n \neq 0.$$

El número natural n es el *grado* del polinomio $p(t)$. La suma de polinomios y la multiplicación de polinomios por escalares se efectúa en la forma usual. ◊

Ejemplo 2.11. La totalidad de *funciones continuas* $f\colon [a,b] \to \mathbb{R}$ definidas en un intervalo acotado $[a,b] \subset \mathbb{R}$ es un espacio vectorial (sobre $\mathbb{R}$) comúnmente denotado por $C[a,b]$. Las operaciones de suma y multiplicación escalar se definen "punto por punto", es decir:

$$f + g(t) := f(t) + g(t), \qquad cf(t) := cf(t),$$

si $f, g \in C[a,b]$, $c \in \mathbb{R}$. (Fíjese que las funciones $f + g$ y cf son también continuas.) ◊

Ejemplo 2.12. Es útil recordar que el conjunto de un solo elemento $\{\mathbf{0}\}$ es un espacio vectorial sobre cualquier cuerpo $\mathbb{F}$: ese elemento es necesariamente el cero, así que las operaciones algebraicas son simplemente: $\mathbf{0} + \mathbf{0} = \mathbf{0}$; y $c\,\mathbf{0} = \mathbf{0}$ para $c \in \mathbb{F}$. ◊

Definición 2.13. Sea V un espacio vectorial sobre $\mathbb{F}$. Un **subespacio** de V es una parte $W \subseteq V$ tal que W es también un espacio vectorial sobre $\mathbb{F}$, con las mismas operaciones de suma y multiplicación escalar. Dicho de otro modo, W es un subespacio de V si $W \subseteq V$ y si para $\boldsymbol{x}, \boldsymbol{y} \in W$, $c \in \mathbb{F}$, se verifican $\boldsymbol{x} + \boldsymbol{y} \in W$, $c\boldsymbol{x} \in W$.

Se usará la notación $\underline{W \leqslant V}$ para significar que W es un subespacio de V. ◊

Por ejemplo, si $\mathbb{F} = \mathbb{R}$, $V = \mathbb{R}[t]$, y si $n \in \mathbb{N}$, sea $\mathbb{R}_n[t]$ la colección de polinomios de grado *no mayor que n*. (Los polinomios constantes a_0 se consideran de grado 0). Como las sumas y los múltiplos de polinomios en $\mathbb{R}_n[t]$ son también de grado menor que n, es decir, pertenecen a $\mathbb{R}_n[t]$, se ve que $\mathbb{R}_n[t]$ es subespacio de $\mathbb{R}[t]$.

Otro ejemplo de subespacio es[5]

$$C^1[a,b] := \{\, f\colon [a,b] \to \mathbb{R} : f'(x) \text{ existe para } a \leqslant x \leqslant b \,\},$$

el cual es un subespacio de $C[a,b]$, porque cualquier función diferenciable en $[a,b]$ es también continua.

2.2. Independencia lineal

Definición 2.14. Sea V un espacio vectorial sobre $\mathbb{F}$. Una expresión del tipo

$$\boxed{a_1\boldsymbol{x}_1 + a_2\boldsymbol{x}_2 + \cdots + a_m\boldsymbol{x}_m,} \tag{2.2}$$

donde $a_k \in \mathbb{F}$, $\boldsymbol{x}_k \in V$ para $k = 1, 2, \dots, m$, representa un vector en V y se llama una **combinación lineal** de los vectores $\boldsymbol{x}_1, \dots, \boldsymbol{x}_m$, con **coeficientes** $a_1, \dots, a_m$. La combinación lineal se dice "trivial" si todos los coeficientes son cero ($a_1 = a_2 = \cdots = a_m = 0$), en cuyo caso la combinación (2.2) vale $\mathbf{0}$.

Se dice que un conjunto de m vectores $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_m\} \subseteq V$ es **linealmente dependiente** si hay un juego de coeficientes $a_1, \dots, a_m$, *no todos cero,* tal que

$$\boxed{a_1\boldsymbol{x}_1 + a_2\boldsymbol{x}_2 + \cdots + a_m\boldsymbol{x}_m = \mathbf{0}.} \tag{2.3}$$

En cambio, si los únicos coeficientes que hacen cumplir (2.3) son $a_1 = a_2 = \cdots = a_m = 0$, se dice que el conjunto de vectores $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_m\}$ es **linealmente independiente**. ◊

Se dice que los vectores $\boldsymbol{x}_1, \dots, \boldsymbol{x}_m$ son linealmente dependientes [o independientes] si el conjunto de ellos es, linealmente dependiente [o independiente], respectivamente.

[5]Las derivadas en los extremos del intervalo $[a,b]$ son unilaterales: $f'(a) := \lim_{h\downarrow 0}(f(a+h) - f(a))/h$ y $f'(b) := \lim_{h\downarrow 0}(f(b-h) - f(b))/h$.

Un conjunto cualquiera $X \subset V$ se llama linealmente independiente si cada parte finita de X es linealmente independiente. Dicho de otro modo, X es linealmente independiente si la ecuación (2.3) admite solamente la solución trivial $a_1 = a_2 = \cdots = a_m = 0$ toda vez que $\boldsymbol{x}_1, \dots, \boldsymbol{x}_m \in X$.

Es oportuno notar que el conjunto[6] de un solo vector $\{\boldsymbol{x}_1\}$ es linealmente independiente si y solo si $\boldsymbol{x}_1 \neq \mathbf{0}$; y que $\{\mathbf{0}\}$ es linealmente *dependiente*. Además, cualquier colección de vectores es linealmente dependiente si uno de sus miembros es igual a $\mathbf{0}$ (tome la coeficiente correspondiente igual a 1 y los demás coeficientes 0). Para prevenir situaciones especiales, se declara (como un convenio) que el conjunto vacío $\emptyset$ es linealmente *independiente*.

Lema 2.15. *Una colección de vectores es linealmente dependiente si y solo si al menos uno de ellos puede expresarse como una combinación lineal de los demás.*

Demostración. Sea $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_m\}$ una colección de vectores en V. Si es linealmente dependiente, la ecuación (2.3) posee una solución con coeficientes $a_1, \dots, a_m$ no todos cero. Después de permutar el orden de los vectores si fuera necesario, se puede suponer que $a_1 \neq 0$. Pero entonces es posible resolver (2.3) para $\boldsymbol{x}_1$:

$$\boldsymbol{x}_1 = -\frac{a_2}{a_1}\boldsymbol{x}_2 - \frac{a_3}{a_1}\boldsymbol{x}_3 - \cdots - \frac{a_m}{a_1}\boldsymbol{x}_m,$$

lo cual expresa $\boldsymbol{x}_1$ como combinación lineal de $\boldsymbol{x}_2, \boldsymbol{x}_3, \dots, \boldsymbol{x}_m$.

Por otro lado, si $\boldsymbol{x}_1 = b_2\boldsymbol{x}_2 + \cdots + b_m\boldsymbol{x}_m$, entonces $-\boldsymbol{x}_1 + b_2\boldsymbol{x}_2 + \cdots + b_m\boldsymbol{x}_m = 0$. Esto es, la ecuación (2.3) se cumple con $a_1 = -1$, de modo que no todos los coeficientes son ceros. Por lo tanto, $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_m\}$ es linealmente dependiente. □

Corolario 2.16. Dos *vectores son linealmente dependientes si y solo si son* proporcionales.

Demostración. Si $\boldsymbol{x}, \boldsymbol{y} \in V$ son linealmente dependientes, entonces $\boldsymbol{x} = a\boldsymbol{y}$ para algún $a \in \mathbb{F}$, o bien $\boldsymbol{y} = b\boldsymbol{x}$ para algún $b \in \mathbb{F}$; en ambos casos, $\boldsymbol{x}, \boldsymbol{y}$ son proporcionales, es decir, uno de ellos es un múltiplo del otro. □

Ejemplo 2.17. En el espacio vectorial $\mathbb{F}^n$, se usan las abreviaturas

$$\boldsymbol{e}_1 := (1, 0, 0, \dots, 0), \quad \boldsymbol{e}_2 := (0, 1, 0, \dots, 0), \ \dots, \quad \boldsymbol{e}_n := (0, 0, \dots, 0, 1).$$

En general, $\boldsymbol{e}_k$ denotará el vector cuyo k-ésima coordenada es 1 y cuyas demás coordenadas son 0. Nótese que

$$a_1\boldsymbol{e}_1 + a_2\boldsymbol{e}_2 + \cdots + a_n\boldsymbol{e}_n = (a_1, a_2, \dots, a_n)$$

y que esta combinación lineal vale $\mathbf{0}$ solo si cada $a_k = 0$. Luego el conjunto $\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}$ es linealmente independiente. ◊

[6] Un conjunto con un solo miembro se llama un *singulete*.

Ejemplo 2.18. En $C[-\pi,\pi]$, las funciones $\underline{\operatorname{sen} t}$ y $\underline{\cos t}$ son linealmente independientes. Para ver eso, supóngase que $a_1, a_2 \in \mathbb{R}$ cumplen

$$a_1 \operatorname{sen} t + a_2 \cos t \equiv 0 \quad \text{para todo } t \in [-\pi,\pi].$$

Al sustituir $t = 0$, se obtiene $a_2 = 0$; al poner $t = \frac{\pi}{2}$, se obtiene $a_1 = 0$. Luego la ecuación se cumple *para todo* t solo si $a_1 = a_2 = 0$. $\lozenge$

Ejemplo 2.19. En $\mathbb{R}[t]$, los *monomios* $1, t, t^2, \dots, t^m$ son linealmente independientes. Porque si $a_0, a_1, \dots, a_m$ son tales que

$$p(t) = a_0 + a_1 t + a_2 t^2 + \cdots + a_m t^m \equiv 0,$$

(el cero de $\mathbb{R}[t]$ es el polinomio constante de valor 0), entonces los coeficientes a_k son todos 0. $\lozenge$

Definición 2.20. Sea $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_m\}$ una colección de vectores en V. El **subespacio generado** por $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_m\}$ es el menor subespacio W de V que contiene cada uno de los vectores $\boldsymbol{x}_k$. Es evidente que W además contiene todas las combinaciones lineales posibles de la forma (2.2); y que las sumas y los múltiplos de estas combinaciones son también de la forma (2.2). Se deduce que W es la totalidad de estas combinaciones lineales; y se escribe

$$W = \boxed{\operatorname{lin}\langle \boldsymbol{x}_1, \dots, \boldsymbol{x}_m \rangle := \{\, a_1\boldsymbol{x}_1 + \cdots + a_m\boldsymbol{x}_m : a_1, \dots, a_m \in \mathbb{F} \,\}} \tag{2.4}$$

para denotar este subespacio generado por e conjunto $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_m\}$.

Si X es una parte (posiblemente infinita) de V, el subespacio $\underline{\operatorname{lin}\langle X\rangle}$ generado por X es el menor subespacio de V que contiene X. Entonces $\operatorname{lin}\langle X\rangle$ es la totalidad de las combinaciones lineales (finitas) de los vectores en X. $\lozenge$

En $\mathbb{R}^3$, el subespacio generado por $\boldsymbol{e}_1 = (1,0,0)$ y $\boldsymbol{e}_2 = (0,1,0)$ es $\operatorname{lin}\langle \boldsymbol{e}_1, \boldsymbol{e}_2\rangle = \{\, (a_1, a_2, 0) : a_1, a_2 \in \mathbb{R} \,\}$. El vector $\boldsymbol{e}_3$ no pertenece a ese subespacio.

2.3. Bases y dimensión de un espacio vectorial

Definición 2.21. Sea V un espacio vectorial sobre un cuerpo $\mathbb{F}$. Una **base** de V es una parte $X \subset V$ tal que:

(a) X *genera* V, esto es, $\operatorname{lin}\langle X\rangle = V$;

(b) X es *linealmente independiente*. $\lozenge$

Ejemplo 2.22. La **base estándar** de $\mathbb{F}^n$ es el conjunto $\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}$ (Ejemplo 2.17). $\lozenge$

Ejemplo 2.23. El conjunto $\{(1,0,-1),\ (-1,2,-1),\ (1,1,1)\}$ es una base de $\mathbb{R}^3$. $\Diamond$

Ejemplo 2.24. El conjunto de monomios $\{1, t, t^2, \dots, t^n\}$ es una base para el espacio vectorial $\mathbb{R}_n[t]$ de los polinomios reales de grado no mayor que n. Ya se ha observado (Ejemplo 2.19) que este conjunto es linealmente independiente; también genera $\mathbb{R}_n[t]$ porque las combinaciones de estos monomios son precisamente todos los polinomios cuyo grado no excede n; o sea, $\operatorname{lin}\langle 1, t, t^2, \dots, t^n \rangle = \mathbb{R}_n[t]$. $\Diamond$

Ejemplo 2.25. El conjunto infinito de monomios $\{1, t, t^2, \dots, t^n, \dots\}$ es una base para el espacio vectorial $\mathbb{R}[t]$ de *todos* los polinomios reales. Cualquier polinomio es (por definición) una combinación lineal de algunos de estos monomios; y este conjunto es linealmente independiente, pues las combinaciones lineales no triviales de monomios son polinomios no nulos. Este ejemplo evidencia que una base puede ser infinito. $\Diamond$

► La Proposición siguiente dice que una base de un espacio vectorial es un conjunto linealmente independiente *maximal*.

Proposición 2.26. *Sea $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ una base para un espacio vectorial V sobre $\mathbb{F}$, y sean $\boldsymbol{y}_1, \dots, \boldsymbol{y}_m \in V$. Si $m > n$, entonces $\boldsymbol{y}_1, \dots, \boldsymbol{y}_m$ son linealmente* dependientes.

Demostración. Supóngase, por el contrario, que $m > n$ y que $\boldsymbol{y}_1, \dots, \boldsymbol{y}_m$ son linealmente independientes. Entonces ningún $\boldsymbol{y}_k$ es $\mathbf{0}$, y hay coeficientes $a_1, \dots, a_m \in \mathbb{F}$, no todos cero, tales que

$$\boldsymbol{y}_1 = a_1\boldsymbol{x}_1 + a_2\boldsymbol{x}_2 + \cdots + a_n\boldsymbol{x}_n.$$

Se puede suponer que $a_1 \neq 0$. (Si fuere necesario, primero se permuta $\boldsymbol{x}_1$ con algún $\boldsymbol{x}_k$ tal que $a_k \neq 0$.) Luego se despeja $\boldsymbol{x}_1$ así:

$$\boldsymbol{x}_1 = \frac{1}{a_1}\,\boldsymbol{y}_1 - \frac{a_2}{a_1}\,\boldsymbol{x}_2 - \cdots - \frac{a_n}{a_1}\,\boldsymbol{x}_n.$$

Así, $\operatorname{lin}\langle \boldsymbol{y}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \dots, \boldsymbol{x}_n \rangle$ contiene $\boldsymbol{x}_1$ y también $\boldsymbol{x}_2, \dots, \boldsymbol{x}_n$: como incluye la base dada, este subespacio es todo: $\operatorname{lin}\langle \boldsymbol{y}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \dots, \boldsymbol{x}_n \rangle = V$.

En consecuencia, hay coeficientes $a'_1, a'_2, \dots, a'_n \in \mathbb{F}$, no todos cero, tales que

$$\boldsymbol{y}_2 = a'_1\boldsymbol{y}_1 + a'_2\boldsymbol{x}_2 + \cdots + a'_n\boldsymbol{x}_n.$$

Como $\boldsymbol{y}_1, \boldsymbol{y}_2$ son linealmente independientes (por hipótesis), los coeficientes $a'_2, \dots, a'_n$ no son todos cero. Después de permutar (si fuere necesario) los elementos de la lista $\{\boldsymbol{x}_2, \dots, \boldsymbol{x}_n\}$, se puede suponer que $a'_2 \neq 0$. Luego

$$\boldsymbol{x}_2 = \frac{1}{a'_2}\,\boldsymbol{y}_2 - \frac{a'_1}{a'_2}\,\boldsymbol{y}_1 - \frac{a'_3}{a'_2}\,\boldsymbol{x}_3 - \cdots - \frac{a'_n}{a'_2}\,\boldsymbol{x}_n.$$

Por lo tanto, $\mathrm{lin}\langle \boldsymbol{y}_1, \boldsymbol{y}_2, \boldsymbol{x}_3, \dots, \boldsymbol{x}_n\rangle$ contiene $\boldsymbol{x}_2$ y $\boldsymbol{x}_1$, y además contiene los otros vectores de la base original, así que $\mathrm{lin}\langle \boldsymbol{y}_1, \boldsymbol{y}_2, \boldsymbol{x}_3, \dots, \boldsymbol{x}_n\rangle = V$.

Al repetir este argumento r veces,[7] se obtiene $\mathrm{lin}\langle \boldsymbol{y}_1, \dots, \boldsymbol{y}_r, \boldsymbol{x}_{r+1}, \dots, \boldsymbol{x}_n\rangle = V$. Por lo tanto, hay coeficientes $b_1, \dots, b_r, c_{r+1}, \dots, c_n$ tales que

$$\boldsymbol{y}_{r+1} = b_1\boldsymbol{y}_1 + \cdots + b_r\boldsymbol{y}_r + c_{r+1}\boldsymbol{x}_{r+1} + \cdots + c_n\boldsymbol{x}_n.$$

Los $c_{r+1}, \dots, c_n$ no son todos cero, ya que $\boldsymbol{y}_1, \dots, \boldsymbol{y}_{r+1}$ serían entonces linealmente dependientes. Después de permutar $\boldsymbol{x}_{r+1}, \dots, \boldsymbol{x}_n$ si fuere necesario, se puede suponer que $c_{r+1} \neq 0$. Entonces

$$\boldsymbol{x}_{r+1} = \frac{1}{c_{r+1}}\boldsymbol{y}_{r+1} - \frac{b_1}{c_{r+1}}\boldsymbol{y}_1 - \cdots - \frac{b_r}{c_{r+1}}\boldsymbol{y}_r - \frac{c_{r+2}}{c_{r+1}}\boldsymbol{x}_{r+2} - \cdots - \frac{c_n}{c_{r+1}}\boldsymbol{x}_n.$$

Por lo tanto, $\mathrm{lin}\langle \boldsymbol{y}_1, \dots, \boldsymbol{y}_{r+1}, \boldsymbol{x}_{r+2}, \dots, \boldsymbol{x}_n\rangle = V$. Después de n repeticiones del argumento, se concluye que $\mathrm{lin}\langle \boldsymbol{y}_1, \dots, \boldsymbol{y}_n\rangle = V$.

Pero entonces hay coeficientes $d_1, \dots, d_n \in \mathbb{F}$ tales que

$$\boldsymbol{y}_{n+1} = d_1\boldsymbol{y}_1 + d_2\boldsymbol{y}_2 + \cdots + d_n\boldsymbol{y}_n,$$

lo cual contradice la supuesta independencia lineal de $\boldsymbol{y}_1, \dots, \boldsymbol{y}_m$. En síntesis: si $m \geqslant n+1$, no es posible que $\boldsymbol{y}_1, \dots, \boldsymbol{y}_m$ sean linealmente independientes. □

Corolario 2.27. *Si* $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ *y* $\{\boldsymbol{y}_1, \dots, \boldsymbol{y}_m\}$ *son dos bases para un espacio vectorial* V*, entonces* $\underline{m = n}$.

Demostración. Como $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ es una base y $\{\boldsymbol{y}_1, \dots, \boldsymbol{y}_m\}$ es linealmente independiente, la Proposición 2.26 garantiza que $m \leqslant n$. Como $\{\boldsymbol{y}_1, \dots, \boldsymbol{y}_m\}$ es una base y $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ es linealmente independiente, se deduce del mismo modo que $n \leqslant m$. □

Definición 2.28. Si un espacio vectorial V sobre $\mathbb{F}$ posee una base finita, *todas* las bases tienen el mismo número (finito) de elementos, por el Corolario 2.27. Este número n es la **dimensión** de V. En símbolos, $\underline{\dim V = n}$ o a veces $\underline{\dim_{\mathbb{F}} V = n}$. Además, se dice que V es *finitodimensional*.

Por el contrario, si una (y por ende toda) base de un espacio vectorial W es infinito, se dice que W es *infinitodimensional* y se escribe $\dim W = \infty$. ◊

En particular, vale $\dim \mathbb{F}^n = n$, ya que $\mathbb{F}^n$ posee la base estándar $\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}$.

En el caso del espacio vectorial trivial $\{\mathbf{0}\}$ (cualquier que sea $\mathbb{F}$), la única parte linealmente independiente de $\{\mathbf{0}\}$ es el vacío $\emptyset$, y el menor subespacio que contiene $\emptyset$ es $\{\mathbf{0}\}$. Luego $\emptyset$ es una base de V y esta base contiene 0 elementos: así que $\dim\{\mathbf{0}\} = 0$.

[7] Esta prueba es un ejemplo de un argumento "por inducción". Sin embargo, no hace falta apelar al principio de inducción porque el número de iteraciones es finito.

► Para poder *construir* bases, no basta con saber que existen; se debe tener algún procedimiento de búsqueda. La Proposición siguiente afirma que *siempre es posible completar una base parcial*.

Proposición 2.29. *Sea V un espacio vectorial finitodimensional y sea $\{\boldsymbol{x}_1,\dots,\boldsymbol{x}_m\}$ un conjunto linealmente independiente de vectores en V. Entonces, bien $\{\boldsymbol{x}_1,\dots,\boldsymbol{x}_m\}$ ya es una base para V, o bien hay otros vectores $\boldsymbol{x}_{m+1},\dots,\boldsymbol{x}_n$ tal que $\{\boldsymbol{x}_1,\dots,\boldsymbol{x}_n\}$ sea una base de V.*

Demostración. Sea $n = \dim V$. Como $\{\boldsymbol{x}_1,\dots,\boldsymbol{x}_m\}$ es linealmente independiente, vale $m \leqslant n$ por la Proposición 2.26. Si $\mathrm{lin}\langle \boldsymbol{x}_1,\dots,\boldsymbol{x}_m\rangle \neq V$, hay un vector no nulo $\boldsymbol{x}_{m+1} \in V$ tal que $\boldsymbol{x}_{m+1} \notin \mathrm{lin}\langle \boldsymbol{x}_1,\dots,\boldsymbol{x}_m\rangle$. Por el Lema 2.15, se ve que $\{\boldsymbol{x}_1,\dots,\boldsymbol{x}_{m+1}\}$ es linealmente independiente.

Al repetir este argumento r veces, se obtiene una colección linealmente independiente de $(m+r)$ vectores $\{\boldsymbol{x}_1,\dots,\boldsymbol{x}_{m+r}\}$ en V, toda vez que $m+r \leqslant n$. Tales repeticiones deben terminar cuando $\mathrm{lin}\langle \boldsymbol{x}_1,\dots,\boldsymbol{x}_{m+r}\rangle = V$; en ese momento, $\{\boldsymbol{x}_1,\dots,\boldsymbol{x}_{m+r}\}$ es una base de V, y por el Corolario 2.27, resulta que $m + r = n$. □

Corolario 2.30. *Si W es un subespacio de V, entonces* $\dim W \leqslant \dim V$.

Demostración. Si $m = \dim W$, entonces W posee una base $\{\boldsymbol{x}_1,\dots,\boldsymbol{x}_m\}$, la cual es un conjunto linealmente independiente en el espacio vectorial V. Por la Proposición 2.29 anterior, se puede completar una base de V, al agregar algunos otros vectores $\boldsymbol{x}_{m+1},\dots,\boldsymbol{x}_n$ a $\{\boldsymbol{x}_1,\dots,\boldsymbol{x}_m\}$. Luego $\dim V = n \geqslant m$. □

Definición 2.31. Un **isomorfismo lineal** entre dos espacios vectoriales V y W sobre un cuerpo $\mathbb{F}$ es una *biyección* $\underline{T\colon V \to W}$ tal que

$$T(\boldsymbol{x}+\boldsymbol{y}) = T(\boldsymbol{x}) + T(\boldsymbol{y}), \quad T(c\boldsymbol{x}) = c\,T(\boldsymbol{x}) \quad \text{para todo} \quad \boldsymbol{x},\boldsymbol{y} \in V,\ c \in \mathbb{F}.$$

Si hay un isomorfismo lineal entre V y W, dícese que V y W son (linealmente) **isomorfos**.

Resulta que si $\{\boldsymbol{x}_1,\dots,\boldsymbol{x}_n\}$ es una base de V, entonces $\{T(\boldsymbol{x}_1), T(\boldsymbol{x}_2),\dots,T(\boldsymbol{x}_n)\}$ es una base de W. Luego V y W son isomorfos solo si tienen la misma dimensión. ◊

Proposición 2.32. *Sea V un espacio vectorial sobre $\mathbb{F}$ con $\dim_{\mathbb{F}} V = n$. Entonces V es isomorfo a $\mathbb{F}^n$.*

Demostración. Sea $\{\boldsymbol{x}_1,\dots,\boldsymbol{x}_n\}$ una base de V. Si $\boldsymbol{x} \in V$, entonces

$$\boxed{\boldsymbol{x} = a_1\boldsymbol{x}_1 + a_2\boldsymbol{x}_2 + \cdots + a_n\boldsymbol{x}_n} \tag{2.5}$$

y los coeficientes $a_1,\dots,a_n \in \mathbb{F}$ son *únicos* (véase el Ejercicio 2.16). Defínase $T\colon V \to \mathbb{F}^n$ por $T(\boldsymbol{x}) := (a_1, a_2,\dots,a_n)$. Es fácil comprobar que T es biyectiva y preserva sumas y múltiplos escalares. Por lo tanto, T es el isomorfismo lineal deseado. □

La Proposición 2.32 justifica *a posteriori* el énfasis que se puso sobre el ejemplo $\mathbb{R}^n$ en el Capítulo 1: ahora se sabe que $\mathbb{R}^n$ es, hasta isomorfismo, el único espacio vectorial real de dimensión n. Sin embargo, es conveniente mantener el punto de vista *abstracto* sobre espacios vectoriales que se ha desarrollado en las últimas secciones, para resaltar *las propiedades estructurales* de los espacios vectoriales, en vez de enfocar los detalles de cálculos numéricos con coordenadas.

2.4. Ejercicios sobre espacios vectoriales

Ejercicio 2.1. Si $m \in \mathbb{N}$ con $m \geqslant 2$, el conjunto de *residuos bajo división por m* es $\{0, 1, 2, \dots, m-1\}$. Se designa un conjunto finito con m elementos así:[8]

$$\mathbb{F}_m := \{\overline{0}, \overline{1}, \overline{2}, \dots, \overline{m-1}\}.$$

Las operaciones aritméticas en $\mathbb{F}_m$ son los residuos de suma y multiplicación de enteros:

$$\left\{ \begin{matrix} \overline{a} + \overline{b} = \overline{c} \\ \overline{a} \cdot \overline{b} = \overline{d} \end{matrix} \right\} \quad \text{en } \mathbb{F}_m \text{ si } \quad \left\{ \begin{matrix} a + b \equiv c \text{ mód } m \\ a \cdot b \equiv d \text{ mód } m \end{matrix} \right\} \quad \text{en } \mathbb{Z};$$

o lo que es lo mismo, si m divide $(a + b) - c$ y $(a \cdot b) - d$.

Por ejemplo, en $\mathbb{F}_7$ valen $\overline{3} + \overline{6} = \overline{2}$ y $\overline{3} \cdot \overline{6} = \overline{4}$.

(a) Demostrar que $\mathbb{F}_5$ es un cuerpo; pero que $\mathbb{F}_6$ **no es** un cuerpo.

(b) ¿Para cuáles $m \in \mathbb{N}$ es $\mathbb{F}_m$ un cuerpo?

Ejercicio 2.2. Si X es un conjunto cualquiera, $\mathcal{P}(X)$ denota la totalidad de las partes[9] de X. Se define como sigue la “suma” de $A \subseteq X$ y $B \subseteq X$ y la “multiplicación escalar” de A por elementos de $\mathbb{F}_2$:

$$A \mathbin{\#} B := (A \cup B) \setminus (A \cap B), \qquad 0 \cdot A = \emptyset, \quad 1 \cdot A = A,$$

Demostrar que $\mathcal{P}(X)$, con estas operaciones, es un espacio vectorial sobre el cuerpo $\mathbb{F}_2$.

Ejercicio 2.3. Demostrar directamente que los cuatro vectores e $\mathbb{R}^3$:

$$(1, 0, 0), \quad (0, 1, 0), \quad (0, 0, 1), \quad (1, 1, 1)$$

son linealmente *dependientes* en $\mathbb{R}^3$, pero que cualquier tres de ellos son linealmente *in*dependientes.

[8]Se denota ‘$a \equiv b$ mód m’ si y solo si $(a - b)$ es un múltiplo entero de m. Esta es una relación de equivalencia sobre $\mathbb{Z}$. Sus *clases de equivalencia* son los elementos de $\mathbb{F}_m$, al identificar la clase de r con $\overline{r} := \{\, qm + r : q \in \mathbb{Z} \,\}$.

[9]Lamentablemente, hay quienes dicen “subconjunto” en vez de *parte*; hasta el DRAE lo tolera. El conjunto $A \mathbin{\#} B$ a veces se llama la **diferencia simétrica** de A y B.

Ejercicio 2.4. Demostrar que los tres vectores

$$(1, 1, 0), \quad (1, 1, 1), \quad (0, 1, -1)$$

son linealmente independientes en $\mathbb{R}^3$. Expresar los vectores $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3$ de la base estándar como combinaciones lineales de esos vectores.

Ejercicio 2.5. Demostrar que los tres polinomios de segundo grado en $\mathbb{R}[t]$:

$$\tfrac{1}{2}t(t-1), \quad 1 - t^2, \quad \tfrac{1}{2}t(t+1)$$

son linealmente independientes en $\mathbb{R}[t]$. Expresar los monomios 1, t, t^2 como combinaciones lineales de ellos.

Ejercicio 2.6. Expresar los vectores $(1, 1, 0)$, $(1, 0, 1)$, $(1, 1, 1)$ en $\mathbb{R}^3$ como combinaciones lineales de los tres vectores $\{(1, 1, 2),\ (2, 2, 1),\ (1, 2, 2)\}$.

Ejercicio 2.7. Si $r, s, t, u \in \mathbb{R}$ son números *distintos*, demostrar que

$$(1, 1, 1, 1), \quad (r, s, t, u), \quad (r^2, s^2, t^2, u^2), \quad (r^3, s^3, t^3, u^3)$$

son vectores linealmente independientes en $\mathbb{R}^4$.

⟦ Indicación: un polinomio cúbico no nulo, $p(t) = a_0 + a_1 t + a_2 t^2 + a_3 t^3$, no puede tener más de tres raíces distintas. ⟧

Ejercicio 2.8. Determinar si los triples de vectores en $\mathbb{R}^3$:

$$A = \{(2, 1, 3),\ (1, 0, 2),\ (1, 2, 0)\} \quad \text{y} \quad B = \{(1, 2, 3),\ (1, 1, 0),\ (1, 0, 1)\}$$

son conjuntos linealmente independientes, o no.

Ejercicio 2.9. ¿Cuales de los siguientes conjuntos forman *subespacios* de $\mathbb{R}^3$?

(a) $\{\, (x_1, x_2, x_3) \in \mathbb{R}^3 : x_2 = x_1 + x_3 \,\}$;

(b) $\{\, (1, x_2, 1) \in \mathbb{R}^3 : x_2 \in \mathbb{R} \,\}$;

(c) $\{\, (x_1, x_2, x_3) \in \mathbb{R}^3 : x_1 + x_2 + x_3 = 4 \,\}$;

(d) $\{\, (x_1, x_2, x_3) \in \mathbb{R}^3 : x_1 x_3 = 0 \,\}$.

Ejercicio 2.10. Sean $\boldsymbol{x}$, $\boldsymbol{y}$ dos vectores en $\mathbb{R}^n$. Demostrar que $\{\, \boldsymbol{z} \in \mathbb{R}^n : \boldsymbol{z} \cdot \boldsymbol{x} = \boldsymbol{z} \cdot \boldsymbol{y} = 0 \,\}$ es un subespacio de $\mathbb{R}^n$.

Ejercicio 2.11. (a) Demostrar que los pares de vectores en $\mathbb{R}^3$:

$$A = \{(2,1,-1),\ (0,1,1)\} \quad \text{y} \quad B = \{(3,1,-2),\ (-1,3,4)\}$$

generan el *mismo* subespacio de $\mathbb{R}^3$.

(b) Demostrar que los conjuntos de vectores

$$C = \{(2,1,-1),\ (1,1,0)\} \quad \text{y} \quad D = \{(3,2,-1),\ (4,3,-1)\ (11,8,-3)\}$$

generan el mismo subespacio de $\mathbb{R}^3$.

Ejercicio 2.12. Si el conjunto de tres vectores $\{\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}\}$ es linealmente independiente en un espacio vectorial V, demostrar que el conjunto $\{\boldsymbol{x}+\boldsymbol{y},\ \boldsymbol{y}+\boldsymbol{z},\ \boldsymbol{z}+\boldsymbol{x}\}$ es también linealmente independiente en V.

Ejercicio 2.13. Escoger dos bases diferentes de $\mathbb{R}^4$ de entre los vectores:

$$(1,1,0,0), \quad (1,0,1,0), \quad (1,0,0,1), \quad (0,1,1,0), \quad (0,1,0,1), \quad (0,0,1,1).$$

¿Cuántas bases de $\mathbb{R}^4$ pueden formarse con cuaternas de estos vectores?

Ejercicio 2.14. Sea V el conjunto de vectores $(x_1, x_2, x_3, x_4) \in \mathbb{R}^4$ tales que

$$\begin{aligned} 2x_1 - 3x_2 - x_3 + x_4 &= 0, \\ x_1 - x_2 + 2x_3 - x_4 &= 0. \end{aligned}$$

Demostrar que V es un subespacio de $\mathbb{R}^4$ y encontrar una base para V.

Ejercicio 2.15. Sean $\boldsymbol{x}$, $\boldsymbol{y}$ dos vectores en $\mathbb{R}^3$.

(a) Demostrar que $\boldsymbol{x} \times \boldsymbol{y} = \boldsymbol{0}$ si y solo si el conjunto $\{\boldsymbol{x}, \boldsymbol{y}\}$ es linealmente dependiente.

(b) Comprobar que $\boldsymbol{x} \times \boldsymbol{y}$ es ortogonal al subespacio $\operatorname{lin}\langle \boldsymbol{x}, \boldsymbol{y}\rangle$.

(c) Concluir que $\{\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{x} \times \boldsymbol{y}\}$ es una base de $\mathbb{R}^3$ si y solo si $\boldsymbol{x} \times \boldsymbol{y} \neq \boldsymbol{0}$.

Ejercicio 2.16. Sea $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ un conjunto de n elementos de un espacio vectorial V. Demostrar que $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ es una *base* de V si y solo si cada vector $\boldsymbol{x} \in V$ puede expresarse como una combinación lineal (2.5) de ellos *de manera única*.

En otras palabras: si $\boldsymbol{x} = a_1\boldsymbol{x}_1 + a_2\boldsymbol{x}_2 + \cdots + a_n\boldsymbol{x}_n$ y si $\boldsymbol{y} = b_1\boldsymbol{x}_1 + b_2\boldsymbol{x}_2 + \cdots + b_n\boldsymbol{x}_n$, mostrar que $\boldsymbol{x} = \boldsymbol{y}$ *si y solo si* $a_j = b_j$ para $j = 1, 2, \dots, n$.

Ejercicio 2.17. Sea $\mathbb{R}_n[t]$ el subespacio de $\mathbb{R}[t]$ que consta de polinomios de grado no mayor que n. Demostrar que $\{1, (t-1), (t-1)^2, \dots, (t-1)^n\}$ es una base para $\mathbb{R}_n[t]$.

Ejercicio 2.18. Sean M, N dos subespacios de un espacio vectorial finitodimensional V. Demostrar que su intersección $\underline{M \cap N}$ y su *suma* $\underline{M + N} := \{\, \boldsymbol{x} + \boldsymbol{y} : \boldsymbol{x} \in M,\ \boldsymbol{y} \in N \,\}$ son también subespacios de V. Verificar la fórmula:

$$\dim(M + N) = \dim M + \dim N - \dim(M \cap N).$$

⟦ Indicación: escoger una base para $M \cap N$ y completarla de dos maneras para formar bases de M y de N. Verificar que la unión de estas dos bases es una base para $M + N$. ⟧

Ejercicio 2.19. Sean V y W espacios vectoriales sobre el mismo cuerpo $\mathbb{F}$. Se define su **suma directa** $\underline{V \oplus W}$ como el producto cartesiano de V y W, dotado de las siguientes operaciones de suma y multiplicación escalar:

$$(\boldsymbol{x}, \boldsymbol{y}) + (\boldsymbol{x}', \boldsymbol{y}') := (\boldsymbol{x} + \boldsymbol{x}', \boldsymbol{y} + \boldsymbol{y}'), \qquad c(\boldsymbol{x}, \boldsymbol{y}) := (c\boldsymbol{x}, c\boldsymbol{y}),$$

para $\boldsymbol{x}, \boldsymbol{x}' \in V$; $\boldsymbol{y}, \boldsymbol{y}' \in W$; $c \in \mathbb{F}$. Verificar que estas reglas cumplen las propiedades (a)–(h) de la Definición 2.6. Además, si V y W son finitodimensionales, demostrar que $\dim(V \oplus W) = \dim V + \dim W$.

Ejercicio 2.20. Sea V un espacio vectorial sobre $\mathbb{F}$ y sea $W \leqslant V$ un subespacio. Otro subespacio $U \leqslant V$ se llama un **suplemento** para W si:

$$U \cap W = \{\mathbf{0}\} \quad \text{y} \quad U + W = V.$$

Verificar que $T : U \oplus W \to V$ dado por $T(u, w) := u + w$ es una biyección lineal. Si V es finitodimensional, demostrar que cada subespacio de V posee un suplemento (aunque en general este suplemento no es único). ⟦ Indicación: compleción de bases. ⟧

Ejercicio 2.21. Sean M, N las partes de $\mathbb{R}^4$ dadas por

$$\begin{aligned} M &:= \{\, (3s, s, s - t, t) \in \mathbb{R}^4 : s, t \in \mathbb{R} \,\}, \\ N &:= \{\, (x_1, x_2, x_3, x_4) \in \mathbb{R}^4 : 2x_1 + 4x_4 = \mathbf{0} \,\}. \end{aligned}$$

Demostrar que M y N son subespacios de $\mathbb{R}^4$ y que $\mathbb{R}^4 = M + N$.

Ejercicio 2.22. Sean V, W las partes de $\mathbb{F}^n$ dadas por

$$\begin{aligned} V &:= \{\, \boldsymbol{x} \in \mathbb{F}^n : x_1 + \cdots + x_n = \mathbf{0} \,\}, \\ W &:= \{\, \boldsymbol{x} \in \mathbb{F}^n : x_1 = \cdots = x_n \,\}. \end{aligned}$$

Demostrar que V y W son subespacios de $\mathbb{F}^n$ y que $\mathbb{F}^n = V \oplus W$.

3 Matrices y ecuaciones lineales

Les règles de calcul algébrique de Heisenberg lui ont été imposées par les résultats expérimentaux de la spectroscopie. Mais Heisenberg n'a pas compris tout de suite que l'algèbre avec laquelle il travaillait était déjà connue des mathématiciens et s'appelait l'algèbre des matrices.

— Alain Connes[1]

La práctica del álgebra lineal se reduce al cálculo de matrices. Una *matriz* es un arreglo rectangular de escalares, cuyas filas o columnas pueden considerarse como vectores según los requerimientos del caso. Las matrices aparecen comúnmente como juegos de coeficientes de sistemas de ecuaciones de primer grado. En este capítulo se utiliza el lenguaje de matrices en la resolución de esos sistemas de ecuaciones; se deja para el capítulo siguiente su interpretación como transformaciones entre espacios vectoriales.

3.1. El álgebra de matrices

Definición 3.1. Una **matriz** $\underline{m \times n}$ con entradas en un cuerpo $\mathbb{F}$ es un arreglo rectangular de elementos de $\mathbb{F}$, que contiene m filas y n columnas. El escalar a_{ij} que corresponde a la fila i y a la columna j de una matriz A se llama la **entrada** (i, j) de A. Se escribe

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}.$$

Para abreviar, a veces se escribe $\underline{A = [a_{ij}]}$, dejando implícito que los índices recorren los rangos $i = 1, 2, \dots, m$ y $j = 1, 2, \dots, n$.

La totalidad de matrices $m \times n$ con entradas en $\mathbb{F}$ se escribe $\underline{M_{m\times n}(\mathbb{F})}$ – o algunas veces $\underline{\mathbb{F}^{m\times n}}$. Este es un espacio vectorial sobre $\mathbb{F}$ de dimensión mn, donde la sumas y los múltiplos escalares se definen *entrada por entrada*, es decir:

$$[a_{ij}] + [b_{ij}] = [a_{ij} + b_{ij}], \qquad c[a_{ij}] = [ca_{ij}].$$

El cero de este espacio vectorial es la matriz 0, cuyas entradas son todas 0.

Si $m = n$, se dice que un elemento de $\underline{M_n(\mathbb{F})} := M_{n\times n}(\mathbb{F})$ es una **matriz cuadrada**. ◊

[1]En su libro *Géometrie non commutative*, InterEditions, Paris, 1990.

Definición 3.2. La **transpuesta** $\underline{A^{\mathsf{t}}}$ de una matriz $A \in M_{m\times n}(\mathbb{F})$ es una matriz $n \times m$, cuya entrada (i, j) es la entrada (j, i) de A. En símbolos, $\underline{A^{\mathsf{t}} = [a_{ji}]}$. Más explícitamente,

$$A^{\mathsf{t}} = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{bmatrix}.$$

Una matriz cuadrada $A \in M_n(\mathbb{F})$ se llama **simétrica** si $\underline{A^{\mathsf{t}} = A}$, es decir, si $a_{ji} = a_{ij}$ para todo $i, j = 1, 2, \dots, n$. ◊

Definición 3.3. Una matriz *cuadrada* $A \in M_n(\mathbb{F})$ se llama **triangular inferior** si $a_{ij} = 0$ para $i < j$; otra matriz cuadrada B se llama **triangular superior** si $b_{ij} = 0$ para $i > j$; una tercera matriz cuadrada D se llama **diagonal** si $d_{ij} = 0$ para $i \neq j$. Explícitamente:

$$A = \begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ a_{21} & a_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}, \quad B = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ 0 & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & b_{nn} \end{bmatrix}, \quad D = \begin{bmatrix} d_{11} & 0 & \cdots & 0 \\ 0 & d_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_{nn} \end{bmatrix}. \qquad \Diamond$$

Se consideran las *columnas* de una matriz $A \in M_{m\times n}(\mathbb{F})$ como vectores en $\mathbb{F}^m$ (esto es conveniente para la manipulación de sistemas de ecuaciones lineales). Una columna típica de A es

$$\boldsymbol{a}_j = \begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{bmatrix}.$$

Dicho de otro modo, se identifican las matrices $m \times 1$ con vectores; y en ese sentido se habla de *vectores de columna*.

Entonces se puede escribir una matriz $m \times n$ como "una fila de n columnas":

$$\boxed{A = [\boldsymbol{a}_1\ \boldsymbol{a}_2\ \cdots\ \boldsymbol{a}_n].} \tag{3.1}$$

De igual manera, se puede considerar las *filas* de A como vectores en $\mathbb{F}^n$. Una fila típica de A es

$$\boldsymbol{a}'_i = \begin{bmatrix} a_{i1} & a_{i2} & \cdots & a_{in} \end{bmatrix}.$$

Para no entorpecer demasiado la notación, se usará la notación $(x_1, x_2, \dots, x_m) \in \mathbb{F}^m$ para referirse bien a un *vector de fila* $1 \times m$ o bien a un vector de columna $m \times 1$, según los requerimientos del caso.

► Una operación algebraica importantísima es el *producto de matrices*. El producto de dos matrices A y B es una nueva matriz AB cuya entrada típica es el producto escalar de una fila de A por una columna de B. Para que esto pueda realizarse, es necesario que las filas de A y las columnas de B tengan el mismo número de entradas cada uno. En otras palabras, si $A \in M_{m\times n}(\mathbb{F})$ y $B \in M_{q\times r}(\mathbb{F})$, el producto AB existe solo si $n = q$.

Definición 3.4. Sea A una matriz $m \times n$ y B una matriz $n \times r$, ambos con entradas en el cuerpo $\mathbb{F}$. El **producto de las matrices** A y B es la matriz $m \times r$ denotado $\underline{AB}$, cuya entrada (i, j) es *el producto escalar de la fila i de A con la columna j de B*, para $i = 1, 2, \dots, m$; $j = 1, 2, \dots, r$.

Al escribir $\underline{C = AB}$, se obtiene la fórmula:

$$\boxed{c_{ij} := \boldsymbol{a}'_i \cdot \boldsymbol{b}_j = \sum_{k=1}^{n} a_{ik} b_{kj}.} \tag{3.2}$$

Por ejemplo:

$$\begin{bmatrix} 2 & 0 & -1 & 1 \\ 0 & 3 & 1 & -2 \\ 1 & 0 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 & -3 \\ 2 & 1 \\ 0 & -2 \\ -1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & -2 \\ 8 & -3 \\ -2 & -1 \end{bmatrix}. \qquad \Diamond$$

Proposición 3.5. *Sean A y D matrices $m \times n$, B y E matrices $n \times r$, C una matriz $r \times s$, con entradas en el cuerpo $\mathbb{F}$. Se verifican las siguientes reglas algebraicas:*

(a) $A(BC) = (AB)C$,

(b) $(A + D)B = AB + DB$ *y* $A(B + E) = AB + AE$,

(c) $(AB)^{\mathsf{t}} = B^{\mathsf{t}} A^{\mathsf{t}}$.

Demostración. Ad (a): Se ve que BC es una matriz $n \times s$ y que AB es una matriz $m \times r$, así que los productos $A(BC)$ y $(AB)C$ existen y ambos son matrices $m \times s$. Si $i = 1, 2, \dots, m$; $j = 1, 2, \dots, s$, las entradas (i, j) de estas matrices son:

$$\sum_{k=1}^{n} a_{ik} \biggl(\sum_{l=1}^{r} b_{kl} c_{lj} \biggr) \quad \text{y} \quad \sum_{l=1}^{r} \biggl(\sum_{k=1}^{n} a_{ik} b_{kl} \biggr) c_{lj},$$

respectivamente; y ellas son evidentemente iguales, al reordenar estas sumas finitas.

Ad (b): De igual modo, la reglas de distributividad se reduce a las igualdades:

$$\sum_{k=1}^{n}(a_{ik}+d_{ik})b_{kj} = \sum_{k=1}^{n} a_{ik}b_{kj} + d_{ik}b_{kj},$$
$$\sum_{k=1}^{n} a_{ik}(b_{kj}+e_{kj}) = \sum_{k=1}^{n} a_{ik}b_{kj} + a_{ik}e_{kj}.$$

Ad (c): Como B^{t} es $r \times n$ y A^{t} es $n \times m$, entonces $(AB)^{\mathrm{t}}$ y $B^{\mathrm{t}}A^{\mathrm{t}}$ son matrices $r \times m$; sus entradas (i,j) son iguales, pues por la fórmula (3.2) estas son:

$$\sum_{k=1}^{n} a_{jk}b_{ki} = \sum_{k=1}^{n} b_{ki}a_{jk}. \qquad \square$$

Es fundamental notar que *el producto de matrices no es conmutativo*. Es decir, en general $AB \neq BA$ (aunque puede haber igualdad en casos particulares). Por ejemplo,

$$\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad \text{pero} \quad \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

Esto indica que el álgebra de matrices requiere cierto cuidado en su manejo.

► Ahora considérese las *matrices cuadradas* de un tamaño fijo, $n \times n$ por ejemplo. Si $A, B \in M_n(\mathbb{F})$, su producto AB está en $M_n(\mathbb{F})$ también. Se busca una **matriz identidad** $\underline{1_n \in M_n(\mathbb{F})}$ tal que

$$\underline{A1_n = A = 1_nA} \quad \text{para todo} \quad A \in M_n(\mathbb{F}).$$

Al escribir $\underline{1_n := [\delta_{ij}]}$, esta ecuación se traduce en

$$\sum_{k=1}^{n} a_{ik}\delta_{kj} = a_{ij} = \sum_{k=1}^{n} \delta_{ik}a_{kj}$$

lo cual significa que

$$\delta_{ij} = \begin{cases} 1 & \text{si } i = j, \\ 0 & \text{si } i \neq j. \end{cases} \tag{3.3}$$

Entonces la identidad 1_n es la matriz diagonal con cada entrada diagonal igual a 1:

$$1_n = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}. \tag{3.4}$$

(A veces se abrevia 1_n en el símbolo 1 simplemente, cuando n es fijo. Algunos autores escriben I_n ó I en vez de 1_n.)

La notación δ_{ij} en (3.3), que asigna 1 a dos índices iguales y 0 a dos índices desiguales, fue introducido por Leopold Kronecker (en 1868), y se llama la "delta de Kronecker". Una alternativa, propuesta por Kenneth Iverson en 1962, es $[\![i = j]\!]$. Si $R(x)$ es una relación lógica que depende de un parámetro x, la notación $[\![R(x)]\!]$ denota la siguiente función booleana:

$$[\![R(x)]\!] := \begin{cases} 1, & \text{si } R(x) \text{ es } \textsc{cierto}; \\ 0, & \text{si } R(x) \text{ es } \textsc{falso}. \end{cases}$$

Así, la "función indicatriz" de un conjunto A se escribe como $1_A(x) := [\![x \in A]\!]$, y la función de signo sobre $\mathbb{R}$, que vale 1, 0 ó -1 cuando t es respectivamente positivo, cero o negativo, se escribe así: $\operatorname{signo}(t) := [\![t > 0]\!] - [\![t < 0]\!]$.

Definición 3.6. Una matriz cuadrada $A \in M_n(\mathbb{F})$ se dice **invertible** o **no singular** si hay una matriz $B \in M_n(\mathbb{F})$ tal que $\underline{AB = 1_n = BA}$. Si A es invertible, la matriz B es única y se llama la **matriz inversa** de A, comúnmente denotada $\underline{A^{-1}}$. ◊

La unicidad de la inversa B (si existe) se ve del siguiente modo: si $AB_1 = 1_n = B_2A$, entonces $B_1 = 1_nB_1 = (B_2A)B_1 = B_2(AB_1) = B_21_n = B_2$.

Ejemplo 3.7. En $M_2(\mathbb{F})$ – cualquiera que sea el cuerpo $\mathbb{F}$ – se verifica la igualdad:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = \begin{bmatrix} ad-bc & 0 \\ 0 & ad-bc \end{bmatrix};$$

Si $ad - bc \neq 0$ en $\mathbb{F}$, se obtiene

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \implies \boxed{A^{-1} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.} \tag{3.5}$$

Por el contrario, si fuera $ad - bc = 0$, la matriz $C = \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$ sería tal que $AC = 0_2$; entonces no podría haber una matriz B con $BA = 1_2$, ya que se cumpliría las igualdades $C = 1_2C = (BA)C = B(AC) = B0_2 = 0_2$, de donde $a = b = c = d = 0$ y por ende $A = 0_2$, lo cual implicaría $1_2 = BA = 0_2$, que es falso.[2] Se concluye que *A es invertible si y solo si $ad - bc \neq 0$.* ◊

Como el producto de matrices no es conmutativo, vale la pena notar que

$$\boxed{(AB)^{-1} = B^{-1}A^{-1}.} \tag{3.6}$$

[2]Aquí el símbolo 0_2 (más adelante 0) denota la **matriz nula**, cuyas entradas son todas iguales a 0.

Basta notar que

$$\begin{aligned} (B^{-1}A^{-1})AB &= B^{-1}(A^{-1}A)B = B^{-1}1_nB = B^{-1}B = 1_n \quad \text{y} \\ AB(B^{-1}A^{-1}) &= A(BB^{-1})A^{-1} = A1_nA^{-1} = AA^{-1} = 1_n, \end{aligned}$$

porque el inverso de AB es único cuando existe. La relación (3.6) también dice que si A^{-1} y B^{-1} existen, entonces el inverso de AB también existe, debido al cálculo anterior. Moraleja: *el producto de dos matrices invertibles es también invertible*.

3.2. Sistemas de ecuaciones lineales

Considérese ahora un sistema de ecuaciones lineales en tres variables x_1, x_2, x_3:

$$\begin{aligned} x_1 + 3x_2 - x_3 &= 1 \\ 3x_1 + 4x_2 - 4x_3 &= 7 \\ 3x_1 + 6x_2 + 2x_3 &= -3. \end{aligned} \tag{3.7a}$$

Para resolverlo, se procede sistemáticamente del siguiente modo. Se elimina x_1 de las ecuaciones segunda y tercera, al restar de estas ecuaciones 3 veces la primera:

$$\begin{aligned} x_1 + 3x_2 - x_3 &= 1 \\ -5x_2 - x_3 &= 4 \\ -3x_2 + 5x_3 &= -6. \end{aligned}$$

Esto es posible, ya que el coeficiente de x_1 (en este ejemplo igual a 1) no es cero; este coeficiente se llama el *pivote* de esta transformación del sistema.

Ahora se elimina x_2 de la tercera ecuación, al restar de esta ecuación $\frac{3}{5}$ veces la segunda; aquí el "pivote" es el coeficiente no cero (-5) de x_2:

$$\begin{aligned} x_1 + 3x_2 - x_3 &= 1 \\ -5x_2 - x_3 &= 4 \\ \tfrac{28}{5}x_3 &= -\tfrac{42}{5}. \end{aligned} \tag{3.7b}$$

Ya es posible despejar $x_3 = -\frac{42}{5} \big/ \frac{28}{5} = -\frac{3}{2}$. Al sustituir esto en la segunda ecuación de (3.7b), se obtiene $-5x_2 + \frac{3}{2} = 4$, de donde $x_2 = -\frac{1}{2}$. Al sustituir los valores obtenidos de x_3 y x_2 en la primera ecuación de (3.7b), se obtiene $x_1 - \frac{3}{2} + \frac{3}{2} = 1$, de donde $x_1 = 1$. La solución (única) del sistema (3.7a) es entonces:

$$x_1 = 1, \quad x_2 = -\tfrac{1}{2}, \quad x_3 = -\tfrac{3}{2}.$$

El método de solución consiste de dos fases:

(I) *eliminación* consecutiva de variables repetidas hasta llegar a un sistema "triangular" como (3.7b);

(II) despeje de las variables por *sustitución regresiva*, es decir, sustitución consecutiva en el orden inverso.

El método que combina estos dos ingredientes se denomina **eliminación gaussiana**. (En honor a Gauss, quien fue el primero[3] en usar el método en forma sistemática.)

► Es importante notar que los sistemas (3.7a) y (3.7b) son *equivalentes*, en el sentido de poseer el mismo conjunto de soluciones. Esto se ve al notar que los pasos dados para transformar (3.7a) en (3.7b) no pueden afectar los valores de las incógnitas x_1, x_2, x_3. Para concretar esta afirmación, se observa que hay tres tipos de "operaciones elementales" sobre sistemas de ecuaciones que no cambian sus conjuntos de soluciones.

Definición 3.8. Cualquiera de las siguientes transformaciones de un sistema de ecuaciones recibe el nombre de **operación elemental**:

(a) $\underline{E_i \leftrightarrow E_k}$: intercambiar dos ecuaciones de la lista;

(b) $\underline{E_i \mapsto c\,E_i}$: multiplicar una ecuación por un escalar $c \neq 0$;

(c) $\underline{E_i \mapsto E_i - c\,E_k}$: sustraer de una ecuación un múltiplo de cualquier otra. ◊

Tres aplicaciones de la operación (c) bastan para transformar (3.7a) en (3.7b).

El nexo entre sistemas de ecuaciones lineales y matrices queda evidente cuando se observa que el lado izquierdo de (3.7a) puede considerarse como un producto de matrices. En efecto, se puede escribir (3.7a) en la forma

$$\begin{bmatrix} 1 & 3 & -1 \\ 3 & 4 & -4 \\ 3 & 6 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 7 \\ -3 \end{bmatrix}.$$

Al nombrar esta matriz y estos dos vectores así:

$$A = \begin{bmatrix} 1 & 3 & -1 \\ 3 & 4 & -4 \\ 3 & 6 & 2 \end{bmatrix}, \qquad \boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \qquad \boldsymbol{b} = \begin{bmatrix} 1 \\ 7 \\ -3 \end{bmatrix},$$

la ecuación (3.7a) se escribe brevemente como

$$\boxed{A\,\boldsymbol{x} = \boldsymbol{b}.} \tag{3.8}$$

[3]Desde luego, hubieron precursores más antiguos. Alrededor del segundo siglo a.C., salió un libro chino de autores desconocidos llamado modernamente *Nueve capítulos del arte matemático*, cuyo octavo capítulo *Fang cheng* ("arreglos rectangulares") es una colección de problemas resueltos de ecuaciones lineales, empleando el método de reducción que aquí se describe.

En los procedimientos de resolución de sistemas de ecuaciones lineales, no hace falta escribir los nombres x_1, x_2, x_3 de las variables, ya que cada uno de estos nombres puede sustituirse por *la columna respectiva de la matriz de coeficientes* A. En efecto, otra manera de reinterpretar (3.7a) es la siguiente:

$$x_1 \begin{bmatrix} 1 \\ 3 \\ 3 \end{bmatrix} + x_2 \begin{bmatrix} 3 \\ 4 \\ 6 \end{bmatrix} + x_3 \begin{bmatrix} -1 \\ -4 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 7 \\ -3 \end{bmatrix},$$

o bien, en notación vectorial:

$$x_1 \boldsymbol{a}_1 + x_2 \boldsymbol{a}_2 + x_3 \boldsymbol{a}_3 = \boldsymbol{b}. \tag{3.9}$$

La ecuación (3.9) es simple pero fundamental. Expresa que un sistema de ecuaciones lineales no es más que una descripción de un vector dado $\boldsymbol{b}$ como una *combinación lineal* de otros vectores dados $\boldsymbol{a}_j$, y que los *coeficientes desconocidos* de esa combinación lineal corresponden a la solución – si existe – del sistema de ecuaciones.

En segundo término, pone de manifiesto que *una solución del sistema de ecuaciones existe* si y solo si el vector dado $\boldsymbol{b}$ *pertenece al subespacio generado por las columnas* de la matriz A de coeficientes.

► El sistema (3.7) es un ejemplo de una cuadro más general, simbolizado por la ecuación matricial (3.8), que postulo un juego de m ecuaciones lineales en n variables, con coeficientes en un cuerpo $\mathbb{F}$ cualquiera. Así, con $\boldsymbol{b} \in \mathbb{F}^m$, $\boldsymbol{x} \in \mathbb{F}^n$ y $A \in M_{m\times n}(\mathbb{F})$, se escribe el sistema general de ecuaciones lineales así:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ \cdots\cdots\cdots\cdots &= \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned} \tag{3.10a}$$

lo cual se resume en el formato de (3.8):

$$\boxed{A\,\boldsymbol{x} = \boldsymbol{b},} \tag{3.10b}$$

y su versión vectorial se escribe como sigue:

$$\boxed{x_1 \boldsymbol{a}_1 + x_2 \boldsymbol{a}_2 + \cdots + x_n \boldsymbol{a}_n = \boldsymbol{b}.} \tag{3.10c}$$

Se requiere, entonces, un procedimiento sistemático de resolver este sistema (3.10), toda vez que el sistema admite soluciones – y si es posible, una solución única.

3.3. El algoritmo de eliminación gaussiana

Para resolver sistemas de ecuaciones más ágilmente, conviene adoptar una notación matricial que permite suprimir una mención explícita de los nombres de las incógnitas $\boldsymbol{x} = (x_1, x_2, \dots, x_n)$. Si A es la matriz $m \times n$ de coeficientes de un sistema de m ecuaciones para n incógnitas, y si $\boldsymbol{b}$ es un vector en $\mathbb{R}^m$ que representa el lado derecho del sistema de ecuaciones (3.8), se escribe $\underline{[A \mid \boldsymbol{b}]}$ para denotar la matriz $m \times (n+1)$ cuyas columnas son $\boldsymbol{a}_1, \dots, \boldsymbol{a}_n, \boldsymbol{b}$. Esta matriz es la **matriz aumentada** del sistema de ecuaciones (3.10).

Por ejemplo, la matriz aumentada del sistema (3.7a) es

$$[A \mid \boldsymbol{b}] = \left[\begin{array}{ccc|c} 1 & 3 & -1 & 1 \\ 3 & 4 & -4 & 7 \\ 3 & 6 & 2 & -3 \end{array}\right].$$

Con esta notación, las transformaciones que convierten (3.7a) en (3.7b) se expresan así:

$$\left[\begin{array}{ccc|c} 1 & 3 & -1 & 1 \\ 3 & 4 & -4 & 7 \\ 3 & 6 & 2 & -3 \end{array}\right] \longmapsto \left[\begin{array}{ccc|c} 1 & 3 & -1 & 1 \\ 0 & -5 & -1 & 4 \\ 0 & -3 & 5 & -6 \end{array}\right] \longmapsto \left[\begin{array}{ccc|c} 1 & 3 & -1 & 1 \\ 0 & -5 & -1 & 4 \\ 0 & 0 & \frac{28}{5} & -\frac{42}{5} \end{array}\right]. \tag{3.11}$$

► Hay tres tipos de “operaciones de fila elementales” sobre matrices aumentadas que corresponden a las transformaciones permitidas de sistemas de ecuaciones.

Definición 3.9. Una **operación de fila elemental** sobre una matriz es cualquier una de las siguientes tres operaciones:

(a) $\underline{F_i \leftrightarrow F_k}$: *intercambiar dos filas*;

(b) $\underline{F_i \mapsto c\,F_i}$: *multiplicar una fila por un escalar* $c \neq 0$;

(c) $\underline{F_i \mapsto F_i - c\,F_k}$: *sustraer de una fila un múltiplo de cualquier otra fila*.

Dos matrices aumentadas son **equivalentes por operaciones de fila** si se transforma una en la otra mediante una cadena finita de operaciones de fila elementales. Por su interpretación como coeficientes de sistemas de ecuaciones lineales, resulta entonces que si $[A \mid \boldsymbol{b}]$ es equivalente a $[A' \mid \boldsymbol{b}']$ por operaciones de fila, entonces $A\boldsymbol{x} = \boldsymbol{b}$ si y solo si $A'\boldsymbol{x} = \boldsymbol{b}'$. En otras palabras, los sistemas de ecuaciones asociados tienen las mismas soluciones $\boldsymbol{x} = (x_1, x_2, \dots, x_n) \in \mathbb{F}^n$. ◊

La clave para entender el proceso de solución en términos del álgebra de matrices es el resultado siguiente.

Proposición 3.10. *Cualquier operación de fila elemental sobre una matriz* $B \in M_{m\times n}(\mathbb{F})$ *se efectúa por* ***premultiplicación*** $\underline{B \mapsto AB}$ *por una matriz cuadrada invertible* $A \in M_m(\mathbb{F})$.

Demostración. Ad (a): Para intercambiar las filas F_i y F_k, sea $\underline{P_{ik}} \in M_m(\mathbb{F})$ la matriz con entradas $p_{ik} = p_{ki} = 1$, $p_{jj} = 1$ si $j \neq i$, $j \neq k$ y las demás entradas 0:

$$
P_{ik} := \begin{array}{cc}
\begin{matrix} 1 & & k & & i & & m \end{matrix} & \\
\begin{bmatrix}
1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\
\vdots & \ddots & \vdots & & \vdots & & \vdots \\
0 & \cdots & 0 & \cdots & 1 & \cdots & 0 \\
\vdots & & \vdots & \ddots & \vdots & & \vdots \\
0 & \cdots & 1 & \cdots & 0 & \cdots & 0 \\
\vdots & & \vdots & & \vdots & \ddots & \vdots \\
0 & \cdots & 0 & \cdots & 0 & \cdots & 1
\end{bmatrix} &
\begin{matrix} 1 \\ \\ k \\ \\ i \\ \\ m \end{matrix}
\end{array}
\tag{3.12a}
$$

Ad (b): Para multiplicar la fila F_i por una constante $c \neq 0$, sea $\underline{M_i(c)} \in M_m(\mathbb{F})$ la matriz diagonal con entradas $m_{ii} = c$, $m_{jj} = 1$ si $j \neq i$ y las demás entradas 0:

$$
M_i(c) := \begin{array}{cc}
\begin{matrix} 1 & 2 & & i & & m \end{matrix} & \\
\begin{bmatrix}
1 & 0 & \cdots & 0 & \cdots & 0 \\
0 & 1 & \cdots & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & & \vdots \\
0 & 0 & \cdots & c & \cdots & 0 \\
\vdots & \vdots & & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & 0 & \cdots & 1
\end{bmatrix} &
\begin{matrix} 1 \\ 2 \\ \\ i \\ \\ m \end{matrix}
\end{array}
\tag{3.12b}
$$

Ad (c): Para sustraer de la fila F_i unas c veces la fila F_k, sea $\underline{R_{ik}(c)} \in M_m(\mathbb{F})$ la matriz con entradas $r_{ik} = -c$, $r_{jj} = 1$ para todo j y las demás entradas 0:

$$
R_{ik}(c) := \begin{array}{cc}
\begin{matrix} 1 & & k & & i & & m \end{matrix} & \\
\begin{bmatrix}
1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\
\vdots & \ddots & \vdots & & \vdots & & \vdots \\
0 & \cdots & 1 & \cdots & 0 & \cdots & 0 \\
\vdots & & \vdots & \ddots & \vdots & & \vdots \\
0 & \cdots & -c & \cdots & 1 & \cdots & 0 \\
\vdots & & \vdots & & \vdots & \ddots & \vdots \\
0 & \cdots & 0 & \cdots & 0 & \cdots & 1
\end{bmatrix} &
\begin{matrix} 1 \\ \\ k \\ \\ i \\ \\ m \end{matrix}
\end{array}
\qquad \underline{\text{si } k < i}.
\tag{3.12c}
$$

La operación de fila $F_i \mapsto F_i - c\,F_k$ se ejecuta con el producto $B \mapsto R_{ik}(c)\,B$.

Todas estas matrices son invertibles: por cálculo directo, se obtiene

$$
P_{ik}^{-1} = P_{ik}, \qquad M_i(c)^{-1} = M_i(1/c), \qquad R_{ik}(c)^{-1} = R_{ik}(-c).
\tag{3.13}
$$

así que sus matrices inversas son otras matrices de los mismos tipos. Las operaciones de fila elementales (3.12) *se deshacen con otras operaciones de fila elementales* (3.13). ◻

Ejemplo 3.11. Las operaciones de fila sobre la matriz aumentada del despliegue (3.11) son $F_2 \mapsto F_2 - 3F_1$ y $F_3 \mapsto F_3 - 3F_1$ (para cambiar la primera matriz aumentada en la segunda), seguidas por $F_3 \mapsto F_3 - \frac{3}{5}F_2$. En símbolos:[4]

$$[A \mid \boldsymbol{b}] \mapsto R_{32}(\tfrac{3}{5})\, R_{31}(3)\, R_{21}(3)\, [A \mid \boldsymbol{b}] =: [V \mid \boldsymbol{c}],$$

donde

$$[V \mid \boldsymbol{c}] = \left[\begin{array}{ccc|c} 1 & 3 & -1 & 1 \\ 0 & -5 & -1 & 4 \\ 0 & 0 & \frac{28}{5} & -\frac{42}{5} \end{array}\right]$$

es la matriz aumentada asociada al sistema equivalente (3.7b).

Como la inversión de matrices revierte el orden de productos, según la fórmula (3.6), se obtiene

$$\begin{aligned} [A \mid \boldsymbol{b}] &= \bigl(R_{32}(\tfrac{3}{5})\, R_{31}(3)\, R_{21}(3)\bigr)^{-1} [V \mid \boldsymbol{c}] \\ &= R_{21}(-3)\, R_{31}(-3)\, R_{32}(-\tfrac{3}{5}) [V \mid \boldsymbol{c}] =: L[V \mid \boldsymbol{c}], \end{aligned}$$

donde esta L es la *matriz triangular inferior*:

$$L = R_{21}(-3)\, R_{31}(-3)\, R_{32}(-\tfrac{3}{5}) = \begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 3 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \frac{3}{5} & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 3 & \frac{3}{5} & 1 \end{bmatrix}.$$

Es muy importante notar que (dejando de lado por ahora la últimas columnas $\boldsymbol{b}$ y $\boldsymbol{c}$) el resultado de este cálculo es una *factorización* de la matriz A como un producto matricial $\underline{A = LV}$, donde L y V son matrices triangulares:

$$\begin{bmatrix} 1 & 3 & -1 \\ 3 & 4 & -4 \\ 3 & 6 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 3 & \frac{3}{5} & 1 \end{bmatrix} \begin{bmatrix} 1 & 3 & -1 \\ 0 & -5 & -1 \\ 0 & 0 & \frac{28}{5} \end{bmatrix}. \qquad \Diamond$$

En el Ejemplo 3.11, las entradas de la matriz L debajo la diagonal son justamente los múltiplos c de filas que fueron escogidos en el proceso de eliminación del sistema de ecuaciones original. Este fenómeno se presenta toda vez que se ejecuta una eliminación con operaciones de fila de tipo (c) solamente. Entonces, si se guarda cuenta de estos múltiplos *en su orden consecutivo*, la matriz triangular inferior L *aparece sin mayor necesidad de calcularla* como producto de matrices.

[4]Cuando se aplican varias premultiplicaciones a una matriz B, el orden de los factores se escribe *de derecha a izquierda*, porque tres premultiplicaciones consecutivas por A_1, A_2, A_3, aplicadas a una matriz B, produce: $A_3(A_2(A_1\, B))) = A_3A_2A_1B$.

La segunda matriz V de la factorización LV es la matriz de coeficientes del sistema "triangular" (3.7b) que resulta de la fase de eliminación. En resumen, el proceso de eliminación gaussiana no sólo resuelve el sistema de ecuaciones lineales (3.10), sino que además *proporciona una factorización*, $A = LV$, como producto de una matriz triangular inferior y una matriz triangular superior.

► La matriz L es una matriz triangular de un tipo especial: tiene solamente entradas 1 en la diagonal (la palabra **unipotente** se aplica a este caso). La matriz triangular superior V generalmente no es unipotente. Sin embargo, si ninguna de sus entradas diagonales es 0, se puede multiplicar cada fila F_i de V por el recíproco de su entrada diagonal v_{ii}, premultiplicando por matrices $M_i(1/v_{ii})$ de tipo (3.12b). La matriz V se puede restaurar con las matrices inversas $M_i(v_{ii})$, cuyo producto es una *matriz diagonal* D, con las mismas entradas diagonales: $d_{ii} = v_{ii}$. Por ejemplo:

$$V = \begin{bmatrix} 1 & 3 & -1 \\ 0 & -5 & -1 \\ 0 & 0 & \frac{28}{5} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -5 & 0 \\ 0 & 0 & \frac{28}{5} \end{bmatrix} \begin{bmatrix} 1 & 3 & -1 \\ 0 & 1 & \frac{1}{5} \\ 0 & 0 & 1 \end{bmatrix} =: D\,U,$$

que conduce a la *triple factorización*

$$\boxed{A = LDU}$$

producto de: una matriz triangular inferior unipotente L, una matriz diagonal D y una matriz triangular superior unipotente U.

Para obtener esta "factorización LDU" en casos más generales, solo hace falta tener un *criterio* que garantiza que ningún elemento diagonal de D será 0. La Proposición 3.12, más abajo, explica cuándo eso será factible.

► Volviendo al sistema de ecuaciones lineales (3.10a):

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ \cdots\cdots\cdots\cdots &= \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

el primer paso en su resolución es el de *eliminar la variable* x_1 de la segunda ecuación y las siguientes, al restarles múltiplos apropiados de la primera ecuación. Eso sólo será factible si $a_{11} \neq 0$. Si se nombran las *ecuaciones* $E_1, E_2, \dots, E_m$, se debe sustituir cada ecuación E_i, para $i = 2, \dots, m$, por la combinación $E_i - (a_{i1}/a_{11})\,E_1$, siempre si $a_{11} \neq 0$.

En seguida, se elimina x_2 de la tercera ecuación y las siguientes, al restar de ellas ciertos múltiplos de la segunda; y así sucesivamente. En el k-ésimo paso, será posible continuar si el coeficiente actual de x_k no es 0.

Para simplificar un poco, considérese el caso $m = n$ (igual número de ecuaciones y variables) y supóngase – por ahora – que todos los coeficientes iniciales obtenidos no son ceros. Al final de la eliminación, se obtendrá un **sistema triangular** de ecuaciones:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{22}^{(2)}x_2 + \cdots + a_{2n}^{(2)}x_n &= b_2^{(2)} \\ \cdots\cdots\cdots\cdots &= \vdots \\ a_{nn}^{(n)}x_n &= b_n^{(n)} \end{aligned}$$

tal que las nuevas ecuaciones $E_1, E_2^{(2)}, E_3^{(3)}, \dots, E_n^{(n)}$ tendrán coeficientes iniciales no ceros:

$$a_{11} \neq 0, \quad a_{22}^{(2)} \neq 0, \quad a_{33}^{(3)} \neq 0, \dots, \quad a_{nn}^{(n)} \neq 0.$$

De inmediato, se puede despejar la última variable: $x_n = b_n^{(n)} \big/ a_{nn}^{(n)}$.

La *penúltima* ecuación se resuelve en seguida, al sustituir en ella este valor de x_n, para luego obtener x_{n-1}. Continuando así para atrás, se despejan consecutivamente las variables $x_n, x_{n-1}, \dots, x_2, x_1$ por **sustitución regresiva**. De este modo se obtendrá una *solución única* del sistema original.

Con eso se llega al problema teórico: si en el transcurso de la eliminación, alguno de los coeficientes $a_{kk}^{(k)}$ es 0, el algoritmo se detiene abruptamente, sin éxito. En tal caso, es de esperar que el sistema de ecuaciones lineales (3.10) *no tenga* solución única.

► En términos de la matriz aumentada, sucede lo siguiente.

Proposición 3.12. *Sea A una matriz* cuadrada *$n \times n$ sobre $\mathbb{F}$ y sea $\boldsymbol{b} \in \mathbb{F}^n$. El algoritmo de **eliminación gaussiana simple** transforma la matriz aumentada $[A \mid \boldsymbol{b}]$ por medio de operaciones de fila de tipo* (c) *solamente, al efectuar transformaciones:*

$$A = A^{(1)} \mapsto A^{(2)} \mapsto \cdots \mapsto A^{(n)}.$$

*En el k-ésimo paso, el **pivote** es el elemento diagonal $a_{kk}^{(k)}$ de $A^{(k)}$; para cada $i > k$, se resta $(a_{ik}^{(k)}/a_{kk}^{(k)})$ veces la fila F_k de la fila F_i, resultando $a_{ik}^{(k+1)} = 0$. Este algoritmo produce una solución única para el sistema de ecuaciones $A\boldsymbol{x} = \boldsymbol{b}$ si y solo si* ninguno de los pivotes $a_{kk}^{(k)}$ es cero.

Demostración. Si algún $a_{kk}^{(k)} = 0$, es claro que el algoritmo *se detiene* en el paso número k, ya que para seguir habría que dividir por cero.

Si ningún $a_{kk}^{(k)}$ resulta ser cero, se puede proseguir el algoritmo hasta el n-ésimo paso. Ahora (para efectos de inducción) se puede suponer que las columnas $1, 2, \dots, k-1$ tienen ceros debajo de la diagonal; luego estas columnas no cambian en el paso k, y las filas $k+1, \dots, n$ se transforman de modo que los elementos subdiagonales de la columna k se vuelven ceros:

$$a_{ik}^{(k+1)} := a_{ik}^{(k)} - \frac{a_{ik}^{(k)}}{a_{kk}^{(k)}}\, a_{kk}^{(k)} = 0.$$

Por inducción, después de $(n-1)$ pasos, la matriz $V := A^{(n)}$ es una matriz triangular superior, con elementos diagonales no ceros.

De hecho, los elementos diagonales de V son $a_{11}^{(1)}, a_{22}^{(2)}, \dots, a_{nn}^{(n)}$, ya que cada paso del algoritmo no cambia ni la fila del pivote $a_{kk}^{(k)}$ ni las filas anteriores. Por hipótesis, estos elementos diagonales no son ceros; luego (al dividir cada fila de V por el elemento diagonal correspondiente) se puede factorizar $V = DU$, donde D es diagonal y U es triangular superior unipotente. En efecto, las entradas diagonales de D son los pivotes $d_{kk} = a_{kk}^{(k)}$ y D es invertible.

Sea L^{-1} el producto de las matrices $R_{ik}(a_{ik}^{(k)}/a_{kk}^{(k)})$ de operaciones de fila, multiplicados en orden (de derecha a izquierda); entonces se ve que $L^{-1}A = V$, así que $\underline{A = LV = LDU}$. Por eso, la ecuación $A\boldsymbol{x} = \boldsymbol{b}$ se descompone en tres ecuaciones consecutivas:

$$L\boldsymbol{c} = \boldsymbol{b}, \qquad D\boldsymbol{y} = \boldsymbol{c}, \qquad U\boldsymbol{x} = \boldsymbol{y}.$$

Entonces la ecuación $A\boldsymbol{x} = \boldsymbol{b}$ se resuelve en tres etapas:

- ◇ El algoritmo de eliminación convierte $[A \mid \boldsymbol{b}]$ en $L^{-1}[A \mid \boldsymbol{b}] = [V \mid \boldsymbol{c}]$.
- ◇ Las divisiones de filas por las entradas de D (los pivotes) convierte la matriz $[V \mid \boldsymbol{c}]$ en $D^{-1}[V \mid \boldsymbol{c}] = [U \mid \boldsymbol{y}]$.
- ◇ Y la ecuación $U\boldsymbol{x} = \boldsymbol{y}$ se resuelve de manera única por sustitución regresiva, ya que cada elemento diagonal de U es 1. □

Fíjese bien que los pivotes $a_{kk}^{(k)}$ se calculan durante el algoritmo, y no se dispone (todavía) de un criterio para decidir *de antemano* si alguno de ellos será cero. El algoritmo puede entonces "fallar": es necesario incorporar alguna modificación para prevenir la posibilidad de que aparezca un pivote igual a cero.[5]

► La discusión anterior puede resumirse en un algoritmo de eliminación gaussiana "simple". Este algoritmo es incompleto pues no contempla pivotes nulos.

El algoritmo que sigue, y otros que vendrán posteriormente, están formulados con un *pseudo-código* de tipo genérico. ⟦ No sería difícil traducirlos a un lenguaje estructurado como Pascal ó C. ⟧ Cualquier ⟨ frase entre corchetes angulares ⟩ representa un bloque o

[5] En la práctica, el algoritmo de eliminación simple falla cuando encuentra un pivote muy pequeño, esto es, de valor absoluto inferior al error de redondeo aceptable.

módulo incompleto, cuya definición puede ser incorporado posteriormente. Las { *frases entre llaves* } son comentarios laterales.[6]

Algoritmo 3.1 (Eliminación Gaussiana Simple)**.** Sean $A \in M_n(\mathbb{F})$, $\boldsymbol{b} \in \mathbb{F}^n$. El algoritmo de eliminación gaussiana "simple" transforma la matriz aumentada $[A \mid \boldsymbol{b}]$ por medio de operaciones de fila de tipo (c) solamente, en el supuesto de que eso sea factible; luego, mediante sustitución regresiva, resuelve el sistema de ecuaciones subyacente.

Como entrada, el algoritmo pide una matriz $[\boldsymbol{a}_1 \ \cdots \ \boldsymbol{a}_n]$ y un vector $\boldsymbol{a}_{n+1} \equiv \boldsymbol{b}$, bajo el formato de una *matrizaumentada*, es decir, una matriz $n \times (n+1)$. Como salida, da un vector $\boldsymbol{x} \in \mathbb{F}^n$ que representa la solución del sistema de ecuaciones $A\boldsymbol{x} = \boldsymbol{b}$.

```
⟨ Algoritmo de eliminación gaussiana simple ⟩ ≡
⟨ Declarar los tipos: matrizaumentada, vectorn ⟩  { se pone a_{i,n+1} := b_i }
procedure ElimGauss(A: matrizaumentada;  X: vectorn)
  ⟨ Declarar las variables locales: i, j, k : 1 . . n + 1; C: matrizaumentada ⟩
  ⟨ Declarar otras variables locales ⟩  { (pendiente) }
begin  { ElimGauss }
    for k ← 1 to n − 1 do  { pivotar en la fila k }
    if A[k, k] = 0 then ⟨ Trocar filas ⟩  { (pendiente) }
              else for i ← k + 1 to n do  { limpiar la fila i }
                      A[i, k] ← C[i, k] ← A[i, k]/A[k, k];
                      for j ← k + 1 to n + 1 do  { limpiar la columna j de esa fila }
                          A[i, j] ← A[i, j] − C[i, k] ∗ A[k, j];
                      endfor { j }
              endfor { i }
    endif
  endfor  { k: la eliminación está completa }
  if A[n, n] = 0 then ⟨ Interrumpir proceso ⟩  { la eliminación no es simple }
  else  for i ← n downto 1 do  { sustitución regresiva }
          X[i] ← A[i, n + 1]/A[i, i];  { x_i := b_i/a_ii para empezar }
          for j ← n downto i + 1 do
              X[i] ← X[i] − A[i, j] ∗ X[j]/A[i, i];   { x_i := (b_i − Σ_{j=i+1}^{n} a_ij x_j)/a_ii }
          endfor { j }
        endfor  { i: la sustitución regresiva está completa }
  endif
end;  { ElimGauss }
```

[6] Los algoritmos en estos apuntes son esquemáticos, ya que asumen el uso de aritmética exacta; para la resolución práctica de sistemas de ecuaciones, es preferible usar software profesional que toma en cuenta errores de redondeo, condición de las matrices, estabilidad numérica y otras dificultades del mundo real.

3.4. Eliminación gaussiana con pivoteo parcial

Si en el proceso de eliminación gaussiana, aparece un pivote $a_{kk}^{(k)} = 0$, el algoritmo antes expuesto deja de ser útil. Hay otro recurso para ese caso, el cual es el *intercambio de filas*. Se busca entre las filas $k+1, \dots, n$ para encontrar una fila r con $a_{rk}^{(k)} \neq 0$ (debajo de la entrada nula $a_{kk}^{(k)}$ en la misma columna); luego se intercambian las filas k y r, de modo que el nuevo pivote es el elemento no cero $a_{rk}^{(k)}$. (En la práctica, es más eficiente dar a cada fila una "etiqueta" e intercambiar las etiquetas de las filas k y r que hacer un intercambio directo de los elementos de esas filas; pero aquí se omiten tales detalles de eficiencia de programación.)

⟦ Para cálculos numéricos, es preferible intercambiar la fila k con otra si $a_{kk}^{(k)}$ es muy pequeño, aunque no fuese cero: cuando se encuentran pivotes muy pequeños, se dice que la matriz original A está "mal colocada"; resulta que los errores de redondeo se propagan rápidamente en matrices mal colocadas debido a la división por pivotes pequeños. No es apropiado entrar en detalle aquí; basta recordar que es deseable que los pivotes sean los más grandes posibles (en valor absoluto) y cuando es necesario cambiar uno, se debe elegir el sustituto más grande. Luego, si $a_{kk}^{(k)} \doteq 0$, se busca la fila $r > k$ tal que el valor absoluto $|a_{rk}^{(k)}|$ sea el mayor posible. ⟧

Si $a_{ik}^{(k)} = 0$ para todo $i \geqslant k$, cualquier intercambio de las filas $k, k+1, \dots, n$ dejará un cero en la entrada diagonal (k,k), y el algoritmo falla sin remedio. En ese caso, no será posible determinar el valor de la incógnita x_k, lo cual nos conduce a dos posibilidades: bien el sistema de ecuaciones (3.8) es *inconsistente*, es decir, no posee solución alguna, o bien posee un juego de soluciones en donde el valor de x_k queda "libre". De estas consideraciones, se obtiene la siguiente conclusión.

Escolio 3.13. *Sea A una matriz cuadrada $n \times n$ sobre $\mathbb{F}$ y sea $b \in \mathbb{F}^n$. Para las soluciones del sistema de ecuaciones* (3.8)*, se admiten tres posibilidades:*

(a) *hay una solución única $\boldsymbol{x} \in \mathbb{F}^n$; o bien*

(b) *no hay solución alguna; o bien*

(c) *hay infinitas soluciones en $\mathbb{F}^n$.* ⊟

Ejemplo 3.14. Considérese el sistema de ecuaciones lineales:

$$\begin{aligned} x_2 + x_3 &= 1 \\ x_1 \phantom{{}+x_2} + x_3 &= 1 \\ x_1 + x_2 \phantom{{}+x_3} &= 1 \end{aligned} \qquad \text{cuya matriz aumentada es} \qquad [A \mid \boldsymbol{b}] = \left[\begin{array}{ccc|c} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \end{array}\right].$$

Al inicio, hay una entrada 0 en la posición $(1,1)$ de A. Por lo tanto, se empieza con un intercambio de las filas F_1 y F_2. El desarrollo apropiado es entonces:

$$\left[\begin{array}{ccc|c} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \end{array}\right] \xrightarrow{F_1 \leftrightarrow F_2} \left[\begin{array}{ccc|c} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 \end{array}\right] \xrightarrow[F_3 - F_1]{F_3 :} \left[\begin{array}{ccc|c} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & -1 & 0 \end{array}\right] \xrightarrow[F_3 - F_2]{F_3 :} \left[\begin{array}{ccc|c} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & -2 & -1 \end{array}\right]$$

en donde se destacan los pivotes con un color rojo. Por sustitución regresiva se obtiene $x_3 = \frac{1}{2}$, luego $x_2 = \frac{1}{2}$, y finalmente $x_1 = \frac{1}{2}$. ◊

Ejemplo 3.15. Por otro lado, con el sistema

$$\begin{aligned} x_1 + 3x_2 + 2x_3 &= 4 \\ 2x_1 + 6x_2 + 5x_3 &= 2 \\ 3x_1 + 9x_2 + 8x_3 &= 0 \end{aligned} \quad \text{cuya matriz aumentada es } [A \mid \boldsymbol{b}] = \left[\begin{array}{ccc|c} 1 & 3 & 2 & 4 \\ 2 & 6 & 5 & 2 \\ 3 & 9 & 8 & 0 \end{array}\right], \tag{3.14}$$

se obtiene:

$$\left[\begin{array}{ccc|c} 1 & 3 & 2 & 4 \\ 2 & 6 & 5 & 2 \\ 3 & 9 & 8 & 0 \end{array}\right] \xrightarrow[F_2 - 2F_1]{F_2 :} \left[\begin{array}{ccc|c} 1 & 3 & 2 & 4 \\ 0 & 0 & 1 & -6 \\ 3 & 9 & 8 & 0 \end{array}\right] \xrightarrow[F_3 - 3F_1]{F_3 :} \left[\begin{array}{ccc|c} 1 & 3 & 2 & 4 \\ 0 & 0 & 1 & -6 \\ 0 & 0 & 2 & -12 \end{array}\right] \xrightarrow[F_3 - 2F_2]{F_3 :} \left[\begin{array}{ccc|c} 1 & 3 & 2 & 4 \\ 0 & 0 & 1 & -6 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

En el tercer paso, hay una entrada 0 en la posición $(2,2)$ y otro en la posición $(3,2)$, así que *no hay solución única* al sistema (3.14); el algoritmo de eliminación gaussiana *simple* se detiene, sin éxito. La tercera transformación indica cómo seguir: se ignora la segunda columna y se pasa directamente a la tercera (siempre en la segunda fila), obteniendo un nuevo pivote 1 en la posición $(2,3)$. Al pivotar allí, se obtiene la última matriz aumentada, que corresponde al sistema de ecuaciones:

$$x_1 + 3x_2 + 2x_3 = 4, \qquad x_3 = -6, \qquad 0 = 0,$$

el cual se reduce al par de ecuaciones:

$$x_3 = -6, \qquad x_1 + 3x_2 = 16. \tag{3.15}$$

El proceso de eliminación (y sustitución regresiva) conduce a la conclusión de que el sistema (3.14) *es consistente* y tiene *varias soluciones*. Además, proporciona una descripción más explícita (3.15) de la naturaleza de esas soluciones.

Por otro lado, si la tercera ecuación de (3.14) se modifica a

$$3x_1 + 9x_2 + 8x_3 = 7,$$

por ejemplo, el proceso eliminatoria da lo siguiente.

$$\left[\begin{array}{ccc|c} 1 & 3 & 2 & 4 \\ 2 & 6 & 5 & 2 \\ 3 & 9 & 8 & 7 \end{array}\right] \xrightarrow[F_2-2F_1]{F_2:} \left[\begin{array}{ccc|c} 1 & 3 & 2 & 4 \\ 0 & 0 & 1 & -6 \\ 3 & 9 & 8 & 7 \end{array}\right] \xrightarrow[F_3-3F_1]{F_3:} \left[\begin{array}{ccc|c} 1 & 3 & 2 & 4 \\ 0 & 0 & 1 & -6 \\ 0 & 0 & 2 & -5 \end{array}\right] \xrightarrow[F_3-2F_2]{F_3:} \left[\begin{array}{ccc|c} 1 & 3 & 2 & 4 \\ 0 & 0 & 1 & -6 \\ 0 & 0 & 0 & 7 \end{array}\right],$$

y la última fila corresponde a la ecuación $\underline{0x_3 = 7}$, que es *inconsistente*. La señal de esa inconsistencia es una fila donde todas las entradas son cero menos la última (que corresponde al lado derecho de la ecuación). ◊

A la luz de los ejemplos anteriores, es necesario modificar el Algoritmo 3.1 para incorporar el intercambio de filas en caso de necesidad; esta modificación se llama *pivoteo parcial*.[7] Solo hace falta precisar la definición del bloque ⟨ Trocar filas ⟩ del Algoritmo 3.1 anterior. Se debe agregar otras variables locales j_1, r, s, t al procedimiento 'ElimGauss', modificando así las declaraciones iniciales.

Algoritmo 3.2 (Eliminación Gaussiana con Pivoteo Parcial)**.** Sean $A \in M_n(\mathbb{F})$, $\boldsymbol{b} \in \mathbb{F}^n$. La eliminación gaussiana con pivoteo parcial transforma la matriz aumentada $[A \mid \boldsymbol{b}]$ mediante operaciones de fila de los tipos (c) y (a) y coloca la matriz aumentada en forma triangular. Si se obtiene suficientes pivotes no ceros, se puede proseguir con la fase de sustitución regresiva. En caso contrario, es necesario modificar o desechar la fase de sustitución regresiva del Algoritmo 3.1.

⟨ Declarar otras variables locales ⟩ ≡
  ⟨ Declarar las variables locales: j_1, r, s: $1\,..\,n+1$; t: real ⟩
⟨ Trocar filas ⟩ ≡
⟨ Definir una subrutina Swap(Y, Z: real) que intercambia $Y \leftrightarrow Z$ ⟩ { *(ejercicio)* }
begin { *TrocarFilas* }
    $r \leftarrow k$; $t \leftarrow 0$; { *inicialización* }
    for $s \leftarrow k+1$ **to** n **do** { *buscar la fila s con mayor* $|a_{sk}|$ }
        if abs($A[s,k]$) $> t$ **then** $r \leftarrow s$;
          $t \leftarrow$ abs($A[s,k]$) { $t := \max_s |a_{sk}|$ }
        endif
    endfor { *s: la búsqueda del pivote ha terminado* }
    if $r = k$ **then** ⟨ Abandonar trueque al no encontrar un pivote no cero ⟩
    else for $j_1 \leftarrow 1$ **to** $n+1$ **do**
            Swap($A[k, j_1]$, $A[r, j_1]$);
        endfor { j_1*: las filas k y r han sido intercambiadas* }
    endif
end; { *TrocarFilas* }

[7] Hay una variante más complicada que incorpora también intercambios de columnas, bajo el nombre de *pivoteo completo*: no es relevante para esta discusión.

3.5. Inversión de matrices

Si en el proceso de eliminación gaussiana, aparece un pivote 0 que no se puede recuperar con intercambio de filas, entonces la ecuación $A\boldsymbol{x} = \boldsymbol{b}$ no tiene una única solución. En cambio, si después de algunos pivoteos parciales se logra obtener n pivotes no ceros – el último será $a_{nn}^{(n)}$ – se calcula en seguida una solución única por sustitución regresiva. Ahora bien, una solución formal de $A\boldsymbol{x} = \boldsymbol{b}$ es la fórmula:

$$\boldsymbol{x} = A^{-1}\boldsymbol{b}. \tag{3.16}$$

El lado derecho de (3.16) tiene sentido en la medida en que el lado izquierdo existe y es unívocamente determinado por $\boldsymbol{b}$. Conviene expresar esta conclusión en forma explícita, como sigue.

Escolio 3.16. *Una matriz cuadrada $A \in M_n(\mathbb{F})$ es invertible si y sólo si la ecuación $A\boldsymbol{x} = \boldsymbol{b}$ tiene solución única para cualquier $\boldsymbol{b} \in \mathbb{F}^n$, si y solo si A produce n pivotes no ceros.* ⊟

Fíjese que la existencia de n pivotes no ceros en la eliminación de $[A \mid \boldsymbol{b}]$ no depende de la última columna $\boldsymbol{b}$ de esta matriz aumentada.

► Resulta que se puede modificar el método de eliminación gaussiana para *calcular* la matriz inversa A^{-1}, cuando esta inversa existe. Es cuestión de abandonar la fase de sustitución regresiva. En su lugar, se convierte la matriz triangular $A^{(n)}$ – el resultado de la primera fase – en una *matriz diagonal* al pivotar sucesivamente sobre $a_{nn}^{(n)}, \dots, a_{22}^{(2)}$ (en el orden inverso) para obtener ceros encima de la diagonal. Finalmente se debe dividir cada fila por su entrada diagonal (no cero), obteniendo como resultado la transformación $[A \mid \boldsymbol{b}] \mapsto [1_n \mid \boldsymbol{x}]$. Este es el *método de Gauss y Jordan*, menos eficiente que la sustitución regresiva pero teóricamente equivalente a ella. Como todas las operaciones de fila se obtienen por premultiplicación de A por unas r matrices de tipo (3.12), se obtiene un par de ecuaciones:

$$M_r \cdots M_2 M_1 A = 1_n, \qquad M_r \cdots M_2 M_1 \boldsymbol{b} = \boldsymbol{x}, \tag{3.17}$$

de donde se concluye que

$$A^{-1} = M_r \cdots M_2 M_1.$$

Este esquema se convierte en un algoritmo eficiente para inversión matricial, *al sustituir el vector de columna $\boldsymbol{b}$ en* (3.17) *por la matriz identidad* 1_n. Se generaliza la idea de matriz aumentada para incluir "matrices de dos bloques" $[A \mid B]$ toda vez que A y B tengan el mismo número de filas. Entonces el algoritmo de eliminación gaussiana (con pivoteo parcial) se adapta fácilmente para producir:

$$[A \mid 1_n] \longmapsto [1_n \mid A^{-1}]$$

cuando A^{-1} existe; y para señalar la falta de inverso cuando A^{-1} no existe.

Ejemplo 3.17. Un ejemplo servirá para ilustrar el procedimiento:

$$
\begin{aligned}
[A \mid 1_3] &= \left[\begin{array}{ccc|ccc} 2 & 1 & 1 & 1 & 0 & 0 \\ 4 & 1 & 0 & 0 & 1 & 0 \\ -2 & 2 & 1 & 0 & 0 & 1 \end{array}\right] \longmapsto \left[\begin{array}{ccc|ccc} 2 & 1 & 1 & 1 & 0 & 0 \\ 0 & -1 & -2 & -2 & 1 & 0 \\ 0 & 3 & 2 & 1 & 0 & 1 \end{array}\right] \\
&\longmapsto \left[\begin{array}{ccc|ccc} 2 & 1 & 1 & 1 & 0 & 0 \\ 0 & -1 & -2 & -2 & 1 & 0 \\ 0 & 0 & -4 & -5 & 3 & 1 \end{array}\right] \longmapsto \left[\begin{array}{ccc|ccc} 2 & 1 & 0 & -\frac14 & \frac34 & \frac14 \\ 0 & -1 & 0 & \frac12 & -\frac12 & -\frac12 \\ 0 & 0 & -4 & -5 & 3 & 1 \end{array}\right] \\
&\longmapsto \left[\begin{array}{ccc|ccc} 2 & 0 & 0 & \frac14 & \frac14 & -\frac14 \\ 0 & -1 & 0 & \frac12 & -\frac12 & -\frac12 \\ 0 & 0 & -4 & -5 & 3 & 1 \end{array}\right] \longmapsto \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & \frac18 & \frac18 & -\frac18 \\ 0 & 1 & 0 & -\frac12 & \frac12 & \frac12 \\ 0 & 0 & 1 & \frac54 & -\frac34 & -\frac14 \end{array}\right] = [1_3 \mid A^{-1}].
\end{aligned}
$$

Se concluye que

$$
A^{-1} = \begin{bmatrix} 2 & 1 & 1 \\ 4 & 1 & 0 \\ -2 & 2 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} \frac18 & \frac18 & -\frac18 \\ -\frac12 & \frac12 & \frac12 \\ \frac54 & -\frac34 & -\frac14 \end{bmatrix}. \qquad \Diamond
$$

3.6. Ejercicios sobre matrices y ecuaciones lineales

Ejercicio 3.1. Sean A, B, C las tres matrices:

$$
A := \begin{bmatrix} 3 & -1 \\ 0 & 2 \\ 1 & -1 \end{bmatrix}, \qquad B := \begin{bmatrix} 2 & -1 & 0 & 1 \\ 3 & 0 & 1 & 2 \end{bmatrix}, \qquad C := \begin{bmatrix} 1 & -1 & 0 \\ 2 & 0 & -1 \\ 0 & 2 & 1 \\ -1 & 0 & 2 \end{bmatrix}.
$$

Calcular los productos[8] de matrices AB, BC, $(AB)C$ y $A(BC)$.

Ejercicio 3.2. Verificar que $AB = 0$ pero que $BA \neq 0$ en $M_3(\mathbb{R})$, para:

$$
A := \begin{bmatrix} 1 & -1 & 1 \\ -3 & 2 & -1 \\ -2 & 1 & 0 \end{bmatrix}, \qquad B := \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 1 & 2 & 3 \end{bmatrix}.
$$

Ejercicio 3.3. Sean C, D y E las matrices:

$$
C := \begin{bmatrix} 1 & -3 & 2 \\ 2 & 1 & -3 \\ 4 & -3 & -1 \end{bmatrix}, \qquad D := \begin{bmatrix} 1 & 4 & 1 & 0 \\ 2 & 1 & 1 & 1 \\ 1 & -2 & 1 & 2 \end{bmatrix}, \qquad E := \begin{bmatrix} 2 & 1 & -1 & -2 \\ 3 & -2 & -1 & -1 \\ 2 & -5 & -1 & 0 \end{bmatrix}.
$$

Verificar que $CD = CE$ en $M_{3\times4}(\mathbb{R})$, aunque $D \neq E$.

[8] Los últimos dos productos deben ser iguales.

Ejercicio 3.4. Para esta matriz $A \in M_3(\mathbb{R})$,

$$A := \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -4 & -6 & -4 \end{bmatrix},$$

calcular el polinomio matricial $A^3 + 4A^2 + 6A + 4\,1_3$.

Ejercicio 3.5. Verificar que $A^4 = 1_3$ en $M_3(\mathbb{R})$, donde

$$A := \frac{1}{9}\begin{bmatrix} 4 & 7 & -4 \\ 1 & 4 & 8 \\ 8 & -4 & 1 \end{bmatrix}.$$

Ejercicio 3.6. Calcular (por inducción sobre $n \in \mathbb{N}$) las potencias A^n, B^n, C^n de las siguientes matrices (donde $a \in \mathbb{R}$ es un número real cualquiera):

$$A := \begin{bmatrix} a & 1 \\ 0 & a \end{bmatrix}, \qquad B := \begin{bmatrix} a & 1 & 0 \\ 0 & a & 1 \\ 0 & 0 & a \end{bmatrix}, \qquad C := \begin{bmatrix} 1 & a \\ 0 & 1 \end{bmatrix}.$$

Ejercicio 3.7. Sea $A := \begin{bmatrix} \cos\theta & -\operatorname{sen}\theta \\ \operatorname{sen}\theta & \cos\theta \end{bmatrix}$ en $M_2(\mathbb{R})$. Demostrar que

$$A^2 = \begin{bmatrix} \cos 2\theta & -\operatorname{sen} 2\theta \\ \operatorname{sen} 2\theta & \cos 2\theta \end{bmatrix}, \qquad A^3 = \begin{bmatrix} \cos 3\theta & -\operatorname{sen} 3\theta \\ \operatorname{sen} 3\theta & \cos 3\theta \end{bmatrix}$$

y determinar A^n (por inducción) para todo $n \in \mathbb{N}$.

Ejercicio 3.8. Hallar todas las matrices $A := \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in M_2(\mathbb{R})$ tales que:

$$\text{(a)} \quad A^2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}; \qquad \text{(b)} \quad A^2 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

Ejercicio 3.9. Hallar todas las matrices $B \in M_3(\mathbb{R})$ tales que $AB = BA$, si $A := \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 3 \end{bmatrix}$.

Ejercicio 3.10. Demostrar que $A(x)A(y) = A(x+y)$ para todo $x, y \in \mathbb{R}$, donde

$$A(x) := \begin{bmatrix} 1 & x & \tfrac{1}{2}x^2 \\ 0 & 1 & x \\ 0 & 0 & 1 \end{bmatrix}.$$

Ejercicio 3.11. Calcular la matriz inversa del producto de matrices:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -c & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -b & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ -a & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

⟦ Indicación: usar la fórmula $(ABC)^{-1} = C^{-1}B^{-1}A^{-1}$. ⟧

Ejercicio 3.12. Demostrar que la totalidad de matrices en $M_2(\mathbb{R})$ de la forma $\begin{bmatrix} a & b \\ -b & a \end{bmatrix}$ forman un *cuerpo*. ⟦ Indicación: sólo hace falta probar conmutatividad y la existencia de inversos de elementos no ceros. ⟧

Ejercicio 3.13. Sea $E_{ij} \in M_n(\mathbb{F})$ la matriz $n \times n$ cuya entrada (i, j) es 1 y cuyas demás entradas son 0. Verificar que el conjunto $\{ E_{ij} : i, j = 1, 2, \dots, n \}$ es una *base* del espacio vectorial $M_n(\mathbb{F})$ (sobre el cuerpo $\mathbb{F}$). Concluir que $\dim M_n(\mathbb{F}) = n^2$.

Ejercicio 3.14. Una matriz cuadrada $A \in M_n(\mathbb{F})$ es **antisimétrica** si $A^t = -A$. Comprobar que las matrices antisimétricas forman un subespacio vectorial de $M_n(\mathbb{F})$; encontrar una base de este subespacio. ¿Cuál es su dimensión?

Ejercicio 3.15. Resolver el sistema de ecuaciones:

$$\begin{aligned} x_1 + x_2 + x_3 + x_4 &= 0 \\ x_1 + x_2 + x_3 - x_4 &= 4 \\ x_1 + x_2 - x_3 + x_4 &= -4 \\ x_1 - x_2 + x_3 + x_4 &= 2 \end{aligned}$$

por el método de eliminación gaussiana.

Ejercicio 3.16. Resolver el sistema de ecuaciones:

$$\begin{aligned} 2x_1 - 3x_2 + 2x_3 + 5x_4 &= 3 \\ x_1 - x_2 + x_3 + 2x_4 &= 1 \\ 3x_1 + 2x_2 + 2x_3 + x_4 &= 0 \\ x_1 + x_2 - 3x_3 - x_4 &= 0 \end{aligned}$$

por el método de eliminación gaussiana.

Ejercicio 3.17. La ecuación general de un círculo en el plano $\mathbb{R}^2$ es:

$$x^2 + y^2 + 2Gx + 2Fy + C = 0.$$

Encontrar la ecuación del círculo que pasa por los tres puntos $(1, 1)$, $(2, -1)$ y $(2, 3)$. ⟦ Indicación: obtener un sistema de tres ecuaciones para los coeficientes G, F, C. ⟧

Ejercicio 3.18. Resolver el sistema de ecuaciones:

$$\begin{aligned} x_1 - 2x_2 + x_3 - 4x_4 + x_5 &= 14 \\ 2x_1 + 2x_2 - x_3 + x_4 + 3x_5 &= 0 \\ -x_1 + x_2 + 2x_3 - x_4 + 2x_5 &= 8 \\ 3x_1 - x_2 + 2x_3 + 2x_4 - 2x_5 &= 5 \\ x_1 + 3x_2 + 2x_3 + 2x_4 - 3x_5 &= -7 \end{aligned}$$

por el método de eliminación gaussiana.

Ejercicio 3.19. Resolver el sistema de ecuaciones:

$$\begin{aligned} x_1 + x_2 + x_3 &= 0 \\ 3x_1 + 3x_2 + 4x_3 &= 2 \\ x_1 + 2x_2 + x_3 &= -4 \end{aligned}$$

por el método de eliminación gaussiana con pivoteo parcial.

Ejercicio 3.20. En este sistema de ecuaciones, $k \in \mathbb{R}$ es un parámetro real:

$$\begin{aligned} kx_1 + x_2 + x_3 &= 1 \\ x_1 + kx_2 + x_3 &= 1 \\ x_1 + x_2 + kx_3 &= 1 \end{aligned}$$

Determinar los valores de k para los cuales el sistema: (a) no tiene solución; o (b) tiene una infinitud de soluciones. Encontrar la solución única para los otros valores de k.

Ejercicio 3.21. (a) Resolver el sistema de ecuaciones $A\boldsymbol{x} = \boldsymbol{b}$, donde

$$A := \begin{bmatrix} 2 & 1 & 0 & 0 \\ 1 & 2 & 1 & 0 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & 1 & 2 \end{bmatrix}, \qquad \boldsymbol{b} := \begin{bmatrix} 2 \\ 1 \\ 4 \\ 8 \end{bmatrix},$$

por el método de eliminación gaussiana. Usar los resultados del cálculo para escribir explícitamente las matrices L, D, U de la factorización $A = LDU$.

(b) Si $\boldsymbol{b}' := [2\ 7\ 12\ 11]^t$, usar la factorización de la parte (a) para resolver el sistema $A\boldsymbol{x}' = \boldsymbol{b}'$, con la misma matriz de coeficientes A.

Ejercicio 3.22. Encontrar los inversos de las siguientes matrices:

$$A := \begin{bmatrix} \cos\theta & -\operatorname{sen}\theta \\ \operatorname{sen}\theta & \cos\theta \end{bmatrix}; \qquad B := \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}; \qquad C := \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}.$$

Ejercicio 3.23. Encontrar los inversos de las matrices:

$$A := \begin{bmatrix} 1 & 3 & 5 \\ 3 & 5 & 1 \\ 5 & 3 & 1 \end{bmatrix}; \qquad B := \begin{bmatrix} 1 & 2 & 1 \\ 1 & 1 & 0 \\ 2 & 0 & -1 \end{bmatrix}; \qquad C := \begin{bmatrix} 2 & 0 & 1 & -1 \\ 1 & -1 & 0 & 2 \\ 0 & -1 & 2 & 1 \\ -2 & 1 & 3 & 0 \end{bmatrix}.$$

Ejercicio 3.24. Encontrar el inverso de las matriz:

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

sin usar métodos de eliminación. ⟦ Indicación: si $\boldsymbol{x} = [x_1\ x_2\ x_3\ x_4]^{\mathsf{t}}$ es un vector de columna, calcular $A\boldsymbol{x}$ y obtener por inspección la matriz B tal que $B(A\boldsymbol{x}) = \boldsymbol{x}$. ⟧

Ejercicio 3.25. Si $A \in M_n(\mathbb{F})$ es invertible, demostrar que $(A^{\mathsf{t}})^{-1} = (A^{-1})^{\mathsf{t}}$.
⟦ Esta es la llamada **matriz contragrediente** de A, a veces denotada por $\underline{A^{-\mathsf{t}}}$. ⟧
Concluir que A^{-1} es una matriz simétrica si A es simétrica.

Ejercicio 3.26. Demostrar que

$$\begin{bmatrix} 1 & 0 & 0 \\ -a_{21} & 1 & 0 \\ -a_{31} & 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ a_{21} & 1 & 0 \\ a_{31} & 0 & 1 \end{bmatrix} \quad \text{pero} \quad \begin{bmatrix} 1 & 0 & 0 \\ -a_{21} & 1 & 0 \\ -a_{31} & -a_{32} & 1 \end{bmatrix}^{-1} \neq \begin{bmatrix} 1 & 0 & 0 \\ a_{21} & 1 & 0 \\ a_{31} & a_{32} & 1 \end{bmatrix}.$$

Ejercicio 3.27. Sea A una matriz triangular superior con ceros en la diagonal:

$$A := \begin{bmatrix} 0 & a_{12} & \cdots & a_{1n} \\ 0 & 0 & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}.$$

(Es decir, $a_{ij} = 0$ para $i \geqslant j$.) Demostrar que $A^n = 0$. Concluir que $1_n + A$ es invertible, con

$$(1_n + A)^{-1} = 1_n - A + A^2 - \cdots + (-1)^{n-1}A^{n-1}.$$

Usar esta relación para calcular el inverso de la matriz $\begin{bmatrix} 1 & a & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{bmatrix}$.

Ejercicio 3.28. Si $A \in M_m(\mathbb{F})$ y $D \in M_n(\mathbb{F})$ son invertibles y si $B \in M_{m\times n}(\mathbb{F})$, defínase una matriz cuadrada $(m+n) \times (m+n)$ por

$$M := \begin{bmatrix} A & B \\ 0 & D \end{bmatrix},$$

donde 0 es la matriz nula en $M_{n\times m}(\mathbb{F})$. Mostrar que M^{-1} existe y está dada por:

$$M^{-1} = \begin{bmatrix} A^{-1} & -A^{-1}BD^{-1} \\ 0 & D^{-1} \end{bmatrix}.$$

Ejercicio 3.29. Demostrar que

$$\begin{bmatrix} \cos\theta & -\operatorname{sen}\theta & \cos\phi & -\operatorname{sen}\phi \\ \operatorname{sen}\theta & \cos\theta & \operatorname{sen}\phi & \cos\phi \\ 0 & 0 & \cos\psi & -\operatorname{sen}\psi \\ 0 & 0 & \operatorname{sen}\psi & \cos\psi \end{bmatrix}^{-1} = \begin{bmatrix} \cos\theta & \operatorname{sen}\theta & -\cos(\theta-\phi+\psi) & -\operatorname{sen}(\theta-\phi+\psi) \\ -\operatorname{sen}\theta & \cos\theta & \operatorname{sen}(\theta-\phi+\psi) & -\cos(\theta-\phi+\psi) \\ 0 & 0 & \cos\psi & \operatorname{sen}\psi \\ 0 & 0 & -\operatorname{sen}\psi & \cos\psi \end{bmatrix}$$

⟦ Indicación: usar el Ejercicio 3.28 anterior. ⟧

4 Aplicaciones lineales y matrices

E la storia della Matematica ci avverte che tutti i concetti fondamentali, venuti ad allargare le idee dominanti nei vari campi di essa, sono stati introdotti nella scienza in un modo analogo, trovando solo più tardi la loro piena ed esatta giustificazione.

— Federigo Enriques[1]

El producto de una matriz fija A por un vector de columna cualquiera, $\boldsymbol{x} \mapsto A\boldsymbol{x}$, preserva sumas y múltiplos de vectores: $A(\boldsymbol{x}+\boldsymbol{y}) = A\boldsymbol{x} + A\boldsymbol{y}$, también $A(c\boldsymbol{x}) = c(A\boldsymbol{x})$; se dice que esta operación es *lineal*. La aplicación $\boldsymbol{x} \mapsto A\boldsymbol{x}$ entonces respeta la estructura vectorial del espacio de las columnas. En este capítulo, se adopta este punto de vista sobre las matrices; en lugar de mirarlas como arreglos rectangulares de entradas sujeto a ciertas reglas de combinación, se las considera como aplicaciones que llevan unos vectores en otros de manera lineal.

4.1. Aplicaciones lineales

Definición 4.1. Sean V y W dos espacios vectoriales sobre un cuerpo $\mathbb{F}$. Se dice que una función $\underline{T \colon V \to W}$ es una **aplicación lineal** (más precisamente, una *aplicación $\mathbb{F}$-lineal*, si es necesario identificar el cuerpo $\mathbb{F}$ de escalares) si T tiene estas dos propiedades:

(a) $T(\boldsymbol{x}+\boldsymbol{y}) = T(\boldsymbol{x}) + T(\boldsymbol{y})$, para todo $\boldsymbol{x}, \boldsymbol{y} \in V$;

(b) $T(c\boldsymbol{x}) = c\,T(\boldsymbol{x})$, para todo $\boldsymbol{x} \in V$ y cada $c \in \mathbb{F}$.

La totalidad de las aplicaciones lineales $T \colon V \to W$ se denota por $\underline{\mathcal{L}(V,W)}$.

Si $V = W$, también se dice que T es un **operador lineal** sobre V y se escribe $\underline{\mathcal{L}(V)}$ como sinónimo de $\mathcal{L}(V,V)$. ◊

Ejemplo 4.2. Obsérvese que cualquier aplicación lineal $T \colon V \to W$ lleva el origen de V en el origen de W:

$$T(\mathbf{0}) = T(0\boldsymbol{x}) = 0\,T(\boldsymbol{x}) = \mathbf{0}.$$

Denótese por $\underline{0}$ la *aplicación lineal nula* de V en W definida por $\underline{0(\boldsymbol{x}) := \mathbf{0}}$ para todo $\boldsymbol{x} \in V$. ◊

Ejemplo 4.3. Denótese por $\underline{1_V} \colon V \to V$ el *operador identidad* sobre V definida por $1_V(\boldsymbol{x}) := \boldsymbol{x}$ para todo $\boldsymbol{x} \in V$. (A veces se escribe 1 simplemente, en vez de 1_V.) ◊

[1]En su libro *Lezioni di Geometria proiettiva*, Zanichelli, Bologna, 1904.

Ejemplo 4.4. Dada una matriz $A \in M_{m\times n}(\mathbb{F})$, defínase una aplicación lineal $\underline{T_A \colon \mathbb{F}^n \to \mathbb{F}^m}$ por

$$\boxed{T_A(\boldsymbol{x}) := A\boldsymbol{x}} \quad \text{para todo} \quad \boldsymbol{x} \in \mathbb{F}^n. \tag{4.1}$$

La linealidad de T_A sigue directamente de las dos propiedades $A(\boldsymbol{x}+\boldsymbol{y}) = A\boldsymbol{x} + A\boldsymbol{y}$ y $A(c\boldsymbol{x}) = c(A\boldsymbol{x})$ del producto de matrices y vectores. ◊

Ejemplo 4.5. Sobre el espacio vectorial de polinomios reales $V = \mathbb{R}[t]$, se definen dos operadores lineales de interés. Dado un polinomio $p = p(t)$, se define su *derivada* $Dp(t)$ y su *primitiva* $Tp(t)$ por las fórmulas:

$$\begin{aligned} D(a_0 + a_1 t + \cdots + a_n t^n) &:= a_1 + 2a_2 t + \cdots + n a_n t^{n-1}, \\ T(a_0 + a_1 t + \cdots + a_n t^n) &:= a_0 t + \tfrac{1}{2} a_1 t^2 + \cdots + \tfrac{1}{n+1} a_n t^{n+1}. \end{aligned}$$

Es fácil verificar que $D \colon \mathbb{R}[t] \to \mathbb{R}[t]$ y $T \colon \mathbb{R}[t] \to \mathbb{R}[t]$ son aplicaciones lineales.[2] También es evidente que $D(Tp)(t) = p(t)$ para todo polinomio $p(t)$. ◊

Lema 4.6. *Sean V y W dos espacios vectoriales sobre un cuerpo $\mathbb{F}$. Entonces $\mathcal{L}(V,W)$ es también un espacio vectorial sobre $\mathbb{F}$.*

Demostración. En $\mathcal{L}(V,W)$ se definen las operaciones de suma y multiplicación escalar (de funciones) por:

$$\underline{T+S} : \boldsymbol{x} \mapsto T(\boldsymbol{x}) + S(\boldsymbol{x}), \qquad \underline{cT} : \boldsymbol{x} \mapsto c\,T(\boldsymbol{x}).$$

Si $\boldsymbol{x}, \boldsymbol{y} \in V$ y si $a, b \in \mathbb{F}$, se verifican las propiedades de linealidad:

$$\begin{aligned} (T+S)(a\boldsymbol{x} + b\boldsymbol{y}) &= T(a\boldsymbol{x} + b\boldsymbol{y}) + S(a\boldsymbol{x} + b\boldsymbol{y}) = a\,T(\boldsymbol{x}) + b\,T(yy) + a\,S(\boldsymbol{x}) + b\,S(yy) \\ &= a(T+S)(\boldsymbol{x}) + b(T+S)(\boldsymbol{y}), \\ (cT)(a\boldsymbol{x} + b\boldsymbol{y}) &= c\,T(a\boldsymbol{x} + b\boldsymbol{y}) = ca\,T(\boldsymbol{x}) + cb\,T(\boldsymbol{y}) = a(cT)(\boldsymbol{x}) + b(cT)(\boldsymbol{y}). \end{aligned}$$ □

► La propiedad clave de las aplicaciones lineales es ésta: *una aplicación lineal queda determinada por sus valores sobre los elementos de una base* de su dominio.

Proposición 4.7. *Sean V y W espacios vectoriales sobre $\mathbb{F}$; y sea $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ una base de V. Si $\boldsymbol{z}_1, \dots, \boldsymbol{z}_n$ son n vectores cualesquiera en W (no necesariamente distintos), hay* una única aplicación lineal $T \in \mathcal{L}(V,W)$ *tal que:*

$$T(\boldsymbol{x}_k) = \boldsymbol{z}_k \quad \textit{para} \quad k = 1, 2, \dots, n.$$

[2] En la teoría de integración, una *primitiva* de una función integrable $f(t)$ es otra función $F(t)$ cuya derivada es $f(t)$. No hay unicidad, porque se puede añadir una “constante de integración” $c \in \mathbb{R}$, de modo que $F(t) + c$ es otra primitiva. Pero en el contexto actual la linealidad de T exige que el coeficiente “constante” de $Tp(t)$ es 0: la primitiva $Tp(t)$ es única.

Demostración. Para que T sea lineal, es obligatorio definir $T\colon V \to W$ por

$$T(a_1\boldsymbol{x}_1 + a_2\boldsymbol{x}_2 + \cdots + a_n\boldsymbol{x}_n) := a_1\boldsymbol{z}_1 + a_2\boldsymbol{z}_2 + \cdots + a_n\boldsymbol{z}_n.$$

Como cada vector $\boldsymbol{x} \in V$ puede expresarse *de manera única* como combinación lineal de los vectores $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$ de la base (Ejercicio 2.16), T está bien definida por esta fórmula.

Para ver la unicidad de T, supóngase que $S \in \mathcal{L}(V, W)$ satisface $S(\boldsymbol{x}_k) = \boldsymbol{z}_k$ para cada k. Entonces la linealidad de S muestra que

$$\begin{aligned} S(\boldsymbol{x}) = S(a_1\boldsymbol{x}_1 + a_2\boldsymbol{x}_2 + \cdots + a_n\boldsymbol{x}_n) &= a_1S(\boldsymbol{x}_1) + a_2S(\boldsymbol{x}_2) + \cdots + a_nS(\boldsymbol{x}_n) \\ &= a_1\boldsymbol{z}_1 + a_2\boldsymbol{z}_2 + \cdots + a_n\boldsymbol{z}_n = T(\boldsymbol{x}) \end{aligned}$$

para todo $\boldsymbol{x} \in V$, así que $S = T$. □

Cada matriz $A \in M_{m\times n}(\mathbb{F})$ define una aplicación lineal $T_A\colon \mathbb{F}^n \to \mathbb{F}^m$. Es importante notar que *toda* aplicación lineal $T \in \mathcal{L}(V, W)$ es de esta forma – es decir, que puede hallarse una matriz A tal que $T = T_A$. Sin embargo, para poder comprobar esa afirmación es necesario identificar V con $\mathbb{F}^n$ y W con $\mathbb{F}^m$ de modo explícito. Estas identificaciones se logran al elegir dos bases, una para V y la otra para W. Por lo tanto, *la matriz A que asociada a T no es única*, sino que *depende de las bases elegidas*.

Proposición 4.8. *Sean V y W dos espacios vectoriales sobre $\mathbb{F}$, con $n = \dim V$, $m = \dim W$. Tómese una base $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ de V y otra base $\{\boldsymbol{y}_1, \dots, \boldsymbol{y}_m\}$ de W. Entonces hay una única matriz $A \in M_{m\times n}(\mathbb{F})$ tal que*

$$\left.\begin{aligned} \boldsymbol{x} &= c_1\boldsymbol{x}_1 + \cdots + c_n\boldsymbol{x}_n \\ T(\boldsymbol{x}) &= b_1\boldsymbol{y}_1 + \cdots + b_m\boldsymbol{y}_m \end{aligned}\right\} \implies \underline{\boldsymbol{b} = A\,\boldsymbol{c}}\,.$$

Demostración. Por la Proposición 2.32, se sabe que la aplicación $\boldsymbol{x} \mapsto \boldsymbol{c}$ es un isomorfismo lineal de V en $\mathbb{F}^n$. Defínase la matriz A como sigue: para cada $j = 1, 2, \dots, n$, se puede expandir $T(\boldsymbol{x}_j)$ en términos de la base de W como

$$\boxed{T(\boldsymbol{x}_j) =: \sum_{i=1}^m a_{ij}\,\boldsymbol{y}_i\,.} \tag{4.2}$$

Entonces, por la linealidad de T,

$$\sum_{i=1}^m b_i\boldsymbol{y}_i = T(\boldsymbol{x}) = T\biggl(\sum_{j=1}^n c_j\boldsymbol{x}_j\biggr) = \sum_{j=1}^n c_j\,T(\boldsymbol{x}_j) = \sum_{j=1}^n c_j \sum_{i=1}^m a_{ij}\,\boldsymbol{y}_i = \sum_{i=1}^m \sum_{j=1}^n a_{ij}c_j\,\boldsymbol{y}_i,$$

y la unicidad de los coeficientes con respecto a la base $\{\boldsymbol{y}_1, \dots, \boldsymbol{y}_m\}$ de W implica

$$b_i = \sum_{j=1}^n a_{ij}\,c_j \quad \text{para } i = 1, 2, \dots, m; \ \text{ esto es,} \quad \boldsymbol{b} = A\,\boldsymbol{c}.$$

Por otro lado, al tomar $\boldsymbol{c} = \boldsymbol{e}_j$, un elemento de la base estándar de $\mathbb{F}^n$, $\boldsymbol{b}$ es el vector (en $\mathbb{F}^m$) de coeficientes de $T(\boldsymbol{x}_j)$ respecto de la base $\{\boldsymbol{y}_1, \dots, \boldsymbol{y}_m\}$, y la ecuación $\boldsymbol{b} = A\boldsymbol{e}_j$ dice que $\sum_{i=1}^m a_{ij}\boldsymbol{y}_i = T(\boldsymbol{x}_j)$. De nuevo, la unicidad de coeficientes para la base $\{\boldsymbol{y}_1, \dots, \boldsymbol{y}_m\}$ muestra que las entradas a_{ij} de A están determinadas por T. □

Definición 4.9. Si $T \in \mathcal{L}(V, W)$, y si $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ y $\{\boldsymbol{y}_1, \dots, \boldsymbol{y}_m\}$ son bases de V y W respectivamente, la matriz $A \in M_{m\times n}(\mathbb{F})$ determinado por (4.2) se llama **la matriz de** T *con respecto a estas bases*. ◊

Corolario 4.10. *Vale* $\dim \mathcal{L}(V, W) = (\dim W)(\dim V)$.

Demostración. Es cuestión de observar que $T \mapsto A$ es un isomorfismo lineal entre los espacios vectoriales $\mathcal{L}(V, W)$ y $M_{m\times n}(\mathbb{F})$, ya que $\dim M_{m\times n}(\mathbb{F}) = mn = (\dim W)(\dim V)$. La ecuación (4.2) afirma que $T \mapsto A$ es uno-a-uno, y el Ejemplo 4.4 anterior afirma que es sobreyectivo. La linealidad de esta correspondencia es fácil de comprobar. Como un isomorfismo lineal de espacios vectoriales no es más que una biyección lineal (por la Definición 2.28), se obtiene que $\mathcal{L}(V, W)$ y $M_{m\times n}(\mathbb{F})$ son isomorfos y por ende tienen la misma dimensión. □

Proposición 4.11. *Sean V, W y Z tres espacios vectoriales sobre $\mathbb{F}$, con bases respectivas $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$, $\{\boldsymbol{y}_1, \dots, \boldsymbol{y}_m\}$ y $\{\boldsymbol{z}_1, \dots, \boldsymbol{z}_r\}$. Si $T \in \mathcal{L}(V, W)$, $S \in \mathcal{L}(W, Z)$, su* ***composición*** *es la aplicación lineal $\underline{ST \in \mathcal{L}(V, Z)}$ dada por*

$$\underline{ST}(\boldsymbol{x}) := S(T(\boldsymbol{x})).$$

Si $A \in M_{r\times m}(\mathbb{F})$, $B \in M_{m\times n}(\mathbb{F})$ son las matrices respectivas de S y de T respecto de las bases dadas, entonces la matriz de ST es el producto matricial AB.

Demostración. La linealidad de ST es evidente de:

$$\begin{gathered} ST(\boldsymbol{x} + \boldsymbol{y}) = S(T(\boldsymbol{x} + \boldsymbol{y})) = S(T(\boldsymbol{x}) + T(\boldsymbol{y})) = S(T(\boldsymbol{x})) + S(T(\boldsymbol{y})) = ST(\boldsymbol{x}) + ST(\boldsymbol{y}), \\ ST(c\boldsymbol{x}) = S(T(c\boldsymbol{x})) = S(c\,T(\boldsymbol{x})) = c\,S(T(\boldsymbol{x})) = c\,ST(\boldsymbol{x}), \end{gathered}$$

usando la linealidad de S y de T.

Colóquese $C := AB \in M_{r\times n}(\mathbb{F})$. Para $j = 1, 2, \dots, n$ se calcula con la fórmula (4.2):

$$\begin{aligned} ST(\boldsymbol{x}_j) = S(T(\boldsymbol{x}_j)) &= S\biggl(\sum_{k=1}^m b_{kj}\,\boldsymbol{y}_k\biggr) = \sum_{k=1}^m b_{kj}\,S(\boldsymbol{y}_k) \\ &= \sum_{k=1}^m b_{kj} \sum_{i=1}^r a_{ik}\,\boldsymbol{z}_i = \sum_{i=1}^r \sum_{k=1}^m a_{ik} b_{kj}\,\boldsymbol{z}_i = \sum_{i=1}^r c_{ij}\,\boldsymbol{z}_i, \end{aligned}$$

así que $C = [c_{ij}]$ es la matriz de ST. □

⟦ La composición de funciones se denota usualmente con un círculo: $S \circ T$. Cuando se componen aplicaciones *lineales*, se suele omitir el círculo y escribir simplemente ST. ⟧

Para ilustrar la composición de T y S, se usan los diagramas:

$$\begin{tikzcd} V \arrow[r, "T"] \arrow[rr, bend right, "ST"'] & W \arrow[r, "S"] & Z \end{tikzcd} \qquad \begin{tikzcd} \mathbb{F}^n \arrow[r, "T_B"] \arrow[rr, bend right, "T_A T_B = T_{AB}"'] & \mathbb{F}^m \arrow[r, "T_A"] & \mathbb{F}^r \end{tikzcd}$$

recordando que una composición se lee "de derecha a izquierda".

► Se ve entonces que las reglas algebraicas para matrices (resumidas en la Proposición 3.5) son un simple reflejo de las reglas algebraicas para funciones de unos conjunto en otros. Falta solamente dar la versión abstracta de la transpuesta de una matriz; así, es necesario definir *la transpuesta de una aplicación lineal*, de modo que sea compatible con la regla de cálculo matricial: $(AB)^{\mathsf{t}} = B^{\mathsf{t}}A^{\mathsf{t}}$.

Sin embargo, no es obvio cómo revertir la flecha $V \to W$ sin apelar a cálculos con coeficientes en $\mathbb{F}$. Para ello, se introducen otros dos espacios vectoriales V^* y W^*, que se denombran *espacios duales* de V y W, de modo que la transpuesta de una aplicación lineal $R\colon V \to W$ es una aplicación lineal $\underline{R^{\mathsf{t}}\colon W^* \to V^*}$. El paso de V a V^* es un ejemplo de **dualidad**, un concepto fundamental del álgebra lineal.

Definición 4.12. Sea V un espacio vectorial finitodimensional sobre un cuerpo $\mathbb{F}$. Una **forma lineal** sobre V es una aplicación lineal $f\colon V \to \mathbb{F}$. El **espacio dual** de V es el espacio vectorial V^* de todas estas formas lineales:

$$\underline{V^* := \mathcal{L}(V, \mathbb{F})}\,.$$

Si $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ es una base de V, se define la **base dual** $\underline{\{f_1, \dots, f_n\}}$ de V^* por

$$f_k(c_1\boldsymbol{x}_1 + c_2\boldsymbol{x}_2 + \cdots + c_n\boldsymbol{x}_n) := c_k \tag{4.3a}$$

para $k = 1, 2, \dots, n$. En palabras: f_k selecciona el coeficiente del vector básico $\boldsymbol{x}_k$ el desarrollo (que es único) de un vector de V en términos de esta base. Se verifica fácilmente que f_k es lineal, y como tal queda determinada por sus valores en la base de V:

$$\boxed{f_k(\boldsymbol{x}_j) := \begin{cases} 1, & \text{si } k = j, \\ 0, & \text{si } k \neq j. \end{cases}} \tag{4.3b}$$

Más brevemente, $f_k(\boldsymbol{x}_j) = [\![k = j]\!] \equiv \delta_{kj}$. ◊

En primer lugar, se debe comprobar que esta "base dual" merece su nombre, debido al lema siguiente.

Lema 4.13. *El conjunto de formas lineales* $\{f_1, \dots, f_n\}$ *sobre* V *definido por las fórmulas* (4.3) es una base *del espacio vectorial* V^*.

Demostración. Para ver que $\{f_1, \dots, f_n\}$ es linealmente independiente, supóngase que $a_1 f_1 + \cdots + a_n f_n = 0$ en V^*, para algunos coeficientes $a_1, \dots, a_n \in \mathbb{F}$. El lado izquierdo es una forma lineal; al evaluarla en el vector básico $\boldsymbol{x}_j$ de V, se obtiene

$$(a_1 f_1 + \cdots + a_n f_n)(\boldsymbol{x}_j) = a_1\, f_1(\boldsymbol{x}_j) + \cdots + a_n\, f_n(\boldsymbol{x}_j) = a_j\, f_j(\boldsymbol{x}_j) = a_j,$$

así que $a_j = 0$ para cada $j \in \{1, \dots, n\}$. Esto establece la independencia lineal.

Para ver que $\{f_1, \dots, f_n\}$ genera V^*, sea $g \in V^*$ una forma lineal cualquiera. Sea $b_k := g(\boldsymbol{x}_k)$ para $k = 1, \dots, n$. *Afirmación*: $g = b_1 f_1 + \cdots + b_n f_n$ en V^*. Para comprobar esta fórmula, evaluamos el lado derecho sobre cada vector básico $\boldsymbol{x}_j$ de V. Se ve que

$$(b_1 f_1 + \cdots + b_n f_n)(\boldsymbol{x}_j) = b_1\, f_1(\boldsymbol{x}_j) + \cdots + b_n\, f_n(\boldsymbol{x}_j) = b_j\, f_j(\boldsymbol{x}_j) = b_j = g(\boldsymbol{x}_j).$$

De la Proposición 4.7, que muestra que una forma lineal en $V^* = \mathcal{L}(V, \mathbb{F})$ queda determinada por sus valores en una base de V, se deduce que $b_1 f_1 + \cdots + b_n f_n = g$. Esto confirma que $\{f_1, \dots, f_n\}$ genera V^*. □

Corolario 4.14. *Si la dimensión* n *de* V *sobre* $\mathbb{F}$ *es finita, entonces* $\underline{\dim V^* = n = \dim V}$. ⊟

La matriz de la forma lineal f_k, con respecto a la base $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ de V y de la base $\{1\}$ de espacio unidimensional $\mathbb{F}$, es una matriz $1 \times n$, esto es, un *vector de fila*. Al comparar la fórmula general (4.2) para una matriz con la fórmula particular (4.3b) para f_k, se ve que esta matriz es el vector de fila:

$$\boldsymbol{e}'_k = [0\ 0\ \cdots\ 1\ \cdots\ 0] = (\boldsymbol{e}_k)^{\mathsf{t}}.$$

Definición 4.15. Sean V, W dos espacios vectoriales. Si $R \in \mathcal{L}(V, W)$, su **transpuesta** es la aplicación lineal $\underline{R^{\mathsf{t}} \colon W^* \to V^*}$ dada por

$$\boxed{R^{\mathsf{t}}(g) := g \circ R} \quad \text{para todo} \quad g \in W^*. \tag{4.4}$$

En otras palabras, si $g \in W^*$, $\boldsymbol{x} \in V$, vale $\underline{R^{\mathsf{t}}(g)(\boldsymbol{x}) := g(R(\boldsymbol{x}))}$. ◊

Proposición 4.16. *Sea* $R \in \mathcal{L}(V, W)$*, y sean* $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$, $\{\boldsymbol{y}_1, \dots, \boldsymbol{y}_m\}$ *bases respectivas de* V *y* W*. Sean* $\{f_1, \dots, f_n\}$ *y* $\{g_1, \dots, g_m\}$ *las* bases duales *correspondientes de* V^* *y de* W^**. Si* $C \in M_{m\times n}(\mathbb{F})$ *es la matriz de* R *respecto de las bases* $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ *y* $\{\boldsymbol{y}_1, \dots, \boldsymbol{y}_m\}$*, entonces la matriz de* R^{t} *respecto de las bases duales* $\{g_1, \dots, g_m\}$ *y* $\{f_1, \dots, f_n\}$ *es la* transpuesta matricial $\underline{C^{\mathsf{t}} \in M_{n\times m}(\mathbb{F})}$.

Demostración. Sea D la matriz deseada de R^{t}, determinada por la fórmula

$$R^{\mathsf{t}}(g_i) =: \sum_{k=1}^{n} d_{ki} f_k$$

el cual es un caso particular de la prescripción (4.2). Al aplicar ambos lados de esta ecuación (que son elementos de V^*) al vector básico $\boldsymbol{x}_j \in V$, se obtiene

$$\begin{aligned} d_{ji} &= \sum_{j=1}^{n} d_{ki} \,[\![j = k]\!] = \sum_{k=1}^{n} d_{ki}\, f_k(\boldsymbol{x}_j) \\ &= R^{\mathsf{t}}(g_i)(\boldsymbol{x}_j) = g_i(R(\boldsymbol{x}_j)) = g_i\Bigl(\sum_{r=1}^{m} c_{rj}\, \boldsymbol{y}_r\Bigr) \\ &= \sum_{r=1}^{m} c_{rj}\, g_i(\boldsymbol{y}_r) = \sum_{r=1}^{m} c_{rj} \,[\![i = r]\!] = c_{ij}, \end{aligned}$$

para cada $j = 1, \dots, n$ y cada $i = 1, \dots, m$. Esto muestra que $D = C^{\mathsf{t}}$. □

Obsérvese que la relación $\underline{(SR)^{\mathsf{t}} = R^{\mathsf{t}} S^{\mathsf{t}}}$, para $R \in \mathcal{L}(V, W)$ y $S \in \mathcal{L}(W, Z)$, es una consecuencia inmediata de (4.4), porque si $h \in Z^*$, resulta que

$$R^{\mathsf{t}} S^{\mathsf{t}}(h) = R^{\mathsf{t}}(S^{\mathsf{t}}(h)) = R^{\mathsf{t}}(h \circ S) = h \circ S \circ R = h \circ SR = (SR)^{\mathsf{t}}(h).$$

4.2. Núcleo e imagen de una aplicación lineal

Si $S \colon V \to W$ es una aplicación lineal, hay dos espacios vectoriales naturalmente asociados con S: su *núcleo*, que es un subespacio de V, y su *imagen*, que es un subespacio de W.

Definición 4.17. Sean V, W dos espacios vectoriales sobre $\mathbb{F}$, y sea $S \in \mathcal{L}(V, W)$.

El **núcleo** $\underline{\ker S}$ y la **imagen** $\underline{\operatorname{im} S}$ de S son los subespacios definidos por:

$$\underline{\ker S} := \{\, \boldsymbol{x} \in V : S(\boldsymbol{x}) = \mathbf{0} \,\} \leqslant V; \tag{4.5}$$

$$\underline{\operatorname{im} S} := S(V) = \{\, S(\boldsymbol{x}) : \boldsymbol{x} \in V \,\} \leqslant W. \tag{4.6}$$

La dimensiones de estos subespacios son la **nulidad** $\underline{n(S)}$ y el **rango** $\underline{r(S)}$, definidos respectivamente por:

$$\boxed{n(S) := \dim(\ker S)} \quad \text{y} \quad \boxed{r(S) := \dim(\operatorname{im} S).}$$

Obsérvese que $n(S) \leqslant \dim V$ y que $r(S) \leqslant \dim W$. ◊

Proposición 4.18. *Sea* $S \in \mathcal{L}(V,W)$*; entonces*

(a) S *es* uno-a-uno *si y solo si* $\underline{\ker S = \{\mathbf{0}\}}$*, si y solo si* $\underline{n(S) = 0}$.

(b) S *es* sobreyectivo *si y solo si* $\underline{\operatorname{im} S = W}$*, si y solo si* $\underline{r(S) = \dim W}$.

Demostración. La parte (b) es evidente, de las definiciones de imagen y rango de S.

Para la parte (a), obsérvese que

$$\begin{aligned}
S \text{ es uno-a-uno} \quad &\text{si y solo si} \quad S(\boldsymbol{x}) = S(\boldsymbol{y}) \implies \boldsymbol{x} = \boldsymbol{y} \\
&\text{si y solo si} \quad S(\boldsymbol{x}) - S(\boldsymbol{y}) = \mathbf{0} \implies \boldsymbol{x} - \boldsymbol{y} = \mathbf{0} \\
&\text{si y solo si} \quad S(\boldsymbol{x} - \boldsymbol{y}) = \mathbf{0} \implies \boldsymbol{x} - \boldsymbol{y} = \mathbf{0} \\
&\text{si y solo si} \quad S(\boldsymbol{z}) = \mathbf{0} \implies \boldsymbol{z} = \mathbf{0} \\
&\text{si y solo si} \quad \boldsymbol{z} \in \ker S \implies \boldsymbol{z} = \mathbf{0} \\
&\text{si y solo si} \quad \ker S = \{\mathbf{0}\}.
\end{aligned}$$

□

► La dualidad de espacios vectoriales permite asociar otros dos espacios vectoriales a una aplicación lineal, que son *el núcleo y la imagen de la aplicación transpuesta*. Antes de examinar este segundo par de espacios, es necesario tomar cuenta de una correspondencia natural entre los subespacios de V y los subespacios de V^*. Recuérdese que la notación $\underline{M \leqslant V}$ dice que M es un subespacio de V (Definición 2.13).

Definición 4.19. Si V es un espacio vectorial sobre $\mathbb{F}$ y si V^* es su espacio vectorial dual, sea $M \leqslant V$ un subespacio de V. El **anulador** de M es el subespacio $\underline{M^\perp \leqslant V^*}$ dado por

$$M^\perp := \{\, f \in V^* : f(\boldsymbol{x}) = 0 \text{ para todo } \boldsymbol{x} \in M \,\}. \tag{4.7a}$$

En cambio, si $N \leqslant V^*$ es un subespacio de V^*, el anulador de N es el subespacio $\underline{{}^\perp N \leqslant V}$ dado por

$${}^\perp N := \{\, \boldsymbol{x} \in V : f(\boldsymbol{x}) = 0 \text{ para todo } f \in N \,\}. \tag{4.7b}$$

Obsérvese que $V^\perp = \{0\}$ y que $\{\mathbf{0}\}^\perp = V^*$. ◊

Proposición 4.20. *Sea V un espacio vectorial con* $\dim V = n$*; sean $M \leqslant V$ y $N \leqslant V^*$. Entonces* $\underline{\dim(M^\perp) = n - \dim M}$ *y además* $\underline{\dim({}^\perp N) = n - \dim N}$.

Demostración. Sea $k := \dim M$ y sea $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_k\}$ una base de M. Por la Proposición 2.29, hay vectores $\boldsymbol{x}_{k+1}, \dots, \boldsymbol{x}_n \in V \setminus M$ tales que $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ es una base de V. Sea $\{f_1, \dots, f_n\}$ la base de V^* que es dual a ésta, dada por (4.3).

Si $i > k$ pero $j \leqslant k$, es claro que $f_i(\boldsymbol{x}_j) = 0$, así que $f_i(\boldsymbol{x}) = 0$ para todo $\boldsymbol{x} \in M$, por la linealidad de f_i. Esto dice que $f_i \in M^\perp$ para $i > k$; o sea, que $\operatorname{lin}\langle f_{k+1}, \dots, f_n \rangle \leqslant M^\perp$.

Por otro lado, si $f \in M^\perp$, entonces $f \in V^*$, así que $f = b_1 f_1 + \cdots + b_n f_n$ para algunos coeficientes b_i. Si $j \leqslant k$, se obtiene

$$b_j = \sum_{i=1}^{n} b_i \, [\![i = j]\!] = \sum_{i=1}^{n} b_i \, f_i(\boldsymbol{x}_j) = f(\boldsymbol{x}_j) = 0$$

por ser $\boldsymbol{x}_j \in M$. Entonces $f = b_{k+1} f_{k+1} + \cdots + b_n f_n \in \operatorname{lin}\langle f_{k+1}, \dots, f_n \rangle$. Se ha mostrado que $M^\perp = \operatorname{lin}\langle f_{k+1}, \dots, f_n \rangle$. Se sigue que $\dim(M^\perp) = n - k$.

Un argumento similar permite concluir que $\dim({}^\perp N) = n - \dim N$. Si $r := \dim N$, sea $\{g_1, \dots, g_r\}$ una base de N; hay formas lineales $g_{r+1}, \dots, g_n \in V^*$ tales que $\{g_1, \dots, g_n\}$ es una base de V^*. Hay vectores linealmente independientes $\boldsymbol{y}_1, \dots, \boldsymbol{y}_n \in V$ tales que $\{g_1, \dots, g_n\}$ es la base dual a la base $\{\boldsymbol{y}_1, \dots, \boldsymbol{y}_n\}$ de V – véase el Ejercicio 4.8. Con el razonamiento del párrafo anterior, se deduce que ${}^\perp N = \operatorname{lin}\langle \boldsymbol{y}_{r+1}, \dots, \boldsymbol{y}_n \rangle$, así que $\dim({}^\perp N) = n - r$. □

Corolario 4.21. *Vale $\underline{{}^\perp(M^\perp) = M}$ (subespacios de V); y $\underline{({}^\perp N)^\perp = N}$ (subespacios de V^*).*

Demostración. Las definiciones (4.7) indican que $M \leqslant {}^\perp(M^\perp)$ y $N \leqslant ({}^\perp N)^\perp$. Al contar dimensiones, se obtiene $\dim({}^\perp(M^\perp)) = n - \dim(M^\perp) = n - (n - k) = k = \dim M$ y $\dim(({}^\perp N)^\perp) = n - \dim({}^\perp N) = \dim N$. Por lo tanto, ${}^\perp(M^\perp) = M$ y $({}^\perp N)^\perp = N$. □

Si $S \in \mathcal{L}(V, W)$, se dispone ahora de 8 espacios vectoriales asociados con S: para empezar, su núcleo $\ker S \leqslant V$, su imagen $\operatorname{im} S \leqslant W$, el núcleo de su transpuesta $\underline{\ker S^{\mathsf{t}} \leqslant W^*}$ y la imagen de su transpuesta $\underline{\operatorname{im} S^{\mathsf{t}} \leqslant V^*}$. También, en principio, se cuenta con los respectivos anuladores: $(\ker S)^\perp \leqslant V^*$, $(\operatorname{im} S)^\perp \leqslant W^*$, ${}^\perp(\ker S^{\mathsf{t}}) \leqslant W$ y ${}^\perp(\operatorname{im} S^{\mathsf{t}}) \leqslant V$.

Afortunadamente, de estos 8 sólo hay 4 espacios distintos, por el resultado siguiente.

Proposición 4.22. *Sean V, W dos espacios vectoriales y sea $S \in \mathcal{L}(V, W)$. Entonces:*

$$\boxed{(\operatorname{im} S)^\perp = \ker S^{\mathsf{t}}} \quad y \quad \boxed{(\ker S)^\perp = \operatorname{im} S^{\mathsf{t}}.} \tag{4.8}$$

Además vale $\underline{{}^\perp(\operatorname{im} S^{\mathsf{t}}) = \ker S}$ y $\underline{{}^\perp(\ker S^{\mathsf{t}}) = \operatorname{im} S}$.

Demostración. Sean $g \in W^*$, $\boldsymbol{x} \in V$. Entonces:

$$\begin{aligned}
g \in (\operatorname{im} S)^\perp &\quad \text{si y solo si} \quad g(S(\boldsymbol{x})) = 0 \text{ para todo } \boldsymbol{x} \in V \\
&\quad \text{si y solo si} \quad S^{\mathsf{t}}(g) = g \circ S = 0 \text{ en } V^* \\
&\quad \text{si y solo si} \quad g \in \ker S^{\mathsf{t}}; \\
\boldsymbol{x} \in {}^\perp(\operatorname{im} S^{\mathsf{t}}) &\quad \text{si y solo si} \quad S^{\mathsf{t}}(g)(\boldsymbol{x}) = 0 \text{ para todo } g \in W^* \\
&\quad \text{si y solo si} \quad g(S(\boldsymbol{x})) = 0 \text{ para todo } g \in W^* \\
&\quad \text{si y solo si} \quad S(\boldsymbol{x}) = \mathbf{0} \text{ en } W \\
&\quad \text{si y solo si} \quad \boldsymbol{x} \in \ker S.
\end{aligned}$$

Las otras dos igualdades son consecuencias de estos dos y del Corolario 4.21:

$$(\ker S)^\perp = ({}^\perp(\operatorname{im} S^{\mathsf{t}}))^\perp = \operatorname{im} S^{\mathsf{t}} \quad \text{y} \quad {}^\perp(\ker S^{\mathsf{t}}) = {}^\perp((\operatorname{im} S)^\perp) = \operatorname{im} S. \qquad \square$$

En consecuencia, las dimensiones $n(S)$, $r(S)$, $n(S^{\mathsf{t}})$ y $r(S^{\mathsf{t}})$ obedecen dos relaciones.

Corolario 4.23. *Vale* $\underline{n(S^{\mathsf{t}}) = \dim W - r(S)}$; *y además* $\underline{r(S^{\mathsf{t}}) = \dim V - n(S)}$.

Demostración. Al combinar las Proposiciones 4.20 y 4.22, resultan:

$$\begin{aligned} n(S^{\mathsf{t}}) &= \dim(\ker S^{\mathsf{t}}) = \dim((\operatorname{im} S)^\perp) = \dim W - \dim(\operatorname{im} S) = \dim W - r(S), \\ r(S^{\mathsf{t}}) &= \dim(\operatorname{im} S^{\mathsf{t}}) = \dim((\ker S)^\perp) = \dim V - \dim(\ker S) = \dim V - n(S). \end{aligned} \qquad \square$$

► El siguiente resultado es "el teorema fundamental del álgebra lineal". Establece dos cosas: (a) que los rangos de una aplicación lineal y de su transpuesta son iguales; (b) que el rango y la nulidad de las aplicaciones en $\mathcal{L}(V, W)$ son complementarios, en el sentido de que su suma es siempre la dimensión *del dominio* V.

Teorema 4.24 (Teorema de Rango y Nulidad)**.** *Para cualquier* $S \in \mathcal{L}(V, W)$*, valen:*

$$\text{(a)} \quad \boxed{r(S) = r(S^{\mathsf{t}})} \qquad y \qquad \text{(b)} \quad \boxed{r(S) + n(S) = \dim V.}$$

Demostración. Como $r(S^{\mathsf{t}}) = \dim V - n(S)$ por el Corolario 4.23, las relaciones (a) y (b) son equivalentes. Basta demostrar la relación (b).

Sea $k := n(S)$ y sea $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_k\}$ una base de $\ker S$. Esta se puede completar a una base $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ de V, por la Proposición 2.29. Si $\boldsymbol{x} \in V$, resulta $\boldsymbol{x} = a_1\boldsymbol{x}_1 + \cdots + a_n\boldsymbol{x}_n$ para algunos coeficientes $a_1, \dots, a_n \in \mathbb{F}$. Entonces

$$S(\boldsymbol{x}) = S\biggl(\sum_{j=1}^n a_j\, \boldsymbol{x}_j\biggr) = \sum_{j=1}^n a_j\, S(\boldsymbol{x}_j) = \sum_{j=\underline{k+1}}^n a_j\, S(\boldsymbol{x}_j).$$

Esto dice que $\{S(\boldsymbol{x}_{k+1}), \dots, S(\boldsymbol{x}_n)\}$ *genera* $\underline{\operatorname{im} S} = S(V)$, un subespacio de W.

Por otro lado, si $c_{k+1}\, S(\boldsymbol{x}_{k+1}) + \cdots + c_n\, S(\boldsymbol{x}_n) = \boldsymbol{0}$, así que $S(c_{k+1}\, \boldsymbol{x}_{k+1} + \cdots + c_n\, \boldsymbol{x}_n) = \boldsymbol{0}$, se deduce que $c_{k+1}\, \boldsymbol{x}_{k+1} + \cdots + c_n\, \boldsymbol{x}_n \in \underline{\ker S}$. Por lo tanto, hay escalares $b_1, \dots, b_k \in \mathbb{F}$ tales que

$$c_{k+1}\, \boldsymbol{x}_{k+1} + \cdots + c_n\, \boldsymbol{x}_n = b_1\boldsymbol{x}_1 + \cdots + b_k\boldsymbol{x}_k.$$

La independencia lineal de $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ implica que esto solo es posible si cada $b_i = 0$ y cada $c_j = 0$. Se ha mostrado que $\{S(\boldsymbol{x}_{k+1}), \dots, S(\boldsymbol{x}_n)\}$ es linealmente independiente.

Luego, $\{S(\boldsymbol{x}_{k+1}), \dots, S(\boldsymbol{x}_n)\}$ es una *base* de $\underline{\operatorname{im} S}$; y $r(S) = n - k = \dim V - n(S)$. $\square$

4.3. Forma escalonada de una matriz

Es oportuno volver ahora al contexto de los sistemas de ecuaciones lineales. Tómese una matriz $A \in M_{m\times n}(\mathbb{F})$, y sea $T_A\colon \mathbb{F}^n \to \mathbb{F}^m$ la aplicación lineal definido por $T_A(\boldsymbol{x}) := A\boldsymbol{x}$. Nótese que A es la matriz de T_A, con respecto a las bases estándares de $\mathbb{F}^n$ y $\mathbb{F}^m$. Las ecuaciones (3.8) y (3.9), generalizadas al caso de n variables $\boldsymbol{x} = (x_1, x_2, \dots, x_n)$, asumen el formato:

$$A\boldsymbol{x} = x_1\boldsymbol{a}_1 + x_2\boldsymbol{a}_2 + \cdots + x_n\boldsymbol{a}_n; \tag{4.9a}$$

esto dice que la imagen de T_A es *el subespacio de $\mathbb{F}^m$ generado por las columnas* de la matriz A. En otras palabras,

$$\boxed{\operatorname{im} T_A = \{\, A\boldsymbol{x} : \boldsymbol{x} \in \mathbb{F}^n \,\} = \operatorname{lin}\langle \boldsymbol{a}_1, \dots, \boldsymbol{a}_n \rangle.} \tag{4.9b}$$

En general, las columnas de A no tienen que ser linealmente independientes. De hecho, son independientes si y solo si forman una *base* de $\operatorname{im} T_A$, si y solo si $n = r(T_A)$.

Definición 4.25. Se definen el **rango** y la **nulidad** de una matriz $A \in M_{m\times n}(\mathbb{F})$ como el rango y la nulidad (respectivamente) de la aplicación lineal $T_A\colon \boldsymbol{x} \mapsto A\boldsymbol{x}$. Esto es,

$$\underline{r(A)} := r(T_A) = \dim\{\, A\boldsymbol{x} \in \mathbb{F}^m : \boldsymbol{x} \in \mathbb{F}^n \,\},$$
$$\underline{n(A)} := n(T_A) = \dim\{\, \boldsymbol{x} \in \mathbb{F}^n : A\boldsymbol{x} = \boldsymbol{0} \,\}.$$

Se debe notar que $\underline{r(A) \leqslant m}$ y $\underline{n(A) \leqslant n}$ porque $\operatorname{im} T_A \leqslant \mathbb{F}^m$ y $\ker T_A \leqslant \mathbb{F}^n$. ◊

En vista de (4.9), $r(A)$ es *el número máximo de columnas linealmente independientes* de entre todas las columnas de A. Ahora el Teorema 4.24, junto con la Proposición 4.16, dice que $\underline{r(A^{\mathsf{t}}) = r(A)}$. Como las columnas de A^{t} no son otra cosa que las filas de A escritas verticalmente, el rango de A es también *el número máximo de filas linealmente independientes* de entre todas las filas de A.

Escolio 4.26. *El rango de una matriz $m \times n$ no excede el número de sus columnas ni el número de sus filas:*

$$r(A) \leqslant \min\{m, n\}. \qquad \boxminus$$

► Considérese ahora sistemas de ecuaciones lineales $A\boldsymbol{x} = \boldsymbol{b}$ de tipo (3.8) donde la matriz de coeficientes $A \in M_{m\times n}(\mathbb{F})$ no necesariamente es cuadrada. El lado derecho es un vector $\boldsymbol{b} \in \mathbb{F}^m$, y se busca una solución $\boldsymbol{x} \in \mathbb{F}^n$.

Dícese que el sistema es **homogéneo** si $\boldsymbol{b} = \boldsymbol{0}$.

Un sistema homogéneo $A\boldsymbol{x} = \boldsymbol{0}$ tiene al menos una solución, el vector nulo $\boldsymbol{x} = \boldsymbol{0}$. Decir que $\boldsymbol{x}$ *es una solución* es afirmar que $\underline{\boldsymbol{x} \in \ker T_A}$. En otras palabras, *las soluciones* del sistema homogéneo $A\boldsymbol{x} = \boldsymbol{0}$ forman un *subespacio* vectorial de $\mathbb{F}^n$, *el cual es precisamente el núcleo de T_A*; y su dimensión es $n(A)$. Eso es el contenido del resultado siguiente.

Proposición 4.27. *Si $A \in M_{m\times n}(\mathbb{F})$, el conjunto de soluciones del sistema* homogéneo *de ecuaciones lineales $\underline{A\boldsymbol{x} = \mathbf{0}}$ es un subespacio de $\mathbb{F}^n$, con dimensión $n(A)$. Hay una* ***solución única*** *(que necesariamente es: $\boldsymbol{x} = \mathbf{0}$) si y solo si $\underline{n(A) = 0}$, si y solo si $\underline{r(A) = n}$.*

Demostración. Hay solución única de $A\boldsymbol{x} = \mathbf{0}$ si y solo si $\boldsymbol{x} = \mathbf{0}$ es la única solución, si y solo si $\ker T_A = \{\mathbf{0}\}$, si y solo si $n(A) = 0$; por el Teorema 4.24 (de rango y nulidad), esto es equivalente a que $r(A) = \dim \mathbb{F}^n = n$. □

Obsérvese que $r(A) \leqslant n$ y además $r(A) = r(A^{\mathsf{t}}) \leqslant m$; luego, si $m < n$, el sistema homogéneo $A\boldsymbol{x} = \mathbf{0}$ tiene infinitas soluciones; más precisamente, *tiene un subespacio de soluciones* cuya dimensión es $\geqslant (n - m)$.

► Las sistemas rectangulares de ecuaciones se resuelven por una variante de la eliminación gaussiana, cuyo objetivo es el de reducir la matriz aumentada $[A \mid \boldsymbol{b}]$ a una forma especial en donde se percibe más fácilmente las dimensiones del espacio de columnas y del espacio de soluciones. Se trata de una forma más general del procedimiento de Gauss y Jordan: este proceso transforma una matriz rectangular $A \in M_{m\times n}(\mathbb{F})$ – o eventualmente $[A \mid \boldsymbol{b}] \in M_{m\times(n+1)}(\mathbb{F})$ – por operaciones de fila solamente, en un formato especial, llamada si *forma escalonada*.

Definición 4.28. Una matriz $A \in M_{m\times n}(\mathbb{F})$ está en **forma escalonada** si satisface los tres requisitos que siguen.[3]

(a) Ciertas columnas, llamadas **columnas básicas** y numeradas $j_1, j_2, \dots, j_k$, son precisamente los vectores $\boldsymbol{e}_1, \dots, \boldsymbol{e}_k$ de la base estándar de $\mathbb{F}^m$.

(b) Las columnas básicas aparecen *en el orden usual*: $j_1 < j_2 < \cdots < j_k$.

(c) – Si $j < j_1$, es $\boldsymbol{a}_j = \mathbf{0}$;
– si $j_r \leqslant j < j_{r+1}$, los últimos $(m - r)$ elementos de $\boldsymbol{a}_j$ son ceros; y
– si $j \geqslant j_k$, los últimos $(m - k)$ elementos de $\boldsymbol{a}_j$ son ceros. ◊

Ejemplo 4.29. La siguiente matriz está en forma escalonada:

$$A = \begin{bmatrix} 0 & 1 & 2 & 0 & 1 & 7 \\ 0 & 0 & 0 & 1 & 5 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Aquí $m = 3$, $n = 6$, $k = 2$, con $j_1 = 2$ y $j_2 = 4$; las columnas $\boldsymbol{a}_2 = \boldsymbol{e}_1$ y $\boldsymbol{a}_4 = \boldsymbol{e}_2$ son básicas. La última fila consta de ceros. De hecho, es una consecuencia de la Definición 4.28 que *las últimas $(m - k)$ filas son filas de ceros*, si $k < m$. ◊

[3] Algunos libros, como el de Horn y Johnson o el de los Katznelson, usan el término más preciso: **forma escalonada reducida por filas**, con una definición equivalente a esta Definición 4.28.

Proposición 4.30. *Si $A \in M_{m\times n}(\mathbb{F})$ es una matriz en forma escalonada con k columnas básicas, entonces $\underline{\operatorname{im} T_A = \operatorname{lin}\langle \boldsymbol{e}_1, \dots, \boldsymbol{e}_k\rangle}$ y por ende $\underline{r(A) = k}$.*

Demostración. Como las últimas filas $\boldsymbol{a}'_{k+1}, \dots, \boldsymbol{a}'_m$ son vectores nulos, es claro que todas las *columnas* de A están en el subespacio $\operatorname{lin}\langle \boldsymbol{e}_1, \dots, \boldsymbol{e}_k\rangle \leqslant \mathbb{F}^m$. Pero además las columnas básicas son $\{\boldsymbol{a}_{j_1}, \dots, \boldsymbol{a}_{j_k}\} = \{\boldsymbol{e}_1, \dots, \boldsymbol{e}_k\}$, que son linealmente independientes por ser parte de la base estándar de $\mathbb{F}^m$. Entonces este conjunto es una base del espacio de columnas $\operatorname{lin}\langle \boldsymbol{a}_1, \dots, \boldsymbol{a}_n\rangle$. Al recordar, según (4.9), que $\operatorname{lin}\langle \boldsymbol{a}_1, \dots, \boldsymbol{a}_n\rangle = \operatorname{im} T_A$, se deduce que

$$\operatorname{im} T_A = \operatorname{lin}\langle \boldsymbol{a}_1, \dots, \boldsymbol{a}_n\rangle = \operatorname{lin}\langle \boldsymbol{a}_{j_1}, \dots, \boldsymbol{a}_{j_k}\rangle = \operatorname{lin}\langle \boldsymbol{e}_1, \dots, \boldsymbol{e}_k\rangle.$$

De ahí, es evidente que $r(A) = \dim(\operatorname{im} T_A) = k$. □

► Cabe recordar (Definición 3.9) que dos matrices A y B del mismo tamaño $m \times n$ son *equivalentes por operaciones de fila* si se transforma A en B (o bien B en A) mediante una o más operaciones de fila elementales.[4]

Proposición 4.31. *Si $A, B \in M_{m\times n}(\mathbb{F})$ son dos matrices equivalentes por operaciones de fila, entonces $r(A) = r(B)$.*

Demostración. Esta equivalencia tiene lugar si y solo si hay matrices invertibles $L_1, \dots, L_p$ en $M_m(\mathbb{F})$, donde cada L_i es de uno de los tres tipos (3.12), tales que $A = L_p \cdots L_2 L_1 B$. El producto de matrices $\underline{L := L_p \cdots L_2 L_1}$ en $M_m(\mathbb{F})$ tiene inverso $L^{-1} = L_1^{-1} L_2^{-1} \cdots L_p^{-1}$. Nótese que $A = LB$ y que $B = L^{-1}A$. Entonces

$$\operatorname{im} T_A = \{\, A\boldsymbol{x} : \boldsymbol{x} \in \mathbb{F}^n \,\} = \{\, LB\boldsymbol{x} : \boldsymbol{x} \in \mathbb{F}^n \,\} = T_L\big(\{\, B\boldsymbol{x} : \boldsymbol{x} \in \mathbb{F}^n \,\}\big) = T_L(\operatorname{im} T_B),$$

y de igual modo $\operatorname{im} T_B = T_{L^{-1}}(\operatorname{im} T_A)$. Ahora $T_L : \boldsymbol{x} \mapsto L\boldsymbol{x}$ es una *biyección lineal*, con inverso $T_{L^{-1}} : \boldsymbol{y} \mapsto L^{-1}\boldsymbol{y}$, así que los espacios vectoriales $\underline{\operatorname{im} T_A}$ y $\underline{\operatorname{im} T_B}$ son isomorfos. Por lo tanto, sus dimensiones son iguales: $r(A) = r(B)$. □

Es más o menos evidente que cualquier operación de fila (exceptuando la identidad) cambia una matriz en forma escalonada en otra matriz *que ya no es* de forma escalonada. Luego cualquier matriz $A \in M_{m\times n}(\mathbb{F})$ es equivalente por operaciones de fila a no más de una matriz en forma escalonada. Por otra parte, el procedimiento de Gauss y Jordan *proporciona un algoritmo* para convertir A, mediante operaciones de fila, en una matriz en forma escalonada. ⟦ Dicho de otra manera: las operaciones de fila reparten $M_{m\times n}(\mathbb{F})$ en clases de equivalencia; y cada clase contiene *exactamente una* forma escalonada. ⟧ (Obsérvese que una matriz $n \times n$ es invertible si y solo si es equivalente a la matriz identidad 1_n.)

[4] En el contexto de la Definición 3.9, se consideraron matrices aumentadas; pero la definición es aplicable a cualquier par de matrices del mismo tamaño. Alternativamente, se puede notar que cualquier matriz es una matriz aumentada, si se destaca su última columna como una columna especial.

Algoritmo 4.1 (Conversión a Forma Escalonada)**.** Sea $A \in M_{m\times n}(\mathbb{F})$. El algoritmo de conversión a forma escalonada transforma la matriz A por medio de operaciones de fila.

$\langle$ Algoritmo de conversión a forma escalonada $\rangle \equiv$
$\langle$ Declarar el tipo: matriz = **array**$[m_{\text{máx}}, n_{\text{máx}}]$ of **real** $\rangle$
procedure Echelon(A: matriz)
  $\langle$ Declarar las variables locales: $i, h, r : 1 \,..\, m_{\text{máx}}$;s $j, k : 1 \,..\, n_{\text{máx}}$ $\rangle$
  $\langle$ Definir una subrutina Swap(Y, Z: real) que intercambia Y con Z $\rangle$
begin { *Echelon* }
    for $i = 1$ **to** m **do** { *pivotear en la fila i* }
      $h \leftarrow i$; { *inicialización* }
      for $j = 1$ **to** n **do** { *limpiar las columnas de* 1 *a* j }
        if $A[i, j] = 0$ **repeat** $h \leftarrow h + 1$
                until $(A[h, j] \neq 0)$ **or** $(h = m)$;
                if $(A[h, j] \neq 0)$ **for** $k \leftarrow j$ **to** n **do**
                          Swap$(A[i, k], A[h, k])$;
                      endfor { k }
                endif { *las filas i, h han sido intercambiadas* }
        endif
        if $A[i, j] \neq 0$ **for** $k \leftarrow j$ **to** n **do**
                  $A[i, k] \leftarrow A[i, k]/A[i, j]$;
                endfor; { *k: se ha dividido la fila i por* a_{ij} }
                for $k \leftarrow j$ **to** n **do**
                  for $r \leftarrow 1$ **to** $i - 1$ **do**
                    $A[r, k] \leftarrow A[r, k] - A[r, j] * A[i, k]$;
                  endfor; { r }
                  for $r \leftarrow i + 1$ **to** m **do**
                    $A[r, k] \leftarrow A[r, k] - A[r, j] * A[i, k]$;
                  endfor; { r }
                endfor { *k: el pivoteo en* a_{ij} *está completo* }
        endif
      endfor { *j: si* $a_{ij} = \cdots = a_{mj} = 0$*, se avanza una columna* }
          { *si no, la columna j ya es básica* }
    endfor { *i: la matriz A ya está en forma escalonada* }
end; { *Echelon* }

La conversión de matrices a forma escalonada ahora permite hacer un análisis completo de las posibilidades de resolver sistemas de ecuaciones lineales en general. Ya es posible abordar sistemas $A\boldsymbol{x} = \boldsymbol{b}$ con m ecuaciones y n variables, aun cuando $m \neq n$.

Proposición 4.32. *Si $A \in M_{m\times n}(\mathbb{F})$, $\boldsymbol{b} \in \mathbb{F}^m$, entonces el sistema de ecuaciones lineales $A\boldsymbol{x} = \boldsymbol{b}$ posee* al menos una solución $\boldsymbol{x} \in \mathbb{F}^n$ *si y solo si $r([A \mid \boldsymbol{b}]) = r(A)$.*

Esta solución es única *si y solo si $r([A \mid \boldsymbol{b}]) = r(A) = n$.*

En cambio, si $r([A \mid \boldsymbol{b}]) = r(A) < n$, el conjunto de soluciones $\{\, \boldsymbol{x} \in \mathbb{F}^n : A\boldsymbol{x} = \boldsymbol{b} \,\}$ *toma la forma $\{\, \boldsymbol{x}_0 + \boldsymbol{z} : \boldsymbol{z} \in \ker T_A \,\}$ donde $\boldsymbol{x}_0$ es alguna* solución particular *de $A\boldsymbol{x} = \boldsymbol{b}$; y* $\ker T_A \leqslant \mathbb{F}^n$ *es un subespacio de dimensión $(n - k)$.*

Demostración. Aplíquese el Algoritmo 4.1 de conversión a forma escalonada a la matriz aumentada $[A \mid \boldsymbol{b}]$; denótese por $[A' \mid \boldsymbol{b}']$ la matriz en forma escalonada resultante. Como la conversión se efectúa por premultiplicación por una matriz invertible $M \in M_m(\mathbb{F})$ que es un producto de matrices de los tipos (3.12), es claro que:

$$A\boldsymbol{x} = \boldsymbol{b} \iff MA\boldsymbol{x} = M\boldsymbol{b} \iff A'\boldsymbol{x} = \boldsymbol{b}'.$$

Fíjese que la matriz A' es un bloque $m\times n$ de la matriz aumentada $[A' \mid \boldsymbol{b}']$, y que está en forma escalonada (ya que el Algoritmo 4.1 se aplica de igual manera a la matriz A, al omitir la última columna $\boldsymbol{b}$ de la matriz aumentada). Denótese los rangos por $k := r(A)$ y $l := r([A \mid \boldsymbol{b}])$. Entonces, por la Proposición 4.31, la matriz A' tiene k filas no nulas y la matriz aumentada $[A' \mid \boldsymbol{b}']$ tiene l filas no nulas. Se ve que $l = k$ o bien $l = k + 1$. Quedan tres posibilidades por considerar.

- Si $l = k + 1$, entonces la fila número $k + 1$ de $[A' \mid \boldsymbol{b}']$ es de la forma $[\mathbf{0} \mid 1]$. En efecto, durante la aplicación del algoritmo el cajón izquierdo se reduce a A', ya en forma escalonada, mientras en la última columna habrá una entrada no cero b'_r con $r > k$. Por un intercambio de filas $F_{k+1} \leftrightarrow F_r$ (si fuera necesario), se puede asegurar que $b'_{k+1} \neq 0$; luego, con una operación $F_{k+1} \mapsto (1/b'_{k+1})\, F_{k+1}$, se puede obtener $b'_{k+1} = 1$; y finalmente, al pivotear en esa entrada se convierte la última columna en una columna básica: $\boldsymbol{b}' = \boldsymbol{e}_{k+1}$, sin afectar el cajón izquierdo A'.

  Ahora bien, la fila $[\mathbf{0} \mid 1]$ representa *la ecuación inconsistente* $\mathbf{0} \cdot \boldsymbol{x} = 1$, y de ahí se ve que el sistema $A'\boldsymbol{x} = \boldsymbol{b}'$ *no posee solución alguna*.

- Si $k = l = n$, entonces $n = r(A) = r(A^{\mathsf{t}}) \leqslant m$: el número n de las variables x_i es menor o igual al número m de las ecuaciones. Todas las n columnas de la matriz A' son básicas y aparecen en su orden usual, y la matriz aumentada tiene el formato:

  $$[A' \mid \boldsymbol{b}'] = \left[\begin{array}{c|c} 1_n & \boldsymbol{c} \\ 0 & \mathbf{0} \end{array}\right],$$

  donde $\boldsymbol{b}' = [c_1 \ \cdots \ c_n \ 0 \ \cdots \ 0]^{\mathsf{t}}$. Las últimas $(m - n)$ filas nulas representan ecuaciones tautológicas $\mathbf{0} \cdot \boldsymbol{x} = 0$, y las primeras n filas representan las ecuaciones desacopladas $x_i = c_i$, para $i = 1, 2, \ldots, n$. Es evidente que esta solución es *única*, y además está dada por las primeras n entradas de la última columna $\boldsymbol{b}'$ del resultante del Algoritmo 4.1.

- Si $l = k < n$, se puede hacer un *reordenamiento de columnas* de A', de tal manera que las columnas básicas pasan a la izquierda (en su orden usual). Esto es equivalente a *permutar el orden de las incógnitas* $x_1, \dots, x_n$, sin afectar la validez del sistema de ecuaciones. ⟦ Se debe notar que esta "operación de columnas" no forma parte del Algoritmo 4.1 que emplea operaciones de filas solamente, y se aplica a la matriz aumentada posteriormente. ⟧ La forma final de la matriz aumentada tiene el formato:
$$[A' \mid \boldsymbol{b}'] \longmapsto \left[\begin{array}{cc|c} 1_k & B & \boldsymbol{c} \\ 0 & 0 & \boldsymbol{0} \end{array}\right],$$
donde $\boldsymbol{b}' = [c_1 \;\cdots\; c_k \; 0 \;\cdots\; 0]^{\mathsf{t}}$ y B es alguna matriz $k \times (n-k)$. Las $(m-k)$ filas finales representan tautologías $\boldsymbol{0} \cdot \boldsymbol{x} = 0$, y las primeras k filas representan las ecuaciones:
$$x_i = c_i - b_{i,k+1}x_{k+1} - \cdots - b_{in}x_n \quad \text{para} \quad i = 1, 2, \dots, k.$$
Los valores de las "últimas" variables (en el nuevo orden) $x_{k+1}, \dots, x_n$ pueden asignarse arbitrariamente y entonces las "primeras" variables $x_1, \dots, x_k$ quedan determinadas. Ahora $\boldsymbol{x}_0 := [c_1 \;\cdots\; c_k \; 0 \;\cdots\; 0]^{\mathsf{t}} \in \mathbb{F}^n$ es la solución particular obtenida al colocar $x_{k+1} = \cdots = x_n := 0$, mientras la solución general es $\boldsymbol{x} = \boldsymbol{x}_0 + \boldsymbol{z}$, donde $\boldsymbol{z} \in \operatorname{lin}\langle \boldsymbol{e}_{k+1}, \dots, \boldsymbol{e}_n \rangle \leqslant \mathbb{F}^n$.

En el tercer caso, si $\boldsymbol{x}$ es una solución cualquiera de la ecuación $A\boldsymbol{x} = \boldsymbol{b}$, el vector $\boldsymbol{z} := \boldsymbol{x} - \boldsymbol{x}_0$ cumple
$$A\boldsymbol{z} = A\boldsymbol{x} - A\boldsymbol{x}_0 = \boldsymbol{b} - \boldsymbol{b} = \boldsymbol{0},$$
así que $\boldsymbol{z} \in \ker T_A$. La dimensión de $\ker T_A$ es $n(A) = n - r(A) = n - k$, por el teorema de rango y nulidad. □

4.4. Cambios de Base

Cada aplicación lineal $T \colon V \to W$ tiene una matriz A respecto de un par de bases dadas para V y W. Si se escoge *un par de bases diferentes* para V y W, se obtendrá otra matriz B que representa T con respecto a las bases nuevas. Es imprescindible determinar la relación entre esas dos matrices A y B. Esto es el asunto de *cambio de base* (o "cambio de bases", si se prefiere).

Es útil etiquetar las bases con una letra específica: $\mathcal{B} = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ denotará una base de V; $\mathcal{C} = \{\boldsymbol{y}_1, \dots, \boldsymbol{y}_m\}$ denotará una base de W, si $\dim V = n$ y $\dim W = m$.

Ahora sean $\mathcal{B} := \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ y $\mathcal{B}' := \{\boldsymbol{x}'_1, \dots, \boldsymbol{x}'_n\}$ dos bases para V. A su vez, sean $\mathcal{C} := \{\boldsymbol{y}_1, \dots, \boldsymbol{y}_m\}$ y $\mathcal{C}' := \{\boldsymbol{y}'_1, \dots, \boldsymbol{y}'_m\}$ dos bases para W. La aplicación identidad $1_V \colon V \to V$ tiene, por (4.2), una matriz $P \in M_n(\mathbb{F})$ con respecto a estas dos bases de V; además, $1_W \colon W \to W$ tiene una matriz $Q \in M_m(\mathbb{F})$ con respecto a las dos bases dadas de W. Explícitamente:

$$\boxed{\boldsymbol{x}'_s = 1_V(\boldsymbol{x}'_s) =: \sum_{j=1}^{n} p_{js}\, \boldsymbol{x}_j} \quad \text{y} \quad \boxed{\boldsymbol{y}_i = 1_W(\boldsymbol{y}_i) =: \sum_{r=1}^{m} q_{ri}\, \boldsymbol{y}'_r} \tag{4.10}$$

observando que estas fórmulas propician cambios $\mathcal{B} \mapsto \mathcal{B}'$ y $\mathcal{C}' \mapsto \mathcal{C}$.

Las matrices P y Q se llaman **matrices de cambio de base**.

Fíjese que $P \neq 1_n$ si $\mathcal{B} \neq \mathcal{B}'$ (bases de V diferentes) y que $Q \neq 1_m$ si $\mathcal{C} \neq \mathcal{C}'$ (bases de W diferentes).

Como las dos aplicaciones 1_V y 1_W son biyecciones lineales, estas matrices cuadradas P y Q son *invertibles*, y sus inversos efectúan los cambios de base en el sentido contrario.

Lema 4.33. *Si A es la matriz de una aplicación lineal $T \colon V \to W$ con respecto a las bases $\mathcal{B}$ de V y $\mathcal{C}$ de W; y si B es la matriz de T con respecto a las bases $\mathcal{B}'$ de V y $\mathcal{C}'$ de W, entonces $B = QAP$.*

Demostración. A partir de la fórmula (4.2) para la matriz A y las fórmulas (4.10), se obtiene por un cálculo directo:

$$\begin{aligned} T(\boldsymbol{x}'_s) &= T\biggl(\sum_{j=1}^{n} p_{js}\, \boldsymbol{x}_j\biggr) = \sum_{j=1}^{n} p_{js}\, T(\boldsymbol{x}_j) = \sum_{j=1}^{n} p_{js} \sum_{i=1}^{m} a_{ij}\, \boldsymbol{y}_i \\ &= \sum_{j=1}^{n} p_{js} \sum_{i=1}^{m} a_{ij} \sum_{r=1}^{m} q_{ri}\, \boldsymbol{y}'_r = \sum_{r=1}^{m} \biggl(\sum_{i=1}^{m} \sum_{j=1}^{n} q_{ri}\, a_{ij}\, p_{js}\biggr) \boldsymbol{y}'_r \end{aligned}$$

en el cual la suma doble entre paréntesis al lado derecho debe ser la entrada b_{rs} de la matriz que expresa T con respecto a las bases $\mathcal{B}'$ y $\mathcal{C}'$:

$$T(\boldsymbol{x}'_s) = \sum_{r=1}^{m} b_{rs}\, \boldsymbol{y}'_r.$$

De ahí se obtiene que las entradas de B son:

$$b_{rs} = \sum_{i=1}^{m} \sum_{j=1}^{n} q_{ri}\, a_{ij}\, p_{js}\,.$$

El lado derecho es la entrada (r, s) del producto de matrices QAP. $\square$

Definición 4.34. Dos matrices $A, B \in M_{m\times n}(\mathbb{F})$ son **equivalentes** si hay dos matrices *invertibles* $P \in M_n(\mathbb{F})$ y $Q \in M_m(\mathbb{F})$ tales que $\underline{B = QAP}$. ◊

Antes de seguir, obsérvese que esta sí es una relación de equivalencia:

- ◇ *Reflexividad*: vale $A = 1_m\, A\, 1_n$.
- ◇ *Simetría*: si $B = QAP$, entonces $A = Q^{-1}BP^{-1}$.
- ◇ *Transitividad*: si $B = QAP$ y si $C = Q'BP'$, entonces $C = (Q'Q)A(PP')$. Las matrices $Q'Q$ y PP' son invertibles, con $(Q'Q)^{-1} = Q^{-1}Q'^{-1}$ y $(PP')^{-1} = P'^{-1}P^{-1}$.

Proposición 4.35. *Si la matriz $A \in M_{m\times n}(\mathbb{F})$ tiene rango k, hay matrices invertibles $Q \in M_m(\mathbb{F})$ y $P \in M_n(\mathbb{F})$ tales que*

$$QAP = \begin{bmatrix} 1_k & 0 \\ 0 & 0 \end{bmatrix}, \tag{4.11}$$

en donde cada 0 *es un bloque rectangular de entradas nulas.*

Demostración. Se puede llevar A en una forma escalonada A' con $k = r(A)$ columnas básicas, al aplicarla el Algoritmo 4.1. En este proceso se ejecuta un número finito de operaciones de fila: por eso, $A \mapsto A'$ es equivalente a *premultiplicar* A por una matriz invertible $Q \in M_m(\mathbb{F})$, de modo que $A' = QA$.

Ahora, nótese que la *postmultiplicación* por una matriz $n \times n$ de tipo (3.12a) hace un intercambio de dos *columnas*; la aplicación de varios intercambios de ese tipo puede efectuar la permutación de las columnas de la matriz original,[5] para llevar las columnas básicas de A' a la izquierda. Esto es equivalente a *postmultiplicación* A' por una matriz invertible $P_1 \in M_n(\mathbb{F})$ para así convertir A en la forma

$$A \longmapsto QAP_1 = A'P_1 = \begin{bmatrix} 1_k & F \\ 0 & 0 \end{bmatrix},$$

donde F es un bloque de tamaño $k \times (n-k)$.

A esta nueva matriz se le puede aplicar "operaciones de columna"; pero es igual de fácil *aplicar operaciones de fila a su transpuesta:*

$$(QAP_1)^{\mathtt{t}} = \begin{bmatrix} 1_k & 0 \\ F^{\mathtt{t}} & 0 \end{bmatrix},$$

en donde $F^{\mathtt{t}}$ es un bloque $(n-k) \times k$. Ahora es claro que (si $F \neq 0$) se puede aplicar eliminación gaussiana *simple* a esta última matriz, convirtiéndola en el formato (4.11).

[5]Cualquier permutación de la lista $(1, 2, \dots, n)$ es un producto de varias *transposiciones* $i \leftrightarrow j$.

Como se trata de una matriz $n \times m$, esto requiere premultiplicar por una matriz invertible $P_2 \in M_n(\mathbb{F})$. Esto es, se ejecuta la operación $(QAP_1)^{\mathtt{t}} \mapsto P_2(QAP_1)^{\mathtt{t}}$:

$$P_2(QAP_1)^{\mathtt{t}} = \begin{bmatrix} 1_k & 0 \\ 0 & 0 \end{bmatrix} \qquad \text{(una matriz } n \times m\text{)}.$$

Finalmente, se transpone la matriz de nuevo, para regresar a matrices $m \times n$. Ahora sea $P := P_1 P_2^{\mathtt{t}} \in M_n(\mathbb{F})$ (invertible); entonces se obtiene lo deseado:

$$QAP = QAP_1P_2^{\mathtt{t}} = \big(P_2(QAP_1)^{\mathtt{t}}\big)^{\mathtt{t}} = \begin{bmatrix} 1_k & 0 \\ 0 & 0 \end{bmatrix}. \qquad \square$$

Corolario 4.36. *Dos matrices $A, B \in M_{m\times n}(\mathbb{F})$ son equivalentes* si y solo si $\underline{r(A) = r(B)}$.

Demostración. Una o más operaciones de fila no cambian el rango de A: si Q es invertible, $T_Q \colon \mathbb{F}^m \to \mathbb{F}^m$ es una biyección lineal y $\dim T_Q(W) = \dim W$ para un subespacio $W \leqslant \mathbb{F}^m$; al tomar $W = \operatorname{im} T_A$, se obtiene

$$r(QA) = r(T_{QA}) = r(T_Q T_A) = \dim T_Q(\operatorname{im} T_A)) = \dim(\operatorname{im} T_A) = r(A).$$

⟦ Alternativamente, se puede aprovechar el Ejercicio 4.15: $r(QA) \leqslant r(A)$; y a la vez, $r(A) = r(Q^{-1}(QA)) \leqslant r(QA)$, así que $r(QA) = r(A)$. ⟧

De igual manera, si $P \in M_n(\mathbb{F})$ es invertible, se obtiene $r(AP) = r(A)$. Por lo tanto, si $B = QAP$, entonces

$$r(B) = r(QAP) = r(AP) = r(A).$$

A la inversa, si se sabe que $r(A) = r(B) = k$, entonces hay cuatro matrices invertibles $Q_1, Q_2 \in M_m(\mathbb{F})$ y $P_1, P_2 \in M_n(\mathbb{F})$ tales que

$$Q_1AP_1 = \begin{bmatrix} 1_k & 0 \\ 0 & 0 \end{bmatrix} = Q_2BP_2$$

por la Proposición 4.35 anterior, y eso dice que A y B son equivalentes. $\square$

4.5. Matrices semejantes

Los cambios de base de mayor interés ocurren cuando $W = V$ y se toman las mismas bases en el dominio y el codominio, esto es, $\boldsymbol{y}_j = \boldsymbol{x}_j$ y $\boldsymbol{y}'_r = \boldsymbol{x}'_r$ para $j, r = 1, \dots, n$. Una mirada a las fórmulas (4.10) hace patente que en este caso $\underline{Q = P^{-1}}$ en $M_n(\mathbb{F})$.

Si $T \colon V \to V$ es un operador lineal sobre V, y si $\mathcal{B} = \{\overline{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n}\}$ es una base de V, se dice que $A \in M_n(\mathbb{F})$ es la matriz de T *con respecto a la base* $\mathcal{B}$ si se cumple (4.2) con $\boldsymbol{y}_j = \boldsymbol{x}_j$ para $j = 1, \dots, n$.

Definición 4.37. Dos *matrices cuadradas* $A, B \in M_n(\mathbb{F})$ son **semejantes** si hay una matriz invertible $P \in M_n(\mathbb{F})$ tal que $\underline{B = P^{-1}AP}$. ◊

Si A y B representan una aplicación lineal $T: V \to V$ con respecto a dos bases $\mathcal{B}$ y $\mathcal{C}$ de V, entonces A y B son semejantes. (Este es un caso particular del Lema 4.33.) Inversamente, si A y B son semejantes mediante la relación $B = P^{-1}AP$, y si se define un cambio de base en V por (4.10), entonces B es la matriz de T_A respecto de la nueva base.

La semejanza de matrices cuadradas es otra *relación de equivalencia*:

- *Reflexividad*: vale $A = (1_n)^{-1}A\,1_n$.
- *Simetría*: si $B = P^{-1}AP$, entonces $A = PBP^{-1}$.
- *Transitividad*: si $B = P^{-1}AP$ y si $C = R^{-1}BR$, entonces $C = (PR)^{-1}A(PR)$, puesto que $(PR)^{-1} = R^{-1}P^{-1}$.

► Es claro que dos matrices semejantes son equivalentes, y por lo tanto tienen el mismo rango; pero la coincidencia de rangos no es suficiente para que dos matrices cuadradas sean semejantes. Cabe preguntar si es posible confeccionar una *forma canónica* de una matriz de modo que dos matrices son semejantes si y solo si pueden ser transformadas, mediante operaciones de la forma $A \mapsto M^{-1}AM$, a la misma forma canónica. La respuesta es afirmativa pero sorprendentemente sutil, y no se ofrecerá una respuesta completa en este curso. Sin embargo, más adelante en el capítulo 6, se asociará un par de números $\underline{\operatorname{tr} A}$ y $\underline{\det A}$ a cada matriz cuadrada A de modo que $\operatorname{tr} A = \operatorname{tr} B$ y $\det A = \det B$ toda vez que A y B sean semejantes.

4.6. Ejercicios sobre aplicaciones lineales

En estos ejercicios, V y W son espacios vectoriales sobre un cuerpo $\mathbb{F}$ con $\dim V = n$ y $\dim W = m$; T y S son aplicaciones lineales en $\mathcal{L}(V, W)$.

Ejercicio 4.1. ¿Cuáles de las siguientes aplicaciones $T: \mathbb{R}^2 \to \mathbb{R}^2$ son lineales?

(a) $T(x, y) := (-y, x)$ $\quad T(x, y) := (-y, x)$ $\quad$ (rotación contrario a reloj por un ángulo recto);

(b) $T(x, y) := (-x, y)$ $\quad$ (reflexión en el eje y);

(c) $T(x, y) := (y, x)$ $\quad$ (reflexión en la diagonal $y = x$);

(d) $T(x, y) := (3x, 3y)$ $\quad$ (dilatación por un factor 3);

(e) $T(x, y) := (2x - y, y + 3x)$;

(f) $T(x, y) := (2x + 1, 2y - 1)$.

Ejercicio 4.2. Si $n \in \mathbb{N}$, $\mathbb{R}_n[t]$ denota el espacio vectorial de polinomios reales de grado no mayor que n.

(a) Sea $D\colon \mathbb{R}_3[t] \to \mathbb{R}_2[t]$ la aplicación lineal de derivación (Ejemplo 4.5). Determinar la matriz de D con respecto a las bases:

$$\mathcal{B} := \{1, t, t^2, t^3\} \quad \text{de } \mathbb{R}_3[t]; \quad \text{y} \quad \mathcal{C} := \{1, t, t^2\} \quad \text{de } \mathbb{R}_2[t].$$

(b) Determinar la matriz de la misma aplicación lineal D con respecto a las otras bases $\mathcal{B}'$ de $\mathbb{R}_3[t]$ y $\mathcal{C}'$ de $\mathbb{R}_2[t]$ dadas por

$$\mathcal{B}' := \{t^3 + 1, t^2 + t, t^2 - t, t^3 - 1\}; \quad \text{y} \quad \mathcal{C}' := \{\tfrac{1}{2}t(t + 1), 1 - t^2, \tfrac{1}{2}t(t - 1)\}.$$

Ejercicio 4.3. Si $T : \mathbb{R}^3 \to \mathbb{R}^4$ es una aplicación lineal tal que

$$\begin{aligned} T(-1, 2, 1) &= (-4, -3, 1, 3), \\ T(-2, 0, 3) &= (1, -8, -5, 3), \\ T(4, 1, -2) &= (0, 8, -1, -1), \end{aligned}$$

determinar el vector $T(4, 1, -3)$ en $\mathbb{R}^4$.

Ejercicio 4.4. Si $T : \mathbb{R}^3 \to \mathbb{R}^3$ es una aplicación lineal tal que

$$\begin{aligned} T(1, 0, 1) &= (2, 3, -1), \\ T(1, -1, 1) &= (3, 0, -2), \\ T(1, 2, -1) &= (-2, 7, -1), \end{aligned}$$

determinar $T(\boldsymbol{e}_1)$, $T(\boldsymbol{e}_2)$ y $T(\boldsymbol{e}_3)$.

Ejercicio 4.5. Sean f_1, f_2, f_3 tres formas lineales sobre $\mathbb{R}^3$ dadas por:

$$\begin{aligned} f_1(\boldsymbol{x}) &= 5x_1 - 2x_2 + 9x_3, \\ f_2(\boldsymbol{x}) &= -x_1 + x_2 + 3x_3, \\ f_3(\boldsymbol{x}) &= x_1 - ax_2 + 2ax_3, \end{aligned}$$

donde $a \in \mathbb{R}$. Hallar el valor de a tal que f_1, f_2 y f_3 sean linealmente dependientes, y determinar unas constantes c_1, c_2, c_3 (no todas cero) tales que $c_1 f_1 + c_2 f_2 + c_3 f_3 = 0$.

Ejercicio 4.6. La *base estándar* del espacio vectorial $M_2(\mathbb{F})$ es $\{E_{11}, E_{12}, E_{21}, E_{22}\}$ donde

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

Si $M = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in M_2(\mathbb{F})$, se definen tres aplicaciones L_M, R_M, $S\colon M_2(\mathbb{F}) \to M_2(\mathbb{F})$ por:

$$L_M(A) := MA, \qquad R_M(A) := AM, \qquad S(A) := A^{\mathsf{t}}.$$

Demostrar que L_M, R_M y S son operadores *lineales* sobre $M_2(\mathbb{F})$; y calcular sus matrices con respecto a la base dada. ⟦ Indicación: estas son matrices 4×4. ⟧

Ejercicio 4.7. Sean $c_0, c_1, \dots, c_n$ números reales distintos. Se definen los **polinomios interpolativos de Lagrange** $\{\, \underline{\pi_k(t)} : k = 0, 1, \dots, n \,\}$ en $\mathbb{R}_n[t]$ por

$$\pi_k(t) := \frac{(t - c_0) \cdots (t - c_{k-1})(t - c_{k+1}) \cdots (t - c_n)}{(c_k - c_0) \cdots (c_k - c_{k-1})(c_k - c_{k+1}) \cdots (c_k - c_n)}.$$

(a) Verificar que $\pi_k(c_j) = [\![j = k]\!]$.[6]

(b) Deducir que $\{\pi_0, \pi_1, \dots, \pi_n\}$ es una *base* para $\mathbb{R}_n[t]$.

(c) Demostrar que la base dual del espacio dual $\mathbb{R}_n[t]^*$ es el conjunto de las *evaluaciones* $f_j\colon \mathbb{R}_n[t] \to \mathbb{R}$ definidas por $\underline{f_j(p(t)) := p(c_j)}$, para $j = 0, 1, \dots, n$.

Ejercicio 4.8. Sea V un espacio vectorial sobre $\mathbb{F}$ con $\dim V = n$ y sea V^* su espacio dual. Sea $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ una base de V y sea $\{f_1, \dots, f_n\}$ la base dual de V^*, dada por (4.3).

(a) Si $\{g_1, \dots, g_n\}$ es otra base de V^*, mostrar que hay matrices $A, C \in M_n(\mathbb{F})$ tales que $g_r = \sum_{i=1}^n a_{ir} f_i$ y $f_i = \sum_{s=1}^n c_{si} g_s$ para cada $i, r = 1, \dots, n$. Deducir que A y C son invertibles, con $C = A^{-1}$.

(b) Si $\{\boldsymbol{y}_1, \dots, \boldsymbol{y}_n\}$ son vectores en V, expresadas en términos de la base $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ por $\boldsymbol{y}_s = \sum_{j=1}^n b_{js} \boldsymbol{x}_j$, calcular los valores $g_r(\boldsymbol{y}_s)$ en términos de las matrices A, B, C.

(c) Obtener una matriz $B = [b_{js}]$ tal que estos vectores $\boldsymbol{y}_s$ cumplan $\underline{g_r(\boldsymbol{y}_s) = [\![r = s]\!]}$ para todo $r, s = 1, \dots, n$.

(d) Demostrar que los vectores $\{\boldsymbol{y}_1, \dots, \boldsymbol{y}_n\}$ obtenidos en (c) son linealmente independientes y por ende forman una *base* de V, cuyo base dual de V^* es $\{g_1, \dots, g_n\}$.

[6] Esto es: $\pi_k(c_k) = 1$ para cada k; y $\pi_k(c_j) = 0$ si $j \neq k$.

Ejercicio 4.9. Encontrar la aplicación lineal $T\colon \mathbb{R}^2 \to \mathbb{R}^2$ determinada por:

$$T(1,1) = (3,3) \quad \text{y} \quad \ker T = \{\, (x_1, x_2) \in \mathbb{R}^2 : x_1 + 2x_2 = 0 \,\}.$$

Ejercicio 4.10. Encontrar los subespacios $\ker S$ e $\operatorname{im} S$ y sus respectivas dimensiones $n(S)$ y $r(S)$, si la matriz de $S\colon \mathbb{R}^3 \to \mathbb{R}^3$ con respecto a la base estándar $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3\}$ de $\mathbb{R}^3$ es:

$$A := \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \end{bmatrix}.$$

Ejercicio 4.11. Sea $S\colon V \to W$ una aplicación lineal.

(a) Si $\boldsymbol{x}_1, \dots, \boldsymbol{x}_k \in V$ son vectores tales que $\{S(\boldsymbol{x}_1), \dots, S(\boldsymbol{x}_k)\}$ es linealmente independiente en W, comprobar que $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_k\}$ es linealmente independiente en V.

(b) Si $\{\boldsymbol{y}_1, \boldsymbol{y}_2, \dots, \boldsymbol{y}_r\}$ es linealmente independiente en V *y si* $\ker S = \{\boldsymbol{0}\}$, demostrar que $\{S(\boldsymbol{y}_1), S(\boldsymbol{y}_2), \dots, S(\boldsymbol{y}_r)\}$ es linealmente independiente en W.

Ejercicio 4.12. Encontrar la dimensión y una base para los cuatro espacios $\ker T_A$, $\operatorname{im} T_A$, $\ker(T_A)^{\mathsf{t}}$ e $\operatorname{im}(T_A)^{\mathsf{t}}$, en los dos casos siguientes:

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}; \qquad A = \begin{bmatrix} 1 & 2 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 2 & 0 & 1 \end{bmatrix}.$$

Ejercicio 4.13. Sea $S\colon V \to W$ una aplicación lineal entre dos espacios vectoriales de la *misma* dimensión finita: $\dim V = \dim W = n$. Demostrar que S es inyectiva *si y solo si* S es sobreyectiva. ⟦ Indicación: usar el teorema de rango y nulidad. ⟧

Ejercicio 4.14. Si $T \in \mathcal{L}(V, W)$ donde V, W son espacios vectoriales finitodimensionales con $\dim V > \dim W$, demostrar que T no puede ser inyectivo.

Sea $T\colon C[a,b] \to C[a,b]$ dado por $(Tf)(t) := \int_a^t f(s)\,ds$. Demostrar que T es lineal e inyectiva, pero no sobreyectiva. Concluir que $C[a,b]$ es infinitodimensional sobre $\mathbb{R}$.

Ejercicio 4.15. Si $T\colon V \to W$ y $S\colon W \to Z$ son aplicaciones lineales, demostrar estas relaciones entre núcleos e imágenes:[7]

$$\begin{aligned} \ker(ST) &\geqslant \ker T, \qquad & \ker((ST)^{\mathsf{t}}) &\geqslant \ker(S^{\mathsf{t}}), \\ \operatorname{im}(ST) &\leqslant \operatorname{im} S, \qquad & \operatorname{im}((ST)^{\mathsf{t}}) &\leqslant \operatorname{im} T^{\mathsf{t}}. \end{aligned}$$

Concluir que $r(ST) \leqslant r(S)$ y que $r(ST) \leqslant r(T)$.

[7] Si M, N son espacios vectoriales, la notación $M \geqslant N$ es sinónimo de $N \leqslant M$.

Ejercicio 4.16. Convertir cada uno de estas matrices a la forma escalonada equivalente:

$$A = \begin{bmatrix} 1 & 2 & 0 & 2 & 1 \\ -1 & -2 & 1 & 1 & 0 \\ 1 & 2 & -3 & -7 & -2 \end{bmatrix}; \qquad B = \begin{bmatrix} 2 & -1 & 3 & 1 & 1 \\ -1 & 0 & -2 & 1 & -3 \\ 1 & 2 & -1 & -4 & 3 \\ 3 & 2 & -2 & -3 & -1 \end{bmatrix}.$$

Ejercicio 4.17. Demostrar que $r(A) = 2$, donde

$$A := \begin{bmatrix} 1 & 2 & -1 & 3 \\ 3 & 4 & 0 & -1 \\ -1 & 0 & -2 & 7 \end{bmatrix}.$$

Expresar la tercera y cuarta columnas como combinaciones lineales de las primeras dos columnas.

Ejercicio 4.18. Hallar los valores de x tales que $r(A) = 3$, donde

$$A := \begin{bmatrix} 4 & 4 & -3 & 1 \\ 1 & 1 & -1 & 0 \\ x & 2 & 2 & 2 \\ 9 & 9 & x & 3 \end{bmatrix}.$$

Ejercicio 4.19. Calcular $r(A)$, encontrar una base para el espacio de soluciones de $A\boldsymbol{x} = \mathbf{0}$ y describir el conjunto de soluciones de $A\boldsymbol{x} = \boldsymbol{b}$, donde

$$[A \mid \boldsymbol{b}] := \left[\begin{array}{ccccc|c} 1 & -1 & 2 & 0 & 3 & -1 \\ 0 & 2 & 1 & 3 & 1 & -4 \\ -1 & 1 & 5 & 1 & 0 & 3 \\ -1 & 0 & 1 & -1 & -2 & 4 \end{array}\right].$$

Ejercicio 4.20. Determinar todas las soluciones del sistema de ecuaciones:

$$\begin{aligned} 2x_1 - 3x_2 + x_3 + x_4 &= 7 \\ 3x_1 - 2x_2 + 2x_3 - 2x_4 &= 5 \\ 6x_1 + x_2 + 5x_3 - 11x_4 &= -1. \end{aligned}$$

Ejercicio 4.21. Determinar todas las soluciones del sistema de ecuaciones:

$$\begin{aligned} x_1 - 3x_2 + x_3 - x_4 &= -8 \\ 2x_1 - x_2 + 2x_3 + 3x_4 &= -1 \\ 7x_1 - 11x_2 + 7x_3 + 3x_4 &= -10. \end{aligned}$$

Ejercicio 4.22. Encontrar una base para el espacio de soluciones de este sistema homogéneo:

$$\begin{aligned} x_1 - 2x_2 - 3x_3 + x_4 + x_5 &= 0 \\ -4x_1 - 2x_2 - 5x_3 + 5x_4 - x_5 &= 0 \\ x_1 + x_2 \qquad\quad - x_4 + 2x_5 &= 0 \\ -2x_2 - 4x_3 + 2x_4 + 3x_5 &= 0 \\ -3x_1 + x_2 + 2x_3 \qquad\quad - x_5 &= 0. \end{aligned}$$

Ejercicio 4.23. Encontrar el valor de a que hace posible resolver el sistema

$$\begin{aligned} x_1 - 2x_2 + x_3 - 3x_4 &= a \\ 2x_1 + x_2 + x_3 - 2x_4 &= 2 \\ 7x_1 - 4x_2 + 5x_3 - 13x_4 &= -8. \end{aligned}$$

Describir el conjunto de soluciones para este valor de a.

Ejercicio 4.24. Demostrar que tres puntos distintos (x_1, y_1), (x_2, y_2), (x_3, y_3) en $\mathbb{R}^2$ son colineales si y solo si la siguiente matriz M tiene rango 2:

$$M = \begin{bmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ 1 & x_3 & y_3 \end{bmatrix}.$$

Ejercicio 4.25. Hallar la forma escalonada A' de la matriz

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

Aplicar operaciones elementales *de columna* a la matriz A' para reducirla a $\begin{bmatrix} 1_k & 0 \\ 0 & 0 \end{bmatrix}$. Ahora, ¿cuál es el rango $r(A)$ de la matriz original?

Ejercicio 4.26. Dada la matriz

$$A := \begin{bmatrix} 1 & 1 & 2 \\ 2 & 2 & 4 \\ 3 & 3 & 6 \end{bmatrix},$$

hallar una matriz $B \in \mathbb{R}^{3\times 3}$ con $r(B) = 2$ tal que $AB = 0$. ⟦ Indicación: las columnas de B serán soluciones de la ecuación $A\boldsymbol{x} = \mathbf{0}$. ⟧

Ejercicio 4.27. Si $A \in M_{m\times n}(\mathbb{F})$ tiene rango $r(A) = k$, demostrar que A es una suma de unas k matrices de rango 1. ⟦ Indicación: usar la Proposición 4.35. ⟧

Ejercicio 4.28. Sea V un espacio vectorial sobre $\mathbb{F}$ con $\dim V = 2$ y sea $\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ la matriz del operador lineal $T \colon V \to V$ con respecto a una base $\{\boldsymbol{x}_1, \boldsymbol{x}_2\}$ de V. Si $\boldsymbol{y}_1 := \boldsymbol{x}_1 - \boldsymbol{x}_2$, $\boldsymbol{y}_2 := 2\boldsymbol{x}_1 + 3\boldsymbol{x}_2$, comprobar que $\{\boldsymbol{y}_1, \boldsymbol{y}_2\}$ es otra base de V y encontrar la matriz de T con respecto a esta nueva base.

Ejercicio 4.29. La fórmula (4.2) para una matriz de una aplicación lineal y las fórmulas de cambio de base (4.10) asumen que las bases tales como $\mathcal{B} = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ *están dadas en orden*: se considera $\mathcal{B}$ como una *lista ordenada* de vectores.[8] Si se define $\boldsymbol{x}'_1 := \boldsymbol{x}_2$, $\boldsymbol{x}'_2 := \boldsymbol{x}_1$, y $\boldsymbol{x}'_j = \boldsymbol{x}_j$ para $j = 3, \dots, n$, ¿cuál es la matriz P de (4.10) correspondiente? Expresar las entradas b_{rs} de la matriz $B = P^{-1}AP$ en términos de las entradas a_{ij} de A.

Ejercicio 4.30. Determinar si las siguientes matrices A y B son semejantes o no:

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

Ejercicio 4.31. Si A y B son matrices $n \times n$ semejantes, escrito $A \sim B$, determinar (con justificaciones) si los pares de matrices que siguen son semejantes o no:

(a) ¿Es cierto que $A^2 \sim B^2$?

(b) ¿Es cierto que $A^{-1} \sim B^{-1}$?

(c) ¿Es cierto que $(A^3 + 7A^2 - 2A + 3\,1_n) \sim (B^3 + 7B^2 - 2B + 3\,1_n)$?

Ejercicio 4.32. Sea $T \colon \mathbb{R}^3 \to \mathbb{R}^3$ el operador lineal cuya matriz con respecto a la base estándar $\mathcal{B} = \{\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3\}$ es

$$A = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 4 \\ 0 & 4 & 9 \end{bmatrix}.$$

Sea $\mathcal{B}' := \{\boldsymbol{x}'_1, \boldsymbol{x}'_2, \boldsymbol{x}'_3\}$ la base de $\mathbb{R}^3$ dada por

$$\boldsymbol{x}'_1 = \begin{bmatrix} 0 \\ -2 \\ 1 \end{bmatrix}, \quad \boldsymbol{x}'_2 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \boldsymbol{x}'_3 = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}.$$

[8] Algunos escriben $\mathcal{B} = (\boldsymbol{x}_1, \dots, \boldsymbol{x}_n)$ para denotar esta base como una lista ordenada.

(a) Encontrar las matrices P y P^{-1} que expresan los elementos de $\mathcal{B}'$ en términos de los elementos de la base $\mathcal{B}$, y viceversa.

(b) Hallar la matriz de T con respecto a las base $\mathcal{B}'$.

5 Espacios vectoriales euclidianos

> *Cuando una línea recta erguida sobre una línea recta hace ángulos adyacentes iguales entre sí, cada uno de los ángulos iguales es* recto *y la línea recta parada se llama* perpendicular *a la subyacente.*
>
> — Euclides de Alejandría[1]

Hasta ahora se ha estudiado los aspectos de espacios vectoriales y sus aplicaciones lineales que se manifiestan con un cuerpo arbitrario $\mathbb{F}$ de escalares: independencia lineal, bases, matrices con respecto a un par de bases, equivalencia y semejanza de matrices. Pero hay otro elemento estructural del espacio $\mathbb{R}^n$, visto en el capítulo 1, que no se ha generalizado aún: el *producto escalar* de un par de vectores.

Es usual emplear esta noción solo si el cuerpo de escalares $\mathbb{F}$ es $\mathbb{R}$, el de números reales, o bien $\mathbb{C}$, el de números complejos. En este capítulo se tomará $\mathbb{F} = \mathbb{R}$ ó $\mathbb{C}$. Inicialmente se introducirá los productos escalares en el caso $\mathbb{F} = \mathbb{R}$, y después se abordará la variante en donde $\mathbb{F} = \mathbb{C}$.

5.1. Productos escalares reales

En esta sección, V y W denotarán *espacios vectoriales sobre* $\mathbb{R}$ (espacios vectoriales reales, espacios $\mathbb{R}$-vectoriales).

Definición 5.1. Un **producto escalar** (real) en V es una función que a cada par de vectores $\underline{\boldsymbol{x}, \boldsymbol{y} \in V}$ asocia un número real $\underline{\langle \boldsymbol{x}, \boldsymbol{y} \rangle \in \mathbb{R}}$ tal que:

(a) $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \langle \boldsymbol{y}, \boldsymbol{x} \rangle$ para todo $\boldsymbol{x}, \boldsymbol{y} \in V$;

(b) $\langle \boldsymbol{x}, \boldsymbol{y} + \boldsymbol{z} \rangle = \langle \boldsymbol{x}, \boldsymbol{y} \rangle + \langle \boldsymbol{x}, \boldsymbol{z} \rangle$ para todo $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in V$;

(c) $\langle \boldsymbol{x}, c\boldsymbol{y} \rangle = \langle c\boldsymbol{x}, \boldsymbol{y} \rangle = c\langle \boldsymbol{x}, \boldsymbol{y} \rangle$ para todo $\boldsymbol{x}, \boldsymbol{y} \in V$, $c \in \mathbb{R}$;

(d) $\langle \boldsymbol{x}, \boldsymbol{x} \rangle \geqslant 0$, con igualdad solo si $\boldsymbol{x} = \boldsymbol{0}$ en V.

Provisto de un producto escalar real, V se llama un **espacio euclidiano**. ◊

Nótese que (a) y (b) implican también que $\langle \boldsymbol{x} + \boldsymbol{z}, \boldsymbol{y} \rangle = \langle \boldsymbol{x}, \boldsymbol{y} \rangle + \langle \boldsymbol{z}, \boldsymbol{y} \rangle$ si $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in V$.

Ejemplo 5.2. Si $V = \mathbb{R}^n$, el *producto punto* de dos vectores (de columna) es:

$$\boxed{\langle \boldsymbol{x}, \boldsymbol{y} \rangle \equiv \boldsymbol{x} \cdot \boldsymbol{y} = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n.} \tag{5.1}$$

La Proposición 1.3 dice que este es un producto escalar sobre $\mathbb{R}^n$. ◊

[1] Su tratado geométrico *Elementos* empieza con 23 definiciones, entre ellas esta, de ángulo recto.

Ejemplo 5.3. Si $V = C[a,b]$ es el espacio de funciones continuas $f\colon [a,b] \to \mathbb{R}$, defínase

$$\langle f, g\rangle := \int_a^b f(t)g(t)\,dt.$$

Es fácil verificar que esta fórmula define un producto escalar sobre $C[a,b]$. ◊

Ejemplo 5.4. La **traza** de una matriz cuadrada $C \in M_n(\mathbb{R})$ es la suma de sus elementos diagonales:

$$\underline{\operatorname{tr}(C)} := c_{11} + c_{22} + \cdots + c_{nn} = \sum_{k=1}^{n} c_{kk}\,. \tag{5.2}$$

La receta

$$\langle A, B\rangle := \underline{\operatorname{tr}(A^{\mathsf{t}}B)} = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}b_{ij}$$

define un producto escalar sobre el espacio vectorial $M_{m\times n}(\mathbb{R})$: si A y B son matrices $m \times n$, entonces $A^{\mathsf{t}}B$ es una matriz $n \times n$. Obsérvese que

$$\langle A, A\rangle = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}^2 \geqslant 0,$$

con igualdad solo si cada $a_{ij} = 0$, solo si $A = 0$ en $M_{m\times n}(\mathbb{R})$. ◊

En el Capítulo 1 las propiedades del producto escalar $\boldsymbol{x} \cdot \boldsymbol{y}$ en $\mathbb{R}^n$ fueron deducidas como consecuencias de las relaciones algebraicas expuestas en la Proposición 1.3 y no con cálculos que dependían de la forma específica (5.1) del producto punto. Entonces, las conclusiones de la sección 1.1 en torno al producto escalar *siguen válidas para cualquier otro producto escalar* que cumple la Definición 5.1, con sólo reemplazar $\boldsymbol{x} \cdot \boldsymbol{y}$ por $\langle \boldsymbol{x}, \boldsymbol{y}\rangle$ en cada caso: se define la norma de un vector, el ángulo entre dos vectores no nulos, la distancia entre dos vectores, la proyección de un vector a lo largo de otro; y la desigualdad de Cauchy sigue válida.

Proposición 5.5. *Si V es un espacio vectorial euclidiano, se define la **norma** (o longitud) de un vector $\boldsymbol{x} \in V$ por $\|\boldsymbol{x}\| := \sqrt{\langle \boldsymbol{x}, \boldsymbol{x}\rangle}$, empleando la fórmula* (1.3).

*La norma así definida verifica la **desigualdad de Schwarz**:*

$$\boxed{\bigl|\langle \boldsymbol{x}, \boldsymbol{y}\rangle\bigr| \leqslant \|\boldsymbol{x}\|\,\|\boldsymbol{y}\|} \quad \textit{para todo} \quad \boldsymbol{x}, \boldsymbol{y} \in V, \tag{5.3}$$

con igualdad si y solo si $\boldsymbol{x}, \boldsymbol{y}$ son proporcionales.

Demostración. La desigualdad (5.3) (con igualdad) es obvio si $\boldsymbol{x} = \mathbf{0}$ o $\boldsymbol{y} = \mathbf{0}$. Supóngase entonces que $\boldsymbol{x} \neq \mathbf{0}$, $\boldsymbol{y} \neq \mathbf{0}$.

Al igual que en la demostración de la Proposición 1.5, defínase $f(t) := \|\boldsymbol{x} + t\boldsymbol{y}\|^2$ para $t \in \mathbb{R}$. Se obtiene

$$\begin{aligned} f(t) &= \langle \boldsymbol{x} + t\boldsymbol{y}, \boldsymbol{x} + t\boldsymbol{y} \rangle \\ &= \langle \boldsymbol{x}, \boldsymbol{x} \rangle + 2t\langle \boldsymbol{x}, \boldsymbol{y} \rangle + t^2 \langle \boldsymbol{y}, \boldsymbol{y} \rangle \quad \text{porque} \quad \langle \boldsymbol{y}, \boldsymbol{x} \rangle = \langle \boldsymbol{x}, \boldsymbol{y} \rangle \in \mathbb{R}, \\ &= \|\boldsymbol{y}\|^2\, t^2 + 2\langle \boldsymbol{x}, \boldsymbol{y} \rangle\, t + \|\boldsymbol{x}\|^2 \\ &=: at^2 + bt + c. \end{aligned}$$

Esta función cuadrática tiene $a > 0$ y cumple $f(t) \geqslant 0$ para todo $t \in \mathbb{R}$. Entonces su discriminante cumple $\underline{b^2 - 4ac \leqslant 0}$, esto es,

$$4\langle \boldsymbol{x}, \boldsymbol{y} \rangle^2 - 4\|\boldsymbol{x}\|^2\|\boldsymbol{y}\|^2 \leqslant 0,$$

esto es, $\langle \boldsymbol{x}, \boldsymbol{y} \rangle^2 \leqslant \|\boldsymbol{x}\|^2\|\boldsymbol{y}\|^2$. Tomando raíces cuadradas (positivas), se obtiene la desigualdad (5.3), con igualdad solo si $b^2 - 4ac = 0$ y $f(t) = 0$ para un solo valor $t = t_0$.

Pero entonces $f(t_0) = \|\boldsymbol{x} + t_0\boldsymbol{y}\|^2 = 0$, de dónde $\boldsymbol{x} + t_0\boldsymbol{y} = \mathbf{0}$, o sea $\boldsymbol{x} = -t_0\boldsymbol{y}$; esto dice que $\boldsymbol{x}$, $\boldsymbol{y}$ son proporcionales. □

Ejemplo 5.6 (Schwarz)**.** Las sucesiones reales $\boldsymbol{x} = (x_0, x_1, x_2, \dots, x_k, \dots)$, tales que la serie $\sum_{k=0}^{\infty} x_k^2$ converge, conforman un espacio vectorial $\ell^2(\mathbb{R})$. Defínase

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle := \sum_{k=0}^{\infty} x_k y_k.$$

Resulta que esta serie converge para todo par de sucesiones $\boldsymbol{x}, \boldsymbol{y} \in \ell^2(\mathbb{R})$ y verifica la desigualdad (5.3). ◊

Definición 5.7. Sea V un espacio vectorial euclidiano. Se define la **distancia** entre dos vectores $\boldsymbol{x}, \boldsymbol{y} \in V$ por

$$\underline{d(\boldsymbol{x}, \boldsymbol{y})} := \|\boldsymbol{x} - \boldsymbol{y}\|.$$

Se define el **ángulo** $\theta \in [0, \pi]$ entre dos vectores no nulos $\boldsymbol{x}, \boldsymbol{y} \in V$ por la fórmula (1.5):

$$\cos\theta := \frac{\langle \boldsymbol{x}, \boldsymbol{y} \rangle}{\|\boldsymbol{x}\|\,\|\boldsymbol{y}\|}\,.$$ ◊

Escolio 5.8. *La* norma *en un espacio euclidiano tiene la siguientes propiedades:*

(a) $\|\alpha\boldsymbol{x}\| = |\alpha|\,\|\boldsymbol{x}\|$ *(homogeneidad positiva),*

(b) $\|\boldsymbol{x} + \boldsymbol{y}\| \leqslant \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$ *(desigualdad triangular),*

(c) $\|\boldsymbol{x}\| \geqslant 0$, *con igualdad solo si* $\boldsymbol{x} = \mathbf{0}$. ⊟

La demostración es idéntica a la de la Proposición 1.6.

5.2. Bases ortonormales

Como el producto escalar permite definir (el coseno de) un ángulo entre dos vectores no nulos en un espacio vectorial real, es legítimo considerar que dos vectores son ortogonales cuando ese coseno es 0.

Definición 5.9. Se dice que dos vectores $\boldsymbol{x}, \boldsymbol{y} \in V$ son **ortogonales** si $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = 0$. (En particular, ese es el caso si $\boldsymbol{x} = \mathbf{0}$ o $\boldsymbol{y} = \mathbf{0}$.) Se escribe $\underline{\boldsymbol{x} \perp \boldsymbol{y}}$ para significar que $\boldsymbol{x}, \boldsymbol{y}$ son ortogonales.

Los vectores no nulos $\boldsymbol{x}_1, \dots, \boldsymbol{x}_m \in V$ forman un **conjunto ortogonal** en V si $\boldsymbol{x}_i \perp \boldsymbol{x}_j$ para $i \neq j$. ◊

Lema 5.10. *Un conjunto ortogonal de vectores en V es linealmente independiente.*

Demostración. Sea $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_m\} \subset V$ un conjunto ortogonal de vectores no nulos. Tómese $c_1, \dots, c_m \in \mathbb{R}$ tales que $c_1\boldsymbol{x}_1 + \cdots + c_m\boldsymbol{x}_m = \mathbf{0}$. Para cada índice $j = 1, \dots, m$, se ve que

$$0 = \langle \boldsymbol{x}_j, \mathbf{0} \rangle = \langle \boldsymbol{x}_j, c_1\boldsymbol{x}_1 + \cdots + c_m\boldsymbol{x}_m \rangle = \langle \boldsymbol{x}_j, c_j\boldsymbol{x}_j \rangle = c_j \, \|\boldsymbol{x}_j\|^2,$$

y luego $c_j = 0$ porque $\|\boldsymbol{x}_j\| > 0$. Así, $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_m\}$ es linealmente independiente en V. □

Definición 5.11. Si $M \subseteq V$ (no necesariamente subespacio), se define su **complemento ortogonal** $\underline{M^\perp} \subseteq V$ por:

$$\boxed{M^\perp := \{\, \boldsymbol{y} \in V : \langle \boldsymbol{y}, \boldsymbol{x} \rangle = 0, \text{ para todo } \boldsymbol{x} \in M \,\}.} \tag{5.4}$$

Nótese que $M^\perp$ es un *subespacio* de V porque, si $\boldsymbol{y}, \boldsymbol{z} \in M^\perp$, $c \in \mathbb{R}$, y si $\boldsymbol{x} \in M$, entonces

$$\begin{aligned} \langle \boldsymbol{y} + \boldsymbol{z}, \boldsymbol{x} \rangle &= \langle \boldsymbol{y}, \boldsymbol{x} \rangle + \langle \boldsymbol{z}, \boldsymbol{x} \rangle = 0 + 0 = 0, \\ \langle c\boldsymbol{y}, \boldsymbol{x} \rangle &= c \, \langle \boldsymbol{y}, \boldsymbol{x} \rangle = c(0) = 0, \end{aligned}$$

así que $\boldsymbol{y} + \boldsymbol{z} \in M^\perp$, $c\boldsymbol{y} \in M^\perp$. ◊

Como $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \langle \boldsymbol{y}, \boldsymbol{x} \rangle$, es fácil ver que $M \subseteq (M^\perp)^\perp =: M^{\perp\perp}$.

Definición 5.12. Una juego de vectores $\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_m\} \subset V$ es una **familia ortonormal** si

(a) $\langle \boldsymbol{e}_j, \boldsymbol{e}_k \rangle = 0$ cuando $j \neq k$ *(ortogonalidad)*;

(b) $\langle \boldsymbol{e}_k, \boldsymbol{e}_k \rangle = 1$ para todo $k = 1, 2, \dots, m$ *(normalización)*.

Una **base ortonormal** de V es una *base* $\mathcal{E} = \{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}$ de V que cumple (a) y (b).

⟦ Es usual *listar* los vectores de la base ortonormal $\mathcal{E}$ en un determinado orden. ⟧ ◊

Dicho con otras palabras: $\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_m\}$ es una familia ortonormal en V si $\boldsymbol{e}_j \perp \boldsymbol{e}_k$ para $j \neq k$ y $\|\boldsymbol{e}_k\| = 1$ para todo k. Las propiedades (a) y (b) de la Definición 5.12 pueden resumirse en la frase: $\langle \boldsymbol{e}_j, \boldsymbol{e}_k \rangle = [\![j = k]\!]$.

Una familia ortonormal es linealmente independiente, por el Lema 5.10. Entonces es una base ortonormal de V si y solo si $m = \dim V$ (esto es, $m = n$).

► Si $\mathcal{E} = \{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}$ es una base ortonormal de V, sean $\boldsymbol{x} = \sum_{j=1}^n x_j \boldsymbol{e}_j$, $\boldsymbol{y} = \sum_{k=1}^n y_k \boldsymbol{e}_k$ las expansiones de dos vectores $\boldsymbol{x}, \boldsymbol{y} \in V$ en esa base. Entonces

$$\begin{aligned} \langle \boldsymbol{x}, \boldsymbol{y} \rangle &= \left\langle \sum_{j=1}^n x_j\, \boldsymbol{e}_j, \sum_{k=1}^n y_k, \boldsymbol{e}_k \right\rangle = \sum_{j=1}^n \sum_{k=1}^n \langle x_j\, \boldsymbol{e}_j, y_k\, \boldsymbol{e}_k \rangle \\ &= \sum_{j=1}^n \sum_{k=1}^n x_j y_k\, \langle \boldsymbol{e}_j, \boldsymbol{e}_k \rangle = \sum_{k=1}^n x_k y_k = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n, \end{aligned}$$

porque los términos de la doble suma con $j \neq k$ se anulan. Luego se recupera la forma explícita (5.1) – de producto punto” – para el producto escalar al usar estas *coordenadas con respecto a una base ortonormal*.

Definición 5.13. Sea V un espacio vectorial euclidiano de dimensión finita n. A partir de una base ortonormal $\mathcal{E} = \{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}$ de V, se puede *identificar el espacio vectorial dual V^* con el espacio vectorial original V*. Defínase un juego de formas lineales $\mathcal{F} := \{f_1, \dots, f_n\} \subset V^*$ por

$$\boxed{f_k(\boldsymbol{x}) := \langle \boldsymbol{e}_k, \boldsymbol{x} \rangle} \quad \text{para} \quad k = 1, \dots, n. \tag{5.5}$$

Si $\boldsymbol{x} = \sum_{j=1}^n x_j \boldsymbol{e}_j$, entonces $f_k(\boldsymbol{x}) = x_k$. Esto dice que $\mathcal{F}$ es la *base dual* en V^* de la base $\mathcal{E}$ de V.

Ahora sea $J \colon V \to V^*$ la aplicación lineal definido por

$$J(\boldsymbol{y}) \equiv J\bigl(\textstyle\sum_{k=1}^n y_k \boldsymbol{e}_k\bigr) := \textstyle\sum_{k=1}^n y_k f_k. \tag{5.6a}$$

Como J está determinada por sus valores sobre una base de V, es suficiente notar que $\underline{J\boldsymbol{e}_k := f_k}$ para $k = 1, 2, \dots, n$. Fíjese que:

$$J(\boldsymbol{y})(\boldsymbol{x}) = \sum_{k=1}^n y_k\, f_k(\boldsymbol{x}) = \sum_{k=1}^n y_k x_k = \langle \boldsymbol{y}, \boldsymbol{x} \rangle, \tag{5.6b}$$

así que $J(\boldsymbol{y})$ es simplemente la forma lineal $\underline{\boldsymbol{x} \mapsto \langle \boldsymbol{y}, \boldsymbol{x} \rangle} \in \mathbb{R}$. ◊

Este resultado aclara que J tiene una descripción que *no depende de la base ortonormal* de V. Dicho de otro modo: se puede identificar V^* con V de manera “canónica”, en la presencia de un producto escalar.

Lema 5.14. *Bajo la identificación* (5.6) *de V con V^*, el complemento ortogonal $M^\perp$ de un* subespacio *$M \subseteq V$ coincide con el* anulador *$M^\perp \leqslant V^*$ de la Definición 4.19.*

En consecuencia, el espacio euclidiano V es la suma directa *de M y su complemento ortogonal: $\underline{M \oplus M^\perp = V}$.*

Demostración. Es simplemente cuestión de notar que, si $M \leqslant V$,

$$\begin{aligned} J(\boldsymbol{y}) \in M^\perp \text{ (anulador)} &\iff J(\boldsymbol{y})(\boldsymbol{x}) = 0 \quad \text{para todo } \boldsymbol{x} \in M \\ &\iff \langle \boldsymbol{y}, \boldsymbol{x} \rangle = 0 \qquad \text{para todo } \boldsymbol{x} \in M \\ &\iff \boldsymbol{y} \in M^\perp \text{ (complemento ortogonal).} \end{aligned}$$

Se sabe, por la Proposición 4.20, que $\dim(M^\perp) = \dim V - \dim M$.

Si $\boldsymbol{x} \in M \cap M^\perp$, entonces $\langle \boldsymbol{x}, \boldsymbol{x} \rangle = 0$, así que $\boldsymbol{x} = \mathbf{0}$; esto es, $M \cap M^\perp = \{\mathbf{0}\}$. La suma de subespacios $M + M^\perp \leqslant V$ entonces es *directa*; se escribe $M \oplus M^\perp$. Por conteo de dimensiones (Ejercicio 2.19) se obtiene

$$\dim(M \oplus M^\perp) = \dim M + \dim M^\perp = \dim V,$$

y se concluye que $M \oplus M^\perp = V$. □

► La definición de la **transpuesta** de una aplicación lineal $S \colon V \to W$ se simplifica, en el caso de espacios euclidianos, al usar esta identificación entre V y V^* y también entre W y W^*. Para evitar ambigüedades, conviene denotar esa segunda identificación por $J' \colon W \to W^*$. Esto es,

$$J(\boldsymbol{z}) : \boldsymbol{x} \mapsto \langle \boldsymbol{z}, \boldsymbol{x} \rangle_V \quad \text{y además} \quad J'(\boldsymbol{y}) : \boldsymbol{y} \mapsto \langle \boldsymbol{y}, \boldsymbol{w} \rangle_W$$

si $\boldsymbol{x}, \boldsymbol{z} \in V$ y si $\boldsymbol{y}, \boldsymbol{w} \in W$. Ahora se puede identificar $S^{\mathsf{t}} \colon W^* \to V^*$ con $J^{-1} S^{\mathsf{t}} J' \colon W \to V$; y se calcula:

$$\langle J^{-1} S^{\mathsf{t}} J'(\boldsymbol{y}), \boldsymbol{x} \rangle_V = S^{\mathsf{t}} J'(\boldsymbol{y})(\boldsymbol{x}) := J'(\boldsymbol{y})(S(\boldsymbol{x})) = \langle \boldsymbol{y}, S(\boldsymbol{x}) \rangle_W \,. \tag{5.7a}$$

De ahora en adelante se escribirá simplemente S^{t} en vez de $J^{-1} S^{\mathsf{t}} J'$, como aplicación lineal de W en V. La ecuación anterior, para $\boldsymbol{x} \in V$, $\boldsymbol{y} \in W$, se simplifica en:

$$\boxed{\langle S^{\mathsf{t}}(\boldsymbol{y}), \boldsymbol{x} \rangle = \langle \boldsymbol{y}, S(\boldsymbol{x}) \rangle.} \tag{5.7b}$$

Esta expresión (5.7b) es *la única fórmula necesaria* para hacer cálculos con transpuestas de aplicaciones lineales entre espacios euclidianos: *una aplicación lineal se traslada de un lado a otro de un producto escalar real al poner su transpuesta al otro lado*.

► En el caso de aplicaciones matriciales $T_A \colon \mathbb{R}^n \to \mathbb{R}^m$, hay otra simplificación. Como de costumbre, $\boldsymbol{x} \in \mathbb{R}^n$ denota un vector de columna. Se escribirá (como ya se ha hecho algunas veces) $\boldsymbol{x}^{\mathsf{t}}$ *para denotar el vector de fila* asociado a un elemento del espacio vectorial dual. El producto escalar se convierte en $\underline{\boldsymbol{x}^{\mathsf{t}}\boldsymbol{y}} = \boldsymbol{x} \cdot \boldsymbol{y}$. La simetría del producto escalar indica que

$$\boldsymbol{x}^{\mathsf{t}}\boldsymbol{y} = \boldsymbol{x} \cdot \boldsymbol{y} = \boldsymbol{y} \cdot \boldsymbol{x} = \boldsymbol{y}^{\mathsf{t}}\boldsymbol{x}.$$

Si A es la matriz $m \times n$ de la aplicación lineal $S \colon V \to W$ (con respecto a determinadas bases), la receta (5.7b) es entonces equivalente a la fórmula matricial:

$$\boldsymbol{x}^{\mathsf{t}}A^{\mathsf{t}}\boldsymbol{y} = (A^{\mathsf{t}}\boldsymbol{y})^{\mathsf{t}}\boldsymbol{x} = \boldsymbol{y}^{\mathsf{t}}A\boldsymbol{x}. \tag{5.7c}$$

⟦ Se obtiene el mismo resultado al transponer $\boldsymbol{y}^{\mathsf{t}}A\boldsymbol{x}$, considerada como matriz 1×1. ⟧

► Para comprobar la *existencia* de una base ortonormal para un espacio vectorial euclidiano, hay un algoritmo que la *construye* a partir de otra base cualquiera. Esa construcción es un proceso iterativo que toma cada vector de la base original y lo proyecta sobre una recta que es ortogonal a cada uno de los vectores anteriores. Este proceso se conoce como el **Algoritmo de Gram y Schmidt**.[2]

Proposición 5.15. *Sea V un espacio euclidiano de dimensión $n > 0$, y sea $\mathcal{B} = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ una base de V. Defínase los vectores $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n \in V$ como sigue:*

$$\begin{aligned}
& & \boldsymbol{e}_1 &:= \boldsymbol{x}_1/\|\boldsymbol{x}_1\|; \\
\boldsymbol{y}_2 &:= \boldsymbol{x}_2 - \langle \boldsymbol{e}_1, \boldsymbol{x}_2\rangle\, \boldsymbol{e}_1, & \boldsymbol{e}_2 &:= \boldsymbol{y}_2/\|\boldsymbol{y}_2\|; \\
\boldsymbol{y}_3 &:= \boldsymbol{x}_3 - \langle \boldsymbol{e}_1, \boldsymbol{x}_3\rangle\, \boldsymbol{e}_1 - \langle \boldsymbol{e}_2, \boldsymbol{x}_3\rangle\, \boldsymbol{e}_2, & \boldsymbol{e}_3 &:= \boldsymbol{y}_3/\|\boldsymbol{y}_3\|; \\
&\vdots & &\vdots \\
\boldsymbol{y}_k &:= \boldsymbol{x}_k - \langle \boldsymbol{e}_1, \boldsymbol{x}_k\rangle\, \boldsymbol{e}_1 - \langle \boldsymbol{e}_2, \boldsymbol{x}_k\rangle\, \boldsymbol{e}_2 - \cdots - \langle \boldsymbol{e}_{k-1}, \boldsymbol{x}_k\rangle\, \boldsymbol{e}_{k-1}, & \boldsymbol{e}_k &:= \boldsymbol{y}_k/\|\boldsymbol{y}_k\|. \\
&\vdots & &\vdots \\
\boldsymbol{y}_n &:= \boldsymbol{x}_n - \langle \boldsymbol{e}_1, \boldsymbol{x}_n\rangle\, \boldsymbol{e}_1 - \langle \boldsymbol{e}_2, \boldsymbol{x}_n\rangle\, \boldsymbol{e}_2 - \cdots - \langle \boldsymbol{e}_{k-1}, \boldsymbol{x}_n\rangle\, \boldsymbol{e}_{n-1}, & \boldsymbol{e}_n &:= \boldsymbol{y}_n/\|\boldsymbol{y}_n\|.
\end{aligned} \tag{5.8}$$

Entonces $\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_k\}$ es una base ortonormal del subespacio $\mathrm{lin}\langle \boldsymbol{x}_1, \dots, \boldsymbol{x}_k\rangle$ de V, para cada $k = 1, 2, \dots, n$; en particular, $\underline{\mathcal{E} := \{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}}$ es una base ortonormal de V.

Demostración. Se procede por inducción sobre k. En el paso inicial, $k = 1$, se debe observar que $\boldsymbol{x}_1 \neq \boldsymbol{0}$ (por la independencia lineal de $\mathcal{B}$), así que $\|\boldsymbol{x}_1\| \neq 0$. Ahora $\boldsymbol{e}_1$ es un múltiplo de $\boldsymbol{x}_1$ tal que $\|\boldsymbol{e}_1\| = 1$, y por eso $\{\boldsymbol{e}_1\}$ es una base ortonormal de $\mathrm{lin}\langle \boldsymbol{x}_1\rangle = \{\, c\boldsymbol{x}_1 : c \in \mathbb{R} \,\}$.

[2] El proceso aparece por primera vez en un libro de Laplace, *Théorie Analytique des Probabilités* (Paris, 1816); una versión modificada fue dada por el danés Jørgen Pedersen Gram en 1883. La versión moderna del algoritmo se debe a Erhard Schmidt, un estudiante de Hilbert, en un trabajo del año 1907.

Supóngase entonces que $\boldsymbol{e}_1, \dots, \boldsymbol{e}_{k-1}$ han sido elegidos por el procedimiento indicado, y que forman una base ortonormal de $\mathrm{lin}\langle \boldsymbol{x}_1, \dots, \boldsymbol{x}_{k-1} \rangle$. Para que $\boldsymbol{e}_1, \dots, \boldsymbol{e}_k$ sea una base ortonormal de $\mathrm{lin}\langle \boldsymbol{x}_1, \dots, \boldsymbol{x}_k \rangle$, basta comprobar que $\boldsymbol{e}_k \perp \boldsymbol{e}_j$ para $j < k$ y que $\|\boldsymbol{e}_k\| = 1$. Ahora si $j < k$, entonces

$$\begin{aligned} \langle \boldsymbol{e}_j, \boldsymbol{y}_k \rangle &= \Bigl\langle \boldsymbol{e}_j, \boldsymbol{x}_k - \sum_{i=1}^{k-1} \langle \boldsymbol{e}_i, \boldsymbol{x}_k \rangle \, \boldsymbol{e}_i \Bigr\rangle = \langle \boldsymbol{e}_j, \boldsymbol{x}_k \rangle - \sum_{i=1}^{k-1} \langle \boldsymbol{e}_i, \boldsymbol{x}_k \rangle \, \langle \boldsymbol{e}_j, \boldsymbol{e}_i \rangle \\ &= \langle \boldsymbol{e}_j, \boldsymbol{x}_k \rangle - \langle \boldsymbol{e}_j, \boldsymbol{x}_k \rangle = 0, \end{aligned}$$

ya que $\langle \boldsymbol{e}_j, \boldsymbol{e}_i \rangle = [\![j = i]\!]$. En consecuencia, cualquier múltiplo de $\boldsymbol{y}_k$ es también ortogonal a $\boldsymbol{e}_j$ cuando $j < k$. Basta entonces comprobar que $\boldsymbol{y}_k \neq \boldsymbol{0}$, para que se pueda dividir por el número positivo $\|\boldsymbol{y}_k\|$ y así definir $\boldsymbol{e}_k$ como un múltiplo de $\boldsymbol{y}_k$ que cumple $\|\boldsymbol{e}_k\| = 1$.

Ahora, si fuera $\boldsymbol{y}_k = \boldsymbol{0}$, $\boldsymbol{x}_k$ sería una combinación lineal de $\boldsymbol{e}_1, \dots, \boldsymbol{e}_{k-1}$ tal como se ve en el despliegue (5.8); y estos vectores $\boldsymbol{e}_j$ son a su vez combinaciones lineales de $\boldsymbol{x}_1, \dots, \boldsymbol{x}_{k-1}$. Pero en tal caso los vectores $\boldsymbol{x}_1, \dots, \boldsymbol{x}_k$ serían linealmente *dependientes* (por el Lema 2.15), lo cual es falso pues por hipótesis $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ es una base de V.

Se deduce que $\boldsymbol{y}_k \neq \boldsymbol{0}$. Luego $\boldsymbol{e}_k$ está bien definida; vale $\|\boldsymbol{e}_k\| = 1$; y $\boldsymbol{e}_k$ es ortogonal a $\mathrm{lin}\langle \boldsymbol{e}_1, \dots, \boldsymbol{e}_{k-1} \rangle$. □

Ejemplo 5.16. En el espacio vectorial $\mathbb{R}_3[t]$ de polinomios reales de grado $\leqslant 3$, el cual es un subespacio de $C[-1,1]$ (los polinomios son funciones continuas sobre este intervalo real), se puede usar el producto escalar del Ejemplo 5.3:

$$\langle p, q \rangle := \int_{-1}^{1} p(t) q(t) \, dt.$$

El juego de monomios $\{1, t, t^2, t^3\}$ es una base de $\mathbb{R}_3[t]$ – véase el Ejemplo 2.19 – y el algoritmo de Gram y Schmidt lo transforma en una base ortonormal $\{p_0, p_1, p_2, p_3\}$ de $\mathbb{R}_3[t]$, en donde cada $p_k(t)$ es un polinomio de grado k. En efecto:

$$\|1\|^2 = \langle 1, 1 \rangle = \int_{-1}^{1} 1 \, dt = 2; \qquad \|1\| = \sqrt{2}; \qquad p_0(t) \equiv \frac{1}{\sqrt{2}}\,.$$

En seguida se calcula:

$$\underline{q_1(t)} := t - \langle p_0, t \rangle p_0 = t - \frac{1}{\sqrt{2}} \int_{-1}^{1} \frac{t}{\sqrt{2}} \, dt = t - \frac{1}{2} \int_{-1}^{1} t \, dt = t - 0 = t.$$

$[\![$ Nótese que la integral $\int_{-a}^{a} g(t) \, dt = 0$ si $g(t)$ es una *función impar*. $]\!]$ Ahora

$$\|q_1\|^2 = \int_{-1}^{1} t^2 \, dt = 2 \int_{0}^{1} t^2 \, dt = \frac{2}{3} \quad \text{así que} \quad \underline{p_1(t)} = \frac{t}{\sqrt{2/3}} = \sqrt{\frac{3}{2}}\, t.$$

$[\![$ Nótese que la integral $\int_{-a}^{a} h(t) \, dt = 2 \int_{0}^{a} h(t) \, dt$ si $h(t)$ es una *función par*. $]\!]$

El siguiente paso pide

$$\begin{aligned} \underline{q_2(t)} := t^2 - \langle p_0, t^2 \rangle p_0 - \langle p_1, t^2 \rangle p_1 &= t^2 - \frac{1}{2} \int_{-1}^{1} t^2\,dt - \frac{3}{2}\,t \int_{-1}^{1} t^3\,dt \\ &= t^2 - \int_0^1 t^2\,dt - 0 = t^2 - \frac{1}{3}\,. \end{aligned}$$

De ahí,

$$\|q_2\|^2 = \int_{-1}^{1} \bigl(t^2 - \tfrac{1}{3}\bigr)^2\,dt = 2 \int_0^1 \bigl(t^4 - \tfrac{2}{3}t^2 + \tfrac{1}{9}\bigr)\,dt = 2\bigl(\tfrac{1}{5} - \tfrac{2}{9} + \tfrac{1}{9}\bigr) = \frac{8}{45},$$

y por ende

$$\|q_2\| = \frac{2}{3}\sqrt{\frac{2}{5}}, \qquad \underline{p_2(t)} = \frac{3}{2}\sqrt{\frac{5}{2}}\,q_2(t) = \frac{1}{2}\sqrt{\frac{5}{2}}\,(3t^2 - 1).$$

Finalmente, se calcula

$$\begin{aligned} \underline{q_3(t)} &:= t^3 - \langle p_0, t^3 \rangle p_0 - \langle p_1, t^3 \rangle p_1 - \langle p_2, t^3 \rangle p_2 \\ &= t^3 - \frac{1}{2} \int_{-1}^{1} t^3\,dt - \frac{3}{2}\,t \int_{-1}^{1} t^4\,dt - \frac{5}{8}\,(3t^2 - 1) \int_{-1}^{1} (3t^5 - t^3)\,dt \\ &= t^3 - 0 - 3t \int_0^1 t^4\,dt - 0 = t^3 - \frac{3}{5}\,t. \end{aligned}$$

Luego

$$\|q_3\|^2 = \int_{-1}^{1} \bigl(t^3 - \tfrac{3}{5}\,t\bigr)^2\,dt = 2 \int_0^1 \bigl(t^6 - \tfrac{6}{5}t^4 + \tfrac{9}{25}t^2\bigr)\,dt = 2\bigl(\tfrac{1}{7} - \tfrac{6}{25} + \tfrac{3}{25}\bigr) = \frac{8}{175},$$

y por ende

$$\|q_3\| = \frac{2}{5}\sqrt{\frac{2}{7}}, \qquad \underline{p_3(t)} = \frac{5}{2}\sqrt{\frac{7}{2}}\,q_2(t) = \frac{1}{2}\sqrt{\frac{7}{2}}\,(5t^3 - 3t).$$

Es costumbre eliminar los radicales en los polinomios finales, al tomar los múltiplos $\underline{P_k(t)} := \sqrt{2/(2k+1)}\,p_k(t)$. Entonces

$$P_0(t) = 1, \quad P_1(t) = t, \quad P_2(t) = \tfrac{1}{2}(3t^2 - 1), \quad P_3(t) = \tfrac{1}{2}(5t^3 - 3t).$$

Fíjese que estos **polinomios de Legendre** son alternadamente funciones pares e impares de la variable $t \in [-1, 1]$. $\Diamond$

La Proposición 5.15 produce la base ortonormal deseada mediante el algoritmo iterativo ilustrado en (5.8), que se detalla a continuación.

Algoritmo 5.1 (Ortogonalización de Gram y Schmidt)**.** Sea $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ una base de $\mathbb{R}^n$. Se construye $\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}$ de la manera indicada en la Proposición 5.15. Se organiza la base dada en *una matriz con columnas* $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$; el algoritmo produce la matriz cuyas columnas son $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$.

```
⟨ Ortogonalizar la base dada ⟩ ≡
⟨ Declarar el tipo: matriznporn ⟩
procedure GramSchmidt(A : matriznporn)
⟨ Declarar las variables locales: i, j, k : 1 .. n; r, t: real ⟩
begin  { GramSchmidt }
      for k ← 1 to n do  { cambiar la columna x_k a e_k }
         for j ← 1 to k − 1 do  { restar su componente e_j de x_k }
            t ← 0;  { inicialización }
            for i ← 1 to n do  { calcular ⟨e_j, x_k⟩ }
               t ← t + A[i, j] ∗ A[i, k]
            endfor;  { i }
            for i ← 1 to n do
               A[i, k] ← A[i, k] − t ∗ A[i, j];
            endfor;  { i }
         endfor  { j: se ha terminado y_k ← x_k }
         r ← 0;  { inicialización }
         for i ← 1 to n do  { calcular ||y_k||^2 }
            r ← r + A[i, k]^2;
         endfor;  { i }
         r ← sqrt(r);  { r = ||y_k|| }
         if r ⩽ 0 then ⟨ Interrumpir ⟩  { las columnas de A eran dependientes }
         else for i ← 1 to n do
                  A[i, k] ← A[i, k]/r
               endfor  { i: se ha terminado e_k ← y_k }
         endif
      endfor  { k: las columnas de A ya son ortonormales }
end;  { GramSchmidt }
```

Al combinar la Proposición 5.15 con la Proposición 2.29, también es posible completar una base ortonormal parcial.

Corolario 5.17. *Una base ortonormal $\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_m\}$ para un subespacio $W \leqslant V$ puede ser completado para obtener así una base ortonormal de V.*

Demostración. Los vectores $\boldsymbol{e}_1, \dots, \boldsymbol{e}_m$ son linealmente independientes y generan el subespacio W. Entonces, por la Proposición 2.29, se puede hallar otros vectores $\boldsymbol{x}_{m+1}, \dots, \boldsymbol{x}_n \in V$ tales que $\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_m, \boldsymbol{x}_{m+1}, \dots, \boldsymbol{x}_n\}$ sea una base de V (de tipo ordinario, no necesariamente ortonormal). Ahora aplíquese el algoritmo de Gram y Schmidt a esta base: *los primeros m vectores no sufren cambio alguno* y el resultado es una base ortonormal $\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}$ de V cuyos primeros m elementos son los vectores originales $\boldsymbol{e}_1, \dots, \boldsymbol{e}_m$. □

5.3. Productos escalares complejos

El segundo tipo de producto escalar que merece consideración se define sobre un espacio vectorial *complejo*. Se presentan muchas similitudes con el caso real, pero también algunas diferencias, debido a la presencia de *conjugados complejos* de escalares en $\mathbb{C}$.

En esta sección, se denotará un número complejo típico por $\underline{\alpha = s + it} \in \mathbb{C}$, donde $s, t \in \mathbb{R}$, e $i^2 = -1$. Su conjugado complejo es $\underline{\bar{\alpha} := s - it} \in \mathbb{C}$. Se debe recordar que

$$\bar{\alpha}\alpha = (s - it)(s + it) = s^2 + t^2 \geqslant 0,$$

con igualdad si y solo si $s = t = 0$, esto es, $\alpha = 0$ en $\mathbb{C}$. El número real no negativo

$$|\alpha| := \sqrt{\bar{\alpha}\alpha} = \sqrt{s^2 + t^2}$$

es el **valor absoluto** (o el *módulo*)[3] de α.

También en esta sección, V y W denotarán *espacios vectoriales sobre* $\mathbb{C}$ (espacios vectoriales complejos, espacios $\mathbb{C}$-vectoriales) de dimensión *finita*.

Definición 5.18. Un **producto escalar** (complejo) en V es una función que a cada par de vectores $\underline{\boldsymbol{z}, \boldsymbol{w} \in V}$ asocia un número complejo $\langle \boldsymbol{z}, \boldsymbol{w} \rangle \in \mathbb{C}$ tal que:

(a) $\langle \boldsymbol{z}, \boldsymbol{w} \rangle = \overline{\langle \boldsymbol{w}, \boldsymbol{z} \rangle}$ para todo $\boldsymbol{z}, \boldsymbol{w} \in V$;

(b) $\langle \boldsymbol{z}, \boldsymbol{w} + \boldsymbol{v} \rangle = \langle \boldsymbol{z}, \boldsymbol{w} \rangle + \langle \boldsymbol{z}, \boldsymbol{v} \rangle$ para todo $\boldsymbol{z}, \boldsymbol{w}, \boldsymbol{v} \in V$;

(c) $\langle \boldsymbol{z}, \alpha \boldsymbol{w} \rangle = \alpha \langle \boldsymbol{z}, \boldsymbol{w} \rangle$, pero $\langle \alpha \boldsymbol{z}, \boldsymbol{w} \rangle = \bar{\alpha} \langle \boldsymbol{z}, \boldsymbol{w} \rangle$ para todo $\boldsymbol{z}, \boldsymbol{w} \in V$, $\alpha \in \mathbb{C}$;

(d) $\langle \boldsymbol{z}, \boldsymbol{z} \rangle \geqslant 0$, con igualdad solo si $\boldsymbol{z} = \boldsymbol{0}$ en V.

Provisto de un producto escalar complejo, V se llama un **espacio hilbertiano**.[4] ◊

[3]Si $t = 0$, entonces $\alpha = s$ es real; su valor absoluto es $|\alpha| = |s| = \sqrt{s^2}$. De este modo las dos definiciones de valor absoluto (en $\mathbb{R}$ y en $\mathbb{C}$) son compatibles.

[4]Un espacio $\mathbb{C}$-vectorial infinitodimensional V provisto de un producto escalar se suele llamar espacio *pre*hilbertiano; solo si V cumple un quinto requisito de "completitud", recibe el nombre de espacio hilbertiano, o *espacio de Hilbert*. (Un espacio finitodimensional es automáticamente completo.)

Las propiedades (b) y (c) dicen que $\langle \boldsymbol{z}, \boldsymbol{w} \rangle$ es $\mathbb{C}$-lineal en la *segunda* variable $\boldsymbol{w}$, pero sólo es semilineal en la *primera* variable $\boldsymbol{z}$. Una aplicación $T \colon V \to W$ entre espacios vectoriales complejos se llama **semilineal** (o *antilineal*) si

$$T(\alpha \boldsymbol{x} + \beta \boldsymbol{y}) = \bar{\alpha}\, T(\boldsymbol{x}) + \bar{\beta}\, T(\boldsymbol{y})$$

para $\boldsymbol{x}, \boldsymbol{y} \in V$, $\alpha, \beta \in \mathbb{C}$. Una función escalar de dos variables vectoriales que obedece las propiedades (b) y (c) se llama una **forma sesquilineal**[5] sobre V.

Ejemplo 5.19. Si $V = \mathbb{C}^n$, el producto escalar de dos vectores (de columna) es

$$\boxed{\langle \boldsymbol{z}, \boldsymbol{w} \rangle \equiv \bar{\boldsymbol{z}} \cdot \boldsymbol{w} \equiv \bar{\boldsymbol{z}}^{\mathsf{t}} \boldsymbol{w} := \bar{z}_1 w_1 + \bar{z}_2 w_2 + \cdots + \bar{z}_n w_n.} \tag{5.9}$$

Nótese que $\langle \boldsymbol{z}, \boldsymbol{z} \rangle = |z_1|^2 + \cdots + |z_n|^2 \geqslant 0$, con igualdad solo si cada $z_i = 0$ en $\mathbb{C}$, solo si $\boldsymbol{z} = \boldsymbol{0}$ en $\mathbb{C}^n$. $\Diamond$

⟦ Algunos autores emplean una variante de la condición (c) arriba, en donde $\langle -, - \rangle$ es lineal en la primera variable y semilineal en la segunda. Esa variante no es recomendable a la hora de calcular con matrices: en la fórmula (5.9) *es el vector de fila* $\boldsymbol{z}^{\mathsf{t}}$ *que recibe la conjugación compleja*; la fórmula debe ser lineal en el vector de columna $\boldsymbol{w}$. ⟧

Ejemplo 5.20. Sea $V := C([a, b]; \mathbb{C})$ es el espacio de funciones continuas $f \colon [a, b] \to \mathbb{C}$, el cual es un espacio vectorial complejo (infinitodimensional). Defínase

$$\langle f, g \rangle := \int_a^b \overline{f(t)}\, g(t)\, dt.$$

No es difícil verificar que esta fórmula cumple los requisitos de la Definición 5.18. $\Diamond$

Definición 5.21. Si $A = [\alpha_{ij}]$ es una matriz compleja $m \times n$, su **matriz conjugada** es $\overline{A} := [\bar{\alpha}_{ij}]$, otra matriz $m \times n$ formada al tomar el conjugado de cada uno de sus entradas.

La transpuesta $A^{\mathsf{t}} = [\alpha_{ji}]$ es una matriz compleja $n \times m$, ya definida. Al combinar estas dos modificaciones, se obtiene su **matriz adjunta** $A^* := [\bar{\alpha}_{ji}]$. Nótese que A^* (también llamado *conjugado hermítico* o *conjugado hermitiano* de A) es una matriz compleja $n \times m$. Obviamente, $A^* = \overline{A^{\mathsf{t}}} = (\overline{A})^{\mathsf{t}}$.

Se dice que una matriz cuadrada $A \in M_n(\mathbb{C})$ es **hermítica** si $A^* = A$. (Las matrices hermíticas juegan un papel análogo al de las matrices simétricas en el caso real.) $\Diamond$

Se ve que $A \mapsto \overline{A}$ es un operador *semilineal* sobre $M_{m \times n}(\mathbb{C})$; y es fácil verificar que $A \mapsto A^*$ es una aplicación *semilineal* de $M_{m \times n}(\mathbb{C})$ en $M_{n \times m}(\mathbb{C})$. La primera transformación conserva el orden de multiplicación, pero la segunda lo revierte:

$$\overline{AB} = \overline{A}\,\overline{B}, \quad \text{mientras} \quad (AB)^* = B^* A^*.$$

[5] El prefijo *sesqui-* significa "$\frac{3}{2}$ veces" (del latín).

Ejemplo 5.22. Si $A = [\alpha_{ij}]$, $B = [\beta_{ij}]$ en $M_{m\times n}(\mathbb{C})$, la receta

$$\langle A, B\rangle := \underline{\operatorname{tr}(A^*B)} = \sum_{i=1}^{m}\sum_{j=1}^{n} \bar{\alpha}_{ij}\beta_{ij}$$

define un producto escalar complejo sobre $M_{m\times n}(\mathbb{C})$: fíjese que A^*B es una matriz compleja $n \times n$. Además

$$\langle A, A\rangle = \sum_{i=1}^{m}\sum_{j=1}^{n} |\alpha_{ij}|^2 \geqslant 0,$$

con igualdad solo si cada $\alpha_{ij} = 0$ en $\mathbb{C}$, solo si $A = 0$ en $M_{m\times n}(\mathbb{C})$. ◊

► La **norma** de un vector $z \in V$ se define por $\|z\| := \sqrt{\langle z, z\rangle}$ (igual que en el caso real). La desigualdad de Schwarz (5.3) también se verifica en el caso complejo, pero la prueba requiere una modificación.

Antes de examinar esa modificación, obsérvese que un número complejo de valor absoluto 1, $\alpha = s + it$ con $s^2 + t^2 = 1$, se puede escribir como $\underline{\alpha = \cos\theta + i \operatorname{sen}\theta}$ para cierto ángulo θ. Además, la *fórmula de de Moivre*:

$$(\cos\theta + i\operatorname{sen}\theta)^n = \cos(n\theta) + i\operatorname{sen}(n\theta) \quad \text{para todo} \quad n \in \mathbb{Z}$$

permite escribir $\underline{e^{i\theta} := \cos\theta + i\operatorname{sen}\theta}$, para que $(e^{i\theta})^n = e^{in\theta}$. ⟦ Esta es una extensión formal de la ley de exponentes para potencias reales a exponentes imaginarios. ⟧

Si $\alpha \neq 0$ en $\mathbb{C}$, se define $r := |\alpha| > 0$; con esto, α/r tiene valor absoluto 1, pues $|\alpha/r| = |\alpha|/|r| = r/r = 1$. Entonces se puede escribir[6]

$$\alpha = r(\cos\theta + i\operatorname{sen}\theta) = r\, e^{i\theta}.$$

Esta es la llamada *forma polar* del número complejo α.

Proposición 5.23. *Si V es un espacio vectorial hilbertiano, su norma $\|z\| := \sqrt{\langle z, z\rangle}$ verifica la desigualdad de Schwarz* (5.3)*:*

$$\big|\langle z, \boldsymbol{w}\rangle\big| \leqslant \|z\|\,\|\boldsymbol{w}\| \quad \textit{para todo} \quad z, \boldsymbol{w} \in V,$$

con igualdad si y solo si $z, \boldsymbol{w}$ son proporcionales.

Demostración. Se verifica esta relación con igualdad si $z = \mathbf{0}$ o si $\boldsymbol{w} = \mathbf{0}$; supóngase, entonces, que $z \neq \mathbf{0}$ y $\boldsymbol{w} \neq \mathbf{0}$ en V.

[6] Si $\alpha = 0$, la forma polar $\alpha = r\,e^{i\theta}$ también es válida con $r = 0$ y θ arbitraria.

Caso 1: Si $\langle z, w\rangle \in \mathbb{R}$, la prueba de la Proposición 5.5 sigue válida. Para $t \in \mathbb{R}$, $f(t) := \|z + tw\|^2 \geqslant 0$ es una función cuadrática $f(t) = at^2 + bt + c$ con $a > 0$ y con $a, b, c \in \mathbb{R}$. Entonces el discriminante cumple $b^2 - 4ac \neq 0$ y se deduce la desigualdad de Schwarz como antes.

Caso 2: Si $\langle z, w\rangle \notin \mathbb{R}$, se puede escribir este número complejo en forma polar, al poner $\langle z, w\rangle =: re^{i\theta}$ con $r > 0$, $\theta \in \mathbb{R}$. Defínase $\underline{v := e^{-i\theta}w} \in V$, y fíjese que

$$\langle z, v\rangle = \langle z, e^{-i\theta}w\rangle = e^{-i\theta}\langle z, w\rangle = e^{-i\theta}(re^{i\theta}) = r \in \mathbb{R}.$$

Ahora se obtiene $\bigl|\langle z, v\rangle\bigr| \leqslant \|z\|\,\|v\|$ por el Caso 1.

Las normas de v y w son iguales, porque[7]

$$\|v\|^2 = \langle v, v\rangle = \langle e^{-i\theta}w, e^{-i\theta}w\rangle = e^{+i\theta}e^{-i\theta}\langle w, w\rangle = \|w\|^2.$$

Entonces, al ser $|e^{i\theta}| = 1$, se obtiene

$$\bigl|\langle z, w\rangle\bigr| = \bigl|\langle z, e^{i\theta}v\rangle\bigr| = \bigl|e^{i\theta}\langle z, v\rangle\bigr| = \bigl|\langle z, v\rangle\bigr| \leqslant \|z\|\,\|v\| = \|z\|\,\|w\|.$$

En el caso de igualdad, los tres vectores z, v, w son proporcionales. $\square$

► Como $\langle z, w\rangle$ es complejo, no es apropiado hablar de un ángulo entre dos vectores; aun así, se dice que dos vectores $z, w \in V$ son **ortogonales** si $\langle z, w\rangle = 0$.

Se define el **complemento ortogonal** $M^\perp$ de un subespacio $M \leqslant V$ por la misma fórmula (5.4) del caso real:

$$M^\perp := \{\, w \in V : \langle w, z\rangle = 0, \text{ para todo } z \in M \,\}.$$

La Definición 5.12 de **base ortonormal** $\{e_1, \dots, e_n\}$ de V se aplica sin cambio en el caso complejo.

► La desigualdad de Schwarz no es la única relación importante entre el producto escalar y la norma. Resulta que se puede *recuperar* el producto escalar a partir de la norma, mediante la siguiente **fórmula de polarización**.

Lema 5.24. *Si V es un espacio hilbertiano, la norma determina el producto escalar por esta fórmula:*

$$4\,\langle w, z\rangle = \|z + w\|^2 - \|z - w\|^2 + i\,\|z + i\,w\|^2 - i\,\|z - i\,w\|^2. \tag{5.10}$$

(No es necesario verificar que el lado derecho cumple las propiedades de la Definición 5.18.)

[7] El conjugado complejo de $e^{-i\theta} = \cos\theta - i\,\mathrm{sen}\,\theta$ es $e^{+i\theta} = \cos\theta + i\,\mathrm{sen}\,\theta$.

Demostración. De la definición de la norma, se obtiene

$$\|z \pm w\|^2 = \langle z \pm w, z \pm w\rangle = \|z\|^2 \pm \langle z, w\rangle \pm \langle w, z\rangle + \|w\|^2.$$

Por lo tanto,

$$\|z + w\|^2 - \|z - w\|^2 = 2\langle z, w\rangle + 2\langle w, z\rangle.$$

De manera similar, se ve que

$$\|z \pm iw\|^2 = \langle z \pm iw, z \pm iw\rangle = \|z\|^2 \pm i\langle z, w\rangle \mp i\langle w, z\rangle + \|w\|^2,$$

y de ahí se obtiene

$$i\|z + iw\|^2 - i\|z - iw\|^2 = -2\langle z, w\rangle + 2\langle w, z\rangle.$$

La fórmula (5.10) sigue directamente. □

► El algoritmo de Gram y Schmidt (y la Proposición 5.15) se pueden aplicar sin cambio en el casos de espacios vectoriales complejos con productos escalares (espacios hilbertianos). Sólo es necesario aclarar que en la fórmula (5.8), los coeficientes $\langle \boldsymbol{e}_j, \boldsymbol{x}_k\rangle$ que allí aparecen en la expansión del vector $\boldsymbol{y}_k$ no deben cambiarse por $\langle \boldsymbol{x}_k, \boldsymbol{e}_j\rangle$, debido a la asimetría del producto escalar complejo.

► Una diferencia significativa entre los dos casos emerge cuando se considera *el espacio dual de un espacio hilbertiano*. Se procede en paralelo con la Definición 5.13. Si V es un espacio hilbertiano de dimensión finita n, con una base ortonormal $\mathcal{E} = \{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}$, la fórmula (5.5) define un juego de formas lineales $\mathcal{F} := \{f_1, \dots, f_n\} \subset V^*$ por

$$f_k(\boldsymbol{x}) := \langle \boldsymbol{e}_k, \boldsymbol{x}\rangle \quad \text{para} \quad k = 1, \dots, n. \tag{5.5$'$}$$

Este conjunto $\mathcal{F}$ es la *base dual* a $\mathcal{E}$, del espacio vectorial complejo, puesto que $f_k(\boldsymbol{e}_j) = 0$ si $j \neq k$ y $f_k(\boldsymbol{e}_k) = 1$.

Definición 5.25. Cuando V es un espacio hilbertiano con una base ortonormal $\mathcal{E} = \{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}$, se define una aplicación *semilineal* $J \colon V \to V^*$ por la correspondencia $J(\boldsymbol{e}_k) := f_k$. Concretamente,

$$J(\boldsymbol{y}) \equiv J\bigl(\textstyle\sum_{k=1}^n y_k \boldsymbol{e}_k\bigr) := \textstyle\sum_{k=1}^n \bar{y}_k f_k. \tag{5.6a$'$}$$

De esta manera, se obtiene

$$J(\boldsymbol{y})(\boldsymbol{x}) = \sum_{k=1}^n \bar{y}_k\, f_k(\boldsymbol{x}) = \sum_{k=1}^n \bar{y}_k x_k = \langle \boldsymbol{y}, \boldsymbol{x}\rangle. \tag{5.6b$'$}$$

Al comparar estas fórmulas con las de (5.6), se observa que cada $J(\boldsymbol{y}) \in V^*$ es la forma lineal $\underline{\boldsymbol{x} \mapsto \langle \boldsymbol{y}, \boldsymbol{x}\rangle}$, al igual que antes (pero ahora con valores en $\mathbb{C}$). ◊

En la Definición 5.25, se puede notar que la semilinealidad de J es necesario (y suficiente) para obtener el mismo resultado final. También, se nota que ese resultado $J(\boldsymbol{y})\colon \boldsymbol{x} \mapsto \langle \boldsymbol{y}, \boldsymbol{x}\rangle$ no depende de cuál base ortonormal $\mathcal{E}$ de V se usa para definir la J.

► La semilinealidad de J también modifica el tratamiento de la *transpuesta* de una aplicación lineal: la fórmula (5.7b) ya no es aplicable en el caso complejo. En su lugar, se debe introducir otro concepto, el de una aplicación lineal *adjunta*.

Definición 5.26. Sea $S\colon V \to W$ una aplicación lineal entre dos espacios hilbertianos V y W. En este caso, la Definición 5.25 determina dos aplicaciones semilineales $J\colon V \to V^*$ y $J'\colon W \to W^*$, definidas por:

$$J(\boldsymbol{z}) : \boldsymbol{x} \mapsto \langle \boldsymbol{z}, \boldsymbol{x}\rangle_V \quad \text{y} \quad J'(\boldsymbol{y}) : \boldsymbol{y} \mapsto \langle \boldsymbol{y}, \boldsymbol{w}\rangle_W$$

si $\boldsymbol{x}, \boldsymbol{z} \in V$ y si $\boldsymbol{y}, \boldsymbol{w} \in W$. Ahora se *define* una aplicación lineal $\underline{S^*\colon W \to V}$ por

$$S^* := J^{-1} S^{\mathsf{t}} J', \qquad \text{esto es,} \qquad \begin{array}{ccc} W & \xrightarrow{S^*} & V \\ {\scriptstyle J'}\big\downarrow & & \big\uparrow{\scriptstyle J^{-1}} \\ W^* & \xrightarrow{S^{\mathsf{t}}} & V^* \end{array}$$

Esta es la aplicación **adjunta** de $S\colon V \to W$. Se calcula, por analogía con (5.7a):

$$\langle J^{-1} S^{\mathsf{t}} J'(\boldsymbol{y}), \boldsymbol{x}\rangle_V = S^{\mathsf{t}} J'(\boldsymbol{y})(\boldsymbol{x}) := J'(\boldsymbol{y})(S(\boldsymbol{x})) = \langle \boldsymbol{y}, S(\boldsymbol{x})\rangle_W \tag{5.11a}$$

En este caso, esto se simplifica en la fórmula clave:

$$\boxed{\langle S^*(\boldsymbol{y}), \boldsymbol{x}\rangle = \langle \boldsymbol{y}, S(\boldsymbol{x})\rangle} \quad \text{para todo} \quad \boldsymbol{x} \in V,\ \boldsymbol{y} \in W. \tag{5.11b}$$

En resumen: *una aplicación lineal se traslada de un lado a otro de un producto escalar* ***complejo*** *al poner su* ***adjunta*** *al otro lado.* ◊

Al tomar el conjugado complejo de ambos lados de (5.11b), se obtiene una fórmula equivalente: $\langle S(\boldsymbol{x}), \boldsymbol{y}\rangle = \langle \boldsymbol{x}, S^*(\boldsymbol{y})\rangle$. De ahí se deduce que la transformación (semilineal) $S \mapsto S^*$ es *involutivo*: es decir, que $\underline{(S^*)^* = S}$.

La transformación $S \mapsto S^*$ también revierte el orden de composición de aplicaciones. Si $R\colon W \to Z$ es otra aplicación lineal entre espacios hilbertianos, es inmediato que

$$\langle (RS)^*(\boldsymbol{z}), \boldsymbol{x}\rangle = \langle \boldsymbol{z}, RS(\boldsymbol{x})\rangle = \langle R^*(\boldsymbol{z}), S(\boldsymbol{x})\rangle = \langle S^*(R^*(\boldsymbol{z})), \boldsymbol{x}\rangle$$

para $\boldsymbol{x} \in V$, $\boldsymbol{z} \in Z$, así que $\underline{(RS)^* = S^* R^*}\colon Z \to V$.

► En el caso de aplicaciones matriciales $T_A\colon \mathbb{C}^n \to \mathbb{C}^m$, se puede comprobar que la matriz de $(T_A)^*$ – con respecto a las bases ortonormales estándares de $\mathbb{C}^n$ y $\mathbb{C}^m$ – no es otra cosa que la matriz adjunta A^* de A.

5.4. Matrices ortogonales y positivas

En esta sección se consideran algunas clases de matrices cuadradas, con entradas reales o complejos, que son especialmente compatibles con los productos escalares usuales en $\mathbb{R}^n$ ó $\mathbb{C}^n$. Es prudente empezar con el caso real.

Definición 5.27. Se dice que una matriz cuadrada $A \in M_n(\mathbb{R})$ es una **matriz ortogonal** si sus columnas forman una base ortonormal de $\mathbb{R}^n$. ◊

Proposición 5.28. *Una matriz $A \in M_n(\mathbb{R})$ es ortogonal si y solo si $A^{\mathsf{t}}A = 1_n$.*

Demostración. El producto escalar usual de $\mathbb{R}^n$ es el producto punto: $\langle \boldsymbol{a}, \boldsymbol{b} \rangle = \boldsymbol{a} \cdot \boldsymbol{b} \in \mathbb{R}$. Como la transpuesta de la columna $\boldsymbol{a}_i$ de una matriz A es la fila número i de A^{t}, la entrada (i, j) del producto $A^{\mathsf{t}}A$ es $\boldsymbol{a}_i^{\mathsf{t}}\, \boldsymbol{a}_j = \boldsymbol{a}_i \cdot \boldsymbol{a}_j$.

Por definición, A es ortogonal si y solo si $\boldsymbol{a}_i \cdot \boldsymbol{a}_j = 0$ para $i \neq j$; $\boldsymbol{a}_j \cdot \boldsymbol{a}_j = \|\boldsymbol{a}_j\|^2 = 1$ para $j = 1, \dots, n$. Esto es equivalente a la afirmación que las matrices $A^{\mathsf{t}}A$ y 1_n tienen las mismas entradas, o sea, son iguales: $A^{\mathsf{t}}A = 1_n$. □

El *rango* de una matriz ortogonal $A \in M_n(\mathbb{R})$ es $r(A) = n$, porque sus n columnas son, en particular, linealmente independientes. Como $n(A) + r(A) = n$ por el teorema de rango y nulidad, su nulidad es $n(A) = 0$. Entonces la aplicación lineal T_A es biyectiva y (lo que es lo mismo) *la matriz A es invertible*. Por lo tanto, el resultado de la Proposición 5.28 puede ser reformulado de esta manera:

$$\boxed{A \text{ es ortogonal} \iff A^{\mathsf{t}}A = 1_n \iff A^{-1} = A^{\mathsf{t}} \iff AA^{\mathsf{t}} = 1_n\,.} \tag{5.12}$$

Nótese una consecuencia: la matriz A^{t} es también ortogonal.

► Como las columnas $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$ de una matriz invertible X son linealmente independientes, el algoritmo de Gram y Schmidt las convierte en las columnas $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$ de alguna matriz ortogonal Q. Si se escribe el k-ésimo paso del algoritmo – del despliegue (5.8) – como sigue:

$$\boldsymbol{x}_k = \langle \boldsymbol{e}_1, \boldsymbol{x}_k \rangle\, \boldsymbol{e}_1 + \langle \boldsymbol{e}_2, \boldsymbol{x}_k \rangle\, \boldsymbol{e}_2 + \cdots + \langle \boldsymbol{e}_{k-1}, \boldsymbol{x}_k \rangle\, \boldsymbol{e}_{k-1} + \|\boldsymbol{y}_k\|\, \boldsymbol{e}_k,$$

se nota que la columna $\boldsymbol{x}_k$ es una combinación lineal de las primeras k columnas de Q. Sea $R \in M_n(\mathbb{R})$ la matriz triangular superior que recoge estos coeficientes:

$$r_{kk} := \|\boldsymbol{y}_k\| > 0, \qquad r_{jk} := \langle \boldsymbol{e}_j, \boldsymbol{x}_k \rangle = \boldsymbol{e}_j \cdot \boldsymbol{x}_k \quad \text{para} \quad j < k,$$

se puede ver que $X = QR$. En otras palabras, el algoritmo de Gram y Schmidt efectúa una *factorización matricial* de una matriz invertible en el producto de una matriz ortogonal y una matriz triangular superior (también invertible).

► En el caso complejo, hay un concepto muy similar, donde el papel de la matriz transpuesta A^{t} es tomado por la *matriz adjunta*.

Definición 5.29. Se dice que una matriz cuadrada $U \in M_n(\mathbb{C})$ es una **matriz unitaria** si sus columnas forman una base ortonormal de $\mathbb{C}^n$. ◊

Proposición 5.30. *Una matriz $U \in M_n(\mathbb{C})$ es unitaria si y solo si $\underline{U^*U = 1_n}$.*

Demostración. El producto escalar usual de $\mathbb{C}^n$ es $\langle \boldsymbol{u}, \boldsymbol{v} \rangle = \bar{\boldsymbol{u}} \cdot \boldsymbol{v} \in \mathbb{C}$. Como *la transpuesta conjugada* de la columna $\boldsymbol{u}_i$ de una matriz U es la fila número i de U^*, la entrada (i, j) del producto U^*U es $\bar{\boldsymbol{u}}_i^{\mathsf{t}} \boldsymbol{u}_j = \bar{\boldsymbol{u}}_i \cdot \boldsymbol{u}_j$.

Por definición, U es unitaria si y solo si $\bar{\boldsymbol{u}}_i \cdot \boldsymbol{u}_j = 0$ para $i \neq j$; $\bar{\boldsymbol{u}}_j \cdot \boldsymbol{u}_j = \|\boldsymbol{u}_j\|^2 = 1$ para $j = 1, \dots, n$. Esto es equivalente a la afirmación que las matrices U^*U y 1_n tienen las mismas entradas, o sea, son iguales: $U^*U = 1_n$. □

Una matriz unitaria $U \in M_n(\mathbb{C})$ tiene rango $r(U) = n$ y nulidad $n(U) = 0$, al igual que las matrices ortogonales $n \times n$. Por lo tanto, *la matriz U es invertible*. Se obtiene así un resultado similar a (5.12) para matrices unitarias:

$$\boxed{U \text{ es unitaria} \iff U^*U = 1_n \iff U^{-1} = U^* \iff UU^* = 1_n\,.} \tag{5.13}$$

En consecuencia, la matriz U^* es también unitaria.

► Volviendo al caso real, sea $A \in M_{m\times n}(\mathbb{R})$ una matriz rectangular cualquiera. Entonces $A^{\mathsf{t}} \in M_{n\times m}(\mathbb{R})$ y $\underline{A^{\mathsf{t}}A}$ es una matriz cuadrada $n \times n$ (no necesariamente igual a 1_n). Pero al menos $A^{\mathsf{t}}A$ es una matriz *simétrica*:

$$(A^{\mathsf{t}}A)^{\mathsf{t}} = A^{\mathsf{t}}A^{\mathsf{tt}} = A^{\mathsf{t}}A.$$

Proposición 5.31. *Si $A \in M_{m\times n}(\mathbb{R})$, las matrices A y $A^{\mathsf{t}}A$ tienen* el mismo rango*: vale $\underline{r(A^{\mathsf{t}}A) = r(A)}$.*

Demostración. Si $\boldsymbol{x} \in \mathbb{R}^n$, entonces $A\boldsymbol{x} = \mathbf{0} \implies A^{\mathsf{t}}A\boldsymbol{x} = \mathbf{0}$ obviamente. Pero además,

$$\begin{aligned} A^{\mathsf{t}}A\boldsymbol{x} = \mathbf{0} &\implies \boldsymbol{x}^{\mathsf{t}}A^{\mathsf{t}}A\boldsymbol{x} = 0 \\ &\implies (A\boldsymbol{x})^{\mathsf{t}}A\boldsymbol{x} = 0 \implies \|A\boldsymbol{x}\|^2 = 0 \implies A\boldsymbol{x} = \mathbf{0}. \end{aligned}$$

De ahí se ve que $A\boldsymbol{x} = \mathbf{0}$ *si y solo si* $A^{\mathsf{t}}A\boldsymbol{x} = \mathbf{0}$. En otras palabras, se ha mostrado que $\ker T_A = \ker T_{A^{\mathsf{t}}A}$. Luego $\underline{n(A) = n(A^{\mathsf{t}}A)}$.

Por el teorema de rango y nulidad, $r(A^{\mathsf{t}}A) = n - n(A^{\mathsf{t}}A) = n - n(A) = r(A)$. □

En consecuencia, $A^{\mathsf{t}}A$ es invertible cuando el rango de A es n.

Corolario 5.32. *Si $A \in M_{m\times n}(\mathbb{R})$, las matrices A y AA^{t} tienen el mismo rango: vale $\underline{r(AA^{\mathsf{t}}) = r(A)}$.*

Demostración. Se puede aplicar la Proposición 5.31 anterior a la matriz A^{t} en vez de A. Esto muestra que $r(AA^{\mathsf{t}}) = r(A^{\mathsf{t}})$. Pero se sabe que $r(A^{\mathsf{t}}) = r(A)$, también por el teorema de rango y nulidad (Teorema 4.24). □

Definición 5.33. Una matriz simétrica $A = A^{\mathsf{t}} \in M_n(\mathbb{R})$ se llama una **matriz positiva** (*semidefinida*) si se cumple

$$\boxed{\langle \boldsymbol{x}, A\boldsymbol{x}\rangle \geqslant 0 \quad \text{para todo} \quad \boldsymbol{x} \in \mathbb{R}^n} \tag{5.14}$$

o bien, lo que es lo mismo: $\boldsymbol{x}^{\mathsf{t}}A\boldsymbol{x} \geqslant 0$ para todo $\boldsymbol{x} \in \mathbb{R}^n$.

Si esta desigualdad es estricta para vectores no nulos: $\underline{\langle \boldsymbol{x}, A\boldsymbol{x}\rangle > 0}$ para $\boldsymbol{x} \neq \mathbf{0}$, se dice que A es **definida positiva**. ◊

⟦ Fíjese que la matriz nula 0 es positiva (semidefinida), según esta definición. ¿Por qué no usar el término *matriz no negativa* en vez de “positiva”? Desafortunadamente, ese término ya está reservada para un concepto diferente: A es *no negativa* si todas sus entradas cumplen $a_{ij} \geqslant 0$. ⟧

Ejemplo 5.34. Si $\underline{A = B^{\mathsf{t}}B}$ para alguna matriz $B \in M_n(\mathbb{R})$, entonces A es una matriz positiva. En efecto,

$$\boldsymbol{x}^{\mathsf{t}}A\boldsymbol{x} = \boldsymbol{x}^{\mathsf{t}}B^{\mathsf{t}}B\boldsymbol{x} = (B\boldsymbol{x})^{\mathsf{t}}B\boldsymbol{x} = \|B\boldsymbol{x}\|^2 \geqslant 0. \tag{5.15}$$

Del mismo modo, cada matriz de la forma $\underline{A = CC^{\mathsf{t}}}$ es positiva: tómese $B := C^{\mathsf{t}}$. ◊

Los valores $\langle \boldsymbol{x}, A\boldsymbol{x}\rangle = \boldsymbol{x}^{\mathsf{t}}A\boldsymbol{x}$ de una matriz simétrica real tienen un interés propio, como ejemplifica el lema siguiente.

Lema 5.35. *Una matriz simétrica real $A = A^{\mathsf{t}} \in M_n(\mathbb{R})$ es nula (esto es, $A = 0$) si y solo si $\langle \boldsymbol{x}, A\boldsymbol{x}\rangle = 0$ para todo $\boldsymbol{x} \in \mathbb{R}^n$.*

Demostración. Es obvio que $\langle \boldsymbol{x}, 0\boldsymbol{x}\rangle = 0$ para todo $\boldsymbol{x} \in \mathbb{R}^n$.

Supóngase, entonces, que $A = A^{\mathsf{t}} \in M_n(\mathbb{R})$ cumple $\langle \boldsymbol{x}, A\boldsymbol{x}\rangle = 0$ para todo $\boldsymbol{x}$. Si $\boldsymbol{x}$, $\boldsymbol{y}$ son dos vectores en $\mathbb{R}^n$, resulta que

$$\langle \boldsymbol{y}, A\boldsymbol{x}\rangle + \langle \boldsymbol{x}, A\boldsymbol{y}\rangle = \langle (\boldsymbol{x}+\boldsymbol{y}), A(\boldsymbol{x}+\boldsymbol{y})\rangle - \langle \boldsymbol{x}, A\boldsymbol{x}\rangle - \langle \boldsymbol{y}, A\boldsymbol{y}\rangle = 0.$$

Como A es simétrico, se obtiene

$$\langle \boldsymbol{y}, A\boldsymbol{x}\rangle + \langle \boldsymbol{x}, A\boldsymbol{y}\rangle = \langle \boldsymbol{y}, A\boldsymbol{x}\rangle + \langle A\boldsymbol{y}, \boldsymbol{x}\rangle = \langle \boldsymbol{y}, A\boldsymbol{x}\rangle + \langle \boldsymbol{y}, A^{\mathsf{t}}\boldsymbol{x}\rangle = 2\langle \boldsymbol{y}, A\boldsymbol{x}\rangle,$$

y se deduce que $\langle \boldsymbol{y}, A\boldsymbol{x}\rangle = 0$ para todo $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$.

En particular, tómese $\boldsymbol{y} := A\boldsymbol{x}$. Entonces $\|A\boldsymbol{x}\|^2 = \langle A\boldsymbol{x}, A\boldsymbol{x}\rangle = 0$, así que $A\boldsymbol{x} = \mathbf{0}$ para todo $\boldsymbol{x} \in \mathbb{R}^n$, lo cual implica $A = 0$. □

Proposición 5.36. *Una matriz cuadrada $A \in M_n(\mathbb{R})$ es definida positiva si y solo si es positiva e invertible.*

Demostración. Si A es definida positiva: $\boldsymbol{x}^{\mathsf{t}}A\boldsymbol{x} > 0$ para todo $\boldsymbol{x} \neq \boldsymbol{0}$ en $\mathbb{R}^n$, entonces

$$A\boldsymbol{x} = \boldsymbol{0} \implies \boldsymbol{x}^{\mathsf{t}}A\boldsymbol{x} = 0 \implies \boldsymbol{x} = \boldsymbol{0},$$

así que $\ker T_A = \{\boldsymbol{0}\}$ y $n(A) = \dim\ker T_A = 0$. Como $r(A) = n - n(A) = n$ por el teorema de rango y nulidad, se sigue que A es invertible.

Inversamente, si A es positiva e invertible, $\boldsymbol{x}^{\mathsf{t}}A\boldsymbol{x} \geqslant 0$ (al menos) para $\boldsymbol{x} \in \mathbb{R}^n$. Defínase una función $\langle\!\langle -, - \rangle\!\rangle$ de dos vectores en $\mathbb{R}^n$ por:

$$\langle\!\langle \boldsymbol{x}, \boldsymbol{y} \rangle\!\rangle := \boldsymbol{x}^{\mathsf{t}}A\boldsymbol{y}. \tag{5.16}$$

Esta función cumple la Definición 5.1 de un producto escalar real, con excepción parcial de la propiedad (d):

$$\begin{aligned}
\langle\!\langle \boldsymbol{y}, \boldsymbol{x} \rangle\!\rangle &= \boldsymbol{y}^{\mathsf{t}}A\boldsymbol{x} = \boldsymbol{x}^{\mathsf{t}}A^{\mathsf{t}}\boldsymbol{y} = \boldsymbol{x}^{\mathsf{t}}A\boldsymbol{y} = \langle\!\langle \boldsymbol{x}, \boldsymbol{y} \rangle\!\rangle, \\
\langle\!\langle \boldsymbol{x}, \boldsymbol{y} + \boldsymbol{z} \rangle\!\rangle &= \boldsymbol{x}^{\mathsf{t}}A(\boldsymbol{y} + \boldsymbol{z}) = \boldsymbol{x}^{\mathsf{t}}A\boldsymbol{y} + \boldsymbol{x}^{\mathsf{t}}A\boldsymbol{z} = \langle\!\langle \boldsymbol{x}, \boldsymbol{y} \rangle\!\rangle + \langle\!\langle \boldsymbol{x}, \boldsymbol{z} \rangle\!\rangle, \\
\langle\!\langle \boldsymbol{x}, c\boldsymbol{y} \rangle\!\rangle &= \boldsymbol{x}^{\mathsf{t}}A(c\boldsymbol{y}) = c\,\boldsymbol{x}^{\mathsf{t}}A\boldsymbol{y} = c\,\langle\!\langle \boldsymbol{x}, \boldsymbol{y} \rangle\!\rangle, \\
\langle\!\langle c\boldsymbol{x}, \boldsymbol{y} \rangle\!\rangle &= (c\boldsymbol{x})^{\mathsf{t}}A\boldsymbol{y} = c\,\boldsymbol{x}^{\mathsf{t}}A\boldsymbol{y} = c\,\langle\!\langle \boldsymbol{x}, \boldsymbol{y} \rangle\!\rangle, \\
\langle\!\langle \boldsymbol{x}, \boldsymbol{x} \rangle\!\rangle &= \boldsymbol{x}^{\mathsf{t}}A\boldsymbol{x} \geqslant 0.
\end{aligned}$$

Esto es suficiente para verificar *la desigualdad de Schwarz* (5.3), en la forma:

$$\bigl|\langle\!\langle \boldsymbol{x}, \boldsymbol{y} \rangle\!\rangle\bigr|^2 \leqslant \langle\!\langle \boldsymbol{x}, \boldsymbol{x} \rangle\!\rangle\,\langle\!\langle \boldsymbol{y}, \boldsymbol{y} \rangle\!\rangle \quad \text{para} \quad \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n,$$

omitiendo el caso de igualdad (por ahora). Basta con repetir la demostración de (5.3), tomando $f(t) := \langle\!\langle \boldsymbol{x} + t\boldsymbol{y}, \boldsymbol{x} + t\boldsymbol{y} \rangle\!\rangle$. Dicho de otra manera:

$$\bigl(\boldsymbol{x}^{\mathsf{t}}A\boldsymbol{y}\bigr)^2 \leqslant (\boldsymbol{x}^{\mathsf{t}}A\boldsymbol{x})(\boldsymbol{y}^{\mathsf{t}}A\boldsymbol{y}) \quad \text{para todo} \quad \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n.$$

Si ahora $\boldsymbol{x}^{\mathsf{t}}A\boldsymbol{x} = 0$ para algún $\boldsymbol{x}$ particular, el lado derecho vale 0, y por lo tanto $\langle \boldsymbol{x}, A\boldsymbol{y} \rangle = \boldsymbol{x}^{\mathsf{t}}A\boldsymbol{y} = 0$ para todo $\boldsymbol{y} \in \mathbb{R}^n$. Como A es invertible, se puede tomar $\boldsymbol{y} := A^{-1}\boldsymbol{x}$, en cuyo caso $\|\boldsymbol{x}\|^2 = \langle \boldsymbol{x}, \boldsymbol{x} \rangle = 0$ y por ende $\boldsymbol{x} = \boldsymbol{0}$.

En resumen, $\boldsymbol{x}^{\mathsf{t}}A\boldsymbol{x} = 0$ implica $\boldsymbol{x} = \boldsymbol{0}$, y se concluye que A es definida positiva. □

Observación. Al final de la demostración anterior, se comprobó que $\langle\!\langle \boldsymbol{x}, \boldsymbol{x} \rangle\!\rangle \geqslant 0$ *con igualdad si y solo si* $\boldsymbol{x} = \boldsymbol{0}$. Entonces, si A es una matriz definida positiva, la fórmula (5.16) define *un nuevo producto escalar real* sobre $\mathbb{R}^n$.

Proposición 5.37. *Una matriz cuadrada $A \in M_n(\mathbb{R})$ es definida positiva si y solo si es simétrica y los pivotes en la eliminación gaussiana simple son todos positivos:*

$$a_{kk}^{(k)} > 0 \quad \textit{para} \quad k = 1, 2, \dots, n.$$

Demostración. Sea $A = A^{\mathsf{t}} \in M_n(\mathbb{R})$ una matriz simétrica e invertible. Si la eliminación gaussiana *simple* es aplicable a esta matriz, producirá pivotes no ceros $a_{kk}^{(k)} \neq 0$, y dará lugar a una factorización $A = LDU$, en donde D es una matriz diagonal y L y U son triangulares unipotentes; ésta descomposición es única.[8]

Además, como A es simétrica, se ve que $A = A^{\mathsf{t}} = U^{\mathsf{t}} D L^{\mathsf{t}}$. Esta es una segunda factorización de tipo LDU de A, con el mismo factor diagonal D; como U^{t} es triangular inferior unipotente y L^{t} es triangular superior unipotente, la unicidad de la factorización muestra que $U = L^{\mathsf{t}}$ (y $L = U^{\mathsf{t}}$ también, desde luego). Esto dice que $\underline{A = LDL^{\mathsf{t}}}$.

Caso 1 Supóngase primero que A es definida positiva.

La eliminación gaussiana es simple cuando ningún pivote $a_{kk}^{(k)}$ se anula durante el proceso. Para $k = 1$, se sabe que

$$a_{11}^{(1)} = a_{11} = \langle \boldsymbol{e}_1, A\,\boldsymbol{e}_1 \rangle > 0.$$

(Aquí $\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}$ es la base estándar de $\mathbb{R}^n$.)

Para $k > 1$, el pivote $a_{kk}^{(k)}$ es el elemento (k, k) de una matriz $\underline{L_k A}$, donde L_k es el producto de varias matrices de tipo $R_{ij}(c)$, véase la fórmula (3.12c). La premultiplicación $A \mapsto L_k A$ coloca ceros debajo de la diagonal en las primeras $(k-1)$ columnas de A.

Como A es simétrica, la *posmultiplicación* $L_k A \mapsto L_k A L_k^{\mathsf{t}}$ ejecuta las *operaciones de columna* análogas a esas operaciones de fila. El resultado es la matriz $\underline{A^{(k)} := L_k A L_k^{\mathsf{t}}}$, que también es simétrica, y tiene ceros *a la derecha de la diagonal* en sus primeras $(k-1)$ *filas*. (Estas operaciones de fila no cambian los elementos diagonales $a_{jj}^{(j)}$ ya calculados, para $j = 1, \dots, k$, porque en cada operación individual su efecto es $a_{jj}^{(j)} \mapsto a_{jj}^{(j)} + 0$.

Siendo así la situación, el siguiente pivote $a_{kk}^{(k)}$ es el elemento (k, k) de la matriz simétrica $A^{(k)}$, dada por:

$$a_{kk}^{(k)} = \langle \boldsymbol{e}_k, A^{(k)} \boldsymbol{e}_k \rangle = \langle \boldsymbol{e}_k, L_k A L_k^{\mathsf{t}}\, \boldsymbol{e}_k \rangle = \langle L_k^{\mathsf{t}}\, \boldsymbol{e}_k, A L_k^{\mathsf{t}}\, \boldsymbol{e}_k \rangle > 0, \tag{5.17}$$

puesto que A es definida positiva. En otras palabras, $a_{kk}^{(k)} = \langle \boldsymbol{x}_k, A\boldsymbol{x}_k \rangle$, donde el vector $\boldsymbol{x}_k := L_k^{\mathsf{t}}\, \boldsymbol{e}_k$ no es $\mathbf{0}$ porque $\langle L_k^{\mathsf{t}}\, \boldsymbol{e}_k, \boldsymbol{e}_k \rangle = \langle \boldsymbol{e}_k, L_k\, \boldsymbol{e}_k \rangle = 1$. (La matriz triangular L_k tiene entradas 1 en su diagonal.)

[8] Se puede suponer que A es *invertible*, porque: (a) si A es definida positiva, entonces A^{-1} existe, por la Proposición 5.36; y (b) si se obtiene $A = LDU$ por eliminación gaussiana simple, la matriz diagonal D es invertible, al igual que L y U; y su producto $A = LDU$ es también invertible.

Antes de pivotear en $a_{kk}^{(k)}$, la matriz $A^{(k)}$ tiene el siguiente aspecto:

$$A^{(k)} = L_k A L_k^{\mathsf{t}} = \begin{bmatrix} a_{11} & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & a_{22}^{(2)} & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_{kk}^{(k)} & * & \dots & * \\ 0 & 0 & \dots & * & * & \dots & * \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & * & * & \dots & * \end{bmatrix}$$

donde cada $*$ representa una entrada desconocida.

Después del último pasos, se llega a una matriz *diagonal*:

$$D := L_n A L_n^{\mathsf{t}} = \operatorname{diag}\bigl[a_{11}, a_{22}^{(2)}, \dots, a_{nn}^{(n)}\bigr].$$

Al poner $\underline{L := L_n^{-1}}$, también triangular inferior, se ha obtenido la factorización $\underline{A = LDL^{\mathsf{t}}}$ por eliminación simple, sin necesidad de hacer intercambios de filas. Entonces la fórmula (5.17) es válida para cada $k = 1, \dots, n$; es decir, todos los pivotes son positivos.

Caso 2 Ahora supóngase, por el contrario, que se obtiene la factorización $A = LDL^{\mathsf{t}}$ por eliminación gaussiana simple, donde la matriz diagonal $D = \operatorname{diag}[d_{11}, d_{22}, \dots, d_{nn}]$ cumple $\underline{d_{kk} > 0}$ para cada k; se debe comprobar que A es definida positiva.

Ahora se ve que $D = D^{1/2} D^{1/2}$, donde se define la “raíz cuadrada positiva” de D como:

$$\underline{D^{1/2}} := \operatorname{diag}[\sqrt{d_{11}}, \dots, \sqrt{d_{nn}}].$$

En tal caso se puede expresar la matriz A como

$$A = LDL^{\mathsf{t}} = LD^{1/2} D^{1/2} L^{\mathsf{t}} = C\,C^{\mathsf{t}}, \quad \text{al poner} \quad \underline{C := LD^{1/2}}.$$

Por el Ejemplo 5.34, se sabe que $C\,C^{\mathsf{t}}$ es una matriz positiva. Como A es también invertible, la Proposición 5.36 muestra que A es definida positiva. □

Definición 5.38. Sea $A \in M_n(\mathbb{R})$ una matriz definida positiva y sea $A = LDL^{\mathsf{t}}$ su factorización por eliminación gaussiana; sea $C := LD^{1/2}$, la cual es una matriz triangular inferior; entonces $\underline{A = C\,C^{\mathsf{t}}}$ se llama la **factorización de Cholesky** de A. ◊

Observación. En el Ejemplo 5.34, se observó que cualquier matriz cuadrada real de la forma $A = B^{\mathsf{t}} B$, para alguna matriz B, es positiva. La Proposición 5.37 dice que si A es *definida* positiva, el resultado inverso es válido: se obtiene $A = B^{\mathsf{t}} B$ al tomar $B := C^{\mathsf{t}}$, una matriz triangular superior. Esto resulta ser cierto también para matrices semidefinidas positivas, aunque se requiere otro tipo de demostración.

Algoritmo 5.2 (Factorización de Cholesky)**.** Sea $A = A^{\mathsf{t}} \in M_n(\mathbb{R})$ una matriz simétrica. Si A es definida positiva, se reemplaza A por la matriz triangular inferior C tal que $A = CC^{\mathsf{t}}$; si A no es positiva, aparece un número negativo en la diagonal de D y el algoritmo termina prematuramente. Se usa la simetría de A para trabajar con el triángulo inferior de A solamente.

⟨ Calcular el factor izquierdo de Cholesky ⟩ ≡
⟨ Declarar el tipo: matriznporn ⟩
procedure Cholesky(A: matriznporn)
⟨ Declarar las variables locales: $i, j, k,$: $1 \,..\, n$; t: real; C: matriznporn ⟩
begin { *Cholesky* }
    for $k \leftarrow 1$ **to** n **do** { *calcular la fila k de C* }
        for $j \leftarrow 1$ **to** $k-1$ **do** { *calcular los elementos no diagonales de C* }
            $C[k,j] \leftarrow A[k,j]/C[j,j]$;
            for $i \leftarrow 1$ **to** $j-1$ **do**
                $C[k,j] \leftarrow C[k,j] - C[j,i] * C[k,i]/C[j,j]$;
            endfor; { i }
            $A[k,j] \leftarrow C[k,j]$;
        endfor; { j }
        $t \leftarrow A[k,k]$; { *inicialización* }
        for $j \leftarrow 1$ **to** $k-1$ **do** { *calcular* d_{kk} }
            $t \leftarrow t - C[k,j]\,\widehat{}\,2$;
        endfor; { j }
        if $t \leqslant 0$ **then** ⟨ Interrumpir proceso ⟩
        else $A[k,k] \leftarrow C[k,k] \leftarrow \mathrm{sqrt}(t)$; { $c_{kk} := \sqrt{a_{kk} - \sum_{j=1}^{k-1} c_{kj}^2}$ }
        endif
        for $j \leftarrow k+1$ **to** n **do**
            $A[k,j] \leftarrow 0$; { *borrar el triángulo superior de A* }
        endfor; { j }
    endfor; { *k: A ha sido reemplazado por C* }
end; { *Cholesky* }

Observación. En el caso complejo, se define matrices positivas (semidefinidas o definidas) de manera análoga. Una matriz cuadrada $A \in M_n(\mathbb{C})$ es una *matriz positiva (semidefinida)* si se cumple esta variante de (5.14): $\langle \boldsymbol{x}, A\boldsymbol{x} \rangle \geqslant 0$ para todo $\underline{\boldsymbol{x} \in \mathbb{C}^n}$. Si $\langle \boldsymbol{x}, A\boldsymbol{x} \rangle > 0$ para $\boldsymbol{x} \neq \mathbf{0}$, se dice que A es *definida positiva*.

Resulta que cada matriz positiva es *hermítica*: $A^* = A$. Cada matriz de la forma B^*B, con $B \in M_n(\mathbb{C})$, es positiva. Y sucede que cualquier matriz positiva es de esa forma.

5.5. Ejercicios sobre espacios vectoriales euclidianos

Ejercicio 5.1. Usar la desigualdad de Schwarz en $\mathbb{R}^n$ para comprobar que la *media aritmética* de n números positivos es menor o igual que su *media cuadrática*; esto es, que

$$\frac{a_1 + a_2 + \cdots + a_n}{n} \leqslant \sqrt{\frac{a_1^2 + a_2^2 + \cdots + a_n^2}{n}}$$

para todo $\boldsymbol{a} \in \mathbb{R}^n$ tal que $a_1 > 0$, $a_2 > 0, \dots, a_n > 0$. Verificar que hay igualdad si y solo si los a_j son todos iguales.

Ejercicio 5.2. Encontrar el complemento ortogonal $M^\perp$ del subespacio $M := \operatorname{lin}\langle \boldsymbol{x}, \boldsymbol{y} \rangle$ de $\mathbb{R}^3$, generado por

$$\boldsymbol{x} := \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}, \qquad \boldsymbol{y} := \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}.$$

⟦ Indicación: sea $A \in M_{2\times3}(\mathbb{R})$ la matriz con filas $\boldsymbol{x}^{\mathsf{t}}$, $\boldsymbol{y}^{\mathsf{t}}$; resolver la ecuación $A\boldsymbol{z} = \boldsymbol{0}$. ⟧

Ejercicio 5.3. Sea $W := \operatorname{lin}\langle \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \rangle$ el subespacio de $\mathbb{R}^4$ generado por los vectores

$$\boldsymbol{x} := \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}, \qquad \boldsymbol{y} := \begin{bmatrix} 2 \\ 1 \\ -1 \\ 1 \end{bmatrix}, \qquad \boldsymbol{z} := \begin{bmatrix} 4 \\ 2 \\ 3 \\ 1 \end{bmatrix}.$$

Encontrar un vector $\underline{\boldsymbol{w} \in W}$ tal que $\langle \boldsymbol{w}, \boldsymbol{x} \rangle = \langle \boldsymbol{w}, \boldsymbol{y} \rangle = 0$.

Ejercicio 5.4. En el espacio vectorial $\mathbb{R}[t]$ de polinomios reales, considérese la fórmula:

$$\langle p(t) \mid q(t) \rangle := \int_{-1}^{1} \frac{p(x)\, q(x)}{\sqrt{1 - x^2}}\, dx.$$

(a) Transformar el lado derecho a una integral $\int_0^\pi \operatorname{algo}(\theta)\, d\theta$ mediante la sustitución $x = \cos\theta$.

(b) Comprobar que este $\langle - \mid - \rangle$ es un producto escalar sobre $\mathbb{R}[t]$ (que no coincide con la receta del Ejemplo 5.3).

(c) El **polinomio de Chebyshev** de grado $n \in \mathbb{N}$ es aquél polinomio $\underline{T_n(t)}$ que cumple

$$T_n(\cos\theta) \equiv \cos n\theta, \quad \text{para} \quad 0 \leqslant \theta \leqslant \pi.$$

Comprobar que $T_0(t) \equiv 1$; $T_1(t) = t$; $T_2(t) = 2t^2 - 1$; $T_3(t) = 4t^3 - 3t$.

(d) Demostrar que estos polinomios son *ortogonales* respecto de este producto escalar.

Ejercicio 5.5. Hallar una base ortonormal para el subespacio $M = \operatorname{lin}\langle \boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3 \rangle \leqslant \mathbb{R}^4$, si

$$\boldsymbol{x}_1 := \begin{bmatrix} 0 \\ 3 \\ 0 \\ 3 \end{bmatrix}, \qquad \boldsymbol{x}_2 := \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \qquad \boldsymbol{x}_3 := \begin{bmatrix} 3 \\ 1 \\ 2 \\ 2 \end{bmatrix}.$$

Ejercicio 5.6. Encontrar una base ortonormal para $M = \operatorname{lin}\langle \boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \boldsymbol{x}_4 \rangle \leqslant \mathbb{R}^5$, donde

$$\boldsymbol{x}_1 := \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \qquad \boldsymbol{x}_2 := \begin{bmatrix} 2 \\ -1 \\ 0 \\ -1 \\ 1 \end{bmatrix}, \qquad \boldsymbol{x}_3 := \begin{bmatrix} 0 \\ 1 \\ 0 \\ -1 \\ 0 \end{bmatrix}, \qquad \boldsymbol{x}_4 := \begin{bmatrix} 2 \\ 1 \\ 1 \\ -1 \\ -1 \end{bmatrix}.$$

Ejercicio 5.7. Obtener una base ortonormal para el subespacio $M \leqslant \mathbb{R}^4$ formado por los vectores $\boldsymbol{x} = (x_1, x_2, x_3, x_4)$ que satisfacen las ecuaciones:

$$\begin{aligned} 3x_1 - 4x_2 + x_3 - 2x_4 &= 0, \\ -3x_1 + 4x_2 - 3x_3 + x_4 &= 0. \end{aligned}$$

Ejercicio 5.8. Usar el algoritmo de Gram y Schmidt para comprobar que los tres vectores

$$\boldsymbol{x}_1 := \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}, \qquad \boldsymbol{x}_2 := \begin{bmatrix} 1 \\ -1 \\ -2 \\ -3 \end{bmatrix}, \qquad \boldsymbol{x}_3 := \begin{bmatrix} 5 \\ 4 \\ 6 \\ 6 \end{bmatrix}$$

son linealmente independientes en $\mathbb{R}^4$.

Ejercicio 5.9. El hiperplano $H := \{\, (x, y, z, w) \in \mathbb{R}^4 : x+y+z+w = 0 \,\}$ tiene dimensión 3. Encontrar (a) una base vectorial de H; y en seguida: (b) una base ortonormal de H.

Ejercicio 5.10. Sea V un espacio vectorial euclidiano (sobre $\mathbb{R}$) o bien hilbertiano (sobre $\mathbb{C}$), con $\dim V = n$. Sea $\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}$ una base ortonormal de V.

(a) Si $x \in V$, verificar la expansión:

$$\boldsymbol{x} = \sum_{j=1}^{n} \langle \boldsymbol{e}_j, \boldsymbol{x} \rangle \, \boldsymbol{e}_j.$$

(b) Demostrar la **igualdad de Parseval**, válida para todo $\boldsymbol{x} \in V$:

$$\|\boldsymbol{x}\|^2 = \sum_{j=1}^{n} \bigl|\langle \boldsymbol{e}_j, \boldsymbol{x} \rangle\bigr|^2.$$

Ejercicio 5.11. Encontrar una tercera columna de modo que esta matriz A sea ortogonal:

$$A := \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & ? \\ \frac{1}{\sqrt{3}} & 0 & ? \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} & ? \end{bmatrix}.$$

Ejercicio 5.12. Decidir (con un razonamiento) si la siguiente matriz es ortogonal o no:

$$A := \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \end{bmatrix}.$$

Ejercicio 5.13. Si $a = \operatorname{sen}\theta \operatorname{sen}\phi$, $b = \operatorname{sen}\theta \cos\phi$, $c = \cos\theta \operatorname{sen}\phi$, $d = \cos\theta\cos\phi$, mostrar que la siguiente matriz es ortogonal:

$$A := \begin{bmatrix} a & b & d & -c \\ b & -a & c & d \\ c & d & -b & a \\ d & -c & -a & -b \end{bmatrix}.$$

Ejercicio 5.14. (a) Si A y B son matrices ortogonales $n \times n$, demostrar que el producto AB es también una matriz ortogonal.

(b) Demostrar que una matriz A es ortogonal si y solo si $\|A\boldsymbol{x}\| = \|\boldsymbol{x}\|$ para todo $\boldsymbol{x} \in \mathbb{R}^n$.

Ejercicio 5.15. Demostrar que una matriz $A \in M_2(\mathbb{R})$ es ortogonal si y solo si hay un ángulo θ tal que

$$A = \begin{bmatrix} \cos\theta & -\operatorname{sen}\theta \\ \operatorname{sen}\theta & \cos\theta \end{bmatrix} \quad \text{o bien} \quad A = \begin{bmatrix} \cos\theta & \operatorname{sen}\theta \\ \operatorname{sen}\theta & -\cos\theta \end{bmatrix}.$$

⟦ Indicación: $a^2 + b^2 = 1$ si y solo si hay un ángulo θ con $a = \cos\theta$, $b = \operatorname{sen}\theta$. ⟧

Ejercicio 5.16. (a) Si $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ con $\boldsymbol{x} \neq \boldsymbol{0}$, $\boldsymbol{y} \neq \boldsymbol{0}$, demostrar que $\boldsymbol{x}\,\boldsymbol{y}^{\mathsf{t}}$ es una matriz $n \times n$ de rango uno: $r(\boldsymbol{x}\,\boldsymbol{y}^{\mathsf{t}}) = 1$.

(b) Sea $\boldsymbol{x} \in \mathbb{R}^n$ tal que $\|\boldsymbol{x}\| = 1$. La **matriz de Householder** determinado por $\boldsymbol{x}$ es $H := 1_n - 2\boldsymbol{x}\,\boldsymbol{x}^{\mathsf{t}} \in M_n(\mathbb{R})$. Demostrar que H es simétrica y ortogonal; y que $H^2 = 1$.

⟦ Indicación: nótese que $\|\boldsymbol{x}\|^2 = \langle \boldsymbol{x}, \boldsymbol{x} \rangle = \boldsymbol{x}^{\mathsf{t}}\boldsymbol{x}$. ⟧

Ejercicio 5.17. Una matriz $A \in M_n(\mathbb{R})$ se llama **no negativa** si $a_{ij} \geqslant 0$ para todo i, j. Sean

$$A := \begin{bmatrix} 2 & 3 \\ 3 & 2 \end{bmatrix}, \qquad B := \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}.$$

Demostrar que A es no negativa pero que no es positiva, mientras que B es definida positiva sin ser no negativa. ⟦ Indicación: para determinar que B es definida positiva, usar eliminación gaussiana para determinar su factorización LDL^{t}. ⟧

Ejercicio 5.18. (a) Sea $A := \begin{bmatrix} a & b \\ b & c \end{bmatrix}$ una matriz simétrica en $M_2(\mathbb{R})$. Demostrar directamente que es posible factorizar $A = LDL^{\mathsf{t}}$, con L triangular inferior unipotente y D diagonal invertible, si y solo si $a \neq 0$ y $c \neq b^2/a$.

(b) Concluir que A es definida positiva si y solo si $a > 0$, $c > 0$ y $ac - b^2 > 0$.

Ejercicio 5.19. Obtener la factorización de Cholesky $A = CC^{\mathsf{t}}$ de la matriz definida positiva

$$A := \begin{bmatrix} 4 & 12 \\ 12 & 45 \end{bmatrix}.$$

Usarla para mostrar que

$$\boldsymbol{x}^{\mathsf{t}} A \boldsymbol{x} = (2x_1 + 6x_2)^2 + (3x_2)^2 \quad \text{si} \quad \boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in \mathbb{R}^2.$$

Ejercicio 5.20. Determinar si cada una de estas matrices simétricas,

$$A := \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \qquad B := \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{bmatrix}, \qquad C := \begin{bmatrix} 9 & -6 & 2 \\ -6 & 8 & -4 \\ 2 & -4 & 4 \end{bmatrix},$$

es (a) positiva; y (b) definida positiva.

Ejercicio 5.21. Un polinomio cuadrático en n variables $\boldsymbol{x} = (x_1, \dots, x_n)$ tiene la forma

$$\underline{P(\boldsymbol{x}) = \tfrac{1}{2}\boldsymbol{x}^{\mathsf{t}} A \boldsymbol{x} + \boldsymbol{x}^{\mathsf{t}}\boldsymbol{b} + c,} \quad \text{donde} \quad A \in M_n(\mathbb{R}),\ \boldsymbol{b} \in \mathbb{R}^n,\ c \in \mathbb{R}.$$

Supóngase que A es definida positiva y $\boldsymbol{x}$ es la solución (única) de la ecuación $\underline{A\boldsymbol{x} = -\boldsymbol{b}}$.

(a) Demostrar que $P(\boldsymbol{y}) - P(\boldsymbol{x}) = \frac{1}{2}(\boldsymbol{y} - \boldsymbol{x})^{\mathsf{t}} A (\boldsymbol{y} - \boldsymbol{x})$ para todo $\boldsymbol{y} \in \mathbb{R}^n$.

(b) Concluir que el polinomio P alcanza un único mínimo absoluto en $\boldsymbol{x} = -A^{-1}\boldsymbol{b}$, y que su valor mínimo es $c - \frac{1}{2}\boldsymbol{b}^{\mathsf{t}} A^{-1} \boldsymbol{b}$.

(c) Calcular el valor mínimo de la función

$$P(x_1, x_2) := \tfrac{1}{2}(x_1^2 + 2x_1x_2 + 2x_2^2) - x_1 + x_2.$$

6 Trazas y determinantes

Datis aequationibus quotcunque sufficientibus ad tollendas quantitates, quae simplicem gradum non egrediuntur, pro aequatione prodeunte primo sumendae sunt omnes combinationes possibiles, quas ingreditur una tantum coefficiens uniuscunque aequationis; secundo eae combinationes opposita habent signa, si in eodem prodeuntis aequationis latere ponantur, quae habent tot coefficentes communes, quot sunt unitates in numero quantitatum tollendarum unitate minuto; caeterae habent eadem signa.

— Gottfried Wilhelm Leibniz[1]

6.1. Traza y determinante de una matriz cuadrada

Este capítulo versa sobre dos funciones escalares de una matriz cuadrada, que resultan ser invariantes bajo la relación de semejanza y en consecuencia son propiedades de operadores lineales. El primero de ellas, *la traza*, es una función lineal; el otro – *el determinante* – es más bien una función polinomial de las entradas matriciales. Esa segunda función recibe su nombre porque determina si una matriz es invertible o no.

Los escalares se toman de cualquier cuerpo $\mathbb{F}$, no solamente de $\mathbb{R}$ o $\mathbb{C}$.

► En el Ejemplo 5.4 se introdujo la traza de una matriz real. Conviene repetir ese concepto ahora en un definición formal, aplicable a cuerpos de escalares cualesquiera.

Definición 6.1. La **traza** de una matriz cuadrada $C \in M_n(\mathbb{F})$ es la suma de sus elementos diagonales:

$$\boxed{\underline{\operatorname{tr} C} := c_{11} + c_{22} + \cdots + c_{nn} = \sum_{k=1}^{n} c_{kk} \in \mathbb{F}.} \tag{6.1}$$

Esta fórmula es idéntica a (5.2), ya visto. ◊

La *propiedad cíclica* de la traza se expresa en el siguiente lema y su corolario.

Lema 6.2. *Dadas dos matrices rectangulares $A \in M_{m\times n}(\mathbb{F})$ y $B \in M_{n\times m}(\mathbb{F})$, resulta que*

$$\boxed{\operatorname{tr}(AB) = \operatorname{tr}(BA).} \tag{6.2a}$$

[1] En una carta del 28 de abril de 1693, dirigida a Guillaume, Marquis de l'Hôpital.

Demostración. Las dos matrices AB y BA son cuadradas, aunque no necesariamente de la misma especie: $AB \in M_m(\mathbb{F})$ mientras $BA \in M_n(\mathbb{F})$. Sin embargo, sus trazas coinciden:

$$\operatorname{tr}(AB) = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}b_{ji} = \sum_{j=1}^{n}\sum_{i=1}^{m} b_{ji}a_{ij} = \operatorname{tr}(BA). \qquad \square$$

Corolario 6.3. *Para tres matrices $A \in M_{m\times n}(\mathbb{F})$, $B \in M_{n\times r}(\mathbb{F})$, $C \in M_{r\times m}(\mathbb{F})$, vale*

$$\operatorname{tr}(ABC) = \operatorname{tr}(BCA) = \operatorname{tr}(CAB). \tag{6.2b}$$

Demostración. Obsérvese que las dimensiones de los espacios de matrices del enunciado son tales que los tres productos de matrices ABC, BCA y CAB existen y dan lugar a matrices cuadradas de lados m, n, r, respectivamente. Si se aplica el Lema 6.2 a las parejas: A y BC; B y CA (o también C y AB), entonces (6.2b) sigue de (6.2a). De modo alternativo, nótese que las tres trazas en cuestión son iguales a la suma triple siguiente:

$$\operatorname{tr}(ABC) = \sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1}^{r} a_{ij}b_{jk}c_{ki}$$

de donde (6.2b) también es evidente. $\square$

Proposición 6.4. *Si $A, B \in M_n(\mathbb{F})$ son matrices semejantes, entonces $\underline{\operatorname{tr} A = \operatorname{tr} B}$.*

Demostración. Se debe recordar (Definición 4.37) que A y B son *semejantes* si y solo si hay una matriz invertible $P \in M_n(\mathbb{F})$ tal que $B = P^{-1}AP$. Entonces, invocando (6.2b),

$$\operatorname{tr} B = \operatorname{tr}(P^{-1}AP) = \operatorname{tr}(APP^{-1}) = \operatorname{tr} A. \qquad \square$$

► Sea $T\colon V \to V$ un operador lineal sobre un espacio vectorial V, de dimensión n sobre $\mathbb{F}$. Si $\mathcal{B} = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ es una base de V, sea $A \in M_n(\mathbb{F})$ la matriz de T con respecto a la base $\mathcal{B}$ dada por la fórmula (4.2). (Véase la discusión al inicio de la subsección 4.5.) Se puede definir la **traza del operador lineal** T por: $\underline{\operatorname{tr} T := \operatorname{tr} A}$.

Esta traza $\operatorname{tr} T$ está bien definida, por lo siguiente: si $\mathcal{B}' = \{\boldsymbol{x}'_1, \dots, \boldsymbol{x}'_n\}$ es otra base de V, con matriz de cambio de base P dada por $\boldsymbol{x}'_s = \sum_{j=1}^{n} p_{js}\,\boldsymbol{x}_j$ según (4.10), ya se sabe que la matriz de T con respecto a la nueva base $\mathcal{B}'$ es $B := P^{-1}AP$. Ahora la Proposición 6.4 dice que $\underline{\operatorname{tr} T = \operatorname{tr} B}$ también.

En resumen: la traza de un operador lineal es la traza de su matriz con respecto a cualquier base de V, puesto que esas trazas son todas iguales.

Observación. Es fácil ver que la traza $\underline{A \mapsto \operatorname{tr} A} \in \mathbb{F}$ es una forma lineal sobre $M_n(\mathbb{F})$. De igual manera, $T \mapsto \operatorname{tr} T \in \mathbb{F}$ es una forma lineal sobre $\mathcal{L}(V)$.

► La otra función de importancia sobre $M_n(A)$ – el determinante – no es lineal. Para definirla, conviene comenzar con matrices 2×2 o 3×3.

Definición 6.5. Sea $A := \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ una matriz en $M_2(\mathbb{F})$; su **determinante** es el escalar $\underline{\det A} \in \mathbb{F}$ definido por

$$\det A \equiv \begin{vmatrix} a & b \\ c & d \end{vmatrix} := \underline{ad - bc}. \tag{6.3}$$

(Se escribe el determinante con barras verticales en vez de corchetes). ◊

Por el Ejemplo 3.7, se ve que $A \in M_2(\mathbb{F})$ es invertible si y solo si $\det A \neq 0$, en cuyo caso el inverso está dado por (3.5):

$$A^{-1} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = \frac{1}{\det A} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

Definición 6.6. Si $A \in M_3(\mathbb{F})$, se define $\underline{\det A}$ por una *expansión en la primera fila*:

$$\begin{aligned} \det A := \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} &:= a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \\ &= a_{11}a_{22}a_{23} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31}. \end{aligned} \tag{6.4}$$

En este caso, $\det A$ es un polinomio homogéneo de grado 3 en las entradas de A. ◊

La fórmula (6.4) para un determinante 3×3 sugiere una receta eficiente de cálculo manual. Se escribe una copia de las primeras dos columnas de A a la derecha de la matriz; luego se calculan seis productos de elementos en las "subdiagonales" de este arreglo 3×5, sumando los tres productos en subdiagonales que bajan hacia la derecha, y restando los tres productos en subdiagonales que suben hacia la derecha. En símbolos:

$$\det A = \begin{array}{c} \begin{array}{ccccc} \searrow^{+} & \searrow^{+} & \searrow^{+} & & \end{array} \\ \left| \begin{array}{ccc} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{array} \right| \begin{array}{cc} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{array} \\ \begin{array}{ccccc} \nearrow_{-} & \nearrow_{-} & \nearrow_{-} & & \end{array} \end{array} \tag{6.5}$$

Ejemplo 6.7. Hallar el determinante 3×3 de la matriz $\begin{bmatrix} -3 & 1 & 2 \\ -1 & 3 & -4 \\ 1 & -2 & 5 \end{bmatrix}$.

Hay tres maneras de efectuar el cálculo de este determinante.

(a) Por expansión en la primera fila (6.4):

$$
\begin{aligned}
\begin{vmatrix} -3 & 1 & 2 \\ -1 & 3 & -4 \\ 1 & -2 & 5 \end{vmatrix} &= -3\begin{vmatrix} 3 & -4 \\ -2 & 5 \end{vmatrix} - 1\begin{vmatrix} -1 & -4 \\ 1 & 5 \end{vmatrix} + 2\begin{vmatrix} -1 & 3 \\ 1 & -2 \end{vmatrix} \\
&= -3(15-8) - 1(-5+4) + 2(2-3) \\
&= -21 + 1 - 2 = -22.
\end{aligned}
$$

(b) También se puede hacer una expansión en la primera *columna*:

$$
\begin{aligned}
\begin{vmatrix} -3 & 1 & 2 \\ -1 & 3 & -4 \\ 1 & -2 & 5 \end{vmatrix} &= -3\begin{vmatrix} 3 & -4 \\ -2 & 5 \end{vmatrix} - (-1)\begin{vmatrix} 1 & 2 \\ -2 & 5 \end{vmatrix} + 1\begin{vmatrix} 1 & 2 \\ 3 & -4 \end{vmatrix} \\
&= -3(15-8) + 1(5+4) + 1(-4-6) \\
&= -21 + 9 - 10 = -22.
\end{aligned}
$$

(c) El método de los productos en subdiagonales produce lo siguiente:

$$
\begin{vmatrix} -3 & 1 & 2 \\ -1 & 3 & -4 \\ 1 & -2 & 5 \end{vmatrix} = \begin{array}{|ccc|cc} \searrow^{+} & \searrow^{+} & \searrow^{+} & & \\ -3 & 1 & 2 & -3 & 1 \\ -1 & 3 & -4 & -1 & 1 \\ 1 & -2 & 5 & 1 & -1 \\ \nearrow_{-} & \nearrow_{-} & \nearrow_{-} & & \end{array}
$$

$$
\begin{aligned}
&= (-45 - 4 + 4) - (6 - 24 - 5) \\
&= -45 - (-23) = -22.
\end{aligned}
$$

Los resultados de los tres cómputos coinciden. ◊

Desafortunadamente, la receta (6.5) no se generaliza a matrices $n \times n$ con $n > 3$.

La definición general del determinante es inductiva. Para enunciarlo, se requiere la terminología siguiente.

Definición 6.8. Sea $A \in M_n(\mathbb{F})$ una matriz $n \times n$. Sea A_{ij} (con A mayúscula) la **submatriz** de A que se obtiene *al borrar la fila* $\boldsymbol{a}'_i$ *y la columna* $\boldsymbol{a}_j$ de A.

Cada A_{ij} es una matriz $(n-1) \times (n-1)$. Sea $M_{ij} := \det A_{ij}$; este escalar $M_{ij} \in \mathbb{F}$ se llama el **menor** de la matriz A correspondiente a la entrada a_{ij}. ◊

Obsérvese que la ecuación (6.4) se escribe en esta notación así:

$$
\det A = a_{11}M_{11} - a_{12}M_{12} + a_{13}M_{13}. \tag{6.4'}
$$

El determinante de una matriz 1×1 es simplemente $\underline{\det [a_{11}] := a_{11}}$. Supóngase entonces que se dispone de un procedimiento para calcular determinantes de matrices $(n-1) \times (n-1)$. La fórmula anterior puede generalizarse como sigue.

Definición 6.9. Sea $A \in M_n(\mathbb{F})$; se define $\underline{\det A}$ por una *expansión en la primera fila* $\boldsymbol{a}'_1$,

$$\begin{aligned} \det A &:= \underline{a_{11}M_{11} - a_{12}M_{12} + \cdots + (-1)^{1+n}a_{1n}M_{1n}} \\ &= \sum_{j=1}^{n} (-1)^{1+j} a_{1j} M_{1j}, \end{aligned} \tag{6.6a}$$

o bien por expansión en cualquier otra fila $\boldsymbol{a}'_i$:

$$\begin{aligned} \det A &= (-1)^{i+1}a_{i1}M_{i1} + (-1)^{i+2}a_{i2}M_{i2} + \cdots + (-1)^{i+n}a_{in}M_{in} \\ &= \sum_{j=1}^{n} (-1)^{i+j} a_{ij} M_{ij}, \end{aligned} \tag{6.6b}$$

o bien por expansión en cualquier columna $\boldsymbol{a}_j$:

$$\begin{aligned} \det A &= (-1)^{1+j}a_{1j}M_{1j} + (-1)^{2+j}a_{2j}M_{2j} + \cdots + (-1)^{n+j}a_{nj}M_{nj} \\ &= \sum_{i=1}^{n} (-1)^{i+j} a_{ij} M_{ij}. \end{aligned} \tag{6.6c}$$

En seguida, se comprobará que todas estas expresiones coinciden. ◊

Para verificar la igualdad de las tres expresiones (6.6), se abordará en seguida otra expresión para el determinante que resulta ser igual a cada una de ellas.

Proposición 6.10. *Sea $A \in M_n(\mathbb{F})$; entonces* $\det A$*, definido por ejemplo por* (6.6a)*, es igual a*

$$\boxed{\det A = \sum_{\sigma} (-1)^{\sigma} a_{1j_1} a_{2j_2} \cdots a_{nj_n},} \tag{6.7}$$

donde la sumatoria recorre todas las $n!$ permutaciones $\sigma = (j_1, \dots, j_n)$ de $(1, 2, \dots, n)$ y el signo $(-1)^{\sigma} = \pm 1$ es $+1$ ó -1 según la permutación σ sea par o impar.

Además, las tres expresiones (6.6) *son iguales a* (6.7) *y por ende son iguales entre sí.*

Demostración. Una permutación σ de $(1, 2, \dots, n)$ se llama **par** si es el producto de un número par de transposiciones $i \leftrightarrow j$; y σ se llama **impar** en el caso contrario. Si σ es el producto de k transposiciones, entonces $\underline{(-1)^{\sigma}} := (-1)^k$ por definición.

Según la fórmula (6.6a), vale $\det A := \sum_{j_1=1}^{n}(-1)^{1+j_1}a_{1j_1}M_{1j_1}$. A su vez, cada menor $M_{1j_1} = \det A_{1j_1}$ es una suma análoga de términos con $\pm a_{2j_2}$ multiplicado por unos menores correspondientes de la submatriz A_{1j_1}.

Al repetir este argumento $(n-1)$ veces, $\det A$ queda expresado como la suma de los productos que aparecen al lado derecho de (6.7), en donde cada producto contiene *un factor tomado de cada fila* y *de columnas distintas*. Hay n términos en (6.6a), $(n-1)$ términos en la expansión correspondiente a cada menor M_{1j_1}, etc., para un total de $n(n-1)(n-2)\cdots 3\cdot 2 = n!$ términos en la expansión final. La suma de estos productos recorre todas las permutaciones posibles de $(1, 2, \dots, n)$.

Quedan por determinarse los *signos* en (6.7). En primer lugar, $(-1)^{1+j_1}$ es el signo de la transposición $1 \leftrightarrow j_1$. Por inducción sobre n, se puede verificar que el producto de los diversos ± 1 que aparecen en la expansión iterativa de (6.6a) es efectivamente $+1$ si y solo si $(j_1, \dots, j_n)$ es una permutación par de $(1, 2, \dots, n)$.

El mismo argumento es aplicable al usar cualquiera de las recetas (6.6b) ó (6.6c). En esos casos, la fila o columna de expansión se puede elegir arbitrariamente en cada iteración de la expansión. A lo sumo, podría ocurrir que el resultado final difiere del lado derecho de (6.7) por un múltiplo global por (± 1), que sería independiente de la matriz A. Un cálculo explícito muestra que todas las recetas en (6.6) dan $\underline{\det 1_n = +1}$, así que el desarrollo (6.7) es correcto en todos los casos. $\square$

Para calcular un determinante $n\times n$ con $n > 3$, es impensable usar (6.7) directamente, ya que habría que calcular $n!$ productos de n términos cada uno y luego sumarlos. Una mejor alternativa es usar uno de los desarrollos (6.6); como todas dan el mismo resultado, *es posible, en cada paso de la expansión, escoger la fila o columna de preferencia para la expansión siguiente*. Obsérvese que el signo $\underline{(-1)^{i+j}}$ se calcula por inspección de un "tablero de damas":

$$\begin{vmatrix} + & - & + & - & + \\ - & + & - & + & - \\ + & - & + & - & + \\ - & + & - & + & - \\ + & - & + & - & + \end{vmatrix} \tag{6.8}$$

y que el signo de cada posición diagonal (j, j) es positivo: $(-1)^{j+j} = (-1)^{2j} = +1$.

Ejemplo 6.11. Calcular el determinante:

$$\begin{vmatrix} 2 & 1 & 1 & 0 \\ 0 & -3 & 1 & 2 \\ 0 & -1 & 3 & -4 \\ 3 & 1 & -2 & 5 \end{vmatrix}$$

por expansiones en filas o columnas.

En cada paso que sigue, se indica la fila o columna de expansión por un juego de signos (apropiados!) en la parte superior izquierda de cada una de sus entradas.

$$
\begin{aligned}
\begin{vmatrix} {}^{(+)}2 & 1 & 1 & 0 \\ {}^{(-)}0 & -3 & 1 & 2 \\ {}^{(+)}0 & -1 & 3 & -4 \\ {}^{(-)}3 & 1 & -2 & 5 \end{vmatrix}
&= 2 \begin{vmatrix} {}^{(+)}-3 & 1 & 2 \\ {}^{(-)}-1 & 3 & -4 \\ {}^{(+)}\ 1 & -2 & 5 \end{vmatrix} - 3 \begin{vmatrix} {}^{(+)}1 & {}^{(-)}1 & {}^{(+)}0 \\ -3 & 1 & 2 \\ -1 & 3 & -4 \end{vmatrix} \\
&= 2(-3) \begin{vmatrix} 3 & -4 \\ -2 & 5 \end{vmatrix} + 2(+1) \begin{vmatrix} 1 & 2 \\ -2 & 5 \end{vmatrix} + 2(+1) \begin{vmatrix} 1 & 2 \\ 3 & -4 \end{vmatrix} \\
&\qquad - 3(+1) \begin{vmatrix} 1 & 2 \\ 3 & -4 \end{vmatrix} - 3(-1) \begin{vmatrix} -3 & 2 \\ -1 & -4 \end{vmatrix} \\
&= -6(15-8) + 2(5+4) + 2(-4-6) - 3(-4-6) + 3(12+2) \\
&= -42 + 18 - 20 + 30 + 42 = 28.
\end{aligned}
$$

En detalle: *la primera columna* de la matriz original *tiene dos ceros*; al expandir allí, hay dos términos menos en el paso siguiente. La submatriz A_{11} no contiene ceros, y se puede expandir en cualquier fila o columna; se usa la primera columna. La submatriz A_{41} contiene un cero en la posición $(1, 3)$, lo cual sugiere una expansión en la primera fila o la tercera columna; se usa la primera fila. Al llegar a menores 2×2, las determinantes se calculan por inspección. $\lozenge$

6.2. Propiedades de determinantes

Proposición 6.12. *Sean $A, B \in M_n(\mathbb{F})$ dos matrices cuadradas. Entonces*

$$\boxed{\det(AB) = (\det A)(\det B).} \tag{6.9}$$

Demostración. Sea $C := AB$. Por la fórmula (6.7), $\underline{\det C}$ es igual a una suma de productos $\pm c_{1j_1} c_{2j_2} \cdots c_{nj_n}$. Cada c_{kj_k} es, a su vez, una suma de términos $\sum_{i_k=1}^{n} a_{ki_k} b_{i_k j_k}$, de tal manera que

$$\det(AB) = \sum (-1)^{\sigma} a_{1i_1} a_{2i_2} \cdots a_{ni_n} b_{i_1 j_1} b_{i_2 j_2} \cdots b_{i_n j_n}. \tag{6.10}$$

Esta suma extiende sobre todas las permutaciones $\sigma = (j_1, \dots, j_n)$ de $(1, 2, \dots, n)$ y sobre *todas las posibilidades* para $i_1, \dots, i_n$. Ahora, si dos de los i_k son iguales, la suma $\sum_{\sigma} (-1)^{\sigma} b_{i_1 j_1} b_{i_2 j_2} \cdots b_{i_n j_n}$ se anula por cancelación de términos, así que aparecen en (6.10) solamente los términos en donde $\tau = (i_1, \dots, i_n)$ es una *permutación* de $(1, 2, \dots, n)$.

Sea $\rho = (r_1, \dots, r_n)$ la permutación de $(1, 2, \dots, n)$ que transforma cada índice i_k en el j_k correspondiente. Entonces la permutación σ es la composición de las dos permutaciones τ y ρ, en cuyo caso vale $(-1)^{\sigma} = (-1)^{\tau}(-1)^{\rho}$.

Esto implica que la expresión (6.10) se simplifica en:

$$\det(AB) = \Bigl(\sum_{\tau}(-1)^{\tau} a_{1i_1}a_{2i_2}\cdots a_{ni_n}\Bigr)\Bigl(\sum_{\rho}(-1)^{\rho} b_{1r_1}b_{2r_2}\cdots b_{nr_n}\Bigr) = (\det A)(\det B). \qquad \square$$

Proposición 6.13. *Sea $A \in M_n(\mathbb{F})$. El determinante de A se transforma bajo operaciones elementales de fila como sigue:*

(a) *al intercambiar dos filas de A,* $\det A$ cambia de signo*;*

(b) *al multiplicar una fila de A por $c \neq 0$,* se multiplica $\det A$ también *por c;*

(c) *al sustraer de una fila de A un múltiplo de otra fila,* $\det A$ permanece igual.

Demostración. En vista de la Proposición 6.12, y habida cuenta de que las operaciones elementales de fila son ejecutadas por premultiplicación por matrices de uno de los tres tipos (3.12), es suficiente mostrar que:

$$\det P_{ik} = -1, \qquad \det M_i(c) = c, \qquad \det R_{ik}(c) = 1,$$

donde P_{ik}, $M_i(c)$ y $R_{ik}(c)$ son las matrices de las ecuaciones (3.12).

Al expandir en una fila i de A *que tenga* 1 *en la diagonal y* 0 *en las demás entradas,* la fórmula (6.6b) muestra que $\det A = (-1)^{i+i} 1\, M_{ii} = M_{ii}$, es decir, pueden eliminarse la fila i y la columna i de A sin cambiar el determinante.

En P_{ik}, $M_i(c)$ y $R_{ik}(c)$, *se eliminan así todas las filas salvo las filas F_i y F_k*, y las columnas correspondientes. (Para $M_i(c)$, se puede tomar $k \neq i$ cualquiera). Luego la expansión (6.6b) *reduce el cálculo* de estos determinantes *al caso* 2×2 y se obtiene:

$$\det P_{ik} = \begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix} = -1, \qquad \det M_i(c) = \begin{vmatrix} 1 & 0 \\ 0 & c \end{vmatrix} = c, \qquad \det R_{ik}(c) = \begin{vmatrix} 1 & 0 \\ -c & 1 \end{vmatrix} = 1. \qquad \square$$

Proposición 6.14. *Sean $A \in M_n(\mathbb{F})$ una matriz cuadrada. Entonces*

$$\boxed{\det A^{\mathsf{t}} = \det A.}$$

Demostración. La fórmula (6.7), aplicada a la matriz A^{t}, da

$$\det A^{\mathsf{t}} = \sum_{\sigma\in S_n} (-1)^{\sigma} a_{j_1 1} a_{j_2 2} \dots a_{j_n n};$$

Sea $\pi \equiv (p_1, \dots, p_n) := \sigma^{-1}$ la permutación recíproca que lleva cada j_k en k. Si σ es el producto de m transposiciones, π es el producto de las mismas m transposiciones en el orden inverso: por lo tanto $(-1)^{\pi} = (-1)^{\sigma}$. Luego,

$$\det A^{\mathsf{t}} = \sum_{\pi\in S_n} (-1)^{\pi} a_{1p_1} a_{2p_2} \dots a_{np_n} = \det A. \qquad \square$$

Proposición 6.15. *Si $A \in M_n(\mathbb{F})$ es triangular, su determinante es el producto de los elementos diagonales:* $\underline{\det A = a_{11}a_{22}\cdots a_{nn}}$. *En particular, esto dice que* $\underline{\det 1_n = 1}$.

Demostración. Por la Proposición 6.14 anterior, se puede suponer que A es una matriz triangular *inferior*. La expansión (6.6a) en la primera fila da $\det A = a_{11}\,M_{11}$. Ahora la submatriz A_{11} (al borrar la primera fila y la primera columna de A) es también triangular inferior, y su determinante es $M_{11} = a_{22}\,M_{12,12}$, así que $\det A = a_{11}a_{22}\,M_{12,12}$, donde $M_{12,12}$ es el menor correspondiente (obtenido al borrar las primeras dos filas y las primeras dos columnas de A). Al repetir este argumento $(n-2)$ veces, se obtiene

$$\det A = a_{11}a_{22}\cdots a_{n-2,n-2}\begin{vmatrix} a_{n-1,n-1} & 0 \\ a_{n,n-1} & a_{nn}\end{vmatrix} = a_{11}a_{22}\cdots a_{nn}. \qquad \square$$

Lema 6.16. *Vale* $\underline{\det A = 0}$ *si una fila o una columna de A es* $\mathbf{0}$*; o si dos filas de A son iguales; o bien si dos columnas de A son iguales.*

Demostración. Si una fila o columna de A tiene todas sus entradas igual a 0, la expansión en esa fila o columna produce 0.

Si dos filas de A son iguales, el intercambio de esas dos filas no modifica A, ni cambia $\det A$. Pero la Proposición 6.13(a) dice que $\det A \mapsto -\det A$; luego, $\det A = 0$.

Si dos columnas de A son iguales, entonces dos filas de A^{t} son iguales y por lo tanto $\det A = \det A^{\mathsf{t}} = 0$. $\square$

Proposición 6.17. $\det A = 0$ *si y solo si A es* singular *(es decir, no invertible).*

Demostración. Si A es invertible, entonces $1 = \det 1_n = \det(AA^{-1}) = (\det A)(\det A^{-1})$, así que $\det A$ no puede ser 0.

Por otro lado, si A es singular, el método de eliminación gaussiana con pivoteo parcial fallará al encontrar un pivote cero que no se puede remediar por trueque de filas.

Sea A' la matriz obtenida por los pasos del algoritmo antes de aparecer el pivote cero, que son operaciones elementales de fila de tipos (a) y (c); vale $\det A' = \pm\det A$ por la Proposición 6.13.

Ahora A' es de la forma

$$A' = \begin{bmatrix} U' & X' \\ 0 & Y' \end{bmatrix},$$

donde U' es *triangular superior* y sus elementos diagonales son los pivotes iniciales $a_{11}, a_{22}^{(2)}, \dots, a_{kk}^{(k)}$; 0 es un bloque rectangular de ceros; y la primera columna de la submatriz Y' es también una columna de ceros ($\implies \det Y' = 0$). Al expandir $\det A'$ en la primera columna k veces, se obtiene

$$\det A' = a_{11}M_{11} = \cdots = a_{11}\,a_{22}^{(2)}\cdots a_{kk}^{(k)}\det Y' = 0,$$

y por lo tanto $\det A = 0$ también. $\square$

Muchas veces, es recomendable *calcular determinantes por eliminación gaussiana*, como sigue. La Proposición 6.17 anterior muestra que $\det A = 0$ si y solo si el algoritmo de eliminación gaussiana con pivoteo parcial (el Algoritmo 3.2) encuentra un pivote 0 que no es posible evitar por intercambio de filas.[2] En el caso de obtener n pivotes no ceros (siendo A una matriz $n \times n$), cada intercambio de filas contribuye un cambio de signo al cómputo del determinante.

Si en el transcurso del algoritmo se debe hacer r intercambios de filas, mediante las matrices $P_{i_1k_1}$, $P_{i_2k_2}$, ..., $P_{i_rk_r}$, habrá un factor $(-1)^r = \pm 1$ en el cálculo final del determinante. Sea $P := P_{i_rk_r} \cdots P_{i_1k_1}$ el producto de esas matrices en el orden inverso. Resulta que *la matriz PA admite eliminación gaussiana simple*. ⟦ Dicho de otro modo: se podría regresar al inicio y ejecutar todos los intercambios de filas necesarios, y luego continuar la eliminación con operaciones de tipo (c) solamente. ⟧ Se obtiene así una factorización matricial $\underline{PA = LV}$ tal que

$$\underline{(-1)^r \det A} = (\det P)(\det A) = \det(PA) = \det(LV) = (\det L)(\det V) = \underline{\det V},$$

pues $\underline{\det L = 1}$ por la Proposición 6.15, siendo L una matriz triangular *unipotente*. Ahora V es una matriz triangular (superior), cuyos elementos diagonales son los pivotes a_{11}, $a_{22}^{(2)}$, ..., $a_{nn}^{(n)}$ obtenidos con el algoritmo; y su producto es $\underline{\det V}$.

Se puede resumir esta discusión en la siguiente Proposición.

Proposición 6.18. *Aplíquese el algoritmo de eliminación gaussiana con pivoteo parcial a la matriz cuadrada $A \in M_n(\mathbb{F})$. Entonces $\det A = 0$ si y solo si es imposible hallar n pivotes no ceros. En cambio, si se logra encontrar n pivotes no ceros a_{11}, $a_{22}^{(2)}$, ..., $a_{nn}^{(n)}$, entonces*

$$\boxed{\det A = (-1)^r\, a_{11}\, a_{22}^{(2)} \cdots a_{nn}^{(n)},} \tag{6.11}$$

donde r es el número de intercambios de filas efectuados. ⊟

► El determinante de una matriz rectangular no se define; sin embargo, los determinantes sí dan información acerca de una matriz rectangular, al considerar la colección de sus *submatrices cuadradas*.

Proposición 6.19. *Sea $A \in M_{m\times n}(\mathbb{F})$ una matriz rectangular. Su **rango** $\overline{r(A)}$ es el mayor entero $\underline{k}$ tal que A posee una* submatriz *B de tamaño $\underline{k \times k}$ tal que* $\det B \neq 0$.

Demostración. Si $r(A) = k$, hay unas k columnas linealmente independientes en A (por ejemplo, aquellas que se convierten en columnas básicas al aplicar el Algoritmo 4.1 de forma escalonada). Sea C la *submatriz* $\underline{m \times k}$ de A obtenida por borrar sus otras columnas.

[2] Si se usa este algoritmo como subrutina para cálculo de determinantes, hay que detenerlo en ese momento y devolver un resultado cero al programa principal.

Ahora $r(C^{\mathsf{t}}) = r(C) = k$; luego C tiene unas k filas linealmente independientes (aplíquese el Algoritmo 4.1 a C^{t}). Sea B la submatriz $k \times k$ de C (y por ende de A) que se obtiene al borrar las otras filas de C. Como $r(B) = r(C) = k$, esta matriz B es invertible y así $\det B \neq 0$.

Si $k < \min\{m, n\}$, cualquier submatriz M de tamaño $(k + 1) \times (k + 1)$ también tiene rango no mayor que k, porque de otro modo las columnas correspondientes de A serían linealmente independientes.[3] En consecuencia, esa submatriz M es no invertible y tiene determinante 0. □

La Proposición 6.19 proporciona un mecanismo alternativo al Algoritmo 4.1 para calcular el rango de una matriz; a veces es útil para averiguar el rango de matrices pequeñas por inspección.

6.3. La regla de Cramer

Los determinantes se pueden emplear también para resolver sistemas de ecuaciones lineales *por fórmulas explícitas* – como una alternativa al procedimiento paso por paso de la eliminación gaussiana. En el caso de sistemas 3×3, esa regla explícita fue propuesta por Gabriel Cramer en 1750. En la práctica, su uso resulta ineficiente para sistemas con más de tres variables, pero tiene cierta importancia teórica. Muestra, por ejemplo, que la solución de un sistema de ecuaciones con coeficientes *enteros* es un vector de números racionales, cuyo denominador común es un divisor del determinante de la matriz de coeficientes del sistema.

Definición 6.20. Sea A una matriz cuadrada en $M_n(\mathbb{F})$. El **cofactor** de la entrada a_{ij} de A es $(-1)^{i+j} M_{ji}$, donde $M_{ji} := \det(A_{ji})$ es el menor de A que corresponde a la entrada a_{ji}. La matriz $\operatorname{adj} A$ cuya entrada (i, j) es este cofactor se llama la **matriz adjugada** de A. Para formar esta matriz adjugada,[4] entonces, se debe:

(i) reemplazar cada elemento a_{ij} de A por el menor M_{ij} correspondiente;

(ii) multiplicar cada entrada por el *signo* $(-1)^{i+j} = \pm 1$ que corresponde a su lugar en el tablero de damas (6.8);

(iii) tomar la *transpuesta* de la matriz resultante. ◊

[3] Si una de estas $k + 1$ columnas de A fuera una combinación lineal de los demás, por ejemplo si fuera $\boldsymbol{a}_{k+1} = c_1\boldsymbol{a}_1 + \cdots + c_k\boldsymbol{a}_k$, esta relación también seguiría válida para las columnas de la submatriz M, al borrar las filas que sobran.

[4] Algunos autores llaman *matriz adjunta* a esta matriz de cofactores $\operatorname{adj} A$. Sin embargo, cuando $\mathbb{F} = \mathbb{C}$, ese término está reservado para el conjugado hermítico A^*, ya visto. La palabra **adjugada** es una mezcla inelegante de “adjunta” y “conjugada” – y no se encuentra en el Diccionario de la Real Academia Española.

Ejemplo 6.21. En el caso 2×2, ya se ha visto la matriz adjugada en Ejemplo 3.7:

$$\operatorname{adj}\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

Los tres pasos de la Definición 6.20 se ejecutan así:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \xmapsto{\text{(i)}} \begin{bmatrix} d & c \\ b & a \end{bmatrix} \xmapsto{\text{(ii)}} \begin{bmatrix} d & -c \\ -b & a \end{bmatrix} \xmapsto{\text{(iii)}} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$ ◊

Proposición 6.22. *Sea $A \in M_n(\mathbb{F})$. Entonces*

$$\boxed{A\,(\operatorname{adj} A) = (\operatorname{adj} A)\,A = (\det A)\,1_n.} \tag{6.12}$$

Demostración. Para efectos de esta prueba, conviene abreviar $B := \operatorname{adj} A$; se requiere verificar que $\underline{AB = BA = (\det A)\,1_n}$.

Con esta notación, las fórmulas (6.6b) y (6.6c) que calculan $\underline{\det A}$ por expansión en la fila $\boldsymbol{a}'_i$ [respectivamente, en la columna $\boldsymbol{a}_j$] de la matriz A se escriben así:

$$\det A = \sum_{j=1}^{n} a_{ij} b_{ji} = \sum_{i=1}^{n} b_{ji} a_{ij}\,. \tag{6.13}$$

La primera sumatoria es el producto punto de la fila $\boldsymbol{a}'_i$ de A por la columna $\boldsymbol{b}_j$ de B. La segunda sumatoria es el producto punto de la fila $\boldsymbol{b}'_j$ de B por la columna $\boldsymbol{a}_j$ de A. Es decir, las sumatorias corresponden a la entrada diagonal (i,i) de AB y a la entrada diagonal (j,j) de BA, respectivamente. En breve, esta fórmula (6.13) dice que cada entrada diagonal de AB o de BA es igual a $\underline{\det A}$.

Si $i \neq k$, la entrada (k,i) de AB es igual a

$$\boldsymbol{a}'_k \cdot \boldsymbol{b}_i = \sum_{j=1}^{n} a_{kj} b_{ji} = \sum_{j=1}^{n} (-1)^{i+j} a_{kj} M_{ij}, \tag{6.14}$$

y esta es, nuevamente por (6.6b), el determinante de la matriz A' obtenida de A al reemplazar su fila $\boldsymbol{a}'_i$ por su fila $\boldsymbol{a}'_k$. Pero entonces A' posee dos filas iguales, así que $\det A' = 0$, por el Lema 6.16.

Por lo tanto, la sumatoria (6.14) vale 0 cuando $i \neq k$. De igual modo, se ve que si $j \neq l$, la entrada (j,l) de BA es $\boldsymbol{b}'_j \cdot \boldsymbol{a}_l = 0$ por la misma razón.

En resumen, las matrices AB y BA son matrices diagonales con entradas diagonales todas iguales a $\underline{\det A}$; y eso dice que $AB = BA = (\det A)\,1_n$. □

La fórmula (6.12) proporciona un método de calcular el inverso de una matriz no singular A (al ser $\det A \neq 0$, por la Proposición 6.17).

Corolario 6.23. *Si* $A \in M_n(\mathbb{F})$ *con* $\det A \neq 0$*, entonces* A *es invertible, con*

$$A^{-1} = \frac{1}{\det A}\, \operatorname{adj} A. \tag{6.15}$$

Como la ejecución de este proceso de inversión requiere el cálculo de un determinante $n \times n$ y unos n^2 determinantes $(n-1)\times(n-1)$ – todos los menores de A – resulta poco eficiente para $n > 3$. En el caso 2×2, la fórmula (6.15) se reduce a (3.5), ya visto:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

► La fórmula explícita para la solución única de un sistema de n ecuaciones lineales para n variables – en el caso en que haya solución única – el la siguiente **regla de Cramer**.

Proposición 6.24 (Regla de Cramer)**.** *Sean* $A \in M_n(\mathbb{F})$*,* $\boldsymbol{b} \in \mathbb{F}^n$*. Para cada* $j = 1, \dots, n$*, denótese por* $B_j := [\boldsymbol{a}_1 \cdots \boldsymbol{a}_{j-1}\, \boldsymbol{b}\, \boldsymbol{a}_{j+1} \cdots \boldsymbol{a}_n]$ *la matriz obtenida de* A *al reemplazar su columna* $\boldsymbol{a}_j$ *por* $\boldsymbol{b}$*. Entonces el sistema de ecuaciones* $\underline{A\boldsymbol{x} = \boldsymbol{b}}$ *tiene solución única* $\boldsymbol{x} \in \mathbb{F}^n$ *si y solo si* $\det A \neq 0$*, en cuyo caso:*

$$\boxed{x_j = \frac{\det B_j}{\det A}} \quad \textit{para} \quad j = 1, 2, \dots, n. \tag{6.16}$$

Demostración. Es oportuno recordar que $A\boldsymbol{x} = \boldsymbol{b}$ tiene solución única si y solo si la ecuación $A\boldsymbol{x} = \boldsymbol{0}$ tiene solución única, si y solo si $\ker T_A = \{\boldsymbol{0}\}$, si y solo si $n(A) = 0$, si y solo si $r(A) = n$, si y solo si A es invertible, si y solo si $\det A \neq 0$. Cuando $\det A \neq 0$, la solución única deseada es $\boldsymbol{x} = A^{-1}\boldsymbol{b}$.

Si si $\det A \neq 0$, la fórmula (6.12) implica que

$$(\det A)\, \boldsymbol{x} = (\operatorname{adj} A)\, A\, \boldsymbol{x} = (\operatorname{adj} A)\, \boldsymbol{b}.$$

La entrada j de este vector de columna es

$$(\det A)\, x_j = (\text{fila } \#j \text{ de } \operatorname{adj} A)\, \boldsymbol{b} = \sum_{i=1}^{n} (-1)^{i+j} M_{ij}\, b_i = \det B_j$$

donde la última igualdad viene de expandir $\det B_j$ en la columna $\#j$ de B_j, la cual es $\boldsymbol{b}$. La fórmula (6.16) sigue enseguida al dividir ambos lados de esta ecuación por $\det A$. □

Ejemplo 6.25. La regla de Cramer resuelve 2 ecuaciones lineales en 2 variables, así:

$$\begin{Bmatrix} ax_1 + bx_2 = p \\ cx_1 + dx_2 = q \end{Bmatrix} \implies x_1 = \frac{pd - bq}{ad - bc}, \quad x_2 = \frac{aq - pc}{ad - bc}.$$ ◊

6.4. Epílogo: Autovalores y autovectores

Definición 6.26. Sea V un espacio vectorial sobre $\mathbb{F}$ y sea $T\colon V \to V$ un operador lineal. Un **autovalor** (o *valor propio*) de T es un escalar $\lambda \in \mathbb{F}$ tal que la siguiente ecuación tenga una solución no nula:

$$\boxed{T(\boldsymbol{x}) = \lambda\,\boldsymbol{x}} \quad \text{con} \quad \boldsymbol{x} \neq \boldsymbol{0}. \tag{6.17a}$$

Un vector *no nulo* $\boldsymbol{x} \in V$, $\boldsymbol{x} \neq \boldsymbol{0}$, que cumple (6.17a) se llama un **autovector** (o *vector propio*) de T asociado al autovalor λ.

Si $A \in M_n(\mathbb{F})$ es una matriz cuadrada, un **autovalor de** A es un escalar $\lambda \in \mathbb{F}$ tal que

$$\boxed{A\boldsymbol{x} = \lambda\,\boldsymbol{x}} \quad \text{para algún} \quad \boldsymbol{x} \neq \boldsymbol{0} \text{ en } \mathbb{F}^n. \tag{6.17b}$$

Un vector $\boldsymbol{x} \in \mathbb{F}^n$, $\boldsymbol{x} \neq \boldsymbol{0}$, tal que $A\boldsymbol{x} = \lambda\,\boldsymbol{x}$ es un **autovector de** A asociado al autovalor λ. En otras palabras, los autovalores y autovectores de A son los del operador T_A. ◊

Una matriz $A \in M_n(\mathbb{F})$ tiene *a lo sumo* $\underline{n}$ autovalores distintos. Para ver eso, nótese que λ es un autovalor de A si y solo si hay $\boldsymbol{x} \neq \boldsymbol{0}$ con $\underline{(\lambda\,1_n - A)\boldsymbol{x} = \boldsymbol{0}}$, si y solo si $\ker(\lambda\,1_n - T_A) \neq \{\boldsymbol{0}\}$, si y solo si[5]

$$\boxed{\det(\lambda\,1_n - A) = 0.} \tag{6.18}$$

Al calcular el determinante $\det(\lambda\,1_n - A)$ por la fórmula de Leibniz (6.7), se obtiene una suma de términos de la forma:

$$(\lambda - a_{11})(\lambda - a_{22})\cdots(\lambda - a_{nn}) + (\text{términos de grado menor que } n \text{ en } \lambda).$$

Definición 6.27. Si $A \in M_n(\mathbb{F})$ es una matriz cuadrada, su **polinomio característico** es el polinomio $\underline{p_A(t)}$ de grado n con coeficientes en $\mathbb{F}$, definido por

$$\underline{p_A(t) := \det(t\,1_n - A)} = t^n + (\text{términos de grado menor que } n). \tag{6.19}$$

Las **raíces** de este polinomio, los $\lambda \in \mathbb{F}$ tales que $\underline{p_A(\lambda) = 0}$, son los autovalores de A. ◊

Este polinomio de grado n tiene a lo sumo n raíces $t = \lambda \in \mathbb{F}$. En el caso especial de que $\mathbb{F} = \mathbb{C}$, el llamada *teorema fundamental del álgebra* asegura que la ecuación $p_A(t) = 0$ tiene exactamente n raíces complejas (no necesariamente distintas). De ahí se deduce que una *matriz compleja* $A \in M_n(\mathbb{C})$ tiene n autovalores, aunque algunos de ellos pueden coincidir. (Estos se llaman "autovalores repetidos".)

[5]Algunos autores prefieren manejar la ecuación $\det(A - \lambda\,1_n) = 0$. Como este determinante es $(-1)^n \det(\lambda\,1_n - A)$, las dos ecuaciones son equivalentes.

Ejemplo 6.28. Para obtener los autovalores de la matriz

$$A = \begin{bmatrix} -5 & 3 \\ -6 & 4 \end{bmatrix},$$

se resuelve la ecuación

$$\det(t\,1_2 - A) = \begin{vmatrix} t+5 & -3 \\ 6 & t-4 \end{vmatrix} = 0,$$

es decir,

$$t^2 + t - 2 = 0, \quad \text{o bien} \quad (t-1)(t+2) = 0,$$

cuyas raíces son

$$\lambda = 1,\ -2.$$

Para encontrar los autovectores correspondientes, hay que resolver los dos sistemas de ecuaciones homogéneas: $(1_2 - A)\boldsymbol{x} = \mathbf{0}$, y también $(-2\,1_2 - A)\boldsymbol{y} = \mathbf{0}$,

$$\begin{Bmatrix} 6x_1 - 3x_2 = 0 \\ 6x_1 - 3x_2 = 0 \end{Bmatrix}, \qquad \begin{Bmatrix} 3y_1 - 3y_2 = 0 \\ 6y_1 - 6y_2 = 0 \end{Bmatrix}.$$

En ambos casos hay una ecuación redundante. Las soluciones son:

$$\boldsymbol{x} = \begin{bmatrix} a \\ 2a \end{bmatrix}, \qquad \boldsymbol{y} = \begin{bmatrix} b \\ b \end{bmatrix}, \qquad \text{con} \quad a, b \in \mathbb{F}.$$

Fíjese que un múltiplo no cero de un autovector es otro autovector del mismo autovalor, pues $A\boldsymbol{x} = \lambda\boldsymbol{x}$ implica que $A(c\boldsymbol{x}) = c(A\boldsymbol{x}) = c(\lambda\boldsymbol{x}) = \lambda(c\boldsymbol{x})$. ◊

Lema 6.29. *Los autovalores de una matriz* triangular $A \in M_n(\mathbb{F})$ *son sus elementos diagonales:* $\lambda = a_{11}, a_{22}, \dots, a_{nn}$.

Demostración. Se puede suponer que A es triangular superior, porque $\det A^{\mathsf{t}} = \det A$ y a la vez $\det(1\,1_n - A^{\mathsf{t}}) = \det(t\,1_n - A)$: las matrices A y A^{t} tienen el mismo polinomio característico (y por ende tienen los mismos autovalores).

Este polinomio característico es

$$p_A(t) = \det(t\,1_n - A) = \begin{vmatrix} t-a_{11} & -a_{12} & \dots & -a_{1n} \\ 0 & t-a_{22} & \dots & -a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & t-a_{nn} \end{vmatrix} = (t-a_{11})(t-a_{22})\cdots(t-a_{nn})$$

porque la matriz $t\,1_n - A$ también es triangular superior. (Se ha usado la Proposición 6.15.) En este caso el polinomio $p_A(t)$ es un producto de factores de primer grado, y sus raíces son los $\lambda \in \{a_{11}, a_{22}, \dots, a_{nn}\}$. □

Lema 6.30. *Dos matrices semejantes A, B tienen el mismo determinante:* $\underline{\det B = \det A}$.

Demostración. Las matrices $A, B \in M_n(\mathbb{F})$ son semejantes si y solo si hay una matriz invertible $P \in M_n(\mathbb{F})$ tal que $B = P^{-1}AP$. Como $(\det P^{-1})(\det P) = \det(P^{-1}P) = \det 1_n = 1$, se ve que $\det P^{-1} = 1/(\det P)$. Entonces

$$\det B = \det(P^{-1}AP) = (\det P^{-1})(\det A)(\det P) = \det A. \qquad \square$$

Después de la Proposición 6.4 (dos matrices semejantes tiene la misma traza), se definió *la traza de un operador lineal* $T : V \to V$ como la traza de su matriz respecto de cualquier base de V. Del mismo modo, se puede definir el **determinante de un operador lineal** como $\underline{\det T := \det A}$, donde A es la matriz de T respecto de una base de V. El Lema 6.30 anterior dice que un cambio de base no modifica ese determinante.

Ejemplo 6.31. En el Ejemplo 6.28, se obtuvieron los autovalores y autovectores para una matriz 2×2 (real):

$$A = \begin{bmatrix} -5 & 3 \\ -6 & 4 \end{bmatrix} \quad \text{tiene autovectores} \quad \boldsymbol{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad \boldsymbol{y} = \begin{bmatrix} 1 \\ 1 \end{bmatrix},$$

que corresponden con los autovalores $\lambda_1 = 1$ y $\lambda_2 = -2$, respectivamente. Estos autovectores son linealmente independientes (pues no son proporcionales) y forman las columnas de una matriz invertible:

$$P = \begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix}, \quad \text{con inverso} \quad P^{-1} = \begin{bmatrix} -1 & 1 \\ 2 & -1 \end{bmatrix}.$$

Bajo el cambio de base de $\mathbb{R}^2$, de la base estándar $\mathcal{E} = \{\boldsymbol{e}_1, \boldsymbol{e}_2\}$ a esta *base de autovectores* $\mathcal{B} = \{\boldsymbol{x}, \boldsymbol{y}\}$, se transforma $A \mapsto B = P^{-1}AP$, así:

$$P^{-1}AP = \begin{bmatrix} -1 & 1 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} -5 & 3 \\ -6 & 4 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} -1 & 1 \\ -4 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & -2 \end{bmatrix}.$$

Esto dice que la matriz original A es **diagonalizable**, esto es, *semejante a una matriz diagonal*. Más aún, esta matriz diagonal tiene los mismos autovalores que A (consecuencia del Lema 6.30) como sus entradas diagonales (consecuencia del Lema 6.29). ◊

Para "diagonalizar" una matriz $n \times n$, entonces, sería suficiente encontrar n autovectores linealmente independientes. Desafortunadamente, eso no siempre es posible: unas matrices son diagonalizables pero otras no. (Hallar criterios que garantizan que una matriz sea diagonalizable, y decidir qué hacer en el caso contrario, serán temas del curso siguiente.)

Ejemplo 6.32. Esta matriz real $J \in M_2(\mathbb{R})$ no tiene autovalores reales:

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

En efecto, su polinomio característico es

$$p_J(t) := \begin{vmatrix} t & -1 \\ 1 & t \end{vmatrix} = t^2 + 1.$$

Este polinomio es irreducible en $\mathbb{R}_2[t]$, pues no tiene raíces reales.

En cambio, la ecuación cuadrática $t^2 + 1 = 0$, esto es, $t^2 = -1$, sí tiene soluciones complejas, que son $t = i$ y $t = -i$. De este modo:

$$p_J(t) = t^2 + 1 = (t - i)(t + i)$$

sí se factoriza en factores de primer grado con coeficientes complejas – esta es una instancia del teorema fundamental del álgebra. Los autovalores complejas de J, entonces, son $\{i, -i\}$.

Se pueden encontrar los autovectores correspondientes en $\mathbb{C}^2$ por el mismo método del Ejemplo 6.28. Como $i^2 = -1$, $(-i)^2 = -1$, y también $i(-i) = -i^2 = +1$, se verifican estos dos cálculos:

$$\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ i \end{bmatrix} = \begin{bmatrix} i \\ -1 \end{bmatrix} = i \begin{bmatrix} 1 \\ i \end{bmatrix}, \qquad \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ -i \end{bmatrix} = \begin{bmatrix} -i \\ -1 \end{bmatrix} = -i \begin{bmatrix} 1 \\ -i \end{bmatrix}.$$

De ahí se ve que $\left[\begin{smallmatrix} 1 \\ i \end{smallmatrix}\right]$ y $\left[\begin{smallmatrix} 1 \\ -i \end{smallmatrix}\right]$ son autovectores respectivos para $i, -i$ (necesariamente en $\mathbb{C}^2$ pero no en $\mathbb{R}^2$, aunque la matriz original J es real).

En este caso también se puede notar que los dos autovectores encontrados no son proporcionales. Por eso son columnas de una matriz invertible $P \in M_2(\mathbb{C})$:

$$P = \begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix}, \quad \text{con inverso} \quad P^{-1} = -\frac{1}{2i} \begin{bmatrix} -i & -1 \\ -i & 1 \end{bmatrix} = \frac{i}{2} \begin{bmatrix} -i & -1 \\ -i & 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & -\frac{i}{2} \\ \frac{1}{2} & \frac{i}{2} \end{bmatrix}.$$

Por lo tanto,

$$P^{-1} J P = \begin{bmatrix} \frac{1}{2} & -\frac{i}{2} \\ \frac{1}{2} & \frac{i}{2} \end{bmatrix} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix} = \begin{bmatrix} \frac{i}{2} & \frac{1}{2} \\ -\frac{i}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix} = \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}$$

y se concluye que J es *diagonalizable en $M_2(\mathbb{C})$ pero no en $M_2(\mathbb{R})$*. ◊

6.5. Ejercicios sobre trazas y determinantes

Ejercicio 6.1. Sea $\{ E_{ij} : i, j = 1, 2, \dots, n \}$ la base de $M_n(\mathbb{F})$ definida en el Ejercicio 3.13.

(a) Comprobar que $\operatorname{tr} E_{ij} = 1$ si $i = j$; y que $\operatorname{tr} E_{ij} = 0$ si $i \neq j$.

(b) Si $\tau \colon M_n(\mathbb{F}) \to \mathbb{F}$ es una forma lineal tal que

$$\tau(AB) = \tau(BA) \quad \text{para todo} \quad A, B \in M_n(\mathbb{F}),$$

demostrar que existe $c \in \mathbb{F}$ tal que $\underline{\tau(A) = c \operatorname{tr} A}$ para toda $A \in M_n(\mathbb{F})$.

Ejercicio 6.2. Demostrar que *no hay* dos matrices $A, B \in M_n(\mathbb{F})$ tales que $AB - BA = 1_n$.

Ejercicio 6.3. Calcular los determinantes:

$$\text{(a)}: \begin{vmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{vmatrix}, \qquad \text{(b)}: \begin{vmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{vmatrix}, \qquad \text{(c)}: \begin{vmatrix} 9 & -6 & 2 \\ -6 & 8 & -4 \\ 2 & -4 & 4 \end{vmatrix}.$$

Ejercicio 6.4. Calcular los determinantes:

$$\text{(a)}: \begin{vmatrix} 3 & -1 & 2 & 0 \\ 0 & 1 & 0 & -1 \\ 2 & 0 & 1 & 3 \\ -1 & -2 & 1 & 0 \end{vmatrix}, \qquad \text{(b)}: \begin{vmatrix} 7 & -3 & 4 & 2 \\ 5 & 2 & 3 & -2 \\ 3 & 7 & 5 & 3 \\ -5 & 4 & -2 & 5 \end{vmatrix}.$$

Ejercicio 6.5. Verificar la relación:

$$\begin{vmatrix} x & 1 & 0 & 0 & 0 \\ 4 & x & 2 & 0 & 0 \\ 0 & 3 & x & 3 & 0 \\ 0 & 0 & 2 & x & 4 \\ 0 & 0 & 0 & 1 & x \end{vmatrix} = x(x^2 - 4)(x^2 - 16).$$

Ejercicio 6.6. Calcular el determinante

$$\begin{vmatrix} 1 & x+2 & (x+2)^2 \\ 1 & x+3 & (x+3)^2 \\ 1 & x+4 & (x+4)^2 \end{vmatrix}$$

y comprobar que su valor no depende de x. ¿Habrá alguna manera de mostrar que este determinante no depende de x, *sin evaluarlo explícitamente*?

Ejercicio 6.7. Verificar que

$$\begin{vmatrix} x^2 & (x+1)^2 & (x+2)^2 \\ y^2 & (y+1)^2 & (y+2)^2 \\ z^2 & (z+1)^2 & (z+2)^2 \end{vmatrix} = 4(x-y)(x-z)(y-z).$$

Ejercicio 6.8. Verificar, por eliminación gaussiana, que

$$\begin{vmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & 1 & -1 \end{vmatrix} = -160.$$

Ejercicio 6.9. Verificar, por eliminación gaussiana, que

$$\begin{vmatrix} 1 & x & x^2 & x^3 \\ 1 & y & y^2 & y^3 \\ 1 & z & z^2 & z^3 \\ 1 & w & w^2 & w^3 \end{vmatrix} = (x-y)(x-z)(x-w)(y-z)(y-w)(z-w).$$

(Este es un **determinante de Vandermonde**.)

Ejercicio 6.10. Si $A \in M_m(\mathbb{F})$, $B \in M_{m\times n}(\mathbb{F})$, $C \in M_{n\times m}(\mathbb{F})$ y $D \in M_n(\mathbb{F})$, demostrar que

$$\det \begin{bmatrix} A & B \\ 0 & D \end{bmatrix} = \det \begin{bmatrix} A & 0 \\ C & D \end{bmatrix} = (\det A)(\det D),$$

si en ambos casos 0 representa un rectángulo de ceros.

Ejercicio 6.11. Sea $U \in M_n(\mathbb{F})$ la matriz tal que $u_{ij} = 1$ para *todo* (i, j). Demostrar que:

$$\det(U - 1_n) = (-1)^{n-1}(n-1), \qquad \det(U + 1_n) = n + 1.$$

Ejercicio 6.12. Sea $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_m\}$ un conjunto de vectores en el espacio vectorial $\mathbb{R}^n$. Su **determinante de Gram** es

$$\det Z := \begin{vmatrix} \boldsymbol{x}_1\cdot\boldsymbol{x}_1 & \boldsymbol{x}_1\cdot\boldsymbol{x}_2 & \cdots & \boldsymbol{x}_1\cdot\boldsymbol{x}_m \\ \boldsymbol{x}_2\cdot\boldsymbol{x}_1 & \boldsymbol{x}_2\cdot\boldsymbol{x}_2 & \cdots & \boldsymbol{x}_2\cdot\boldsymbol{x}_m \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{x}_m\cdot\boldsymbol{x}_1 & \boldsymbol{x}_m\cdot\boldsymbol{x}_2 & \cdots & \boldsymbol{x}_m\cdot\boldsymbol{x}_m \end{vmatrix},$$

el determinante de la matriz $Z \in M_m(\mathbb{R})$ cuya entrada (i, j) es $z_{ij} := \boldsymbol{x}_i \cdot \boldsymbol{x}_j$. Demostrar que este determinante es 0 si y solo si $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_m\}$ es linealmente *dependiente*, y es positivo cuando $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_m\}$ es linealmente *independiente*.

⟦ Indicación: encontrar una matriz X tal que $Z = X^{\mathrm{t}}X$. ⟧

Ejercicio 6.13. Obtener el rango de esta matriz A, de dos maneras:

$$A := \begin{bmatrix} 3 & 5 & 1 & 4 \\ 2 & -1 & 1 & 1 \\ 5 & 4 & 2 & 5 \end{bmatrix},$$

(a) por cálculo de menores; (b) por cambio a forma escalonada.

Ejercicio 6.14. Calcular la matriz adjugada $(\operatorname{adj} A)$ y la matriz inversa A^{-1}, para

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

Ejercicio 6.15. Resolver estos sistemas de ecuaciones por la regla de Cramer:

$$\left\{\begin{matrix} 3x_1 - 4x_2 = 1 \\ 5x_1 + 7x_2 = 29 \end{matrix}\right\} \quad \text{y} \quad \left\{\begin{matrix} 4y_1 - 3y_2 = 8 \\ 5y_1 + 4y_2 = 41 \end{matrix}\right\}.$$

Ejercicio 6.16. Resolver este sistema de ecuaciones por la regla de Cramer:

$$\begin{aligned} x_1 + 4x_2 - x_3 &= 1 \\ x_1 + x_2 + x_3 &= 0 \\ 2x_1 + 3x_3 &= 0\,. \end{aligned}$$

Ejercicio 6.17. Demostrar que el círculo en $\mathbb{R}^2$ que pasa por tres puntos (x_1, y_1), (x_2, y_2), (x_3, y_3) tiene ecuación:

$$\begin{vmatrix} 1 & x & y & x^2 + y^2 \\ 1 & x_1 & y_1 & x_1^2 + y_1^2 \\ 1 & x_2 & y_2 & x_2^2 + y_2^2 \\ 1 & x_3 & y_3 & x_3^2 + y_3^2 \end{vmatrix} = 0.$$

⟦ Indicación: comprobar primero que la ecuación representa un círculo, y luego que pasa por los tres puntos dados. ⟧

Ejercicio 6.18. (a) Si $A \in M_n(\mathbb{F})$ con $n \geqslant 2$, demostrar que

$$\det(\operatorname{adj} A) = (\det A)^{n-1}.$$

(b) Concluir que $\operatorname{adj}(\operatorname{adj} A) = (\det A)^{n-2} A$, si $n \geqslant 3$.

⟦ Indicación: considerar el producto de tres matrices: $A\,(\operatorname{adj} A)\ \operatorname{adj}(\operatorname{adj} A)$. ⟧

Ejercicio 6.19. Si $A \in M_n(\mathbb{F})$ es una matriz (i) diagonal; (ii) triangular; o (iii) simétrica, demostrar que $\operatorname{adj} A$ es, respectivamente: (i) diagonal; (ii) triangular; (iii) simétrica.

(Por eso, si una matriz invertible es triangular, su inverso es también triangular.)

Ejercicio 6.20. Demostrar (por eliminación gaussiana) que la matriz[6]

$$A := \begin{bmatrix} b & 0 & -1 & 0 & 0 & 0 & 0 & 1 \\ 0 & b & 0 & -1 & 0 & 0 & -1 & 0 \\ -1 & 0 & c & 0 & 0 & -1 & 0 & 0 \\ 0 & -1 & 0 & c & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & b & 0 & 1 & 0 \\ 0 & 0 & -1 & 0 & 0 & b & 0 & 1 \\ 0 & -1 & 0 & 0 & 1 & 0 & c & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & c \end{bmatrix} \tag{6.20}$$

es definida positiva si y solo si: $b > 0$, $c > 0$, $bc > 4$.

Ejercicio 6.21. Si $J = \{j_1, \dots, j_k\} \subseteq \{1, 2, \dots, n\}$, el **menor principal** de una matriz $A \in M_n(\mathbb{F})$ es el determinante de una submatriz A_J formada al borrar las filas y también las columnas cuyos índices *no están en* J. (La diagonal de la submatriz A_J es una parte de la diagonal de A). Por ejemplo, si $n = 3$, los menores principales de A son: a_{11}, a_{22}, a_{33}, $(a_{11}a_{22} - a_{12}a_{21})$, $(a_{11}a_{33} - a_{13}a_{31})$, $(a_{22}a_{33} - a_{23}a_{32})$, $\det A$. Los **menores principales delanteros** son a_{11}, $\det A_{12}$, $\det A_{123}, \dots, \det A$, formados por borrar las *últimas* $(n-k)$ filas y columnas de A; para $n = 3$, éstas son a_{11}, $(a_{11}a_{22} - a_{12}a_{21})$, $\det A$.

Es sabido[7] que una matriz simétrica real $A \in M_n(\mathbb{R})$ es *positiva* si y solo si cada $\det A_J \geqslant 0$; y que A es *definida positiva* si y solo si todos sus menores principales delanteros son > 0.

Usar este criterio (en vez de eliminación gaussiana) para demostrar que la matriz del Ejercicio 6.20 anterior es definida positiva si y solo si: $b > 0$, $c > 0$, $bc > 4$.

Ejercicio 6.22. Calcular los tres autovalores de la matriz

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

Ejercicio 6.23. Calcular los tres autovalores distintos λ_1, λ_2, λ_3 de la matriz

$$A := \begin{bmatrix} 3 & 2 & 2 \\ 2 & 2 & 0 \\ 2 & 0 & 4 \end{bmatrix}.$$

[6] Referencia: Ileana Castillo, *Productos cuánticos en espacios de funciones analíticas*, tesis de licenciatura, UCR, 1988.

[7] Referencia: véase el libro de Gantmacher, § X.4; o el de Horn y Johnson, Teorema 7.2.5.

Resolver las ecuaciones $(\lambda_j 1_3 - A)\boldsymbol{x}_j = \mathbf{0}$ para $j = 1, 2, 3$; estos autovectores $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3\}$ de A son columnas de una matriz invertible $P := [\boldsymbol{x}_1\ \boldsymbol{x}_2\ \boldsymbol{x}_3]$. Verificar que la matriz $P^{-1}AP$ es diagonal y que sus entradas diagonales son los autovalores de A.

Ejercicio 6.24. Si λ es un autovalor de una matriz $A \in M_n(\mathbb{F})$, comprobar que λ^k es autovalor de la matriz A^k.

Si además A es invertible, mostrar que $1/\lambda$ es un autovalor de A^{-1}.

Ejercicio 6.25. Calcular los polinomios característicos y determinar los autovalores de las matrices

$$A = \begin{bmatrix} \cos\theta & -\operatorname{sen}\theta \\ \operatorname{sen}\theta & \cos\theta \end{bmatrix} \quad \text{y} \quad B = \begin{bmatrix} \cosh a & \operatorname{senh} a \\ \operatorname{senh} a & \cosh a \end{bmatrix}.$$

donde $-\pi < \theta \leqslant \pi$, y $a \in \mathbb{R}$.