

# Algebra Lineal



Alonso Takahashi



UNIVERSIDAD  
**NACIONAL**  
DE COLOMBIA

Departamento de Matemáticas

Facultad de Ciencias

# Algebra Lineal

Machu Picchu, octubre de 1979

# Algebra Lineal

Alonso Takahashi

PROFESOR EMÉRITO

Departamento de Matemáticas – Facultad de Ciencias

Universidad Nacional de Colombia

Algebra Lineal

Takahashi Orozco, Alonso R.  
Algebra Lineal  
512.5 2002

© Alonso Takahashi  
Facultad de Ciencias  
Universidad Nacional de Colombia

Primera reimpresión, 2002  
ISBN 958-628-080-2

Portada: Clara I. Bermúdez S.  
Con detalle de un grabado de Maurits Cornelis Escher.

Impresión:  
Unibiblos  
Universidad Nacional de Colombia  
Bogotá, D.C.  
COLOMBIA

# Prefacio

Esta es una obra introductoria, orientada a exponer nociones, procedimientos y resultados básicos. Su propósito no es el desarrollo completo y minucioso de una materia sino la organización de un trabajo autónomo de lectura; en lugar de suplantar al profesor y al estudiante se ha buscado conseguir un lector.

Manteniendo un nivel elemental, y en poco más de cien páginas, se han destacado los aspectos que hacen del álgebra lineal una materia privilegiada: su simplicidad conceptual, en contraste con su poder casi sorprendente; la naturaleza geométrica de sus nociones cuya ulterior algebrización permite aplicarlas en los ámbitos más diversos; el carácter algorítmico de sus procedimientos y, en especial, la magia de la forma escalonada.

En la selección de cada tema y en su presentación se ha intentado revelar la esencia del asunto con el fin de configurar un núcleo alrededor del cual se puedan estructurar distintos cursos de acuerdo con necesidades especiales. No se dan muchos pormenores en las deducciones pero se ha realizado un esfuerzo considerable para señalar en cada caso los puntos de partida y el camino a seguir para llegar a las conclusiones enunciadas. Aunque no hay ejercicios explícitos, la lectura del texto es en sí misma un ejercicio continuo del cual hace parte la verificación de los ejemplos, los cuales esperan, en su mayoría, una elaboración por parte del lector. El resultado es un texto cuya lectura supone mucha escritura.

En el conocimiento genuino de un tema, así como en los seres más organizados, pueden distinguirse tres componentes: un esqueleto, un sistema muscular y un fluido vital. Este libro tiene más de lo primero que de lo

segundo y, por ello, necesita ser complementado con problemas rutinarios, extensiones de la teoría, desarrollos específicos y aplicaciones. Lo tercero sólo puede aportarlo, por supuesto, el lector.

Para facilitar el uso del libro se incluye una tabla de abreviaturas y símbolos y un índice alfabético detallado. Las fórmulas se numeran dentro de cada sección; por ejemplo, (2), §19 (o §19, (2)), se refiere a la ecuación (2) de la sección 19, mientras que (2), §36 es la ecuación (2) de la sección 36; cuando la referencia tiene lugar en la misma sección ésta no se menciona.

La realización de este trabajo contó con el apoyo del Departamento de Matemáticas de la Universidad Nacional. Al profesor Gustavo Rubiano le agradezco el estímulo y la ayuda para llevar a cabo esta edición.

# Índice

<b>Prefacio</b>	<b>-4</b>
<b>Preliminares</b>	<b>1</b>
<b>1 Eliminación</b>	<b>9</b>
§ 1 Soluciones de ecuaciones lineales . . . . .	9
§ 2 Problemas equivalentes . . . . .	10
§ 3 Notaciones . . . . .	11
§ 4 El método de eliminación . . . . .	12
§ 5 Descripción general del método de eliminación . . . . .	17
§ 6 Conclusiones . . . . .	21
§ 7 Comentarios . . . . .	21
<b>2 Dimensión</b>	<b>23</b>
§ 8 Espacios vectoriales . . . . .	23
§ 9 Subespacios . . . . .	25
§ 10 Transformaciones lineales e isomorfismos . . . . .	25
§ 11 Sistemas de vectores. Combinaciones lineales . . . . .	27
§ 12 La transformación asociada a un sistema . . . . .	28
§ 13 Generación . . . . .	29
§ 14 Libertad . . . . .	30
§ 15 Bases . . . . .	32
§ 16 Sistemas equivalentes . . . . .	35
§ 17 Rango . . . . .	37
<b>3 Linealidad</b>	<b>39</b>
§ 18 Álgebra de transformaciones . . . . .	39
§ 19 Álgebra de matrices . . . . .	40
§ 20 Transformaciones y matrices . . . . .	44
§ 21 Nulidad y rango . . . . .	45
§ 22 Inversión . . . . .	45
§ 23 Cambio de base . . . . .	47



---

§ 24	Matrices elementales . . . . .	48
§ 25	Formas bilineales . . . . .	50
<b>4</b>	<b>Diagonalización</b>	<b>52</b>
§ 26	Autovalores y autovectores . . . . .	53
§ 27	Operadores y matrices diagonalizables . . . . .	56
§ 28	Descomposición espectral . . . . .	58
§ 29	Diagonalización simultánea . . . . .	58
<b>5</b>	<b>Ortogonalidad</b>	<b>60</b>
§ 30	Producto interno . . . . .	60
§ 31	Proyección ortogonal de un vector sobre otro . . . . .	62
§ 32	Proyección ortogonal de un vector sobre un subespacio . . . . .	64
§ 33	Ortonormalización . . . . .	65
§ 34	Bases ortonormales . . . . .	65
§ 35	Proyecciones ortogonales . . . . .	66
§ 36	Diagonalización ortonormal . . . . .	67
§ 37	Representación de funcionales. Adjunción . . . . .	68
§ 38	El álgebra $\mathcal{L}(\mathcal{U})$ . . . . .	70
§ 39	Formas sesquilineales . . . . .	72
<b>6</b>	<b>Normalidad</b>	<b>74</b>
§ 40	Operadores normales . . . . .	74
§ 41	El teorema espectral . . . . .	75
§ 42	Operadores autoadjuntos, unitarios y positivos . . . . .	76
§ 43	Matrices simétricas . . . . .	79
§ 44	Matrices ortogonales . . . . .	80
§ 45	Matrices positivas . . . . .	81
<b>7</b>	<b>Aplicaciones</b>	<b>84</b>
§ 46	Rectas y planos . . . . .	84
§ 47	Proyecciones y mínimos cuadrados . . . . .	87
§ 48	Ángulos . . . . .	89
§ 49	Volúmenes . . . . .	89
§ 50	El producto externo en $\mathbb{R}^3$ . . . . .	91
§ 51	Aplicaciones a la estadística . . . . .	92
	<b>Apéndice</b>	<b>97</b>
§ 52	Determinantes . . . . .	97
§ 53	La forma de Jordan . . . . .	105
§ 54	Soluciones óptimas . . . . .	110
	<b>Símbolos y abreviaturas</b>	<b>114</b>



# Preliminares

En esta sección se recuerdan varias nociones y se establecen algunas convenciones.

DEMOSTRACIONES. En el tratamiento cotidiano de la matemática la formalización completa es, de hecho, inabordable. No hay una descripción rigurosa del rigor; una demostración se considera válida cuando hace ver la certeza de sus afirmaciones a un público convenientemente familiarizado con los términos y las relaciones empleados y con los procedimientos aceptados para su encadenamiento.

DEFINICIONES. Los términos técnicos tienen como fin abreviar el texto. Su adopción se indica **destacándolos** tipográficamente y fijando su sentido la primera vez que aparecen.

FUNCIONES. Una **función**  $T$  **de** un conjunto  $\mathcal{U}$  **en** un conjunto  $\mathcal{V}$  asigna a cada elemento  $u$  de  $\mathcal{U}$  un elemento, bien determinado,  $v$  de  $\mathcal{V}$ , el cual se llama la **imagen** de  $u$  por  $T$ , o el **valor** de  $T$  en  $u$  y se denota  $T(u)$  o  $Tu$ . Para aludir a esta situación se usan notaciones como

$$\mathcal{U} \xrightarrow{T} \mathcal{V} \quad , \quad v = Tu .$$

Dadas  $\mathcal{U} \xrightarrow{T} \mathcal{V}$  y  $\mathcal{V} \xrightarrow{S} \mathcal{W}$ , se define la función **compuesta**  $S \circ T$  (o, simplemente  $ST$ ),

$$\mathcal{U} \xrightarrow{\quad ST \quad} \mathcal{W}$$

conviniendo que, para cada  $u$  de  $\mathcal{U}$ :

$$(ST)(u) = S(T(u)).$$

De acuerdo con esta definición, si  $\mathcal{U} \xrightarrow{T} \mathcal{V} \xrightarrow{S} \mathcal{W} \xrightarrow{R} \mathcal{X}$  entonces

$$R(ST) = (RS)T.$$

Para cada conjunto  $\mathcal{U}$  existe una única función  $R$ , de  $\mathcal{U}$  en  $\mathcal{U}$ , tal que, cualesquiera sean  $T$  y  $S$ , si

$$\mathcal{X} \xrightarrow{T} \mathcal{U} \xrightarrow{S} \mathcal{Y},$$

entonces  $RT = T$  y  $SR = S$ . Esta función  $R$  se llama la **identidad** de  $\mathcal{U}$ , se denota  $1$  (ó  $1_{\mathcal{U}}$ , si hay lugar a confusión) y está definida por  $1(u) = u$  (para todo  $u$  de  $\mathcal{U}$ ).

Si  $T : \mathcal{U} \rightarrow \mathcal{V}$  es una función y  $\mathcal{M} \subset \mathcal{U}$ , la **imagen** de  $\mathcal{M}$  por  $T$  es el subconjunto  $T\langle\mathcal{M}\rangle$  de  $\mathcal{V}$  formado por las imágenes de los elementos de  $\mathcal{M}$ . La imagen de  $\mathcal{U}$  se llama el **recorrido** de  $T$  y se denota  $\mathcal{R}_T$ . Para todo  $v$  de  $\mathcal{V}$  se tiene que

$$v \in \mathcal{R}_T \text{ si y sólo si existe un } u \text{ en } \mathcal{U} \text{ tal que } v = T(u).$$

A menudo escribiremos  $\langle T \rangle$  en lugar de  $\mathcal{R}_T$ . Cuando  $\mathcal{R}_T = \mathcal{V}$  se dice que  $T$  es **sobreyectiva**. Por otra parte, si siempre que  $u_1 \neq u_2$  se tiene que  $T(u_1) \neq T(u_2)$  (es decir, si  $T(u_1) = T(u_2)$  implica  $u_1 = u_2$ ), se dice que  $T$  es **inyectiva**. Cuando  $T$  es inyectiva y sobreyectiva se dice que es **biyectiva**; en este caso, cada elemento  $v$  de  $\mathcal{V}$  determina un único elemento  $u$  de  $\mathcal{U}$  tal que  $T(u) = v$ . Se define así una función  $T^{-1} : \mathcal{V} \rightarrow \mathcal{U}$ , la **inversa** de  $T$ , tal que, para todo  $u$  de  $\mathcal{U}$  y todo  $v$  de  $\mathcal{V}$ :

$$u = T^{-1}(v) \text{ si y sólo si } T(u) = v.$$

Si dos funciones,  $T : \mathcal{U} \rightarrow \mathcal{V}$  y  $S : \mathcal{V} \rightarrow \mathcal{U}$ , son tales que  $ST = 1_{\mathcal{U}}$  entonces  $T$  es inyectiva y  $S$  es sobreyectiva.

De una función  $T$  de  $\mathcal{U}$  en  $\mathcal{U}$  se dice que es un **operador** en  $\mathcal{U}$ ; un subconjunto  $\mathcal{M}$  de  $\mathcal{U}$  es  $T$ -**invariante** si  $T\langle\mathcal{M}\rangle \subset \mathcal{M}$ . En este caso  $T$  puede considerarse como un operador en  $\mathcal{M}$ , esto es,  $T$  define, de manera obvia, un operador en  $\mathcal{M}$ , a saber, la restricción de  $T$  a  $\mathcal{M}$ .

VECTORES GEOMÉTRICOS Y SU REPRESENTACIÓN ANALÍTICA. En el plano los vectores se representan por medio de segmentos orientados (flechas). Se considera que dos flechas representan el mismo vector si son paralelas y tienen la misma longitud y sentido (cfr. § 46). Dados dos vectores  $\mathbf{a}$  y  $\mathbf{b}$ , y un escalar  $t$ , la suma  $\mathbf{a} + \mathbf{b}$  corresponde a la diagonal del paralelogramo  $\mathcal{M}$  determinado por  $\mathbf{a}$  y  $\mathbf{b}$  (ley del paralelogramo); el producto  $t\mathbf{a}$  es el vector paralelo a  $\mathbf{a}$ , cuya longitud es  $|t|$  veces la longitud de  $\mathbf{a}$  y cuyo sentido es igual u opuesto al de  $\mathbf{a}$  según que  $t > 0$  ó  $t < 0$ . Por último, el producto interno  $\mathbf{a} \cdot \mathbf{b}$ , escrito también  $\langle \mathbf{a}, \mathbf{b} \rangle$ , es igual al producto de la longitud de  $\mathbf{a}$  por la longitud de  $\mathbf{b}$  por el coseno del ángulo entre  $\mathbf{a}$  y  $\mathbf{b}$ .

Llamando  $\mathcal{U}$  al plano e introduciendo un sistema rectangular de coordenadas cartesianas, cada vector está representado por una flecha que parte del origen, pudiendo entonces identificarse con un punto del plano (el extremo de la flecha). En consecuencia, los vectores del plano pueden representarse analíticamente por medio de **parejas ordenadas** de números:

$$\mathbf{a} = (x, y) \quad , \quad \mathbf{b} = (u, v) \quad , \quad \text{etc.}$$

Estas parejas son *ordenadas* porque

$$\mathbf{a} = \mathbf{b} \quad \text{si y sólo si} \quad x = u \quad \text{y} \quad y = v \quad . \quad (1)$$

La pareja  $(0, 0)$  es el **vector cero** y se denota  $\mathbf{0}$ .

Consideraciones geométricas sencillas revelan la expresión analítica de las operaciones, a saber:

$$\mathbf{a} + \mathbf{b} = (x, y) + (u, v) = (x + u, y + v) \quad (2)$$

$$t\mathbf{a} = t(x, y) = (tx, ty) \quad , \quad (3)$$

y la fórmula para  $\cos(\alpha - \beta)$  muestra que

$$\mathbf{a} \cdot \mathbf{b} = xu + yv \quad . \quad (4)$$

Por último, la **longitud**  $\|\mathbf{a}\|$  del vector  $\mathbf{a}$  viene dada por

$$\|\mathbf{a}\| = \sqrt{x^2 + y^2} \quad . \quad (5)$$

Los vectores unitarios

$$\mathbf{e}_1 = (1, 0) \quad \text{y} \quad \mathbf{e}_2 = (0, 1)$$

permiten escribir un vector cualquiera  $\mathbf{a} = (x, y)$  en la forma

$$\mathbf{a} = x\mathbf{e}_1 + y\mathbf{e}_2 .$$

TRANSFORMACIONES. Una transformación  $\tau$  del plano transforma cada punto  $a$  del plano en un punto  $b$  del plano (el cual puede coincidir con el primero). En otros términos,  $\tau$  es un operador en el plano  $\mathcal{U}$ , y puede visualizarse por medio de diagramas como el de la figura 1.

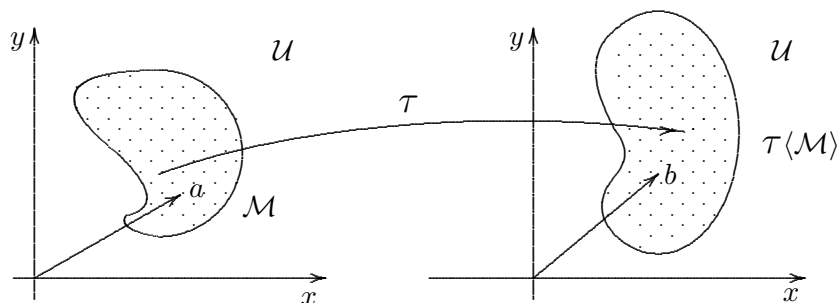


Figura 1

Supongamos, por ejemplo, que  $\rho$  es una rotación de  $30^\circ$  en sentido positivo, esto es,  $\rho$  es la transformación que a cada vector lo hace girar  $30^\circ$  en sentido contrario al de las agujas del reloj, y que  $\pi$  es la proyección (ortogonal) sobre la recta  $y = x$ . Pueden, entonces, considerarse las transformaciones  $\rho\pi$ ,  $\pi\rho$ ,  $\rho^2 (= \rho\rho)$ ,  $\rho^3$ ,  $\pi^2$ , etc; también  $\rho^{-1}$  (pero no  $\pi^{-1}$ ). Tomando por ejemplo,  $\mathbf{a} = (-1, 11)$  y usando un poco de geometría y trigonometría se puede ver que

$$\begin{aligned} \rho\pi(\mathbf{a}) &= \frac{5}{2}(\sqrt{3} - 1, \sqrt{3} + 1) \quad \text{y} \\ \pi\rho(\mathbf{a}) &= \frac{1}{2}(5\sqrt{3} - 6)(1, 1) . \end{aligned}$$

Observaciones elementales permiten representar y manipular más eficientemente estas transformaciones: considerando el paralelogramo  $\mathcal{M}$  determinado por dos vectores  $\mathbf{a}$  y  $\mathbf{b}$  y su imagen  $\rho(\mathcal{M})$  bajo la rotación  $\rho$  (ver figura 2), se ve que:

$$\rho(\mathbf{a} + \mathbf{b}) = \rho\mathbf{a} + \rho\mathbf{b} .$$

Análogamente se observa que

$$\rho(t\mathbf{a}) = t\rho\mathbf{a}$$

para cualquier vector  $\mathbf{a}$  y cualquier número  $t$ .

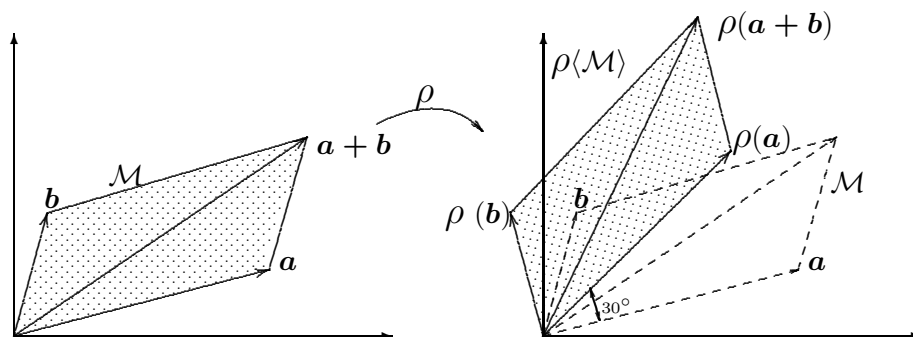


Figura 2

Se puede comprobar que la transformación  $\pi$  también goza de estas dos propiedades. De una función que tenga estas propiedades diremos que es una **transformación lineal**.

Para conocer una transformación lineal basta conocerla en dos vectores no colineales. Los vectores  $\mathbf{e}_1$  y  $\mathbf{e}_2$  constituyen un caso particularmente simple; en efecto, si  $\mathcal{T}$  es una transformación lineal se tiene que, cualquiera sea  $\mathbf{a} = (x, y)$ :

$$\mathcal{T}\mathbf{a} = \mathcal{T}(xe_1 + ye_2) = \mathcal{T}(xe_1) + \mathcal{T}(ye_2) = x\mathcal{T}e_1 + y\mathcal{T}e_2 .$$

Si escribimos  $\mathcal{T}\mathbf{a} = (u, v)$  y además  $\mathcal{T}e_1 = (\alpha_1, \alpha_2)$  y  $\mathcal{T}e_2 = (\beta_1, \beta_2)$  se puede ver que

$$\begin{cases} u = \alpha_1 x + \beta_1 y \\ v = \alpha_2 x + \beta_2 y . \end{cases} \quad (6)$$

Los cuatro números  $\alpha_1, \beta_1, \alpha_2, \beta_2$  determinan completamente la transformación  $\mathcal{T}$  y constituyen su **matriz**,

$$M = \begin{bmatrix} \alpha_1 & \beta_1 \\ \alpha_2 & \beta_2 \end{bmatrix} ,$$

la cual también se denota  $[\mathcal{T}]$ .

Nótese que los elementos que aparecen en la primera (resp. segunda) columna de la matriz de  $\mathcal{T}$  son, precisamente, las componentes del vector  $\mathcal{T}\mathbf{e}_1$  (resp. del vector  $\mathcal{T}\mathbf{e}_2$ ).

Conviniendo representar los vectores  $(x, y)$  en la forma

$$\begin{bmatrix} x \\ y \end{bmatrix},$$

las ecuaciones (6), las cuales expresan las componentes de la imagen  $\mathcal{T}\mathbf{a}$  en términos de las componentes del vector original  $\mathbf{a}$ , pueden escribirse, en forma más compacta, así:

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \alpha_1 & \beta_1 \\ \alpha_2 & \beta_2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}. \quad (7)$$

Si  $\sigma$  es otra transformación lineal cuya matriz es

$$L = \begin{bmatrix} \xi_1 & \eta_1 \\ \xi_2 & \eta_2 \end{bmatrix},$$

puede comprobarse que  $\sigma\mathcal{T}$  es de nuevo una transformación lineal y que su matriz  $K$  puede obtenerse a partir de  $L$  y de  $M$  mediante la siguiente regla de multiplicación:

$$K = LM = \begin{bmatrix} \xi_1 & \eta_1 \\ \xi_2 & \eta_2 \end{bmatrix} \begin{bmatrix} \alpha_1 & \beta_1 \\ \alpha_2 & \beta_2 \end{bmatrix} = \begin{bmatrix} \xi_1\alpha_1 + \eta_1\alpha_2 & \xi_1\beta_1 + \eta_1\beta_2 \\ \xi_2\alpha_1 + \eta_2\alpha_2 & \xi_2\beta_1 + \eta_2\beta_2 \end{bmatrix}.$$

Resulta entonces que  $[\sigma\mathcal{T}] = [\sigma][\mathcal{T}]$ , entendiéndose este producto en el sentido que acaba de indicarse. Esta forma de multiplicación es del mismo tipo de la adoptada implícitamente en las fórmulas (6) y (7).

Por analogía con los vectores, puede también definirse la suma de dos matrices y el producto de una matriz por un número:

$$\begin{bmatrix} \xi_1 & \eta_1 \\ \xi_2 & \eta_2 \end{bmatrix} + \begin{bmatrix} \alpha_1 & \beta_1 \\ \alpha_2 & \beta_2 \end{bmatrix} = \begin{bmatrix} \xi_1 + \alpha_1 & \eta_1 + \beta_1 \\ \xi_2 + \alpha_2 & \eta_2 + \beta_2 \end{bmatrix},$$

$$\lambda \begin{bmatrix} \xi_1 & \eta_1 \\ \xi_2 & \eta_2 \end{bmatrix} = \begin{bmatrix} \lambda\xi_1 & \lambda\eta_1 \\ \lambda\xi_2 & \lambda\eta_2 \end{bmatrix}.$$



Volviendo al ejemplo inicial, puede verse que

$$[\rho] = \begin{bmatrix} \sqrt{3}/2 & -1/2 \\ 1/2 & \sqrt{3}/2 \end{bmatrix}, \quad [\pi] = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix},$$

con lo cual puede calcularse,  $\rho\pi(\mathbf{a})$ , para cualquier  $\mathbf{a}$ . En efecto,

$$\rho\pi(\mathbf{a}) = [\rho\pi] \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \sqrt{3}/2 & -1/2 \\ 1/2 & \sqrt{3}/2 \end{bmatrix} \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}.$$

En particular, si  $\mathbf{a} = (-1, 11)$  entonces  $\rho\pi(\mathbf{a}) = \frac{5}{2}(\sqrt{3} - 1, \sqrt{3} + 1)$ .

ESPACIOS EUCLIDIANOS. Las consideraciones hechas en relación con el plano pueden generalizarse introduciendo, para cada  $n \geq 1$ , el espacio euclidiano  $n$ -dimensional cuyos elementos son las  $n$ -plas ordenadas

$$(x_1, \dots, x_n),$$

de números reales, a las cuales se extienden directamente las definiciones (1) a (5).

MATRICES. Según se ha observado, la necesidad de manejar entes cuyas características numéricas no pueden disponerse eficientemente en forma de  $n$ -plas, motiva la introducción de dispositivos más complejos: una **matriz**  $m \times n$  es una tabla formada por  $mn$  componentes numéricas dispuestas en  $m$  filas y  $n$  columnas. Si  $A$  es una matriz, sus **componentes** se denotan  $a_j^i$  ó  $a_{ij}$ , lo cual se lee “a,i,j”. Nótese que  $a_j^i$ , la  $(i, j)$ -componente de  $A$ , es el elemento situado en el *lugar*  $(i, j)$ , es decir, en la fila  $i$  y la columna  $j$ . La fórmula  $A = [a_j^i]$  es una abreviación de

$$A = \begin{bmatrix} a_1^1 & \cdots & a_n^1 \\ \vdots & \cdots & \vdots \\ a_1^m & \cdots & a_n^m \end{bmatrix}.$$

Denotando por  $\mathbf{a}^i$  la  $i$ -ésima fila y por  $\mathbf{a}_j$  la  $j$ -ésima columna, puede escribirse

$$A = \begin{bmatrix} \mathbf{a}^1 \\ \vdots \\ \mathbf{a}^m \end{bmatrix} \quad \text{y también} \quad A = [\mathbf{a}_1, \dots, \mathbf{a}_n].$$

Esto quiere decir que no queremos distinguir, por ejemplo, entre

$$\left[ \begin{bmatrix} 2 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 8 \end{bmatrix} \right] \quad \text{y} \quad \begin{bmatrix} 2 & -1 & 0 \\ 0 & 1 & 8 \end{bmatrix} .$$

El uso de superíndices tiene muchas ventajas y una desventaja: la posible confusión con exponentes; los usaremos, a discreción, cuando no exista este riesgo. Nótese que  $a_j^i$  es la  $i$ -ésima componente de  $\mathbf{a}_j$  y la  $j$ -ésima componente de  $\mathbf{a}^i$ .

El conjunto de todas las matrices  $m \times n$  se denota  $\mathcal{M}_{m,n}$ , y a él pueden extenderse las nociones de igualdad, suma y multiplicación por escalar.

La matriz  $B = [b_j^i]$ , donde  $b_j^i = a_i^j$  para todo  $i$  y todo  $j$ , es la **traspuesta** de  $A$  y se denota  $A^t$ . Es claro que si  $A$  está en  $\mathcal{M}_{m,n}$  entonces  $A^t$  está en  $\mathcal{M}_{n,m}$ .

Para cada  $n$  el conjunto  $\mathcal{M}_{n,1}$  (resp.  $\mathcal{M}_{1,n}$ ) se denota  $\mathbb{R}^n$  (resp.  $\mathbb{R}_n$ ) y sus elementos se llaman  $n$ -**vectores columna** (resp.  $n$ -**vectores fila**). El **producto** de un elemento  $\mathbf{a}$  de  $\mathbb{R}_n$  (de componentes  $a_i$ ) por un elemento  $\mathbf{x}$  de  $\mathbb{R}^n$  (de componentes  $x^i$ ) es la suma de los productos de las componentes correspondientes:

$$\mathbf{a}\mathbf{x} = [a_1, \dots, a_n] \begin{bmatrix} x^1 \\ \vdots \\ x^n \end{bmatrix} = a_1x^1 + \dots + a_nx^n . \quad (8)$$

Esto generaliza la convención implícita en las fórmulas (7) y (6).

Tanto  $\mathbb{R}^n$  como  $\mathbb{R}_n$  son representaciones matriciales del espacio euclidiano  $n$ -dimensional. Los elementos de  $\mathbb{R}^1$  y de  $\mathbb{R}_1$  se identifican con los números reales: tampoco queremos distinguir entre  $[a]$  y  $a$ .

En lugar de  $\mathcal{M}_{n,n}$  se escribe  $\mathcal{M}_n$  y sus elementos se llaman **matrices** (cuadradas) de **orden**  $n$ . En una matriz cuadrada  $A$  los elementos  $a_i^i$  forman la **diagonal** (principal) de  $A$ .

# Capítulo 1

## Eliminación

En este capítulo se describe un algoritmo, es decir, un procedimiento sistemático, una receta, para analizar y resolver ecuaciones lineales simultáneas.

### § 1 Soluciones de ecuaciones lineales

La expresión

$$-5 \{ \ }_1 + 3 \{ \ }_2 - \{ \ }_3 = -2$$

es una manera, sin duda extravagante, de escribir una **ecuación lineal** con tres **incógnitas**. Las ternas *ordenadas*  $(-17, 1, 90)$ ,  $(1, 2, 3)$ ,  $(0, 0, 2)$ ,  $(2, 3, 1)$  son **soluciones** de ella. En particular,  $(1, 2, 3)$  y  $(2, 3, 1)$  son dos soluciones *distintas*. En cambio  $(3, 1, 2)$  *no* es una solución.

En general, una ecuación lineal con  $n$  incógnitas puede escribirse en la forma

$$E : \quad a_1 \{ \ }_1 + \cdots + a_n \{ \ }_n = b .$$

Los números  $a_1, \dots, a_n$  son los **coeficientes** y  $b$  es el **término independiente**. Cuando éste es cero se dice que la ecuación es **homogénea**. Una solución de  $E$  es una  $n$ -pla ordenada  $(s^1, \dots, s^n)$  de números que

**satisface** la ecuación, es decir, tal que

$$a_1 s^1 + \cdots + a_n s^n = b .$$

Un **sistema** de ecuaciones lineales o **problema lineal** (P) consta de un cierto número  $m$  de ecuaciones (lineales) con un cierto número  $n$  de incógnitas y está determinado por su **matriz**,

$$\begin{array}{cccc} & (1) & \cdots & (n) \\ \hline \left[ \begin{array}{ccc|c} a_1^1 & \cdots & a_n^1 & b^1 \\ \cdot & \cdots & \cdot & \vdots \\ a_1^m & \cdots & a_n^m & b^m \end{array} \right] & & & (P) \end{array}$$

donde, por ejemplo, la segunda fila

$$f_2 \quad : \quad [a_1^2, \dots, a_n^2 ; b^2]$$

representa la segunda ecuación del problema, a saber

$$a_1^2 \{ \quad \}_1 + \cdots + a_n^2 \{ \quad \}_n = b^2 .$$

Una **solución** de (P) es una  $n$ -pla  $(s^1, \dots, s^n)$  que satisface *cada una* de las  $m$  ecuaciones del problema; las soluciones de (P) constituyen su **conjunto solución**. Un problema es **consistente** si tiene, por lo menos, una solución. En caso contrario, es decir, cuando el conjunto solución es vacío, se dice que es **inconsistente**.

Al sustituir por ceros todos los términos independientes se obtiene el sistema o problema **homogéneo** asociado a (P); su conjunto solución se llama el **núcleo** del sistema. Un problema homogéneo es siempre consistente pues  $(0, 0, \dots, 0)$  es una solución (es la llamada solución **trivial**).

## § 2 Problemas equivalentes

Dos problemas se dicen **equivalentes** si tienen el mismo conjunto solución. Es claro que al cambiar el orden en que aparecen las ecuaciones en

un problema se obtiene otro equivalente; y también que si a dos problemas equivalentes se les agregan *las mismas* ecuaciones se obtienen dos problemas que también son equivalentes (entre sí). Por último, al multiplicar una ecuación (esto es, cada uno de sus coeficientes y su término independiente) de un problema por un número, distinto de 0, resulta otro equivalente. Dos problemas inconsistentes son equivalentes (no hay una solución de uno de ellos que no lo sea del otro).

Dadas dos ecuaciones

$$E : \quad a_1 x_1 + \cdots + a_n x_n = h ,$$

$$F : \quad b_1 x_1 + \cdots + b_n x_n = k ,$$

y un número  $\lambda$ , la ecuación

$$E' : \quad (a_1 + \lambda b_1)x_1 + \cdots + (a_n + \lambda b_n)x_n = h + \lambda k,$$

obtenida sumándole, a la primera,  $\lambda$  veces la segunda, se denota  $E + \lambda F$ . Por cómputo directo se comprueba que toda solución de  $E$  y  $F$  lo es también de  $E'$ ; como, a su vez,  $E$  es precisamente  $E' + (-\lambda)F$ , se concluye que los problemas  $\{E, F\}$  y  $\{E', F\}$  son equivalentes. Por lo tanto: si a una de las ecuaciones de un problema lineal se le suma un múltiplo de otra de ellas, el nuevo problema es equivalente al original.

### § 3 Notaciones

Las incógnitas se denotan usualmente por medio de letras como  $t, x, y$ , etc con o sin índices. Así, en la formulación matricial (P) podemos escribir  $x^1, \dots, x^n$  en lugar de  $(1), \dots, (n)$  resultando la siguiente forma explícita del problema:

$$\begin{cases} a_1^1 x^1 + \cdots + a_n^1 x^n = b^1 \\ \vdots \\ a_1^m x^1 + \cdots + a_n^m x^n = b^m \end{cases}$$

lo cual también puede escribirse en la forma

$$x^1 \mathbf{a}_1 + \cdots + x^n \mathbf{a}_n = \mathbf{b} ,$$

siendo  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$  las columnas de coeficientes y  $\mathbf{b}$  la de términos independientes.

Otra forma de escribir el problema es

$$A\mathbf{x} = \mathbf{b} ,$$

explícitamente

$$\underbrace{\begin{bmatrix} a_1^1 & \cdots & a_n^1 \\ \vdots & \cdots & \vdots \\ a_1^m & \cdots & a_n^m \end{bmatrix}}_A \underbrace{\begin{bmatrix} x^1 \\ \vdots \\ x^n \end{bmatrix}}_{\mathbf{x}} = \underbrace{\begin{bmatrix} b^1 \\ \vdots \\ b^m \end{bmatrix}}_{\mathbf{b}} ,$$

donde el *producto*  $A\mathbf{x}$  es el vector columna cuyas componentes se obtienen multiplicando cada una de las filas de  $A$  por  $\mathbf{x}$ . Anotemos de paso que esta multiplicación (de una matriz  $m \times n$  por un  $n$ -vector columna) goza de las propiedades

$$A(\mathbf{x}_1 + \mathbf{x}_2) = A\mathbf{x}_1 + A\mathbf{x}_2 \quad , \quad A(\lambda\mathbf{x}) = \lambda A\mathbf{x} \quad , \quad (\text{L})$$

según se comprueba efectuando las operaciones indicadas.

Con esta notación el problema homogéneo asociado es

$$A\mathbf{x} = \mathbf{0} . \quad (\text{H})$$

Su conjunto solución y, en general, el conjunto solución de  $A\mathbf{x} = \mathbf{b}$ , es un subconjunto de  $\mathbb{R}^n$ .

## § 4 El método de eliminación

Si en un problema lineal las incógnitas pueden *clasificarse en dos grupos* de tal manera que cada una de las primeras (*incógnitas principales*) pueda expresarse en términos de las segundas (*parámetros*) el conjunto solución puede describirse inmediatamente. Ejemplo: en el problema

$$\begin{cases} -3x^1 + \sqrt{2}x^2 & - 6x^4 + 4x^5 = -1 \\ & 6x^2 + 2x^3 + 2x^4 - x^5 = 4, \end{cases}$$

las incógnitas  $x^1$  y  $x^3$  pueden despejarse en términos de  $x^2$ ,  $x^4$  y  $x^5$  :

$$\begin{aligned}x^1 &= \frac{\sqrt{2}}{3}x^2 - 2x^4 + \frac{4}{3}x^5 + \frac{1}{3} \\x^3 &= -3x^2 - x^4 + \frac{1}{2}x^5 + 2 .\end{aligned}$$

Las soluciones se obtienen dándole a los parámetros  $x^2$ ,  $x^4$  y  $x^5$  valores arbitrarios y calculando los correspondientes valores de  $x^1$  y  $x^3$ . Por ejemplo, para  $x^2 = 3$ ,  $x^4 = 0$  y  $x^5 = -1/4$  se obtiene la solución  $(\sqrt{2}, 3, -57/8, 0, -1/4)$ .

El **método de eliminación** consiste en transformar el problema dado en otro equivalente y del tipo anterior. Esto se logra por medio de las siguientes **operaciones elementales**, las cuales, según se observó en la § 2, no alteran el conjunto solución:

- (I) : intercambiar ecuaciones,
- (M) : multiplicar una ecuación por un número *distinto* de 0,
- (S) : sumar a una ecuación un múltiplo de otra ecuación (del mismo problema).

Estas operaciones corresponden a operaciones similares sobre las filas de la matriz del problema y se indican en la forma siguiente:

- $[f_i \leftrightarrow f_j]$  : intercambiar las filas  $i$ -ésima y  $j$ -ésima,
- $[cf_i]$  : multiplicar la  $i$ -ésima fila por  $c$  ( $c \neq 0$ ),
- $[f_i + cf_j]$  : sumar a la fila  $i$ -ésima la fila  $j$ -ésima multiplicada por  $c$ . Obsérvese que, por ejemplo,  $[f_1 + f_2]$  es diferente de  $[f_2 + f_1]$ .

Con el fin de facilitar la discusión general pueden utilizarse las siguientes operaciones *auxiliares*:

- (C): intercambiar dos columnas de coeficientes (teniendo en cuenta que hay un *cambio* correspondiente en el *orden* de las incógnitas). Para indicar el intercambio de las columnas  $i$ -ésima y  $j$ -ésima (de coeficientes) se escribe  $[(i) \leftrightarrow (j)]$ ,

(O): suprimir filas nulas (filas de ceros), pues estas corresponden a ecuaciones triviales y superfluas. Se escribe  $]f_i[$  para indicar la supresión de  $f_i$ .

Puede suponerse que ninguna columna de coeficientes es nula pues ello indicaría que la incógnita correspondiente no figura en el problema.

Por otra parte, la columna de términos independientes puede ser nula (problemas homogéneos).

**Ejemplo A.** Para resolver el problema lineal

$$\begin{cases} 2y^1 & & & & 2y^4 & + & y^5 & = & 0 \\ 2y^1 & & + & 2y^3 & - & 4y^4 & & = & -2 \\ 2y^1 & + & y^2 & + & 3y^3 & - & 5y^4 & & = & -1 \\ 3y^1 & + & y^2 & + & 4y^3 & - & 5y^4 & + & y^5 & = & -2 \end{cases},$$

escribimos su matriz y efectuamos, sucesivamente, las siguientes operaciones:

**Primer paso** (*reducción de la 1a. columna*): poner un 1 en el lugar (1, 1) (1a. fila, 1a. columna) y ceros en el resto de la 1a. columna,

$$[f_1 \leftrightarrow f_2], [1/2 f_1], [f_3 - 2f_1], [f_4 - 3f_1].$$

**Segundo paso** (*reducción de la 2a. columna*): poner un 1 en el lugar (2, 2) (2a. fila, 2a. columna) y ceros en el resto de la 2a. columna,

$$[f_2 \leftrightarrow f_3], [f_4 - f_2].$$

**Tercer paso** (*reducción de la 3a. columna*): poner un 1 en el lugar (3, 3) y ceros en el resto de la 3a. columna,

$$[(3) \leftrightarrow (4)], [1/2 f_3], [f_1 + 2f_3], [f_2 + f_3], [f_4 - 2f_3], ]f_4[.$$

Se obtiene así un problema equivalente, a saber,



$$\begin{array}{ccccc} y^1 & y^2 & y^4 & y^3 & y^5 \\ \left[ \begin{array}{ccccc|c} 1 & 0 & 0 & 1 & 1 & -1 \\ 0 & 1 & 0 & 1 & 1/2 & 1 \\ 0 & 0 & 1 & 0 & 1/2 & 0 \end{array} \right], \end{array}$$

en el cual  $y^3$  y  $y^5$  pueden tomarse como parámetros. Llamándolas  $\lambda$  y  $\mu$ , respectivamente, se observa que todas las soluciones pueden obtenerse de la fórmula

$$\begin{bmatrix} y^1 \\ y^2 \\ y^4 \\ y^3 \\ y^5 \end{bmatrix} = \begin{bmatrix} -\lambda & -\mu & -1 \\ -\lambda & -1/2\mu & +1 \\ & -1/2\mu & \\ \lambda & & \\ & \mu & \end{bmatrix} = \lambda \begin{bmatrix} -1 \\ -1 \\ 0 \\ 1 \\ 0 \end{bmatrix} + \mu \begin{bmatrix} -1 \\ -1/2 \\ -1/2 \\ 0 \\ 1 \end{bmatrix} + \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix},$$

dándole a  $\lambda$  y  $\mu$  valores arbitrarios. El problema tiene infinitas soluciones; en realidad, puede decirse que hay un número *doblemente infinito* de soluciones, ya que hay *dos* parámetros que pueden variar *independientemente*. Obsérvese, por último, que los vectores que aparecen en la fórmula final provienen directamente de la forma reducida de la matriz.

**Ejemplo B.** Para hallar el conjunto solución de

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 5 & 3 & 3 \\ 1 & 1 & -2 & 1 \\ 1 & 2 & 1 & -1 \\ 1 & -1 & -3 & -2 \end{bmatrix} \begin{bmatrix} r \\ s \\ t \\ u \end{bmatrix} = \begin{bmatrix} 5 \\ 12 \\ -4 \\ 0 \\ -11 \end{bmatrix},$$

escribimos la matriz y seguimos la misma receta.

**Primer paso** (*reducción de la 1a. columna*):

$$[f_2 - 2f_1], [f_3 - f_1], [f_4 - f_1], [f_5 - f_1].$$

**Segundo paso** (*reducción de la 2a. columna*):

$$[f_2 \leftrightarrow f_4], [f_1 - f_2], [f_4 - 3f_2], [f_5 + 2f_2].$$

**Tercer paso** (*reducción de la 3a. columna*):

$$[-1/3 f_3], [f_1 - f_3], [f_4 - f_3], [f_5 + 4f_3].$$

**Cuarto paso** (*reducción de la 4a. columna*):

$$[1/7 f_4], [f_1 - 3f_4], [f_2 + 2f_4], [f_5 + 7f_4], ]f_5[.$$

La forma final es

$$\left[ \begin{array}{cccc|c} r & s & t & u & \\ \hline 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 & 3 \\ 0 & 0 & 0 & 1 & 2 \end{array} \right], \text{ es decir } \begin{bmatrix} r \\ s \\ t \\ u \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \\ 3 \\ 2 \end{bmatrix}.$$

El problema tiene una sola solución, i.e. su conjunto solución tiene un solo elemento, a saber, el vector de componentes 1, -1, 3 y 2 (en este orden).

**Ejemplo C.**

$$\begin{cases} 2t_1 + 2t_2 - 4t_3 + 2t_4 = 8 \\ 2t_1 + 2t_2 + t_3 + t_4 = 6 \\ -2t_1 - 2t_2 + 4t_3 - 2t_4 = 3. \end{cases}$$

Al cabo del primer paso,  $[1/2 f_1], [f_2 - 2f_1], [f_3 + 2f_1]$ , se observa ya que el problema es inconsistente. Sin embargo, en aras de la uniformidad, puede continuarse el proceso.

El segundo paso,  $[(2) \leftrightarrow (3)], [1/5, f_2], [f_1 + 2f_2]$ , conduce a la matriz

$$\left[ \begin{array}{cccc|c} t_1 & t_3 & t_2 & t_4 & \\ \hline 1 & 0 & 1 & 3/5 & 16/5 \\ 0 & 1 & 0 & -1/5 & -2/5 \\ 0 & 0 & 0 & 0 & 11 \end{array} \right].$$

Como no se pueden reducir más columnas de coeficientes se procede con

la de términos independientes, obteniéndose finalmente

$$\begin{array}{cccc|c} t_1 & t_3 & t_2 & t_4 & \\ \hline \left[ \begin{array}{cccc|c} 1 & 0 & 1 & 3/5 & 0 \\ 0 & 1 & 0 & -1/5 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{array} \right] . \end{array}$$

La última ecuación es  $0 = 1$  y, por lo tanto, el problema carece de soluciones.

## § 5 Descripción general del método de eliminación

Consideramos un problema con  $m$  ecuaciones y  $n$  incógnitas. El símbolo  $\times$  se usará para señalar lugares ocupados por números, en general *distintos* entre sí.

**Primer paso** (*reducción de la 1a. columna*):

$$\begin{array}{cccc|c} x^1 & \cdots & x^n & & \\ \hline \left[ \begin{array}{cccc|c} a_1^1 & \cdots & a_n^1 & \vdots & b^1 \\ \cdot & \cdots & \cdot & \vdots & \cdot \\ a_1^m & \cdots & a_n^m & \vdots & b^m \end{array} \right] . \end{array} \quad (\text{P})$$

Una operación ( $I$ ) permite llevar un coeficiente  $c_1 \neq 0$  al lugar  $(1, 1)$ . Multiplicando la 1a. fila por  $1/c_1$  se obtiene un 1 en el lugar  $(1, 1)$ . Usando la primera fila y operaciones ( $S$ ) pueden anularse los restantes elementos de la 1a. columna.

**Segundo paso** (*reducción de la 2a. columna*):

$$\left[ \begin{array}{cccc|c} \boxed{1} & \times & \times & \cdots & \times & \times \\ 0 & \times & \times & \cdots & \times & \times \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ 0 & \times & \times & \cdots & \times & \times \end{array} \right]$$

Si en la matriz que queda, al omitir la 1a. fila y la 1a. columna (y también la última columna), hay un coeficiente  $c_2 \neq 0$ , éste puede llevarse al lugar (2,2) mediante una operación (I) y una operación (C). Multiplicando la 2a. fila por  $1/c_2$  se obtiene un 1 en el lugar (2,2). Usando la 2a. fila y operaciones (S) pueden anularse los restantes elementos de la 2a. columna.

**Tercer paso** (*reducción de la 3a. columna*):

$$\left[ \begin{array}{cc|ccc|c} \boxed{1} & 0 & \times & \cdots & \times & \times \\ 0 & 1 & \times & \cdots & \times & \times \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ 0 & 0 & \times & \cdots & \times & \times \end{array} \right].$$

Si en la matriz que queda, al omitir la 1a. y 2a. filas y la 1a. y 2a. columnas (y también la última columna), hay un coeficiente  $c_3 \neq 0$ , éste puede llevarse al lugar (3,3) mediante ... etcétera.

La operación (O) se aplica cada vez que haya lugar. Como paso final, y si es el caso, se reduce la columna de términos independientes.

Debido a que en cada paso se desciende un escalón, el proceso termina después de un cierto número  $r$  de pasos, siendo  $r \leq m$ ; a este número  $r$  lo llamamos la **característica** del problema. La forma final, **escalonada** (o **reducida**), de la matriz, debe ser una de las siguientes:

A)

$$r \left\{ \left[ \begin{array}{cccccc|c} 1 & \cdots & 0 & c_1^1 & \cdots & c_p^1 & \vdots & d^1 \\ \cdot & \cdots & \cdot & \cdot & \cdots & \cdot & \cdot & \cdot \\ 0 & \cdots & 1 & c_1^r & \cdots & c_p^r & \cdot & d^r \end{array} \right] \right.$$

$\underbrace{\hspace{15em}}_n$

En este caso  $r < n$  y  $(p = n - r > 1)$ .

B)

$$r \left\{ \left[ \begin{array}{ccc|c} 1 & \cdots & 0 & \vdots & d^1 \\ \cdot & \cdots & \cdot & \cdot & \cdot \\ 0 & \cdots & 1 & \cdot & d^r \end{array} \right] \right.$$

$\underbrace{\hspace{10em}}_n$

En este caso  $r = n$ .

C)

$$r \left\{ \left[ \begin{array}{cccccc|c} 1 & \cdots & 0 & \times & \cdots & \times & \vdots & 0 \\ \cdot & \cdots & \cdot & \cdot & \cdots & \cdot & \cdot & \cdot \\ 0 & \cdots & 1 & \times & \cdots & \times & \cdot & 0 \\ 0 & \cdots & 0 & 0 & \cdots & 0 & \cdot & 1 \end{array} \right] \right.$$

$\underbrace{\hspace{15em}}_n$

En el segundo caso el problema tiene una sola solución, a saber, el vector de componentes  $d^1, \dots, d^n$  (ver ejemplo B). En el tercero el problema es inconsistente (ver ejemplo C).

Consideremos más detenidamente el primer caso (ver ejemplo A):

Sea  $t^1, \dots, t^r, t^{r+1}, \dots, t^{r+p}$  ( $r + p = n$ ) el ordenamiento de las incógnitas en la forma final, la cual corresponde al problema equivalente

$$\begin{cases} t^1 & + c_1^1 t^{r+1} + c_2^1 t^{r+2} + \cdots + c_p^1 t^{r+p} = d^1 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ t^r & + c_1^r t^{r+1} + c_2^r t^{r+2} + \cdots + c_p^r t^{r+p} = d^r, \end{cases}$$

en donde se pueden despejar inmediatamente  $t^1, \dots, t^r$  en términos de las incógnitas restantes. Simplificando la notación mediante las definiciones

$$\lambda_1 = t^{r+1}, \lambda_2 = t^{r+2}, \dots, \lambda_p = t^{r+p} \quad (= t^n),$$

todas las soluciones del problema pueden obtenerse de la fórmula

$$\begin{bmatrix} t^1 \\ \vdots \\ t^r \\ t^{r+1} \\ t^{r+2} \\ \vdots \\ t^n \end{bmatrix} = \begin{bmatrix} -c_1^1 \lambda_1 & -c_2^1 \lambda_2 & \cdots & -c_p^1 \lambda_p & +d^1 \\ & & & & \\ & & & & \\ -c_1^r \lambda_1 & -c_2^r \lambda_2 & \cdots & -c_p^r \lambda_p & +d^r \\ & \lambda_1 & & & \\ & & \lambda_2 & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \lambda_p \end{bmatrix},$$

la cual puede también escribirse en la forma

$$\begin{bmatrix} t^1 \\ \vdots \\ t^r \\ t^{r+1} \\ t^{r+2} \\ \vdots \\ t^n \end{bmatrix} = \lambda_1 \begin{bmatrix} -c_1^1 \\ \vdots \\ -c_1^r \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \lambda_2 \begin{bmatrix} -c_2^1 \\ \vdots \\ -c_2^r \\ 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} + \cdots + \lambda_p \begin{bmatrix} -c_p^1 \\ \vdots \\ -c_p^r \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} + \begin{bmatrix} d^1 \\ \vdots \\ d^r \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (1)$$

dándole a  $\lambda_1, \lambda_2, \dots, \lambda_p$  valores arbitrarios. El problema tiene infinitas soluciones; puede decirse que tiene un número  $p$  veces *infinito* de soluciones pues hay  $p$  parámetros independientes.

La fórmula (1) se llama la **solución general** del problema. Obsérvese que los vectores que allí aparecen pueden escribirse inmediatamente a partir de la forma final de la matriz.

Devolviendo las incógnitas a su orden original, y llamando  $\mathbf{x}$  al vector que las tiene por componentes, la solución general toma la forma

$$\mathbf{x} = \lambda_1 \mathbf{h}_1 + \cdots + \lambda_p \mathbf{h}_p + \mathbf{c}. \quad (2)$$

Haciendo  $\lambda_1 = \dots = \lambda_p = 0$  se observa que  $\mathbf{c}$  es una solución (particular) del problema (P). Un instante de reflexión muestra que la expresión

$$\lambda_1 \mathbf{h}_1 + \dots + \lambda_p \mathbf{h}_p, \quad (3)$$

es la solución general del problema homogéneo asociado (H).

## § 6 Conclusiones

Si un problema es *consistente*, sólo caben *dos* posibilidades:

- a) si  $r < n$  el problema tiene *infinitas* soluciones,
- b) si  $r = n$  el problema tiene *una sola* solución.

Como  $r \leq m$ , siempre que  $m < n$  debe tenerse que  $r < n$ . Se concluye que si, en un sistema consistente, el número  $m$  de ecuaciones es menor que el número  $n$  de incógnitas entonces el problema tiene infinitas soluciones. Esto se aplica, en particular, a los problemas homogéneos, los cuales son siempre consistentes:

**Teorema.** *Todo problema homogéneo con más incógnitas que ecuaciones tiene soluciones diferentes de la trivial.*

En otros términos: cuando hay más grados de libertad (incógnitas) que restricciones (ecuaciones) el problema tiene soluciones no triviales.

## § 7 Comentarios

(1) Las operaciones (C), responsables del posible desordenamiento de las incógnitas en la forma final, no son esenciales y se han introducido aquí sólo para simplificar la descripción general del proceso de eliminación. Una vez comprendido éste puede prescindirse de aquellas.

(2) Cuando se trata de un problema homogéneo  $A\mathbf{x} = \mathbf{0}$ , no es necesario escribir la columna de términos independientes y el método se limita a obtener la forma reducida de la matriz  $A$ .

(3) La forma final (escalonada, reducida) de una matriz es única. Demostrarlo no es difícil aunque la descripción del argumento y la notación pueden ser un poco engorrosas. Por otra parte, los pasos del algoritmo de eliminación pueden especificarse de tal modo que el proceso mismo sea único. Por ejemplo: en el primer paso (ver § 5) se mira la primera columna y se busca, de arriba hacia abajo, su primer elemento no nulo; éste será  $c_1$ . . . . En el segundo paso, y después de las omisiones indicadas, se busca, de izquierda a derecha, la primera columna no nula, y en ella se busca, de arriba hacia abajo, el primer elemento no nulo; éste será  $c_2$ . . . etcétera.



## Capítulo 2

# Dimensión

Los *vectores* son objetos matemáticos que pueden ser sumados entre sí y multiplicados por *escalares* (números) y que se agrupan en conjuntos llamados *espacios* vectoriales. En cada espacio se trata de seleccionar un puñado *básico* de vectores, los menos posibles que, *combinados* adecuadamente, *generen* cualquier otro vector. El número de vectores básicos puede entenderse como el número de *grados de libertad* que hay en el espacio.

La recta y el plano (y también la recta y el espacio) pueden ponerse en correspondencia biunívoca, pudiendo entonces decirse que los tres tienen *el mismo número* de elementos. Sin embargo, “espacialmente” son muy diferentes: en la recta hay un solo grado de libertad, en el plano dos y en el espacio tres. Esto nos indica que, a pesar de tener el mismo número de elementos o puntos, dos conjuntos pueden diferir notablemente en su “forma de extenderse”.

### § 8 Espacios vectoriales

El conjunto solución, o núcleo, de un problema homogéneo  $A\mathbf{x} = \mathbf{0}$ , con  $n$  incógnitas, es un subconjunto  $\mathcal{U}$  de  $\mathbb{R}^n$  que goza de las propiedades siguientes:

- (A.1) Si  $\mathbf{u}$  y  $\mathbf{v}$  son elementos de  $\mathcal{U}$  entonces  $\mathbf{u} + \mathbf{v}$  también está en  $\mathcal{U}$ .
- (A.2)  $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$  y  $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$  para todo  $\mathbf{u}, \mathbf{v}, \mathbf{w}$  de  $\mathcal{U}$ .
- (A.3) Hay un elemento neutro  $\mathbf{0}$  (es el *vector cero*) de  $\mathcal{U}$  tal que  $\mathbf{u} + \mathbf{0} = \mathbf{u}$ , para todo  $\mathbf{u}$  de  $\mathcal{U}$ .
- (A.4) Para cada  $\mathbf{u}$  de  $\mathcal{U}$  hay un elemento,  $-\mathbf{u}$ , de  $\mathcal{U}$  tal que  $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$
- (M.1) Si  $\mathbf{u}$  es un elemento de  $\mathcal{U}$  entonces  $\lambda\mathbf{u}$  también está en  $\mathcal{U}$ , cualquiera sea el número  $\lambda$ .

Además, si  $\lambda$  y  $\mu$  son números cualesquiera y  $\mathbf{u}$  y  $\mathbf{v}$  son elementos cualesquiera de  $\mathcal{U}$ , entonces:

- (M.2)  $\lambda(\mu\mathbf{u}) = (\lambda\mu)\mathbf{u}$  ,  $1 \cdot \mathbf{u} = \mathbf{u}$  ,  $0 \cdot \mathbf{u} = \mathbf{0}$ .
- (AM)  $\lambda(\mathbf{u} + \mathbf{v}) = \lambda\mathbf{u} + \lambda\mathbf{v}$  y  $(\lambda + \mu)\mathbf{u} = \lambda\mathbf{u} + \mu\mathbf{u}$ .

De § 3 (L) se deducen (A.1), (M.1) y parte de (A.3) y (A.4). Las demás son propiedades generales de las operaciones en  $\mathbb{R}^n$ .

Un conjunto  $\mathcal{U}$  en el cual estén definidas una *adición* y una *multiplicación* por *escalares* (números) que satisfagan las propiedades enumeradas, se llama un **espacio vectorial** y sus elementos se llaman **vectores**. Dentro de lo posible se escribirá  $\mathbf{0}$ ,  $\mathbf{a}$ ,  $\mathbf{u}$ ,  $\mathbf{v}$ , etc para denotar elementos de un espacio vectorial.

En lugar de  $\mathbf{u} + (-\mathbf{v})$  se escribe  $\mathbf{u} - \mathbf{v}$ . En lugar de  $\lambda\mathbf{u}$  se puede escribir  $\mathbf{u}\lambda$ . Si  $\lambda \neq 0$ , en lugar de  $(1/\lambda)\mathbf{u}$  se puede escribir  $\mathbf{u}/\lambda$ .

Las propiedades básicas de los espacios vectoriales implican otras relaciones familiares, como:  $(-1)\mathbf{u} = -\mathbf{u}$ ,  $-(\mathbf{u} + \mathbf{v}) = -\mathbf{u} - \mathbf{v}$ ,  $(\lambda - \mu)\mathbf{u} = \lambda\mathbf{u} - \mu\mathbf{u}$ , etc.

**Ejemplos.** (1) Las matrices  $m \times n$  forman un espacio vectorial, en particular  $\mathbb{R}^n$  y  $\mathbb{R}_n$  son espacios vectoriales.

(2) El conjunto  $\mathcal{F}(S, \mathcal{U})$  de todas las funciones de un conjunto dado  $S$  en un espacio vectorial dado  $\mathcal{U}$  (con las operaciones definidas puntualmente) es un espacio vectorial.

**Nota.** En lo que sigue  $\mathcal{U}$ ,  $\mathcal{V}$ ,  $\mathcal{W}$  designarán espacios vectoriales.

## § 9 Subespacios

Si  $\mathcal{M}$  es un subconjunto no vacío de  $\mathcal{U}$  tal que  $\mathbf{u} + \mathbf{v}$  y  $\lambda\mathbf{u}$  están en  $\mathcal{M}$  siempre que  $\mathbf{u}$  y  $\mathbf{v}$  están en  $\mathcal{M}$  (y  $\lambda$  es un escalar), entonces  $\mathcal{M}$  es a su vez un espacio vectorial, con las operaciones heredadas de  $\mathcal{U}$ . En este caso se dice que  $\mathcal{M}$  es un **subespacio** de  $\mathcal{U}$ . Tanto  $\mathcal{U}$  como  $\{\mathbf{0}\}$  son siempre subespacios de  $\mathcal{U}$ ;  $\{\mathbf{0}\}$  es el subespacio **trivial**.

Los conjuntos de la forma  $\mathcal{M} + \mathbf{c} = \{\mathbf{u} + \mathbf{c} : \mathbf{u} \in \mathcal{M}\}$ , obtenidos por *traslación* de un subespacio  $\mathcal{M}$  a lo largo de un vector  $\mathbf{c}$  se llaman **variedades** (lineales).

**Ejemplos.** (1) El conjunto solución de un problema lineal es una variedad de  $\mathbb{R}^n$  ( $n =$  número de incógnitas). Si el problema es homogéneo, su conjunto solución (núcleo) es un subespacio. Puede demostrarse que, recíprocamente, todo subespacio de  $\mathbb{R}^n$  es un núcleo.

(2) El conjunto  $\mathcal{P}_n$  de todas las funciones polinómicas de grado menor que  $n$  es un subespacio de  $\mathcal{F}(\mathbb{R}, \mathbb{R})$ .

(3) Si  $\mathcal{N}_1, \dots, \mathcal{N}_s$  son subespacios de  $\mathcal{U}$ , entonces:

a) el conjunto  $\mathcal{N}_1 + \dots + \mathcal{N}_s$  de todas las sumas  $\mathbf{u}_1 + \dots + \mathbf{u}_s$ , con  $\mathbf{u}_i$  en  $\mathcal{N}_i$  es el menor de los subespacios de  $\mathcal{U}$  que contienen a todos los  $\mathcal{N}_i$ .

b) el conjunto  $\mathcal{N}_1 \cap \dots \cap \mathcal{N}_s$  formado por los vectores que están, a la vez, en *todos* los  $\mathcal{N}_i$  es el mayor de los subespacios contenidos en todos los  $\mathcal{N}_i$ . En particular, si  $\mathcal{M}$  y  $\mathcal{N}$  son subespacios de  $\mathcal{U}$ , entonces  $\mathcal{M} + \mathcal{N}$  y  $\mathcal{M} \cap \mathcal{N}$  también lo son.

## § 10 Transformaciones lineales e isomorfismos

Una función  $T : \mathcal{U} \rightarrow \mathcal{V}$  es una **transformación lineal** (t.l.) si, para todo  $\mathbf{u}$ ,  $\mathbf{u}_1$ ,  $\mathbf{u}_2$  de  $\mathcal{U}$  y todo escalar  $\lambda$ ,

$$T(\mathbf{u}_1 + \mathbf{u}_2) = T(\mathbf{u}_1) + T(\mathbf{u}_2) \quad \text{y} \quad T(\lambda\mathbf{u}) = \lambda T(\mathbf{u}) .$$

En otras palabras,  $T$  preserva las operaciones (sumas y productos por escalar). De estas propiedades básicas se deducen otras como:  $T(\mathbf{0}) = \mathbf{0}$ ,  $T(\mathbf{u}_1 - \mathbf{u}_2) = T(\mathbf{u}_1) - T(\mathbf{u}_2)$ , etc.

Si  $T$  es además biyectiva se dice que es un **isomorfismo**. En este caso  $T^{-1} : \mathcal{V} \longrightarrow \mathcal{U}$  es también un isomorfismo. Toda relación en  $\mathcal{U}$  puede trasladarse a  $\mathcal{V}$  y recíprocamente. Estos espacios son, esencialmente, la misma cosa. Se dice que son **isomorfos**.

**Ejemplos.** (1) La identidad  $1$  de  $\mathcal{U}$  es una t.l. de  $\mathcal{U}$  en  $\mathcal{U}$ .

(2) Toda t.l.  $T$  de  $\mathbb{R}$  en  $\mathbb{R}$  es de la forma  $T(x) = ax$ . A menos que  $a = 0$ , ésta es un isomorfismo.

(3) Si  $A$  es una matriz  $m \times n$  y definimos  $T(\mathbf{u}) = A\mathbf{u}$ , para cada  $\mathbf{u}$  de  $\mathbb{R}^n$ , entonces  $T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$  es una transformación lineal.

(4)  $\mathbb{R}^m$  y  $\mathbb{R}_m$  son **isomorfos**. Un intercambio de coordenadas en  $\mathbb{R}^m$  (resp. en  $\mathbb{R}_m$ ) es un isomorfismo de este espacio sobre sí mismo.

(5) Si  $k \leq n$  y, para cada  $\mathbf{x} \in \mathbb{R}^k$ , definimos  $\mathbf{x}_{(n)} = [x^1, \dots, x^k, 0, \dots, 0]^t$  ( $\in \mathbb{R}^n$ ), entonces el conjunto  $\mathbb{R}_{(n)}^k$  formado por todos estos vectores es un subespacio de  $\mathbb{R}^n$ , isomorfo a  $\mathbb{R}^k$ : es una *copia* de  $\mathbb{R}^k$  inmersa en  $\mathbb{R}^n$ .

Hay una caracterización sencilla de la inyectividad para transformaciones lineales:

[1] Una condición necesaria y suficiente para que una transformación lineal  $T : \mathcal{U} \longrightarrow \mathcal{V}$  sea inyectiva es que, para todo  $\mathbf{u}$  de  $\mathcal{U}$ ,

$$T(\mathbf{u}) = \mathbf{0} \implies \mathbf{u} = \mathbf{0} .$$

DEMOSTRACIÓN. (i) Si  $T$  es inyectiva y  $T(\mathbf{u}) = \mathbf{0}$ , entonces  $T(\mathbf{u}) = T(\mathbf{0})$ . Luego  $\mathbf{u} = \mathbf{0}$ .

(ii) Si  $T$  satisface la condición y  $T(\mathbf{u}_1) = T(\mathbf{u}_2)$  entonces  $T(\mathbf{u}_1 - \mathbf{u}_2) = T(\mathbf{u}_1) - T(\mathbf{u}_2) = \mathbf{0}$ . Luego  $\mathbf{u}_1 - \mathbf{u}_2 = \mathbf{0}$ , es decir,  $\mathbf{u}_1 = \mathbf{u}_2$ . ♠

Nociones fundamentales, como las de espacio vectorial y transformación lineal, esquematizan los rasgos esenciales de una teoría. Al revelar los

elementos que actúan en la obtención de los resultados centrales, permiten generalizar los enunciados y simplificar los argumentos. El precio ineludible es una mayor abstracción. Algo de esto puede percibirse en la § 54.

## § 11 Sistemas de vectores. Combinaciones lineales

Un **sistema** de  $n$  vectores de un espacio  $\mathcal{U}$  es una  $n$ -pla ordenada  $\mathcal{A}$ , de vectores de  $\mathcal{U}$ :

$$\mathcal{A} = [\mathbf{u}_1, \dots, \mathbf{u}_n] .$$

Cuando no hay riesgo de confusión se escribe  $\mathbf{u}_1, \dots, \mathbf{u}_n$  en lugar de  $[\mathbf{u}_1, \dots, \mathbf{u}_n]$ .

De todo vector de la forma

$$x^1 \mathbf{u}_1 + \dots + x^n \mathbf{u}_n ,$$

donde  $x^1, \dots, x^n$  son escalares, diremos que es una **combinación lineal** (c.l.) de  $\mathcal{A}$  (o de los vectores  $\mathbf{u}_1, \dots, \mathbf{u}_n$ ), con coeficientes  $x^1, \dots, x^n$ . En particular  $\mathbf{u} + \mathbf{v}$  es una c.l. de  $\mathbf{u}$  y  $\mathbf{v}$  y  $\lambda \mathbf{u}$  es una c.l. de  $\mathbf{u}$ .

Varias situaciones pueden describirse en términos de combinaciones lineales. Por ejemplo:

(1) Un subconjunto  $\mathcal{M}$  de  $\mathcal{U}$  es un subespacio si y sólo si toda c.l. de elementos de  $\mathcal{M}$  está de nuevo en  $\mathcal{M}$ .

(2) Una función  $T : \mathcal{U} \rightarrow \mathcal{V}$  es una t.l. si y sólo si la imagen de toda c.l. de elementos de  $\mathcal{U}$  es la c.l. de sus imágenes, con los mismos coeficientes. Esto es,

$$T(x^1 \mathbf{u}_1 + \dots + x^n \mathbf{u}_n) = x^1 T(\mathbf{u}_1) + \dots + x^n T(\mathbf{u}_n)$$

(es decir, la función  $T$  *preserva* combinaciones lineales).

Un sistema  $[\mathbf{a}_1, \dots, \mathbf{a}_n]$  en  $\mathbb{R}^m$  puede visualizarse como una matriz  $m \times n$ . A su vez, una tal matriz puede considerarse como un sistema de  $n$  vectores de  $\mathbb{R}^m$  o como un sistema de  $m$  vectores de  $\mathbb{R}_n$  (ver Preliminares).

**Operaciones elementales** (sobre un sistema  $\mathcal{A}$  de vectores). Son de tres clases:

- (I) intercambiar dos vectores de  $\mathcal{A}$ ,
- (M) multiplicar un vector (de  $\mathcal{A}$ ) por un escalar distinto de 0,
- (S) sumar a un vector un múltiplo de otro vector (del sistema).

Para indicar que  $\mathcal{B}$  es el sistema obtenido a partir de  $\mathcal{A}$ , por medio de una operación elemental, escribimos

$$\mathcal{A} \rightsquigarrow \mathcal{B} .$$

Es claro que, entonces, cada elemento de  $\mathcal{B}$  es una c.l. de  $\mathcal{A}$ .

Considerando separadamente los tres casos se ve que si  $\mathcal{A} \rightsquigarrow \mathcal{B}$  entonces también  $\mathcal{B} \rightsquigarrow \mathcal{A}$ .

## § 12 La transformación asociada a un sistema

Con un sistema  $\mathcal{A} = [\mathbf{u}_1, \dots, \mathbf{u}_n]$  en un espacio  $\mathcal{U}$  quizás lo único que uno puede hacer es formar combinaciones lineales. De hecho, estamos en presencia de una función, a saber, la función  $\tilde{\mathcal{A}}$  que a cada elemento  $\mathbf{x} = [x^1, \dots, x^n]^t$  de  $\mathbb{R}^n$  le hace corresponder el vector  $x^1\mathbf{u}_1 + \dots + x^n\mathbf{u}_n$  de  $\mathcal{U}$ .

Esta función

$$\tilde{\mathcal{A}} : \mathbb{R}^n \longrightarrow \mathcal{U} ,$$

es claramente una t.l. y se llama la **transformación asociada** al sistema  $\mathcal{A}$  (o la transformación lineal de  $\mathcal{A}$ ). En realidad, puede considerarse que  $\tilde{\mathcal{A}}$  es lo mismo que  $\mathcal{A}$ , sólo que visto en su aspecto dinámico. Cuando no sea útil destacar la diferencia se puede escribir  $\mathcal{A}$  en lugar de  $\tilde{\mathcal{A}}$ .

Un **problema lineal** en  $\mathcal{U}$  es una ecuación de la forma  $\mathcal{A}\mathbf{x} = \mathbf{b}$ , esto es,

$$x^1\mathbf{u}_1 + \dots + x^n\mathbf{u}_n = \mathbf{b} ,$$

donde  $\mathcal{A} = [\mathbf{u}_1, \dots, \mathbf{u}_n]$  es un sistema y  $\mathbf{b}$  un elemento de  $\mathcal{U}$ . Su **conjunto solución**,  $\{\mathbf{x} \in \mathbb{R}^n : \mathcal{A}\mathbf{x} = \mathbf{b}\}$ , es una variedad de  $\mathbb{R}^n$ . Si  $\mathbf{b} = \mathbf{0}$  el problema es **homogéneo**.

Dos problemas,  $\mathcal{A}\mathbf{x} = \mathbf{b}$  y  $\mathcal{B}\mathbf{x} = \mathbf{d}$ , donde  $\mathcal{A}$  (resp.  $\mathcal{B}$ ), es un sistema de  $n$  vectores de  $\mathcal{U}$  (resp.  $\mathcal{V}$ ), son **equivalentes** cuando dichos problemas tienen el mismo conjunto solución. En los espacios euclidianos estas nociones coinciden con las introducidas en las §§ 1 y 2, y la notación  $\mathcal{A}\mathbf{x}$  corresponde a la de la § 3.

## § 13 Generación

El conjunto  $\langle \mathcal{A} \rangle$  de todas las combinaciones lineales de un sistema dado  $\mathcal{A}$  de vectores de  $\mathcal{U}$  es un subespacio de  $\mathcal{U}$  y se llama el **subespacio generado** por  $\mathcal{A}$ .

Toda c.l. de elementos de  $\langle \mathcal{A} \rangle$  está de nuevo en  $\langle \mathcal{A} \rangle$ . En otros términos: toda combinación lineal de combinaciones lineales de un sistema  $\mathcal{A}$  es de nuevo una c.l. de  $\mathcal{A}$ . Esta observación permite afirmar que si  $\mathcal{A} \rightsquigarrow \mathcal{B}$  entonces  $\langle \mathcal{B} \rangle \subset \langle \mathcal{A} \rangle$  y, finalmente, que

[2] Si  $\mathcal{A} \rightsquigarrow \mathcal{B}$  entonces  $\langle \mathcal{A} \rangle = \langle \mathcal{B} \rangle$ .

Si  $\mathcal{A}$  es un sistema de vectores de un subespacio  $\mathcal{M}$  de  $\mathcal{U}$  entonces  $\langle \mathcal{A} \rangle \subset \mathcal{M}$ . Cuando  $\langle \mathcal{A} \rangle = \mathcal{M}$  se dice que  $\mathcal{A}$  **genera** a  $\mathcal{M}$ . Si  $\mathcal{A}$  genera al espacio total  $\mathcal{U}$  se dice que es **generador**. Es claro que  $\langle \mathcal{A} \rangle$  no es otra cosa que el recorrido de la función  $\tilde{\mathcal{A}}$ . En consecuencia:

[3] El sistema  $\mathcal{A}$  es generador si y sólo si la transformación  $\tilde{\mathcal{A}}$  es sobreyectiva.

**Notas.** (1) En lugar de decir que *el sistema*  $[\mathbf{u}_1, \dots, \mathbf{u}_n]$  genera a un subespacio  $\mathcal{M}$  se acostumbra decir que *los vectores*  $\mathbf{u}_1, \dots, \mathbf{u}_n$  generan a  $\mathcal{M}$ . En consonancia con esto, y con lo dicho en la § 11, en lugar de  $\langle [\mathbf{u}_1, \dots, \mathbf{u}_n] \rangle$  se escribe  $\langle \mathbf{u}_1, \dots, \mathbf{u}_n \rangle$ .

(2) Si  $T : \mathcal{U} \rightarrow \mathcal{V}$  es una t.l. y  $\mathcal{A} = [\mathbf{u}_1, \dots, \mathbf{u}_n]$  es un sistema de

$n$  vectores de  $\mathcal{U}$  entonces, por abuso de notación, escribiremos  $T\langle\mathcal{A}\rangle = [T\mathbf{u}_1, \dots, T\mathbf{u}_n]$ ; éste es un sistema de  $n$  vectores de  $\mathcal{V}$ . Si  $\mathcal{A}$  genera a un subespacio  $\mathcal{M}$  de  $\mathcal{U}$  entonces  $T\langle\mathcal{A}\rangle$  genera al subespacio  $T\langle\mathcal{M}\rangle$  de  $\mathcal{V}$ .

## § 14 Libertad

Una **relación de dependencia** en un sistema  $\mathcal{A} = [\mathbf{u}_1, \dots, \mathbf{u}_n]$  es una relación de la forma

$$x^1\mathbf{u}_1 + \dots + x^n\mathbf{u}_n = \mathbf{0} , \quad (1)$$

en la cual hay, por lo menos, un coeficiente diferente de cero.

Si en  $\mathcal{A}$  existe una relación de dependencia se dice que es **linealmente dependiente** (l.d.). Esto equivale a decir que alguno de los vectores de  $\mathcal{A}$  es combinación lineal de los restantes. En efecto:

(i) Si en (1) tenemos que  $x^i \neq 0$  entonces el vector correspondiente  $\mathbf{u}_i$  puede expresarse como c.l. de los restantes. (Obsérvese que, en este caso,  $\mathbf{u}_i$  puede excluirse del sistema sin alterar el subespacio generado).

(ii) La recíproca es inmediata.

Es claro que si  $\mathbf{0}$  está en  $\mathcal{A}$  entonces  $\mathcal{A}$  es l.d. y  $\mathbf{0}$  puede ser excluido sin alterar el subespacio generado.

Cuando  $\mathcal{A}$  *no* es linealmente dependiente se dice que es **linealmente independiente** (l.i.) o **libre**. Esto es:  $\mathcal{A}$  es l.i. si y sólo si ninguno de los  $\mathbf{u}_i$  es c.l. de los restantes.

Explícitamente:  $\mathcal{A}$  es l.i. si y sólo si, para todo  $x^1, \dots, x^n$ ,

$$x^1\mathbf{u}_1 + \dots + x^n\mathbf{u}_n = \mathbf{0} \implies x^1 = \dots = x^n = 0 . \quad (2)$$

En otros términos: una condición necesaria y suficiente para que  $\mathcal{A}$  sea l.i. es que la única c.l. de  $\mathcal{A}$  igual al vector  $\mathbf{0}$  sea la que tiene todos los coeficientes nulos (la llamada combinación lineal **trivial**).

Nótese que si  $\mathbf{u} \neq \mathbf{0}$  entonces  $[\mathbf{u}]$  es linealmente independiente.



Geoméricamente, dos vectores son l.d. cuando son colineales, tres vectores son l.d. cuando son coplanarios.

Si llamamos  $\mathbf{x}$  al elemento de  $\mathbb{R}^n$  de componentes  $x^1, \dots, x^n$ , la condición (2), en términos de  $\tilde{\mathcal{A}}$ , se formula así: para todo  $\mathbf{x}$  de  $\mathbb{R}^n$

$$\tilde{\mathcal{A}}(\mathbf{x}) = \mathbf{0} \implies \mathbf{x} = \mathbf{0}.$$

Por lo tanto, según [1] (§ 10):

[4] *El sistema  $\mathcal{A}$  es libre si y sólo si la transformación  $\tilde{\mathcal{A}}$  es inyectiva.*

Como toda t.l.  $T : \mathcal{U} \rightarrow \mathcal{V}$  preserva combinaciones lineales, también preserva la dependencia, i.e. : si  $\mathcal{A} = [\mathbf{a}_1, \dots, \mathbf{a}_p] \subset \mathcal{U}$  es l.d. entonces  $T\langle \mathcal{A} \rangle = [T\mathbf{a}_1, \dots, T\mathbf{a}_p] \subset \mathcal{V}$  también lo es. O, lo que es lo mismo: si  $T\langle \mathcal{A} \rangle$  es libre entonces  $\mathcal{A}$  es libre.

En  $\mathbb{R}^m$ , un sistema  $[\mathbf{a}_1, \dots, \mathbf{a}_n]$  es libre si y sólo si el problema homogéneo

$$x^1 \mathbf{a}_1 + \dots + x^n \mathbf{a}_n = \mathbf{0} \tag{3}$$

no tiene soluciones diferentes de la trivial. Esto permite usar el método de eliminación para dilucidar asuntos de libertad (!).

Cuando  $n > m$  el sistema (3) tiene más incógnitas que ecuaciones y el teorema de la § 6 implica que

[5] *Un sistema libre en  $\mathbb{R}^m$  no puede tener más de  $m$  elementos.*

**Nota.** En lugar de decir que el sistema  $[\mathbf{u}_1, \dots, \mathbf{u}_n]$  es l.i. (resp. es l.d.), se acostumbra decir que (los vectores)  $\mathbf{u}_1, \dots, \mathbf{u}_n$  son l.i. (resp. son l.d.).

Se dice que los subespacios, no triviales,  $\mathcal{N}_1, \dots, \mathcal{N}_s$ , de  $\mathcal{U}$ , son (linealmente) **independientes** cuando *todo* sistema de vectores *no nulos*  $[\mathbf{u}_1, \dots, \mathbf{u}_s]$ , con  $\mathbf{u}_i \in \mathcal{N}_i$ , es linealmente independiente. Esto equivale a decir que una suma de la forma  $\mathbf{u}_1 + \dots + \mathbf{u}_s$ , con  $\mathbf{u}_i \in \mathcal{N}_i$ , sólo puede ser  $\mathbf{0}$  cuando *cada uno* de los  $\mathbf{u}_i$  es  $\mathbf{0}$ .

**Ejemplos.** (1) Si *todo* elemento de  $\mathcal{M} = \mathcal{N}_1 + \dots + \mathcal{N}_s$  puede escribirse de manera *única* en la forma  $\mathbf{u}_1 + \dots + \mathbf{u}_s$ , con  $\mathbf{u}_i \in \mathcal{N}_i$ , entonces los  $\mathcal{N}_i$

son independientes; y recíprocamente. En este caso se dice que  $\mathcal{M}$  es la **suma directa** de los  $\mathcal{N}_i$  y se escribe

$$\mathcal{M} = \mathcal{N}_1 \oplus \cdots \oplus \mathcal{N}_s .$$

(2) Los vectores  $\mathbf{u}_1, \dots, \mathbf{u}_s$  constituyen una base de un subespacio  $\mathcal{M}$  si y sólo si  $\mathcal{M}$  es la suma directa de los subespacios  $\mathcal{N}_i = \langle \mathbf{u}_i \rangle$ .

## § 15 Bases

Al formular las nociones de generación y libertad en términos de la transformación asociada se observa cierta dualidad entre ellas. Existe una relación más cuantitativa. En un espacio  $\mathcal{V}$  tomemos  $\mathcal{A} = [\mathbf{u}_1, \dots, \mathbf{u}_m]$  generador y  $\mathcal{B} = [\mathbf{v}_1, \dots, \mathbf{v}_p]$  libre. Como  $T = \tilde{\mathcal{A}} : \mathbb{R}^m \rightarrow \mathcal{V}$  es sobreyectiva, existen  $\mathbf{a}_i \in \mathbb{R}^m$  tales que  $T(\mathbf{a}_i) = \mathbf{v}_i$ ,  $i = 1, \dots, p$ . El sistema  $A = [\mathbf{a}_1, \dots, \mathbf{a}_p]$  de  $\mathbb{R}^m$  es libre puesto que  $T\langle A \rangle = \mathcal{B}$  es libre. Entonces  $p \leq m$ .

Acabamos de probar que

[6] *En un espacio vectorial un sistema libre no puede tener más elementos que un sistema generador.*

Cuando un sistema es libre y generador se dice que es una **base**.

Lo anterior nos permite afirmar que:

[7] *Dos bases cualesquiera (de un mismo espacio) tienen igual número de elementos.*

Se dice que  $\mathcal{U}$  tiene **dimensión** finita  $n$ , y se escribe

$$\dim \mathcal{U} = n ,$$

si  $\mathcal{U}$  tiene una base con  $n$  elementos. En lo que sigue sólo consideraremos espacios de dimensión finita; para  $\mathcal{U} = \{\mathbf{0}\}$  se conviene que  $\dim \mathcal{U} = 0$ .

**Ejemplos.** (1) La llamada **base canónica** de  $\mathbb{R}^n$ , denotada por  $I$ , está formada por los vectores columna  $\mathbf{e}_1 = [1, 0, \dots, 0]^t, \dots, \mathbf{e}_n = [0, 0, \dots, 1]^t$ . Entonces  $\dim \mathbb{R}^n = n$ ; también  $\dim \mathbb{R}_n = n$ . Obsérvese

que  $I = [\mathbf{e}_1, \dots, \mathbf{e}_n]$  es la matriz  $[\delta_j^i]$ , donde  $\delta_j^i$  es la *delta de Kronecker*. Una matriz de la forma  $[\lambda_1 \mathbf{e}_1, \dots, \lambda_n \mathbf{e}_n]$  se llama matriz **diagonal** y se denota  $\Delta(\lambda_1, \dots, \lambda_n)$ .

(2) Los vectores  $\mathbf{h}_1, \dots, \mathbf{h}_p$  que figuran en la solución general de un sistema de ecuaciones (ver § 5, (2) y (3)) constituyen una base del núcleo.

(3) Si  $\mathcal{B}_i$  es una base del subespacio no trivial  $\mathcal{N}_i$  (para  $i = 1, \dots, s$ ) entonces los  $\mathcal{N}_i$  son independientes si y sólo si la “reunión”  $\mathcal{B} = [\mathcal{B}_1, \dots, \mathcal{B}_s]$  es una base de  $\mathcal{N}_1 + \dots + \mathcal{N}_s$  (suma que, en este caso, es directa).

Considerando un sistema  $\mathcal{B} = [\mathbf{u}_1, \dots, \mathbf{u}_n]$  de vectores de un espacio  $\mathcal{U}$  y aplicando los resultados [3] (§ 13) y [4] (§ 14) se deduce que

[8]  $\mathcal{B}$  es una base si y sólo si  $\tilde{\mathcal{B}}$  es un isomorfismo.

Por lo tanto  $\mathcal{B}$  es una base si y sólo si cada  $\mathbf{u}$  de  $\mathcal{U}$  puede escribirse, de manera única, en la forma  $\mathbf{u} = x^1 \mathbf{u}_1 + \dots + x^n \mathbf{u}_n$ ; los números  $x^i$  son las **coordenadas** o **componentes** de  $\mathbf{u}$  en la base  $\mathcal{B}$ ; el elemento  $[x^1, \dots, x^n]^t$  de  $\mathbb{R}^n$  es el **vector de coordenadas** de  $\mathbf{u}$  en la base  $\mathcal{B}$  y se denota  $[\mathbf{u}]_{\mathcal{B}}$ . Escribiendo  $\mathcal{B}$  en lugar de  $\tilde{\mathcal{B}}$  (ver § 12) se tiene que

$$[\mathbf{u}]_{\mathcal{B}} = \mathcal{B}^{-1} \mathbf{u} .$$

El estudio de los espacios de dimensión finita puede, en gran parte, reducirse al estudio de los espacios euclidianos. Cada uno de ellos, siendo isomorfo a un  $\mathbb{R}^n$ , tiene la potencia del continuo. La diferencia radica en la dimensión: si  $\dim \mathcal{V} > \dim \mathcal{U}$  se necesitan más vectores para generar a  $\mathcal{V}$  que para generar a  $\mathcal{U}$ ; en este sentido el primero es “más voluminoso”.

Para que un sistema sea libre no debe tener demasiados elementos mientras que para ser generador debe tener suficientes. Se comprueba sin dificultad que todo sistema más pequeño que uno libre es libre y todo sistema más grande que un generador es generador.

Parece también natural que a un sistema libre que no genere el espacio se le puedan agregar “cuidadosamente” otros elementos sin perturbar la independencia y que en un sistema generador pero dependiente haya elementos superfluos que puedan quitarse sin alterar su carácter generador.

Para precisar lo anterior consideremos un espacio  $\mathcal{U}$  y supongamos que  $\mathcal{A} = [\mathbf{u}_1, \dots, \mathbf{u}_p]$  es libre y que  $\mathbf{v}$  es un vector de  $\mathcal{U}$  que *no* está en  $\langle \mathcal{A} \rangle$ . Entonces el sistema  $\mathcal{A}_1 = [\mathbf{u}_1, \dots, \mathbf{u}_p, \mathbf{v}]$  es aún libre. En efecto: supongamos que

$$x^1 \mathbf{u}_1 + \dots + x^p \mathbf{u}_p + x \mathbf{v} = \mathbf{0}$$

y veamos que  $x^1 = \dots = x^p = x = 0$ . En primer lugar,  $x$  no puede ser diferente de 0 pues en tal caso  $\mathbf{v}$  estaría en  $\langle \mathcal{A} \rangle$ . Entonces  $x = 0$  y por lo tanto  $x^1 \mathbf{u}_1 + \dots + x^p \mathbf{u}_p = \mathbf{0}$  resultando que también  $x^1 = \dots = x^p = 0$  (puesto que  $\mathcal{A}$  es libre).

(Si  $\mathcal{A}$  puede obtenerse a partir de  $\mathcal{B}$  suprimiendo algunos vectores, escribimos  $\mathcal{A} \subset \mathcal{B}$ .)

Ahora podemos probar la proposición esbozada arriba:

**[9]** Si  $\mathcal{A} \subset \mathcal{C}$ , siendo  $\mathcal{A}$  libre y  $\mathcal{C}$  generador, existe una base  $\mathcal{B}$  tal que  $\mathcal{A} \subset \mathcal{B} \subset \mathcal{C}$ .

DEMOSTRACIÓN. Sean  $\mathcal{A} = [\mathbf{u}_1, \dots, \mathbf{u}_p]$  y  $\mathcal{C} = [\mathbf{u}_1, \dots, \mathbf{u}_p, \mathbf{v}_1, \dots, \mathbf{v}_q]$ .

(i) Si  $\mathbf{v}_1 \in \langle \mathcal{A} \rangle$ , este vector puede descartarse sin alterar el carácter generador de  $\mathcal{C}$ .

(ii) Si  $\mathbf{v}_1 \notin \langle \mathcal{A} \rangle$  entonces el sistema  $\mathcal{A}_1 = [\mathbf{u}_1, \dots, \mathbf{u}_p, \mathbf{v}_1]$  es libre y está contenido en  $\mathcal{C}$ . Y estamos en la situación original.

Continuando así llegamos a  $\mathbf{v}_q$  obteniéndose al final el  $\mathcal{B}$  deseado. ♠

Enunciemos algunos casos particulares:

**[10]** *Todo sistema  $\mathcal{A}$ , libre pero no generador, puede completarse hasta obtener una base.*

Basta aplicar [9] a  $\mathcal{A}$  y al sistema generador  $\mathcal{C}$  obtenido agregando a continuación de los elementos de  $\mathcal{A}$  los elementos de cualquier generador.

**[11]** *Todo sistema  $\mathcal{C}$ , no reducido a  $[\mathbf{0}]$ , contiene un subconjunto libre maximal.*

Es decir: existe un subconjunto  $\mathcal{B}$  el cual es libre y tal que no existe

ningún  $\mathcal{D}$ , libre, con

$$\mathcal{B} \subsetneq \mathcal{D} \subset \mathcal{C} .$$

En efecto, basta tomar  $\mathbf{v} \in \mathcal{C}$ ,  $\mathbf{v} \neq \mathbf{0}$ , y aplicar [9] al espacio  $\langle \mathcal{C} \rangle$  con  $\mathcal{A} = [\mathbf{v}]$ . Se ve inmediatamente que un tal  $\mathcal{B}$  es, siempre, una base de  $\langle \mathcal{C} \rangle$ .

[12] *Todo sistema generador pero dependiente puede reducirse hasta obtener una base.*

**Ejemplos.** (4) Sean  $\mathcal{M}$  y  $\mathcal{N}$  dos subespacios de  $\mathcal{U}$ . Una base  $[\mathbf{u}_1, \dots, \mathbf{u}_s]$  de  $\mathcal{M} \cap \mathcal{N}$  puede extenderse hasta una base  $[\mathbf{u}_1, \dots, \mathbf{u}_s, \mathbf{v}_1, \dots, \mathbf{v}_p]$  de  $\mathcal{M}$  (resp.  $[\mathbf{u}_1, \dots, \mathbf{u}_s, \mathbf{w}_1, \dots, \mathbf{w}_q]$  de  $\mathcal{N}$ ); es fácil ver que  $[\mathbf{u}_1, \dots, \mathbf{u}_s, \mathbf{v}_1, \dots, \mathbf{v}_p, \mathbf{w}_1, \dots, \mathbf{w}_q]$  es una base de  $\mathcal{M} + \mathcal{N}$ . Entonces, la igualdad  $(s+p) + (s+q) = (s+p+q) + s$  indica que

$$\dim \mathcal{M} + \dim \mathcal{N} = \dim (\mathcal{M} + \mathcal{N}) + \dim (\mathcal{M} \cap \mathcal{N}) .$$

Si  $\mathcal{M} \cap \mathcal{N} = \{\mathbf{0}\}$ , la suma es directa y se tiene que

$$\dim (\mathcal{M} \oplus \mathcal{N}) = \dim \mathcal{M} + \dim \mathcal{N} .$$

Por lo tanto, si  $\dim \mathcal{M} + \dim \mathcal{N} > \dim (\mathcal{M} + \mathcal{N})$  (p.ej. si  $\dim \mathcal{M} + \dim \mathcal{N} > \dim \mathcal{U}$ ), entonces  $\mathcal{M} \cap \mathcal{N} \neq \{\mathbf{0}\}$ .

(5) Si  $T : \mathcal{U} \rightarrow \mathcal{V}$  es una t.l. y  $[\mathbf{u}_1, \dots, \mathbf{u}_n]$  es una base de  $\mathcal{U}$ , entonces el sistema  $[T\mathbf{u}_1, \dots, T\mathbf{u}_n]$  contiene una base de  $\mathcal{R}_T$ .

## § 16 Sistemas equivalentes

Diremos que dos sistemas

$$\mathcal{A} = [\mathbf{u}_1, \dots, \mathbf{u}_p] \subset \mathcal{U} \quad \text{y} \quad \mathcal{B} = [\mathbf{v}_1, \dots, \mathbf{v}_p] \subset \mathcal{V}$$

son **equivalentes** si los problemas homogéneos por ellos determinados lo son (ver § 12). En este caso la correspondencia  $\mathbf{u}_i \leftrightarrow \mathbf{v}_i$  preserva todas las relaciones de dependencia. En efecto, si para ciertos números  $x^1, \dots, x^p$  se cumple que  $x^1\mathbf{u}_1 + \dots + x^p\mathbf{u}_p = \mathbf{0}$  entonces también  $x^1\mathbf{v}_1 + \dots + x^p\mathbf{v}_p = \mathbf{0}$ , y recíprocamente.

En particular, si  $\mathcal{C}$  es un subconjunto libre maximal de  $\mathcal{A}$ , su correspondiente es un subconjunto libre maximal de  $\mathcal{B}$ .

Si a un sistema  $A$  de vectores de  $\mathbb{R}^m$ ,

$$A = [\mathbf{a}_1, \dots, \mathbf{a}_n],$$

se le aplica el algoritmo de eliminación, sin intercambio de columnas, se obtiene, al final, un sistema equivalente

$$B = [\mathbf{b}_1, \dots, \mathbf{b}_n],$$

de vectores de  $\mathbb{R}^r$  (donde  $r$  es la característica del sistema). En otras palabras: si sobre las filas de una matriz se efectúan operaciones elementales, incluyendo supresión de filas nulas, el sistema formado por sus columnas se transforma en otro equivalente. Lo mismo vale intercambiando los términos “fila” y “columna”.

En la forma final  $B$ , las  $r$  columnas reducidas forman un subconjunto libre maximal. Los vectores  $\mathbf{a}_i$  correspondientes forman entonces un subconjunto libre maximal de  $A$ .

Un sistema libre  $L = [\mathbf{l}_1, \dots, \mathbf{l}_p] \subset \mathbb{R}^m$  puede completarse procediendo en forma análoga con la matriz

$$[\mathbf{l}_1, \dots, \mathbf{l}_p, \mathbf{e}_1, \dots, \mathbf{e}_m].$$

**Ejemplo.** Dados los siguientes elementos de  $\mathbb{R}^5$

$$\mathbf{a}_1 = \begin{bmatrix} 1 \\ 2 \\ -2 \\ 12 \\ -4 \end{bmatrix}, \quad \mathbf{a}_2 = \begin{bmatrix} -1 \\ -3 \\ 2 \\ -15 \\ 5 \end{bmatrix}, \quad \mathbf{a}_3 = \begin{bmatrix} -1 \\ 0 \\ 2 \\ -6 \\ 2 \end{bmatrix}, \quad \mathbf{a}_4 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ -3 \\ 1 \end{bmatrix}, \quad \mathbf{a}_5 = \begin{bmatrix} -3 \\ -10 \\ 9 \\ -57 \\ 19 \end{bmatrix};$$

$$\mathbf{x} = \begin{bmatrix} 0 \\ -1 \\ 0 \\ -2 \\ 1 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} -3 \\ -8 \\ 9 \\ -51 \\ 17 \end{bmatrix},$$

sea  $A = [\mathbf{a}_1, \dots, \mathbf{a}_5]$  y sea  $I = [\mathbf{e}_1, \dots, \mathbf{e}_5]$  la base canónica. Aplicándole a la matriz  $B = [A, I, \mathbf{x}, \mathbf{y}]$  el algoritmo de eliminación, sin intercambio de

columnas, p.ej.  $[f_2 - 2f_1]$  ,  $[f_3 + 2f_1]$  ,  $[f_4 - 12f_1]$  ,  $[f_5 + 4f_1]$  ,  $[-f_2]$  ,  $[f_1 + f_2]$  ,  $[f_4 + 3f_2]$  ,  $[f_5 - f_2]$  ,  $[f_4 + 3f_3]$  ,  $[f_5 - f_3]$  ,  $[f_4 \leftrightarrow f_5]$  ,  $[f_1 + f_4]$  ,  $[f_2 + f_4]$  ,  $[f_5 + 3f_4]$ , se llega a la forma final  $C$ ,

$$B = [ \mathbf{a}_1 \quad \mathbf{a}_2 \quad \mathbf{a}_3 \quad \mathbf{a}_4 \quad \mathbf{a}_5 \quad | \quad \mathbf{e}_1 \quad \mathbf{e}_2 \quad \mathbf{e}_3 \quad \mathbf{e}_4 \quad \mathbf{e}_5 \quad | \quad \mathbf{x} \quad \mathbf{y} ]$$

$$C = \begin{bmatrix} \uparrow & \uparrow & \uparrow & \uparrow & \uparrow & | & \uparrow & \uparrow & \uparrow & \uparrow & \uparrow & | & \uparrow & \uparrow \\ 1 & 0 & -3 & 0 & 1 & | & 3 & 0 & -1 & 0 & 1 & | & 1 & -1 \\ 0 & 1 & -2 & 0 & 4 & | & 2 & 0 & -1 & 0 & 1 & | & 1 & 2 \\ 0 & 0 & 0 & 1 & 3 & | & 2 & 0 & 1 & 0 & 0 & | & 0 & 3 \\ 0 & 0 & 0 & 0 & 0 & | & 0 & 1 & -1 & 0 & 1 & | & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & | & 0 & 0 & 0 & 1 & 3 & | & 1 & 0 \end{bmatrix},$$

en la cual se lee (¿magia?) la siguiente información sobre  $A$ :

1.  $A$  no es libre, pues, p.ej.,  $\mathbf{a}_3 = -3\mathbf{a}_1 - 2\mathbf{a}_2$ . También  $\mathbf{a}_5 = \mathbf{a}_1 + 4\mathbf{a}_2 + 3\mathbf{a}_4$ .
2.  $A$  no es generador, pues, p.ej.,  $\mathbf{e}_2$  no está en  $\langle A \rangle$ .
3.  $[\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_4]$  es un subconjunto libre maximal de  $A$ , el cual puede completarse obteniéndose la base  $[\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_4, \mathbf{e}_2, \mathbf{e}_4]$ .
4.  $\mathbf{x}$  no está en  $\langle A \rangle$ .
5.  $\mathbf{y}$  sí está en  $\langle A \rangle$ ; explícitamente:  $\mathbf{y} = -\mathbf{a}_1 + 2\mathbf{a}_2 + 3\mathbf{a}_4$ .

## § 17 Rango

Se llama **rango** de un sistema  $\mathcal{A}$ , y se denota  $rg \mathcal{A}$ , la dimensión del subespacio generado por  $\mathcal{A}$ :

$$rg \mathcal{A} = \dim \langle \mathcal{A} \rangle .$$

Como cualquier subconjunto libre maximal de un sistema es una base del subespacio por él generado, vemos que el rango de un sistema  $\mathcal{A}$  de  $n$  vectores de  $\mathbb{R}^m$  es igual al número de columnas reducidas al aplicar el algoritmo de eliminación a la matriz  $A$ , y este número es, precisamente, la característica  $r$  del problema homogéneo asociado a  $A$ .

Teniendo en cuenta que las operaciones elementales no alteran el subespacio generado y que las filas (no nulas) que aparecen en la forma final son linealmente independientes, se ve que su número (que es de nuevo igual a la característica del problema), es también el rango del sistema de vectores de  $\mathbb{R}_n$  formado por las filas de  $A$ .

Al aplicar el algoritmo de eliminación a una matriz  $m \times n$  se está entonces calculando, simultáneamente, el rango del sistema formado por sus columnas y el del sistema formado por sus filas, y ambos resultan ser iguales. Este número se llama el *rango* de la matriz.

En el ejemplo de la § 16 es claro que  $rg A = 3$ .



## Capítulo 3

# Linealidad

Diversas situaciones pueden idealizarse bajo la forma de un dispositivo que, a partir de ciertas causas o estímulos, produce unos efectos o respuestas que son proporcionales a esos estímulos o, más general, que obedecen al principio de superposición de los efectos. Con más precisión: al aplicar una suma de estímulos se obtiene como efecto la suma de los efectos individuales y al multiplicar un estímulo el efecto se multiplica por el mismo factor. El modelo matemático adecuado es una transformación lineal  $T : \mathcal{U} \rightarrow \mathcal{V}$  entre dos espacios vectoriales  $\mathcal{U}$  y  $\mathcal{V}$ . El mismo modelo incluye las transformaciones geométricas como rotaciones, proyecciones y simetrías.

### § 18 Álgebra de transformaciones

El conjunto  $\mathcal{L}(\mathcal{U}, \mathcal{V})$  de todas las transformaciones lineales de un espacio vectorial  $\mathcal{U}$  en otro,  $\mathcal{V}$ , es un espacio vectorial; es a su vez un subespacio de  $\mathcal{F}(\mathcal{U}, \mathcal{V})$  (ver § 8, ej. (2)) y, por lo tanto, para  $S, T$  en  $\mathcal{L}(\mathcal{U}, \mathcal{V})$ ,  $\lambda$  escalar, y para todo  $\mathbf{u}$  de  $\mathcal{U}$ :

$$(S + T)\mathbf{u} = S\mathbf{u} + T\mathbf{u} \quad \text{y} \quad (\lambda T)\mathbf{u} = \lambda(T\mathbf{u}) .$$

El espacio  $\mathcal{L}(\mathcal{U}, \mathbb{R})$  es el **dual** de  $\mathcal{U}$  y se denota  $\mathcal{U}'$ ; sus elementos se llaman **funcionales** o **formas lineales** en  $\mathcal{U}$ .

La compuesta de dos transformaciones lineales,

$$\begin{array}{ccccc} \mathcal{U} & \xrightarrow{T} & \mathcal{V} & \xrightarrow{S} & \mathcal{W} \\ & \underbrace{\hspace{10em}}_{ST} & & & \end{array}$$

es también una transformación lineal llamada el **producto** de  $S$  y  $T$  y denotada  $ST$ . Directamente se comprueba que

$$S(T_1 + T_2) = ST_1 + ST_2 \quad , \quad (S_1 + S_2)T = S_1T + S_2T \quad , \quad (1)$$

$$\lambda(ST) = (\lambda S)T = S(\lambda T) \quad , \quad R(ST) = (RS)T \quad ,$$

siempre que las operaciones estén definidas. Esto ocurre, en particular, cuando  $\mathcal{V} = \mathcal{U}$ ; el espacio  $\mathcal{L}(\mathcal{U}, \mathcal{U})$  se llama el álgebra de  $\mathcal{U}$  y se denota  $\mathcal{L}(\mathcal{U})$ ; sus elementos se llaman *operadores* lineales (de  $\mathcal{U}$ ).

**Ejemplo.** Si  $\mathcal{U} = \mathcal{N}_1 \oplus \cdots \oplus \mathcal{N}_s$  y, para cada  $j$  ( $= 1, \dots, s$ ) y cada  $\mathbf{u} = \mathbf{u}_1 + \cdots + \mathbf{u}_s$ , con  $\mathbf{u}_i$  en  $\mathcal{N}_i$ , se define  $P_j \mathbf{u} = \mathbf{u}_j$  entonces las  $P_j$  son operadores lineales de  $\mathcal{U}$ . Estas  $P_j$  son las **proyecciones asociadas** a la descomposición de  $\mathcal{U}$  como suma directa de los  $\mathcal{N}_i$  y tienen las siguientes propiedades: a)  $P_j^2 = P_j$ , b)  $P_j P_k = 0$  si  $j \neq k$ , c)  $\sum_{j=1}^s P_j = 1$ .

**Nota.** Un **álgebra** es un espacio vectorial en el cual está definida una *multiplicación* asociativa y bien relacionada con la suma y el producto por escalar, pero no necesariamente conmutativa; es en este sentido que  $\mathcal{L}(\mathcal{U})$  es un álgebra. Dos álgebras son *isomorfas* cuando existe entre ellas una biyección que preserve todas las operaciones (1).

## § 19 Álgebra de matrices

Un sistema  $\mathcal{B} = [\mathbf{v}_1, \dots, \mathbf{v}_n]$  de  $n$  vectores de un espacio  $\mathcal{V}$  da origen a la transformación lineal  $T = \tilde{\mathcal{B}} : \mathbb{R}^n \rightarrow \mathcal{V}$ . El sistema  $\mathcal{B}$  puede recuperarse a partir de  $T$ ; en efecto,  $\mathcal{B} = T(I) = [T\mathbf{e}_1, \dots, T\mathbf{e}_n]$ , donde  $I = [\mathbf{e}_1, \dots, \mathbf{e}_n]$  es la base canónica de  $\mathbb{R}^n$ .

En general, si  $\mathcal{A} = [\mathbf{u}_1, \dots, \mathbf{u}_n]$  es una base de un espacio  $\mathcal{U}$ , puede definirse una t.l.  $T : \mathcal{U} \rightarrow \mathcal{V}$  en la forma siguiente:

1. Se define en la base, conviniendo que

$$T\mathbf{u}_1 = \mathbf{v}_1, \dots, T\mathbf{u}_n = \mathbf{v}_n.$$

2. Se extiende por linealidad: como cada  $\mathbf{u}$  de  $\mathcal{U}$  puede escribirse, de manera única, en la forma  $\mathbf{u} = \lambda^1\mathbf{u}_1 + \dots + \lambda^n\mathbf{u}_n$ , se conviene que

$$T\mathbf{u} = \lambda^1\mathbf{v}_1 + \dots + \lambda^n\mathbf{v}_n.$$

De esta manera  $T$  queda bien definida y es lineal. Se observa además que  $\mathcal{B} = T\langle\mathcal{A}\rangle = [T\mathbf{u}_1, \dots, T\mathbf{u}_n]$ .

Estos hechos se usarán enseguida para relacionar matrices y transformaciones.

A) Tomemos una matriz  $A$  de  $m$  filas y  $n$  columnas, es decir, un elemento del conjunto  $\mathcal{M}_{m,n}$ :

$$A = [\mathbf{a}_1, \dots, \mathbf{a}_n], \quad \mathbf{a}_i \in \mathbb{R}^m.$$

La t.l. asociada  $\tau = \tilde{A}$  es un elemento de  $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$  tal que, para cada  $\mathbf{x} = [x^1, \dots, x^n]^t$  de  $\mathbb{R}^n$  (ver § 12):

$$\tau(\mathbf{x}) = A\mathbf{x} = \mathbf{a}_1x^1 + \dots + \mathbf{a}_nx^n. \quad (1)$$

En particular

$$A\mathbf{e}_i = \mathbf{a}_i \quad \text{para } i = 1, \dots, n. \quad (2)$$

Cuando  $m = 1$  la matriz  $A$  es un vector fila, digamos  $\mathbf{a} = [a_1, \dots, a_n]$ , y entonces (1) conduce a la siguiente regla para *multiplicar* filas por columnas (cfr. Preliminares, (8)):

$$\mathbf{a}\mathbf{x} = [a_1, \dots, a_n] \begin{bmatrix} x^1 \\ \vdots \\ x^n \end{bmatrix} = a_1x^1 + \dots + a_nx^n. \quad (3)$$

Es claro que  $\mathbf{a}\mathbf{x} = \mathbf{x}^t \mathbf{a}^t$ .

De la relación (1) se deduce que

[1] La  $i$ -ésima componente de  $A\mathbf{x}$  es  $\mathbf{a}^i\mathbf{x}$ .

B) Tomemos ahora una t.l.  $\tau : \mathbb{R}^n \rightarrow \mathbb{R}^m$ , es decir, un elemento de  $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ .

Entonces  $\tau \langle I \rangle = [\tau e_1, \dots, \tau e_n]$  es un elemento de  $\mathcal{M}_{m,n}$  y se llama *la matriz de  $\tau$* . Designándola  $[\tau]$  tenemos que

$$[\tau] = [\tau e_1, \dots, \tau e_n] .$$

La matriz de 1, la identidad de  $\mathbb{R}^n$ , es  $I$  y se llama la **matriz unidad** de orden  $n$  (cfr. § 15, ejemplo (1)).

La relación  $\tau = \tilde{A}$  es equivalente a la relación  $A = [\tau]$ .

Si  $A$  es la matriz de  $\tau$  entonces la t.l. asociada a  $A$  es  $\tau$ . Esto es: la transformación de  $[\tau]$  es  $\tau$ .

Por otra parte si  $\tau$  es la transformación de una matriz  $A$  entonces la matriz asociada a  $\tau$  es precisamente  $A$ . Esto es: la matriz de  $\tilde{A}$  es  $A$ .

Se concluye que la correspondencia entre transformaciones lineales de  $\mathbb{R}^n$  en  $\mathbb{R}^m$ , por un lado, y matrices  $m \times n$ , por el otro, tiene carácter biyectivo permitiendo *trasladar* a las matrices todas las operaciones entre transformaciones lineales. Por ejemplo, la suma  $A + B$  de dos matrices  $m \times n$  es la matriz  $m \times n$  de la suma  $\rho + \sigma$  de la t.l.  $\rho$  determinada por  $A$  y la t.l.  $\sigma$  determinada por  $B$ . Similarmente se define  $\lambda A$ . Con estas operaciones se tiene que:

[2]  $\mathcal{M}_{m,n}$  es un espacio vectorial isomorfo a  $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ .

En particular  $(\mathbb{R}^n)'$ , es decir,  $\mathcal{L}(\mathbb{R}^n, \mathbb{R})$  (ver § 18), es isomorfo a  $\mathcal{M}_{1,n}$ , es decir, a  $\mathbb{R}_n$ .

El **producto**  $AB$ , donde  $A$  es una matriz  $m \times n$  y  $B$  es una matriz  $n \times p$  es, por definición, la matriz  $m \times p$  del producto  $\rho\sigma : \mathbb{R}^p \rightarrow \mathbb{R}^m$  de la t.l.  $\rho : \mathbb{R}^n \rightarrow \mathbb{R}^m$ , determinada por  $A$ , y la t.l.  $\sigma : \mathbb{R}^p \rightarrow \mathbb{R}^n$ , determinada por  $B$ .

Como  $B = [\mathbf{b}_1, \dots, \mathbf{b}_p]$  tenemos que

$$AB = [(\rho\sigma)\mathbf{e}_1, \dots, (\rho\sigma)\mathbf{e}_p] = [A\mathbf{b}_1, \dots, A\mathbf{b}_p] ,$$

o también:

$$A[\mathbf{b}_1, \dots, \mathbf{b}_p] = [A\mathbf{b}_1, \dots, A\mathbf{b}_p] . \tag{4}$$

Escribiendo  $C = AB$  vemos que  $c_k^i$  es la  $i$ -ésima componente de  $A\mathbf{b}_k$ . Entonces, de acuerdo con [1] y (3) :

$$c_k^i = \mathbf{a}^i \mathbf{b}_k = a_1^i b_k^1 + \dots + a_n^i b_k^n .$$

Esta regla de multiplicación puede esquematizarse así:

$$\begin{array}{c}
 \left[ \begin{array}{c} \text{---} \\ \text{---} \mathbf{a}^i \text{---} \\ \text{---} \end{array} \right] \left[ \begin{array}{c} | \\ | \\ | \mathbf{b}_k \\ | \\ | \end{array} \right] = \left[ \begin{array}{c} \dots \\ \dots \mathbf{a}^i \mathbf{b}_k \dots \\ \dots \end{array} \right] \\
 \underbrace{\hspace{10em}}_{(m \times n) \quad (n \times p)} \hspace{1em} \underbrace{\hspace{10em}}_{(m \times p)} \tag{i}
 \end{array}$$

De acuerdo con la definición dada, las propiedades de la multiplicación matricial son las mismas de la composición entre transformaciones lineales (ver (1), § 18) y de ellas se derivan. En realidad, se ha aprovechado la correspondencia biunívoca entre  $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$  y  $\mathcal{M}_{m,n}$  para trasladar a este último conjunto toda la estructura del primero. En particular, resulta que  $\mathcal{M}_n$  es un álgebra isomorfa a  $\mathcal{L}(\mathbb{R}^n)$ .

Anotemos algunos hechos y peculiaridades:

1. El producto  $AB$  sólo está definido cuando el número de columnas de  $A$  es igual al número de filas de  $B$ .
2. En general  $AB \neq BA$ .
3. Es posible que  $AB = 0$  con  $A \neq 0$  y  $B \neq 0$ .
4. Es posible que  $AB = 0$  pero  $BA \neq 0$ .
5. Si  $AB$  está definida entonces  $(AB)^t = B^t A^t$ .
6.  $AI = A$ ,  $IB = B$  siempre que el producto esté definido.

## § 20 Transformaciones y matrices

Consideremos ahora una transformación lineal  $T : \mathcal{U} \rightarrow \mathcal{V}$ , una base  $\mathcal{A} = [\mathbf{u}_1, \dots, \mathbf{u}_n]$  de  $\mathcal{U}$  y una base  $\mathcal{B} = [\mathbf{v}_1, \dots, \mathbf{v}_m]$  de  $\mathcal{V}$ . En esta situación, la transformación  $T : \mathcal{U} \rightarrow \mathcal{V}$  queda representada por una transformación  $\tau : \mathbb{R}^n \rightarrow \mathbb{R}^m$ , de tal manera que, el diagrama

$$\begin{array}{ccc}
 \mathcal{U} & \xrightarrow{T} & \mathcal{V} \\
 \mathcal{A} \uparrow & & \uparrow \mathcal{B} \\
 \mathbb{R}^n & \xrightarrow{\tau} & \mathbb{R}^m
 \end{array}$$

es conmutativo, es decir,  $T\mathcal{A} = \mathcal{B}\tau$  (nótese que estamos escribiendo  $\mathcal{A}, \mathcal{B}$  en lugar de  $\tilde{\mathcal{A}}, \tilde{\mathcal{B}}$  (ver § 12)). Explícitamente,  $\tau = \mathcal{B}^{-1}T\mathcal{A}$  y se llama la transformación lineal **inducida** por  $T$ , *con respecto a* las bases  $\mathcal{A}$  y  $\mathcal{B}$ . La matriz  $C$  de  $\tau$  se llama la matriz de  $T$  *con respecto a* (las bases)  $\mathcal{A}$  y  $\mathcal{B}$ ; cuando  $\mathcal{U} = \mathcal{V}$  y  $\mathcal{A} = \mathcal{B}$  se dice, sencillamente, “con respecto a la base  $\mathcal{A}$ ”.

Como  $C = \tau\langle I \rangle = \mathcal{B}^{-1}T\mathcal{A}\langle \mathbf{e}_1, \dots, \mathbf{e}_n \rangle = [\mathcal{B}^{-1}(T\mathbf{u}_1), \dots, \mathcal{B}^{-1}(T\mathbf{u}_n)]$ , tenemos que  $C$  es una matriz  $m \times n$  cuyas *columnas* son

$$\mathbf{c}_i = [T\mathbf{u}_i]_{\mathcal{B}} \quad , \quad i = 1, \dots, n .$$

Además, como  $T = \mathcal{B}\tau\mathcal{A}^{-1}$ , entonces  $\mathbf{v} = T\mathbf{u}$  es equivalente a  $\tau(\mathcal{A}^{-1}\mathbf{u}) = \mathcal{B}^{-1}\mathbf{v}$ , es decir, a

$$[\mathbf{v}]_{\mathcal{B}} = C [\mathbf{u}]_{\mathcal{A}} .$$

Si  $A \in \mathcal{M}_{m,n}$  entonces  $A$  es la matriz de  $\tilde{A}$  con respecto a las bases canónicas de  $\mathbb{R}^n$  y  $\mathbb{R}^m$ .

Al tratar de desentrañar la estructura de un operador salta a la vista la utilidad de considerar bases que, aunque no sean simples (como las canónicas), sean convenientes para analizar ese operador particular, debido a que la matriz correspondiente es sencilla, por ejemplo, diagonal.

## § 21 Nulidad y rango

Dada una transformación lineal  $T : \mathcal{U} \rightarrow \mathcal{V}$ , el conjunto  $\mathcal{N}_T = \{\mathbf{u} \in \mathcal{U} : T(\mathbf{u}) = \mathbf{0}\}$  es un subespacio de  $\mathcal{U}$  y se llama el **núcleo** de  $T$ ; su dimensión es la **nulidad** de  $T$  y se denota  $\nu(T)$ . Es claro que  $T$  es inyectiva si y sólo si  $\mathcal{N}_T = \{\mathbf{0}\}$ , es decir, si y sólo si  $\nu(T) = 0$ .

Por otra parte, el recorrido  $\mathcal{R}_T$  de  $T$  es un subespacio de  $\mathcal{V}$  cuya dimensión se llama el **rango** de  $T$  y se denota  $\rho(T)$ . La t.l.  $T$  es sobreyectiva si y sólo si  $\mathcal{R}_T = \mathcal{V}$ , es decir, si y sólo si  $\rho(T) = \dim \mathcal{V}$ .

Cuando  $\mathcal{V} = \mathcal{U}$ , tanto  $\mathcal{N}_T$  como  $\mathcal{R}_T$  son subespacios  $T$ -invariantes.

Para una t.l.  $\tau : \mathbb{R}^n \rightarrow \mathbb{R}^m$ , con matriz  $A$ , se tiene que  $\mathcal{N}_\tau$  es el núcleo o conjunto solución del problema homogéneo

$$A\mathbf{x} = \mathbf{0} , \quad (1)$$

de tal manera que  $\nu(\tau) = n - r$ , siendo  $r$  la característica del sistema.

Por otra parte, como  $A = [\tau\mathbf{e}_1, \dots, \tau\mathbf{e}_n]$ , resulta que  $\mathcal{R}_\tau$  es el subespacio de  $\mathbb{R}^m$  generado por las columnas de  $A$ . El rango de este sistema (que es la característica  $r$  del problema (1) y también el rango de la matriz  $A$ ) es entonces  $\rho(\tau)$ . En consecuencia  $\nu(\tau) + \rho(\tau) = n$ .

Para una t.l. cualquiera  $T : \mathcal{U} \rightarrow \mathcal{V}$ , si  $\mathcal{A}$  y  $\mathcal{B}$  son bases de  $\mathcal{U}$  y de  $\mathcal{V}$ , respectivamente, se tiene que  $\mathcal{N}_T = \mathcal{A}\langle\mathcal{N}_T\rangle$  y  $\mathcal{R}_T = \mathcal{B}\langle\mathcal{R}_T\rangle$ , siendo  $\tau$  la t.l. inducida por  $T$ . El resultado puede entonces trasladarse inmediatamente a  $T$  obteniéndose la llamada **ecuación dimensional**,  $\nu(T) + \rho(T) = n$ , esto es:

$$\dim \mathcal{N}_T + \dim \mathcal{R}_T = \dim \mathcal{U} . \quad (2)$$

## § 22 Inversión

Aplicando la ecuación dimensional a un operador  $T : \mathcal{U} \rightarrow \mathcal{U}$  se observa que

[3] *Un operador lineal es inyectivo si y sólo si es sobreyectivo.*

Si  $S$  y  $T$  son operadores lineales de  $\mathcal{U}$  tales que

$$ST = 1 ,$$

entonces  $T$  es inyectivo y  $S$  es sobreyectivo, es decir, ambos son biyectivos e inversos uno del otro. Y entonces también se tiene que

$$TS = 1 .$$

Estos resultados se trasladan de la manera usual a las matrices: si  $A$  es una matriz  $n \times n$  y existe una matriz  $X$  tal que  $AX = I$ , entonces también  $XA = I$ . Si tal  $X$  existe es única: si  $Y$  fuera otra, tendríamos que  $YAX = (YA)X = IX = X$  y también  $YAX = Y(AX) = YI = Y$ . Se dice entonces que  $A$  es **regular** o **invertible**; la matriz  $X$  se llama la *inversa* de  $A$  y se denota  $A^{-1}$ . Obsérvese que  $A$  es regular si y sólo si  $\tilde{A}$  es biyectiva y, en este caso, la matriz de  $\tilde{A}^{-1}$  es precisamente la inversa de  $A$ .

Es claro que  $A$  es regular si y sólo si (el sistema formado por sus columnas) es linealmente independiente (y, por lo tanto, una base de  $\mathbb{R}^n$ ), es decir, si y sólo si el problema

$$Ax = \mathbf{0}$$

no tiene soluciones diferentes de la trivial.

Si  $A$  y  $B$  son regulares (y del mismo orden), entonces  $AB$  es regular y

$$(AB)^{-1} = B^{-1}A^{-1} .$$

Si  $A$  es regular  $A^t$  también lo es y

$$(A^t)^{-1} = (A^{-1})^t .$$

Entonces:  $A$  es regular si y sólo si el sistema formado por sus filas es linealmente independiente (y es, por lo tanto, una base de  $\mathbb{R}_n$ ).

Si  $A = [\mathbf{a}_1, \dots, \mathbf{a}_n]$  es regular, entonces (ver § 19, (2)):

$$A^{-1}\mathbf{a}_i = \mathbf{e}_i \quad , \quad (i = 1, \dots, n) . \tag{1}$$



Cuando  $A$  no es regular se dice que es **singular**.

Como  $AX = A[\mathbf{x}_1, \dots, \mathbf{x}_n] = [A\mathbf{x}_1, \dots, A\mathbf{x}_n]$  y, por otra parte,  $I = [\mathbf{e}_1, \dots, \mathbf{e}_n]$ , la condición  $AX = I$  es equivalente a los siguientes  $n$  sistemas de ecuaciones

$$A\mathbf{x}_1 = \mathbf{e}_1, \dots, A\mathbf{x}_n = \mathbf{e}_n.$$

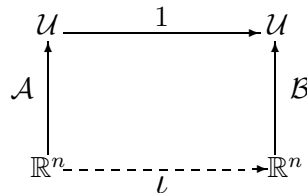
Como todos ellos tienen la misma matriz, pueden analizarse simultáneamente aplicando el algoritmo de eliminación a la matriz  $[A, I]$  hasta reducir todas las columnas de  $A$ . La matriz  $A$  será invertible si y sólo si la forma final es del tipo  $[I, B]$  y, en este caso,  $B$  es  $A^{-1}$ .

Por último, se observa que:

[4] Una matriz  $A$  es regular si y sólo si puede transformarse en  $I$  mediante operaciones elementales sobre sus filas o columnas.

## § 23 Cambio de base

Supongamos que en un espacio  $n$ -dimensional  $\mathcal{U}$  hay una base  $\mathcal{A}$ , llamémosla *azul*, y una base  $\mathcal{B}$ , *blanca*. ¿Cómo se expresan las coordenadas blancas en términos de las azules? Basta considerar la matriz  $P$  de la identidad, con respecto a las bases  $\mathcal{A}$  y  $\mathcal{B}$  (ver §§ 15 y 20):



$$[\mathbf{u}]_{\mathcal{B}} = P[\mathbf{u}]_{\mathcal{A}}.$$

Receta: para hallar la matriz  $P$  que expresa las coordenadas blancas en términos de las azules se expresan los vectores azules en términos de los blancos. Los coeficientes así obtenidos constituyen las columnas de  $P$ .

Como  $\iota$  es un isomorfismo,  $P$  es invertible y

$$[\mathbf{u}]_{\mathcal{A}} = P^{-1}[\mathbf{u}]_{\mathcal{B}}.$$

Ahora puede demostrarse que dos matrices  $A$  y  $B$  representan una *misma* transformación  $T : \mathcal{U} \rightarrow \mathcal{V}$ , con respecto a *bases diferentes* (tanto en  $\mathcal{U}$  como en  $\mathcal{V}$ ), si y sólo si existen matrices  $P$  y  $Q$ , invertibles, tales que

$$QA = BP .$$

En particular,  $A$  y  $B$  representan a un mismo operador  $T$  de  $\mathcal{U}$  si y sólo si existe  $P$  invertible tal que

$$B = P^{-1}AP .$$

En este caso se dice que  $A$  y  $B$  son *semejantes*.

## § 24 Matrices elementales

Son las obtenidas aplicando operaciones elementales a la matriz unidad  $I = [\mathbf{e}_1, \dots, \mathbf{e}_n]$ . Hay tres clases:

- (I)  $E_{ij} = [\dots, \underset{(i)}{\mathbf{e}_j}, \dots, \underset{(j)}{\mathbf{e}_i}, \dots]$  : intercambiando las columnas  $i$ -ésima y  $j$ -ésima de  $I$  (los índices entre paréntesis indican el lugar),
- (M)  $E_i(\lambda) = [\dots, \underset{(i)}{\lambda \mathbf{e}_i}, \dots]$  : multiplicando la  $i$ -ésima columna de  $I$  por  $\lambda (\neq 0)$ ,
- (S)  $E_{ij}(\lambda) = [\dots, \underset{(i)}{\mathbf{e}_i + \lambda \mathbf{e}_j}, \dots, \underset{(j)}{\mathbf{e}_j}, \dots]$  : sumándole a la  $i$ -ésima columna  $\lambda$  veces la  $j$ -ésima columna (de  $I$ ).

Directamente se comprueba que:

[5] *La traspuesta de una matriz elemental es una matriz elemental de la misma clase. Específicamente:*

$$E_{ij}^t = E_{ij} \quad , \quad (E_i(\lambda))^t = E_i(\lambda) \quad , \quad (E_{ij}(\lambda))^t = E_{ji}(\lambda) .$$

Sea ahora  $A = [\mathbf{a}_1, \dots, \mathbf{a}_n]$  una matriz de  $n$  columnas. Como

$$\begin{aligned}
 AE_{ij} &= [A\mathbf{e}_1, \dots, \underset{(i)}{A\mathbf{e}_j}, \dots, \underset{(j)}{A\mathbf{e}_i}, \dots, A\mathbf{e}_n] \\
 &= [\mathbf{a}_1, \dots, \underset{(i)}{\mathbf{a}_j}, \dots, \underset{(j)}{\mathbf{a}_i}, \dots, \mathbf{a}_n] ,
 \end{aligned}$$

se observa que:

(I) multiplicar  $A$  por  $E_{ij}$  equivale a intercambiar las columnas  $i$ -ésima y  $j$ -ésima de  $A$ .

Análogamente se comprueba que:

(M) multiplicar  $A$  por  $E_i(\lambda)$  equivale a multiplicar la  $i$ -ésima columna de  $A$  por  $\lambda$ , y

(S) multiplicar  $A$  por  $E_{ij}(\lambda)$  equivale a sumarle a la  $i$ -ésima columna  $\lambda$  veces la  $j$ -ésima columna (de  $A$ ).

Por ejemplo, la secuencia:

$$\begin{aligned}
 E_j(\lambda) &= [\dots, \mathbf{e}_i, \dots, \lambda\mathbf{e}_j, \dots] \rightarrow [\dots, \mathbf{e}_i + \lambda\mathbf{e}_j, \dots, \lambda\mathbf{e}_j, \dots] \\
 &\rightarrow [\dots, \mathbf{e}_i + \lambda\mathbf{e}_j, \dots, \frac{1}{\lambda}(\lambda\mathbf{e}_j), \dots] = E_{ij}(\lambda) ,
 \end{aligned}$$

muestra que, al multiplicar  $E_j(\lambda)$  por  $E_{ij}(1)$  y luego por  $E_j(1/\lambda)$ , se obtiene  $E_{ij}(\lambda)$ ; es decir

$$E_{ij}(\lambda) = E_j(\lambda) E_{ij}(1) E_j(1/\lambda) . \quad (1)$$

Así mismo, las tres secuencias siguientes:

$$(a) \ E_{ij} = [\dots, \underset{(i)}{\mathbf{e}_j}, \dots, \underset{(j)}{\mathbf{e}_i}, \dots] \rightarrow [\dots, \underset{(i)}{\mathbf{e}_i}, \dots, \underset{(j)}{\mathbf{e}_j}, \dots] = I ,$$

$$(b) \ E_i(\lambda) = [\dots, \underset{(i)}{\lambda\mathbf{e}_i}, \dots] \rightarrow [\dots, \underset{(i)}{\frac{1}{\lambda}(\lambda\mathbf{e}_i)}, \dots] = I ,$$

$$\begin{aligned}
 (c) \ E_{ij}(\lambda) &= [\dots, \underset{(i)}{\mathbf{e}_i + \lambda\mathbf{e}_j}, \dots, \underset{(j)}{\mathbf{e}_j}, \dots] \rightarrow [\dots, \underset{(i)}{(\mathbf{e}_i + \lambda\mathbf{e}_j) + \mu\mathbf{e}_j}, \dots, \underset{(j)}{\mathbf{e}_j}, \dots] \\
 &= E_{ij}(\lambda + \mu)
 \end{aligned}$$

muestran, respectivamente, que

$$\begin{aligned}
 \text{(a)} \quad & E_{ij}E_{ij} = I \\
 \text{(b)} \quad & E_i(\lambda)E_i(1/\lambda) = I \\
 \text{(c)} \quad & E_{ij}(\lambda)E_{ij}(\mu) = E_{ij}(\lambda + \mu); \text{ en particular } E_{ij}(\lambda)E_{ij}(-\lambda) = I,
 \end{aligned} \tag{2}$$

permitiendo afirmar que

[6] *Toda matriz elemental es regular y su inversa es una matriz elemental de la misma clase. Específicamente*

$$E_{ij}^{-1} = E_{ij} \quad , \quad (E_i(\lambda))^{-1} = E_i(1/\lambda) \quad , \quad (E_{ij}(\lambda))^{-1} = E_{ij}(-\lambda) .$$

**Descomposición de matrices invertibles.** Si  $A$  es invertible entonces puede transformarse en  $I$  mediante operaciones elementales sobre sus columnas (ver [4], § 22), es decir, existen matrices elementales  $E_1, E_2, \dots, E_s$  tales que

$$A E_1 E_2 \cdots E_s = I ,$$

y entonces  $A = E_s^{-1} \cdots E_2^{-1} E_1^{-1}$ . Teniendo en cuenta [6] se ve que

[7] *Toda matriz regular es producto de matrices elementales.*

## § 25 Formas bilineales

El tránsito de la linealidad a la multilinealidad, o sea la *linealidad en varias variables*, puede ilustrarse en un caso sencillo, pero de importancia especial, considerando las funciones reales definidas en el conjunto  $\mathcal{U} \times \mathcal{U}$ , de todas las parejas ordenadas  $(\mathbf{u}, \mathbf{v})$  de elementos de  $\mathcal{U}$ , y que son lineales en cada una de sus variables. Con más precisión: diremos que una función

$$f : \mathcal{U} \times \mathcal{U} \rightarrow \mathbb{R}$$

es una **forma bilineal** en  $\mathcal{U}$  si para,  $\mathbf{u}, \mathbf{v}, \mathbf{u}_i, \mathbf{v}_i$  en  $\mathcal{U}$  y  $\lambda \in \mathbb{R}$ ,

$$\begin{aligned}
 \text{(B.1)} \quad & f(\mathbf{u}_1 + \mathbf{u}_2, \mathbf{v}) = f(\mathbf{u}_1, \mathbf{v}) + f(\mathbf{u}_2, \mathbf{v}) \quad \text{y} \\
 & f(\mathbf{u}, \mathbf{v}_1 + \mathbf{v}_2) = f(\mathbf{u}, \mathbf{v}_1) + f(\mathbf{u}, \mathbf{v}_2) ,
 \end{aligned}$$

$$(B.2) \quad f(\lambda \mathbf{u}, \mathbf{v}) = \lambda f(\mathbf{u}, \mathbf{v}) \quad \text{y} \quad f(\mathbf{u}, \lambda \mathbf{v}) = \lambda f(\mathbf{u}, \mathbf{v}).$$

Las formas bilineales forman un subespacio de  $\mathcal{F}(\mathcal{U} \times \mathcal{U}, \mathbb{R})$ , el cual es isomorfo a  $\mathcal{M}_n$ : dada una base  $[\mathbf{u}_1, \dots, \mathbf{u}_n]$  de  $\mathcal{U}$ , a cada forma bilineal  $f$  se le hace corresponder la matriz  $A$  tal que  $a_{ij} = f(\mathbf{u}_i, \mathbf{u}_j)$ .

**Nota.** Debido a la eventual aparición de exponentes en el tratamiento de formas bilineales y temas relacionados, muchas veces escribiremos  $a_{ij}$  (resp.  $x_i$ ) en lugar de  $a_j^i$  (resp.  $x^i$ ).

Si  $f(\mathbf{u}, \mathbf{v}) = f(\mathbf{v}, \mathbf{u})$ , para todo  $\mathbf{u}$  y  $\mathbf{v}$ , se dice que  $f$  es **simétrica**. La **forma cuadrática** asociada a una forma bilineal simétrica  $f$  es la función  $q: \mathcal{U} \rightarrow \mathbb{R}$  tal que  $q(\mathbf{u}) = f(\mathbf{u}, \mathbf{u})$  para todo  $\mathbf{u}$  de  $\mathcal{U}$ .

Se dice que  $f$  (o que  $q$ ) es:

- (a) **positiva** si  $q(\mathbf{u}) \geq 0$  para todo  $\mathbf{u}$ ,
- (b) **definida** si  $q(\mathbf{u}) = 0$  implica  $\mathbf{u} = \mathbf{0}$ .

En  $\mathbb{R}^n$  una forma cuadrática está determinada por un polinomio de segundo grado en  $n$  variables:

$$q(x_1, \dots, x_n) = \sum_{i,j} a_{ij} x_i x_j$$

o, en forma matricial,  $q(\mathbf{x}) = \mathbf{x}^t A \mathbf{x}$ .

Una interpretación obvia de (B.1) y (B.2) conduce a la noción general de función bilineal sobre  $\mathcal{U} \times \mathcal{V}$ , con valores en  $\mathcal{W}$ :

$$f: \mathcal{U} \times \mathcal{V} \longrightarrow \mathcal{W},$$

donde  $\mathcal{U}, \mathcal{V}$  y  $\mathcal{W}$  son espacios vectoriales con los mismos escalares.

**Ejemplo.** La función  $f: \mathcal{U}' \times \mathcal{U} \longrightarrow \mathbb{R}$ ,  $f(\phi, \mathbf{u}) = \phi(\mathbf{u})$ , llamada *evaluación*, es una forma bilineal: es lineal en la primera variable por la definición de las operaciones (ver § 18) y en la segunda por la linealidad de  $\phi$ .

## Capítulo 4

# Diagonalización

Todo operador lineal en  $\mathbb{R}$  está definido por una ecuación de la forma

$$y = ax \quad , \quad \text{donde } a \text{ es una constante (escalar) .}$$

En  $\mathbb{R}^2$  cada operador  $T$  está definido por dos ecuaciones

$$\begin{aligned} y_1 &= a_{11}x_1 + a_{12}x_2 \\ y_2 &= a_{21}x_1 + a_{22}x_2 \quad , \end{aligned}$$

observándose que, en general, cada  $y_i$  depende tanto de  $x_1$  como de  $x_2$ .

Cuando cada  $y_i$  depende únicamente del  $x_i$  correspondiente,

$$\begin{aligned} y_1 &= \lambda_1 x_1 \\ y_2 &= \lambda_2 x_2 \quad , \end{aligned}$$

es decir, cuando la matriz de  $T$  es diagonal, se ve que  $T(\mathbf{u}) = \lambda_1 \mathbf{u}$  para todo  $\mathbf{u}$  en el primer eje coordenado,  $\mathcal{U}_1$  y  $T(\mathbf{u}) = \lambda_2 \mathbf{u}$  para todo  $\mathbf{u}$  en el segundo,  $\mathcal{U}_2$ : estas dos rectas son subespacios  $T$ -invariantes, unidimensionales, en los cuales  $T$  se comporta como lo hace un operador lineal de  $\mathbb{R}$ .

Con miras a descifrar la estructura de un operador  $T$  se trata de descomponer el espacio total en subespacios  $T$ -invariantes unidimensionales. Si  $\mathcal{N}$  es un tal subespacio, y  $\mathbf{u}$  es un elemento de  $\mathcal{N}$  entonces  $T\mathbf{u}$  debe ser de la forma  $\lambda\mathbf{u}$ , esto es :  $T$  actúa en  $\mathcal{N}$  como simple multiplicación por escalar, o cambio de escala, como sucede en  $\mathbb{R}$ .

## § 26 Autovalores y autovectores

Sea  $T$  un operador lineal en un espacio vectorial  $\mathcal{U}$ . Si  $\mathbf{u}$  es un vector y  $\lambda$  un escalar tales que

$$\mathbf{u} \neq \mathbf{0} \quad \text{y} \quad T\mathbf{u} = \lambda\mathbf{u} ,$$

se dice que  $\lambda$  es un **autovalor** (o **valor propio**) y que  $\mathbf{u}$  es un **autovector** (o **vector propio**) de  $T$ , que se corresponden. También se dice en este caso que  $\mathbf{u}$  es un  $\lambda$ -*autovector* de  $T$ . Obsérvese que los dos conceptos se definen *simultáneamente*.

La idea subyacente es geométrica: un autovector de  $T$  es un vector (no nulo) que no cambia de dirección al ser transformado por  $T$ .

Con frecuencia escribiremos  $T - \lambda$  en lugar de  $T - \lambda 1$ , donde  $1$  es la identidad  $1_{\mathcal{U}}$  de  $\mathcal{U}$ . Así mismo, si  $A$  es una matriz cuadrada entonces se escribe  $A - \lambda$  en lugar de  $A - \lambda I$ , donde  $I$  es la matriz unidad correspondiente.

Es claro que  $\mathbf{u}$  es un  $\lambda$ -autovector de  $T$  si y sólo si  $\mathbf{u}$  es un  $0$ -autovector de  $T - \lambda$ , de modo que los  $\lambda$ -autovectores de  $T$  son precisamente los vectores no nulos del núcleo de  $T - \lambda$ .

Para cada escalar  $\lambda$ , el núcleo  $\mathcal{N}_{\lambda}$  de  $T - \lambda$  es  $T$ -invariante y se llama el  $\lambda$ -**subespacio propio** de  $T$ ; su dimensión es la **multiplicidad geométrica** de  $\lambda$ . Un escalar  $\lambda$  es un autovalor de  $T$  si y sólo si  $\mathcal{N}_{\lambda} \neq \{\mathbf{0}\}$ .

También es fácil ver que  $\mathbf{u}$  es un autovector de  $T$  si y sólo si el subespacio generado por  $\mathbf{u}$  es unidimensional y  $T$ -invariante.

Estas mismas definiciones se aplican a una matriz  $A \in \mathcal{M}_n$ , considerada como un operador en  $\mathbb{R}^n$ , a saber, el operador asociado  $\tilde{A}$ .

Si  $A$  es la matriz de un operador  $T$  de  $\mathcal{U}$ , con respecto a una base  $\mathcal{B}$  de  $\mathcal{U}$ , entonces  $A$  y  $T$  tienen los mismos autovalores y el isomorfismo  $\tilde{\mathcal{B}}$  establece una correspondencia biunívoca entre los autovectores de  $A$  y los de  $T$ .

Sea  $A \in \mathcal{M}_n$ . Un escalar  $\lambda$  es un autovalor de  $A$  si y sólo si el problema homogéneo

$$(A - \lambda)\mathbf{x} = \mathbf{0} , \tag{1}$$

tiene soluciones diferentes de la trivial, es decir (ver [1], § 52) si y sólo si

$$\det (A - \lambda) = 0 . \quad (2)$$

De acuerdo con la fórmula (7) de la § 52, ésta es una ecuación polinómica de grado  $n$  en  $\lambda$ ; es la llamada **ecuación característica** de  $A$  y tiene la forma

$$p(\lambda) = \det (A - \lambda) = a_0 + a_1\lambda + a_2\lambda^2 + \dots + (-1)^n\lambda^n = 0 .$$

Si  $\lambda_1, \dots, \lambda_n$  son sus raíces tenemos que  $p(\lambda) = (\lambda_1 - \lambda) \dots (\lambda_n - \lambda)$  y, haciendo  $\lambda = 0$ , se observa que

$$\det A = \lambda_1 \dots \lambda_n . \quad (3)$$

Si suponemos que  $\lambda_1, \dots, \lambda_s$  son las raíces *distintas* de  $p(\lambda)$ , entonces

$$p(\lambda) = (\lambda_1 - \lambda)^{m_1} \dots (\lambda_s - \lambda)^{m_s} ,$$

siendo  $m_j$  la **multiplicidad (algebraica)** de  $\lambda_j$ . Esta multiplicidad nunca es inferior a la geométrica. Para verlo, escribamos  $\alpha = \lambda_j$  y tomemos una base  $[\mathbf{b}_1, \dots, \mathbf{b}_k]$  de  $\mathcal{N}_\alpha = \{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} = \alpha\mathbf{x}\}$ . Completándola obtenemos una base  $B = [\mathbf{b}_1, \dots, \mathbf{b}_n]$  de  $\mathbb{R}^n$ . Si  $\mathbf{c}_i$  es la  $i$ -ésima columna de la matriz  $C = B^{-1}AB$ , entonces, para  $1 \leq i \leq k$ :

$$\mathbf{c}_i = (B^{-1}AB)\mathbf{e}_i = B^{-1}A\mathbf{b}_i = B^{-1}(\alpha\mathbf{b}_i) = \alpha B^{-1}\mathbf{b}_i = \alpha\mathbf{e}_i ,$$

(ver § 19, (2) y § 22, (1)), por lo tanto:

$$p(\lambda) = \det (A - \lambda) = \det B(C - \lambda)B^{-1} = \det (C - \lambda) = (\alpha - \lambda)^k q(\lambda) ,$$

lo cual prueba que  $k$ , la multiplicidad geométrica de  $\alpha$ , es menor o igual que su multiplicidad algebraica.

Recapitulando: las raíces de (2) son precisamente los autovalores de  $A$ . Para hallar los autovectores primero se hallan las raíces de (2) y luego, para cada raíz  $\lambda = \lambda_j$ , se resuelve el problema (1), es decir se halla una base de su núcleo; los elementos no nulos de éste son los  $\lambda_j$ -autovectores de la matriz  $A$ .

**Notas.** (1) En este punto irrumpe un elemento extraño al espíritu de nuestro tema pues, salvo en dimensión uno, la ecuación característica *no*



es lineal. Esto a su vez franquea la entrada a los números complejos, pues los ceros de un polinomio pueden no ser reales.

Admitiendo a los complejos como campo de escalares, la *existencia* de valores propios se basa en el teorema fundamental del álgebra, el cual es, esencialmente, un resultado del análisis complejo. La existencia de autovalores y de autovectores no es entonces un hecho trivial. Su profundidad se refleja en sus formidables consecuencias.

**Espacios complejos.** Puede comprobarse sin dificultad que los desarrollos realizados hasta ahora continúan siendo válidos al sustituir  $\mathbb{R}$  por  $\mathbb{C}$ . Los espacios vectoriales cuyos escalares están en  $\mathbb{C}$  se llaman **espacios complejos**.

**El espacio  $\mathbb{C}^n$ .** El conjunto de todas las  $n$ -plas ordenadas de números complejos se denota  $\mathbb{C}^n$ . Su estructura es similar a la de  $\mathbb{R}^n$ , salvo por el producto interno (ver § 30, ejemplo (1)).

Como los complejos incluyen a los reales, resulta que  $\mathbb{R}^n$  es un subconjunto (mas no un subespacio) de  $\mathbb{C}^n$ . En particular  $[e_1, \dots, e_n]$  es una base de  $\mathbb{C}^n$ , de modo que  $\dim \mathbb{C}^n = n$ . (También es posible considerar a  $\mathbb{C}^n$  como un espacio vectorial sobre  $\mathbb{R}$  y en este caso su dimensión es  $2n$  y  $\mathbb{R}^n$  es un subespacio. Pero este punto de vista no es útil aquí).

Cada elemento  $z$  de  $\mathbb{C}^n$  puede escribirse, de manera única, en la forma

$$z = x + iy \quad \text{con} \quad x, y \in \mathbb{R}^n .$$

Si, además,  $w = u + iv$  (con  $u, v \in \mathbb{R}^n$ ) y  $\lambda = \mu + i\nu$  (con  $\mu, \nu \in \mathbb{R}$ ), inmediatamente se comprueba que:

- (i)  $\bar{z} = x - iy$ ,
- (ii)  $z + w = x + u + i(y + v)$ ,
- (iii)  $\lambda z = \mu x - \nu y + i(\nu x + \mu y)$ .
- (iv) Si  $A$  es una matriz real y  $z = x + iy$  es un  $\lambda$ -autovector de  $A$ , con  $\lambda = \mu + i\nu$ , entonces

$$Ax = \mu x - \nu y \quad \text{y} \quad Ay = \nu x + \mu y , \quad (4)$$

y además  $\bar{z}$  es un  $\bar{\lambda}$ -autovector de  $A$ .

**Notas.** (2) Algunas veces, en lugar de  $\mathbb{R}$  ó  $\mathbb{C}$  escribiremos  $\mathbb{K}$ ; en lo sucesivo  $\mathcal{U}$  será un espacio complejo de dimensión  $n$  y  $A, B, \dots$  serán matrices complejas  $n \times n$ , salvo que se diga otra cosa.

(3) Si  $A = [a_j^i]$  entonces la matriz  $\bar{A} = [b_j^i]$  con  $b_j^i = \bar{a_j^i}$  se llama la **conjugada** de  $A$ .

## § 27 Operadores y matrices diagonalizables

Se dice que un operador  $T$  en  $\mathcal{U}$  es **diagonalizable** si existe una base  $\mathcal{B}$  de  $\mathcal{U}$  tal que la matriz de  $T$  con respecto a  $\mathcal{B}$  es diagonal; se dice entonces que  $\mathcal{B}$  es una base **adecuada** para  $T$ .

Directamente se comprueba que  $\mathcal{B}$  es una base adecuada para  $T$  si y sólo si la base canónica es adecuada para el operador  $\tau$ , inducido por  $T$ , con respecto a la base  $\mathcal{B}$  (ver § 20).

Decimos que una matriz  $A$  es diagonalizable si el operador  $\tilde{A}$  lo es, es decir (ver § 23), si existe una matriz *regular*  $U$  tal que

$$U^{-1}AU$$

es una matriz diagonal.

Esta propiedad depende de la abundancia de autovectores:

[1] Una matriz  $A$ , de orden  $n$ , es diagonalizable si y sólo si existe una base  $U$  de  $\mathbb{K}^n$  constituida exclusivamente por autovectores de  $A$ .

DEMOSTRACIÓN. Sea  $U = [\mathbf{u}_1, \dots, \mathbf{u}_n]$  una base de  $\mathbb{K}^n$  tal que cada  $\mathbf{u}_i$  es un  $\lambda_i$ -autovector y sea  $\Lambda = \Delta(\lambda_1, \dots, \lambda_n)$ . Entonces, (ver § 19, (4)):

$$\begin{aligned} U\Lambda &= U[\lambda_1\mathbf{e}_1, \dots, \lambda_n\mathbf{e}_n] = [\lambda_1U\mathbf{e}_1, \dots, \lambda_nU\mathbf{e}_n] \\ &= [\lambda_1\mathbf{u}_1, \dots, \lambda_n\mathbf{u}_n], \\ AU &= A[\mathbf{u}_1, \dots, \mathbf{u}_n] = [A\mathbf{u}_1, \dots, A\mathbf{u}_n]. \end{aligned}$$

Como  $A\mathbf{u}_i = \lambda_i\mathbf{u}_i$  y, además,  $U$  es invertible, entonces  $U^{-1}AU = \Lambda$ . Los mismos cálculos prueban la recíproca. ♠

Este resultado se traslada sin dificultad a operadores:

[1'] *Un operador  $T$  en  $\mathcal{U}$  es diagonalizable si y sólo si existe una base de  $\mathcal{U}$  constituida exclusivamente por autovectores de  $T$ .*

En otros términos : *Una base es adecuada para  $T$  si y sólo si todos sus elementos son autovectores de  $T$ .*

Dicho de otro modo: para que  $T$  sea diagonalizable basta que haya suficientes autovectores para generar a  $\mathcal{U}$  (ver [12], § 15).

Hay un criterio de independencia de autovectores:

[2] *Autovectores correspondientes a autovalores distintos son linealmente independientes.*

DEMOSTRACIÓN. Supongamos que  $\lambda_1, \dots, \lambda_p$  son *distintos* y que, para cada  $i$ ,  $\mathbf{u}_i$  es un  $\lambda_i$ -autovector. Como  $\mathbf{u}_1 \neq \mathbf{0}$  se tiene que  $[\mathbf{u}_1]$  es linealmente independiente.

Suponiendo ahora que  $\mathcal{A}_k = [\mathbf{u}_1, \dots, \mathbf{u}_k]$ ,  $k < p$ , es l.i., probemos que  $\mathcal{A}_{k+1} = [\mathbf{u}_1, \dots, \mathbf{u}_k, \mathbf{u}_{k+1}]$  es l.i. En caso contrario  $\mathbf{u}_{k+1} \in \langle \mathcal{A}_k \rangle$  :

$$\mathbf{u}_{k+1} = \mu_1 \mathbf{u}_1 + \dots + \mu_k \mathbf{u}_k . \quad (1)$$

Aplicando  $A$  obtenemos:

$$\lambda_{k+1} \mathbf{u}_{k+1} = \lambda_1 \mu_1 \mathbf{u}_1 + \dots + \lambda_k \mu_k \mathbf{u}_k . \quad (2)$$

Multiplicando (1) por  $\lambda_{k+1}$  y restando de (2) se obtiene:

$$\mathbf{0} = (\lambda_1 - \lambda_{k+1}) \mu_1 \mathbf{u}_1 + \dots + (\lambda_k - \lambda_{k+1}) \mu_k \mathbf{u}_k .$$

Como  $\mathcal{A}_k$  es libre y  $\lambda_{k+1} \neq \lambda_i$  para  $i = 1, \dots, k$  tendremos que  $\mu_1 = \dots = \mu_k = 0$ , lo cual implica que  $\mathbf{u}_{k+1} = \mathbf{0}$ . Pero esto es imposible pues  $\mathbf{u}_{k+1}$  es un autovector. ♠

En consecuencia:

[3] *Si una matriz  $n \times n$  tiene  $n$  autovalores distintos es diagonalizable.*

## § 28 Descomposición espectral

Si  $\lambda_1, \dots, \lambda_s$  son los *distintos* autovalores de  $T$ , entonces los subespacios propios correspondientes,  $\mathcal{N}_j = \{\mathbf{u} \in \mathcal{U} : T\mathbf{u} = \lambda_j\mathbf{u}\}$ ,  $j = 1, \dots, s$ , son independientes.

El operador  $T$  es diagonalizable si y sólo si  $\mathcal{U} = \mathcal{N}_1 \oplus \dots \oplus \mathcal{N}_s$ . Esta es la **descomposición espectral** del espacio  $\mathcal{U}$  **subordinada** al operador diagonalizable  $T$ .

Si  $P_1, \dots, P_s$  son las proyecciones asociadas a esta descomposición, (ver § 18, ejemplo), entonces, para cada  $j$ ,  $TP_j = \lambda_j P_j$  y, como  $1 = P_1 + \dots + P_s$ , entonces  $T = \lambda_1 P_1 + \dots + \lambda_s P_s$ . Esta es la **forma espectral** de  $T$ .

Cuando  $T$  *no* es diagonalizable, la suma directa  $\mathcal{N}_1 \oplus \dots \oplus \mathcal{N}_s$  no alcanza a ser igual a  $\mathcal{U}$ . El sistema (libre)  $\mathcal{B}$  obtenido reuniendo sendas bases de los  $\mathcal{N}_j$  *no* es una base de  $\mathcal{U}$ . No obstante, es posible completar  $\mathcal{B}$  hasta obtener una base en la cual la matriz de  $T$  sea aproximadamente diagonal. Hablamos de la *forma canónica de Jordan* (ver § 53).

## § 29 Diagonalización simultánea

Supongamos que  $T$  y  $S$  conmutan (i.e.  $TS = ST$ ) y son diagonalizables; podemos entonces, por una parte, usar la descomposición espectral de  $\mathcal{U}$  subordinada a  $T$  y, por la otra, tomar una base  $[\mathbf{v}_1, \dots, \mathbf{v}_n]$  adecuada para  $S$ . Empecemos escribiendo

$$\mathbf{v}_1 = \mathbf{w}_1 + \dots + \mathbf{w}_s, \quad \text{con} \quad \mathbf{w}_j \in \mathcal{N}_j, \quad (j = 1, \dots, s),$$

y observemos que  $T(S\mathbf{w}_1) = S(T\mathbf{w}_1) = S(\lambda_1\mathbf{w}_1) = \lambda_1(S\mathbf{w}_1)$ , es decir,  $S\mathbf{w}_1 \in \mathcal{N}_1$ . En general,  $S\mathbf{w}_j \in \mathcal{N}_j$  para  $j = 1, \dots, s$ .

Si  $\mu_1$  es el autovalor de  $\mathbf{v}_1$  tenemos que

$$S\mathbf{v}_1 = S\mathbf{w}_1 + \dots + S\mathbf{w}_s \quad \text{y también} \quad S\mathbf{v}_1 = \mu_1\mathbf{v}_1 = \mu_1\mathbf{w}_1 + \dots + \mu_1\mathbf{w}_s.$$

Debido a la unicidad de la descomposición tenemos que

$$S\mathbf{w}_j = \mu_1\mathbf{w}_j \quad \text{para} \quad j = 1, \dots, s,$$

---

o sea que, cada uno de los vectores  $w_j$ , componentes de  $v_1$ , es un autovector de  $S$ , a menos que sea  $\mathbf{0}$ .

Descomponiendo igualmente a  $v_2, \dots, v_n$ , y excluyendo los componentes nulos, se obtiene un sistema de vectores que genera a  $\mathcal{U}$  (pues los  $v_i$  lo hacen) y son, a la vez, autovectores de  $S$  y de  $T$ . Dicho sistema puede reducirse a una base que será adecuada tanto para  $S$  como para  $T$ .

Suponiendo que existe una base adecuada para los operadores  $S_1, \dots, S_n$  y que  $T$  conmuta con todos ellos, se observa que el resultado es válido para cualquier número de operadores diagonalizables que conmutan dos a dos. Para matrices tenemos entonces que: si  $A_1, \dots, A_n$  son diagonalizables y conmutan dos a dos, entonces existe una matriz regular  $U$  tal que, para cada  $i = 1, \dots, n$ , la matriz  $U^{-1}A_iU$  es diagonal (y recíprocamente).

## Capítulo 5

# Ortogonalidad

Las nociones discutidas hasta ahora, en el contexto de los espacios vectoriales, tienen bastante sabor geométrico. Han quedado por fuera, sin embargo, conceptos como los de ángulo, perpendicularidad, proyección ortogonal y distancia, entre los cuales hay conexiones significativas. En el caso del plano se observa claramente que estas ideas están relacionadas con la de producto interno. La extensión a  $\mathbb{R}^n$  de las nociones correspondientes es bastante natural y sirve de enlace con la formulación general que desarrollaremos en este capítulo.

### § 30 Producto interno

Un **producto interno** (p.i.) en un espacio complejo  $\mathcal{U}$  es una función que a cada par de vectores  $\mathbf{u}, \mathbf{v}$  de  $\mathcal{U}$  le hace corresponder un número complejo, denotado  $\langle \mathbf{u}, \mathbf{v} \rangle$ , de tal manera que:

$$(PI.1) \quad \langle \mathbf{u}, \mathbf{v}_1 + \mathbf{v}_2 \rangle = \langle \mathbf{u}, \mathbf{v}_1 \rangle + \langle \mathbf{u}, \mathbf{v}_2 \rangle,$$

$$(PI.2) \quad \langle \lambda \mathbf{u}, \mathbf{v} \rangle = \lambda \langle \mathbf{u}, \mathbf{v} \rangle,$$

$$(PI.3) \quad \langle \mathbf{v}, \mathbf{u} \rangle = \overline{\langle \mathbf{u}, \mathbf{v} \rangle},$$

$$(PI.4) \quad \langle \mathbf{u}, \mathbf{u} \rangle \geq 0. \text{ Además } \langle \mathbf{u}, \mathbf{u} \rangle = 0 \text{ si y sólo si } \mathbf{u} = \mathbf{0}.$$

Otras propiedades útiles del producto interno se derivan fácilmente de las anteriores propiedades básicas. Por ejemplo :

$$\begin{aligned} \langle \mathbf{u}_1 + \mathbf{u}_2, \mathbf{v} \rangle &= \langle \mathbf{u}_1, \mathbf{v} \rangle + \langle \mathbf{u}_2, \mathbf{v} \rangle , \\ \langle \mathbf{u}, \lambda \mathbf{v} \rangle &= \bar{\lambda} \langle \mathbf{u}, \mathbf{v} \rangle , \\ \left\langle \sum_i \lambda_i \mathbf{u}_i, \sum_j \mu_j \mathbf{v}_j \right\rangle &= \sum_{i,j} \lambda_i \bar{\mu}_j \langle \mathbf{u}_i, \mathbf{v}_j \rangle , \\ \langle \mathbf{u}, \mathbf{0} \rangle &= \langle \mathbf{0}, \mathbf{u} \rangle = 0 . \end{aligned}$$

Además,  $\langle \mathbf{u}, \mathbf{v} \rangle = 0$ , para todo  $\mathbf{v}$ , si y sólo si  $\mathbf{u} = \mathbf{0}$ .

**Ejemplo.** (1) Para  $\mathbf{z} = [z_1, \dots, z_n]^t$ ,  $\mathbf{w} = [w_1, \dots, w_n]^t$  en  $\mathbb{C}^n$  se define

$$\langle \mathbf{z}, \mathbf{w} \rangle = z_1 \bar{w}_1 + \dots + z_n \bar{w}_n = \mathbf{w}^* \mathbf{z} , \text{ donde } \mathbf{w}^* = \bar{\mathbf{w}}^t .$$

Este es un producto interno en  $\mathbb{C}^n$  y coincide con el producto interno usual en  $\mathbb{R}^n$  (cuando  $\mathbf{z}, \mathbf{w} \in \mathbb{R}^n$ ).

Si  $\mathbf{z} = \mathbf{x} + i\mathbf{y}$  y  $\mathbf{w} = \mathbf{u} + i\mathbf{v}$  ( $\mathbf{x}, \mathbf{y}, \mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ ) entonces

$$\langle \mathbf{z}, \mathbf{w} \rangle = \langle \mathbf{x}, \mathbf{u} \rangle + \langle \mathbf{y}, \mathbf{v} \rangle - i (\langle \mathbf{x}, \mathbf{v} \rangle - \langle \mathbf{y}, \mathbf{u} \rangle) .$$

**Norma.** La **norma** o **longitud** asociada a un producto interno es la función

$$\|\mathbf{u}\| = \langle \mathbf{u}, \mathbf{u} \rangle^{1/2} .$$

De esta definición se deducen las siguientes propiedades básicas:

(N.1) Para todo  $\mathbf{u}$ ,  $\|\mathbf{u}\| \geq 0$ . Además,  $\|\mathbf{u}\| = 0$  si y sólo si  $\mathbf{u} = \mathbf{0}$ .

(N.2) Para todo  $\mathbf{u}$  y todo  $\lambda$ ,  $\|\lambda \mathbf{u}\| = |\lambda| \|\mathbf{u}\|$ .

La **distancia** entre  $\mathbf{u}$  y  $\mathbf{v}$  es la longitud de su diferencia:

$$\|\mathbf{u} - \mathbf{v}\| .$$

Si  $\mathbf{u}$  es tal que  $\|\mathbf{u}\| = 1$ , se dice que es un **vector unitario**; para todo  $\mathbf{u} \neq \mathbf{0}$  el vector  $\mathbf{u}/\|\mathbf{u}\|$  es unitario.

**Vectores ortogonales.** Cuando  $\langle \mathbf{u}, \mathbf{v} \rangle = 0$  se dice que  $\mathbf{u}$  y  $\mathbf{v}$  son **ortogonales**, y se escribe  $\mathbf{u} \perp \mathbf{v}$ .

Si  $\mathcal{R}$  y  $\mathcal{S}$  son subconjuntos no vacíos de  $\mathcal{U}$ , tales que  $\mathbf{u} \perp \mathbf{v}$  para todo  $\mathbf{u} \in \mathcal{R}$  y todo  $\mathbf{v} \in \mathcal{S}$ , se dice que  $\mathcal{R}$  y  $\mathcal{S}$  son *ortogonales* y se escribe  $\mathcal{R} \perp \mathcal{S}$ . En lugar de  $\{\mathbf{u}\} \perp \mathcal{S}$  se escribe  $\mathbf{u} \perp \mathcal{S}$  y se dice que  $\mathbf{u}$  es *ortogonal* a  $\mathcal{S}$ . El conjunto de todos los vectores ortogonales a  $\mathcal{S}$  es un *subespacio* llamado el **ortogonal** de  $\mathcal{S}$  y denotado  $\mathcal{S}^\perp$ . En general  $\mathcal{S} \subset \mathcal{S}^{\perp\perp} (= (\mathcal{S}^\perp)^\perp)$ .

**Ejemplos.** (2) El conjunto solución de un sistema homogéneo,  $A\mathbf{x} = \mathbf{0}$ , es el ortogonal de (las columnas de)  $A^t$ .

(3) Para que un vector sea ortogonal a un subespacio  $\mathcal{N}$  basta que sea ortogonal a un sistema de generadores de  $\mathcal{N}$ .

**Sistemas ortonormales.** Se dice que un sistema  $\mathcal{A} = [\mathbf{u}_1, \dots, \mathbf{u}_n]$  de vectores *no nulos* es **ortogonal** si  $\mathbf{u}_i \perp \mathbf{u}_j$  siempre que  $i \neq j$ . Si, además, los  $\mathbf{u}_i$  son todos unitarios se dice que  $\mathcal{A}$  es un sistema **ortonormal**.

Hay relación entre ortogonalidad y libertad:

[1] *Todo sistema ortogonal es libre.*

DEMOSTRACIÓN. Supongamos que  $x_1\mathbf{u}_1 + \dots + x_k\mathbf{u}_k = \mathbf{0}$  y escribamos  $\mathbf{u} = x_1\mathbf{u}_1 + \dots + x_k\mathbf{u}_k$ . Multiplicando por  $\mathbf{u}_1$  tenemos que, por una parte,  $\langle \mathbf{u}, \mathbf{u}_1 \rangle = \langle \mathbf{0}, \mathbf{u}_1 \rangle = 0$ , y, por la otra,  $\langle \mathbf{u}, \mathbf{u}_1 \rangle = x_1\langle \mathbf{u}_1, \mathbf{u}_1 \rangle + \dots + x_k\langle \mathbf{u}_k, \mathbf{u}_1 \rangle = x_1\|\mathbf{u}_1\|^2$ .

Como  $\|\mathbf{u}_1\| \neq 0$  se concluye que  $x_1 = 0$ . Similarmente se deduce que los otros coeficientes deben ser nulos. ♠

**Ejemplo.** (4) Dados  $\mathbf{z}$  y  $\mathbf{w}$  como en el ejemplo (1), si  $\mathbf{z} \perp \bar{\mathbf{z}}$  entonces  $\mathbf{x} \perp \mathbf{y}$ ,  $\|\mathbf{x}\| = \|\mathbf{y}\|$  y  $\|\mathbf{z}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2$ . Si el sistema  $[\mathbf{z}, \bar{\mathbf{z}}, \mathbf{w}, \bar{\mathbf{w}}]$  es ortonormal entonces  $[\sqrt{2}\mathbf{x}, \sqrt{2}\mathbf{y}, \sqrt{2}\mathbf{u}, \sqrt{2}\mathbf{v}]$  también lo es y genera el mismo subespacio. La generalización es directa.

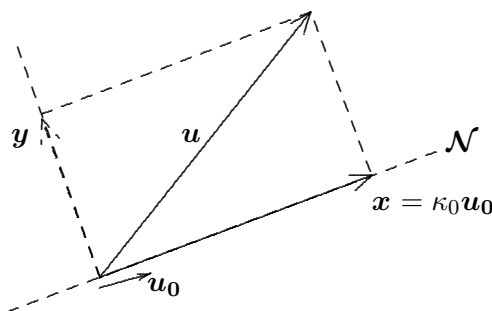
### § 31 Proyección ortogonal de un vector sobre otro

Sea  $\mathbf{u}_0$  un vector unitario. Para cada vector  $\mathbf{u}$  existe un *único* escalar  $\kappa_0$  tal que  $\mathbf{u} - \kappa_0\mathbf{u}_0$  es ortogonal a  $\mathbf{u}_0$ .



En efecto, la condición  $\mathbf{u} - \kappa_0 \mathbf{u}_0 \perp \mathbf{u}_0$ , i.e.  $\langle \mathbf{u} - \kappa_0 \mathbf{u}_0, \mathbf{u}_0 \rangle = 0$ , es equivalente a

$$\kappa_0 = \langle \mathbf{u}, \mathbf{u}_0 \rangle .$$



Este número  $\kappa_0$  es la **componente** de  $\mathbf{u}$  a lo largo de  $\mathbf{u}_0$ . El vector  $\mathbf{x} = \kappa_0 \mathbf{u}_0 = \langle \mathbf{u}, \mathbf{u}_0 \rangle \mathbf{u}_0$  es la **proyección** (ortogonal) de  $\mathbf{u}$  sobre  $\mathbf{u}_0$ , o la **parte** de  $\mathbf{u}$  paralela a  $\mathbf{u}_0$ , su longitud es  $|\kappa_0|$  y se caracteriza por ser el elemento de la recta  $\mathcal{N}$ , determinada por  $\mathbf{u}_0$ , *más próximo* a  $\mathbf{u}$ . En efecto, para un elemento cualquiera  $\lambda \mathbf{u}_0$  de  $\mathcal{N}$  se tiene que

$$\begin{aligned} \|\mathbf{u} - \lambda \mathbf{u}_0\|^2 &= \langle \mathbf{u} - \lambda \mathbf{u}_0, \mathbf{u} - \lambda \mathbf{u}_0 \rangle \\ &= \langle \mathbf{u}, \mathbf{u} \rangle - \bar{\lambda} \langle \mathbf{u}, \mathbf{u}_0 \rangle - \lambda \langle \mathbf{u}_0, \mathbf{u} \rangle + \lambda \bar{\lambda} \langle \mathbf{u}_0, \mathbf{u}_0 \rangle \\ &= \|\mathbf{u}\|^2 - \bar{\lambda} \kappa_0 - \lambda \bar{\kappa}_0 + \lambda \bar{\lambda} . \end{aligned}$$

Restando y sumando  $\kappa_0 \bar{\kappa}_0$ , es decir,  $|\kappa_0|^2$ , se observa que

$$\|\mathbf{u} - \lambda \mathbf{u}_0\|^2 = \|\mathbf{u}\|^2 - |\kappa_0|^2 + |\lambda - \kappa_0|^2 ,$$

donde se ve que  $\|\mathbf{u} - \lambda \mathbf{u}_0\|^2$ , y por lo tanto  $\|\mathbf{u} - \lambda \mathbf{u}_0\|$ , es *mínima* precisamente cuando  $\lambda = \kappa_0$ , es decir, cuando  $\lambda \mathbf{u}_0 = \mathbf{x}$ .

También se ve allí (haciendo  $\lambda = \kappa_0$ ) que  $|\kappa_0| \leq \|\mathbf{u}\|$  (i.e. “la proyección (ortogonal) no puede ser mayor que lo proyectado”), es decir

$$|\langle \mathbf{u}, \mathbf{u}_0 \rangle| \leq \|\mathbf{u}\| .$$

Si  $\mathbf{v}$  es un vector cualquiera, no nulo, y tomamos  $\mathbf{u}_0 = \mathbf{v}/\|\mathbf{v}\|$  obtenemos la célebre

**Desigualdad de Schwarz.** Para todo  $\mathbf{u}$  y todo  $\mathbf{v}$

$$|\langle \mathbf{u}, \mathbf{v} \rangle| \leq \|\mathbf{u}\| \|\mathbf{v}\| .$$

Si alguno de los vectores es nulo la desigualdad se cumple trivialmente.

Este hecho permite obtener la última de las propiedades básicas de la norma, a saber, la *desigualdad del triángulo*:

$$(N.3) \quad \text{Para todo } \mathbf{u} \text{ y todo } \mathbf{v}, \|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|.$$

## § 32 Proyección ortogonal de un vector sobre un subespacio

Para generalizar la noción de proyección consideremos ahora un vector  $\mathbf{u}$  y un subespacio  $\mathcal{N}$ , generado por un sistema ortonormal  $[\mathbf{u}_1, \dots, \mathbf{u}_k]$ . Buscar un elemento de  $\mathcal{N}$  cuya distancia a  $\mathbf{u}$  sea mínima significa buscar escalares  $\lambda_i$  tales que  $\|\mathbf{u} - \sum \lambda_i \mathbf{u}_i\|$  sea mínima. Un cómputo similar al anteriormente efectuado muestra que

$$\|\mathbf{u} - \sum \lambda_i \mathbf{u}_i\|^2 = \|\mathbf{u}\|^2 - \sum |\kappa_i|^2 + \sum |\lambda_i - \kappa_i|^2,$$

donde  $\kappa_i = \langle \mathbf{u}, \mathbf{u}_i \rangle$  es la *componente* de  $\mathbf{u}$  a lo largo de  $\mathbf{u}_i$ . Se concluye que el vector

$$\mathbf{x} = \kappa_1 \mathbf{u}_1 + \dots + \kappa_k \mathbf{u}_k$$

es el elemento de  $\mathcal{N}$  más próximo a  $\mathbf{u}$ . Este vector  $\mathbf{x}$  se llama la **proyección** (ortogonal) de  $\mathbf{u}$  sobre  $\mathcal{N}$  o la **parte** de  $\mathbf{u}$  en  $\mathcal{N}$  y está caracterizado por la propiedad de que el vector

$$\mathbf{y} = \mathbf{u} - \mathbf{x}$$

es ortogonal a  $\mathcal{N}$ , es decir, está en  $\mathcal{N}^\perp$ . Este vector  $\mathbf{y}$  se llama la **parte** de  $\mathbf{u}$  ortogonal a  $\mathcal{N}$ .

Es claro que  $\mathbf{x} = \mathbf{0}$  si y sólo si  $\mathbf{u} \perp \mathcal{N}$ . También es claro que  $\mathbf{u} \in \mathcal{N}$  si y sólo si  $\mathbf{y} = \mathbf{0}$ , de modo que si  $\mathbf{u} \notin \mathcal{N}$  y definimos

$$\mathbf{u}_{k+1} = \mathbf{y}/\|\mathbf{y}\|$$

entonces  $\mathbf{u}_{k+1}$  es una combinación lineal de  $\mathbf{u}_1, \dots, \mathbf{u}_k, \mathbf{u}$ , y el sistema ampliado  $[\mathbf{u}_1, \dots, \mathbf{u}_k, \mathbf{u}_{k+1}]$  es también ortonormal.

### § 33 Ortonormalización

A partir de un sistema libre  $\mathcal{B} = [\mathbf{v}_1, \dots, \mathbf{v}_p]$  puede construirse, paso a paso, uno ortonormal  $\mathcal{A} = [\mathbf{u}_1, \dots, \mathbf{u}_p]$ , de tal manera que, para cada  $i$ ,

$$\langle \mathcal{A}_i \rangle = \langle \mathcal{B}_i \rangle ,$$

donde  $\mathcal{A}_i = [\mathbf{u}_1, \dots, \mathbf{u}_i]$  y  $\mathcal{B}_i = [\mathbf{v}_1, \dots, \mathbf{v}_i]$ , para  $i = 1, \dots, p$ . En particular  $\langle \mathcal{A} \rangle = \langle \mathcal{B} \rangle$ .

El primer paso es obvio: como  $\mathcal{B}$  es libre entonces  $\mathbf{v}_1 \neq \mathbf{0}$  y podemos definir

$$\mathbf{u}_1 = \mathbf{v}_1 / \|\mathbf{v}_1\| .$$

Suponiendo ahora que ya hemos construido  $\mathcal{A}_k$ ,  $k < p$ , que satisface los requerimientos, observamos que  $\mathbf{v}_{k+1} \notin \langle \mathcal{A}_k \rangle$  y entonces el último resultado de la § 32, con  $\mathbf{u} = \mathbf{v}_{k+1}$  y  $\mathcal{N} = \langle \mathcal{A}_k \rangle$ , permite obtener un vector  $\mathbf{u}_{k+1}$  tal que  $[\mathbf{u}_1, \dots, \mathbf{u}_k, \mathbf{u}_{k+1}]$  es ortonormal y  $\mathbf{u}_{k+1}$ , es c.l. de  $\mathbf{u}_1, \dots, \mathbf{u}_k$  y  $\mathbf{v}_{k+1}$  y, por lo tanto, de  $\mathbf{v}_1, \dots, \mathbf{v}_k$  y  $\mathbf{v}_{k+1}$ ; en consecuencia  $\langle \mathcal{A}_{k+1} \rangle \subset \langle \mathcal{B}_{k+1} \rangle$  y entonces  $\langle \mathcal{B}_{k+1} \rangle = \langle \mathcal{A}_{k+1} \rangle$ , ya que tienen la misma dimensión. Es decir,  $\langle \mathcal{A}_{k+1} \rangle$  satisface los requerimientos. De esta manera llegamos a  $\mathcal{A}_p$  que es el  $\mathcal{A}$  buscado.

Este es el llamado *proceso* (de ortogonalización) *de Gram-Schmidt*.

### § 34 Bases ortonormales

El proceso de ortogonalización garantiza la existencia de bases ortonormales, las cuales brindan ventajas en un espacio con producto interno. En efecto, el mismo procedimiento seguido para establecer [1] en la § 30 muestra que:

[2] *Las componentes  $x_i$  de un vector  $\mathbf{u}$  en una base ortonormal  $[\mathbf{u}_1, \dots, \mathbf{u}_n]$  vienen dadas por*

$$x_i = \langle \mathbf{u}, \mathbf{u}_i \rangle \quad , \quad i = 1, \dots, n .$$

Además:  $\|\mathbf{u}\|^2 = |x_1|^2 + \cdots + |x_n|^2$  (Pitágoras).

En particular, si  $\mathbf{x} \in \mathbb{K}^n$  y  $[\mathbf{e}_1, \dots, \mathbf{e}_n]$  es la base canónica, entonces

$$x_i = \langle \mathbf{x}, \mathbf{e}_i \rangle, \quad i = 1, \dots, n \quad \text{y} \quad \|\mathbf{x}\|^2 = |x_1|^2 + \cdots + |x_n|^2.$$

**Nota.** A menudo se usa el término *ortogonal* (resp. *unitario*) para referirse a un sistema (o matriz)  $[\mathbf{a}_1, \dots, \mathbf{a}_n]$  *ortonormal* en  $\mathbb{R}^n$  (resp.  $\mathbb{C}^n$ ).

## § 35 Proyecciones ortogonales

Dado un subespacio  $\mathcal{N}$  de  $\mathcal{U}$ , la función  $P : \mathcal{U} \rightarrow \mathcal{U}$  que a cada  $\mathbf{u}$  de  $\mathcal{U}$  le hace corresponder su proyección ortogonal sobre  $\mathcal{N}$  (ver § 32), es un operador lineal y se llama, por supuesto, la **proyección** (ortogonal) sobre  $\mathcal{N}$ ; su recorrido es  $\mathcal{N}$  y su núcleo es  $\mathcal{N}^\perp$ , de modo que

$$\dim \mathcal{N} + \dim \mathcal{N}^\perp = \dim \mathcal{U}. \quad (1)$$

Nótese que, como  $\mathcal{N} \subset \mathcal{N}^{\perp\perp}$  y  $\dim \mathcal{N}^{\perp\perp} = \dim \mathcal{U} - \dim \mathcal{N}^\perp = \dim \mathcal{N}$ , entonces  $\mathcal{N} = \mathcal{N}^{\perp\perp}$ .

La proyección (ortogonal) sobre  $\mathcal{N}^\perp$  es  $Q = 1 - P$ . En realidad  $\mathcal{U} = \mathcal{N} \oplus \mathcal{N}^\perp$  y  $P$  y  $Q$  son las proyecciones asociadas a esta descomposición de  $\mathcal{U}$ , (ver § 18, ejemplo).

Dos subespacios  $\mathcal{N}_1$  y  $\mathcal{N}_2$ , con proyecciones  $P_1$  y  $P_2$ , respectivamente, son ortogonales si y sólo si  $P_1 P_2 = 0$ . En particular,  $PQ = 0$ .

Si  $\mathcal{U} = \mathbb{K}^n$  y  $P_0$  es la proyección sobre el subespacio (unidimensional) generado por un vector unitario  $\mathbf{u}_0$ , se tiene que

$$\begin{aligned} P_0(\mathbf{u}) &= \langle \mathbf{u}, \mathbf{u}_0 \rangle \mathbf{u}_0 = \mathbf{u}_0 \langle \mathbf{u}, \mathbf{u}_0 \rangle \\ &= \mathbf{u}_0 (\mathbf{u}_0^* \mathbf{u}) = (\mathbf{u}_0 \mathbf{u}_0^*) \mathbf{u}, \quad \text{para todo } \mathbf{u} \text{ de } \mathbb{K}^n, \end{aligned}$$

de modo que  $\mathbf{u}_0 \mathbf{u}_0^*$  es precisamente la matriz de  $P_0$ .

Si  $[\mathbf{u}_1, \dots, \mathbf{u}_n]$  es una base ortonormal de  $\mathbb{K}^n$ , cada  $\mathbf{u}$  de  $\mathbb{K}^n$  se expresa en la forma

$$\mathbf{u} = x_1 \mathbf{u}_1 + \cdots + x_n \mathbf{u}_n,$$

donde (ver [2], § 34)  $x_i \mathbf{u}_i = \langle \mathbf{u}, \mathbf{u}_i \rangle \mathbf{u}_i = (\mathbf{u}_i \mathbf{u}_i^*) \mathbf{u}$ . Entonces:

$$\mathbf{u} = \langle \mathbf{u}, \mathbf{u}_1 \rangle \mathbf{u}_1 + \cdots + \langle \mathbf{u}, \mathbf{u}_n \rangle \mathbf{u}_n = (\mathbf{u}_1 \mathbf{u}_1^* + \cdots + \mathbf{u}_n \mathbf{u}_n^*) \mathbf{u} , \quad (2)$$

para todo  $\mathbf{u}$  de  $\mathbb{K}^n$ . Por lo tanto

$$\mathbf{u}_1 \mathbf{u}_1^* + \cdots + \mathbf{u}_n \mathbf{u}_n^* = I .$$

**Ejemplo.** En toda variedad  $\mathcal{S} = \mathcal{N} + \mathbf{c}$  (ver § 9) hay un vector de norma mínima, a saber, la proyección ortogonal de  $\mathbf{c}$  (o de cualquier otro elemento de  $\mathcal{S}$ ) sobre  $\mathcal{N}^\perp$ . Tal vector es el único elemento de  $\mathcal{S} \cap \mathcal{N}^\perp$ .

## § 36 Diagonalización ortonormal

Como las bases ortonormales son muy convenientes, dado un operador  $T$  nos preguntamos si existe una base adecuada para  $T$  (ver § 27) que sea ortonormal. Si éste es el caso diremos que  $T$  es **ortonormalmente diagonalizable** ( $\perp$ -diagonalizable). La noción se aplica a matrices vía el operador asociado (cfr. § 26).

Los resultados de las §§ 27 y 28 pueden ajustarse a esta situación. En particular, si  $T$  es  $\perp$ -diagonalizable, en la descomposición subordinada a  $T$  (ver § 28) los subespacios  $\mathcal{N}_j$  son mutuamente ortogonales y cada  $P_j$  es la proyección ortogonal sobre  $\mathcal{N}_j$ . A la descomposición espectral de  $\mathcal{U}$ ,

$$\mathcal{U} = \mathcal{N}_1 \oplus \cdots \oplus \mathcal{N}_s ,$$

corresponde la forma espectral de  $T$ ,

$$T = \sum_1^s \lambda_j P_j , \quad (1)$$

donde (ver § 18, ejemplo): a)  $P_j^2 = P_j$  para todo  $j = 1, \dots, s$ , b)  $P_j P_k = 0$  si  $j \neq k$ , c)  $\sum_1^s P_j = 1_{\mathcal{U}}$ .

Puede demostrarse que tal descomposición es única.

Estas relaciones implican que, para cualquier polinomio  $f(t)$ ,

$$f(T) = \sum_1^s f(\lambda_j) P_j . \quad (2)$$

Si los  $\lambda_j$  son, todos, distintos de 0 entonces  $T$  es invertible, y recíprocamente. En este caso  $T^{-1} = \sum \lambda_j^{-1} P_j$ .

Los resultados anteriores se reformulan de la manera usual para matrices. En este caso se puede también proceder directamente como sigue: sea  $A$  una matriz de orden  $n$  y sea  $U = [\mathbf{u}_1, \dots, \mathbf{u}_n]$  una base ortonormal, adecuada para  $A$ , con  $A\mathbf{u}_i = \xi_i \mathbf{u}_i$  ( $i = 1, \dots, n$ ). Para todo  $\mathbf{u}$  de  $\mathbb{K}^n$  se tiene que (ver § 35, (2)):

$$\begin{aligned} A\mathbf{u} &= A(\langle \mathbf{u}, \mathbf{u}_1 \rangle \mathbf{u}_1 + \dots + \langle \mathbf{u}, \mathbf{u}_n \rangle \mathbf{u}_n) \\ &= \langle \mathbf{u}, \mathbf{u}_1 \rangle \xi_1 \mathbf{u}_1 + \dots + \langle \mathbf{u}, \mathbf{u}_n \rangle \xi_n \mathbf{u}_n \\ &= [\xi_1 (\mathbf{u}_1 \mathbf{u}_1^*) + \dots + \xi_n (\mathbf{u}_n \mathbf{u}_n^*)] \mathbf{u} \quad , \end{aligned}$$

y, finalmente,

$$A = \sum_1^n \xi_i K_i \quad ,$$

donde  $K_i$  es la matriz de la proyección ortogonal sobre la *recta* (subespacio unidimensional) generada por  $\mathbf{u}_i$ . Agrupando los sumandos correspondientes a los  $\xi_i$  repetidos, se obtiene la expresión correspondiente a la forma espectral (1).

## § 37 Representación de funcionales. Adjunción

Cada elemento  $\mathbf{v}$  de  $\mathcal{U}$  presenta una faceta dinámica cuando observamos que, con  $\mathbf{v}$  fijo, la expresión  $\langle \mathbf{u}, \mathbf{v} \rangle$  define un funcional de  $\mathcal{U}$ :

$$\varphi_{\mathbf{v}} : \mathcal{U} \rightarrow \mathbb{K} \quad ,$$

donde  $\varphi_{\mathbf{v}}(\mathbf{u}) = \langle \mathbf{u}, \mathbf{v} \rangle$ , para todo  $\mathbf{u}$  de  $\mathcal{U}$ . Dos vectores distintos no pueden originar un mismo funcional.

Nos preguntamos si todo funcional  $\varphi : \mathcal{U} \rightarrow \mathbb{K}$  es de esta forma. Es claro que, si hay un elemento  $\mathbf{v}$  de  $\mathcal{U}$  tal que  $\varphi(\mathbf{u}) = \langle \mathbf{u}, \mathbf{v} \rangle$  (para todo  $\mathbf{u}$  de  $\mathcal{U}$ ), entonces  $\mathbf{v}$  es ortogonal a todo elemento  $\mathbf{u}$  del núcleo  $\mathcal{N}$  de  $\varphi$ , de modo que  $\mathbf{v}$  debe ser buscado en  $\mathcal{N}^\perp$  y, si existe, será tanto más fácil hallarlo cuanto más pequeño sea  $\mathcal{N}^\perp$ ; en particular, si  $\varphi = 0$  necesariamente  $\mathbf{v} = \mathbf{0}$ .

Cuando  $\varphi \neq 0$ , su recorrido es  $\mathbb{K}$  (pues no es  $\{0\}$ ) y entonces la ecuación dimensional indica que  $\dim \mathcal{N} = \dim \mathcal{U} - 1$ , concluyéndose que  $\dim \mathcal{N}^\perp = 1$  (ver § 35, (1)); es decir,  $\mathcal{N}^\perp = \langle \mathbf{u}_0 \rangle$ , para algún  $\mathbf{u}_0$  con  $\|\mathbf{u}_0\| = 1$  y entonces el  $\mathbf{v}$  buscado debe ser de la forma  $\lambda \mathbf{u}_0$ . Como, para cada  $\mathbf{u}$  de  $\mathcal{U}$ , debe cumplirse que

$$\varphi(\mathbf{u}) = \langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{u}, \lambda \mathbf{u}_0 \rangle = \bar{\lambda} \langle \mathbf{u}, \mathbf{u}_0 \rangle \quad (1)$$

y, por otra parte,  $\mathbf{u} = \mathbf{x} + \mathbf{y}$  con  $\mathbf{x} \in \mathcal{N}^\perp$  y  $\mathbf{y} \in \mathcal{N}$ , donde  $\mathbf{x} = \langle \mathbf{u}, \mathbf{u}_0 \rangle \mathbf{u}_0$  (ver § 31), entonces

$$\varphi(\mathbf{u}) = \varphi(\mathbf{x}) + \varphi(\mathbf{y}) = \varphi(\mathbf{x}) = \langle \mathbf{u}, \mathbf{u}_0 \rangle \varphi(\mathbf{u}_0) . \quad (2)$$

Comparando (1) y (2) se ve que basta tomar  $\lambda = \overline{\varphi(\mathbf{u}_0)}$ .

Se establece así una correspondencia biyectiva, entre  $\mathcal{U}$  y  $\mathcal{U}'$ , la cual permite visualizar los elementos de  $\mathcal{U}$  como funcionales y recíprocamente.

Supongamos ahora que  $T : \mathcal{U} \rightarrow \mathcal{V}$  es una transformación lineal entre espacios con producto interno. Cada elemento  $\mathbf{v}$  de  $\mathcal{V}$  determina un funcional  $\psi = \psi_{\mathbf{v}} : \mathcal{U} \rightarrow \mathbb{K}$ ,  $\psi(\mathbf{u}) = \langle T\mathbf{u}, \mathbf{v} \rangle$ , el cual, según la discusión anterior, determina un elemento de  $\mathcal{U}$  (que depende tanto de  $T$  como de  $\mathbf{v}$ , y por ello se denota  $T^*\mathbf{v}$ ) caracterizado por la condición

$$\psi(\mathbf{u}) = \langle \mathbf{u}, T^*\mathbf{v} \rangle ,$$

para todo  $\mathbf{u}$  de  $\mathcal{U}$ . Queda así definida una función  $T^* : \mathcal{V} \rightarrow \mathcal{U}$ , llamada la **adjunta** de  $T$  que verifica la siguiente *relación de adjunción*: para *todo*  $\mathbf{u}$  de  $\mathcal{U}$  y *todo*  $\mathbf{v}$  de  $\mathcal{V}$ ,

$$(RA) \quad \langle \mathbf{u}, T^*\mathbf{v} \rangle = \langle T\mathbf{u}, \mathbf{v} \rangle .$$

Esta condición caracteriza a  $T^*$ , sin involucrar explícitamente la identificación entre vectores y funcionales, y permite demostrar las propiedades básicas de la operación  $\star$ ; por ejemplo:

$$\begin{aligned} (S \pm T)^* &= S^* \pm T^* & (\lambda T)^* &= \bar{\lambda} T^* \\ 1^* &= 1 & T^{**} &= (T^*)^* = T \\ (ST)^* &= T^* S^* , \end{aligned} \quad (3)$$

siempre que las sumas y productos estén definidos. Además,  $T$  es invertible si y sólo si  $T^*$  lo es y, en este caso,

$$(T^{-1})^* = (T^*)^{-1} .$$

Si  $A = [\mathbf{a}_1, \dots, \mathbf{a}_n] = [a_{ij}]$  es una matriz  $m \times n$ , su adjunta  $A^*$  es, por definición, la matriz de la adjunta de la transformación  $T$  definida por  $A$ .

Sea  $A^* = [\mathbf{b}_1, \dots, \mathbf{b}_m] = [b_{ij}]$ . Aplicando (RA) con  $T = \tilde{A}$ ,  $\mathbf{u} = \mathbf{e}_i$ , ( $1 \leq i \leq n$ ) y  $\mathbf{v} = \mathbf{f}_j$ , ( $1 \leq j \leq m$ ) (vectores de las bases canónicas de  $\mathbb{K}^n$  y  $\mathbb{K}^m$ , respectivamente) y, teniendo en cuenta [2], de la § 34, obtenemos:

$$\left. \begin{array}{l} \langle \mathbf{e}_i, T^* \mathbf{f}_j \rangle = \langle \mathbf{e}_i, A^* \mathbf{f}_j \rangle = \langle \mathbf{e}_i, \mathbf{b}_j \rangle = \overline{b_{ij}} \\ \parallel \\ \langle T \mathbf{e}_i, \mathbf{f}_j \rangle = \langle A \mathbf{e}_i, \mathbf{f}_j \rangle = \langle \mathbf{a}_i, \mathbf{f}_j \rangle = a_{ji} \end{array} \right\} \implies b_{ij} = \overline{a_{ji}}$$

En otros términos  $A^*$  es la conjugada (ver § 26, nota (3)) de la traspuesta (ver Preliminares) de  $A$ :

$$A^* = \overline{A^t} . \quad (4)$$

En particular, cada fila de  $A^*$  es la conjugada de la traspuesta de la correspondiente columna de  $A$ . Al multiplicar  $A^*$  por  $A$ , el elemento situado en la fila  $i$  y la columna  $k$  del producto es igual al producto interno de  $\mathbf{a}_k$  por  $\mathbf{a}_i$  (ver § 30, ej. (1)). Por lo tanto:  $A$  es ortonormal si y sólo si  $A^*A = I$ .

En el caso real

$$A^* = A^t , \quad (5)$$

y la matriz real  $A$  es ortogonal si y sólo si  $A^tA = I$ .

### § 38 El álgebra $\mathcal{L}(\mathcal{U})$

Cuando  $\mathcal{V} = \mathcal{U}$ , la adjunción  $\star$  es una operación en  $\mathcal{L}(\mathcal{U})$ , la cual, debido a que  $T^{\star\star} = T$ , se llama una *involución*. Ahora  $\mathcal{L}(\mathcal{U})$  (y por lo tanto  $\mathcal{M}_n$ ) ha adquirido una estructura polifacética: es un espacio vectorial,



posee una multiplicación con elemento unidad, una involución (y hasta es posible dotarla de una norma) y todo ello muy bien armonizado (ver § 37, (3) y § 18, (1)). Tiene, no obstante algunas peculiaridades, como la no conmutatividad, que provienen de su misma naturaleza. Con esta salvedad,  $\mathcal{L}(\mathcal{U})$  es muy parecida al álgebra de todas las funciones complejas definidas sobre un conjunto (y, más particularmente, al campo complejo  $\mathbb{C}$ ) con la conjugación como involución.

En  $\mathcal{L}(\mathcal{U})$  surgen posibilidades de interacción entre un operador y su adjunto, las cuales resultan especialmente fructíferas. Por ejemplo, es claro que, según (RA) (ver § 37), para todo  $\mathbf{u}, \mathbf{v}$  de  $\mathcal{U}$ ,

$$T\mathbf{u} \perp \mathbf{v} \quad \text{si y sólo si} \quad \mathbf{u} \perp T^*\mathbf{v} ,$$

lo cual, aunque no parece a primera vista, implica que

[3] *Un subespacio  $\mathcal{N}$  es  $T$ -invariante si y sólo si su ortogonal  $\mathcal{N}^\perp$  es  $T^*$ -invariante.*

DEMOSTRACIÓN. Supongamos que  $\mathcal{N}$  es  $T$ -invariante y tomemos  $\mathbf{v} \in \mathcal{N}^\perp$ . Para todo  $\mathbf{u}$  de  $\mathcal{N}$  se cumple que  $T\mathbf{u} \in \mathcal{N}$  y entonces  $T\mathbf{u} \perp \mathbf{v}$ , es decir,  $\mathbf{u} \perp T^*\mathbf{v}$ . En otros términos,  $T^*\mathbf{v} \in \mathcal{N}^\perp$ .

La recíproca es inmediata ya que  $\mathcal{N}^{\perp\perp} = \mathcal{N}$  y  $T^{**} = T$ . ♠

Es posible que un operador coincida con su adjunto. Por ejemplo, si  $P$  es la proyección (ortogonal) sobre un subespacio  $\mathcal{N}$  entonces, dados  $\mathbf{u}, \mathbf{v}$  en  $\mathcal{U}$ , puede escribirse

$$\mathbf{u} = \mathbf{u}_1 + \mathbf{u}_2 \quad \text{y} \quad \mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2 \quad \text{con} \quad \mathbf{u}_1, \mathbf{v}_1 \in \mathcal{N} \quad \text{y} \quad \mathbf{u}_2, \mathbf{v}_2 \in \mathcal{N}^\perp$$

y entonces

$$\langle P\mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{u}_1, \mathbf{v}_1 + \mathbf{v}_2 \rangle = \langle \mathbf{u}_1, \mathbf{v}_1 \rangle + \langle \mathbf{u}_1, \mathbf{v}_2 \rangle = \langle \mathbf{u}_1, \mathbf{v}_1 \rangle$$

$$\text{y} \quad \langle \mathbf{u}, P\mathbf{v} \rangle = \langle \mathbf{u}_1 + \mathbf{u}_2, \mathbf{v}_1 \rangle = \langle \mathbf{u}_1, \mathbf{v}_1 \rangle + \langle \mathbf{u}_2, \mathbf{v}_1 \rangle = \langle \mathbf{u}_1, \mathbf{v}_1 \rangle .$$

Por lo tanto  $P^* = P$ .

Esto llama la atención sobre los operadores  $\perp$ -diagonalizables pues,

para un tal operador  $T$ , la forma espectral (§ 36, (1)) indica que

$$T^* = \sum_1^s \bar{\lambda}_j P_j, \quad (1)$$

de donde se deduce que los operadores  $\perp$ -diagonalizables conmutan con su adjunto. Específicamente

$$T^*T = TT^* = \sum_1^s |\lambda_j|^2 P_j.$$

Por último, comparando (1) con (§ 36, (2)) se concluye que, si  $T$  es  $\perp$ -diagonalizable, entonces su adjunto es un polinomio en  $T$  (basta tomar un polinomio  $f(t)$  tal que  $f(\lambda_j) = \bar{\lambda}_j$  para  $j = 1, \dots, s$ ).

### § 39 Formas sesquilineales

El producto interno usual de  $\mathbb{R}^n$ ,  $\langle \mathbf{x}, \mathbf{y} \rangle = x_1 y_1 + \dots + x_n y_n$ , el cual es una forma bilineal simétrica (ver § 25), no puede extenderse directamente a  $\mathbb{C}^n$ , pues no siempre se tendría  $\langle \mathbf{x}, \mathbf{x} \rangle \geq 0$ , lo cual es necesario para introducir una noción razonable de *longitud* (por ejemplo, en  $\mathbb{C}^2$  el vector  $[1, i]^t$  tendría norma 0). La modificación necesaria no es mayor (ver § 30, ejemplo (1)) pero altera la simetría, y parte de la linealidad en la segunda variable. Esto conduce a modificar la noción de forma bilineal con el fin de hacerla útil en espacios complejos: de una función  $f : \mathcal{U} \times \mathcal{U} \rightarrow \mathbb{C}$  que satisface las propiedades (B.1) de la § 25 y

$$(B.2') : f(\lambda \mathbf{u}, \mathbf{v}) = \lambda f(\mathbf{u}, \mathbf{v}) \quad \text{y} \quad f(\mathbf{u}, \lambda \mathbf{v}) = \bar{\lambda} f(\mathbf{u}, \mathbf{v}),$$

se dice que es una forma **sesquilineal** (sesqui =  $1\frac{1}{2}$ ). Si, además,  $f(\mathbf{u}, \mathbf{v}) = \overline{f(\mathbf{v}, \mathbf{u})}$  se dice que  $f$  es **hermitiana**; las otras nociones de la § 25 se aplican aquí y todas coinciden, en el caso real, con las allí tratadas.

**Representación.** Para cada operador  $T$  en  $\mathcal{U}$ , la ecuación

$$f(\mathbf{u}, \mathbf{v}) = \langle T\mathbf{u}, \mathbf{v} \rangle \quad (1)$$

define una forma sesquilineal. Recíprocamente, si  $f$  es una forma sesquilineal, cada  $\mathbf{u}$  de  $\mathcal{U}$  define un funcional lineal  $\varphi(\mathbf{v}) = \overline{f(\mathbf{u}, \mathbf{v})}$ , determinando

entonces un elemento  $\mathbf{w}$  de  $\mathcal{U}$  tal que  $\varphi(\mathbf{v}) = \langle \mathbf{v}, \mathbf{w} \rangle$  (para todo  $\mathbf{v} \in \mathcal{U}$ ): queda definido así un operador  $T$ , que a  $\mathbf{u}$  le hace corresponder  $\mathbf{w}$ , y este operador obviamente satisface (1).

La modificación introducida tiene por fin tratar el caso complejo. Sin embargo, la idea básica sigue siendo la multilinealidad en el producto cartesiano.

**Nota.** El producto cartesiano permite juntar elementos de clases distintas, sin revolverlos. Por eso es útil para combinar cosas que, en principio, no son combinables; por ejemplo, sumar o multiplicar vectores de espacios distintos. Para ilustrarlo tomemos dos espacios vectoriales  $\mathcal{U}$  y  $\mathcal{V}$ , con bases  $[\mathbf{u}_1, \dots, \mathbf{u}_m]$  y  $[\mathbf{v}_1, \dots, \mathbf{v}_n]$ , respectivamente.

a) la **suma directa (externa)**  $\mathcal{U} \oplus \mathcal{V}$  es el espacio vectorial formado por el conjunto  $\mathcal{U} \times \mathcal{V}$  con las operaciones

$$(\mathbf{x}, \mathbf{y}) + (\mathbf{u}, \mathbf{v}) = (\mathbf{x} + \mathbf{u}, \mathbf{y} + \mathbf{v}), \quad \lambda(\mathbf{u}, \mathbf{v}) = (\lambda\mathbf{u}, \lambda\mathbf{v}).$$

Cada  $\mathbf{u} \in \mathcal{U}$  (resp.  $\mathbf{v} \in \mathcal{V}$ ) puede identificarse con  $(\mathbf{u}, \mathbf{0})$  (resp.  $(\mathbf{0}, \mathbf{v})$ ); además  $(\mathbf{u}, \mathbf{v}) = (\mathbf{u}, \mathbf{0}) + (\mathbf{0}, \mathbf{v})$ ; en este sentido  $(\mathbf{u}, \mathbf{v})$  representa la *suma* de  $\mathbf{u}$  y  $\mathbf{v}$ .

Los vectores  $(\mathbf{u}_i, \mathbf{0})$ ,  $(\mathbf{0}, \mathbf{v}_j)$  forman una base de  $\mathcal{U} \oplus \mathcal{V}$ . Entonces:

$$\dim \mathcal{U} \oplus \mathcal{V} = \dim \mathcal{U} + \dim \mathcal{V}.$$

b) el **producto tensorial**  $\mathcal{U} \otimes \mathcal{V}$  es el dual del espacio vectorial  $\mathcal{W}$  de todas las formas bilineales sobre  $\mathcal{U} \times \mathcal{V}$ . Si  $\mathbf{u} \in \mathcal{U}$  y  $\mathbf{v} \in \mathcal{V}$ , definimos:

$$(\mathbf{u} \otimes \mathbf{v})(\varphi) = \varphi(\mathbf{u}, \mathbf{v}), \text{ para todo } \varphi \in \mathcal{W}.$$

Entonces  $\mathbf{u} \otimes \mathbf{v} \in \mathcal{U} \otimes \mathcal{V}$ ; este vector es el *producto (tensorial)* de  $\mathbf{u}$  y  $\mathbf{v}$ . Los vectores  $\mathbf{u}_i \otimes \mathbf{v}_j$  forman una base de  $\mathcal{U} \otimes \mathcal{V}$ . Por lo tanto:

$$\dim \mathcal{U} \otimes \mathcal{V} = \dim \mathcal{U} \cdot \dim \mathcal{V}.$$

## Capítulo 6

# Normalidad

Con la descomposición espectral un operador  $\perp$ -diagonalizable queda disecado en fragmentos (proyecciones) a partir de los cuales puede inferirse su comportamiento global, colmándose así prácticamente todas las aspiraciones relativas al análisis individualizado del operador. Falta, no obstante, demarcar efectivamente la clase de los operadores  $\perp$ -diagonalizables. Por lo pronto sólo sabemos que un tal operador debe conmutar con su adjunto, razón por la cual dirigimos nuestra atención hacia esa propiedad.

### § 40 Operadores normales

De un operador  $T : \mathcal{U} \rightarrow \mathcal{U}$  tal que

$$T^*T = TT^*$$

se dice que es un operador **normal**. Esta condición vincula estrechamente a  $T$  con  $T^*$ , vía la relación de adjunción. Tomemos, por ejemplo,  $\mathbf{v} \in \mathcal{U}$  :

$$\begin{aligned} \|T\mathbf{v}\|^2 &= \langle T\mathbf{v}, T\mathbf{v} \rangle = \langle \mathbf{v}, T^*T\mathbf{v} \rangle \\ &= \langle \mathbf{v}, TT^*\mathbf{v} \rangle = \langle T^*\mathbf{v}, T^*\mathbf{v} \rangle = \|T^*\mathbf{v}\|^2 . \end{aligned}$$

Esto muestra que:

[1] Si  $T$  es un operador normal, entonces, para todo  $\mathbf{v}$  de  $\mathcal{U}$  :

$$\|T\mathbf{v}\| = \|T^*\mathbf{v}\| .$$

De aquí se deduce que

$$T\mathbf{v} = \mathbf{0} \quad \text{si y sólo si} \quad T^*\mathbf{v} = \mathbf{0} ,$$

y entonces:  $\mathbf{v}$  es un 0-autovector de  $T$  si y sólo si  $\mathbf{v}$  es un 0-autovector de  $T^*$ . Esta relación se generaliza con sólo observar que  $(T - \mu)^* = T^* - \bar{\mu}$  y que  $T - \mu$  es normal si y sólo si  $T$  lo es:

[2] Si  $T$  es normal entonces  $\mathbf{v}$  es un  $\mu$ -autovector de  $T$  si y sólo si  $\mathbf{v}$  es un  $\bar{\mu}$  autovector de  $T^*$ .

El conjunto de todos los autovalores de  $T$  es el **espectro** de  $T$  y se denota  $Spec T$ . Con esta notación  $Spec T^* = \overline{Spec T}$ , para  $T$  normal.

Ahora se puede afinar un resultado anterior (ver [2], § 27): supongamos que  $\mathbf{u}$  es un  $\lambda$ -autovector y  $\mathbf{v}$  es un  $\mu$ -autovector de  $T$ , con  $\lambda \neq \mu$ , entonces:

$$\left. \begin{array}{l} \langle T\mathbf{u}, \mathbf{v} \rangle = \langle \lambda\mathbf{u}, \mathbf{v} \rangle = \lambda\langle \mathbf{u}, \mathbf{v} \rangle \\ \parallel \\ \langle \mathbf{u}, T^*\mathbf{v} \rangle = \langle \mathbf{u}, \bar{\mu}\mathbf{v} \rangle = \bar{\mu}\langle \mathbf{u}, \mathbf{v} \rangle \end{array} \right\} \implies \langle \mathbf{u}, \mathbf{v} \rangle = 0 .$$

Hemos demostrado que:

[3] Autovectores correspondientes a autovalores distintos, de un operador normal, son ortogonales.

En otros términos: para autovalores  $\lambda, \mu$  de un operador normal  $T$  se cumple que  $\mathcal{N}_\lambda \perp \mathcal{N}_\mu$ , siempre que  $\lambda \neq \mu$ .

## § 41 El teorema espectral

Consideremos un operador  $T$  en un espacio complejo  $\mathcal{U}$  de dimensión  $n$ . Entonces  $T$  tiene, por lo menos, un autovector  $\mathbf{u}_1$  (ver § 26, nota), y puede suponerse que  $\|\mathbf{u}_1\| = 1$ .

Si  $T$  es normal,  $\mathbf{u}_1$  es un autovector de  $T^*$ , es decir, el subespacio  $\mathcal{N}$  generado por  $\mathbf{u}_1$  es unidimensional y  $T^*$ -invariante. Entonces  $\mathcal{N}^\perp$  tiene dimensión  $n - 1$  y es  $T$  (y  $T^*$ )-invariante. Esto permite considerar a  $T$  como un operador en  $\mathcal{N}^\perp$ , el cual es normal (pues la relación  $T^*T = TT^*$  se mantiene), pudiendo reiterarse el proceso. Esta es la clave para el siguiente teorema fundamental.

**Teorema espectral.** *Todo operador normal en un espacio complejo es ortonormalmente diagonalizable.*

DEMOSTRACIÓN. Después de las consideraciones anteriores sólo falta organizar el argumento inductivo:

1. Si  $\dim \mathcal{U} = 1$  la afirmación es trivial.
2. Supongamos que el enunciado ha sido demostrado para espacios de dimensión  $n - 1$  y tomemos  $\mathcal{U}$  con  $\dim \mathcal{U} = n$ .

Según se ha visto,  $T$  es normal en  $\mathcal{N}^\perp$  y  $\dim \mathcal{N}^\perp = n - 1$ , de modo que existe una base ortonormal  $[\mathbf{u}_2, \dots, \mathbf{u}_n]$  de  $\mathcal{N}^\perp$  formada por autovectores de  $T$ . Entonces  $[\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n]$  es una base ortonormal de  $\mathcal{U}$  formada por autovectores de  $T$ . ♠

## § 42 Operadores autoadjuntos, unitarios y positivos

La analogía entre operadores y números complejos puede usarse como principio heurístico para identificar y estudiar clases importantes de operadores.

El cuerpo  $\mathbb{C}$  tiene varios subconjuntos notables: los reales,  $\mathbb{R}$ ; los complejos de módulo uno,  $\mathbb{U}$ ; los reales positivos (en el sentido de no negativos),  $\mathbb{R}^+$ ; los reales estrictamente positivos,  $\mathbb{R}_0^+$ ; los complejos no nulos,  $\mathbb{C}_0$ .

Los tres primeros se caracterizan así:

- (1)  $z \in \mathbb{R}$  si y sólo si  $\bar{z} = z$ ,
- (2)  $z \in \mathbb{U}$  si y sólo si  $\bar{z}z = 1$ ,

(3)  $z \in \mathbb{R}^+$  si y sólo si  $z = \bar{w}w$  para algún  $w$  de  $\mathbb{C}$  (o también, si y sólo si  $z = x^2$  para algún  $x$  de  $\mathbb{R}$ ).

El cuarto consta de los elementos no nulos de  $\mathbb{R}^+$  y, debido a que, en los números, no ser nulo equivale a tener inverso, el quinto no es otro que el conjunto de los complejos *invertibles*.

En consonancia con tales caracterizaciones se dice que un operador  $T$  es

(1) **autoadjunto** (**simétrico** en el caso real, **hermitiano** en el caso complejo) si  $T^* = T$ ,

(2) **unitario** (**ortogonal**, en el caso real) si  $T^*T = 1$ . Esto equivale a decir que  $T$  es invertible y  $T^{-1} = T^*$ ,

(3) **positivo**, y se escribe  $T \geq 0$ , si existe un operador  $S$  tal que  $S^*S = T$ . Si  $T$  es positivo e invertible se dice que es **estrictamente positivo** y se escribe  $T > 0$ .

Estos operadores son todos normales, es decir,  $\perp$ -diagonalizables y la naturaleza de sus espectros corrobora la analogía que ha inspirado sus definiciones. En efecto, sean  $T$  un operador,  $\lambda$  un autovalor de  $T$  y  $\mathbf{u}$  un  $\lambda$ -autovector. Podemos suponer que  $\|\mathbf{u}\| = 1$ , y así  $\langle T\mathbf{u}, \mathbf{u} \rangle = \langle \lambda\mathbf{u}, \mathbf{u} \rangle = \lambda$ . Con base en las hipótesis 1)  $T^* = T$ , 2)  $T^*T = 1$ , 3)  $T = S^*S$ , pueden efectuarse los siguientes cálculos, respectivamente:

$$1) \lambda = \langle T\mathbf{u}, \mathbf{u} \rangle = \langle \mathbf{u}, T^*\mathbf{u} \rangle = \langle \mathbf{u}, T\mathbf{u} \rangle = \bar{\lambda}.$$

$$2) \langle T\mathbf{u}, T\mathbf{u} \rangle = \langle \mathbf{u}, T^*T\mathbf{u} \rangle = \langle \mathbf{u}, \mathbf{u} \rangle = 1$$

$$3) \lambda = \langle T\mathbf{u}, \mathbf{u} \rangle = \langle S^*S\mathbf{u}, \mathbf{u} \rangle = \langle S\mathbf{u}, S\mathbf{u} \rangle \geq 0 \text{ (si } T \text{ es invertible, } S \text{ es invertible y entonces } S\mathbf{u} \neq \mathbf{0}, \text{ de modo que } \langle S\mathbf{u}, S\mathbf{u} \rangle > 0),$$

los cuales indican que, si  $T$  es autoadjunto (resp. unitario, positivo, estrictamente positivo), entonces sus autovalores son reales (resp. de módulo 1, positivos, estrictamente positivos), esto es,  $\text{Spec } T$  está en  $\mathbb{R}$  (resp.  $\mathbb{U}$ ,  $\mathbb{R}^+$ ,  $\mathbb{R}_0^+$ ). Para operadores normales, la recíproca también es cierta y se demuestra usando la forma espectral.

Si  $T = S^2$ , con  $S$  autoadjunto, entonces  $T \geq 0$ ; si, además,  $S$  es invertible, entonces  $T$  es invertible y, por lo tanto,  $T > 0$ . Recíprocamente, si

$T \geq 0$ , su forma espectral,  $T = \sum \lambda_j P_j$ , permite hallar un  $S$ , autoadjunto, tal que  $S^2 = T$ , a saber

$$S = \sum \sqrt{\lambda_j} P_j . \quad (1)$$

Si  $T > 0$  entonces  $\lambda_j > 0$ , para todo  $j$ , luego  $S$  es invertible. En consecuencia, un operador  $T$  es positivo (resp. estrictamente positivo) si y sólo si  $T = S^2$  para algún  $S$  autoadjunto (resp. autoadjunto e invertible). El operador  $S$  puede escogerse positivo (como se hizo en (1)); en tal caso es único y se llama *la raíz cuadrada*,  $\sqrt{T}$ , de  $T$ . La analogía con los números es notable:  $z \in \mathbb{R}^+$  (resp.  $z \in \mathbb{R}_0^+$ ) si y sólo si  $z = x^2$  para algún  $x$  real (resp. real y distinto de cero); en tal caso existe un único  $x > 0$  tal que  $x^2 = z$ , a saber,  $x = \sqrt{z}$ .

Recordemos también que un operador cualquiera  $T$  es invertible si y sólo si sus autovalores son todos diferentes de cero (i.e. están en  $\mathbb{C}_0$ ).

El carácter de  $T$  se refleja, además, en su **forma cuadrática** asociada  $q(\mathbf{u}) = \langle T\mathbf{u}, \mathbf{u} \rangle$ . En efecto, si  $T$  es autoadjunto (resp. positivo, estrictamente positivo), los mismos cálculos 1) y 3) muestran que  $q(\mathbf{u})$  es real (resp. positiva, positiva definida). Para la recíproca basta tomar  $T$ ,  $\lambda$  y  $\mathbf{u}$  como antes; en efecto, como  $q(\mathbf{u}) = \langle T\mathbf{u}, \mathbf{u} \rangle = \lambda$  se ve que, si  $q(\mathbf{u})$  es real (resp. positiva, positiva definida), entonces los autovalores de  $T$  son reales (resp. positivos, estrictamente positivos), de modo que, si  $T$  es normal, entonces es autoadjunto (resp. positivo, estrictamente positivo).

Para un operador autoadjunto  $T : \mathcal{U} \rightarrow \mathcal{U}$ , el número  $p$  de autovalores mayores que 0 puede considerarse como una medida de su *grado de positividad*. Para el mismo efecto, podría también considerarse el máximo  $m$  de la dimensión de los subespacios en los cuales su forma cuadrática es positiva definida. Por fortuna estos valores coinciden. Para verlo, tómesese una base *ortonormal* de  $\mathcal{U}$  formada por autovectores  $\mathbf{u}_i$ , con autovalores correspondientes  $\lambda_i$ , de los cuales los primeros  $p$  son los positivos:  $\lambda_i > 0$  para  $i = 1, \dots, p$  y  $\lambda_i \leq 0$  para  $i = p + 1, \dots, n$ .

Sea  $\mathcal{M}$  un subespacio de dimensión  $s$  tal que  $q(\mathbf{u}) > 0$  para todo  $\mathbf{u} \in \mathcal{M}$ , siempre que  $\mathbf{u} \neq \mathbf{0}$ . Si fuera  $s > p$  entonces (ver § 15, ejemplo



(4) existiría un  $\mathbf{u} \in \mathcal{M} \cap \langle \mathbf{u}_{p+1}, \dots, \mathbf{u}_n \rangle$ ,  $\mathbf{u} \neq \mathbf{0}$ :

$$\mathbf{u} = \sum_{p+1}^n x_i \mathbf{u}_i \quad , \quad \text{con } x_i \neq 0 \quad , \quad \text{para algún } i.$$

Entonces  $T\mathbf{u} = \sum_{p+1}^n x_i \lambda_i \mathbf{u}_i$  y, por lo tanto:

$$0 < q(\mathbf{u}) = \langle T\mathbf{u}, \mathbf{u} \rangle = \sum_{p+1}^n x_i^2 \lambda_i \leq 0 .$$

Esta contradicción prueba que  $s \leq p$ . Luego  $m \leq p$ .

Por otra parte, si  $\mathbf{u} \in \langle \mathbf{u}_1, \dots, \mathbf{u}_p \rangle$ ,  $\mathbf{u} \neq \mathbf{0}$ , entonces  $q(\mathbf{u}) = \sum_1^p x_i^2 \lambda_i$  es mayor que 0. Esto prueba que  $m \geq p$ . Conclusión:  $m = p$ .

**Nota.** Las nociones y resultados anteriores se trasladan de la manera usual a las matrices. En particular,  $A$  es normal si y sólo si  $A^*A = AA^*$  y es autoadjunta si coincide con  $A^*$ , es decir, con la conjugada de su traspuesta (ver § 37, (4)).

## § 43 Matrices simétricas

De una matriz real autoadjunta, es decir, tal que  $A^t = A$ , se dice que es **simétrica** (ver § 42 y § 37, (5)). Como, en este caso, sus autovalores son reales, el teorema espectral indica que existe una matriz ortogonal  $P$  tal que la matriz  $\Lambda = P^t A P$  es diagonal.

Esto puede aplicarse a la matriz de una forma cuadrática

$$q(x_1, \dots, x_n) = \sum a_{ij} x_i x_j \quad , \quad \text{esto es, } q(\mathbf{x}) = \mathbf{x}^t A \mathbf{x} \quad ,$$

interpretando la ecuación

$$\mathbf{x} = P \mathbf{y}$$

como un cambio de coordenadas ortogonales (de las  $x_i$  a las  $y_i$ , o sea, de la base canónica  $I$  a la base  $P$ ) que transforma a  $q$  en la forma cuadrática

$$q_0(\mathbf{y}) = q(P\mathbf{y}) = \mathbf{y}^t \Lambda \mathbf{y} \quad ,$$

es decir:

$$q_0(y_1, \dots, y_n) = \sum \lambda_i y_i^2.$$

Se obtiene así el llamado **teorema de los ejes principales**:

[4] *Para toda forma cuadrática real  $\sum a_{ij}x_i x_j$  hay un cambio de coordenadas ortogonales que la lleva a la forma  $\sum \lambda_i y_i^2$  (donde los  $\lambda_i$  son los autovalores de  $A$ ).*

Hay también una versión para el caso complejo, es decir, para  $A$  hermitiana.

## § 44 Matrices ortogonales

Una matriz real  $A$  es ortogonal si  $A^t A = I$ , es decir, si  $A$  es invertible y  $A^{-1} = A^t$  (ver § 42); como sus autovalores están en  $\mathbb{U}$ , son de la forma  $\lambda = \mu + i\nu$ , con  $\mu^2 + \nu^2 = 1$ . En particular, pueden ser 1 ó  $-1$  (cuando  $\nu = 0$ ). Si  $\nu \neq 0$  entonces  $\lambda$  no es real y  $\bar{\lambda}$  es también un autovalor.

Por el teorema espectral existe una base ortonormal de  $\mathbb{C}^n$  compuesta por autovectores de  $A$ , la cual será de la forma

$$[\mathbf{p}_1, \dots, \mathbf{p}_r, \mathbf{q}_1, \dots, \mathbf{q}_s, \mathbf{z}_1, \bar{\mathbf{z}}_1, \dots, \mathbf{z}_t, \bar{\mathbf{z}}_t] \quad , \quad r + s + 2t = n \quad ,$$

donde los  $\mathbf{p}_j$  (resp. los  $\mathbf{q}_j$ ) son 1-autovectores (resp.  $(-1)$ -autovectores) y los  $\mathbf{z}_j = \mathbf{x}_j + i\mathbf{y}_j$  son  $\lambda_j$ -autovectores, con  $\lambda_j = \mu_j + i\nu_j$  ( $\nu_j \neq 0$ ). Es posible que  $r$  ó  $s$  ó  $t$  sean nulos.

De acuerdo con el ejemplo (4) de la § 30, el sistema  $[\mathbf{z}_1, \bar{\mathbf{z}}_1, \dots, \mathbf{z}_t, \bar{\mathbf{z}}_t]$  puede sustituirse por el sistema  $[\mathbf{u}_1, \mathbf{v}_1, \dots, \mathbf{u}_t, \mathbf{v}_t]$ , donde  $\mathbf{u}_j = \sqrt{2}\mathbf{x}_j$  y  $\mathbf{v}_j = \sqrt{2}\mathbf{y}_j$ , y por lo tanto (ver § 26, (4)):

$$A\mathbf{u}_j = \mu_j\mathbf{u}_j - \nu_j\mathbf{v}_j \quad \text{y} \quad A\mathbf{v}_j = \nu_j\mathbf{u}_j + \mu_j\mathbf{v}_j. \quad (1)$$

Entonces  $P = [\mathbf{p}_1, \dots, \mathbf{p}_r, \mathbf{q}_1, \dots, \mathbf{q}_s, \mathbf{u}_1, \mathbf{v}_1, \dots, \mathbf{u}_t, \mathbf{v}_t]$  es una base ortonormal de  $\mathbb{R}^n$ . Calculando  $AP$ , y teniendo en cuenta (1) y que  $A\mathbf{p}_j = \mathbf{p}_j$  y  $A\mathbf{q}_j = -\mathbf{q}_j$ , se obtiene:

$$AP = [\mathbf{p}_1, \dots, \mathbf{p}_r, -\mathbf{q}_1, \dots, -\mathbf{q}_s, \mu_1\mathbf{u}_1 - \nu_1\mathbf{v}_1, \nu_1\mathbf{u}_1 + \mu_1\mathbf{v}_1, \dots, \mu_t\mathbf{u}_t - \nu_t\mathbf{v}_t, \nu_t\mathbf{u}_t + \mu_t\mathbf{v}_t].$$

Recordando que la  $j$ -ésima columna de  $P$  es  $P\mathbf{e}_j$  se tiene que  $AP = P\Gamma$ , donde  $\Gamma$  es la matriz en cuya diagonal aparecen los números  $1, \dots, 1$  ( $r$  veces),  $-1, \dots, -1$  ( $s$  veces), luego los  $t$  bloques

$$M_1 = \begin{bmatrix} \mu_1 & \nu_1 \\ -\nu_1 & \mu_1 \end{bmatrix}, \dots, M_t = \begin{bmatrix} \mu_t & \nu_t \\ -\nu_t & \mu_t \end{bmatrix}$$

y ceros en los otros lugares. Para indicar esto escribimos

$$\Gamma = \Delta(1, \dots, 1, -1, \dots, -1, M_1, \dots, M_t).$$

Como  $\mu_j^2 + \nu_j^2 = 1$  puede escogerse  $\theta_j$  tal que  $\cos \theta_j = \mu_j$  y  $\sin \theta_j = -\nu_j$ , y entonces tenemos que:

[5] Para cada matriz ortogonal  $A$  existe una matriz ortogonal  $P$  tal que:

$$P^t A P = \Delta(\underbrace{1, \dots, 1}_r, \underbrace{-1, \dots, -1}_s, M_1, \dots, M_t), \quad r + s + 2t = n,$$

donde

$$M_j = \begin{bmatrix} \cos \theta_j & -\sin \theta_j \\ \sin \theta_j & \cos \theta_j \end{bmatrix}, \quad j = 1, \dots, t.$$

## § 45 Matrices positivas

Si  $A$  es positiva (ver § 42, Nota) entonces la forma cuadrática asociada  $\langle A\mathbf{x}, \mathbf{x} \rangle$  es positiva. La positividad de esta forma se toma a veces como definición. Hemos visto que, para matrices *normales*, esta condición implica la positividad de  $A$  en el sentido aquí adoptado. En realidad,  $A$  es entonces autoadjunta (simétrica, en el caso real).

Resumimos algunas equivalencias útiles, ya obtenidas:

[6] Para una matriz normal  $A$  los enunciados siguientes son equivalentes:

- (i)  $A$  es estrictamente positiva ( $A > 0$ ), es decir,  $A$  es invertible y existe  $B$  tal que  $A = B^*B$ .
- (ii) Los autovalores de  $A$  son mayores que 0.

- (iii) *Existe una matriz  $C$  autoadjunta e invertible tal que  $C^2 = A$  (una raíz cuadrada de  $A$ ).*
- (iv) *La forma cuadrática  $\langle A\mathbf{x}, \mathbf{x} \rangle$  es definida positiva.*

Para la positividad simple ( $A \geq 0$ ) es válido un enunciado similar.

Estos conceptos son útiles en el campo de la optimización ya que, según el criterio basado en la fórmula de Taylor, el carácter de una función de varias variables en un punto crítico (primeras derivadas nulas) depende de la matriz de las segundas derivadas en ese punto. Es entonces muy deseable tener un criterio para la positividad de una matriz  $A$ .

Para llegar a una prueba efectiva de positividad, obsérvese, en primer lugar, que el determinante de una matriz  $A > 0$ , siendo el producto de sus autovalores (ver § 26, (3)), es también mayor que 0.

Si  $A$  es una matriz de orden  $n$  y  $k \leq n$ , la matriz  $A_k$  obtenida suprimiendo las filas y columnas de índice superior a  $k$  es la  $k$ -ésima **submatriz principal** de  $A$ ; su determinante,  $\det A_k$ , es el  $k$ -ésimo **menor principal**.

Para cada  $\mathbf{x}$  de  $\mathbb{R}^k$  (ver § 10, ejemplo (5))

$$\langle A\mathbf{x}_{(n)}, \mathbf{x}_{(n)} \rangle = \langle A_k\mathbf{x}, \mathbf{x} \rangle. \quad (1)$$

Aplicando [6], (iv), se ve que, si  $A > 0$  entonces  $A_k > 0$  y, por la observación anterior,  $\det A_k > 0$ .

El mismo argumento puede generalizarse: sea  $A_K$  la matriz obtenida suprimiendo las filas y columnas cuyos índices están en un subconjunto  $K$  de  $\{1, \dots, n\}$  (caso particular: si  $K = \{k+1, \dots, n\}$ ,  $A_K = A_k$ ); entonces  $A > 0$  implica  $\det A_K > 0$ .

En realidad, estamos en presencia de un criterio de positividad:

- [7] *Una matriz autoadjunta  $A$  es estrictamente positiva si y sólo si todos sus menores principales son mayores que 0.*

Para demostrar la implicación que falta, se procede por inducción. El caso  $n = 1$  es trivial. Supongamos entonces que el criterio es válido para matrices de orden  $n-1$  y supongamos que  $A$  es de orden  $n$  y que  $\det A_k > 0$

---

para todo  $k$  (en particular,  $\det A > 0$ ). Como lo propio ocurre con la matriz  $A_{n-1}$ , la hipótesis implica que  $A_{n-1} > 0$ . La relación (1), con  $k = n - 1$ , muestra que

$$q(\mathbf{x}) = \langle A\mathbf{x}, \mathbf{x} \rangle > 0 ,$$

para todo  $\mathbf{x} \neq \mathbf{0}$ , en el subespacio  $\mathbb{R}_{(n)}^{n-1}$  de  $\mathbb{R}^n$ . Como  $\dim \mathbb{R}_{(n)}^{n-1} = n - 1$ , por lo menos  $n - 1$  de los  $n$  autovalores  $\lambda_1, \dots, \lambda_n$  de  $A$  son mayores que 0 (ver § 42). Pero  $\lambda_1 \dots \lambda_n = \det A > 0$ . Por lo tanto  $\lambda_i > 0$ , para todo  $i$ . Entonces  $A > 0$ .

Para la positividad ( $A \geq 0$ ) hay un criterio similar en el cual, sin embargo, hay que tener en cuenta *todas* las  $A_K$  y no sólo las principales.

## Capítulo 7

# Aplicaciones

La utilidad del álgebra lineal se manifiesta tanto en la misma matemática como fuera de ella. Muchas situaciones, una vez desechados los aspectos incidentales, pueden describirse apropiadamente en términos de espacios vectoriales, bases, transformaciones lineales, ortogonalidad, álgebra matricial, etc.

Este capítulo sólo pretende dar una idea sobre algunas aplicaciones, mostrando que los desarrollos efectuados anteriormente proporcionan herramientas eficaces para resolver diversos problemas y para formular conceptos arrojando nueva luz sobre su significado.

### § 46 Rectas y planos

Si en el plano (o en el espacio) se introduce un sistema ortogonal de coordenadas, cada elemento  $\boldsymbol{x}$  de  $\mathbb{R}^2$  (resp. de  $\mathbb{R}^3$ ) puede interpretarse:

- 1) como un punto  $P$  (de coordenadas  $x_i$ ),
- 2) como un segmento orientado (flecha) que va del origen  $0$  al punto  $P$ .

Es útil convenir que dos segmentos orientados  $\overrightarrow{P_1Q_1}$  y  $\overrightarrow{P_2Q_2}$  representan el mismo vector cuando tienen igual longitud, dirección (están sobre rectas

paralelas) y sentido (el segmento  $\overline{P_1P_2}$  no encuentra a  $\overline{Q_1Q_2}$ ). Las flechas de origen 0 son entonces las representantes canónicas de los vectores. El término *punto* se usa como sinónimo de *vector*.

**Rectas.** Si  $\mathbf{a}$  y  $\mathbf{u}$  son vectores fijos, con  $\mathbf{u} \neq \mathbf{0}$ , el conjunto de los vectores  $\mathbf{x}$  tales que

$$\mathbf{x} = \mathbf{a} + s\mathbf{u}, \quad (s \in \mathbb{R}) \tag{1}$$

es la **recta** que *pasa* por  $\mathbf{a}$  y es *paralela* a  $\mathbf{u}$  (ver fig. 3). Se dice que (1) es la *ecuación* de esta recta.

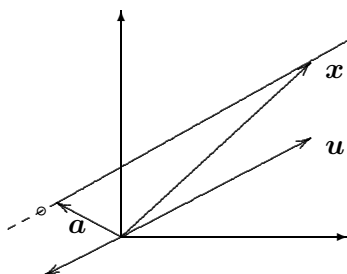


Figura 3

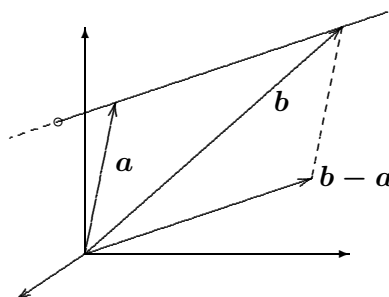


Figura 4

Dados dos puntos  $\mathbf{a}$  y  $\mathbf{b}$ , la recta que pasa por  $\mathbf{a}$  y  $\mathbf{b}$  es la recta que pasa por  $\mathbf{a}$  y es paralela a  $\mathbf{b} - \mathbf{a}$  (ver fig. 4); su ecuación es entonces  $\mathbf{x} = \mathbf{a} + s(\mathbf{b} - \mathbf{a})$  o también,  $\mathbf{x} = (1 - s)\mathbf{a} + s\mathbf{b}$ .

Cuando  $0 \leq s \leq 1$  se obtienen los puntos de esta recta que están *entre*  $\mathbf{a}$  y  $\mathbf{b}$ . Estos constituyen el **segmento**  $\overline{\mathbf{ab}}$  :

$$\mathbf{x} \in \overline{\mathbf{ab}} \iff \mathbf{x} = (1 - s)\mathbf{a} + s\mathbf{b}, \quad (0 \leq s \leq 1).$$

La longitud de  $\overline{\mathbf{ab}}$  es  $\|\mathbf{b} - \mathbf{a}\|$ .

**Planos.** Un plano puede determinarse de dos maneras:

(1a) Dando uno de sus puntos  $\mathbf{a}$ , y dos vectores,  $\mathbf{u}$  y  $\mathbf{v}$ , linealmente independientes y paralelos a él (ver fig. 5); la ecuación en este caso es

$$\mathbf{x} = \mathbf{a} + s\mathbf{u} + t\mathbf{v}, \quad (s, t \in \mathbb{R}).$$

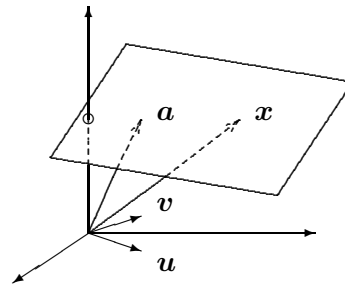


Figura 5

(2a) Dando uno de sus puntos  $\mathbf{a}$  y un vector  $\mathbf{p}$  perpendicular. Un punto  $\mathbf{x}$  estará en el plano si y sólo si  $(\mathbf{x} - \mathbf{a}) \perp \mathbf{p}$  :

$$\langle \mathbf{x} - \mathbf{a}, \mathbf{p} \rangle = 0 ,$$

o también

$$p_1x_1 + p_2x_2 + p_3x_3 = q ,$$

donde  $q = p_1a_1 + p_2a_2 + p_3a_3$ . Se observa que el plano en cuestión no es otra cosa que el conjunto solución de una ecuación lineal. Entonces el conjunto solución de un problema lineal es la intersección de varios planos.

Estas nociones pueden introducirse sin modificación en  $\mathbb{R}^n$ ; en este caso se habla de **hiperplanos**. Por ejemplo: el conjunto solución de un problema lineal es una intersección de hiperplanos.

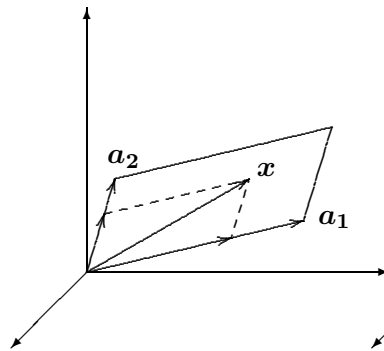


Figura 6

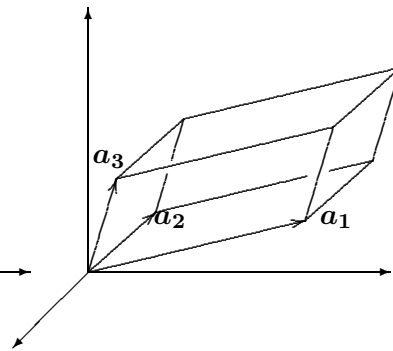


Figura 7



**Paralelepípedos.** Dado un sistema  $A = [\mathbf{a}_1, \dots, \mathbf{a}_k]$  de vectores de  $\mathbb{R}^n$ , con  $1 \leq k \leq n$ , el **paralelepípedo**  $k$ -dimensional determinado por  $A$  es el conjunto *prl*  $A$  formado por los vectores  $\mathbf{x}$  de la forma  $s_1\mathbf{a}_1 + \dots + s_k\mathbf{a}_k$  con  $0 \leq s_i \leq 1$  para todo  $i = 1, \dots, k$ . Para  $k = 1, 2, 3$  (en  $\mathbb{R}^3$ ) esto corresponde a un segmento, un paralelogramo (ver fig. 6) y un paralelepípedo (ver fig. 7), respectivamente.

## § 47 Proyecciones y mínimos cuadrados

Sea  $A$  un sistema de  $n$  vectores de  $\mathbb{K}^m$  y  $\mathcal{R}$  el subespacio generado por  $A$ , es decir, el recorrido de la transformación lineal  $\tilde{A} : \mathbb{K}^n \rightarrow \mathbb{K}^m$ . Dado  $\mathbf{b} \in \mathbb{K}^m$ , el problema lineal

$$A\mathbf{x} = \mathbf{b} \quad (1)$$

es consistente si y sólo si  $\mathbf{b} \in \mathcal{R}$ . En su defecto, es decir, cuando no existe ningún  $\mathbf{x}$  de  $\mathbb{R}^n$  tal que  $A\mathbf{x} = \mathbf{b}$  (i.e. tal que  $\|\mathbf{b} - A\mathbf{x}\| = 0$ ) se busca un  $\mathbf{x}_0$  de  $\mathbb{K}^n$  tal que  $\|\mathbf{b} - A\mathbf{x}_0\|$  sea mínima, esto es (ver § 32), tal que el vector  $A\mathbf{x}_0$  sea la proyección (ortogonal) de  $\mathbf{b}$  sobre el subespacio  $\mathcal{R}$ ; esto equivale a decir que, para todo  $\mathbf{x}$  de  $\mathbb{K}^n$ ,

$$\langle A\mathbf{x}, \mathbf{b} - A\mathbf{x}_0 \rangle = 0 .$$

En otros términos,  $\langle \mathbf{x}, A^*(\mathbf{b} - A\mathbf{x}_0) \rangle = 0$ , para todo  $\mathbf{x}$  de  $\mathbb{K}^n$ , es decir,  $A^*(\mathbf{b} - A\mathbf{x}_0) = 0$ , o también

$$A^*A\mathbf{x}_0 = A^*\mathbf{b} , \quad (2)$$

el cual será, necesariamente, un problema consistente. Habrá solución única cuando  $A^*A$  sea regular:

$$\mathbf{x}_0 = (A^*A)^{-1}A^*\mathbf{b} . \quad (3)$$

En este caso, la proyección  $\mathbf{p}$  de  $\mathbf{b}$  sobre  $\mathcal{R}$  viene dada por

$$\mathbf{p} = A\mathbf{x}_0 = A(A^*A)^{-1}A^*\mathbf{b} , \quad (4)$$

lo cual muestra, en particular, que la matriz de la proyección  $P : \mathbb{K}^m \rightarrow \mathbb{K}^m$  sobre  $\mathcal{R}$  (ver § 35) es  $A(A^*A)^{-1}A^*$ .

El *error*,  $\mathbf{h} = \mathbf{b} - \mathbf{p}$ , es la parte de  $\mathbf{b}$  ortogonal a  $\mathcal{R}$ . Para hallar su magnitud  $\|\mathbf{h}\|$  observamos que, como  $\mathbf{b} = \mathbf{p} + \mathbf{h}$  y  $\mathbf{p} \perp \mathbf{h}$ , entonces

$$\|\mathbf{h}\|^2 = \langle \mathbf{h}, \mathbf{h} \rangle = \langle \mathbf{h}, \mathbf{b} \rangle = \langle \mathbf{b} - \mathbf{p}, \mathbf{b} \rangle = \langle \mathbf{b}, \mathbf{b} \rangle - \langle \mathbf{p}, \mathbf{b} \rangle .$$

Usando (4), la relación de adjunción (§ 37, (RA)) y la fórmula para la inversa (§ 52, 8), se obtiene:

$$\|\mathbf{h}\|^2 = \langle \mathbf{b}, \mathbf{b} \rangle - \frac{1}{\det A^*A} \langle (\text{Cof } A^*A)^t A^* \mathbf{b}, A^* \mathbf{b} \rangle . \quad (5)$$

Es obvia la utilidad de criterios que garanticen la regularidad de  $A^*A$ . Hay uno que se deriva del hecho siguiente:

[1] *El núcleo de  $A^*A$  coincide con el de  $A$  (y, por lo tanto,  $A$  y  $A^*A$  tienen la misma nulidad).*

DEMOSTRACIÓN. (i) Si  $A\mathbf{x} = \mathbf{0}$  entonces  $(A^*A)\mathbf{x} = A^*(A\mathbf{x}) = \mathbf{0}$ .

(ii) Si  $A^*A\mathbf{x} = \mathbf{0}$  entonces  $\langle \mathbf{x}, A^*A\mathbf{x} \rangle = 0$ , es decir,  $\langle A\mathbf{x}, A\mathbf{x} \rangle = 0$ , esto es,  $A\mathbf{x} = \mathbf{0}$ . ♠

Como  $A$  y  $A^*A$  tienen el mismo número de columnas se deduce que

[2] *El rango de  $A$  coincide con el de  $A^*A$ .*

En particular, si  $A$  tiene rango  $n$ , es decir, si las columnas de  $A$  son linealmente independientes, entonces  $A^*A$  es regular, y recíprocamente.

Se dice que (3) es la **solución óptima**, del problema (1), en el sentido de los **mínimos cuadrados**. Si  $A$  es regular  $\mathbf{x}_0$  coincide con la solución usual. Cuando  $A^*A$  no es regular se llama **mínima solución óptima** de (1) a la solución  $\mathbf{x}_0$  de (2) que tiene norma mínima; también en este caso hay una matriz  $A^-$  (que coincide con  $(A^*A)^{-1}A^*$  cuando  $A^*A$  es regular) tal que  $\mathbf{x}_0 = A^- \mathbf{b}$ . Esta matriz se llama la **seudoinversa** de  $A$  (ver § 54, 2).

**Ejemplo.** En el plano cartesiano, hallar la recta  $y = \alpha x + \beta$  que mejor se ajusta a  $n$  puntos dados  $(a_1, b_1), \dots, (a_n, b_n)$ , en el sentido de los *mínimos cuadrados*, equivale a encontrar la solución óptima del problema

$$\alpha a_i + \beta = b_i \quad , \quad i = 1, \dots, n .$$

es decir,  $A\mathbf{x} = \mathbf{b}$ , donde  $A = [\mathbf{a}, \mathbf{1}]$ , con  $\mathbf{a} = [a_1, \dots, a_n]^t$ ,  $\mathbf{1} = [1, \dots, 1]^t$ ,  $\mathbf{x} = [\alpha, \beta]^t$  y  $\mathbf{b} = [b_1, \dots, b_n]^t$ .

## § 48 Ángulos

En  $\mathbb{R}^2$  o en  $\mathbb{R}^3$  se tiene que

$$\langle \mathbf{a}, \mathbf{b} \rangle = \|\mathbf{a}\| \|\mathbf{b}\| \cos \theta ,$$

siendo  $\theta$  el ángulo entre  $\mathbf{a}$  y  $\mathbf{b}$ .

La desigualdad de Schwarz permite *definir* el **ángulo**  $\theta$  entre dos vectores, no nulos,  $\mathbf{a}$  y  $\mathbf{b}$  de  $\mathbb{R}^n$ , mediante la relación

$$\cos \theta = \frac{\langle \mathbf{a}, \mathbf{b} \rangle}{\|\mathbf{a}\| \|\mathbf{b}\|} .$$

La identidad  $\operatorname{sen}^2 \theta + \cos^2 \theta = 1$  conduce a la siguiente ecuación para  $\operatorname{sen} \theta$ :

$$\|\mathbf{a}\|^2 \|\mathbf{b}\|^2 \operatorname{sen}^2 \theta = \|\mathbf{a}\|^2 \|\mathbf{b}\|^2 - \langle \mathbf{a}, \mathbf{b} \rangle^2 . \quad (1)$$

Estas nociones pueden formularse en forma idéntica en cualquier espacio real con producto interno.

## § 49 Volúmenes

El  **$n$ -volumen**,  $|prl A|$ , del paralelepípedo  $n$ -dimensional  $prl A$ , determinado por un sistema  $A$  de  $n$  vectores de  $\mathbb{R}^m$ , se define inductivamente:

- (i) para  $A = [\mathbf{a}]$ , el 1-volumen de  $prl [\mathbf{a}]$  es la norma de  $\mathbf{a}$ , es decir,  $\langle \mathbf{a}, \mathbf{a} \rangle^{1/2}$ ;
- (ii) suponiendo definido el  $n$ -volumen, el  $(n+1)$ -volumen de  $prl B$  (siendo  $B = [\mathbf{b}, \mathbf{a}_1, \dots, \mathbf{a}_n]$ ) es, por definición, el  $n$ -volumen ("área") de la base  $prl A$  (donde  $A = [\mathbf{a}_1, \dots, \mathbf{a}_n]$ ) multiplicado por la altura, es decir, por la

magnitud (norma) de la parte  $\mathbf{h}$ , de  $\mathbf{b}$ , ortogonal al subespacio  $\mathcal{R}$  generado por  $A$  (ver fig. 8):

$$|\text{prl } B| = |\text{prl } A| \|\mathbf{h}\| .$$

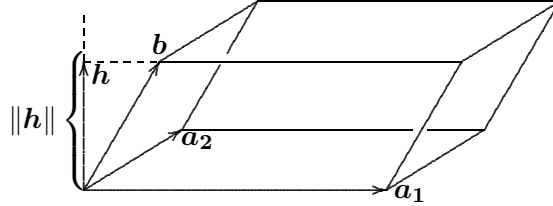


Figura 8

Como era de esperarse, esta noción está relacionada con la de determinante :

[3] Para todo sistema  $A$  en  $\mathbb{R}^m$  se tiene que

$$|\text{prl } A| = (\det A^t A)^{1/2} .$$

(Si  $A$  no es libre, la definición de  $\text{prl } A$  y el criterio [1'] de la § 52, muestran que ambos miembros de la igualdad son nulos. Por ello, y con miras a la demostración, puede suponerse que  $A$  es libre).

DEMOSTRACIÓN. Por inducción: (i) Si  $A = [\mathbf{a}]$  entonces, por definición,  $|\text{prl } A| = \langle \mathbf{a}, \mathbf{a} \rangle^{1/2} = (\mathbf{a}^t \mathbf{a})^{1/2} = (\det \mathbf{a}^t \mathbf{a})^{1/2}$ .

(ii) Tomemos ahora  $B = [\mathbf{b}, \mathbf{a}_1, \dots, \mathbf{a}_k]$ . Efectuando los cálculos pertinentes se comprueba que

$$B^t B = \begin{bmatrix} \lambda & \mathbf{x}^t \\ \mathbf{x} & A^t A \end{bmatrix}, \quad \text{donde } A = [\mathbf{a}_1, \dots, \mathbf{a}_n], \lambda = \langle \mathbf{b}, \mathbf{b} \rangle \text{ y } \mathbf{x} = A^t \mathbf{b} .$$

De acuerdo con la relación (6) de la § 52, tenemos que

$$\det B^t B = \langle \mathbf{b}, \mathbf{b} \rangle \det A^t A - \langle (\text{Cof } A^t A)^t A^t \mathbf{b}, A^t \mathbf{b} \rangle ,$$

lo cual, teniendo en cuenta § 47 (5), muestra que

$$|\text{prl } B|^2 = \|\mathbf{h}\|^2 \det A^t A = \det B^t B . \spadesuit$$

**Ejemplo.** Si  $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$ , el área  $|\text{prl } A|$  del paralelogramo determinado por  $A = [\mathbf{a}, \mathbf{b}]$  viene dada por

$$|\text{prl } A|^2 = \det \begin{vmatrix} \langle \mathbf{a}, \mathbf{a} \rangle & \langle \mathbf{a}, \mathbf{b} \rangle \\ \langle \mathbf{b}, \mathbf{a} \rangle & \langle \mathbf{b}, \mathbf{b} \rangle \end{vmatrix} = \|\mathbf{a}\|^2 \|\mathbf{b}\|^2 - \langle \mathbf{a}, \mathbf{b} \rangle^2 . \quad (1)$$

## § 50 El producto externo en $\mathbb{R}^3$

Cada par  $\mathbf{a}, \mathbf{b}$  de vectores de  $\mathbb{R}^3$  definen (ver [5], § 52) un funcional  $\varphi : \mathbb{R}^3 \rightarrow \mathbb{R}$ , a saber,

$$\varphi(\mathbf{x}) = \det [\mathbf{x}, \mathbf{a}, \mathbf{b}] \quad , \quad \mathbf{x} \in \mathbb{R}^3 ,$$

determinando, por lo tanto, un vector  $\mathbf{v}$  tal que  $\varphi(\mathbf{x}) = \langle \mathbf{x}, \mathbf{v} \rangle$  (ver § 37). Este vector  $\mathbf{v}$  se llama el **producto externo** (o **vectorial**) de  $\mathbf{a}$  y  $\mathbf{b}$ , en este orden, y se denota  $\mathbf{a} \wedge \mathbf{b}$ ; por consiguiente, para todo  $\mathbf{x}$  de  $\mathbb{R}^3$  :

$$\langle \mathbf{x}, \mathbf{a} \wedge \mathbf{b} \rangle = \det [\mathbf{x}, \mathbf{a}, \mathbf{b}] . \quad (1)$$

Las propiedades de la función  $\det$  indican que  $\mathbf{a} \wedge \mathbf{b}$  es ortogonal a los vectores  $\mathbf{a}$  y  $\mathbf{b}$  y que:

$$\mathbf{a} \wedge (\mathbf{b} + \mathbf{c}) = (\mathbf{a} \wedge \mathbf{b}) + (\mathbf{a} \wedge \mathbf{c}) \quad , \quad (\mathbf{a} + \mathbf{b}) \wedge \mathbf{c} = (\mathbf{a} \wedge \mathbf{c}) + (\mathbf{b} \wedge \mathbf{c})$$

$$\lambda(\mathbf{a} \wedge \mathbf{b}) = (\lambda\mathbf{a}) \wedge \mathbf{b} = \mathbf{a} \wedge (\lambda\mathbf{b}) \quad , \quad \mathbf{a} \wedge \mathbf{b} = -\mathbf{b} \wedge \mathbf{a} .$$

Además,  $\mathbf{a} \wedge \mathbf{b} = \mathbf{0}$  si y sólo si  $\mathbf{a}$  y  $\mathbf{b}$  son linealmente dependientes.

Para hallar la longitud de  $\mathbf{a} \wedge \mathbf{b}$  usamos (1) con  $\mathbf{x} = \mathbf{v}$  ( $= \mathbf{a} \wedge \mathbf{b}$ ) obteniendo  $\|\mathbf{a} \wedge \mathbf{b}\|^2 = \det A$ , donde  $A = [\mathbf{v}, \mathbf{a}, \mathbf{b}]$ . Si ahora calculamos  $A^t A$  (teniendo en cuenta que  $\langle \mathbf{v}, \mathbf{a} \rangle = \langle \mathbf{v}, \mathbf{b} \rangle = 0$ ) y luego  $\det A^t A$  (el cual, por otra parte, es igual a  $(\det A)^2$ ), vemos que, si  $\mathbf{a} \wedge \mathbf{b} \neq \mathbf{0}$ , entonces

$$\|\mathbf{a} \wedge \mathbf{b}\|^2 = \|\mathbf{a}\|^2 \|\mathbf{b}\|^2 - \langle \mathbf{a}, \mathbf{b} \rangle^2 .$$

Es decir: la norma de  $\mathbf{a} \wedge \mathbf{b}$  es igual al área del paralelogramo de lados  $\mathbf{a}$  y  $\mathbf{b}$  (ver § 49 (1)). Esta relación también es válida cuando  $\mathbf{a} \wedge \mathbf{b} = \mathbf{0}$  pues, en

este caso,  $\mathbf{a}$  y  $\mathbf{b}$  son linealmente dependientes. Por lo tanto, y de acuerdo con (§ 48, (1)),

$$\|\mathbf{a} \wedge \mathbf{b}\| = \|\mathbf{a}\| \|\mathbf{b}\| \operatorname{sen} \theta \quad , \quad (0 \leq \theta \leq \pi) .$$

Haciendo  $\mathbf{x} = \mathbf{e}_i$  en (1) se ve que las componentes de  $\mathbf{a} \wedge \mathbf{b}$  (ver § 34) son los cofactores (ver § 52, 5) de la primera columna de  $[\mathbf{x}, \mathbf{a}, \mathbf{b}]$ , justificándose así la escritura

$$\mathbf{a} \wedge \mathbf{b} = \det \begin{bmatrix} \mathbf{e}_1 & a^1 & b^1 \\ \mathbf{e}_2 & a^2 & b^2 \\ \mathbf{e}_3 & a^3 & b^3 \end{bmatrix} = \det \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \\ a^1 & a^2 & a^3 \\ b^1 & b^2 & b^3 \end{bmatrix} .$$

## § 51 Aplicaciones a la estadística

Para describir un fenómeno aleatorio con  $n$  resultados elementales posibles,  $1, \dots, n$ , cada uno de los cuales tiene una probabilidad  $P_i > 0$ ,  $\sum P_i = 1$ , se considera el *espacio muestral*  $\Omega = \{1, \dots, n\}$  y la *función de probabilidad*  $P = (P_1, \dots, P_n)$ .

**Variables aleatorias.** Una función  $X : \Omega \rightarrow \mathbb{R}$ , que a cada resultado posible le asigna un valor numérico, se llama una **variable aleatoria** (v.a.). Si  $x_i = X(i)$ ,  $i = 1, \dots, n$ , la v.a.  $X$  puede identificarse con el vector  $[x_1, \dots, x_n]^t$  de modo que las variables aleatorias de  $\Omega$  son, simplemente, los elementos de  $\mathbb{R}^n$ .

Además de las operaciones usuales de  $\mathbb{R}^n$ , se define el producto de dos variables aleatorias  $X = [x_1, \dots, x_n]^t$  y  $Y = [y_1, \dots, y_n]^t$ :

$$XY = [x_1 y_1, \dots, x_n y_n]^t .$$

La variable aleatoria  $[\lambda, \dots, \lambda]^t$  se denota  $\boldsymbol{\lambda}$ .

**Esperanza.** La **esperanza**, valor esperado o media  $EX$  de la v.a.  $X$  es el promedio de sus valores, ponderado con los  $P_i$  como pesos:

$$EX = \sum x_i P_i .$$

La función  $E : \mathbb{R}^n \rightarrow \mathbb{R}$  es un funcional (lineal).

Si  $\mu = EX$  entonces la v.a.

$$\hat{X} = X - \boldsymbol{\mu} = [x_1 - \mu, \dots, x_n - \mu]^t$$

representa la dispersión de los valores de  $X$  con respecto al valor promedio  $\mu$ . Es claro que la media de  $\hat{X}$  es 0.

**Producto interno.** Si  $X$  y  $Y$  son variables aleatorias se define

$$\langle X, Y \rangle = E(XY) = \sum x_i y_i P_i .$$

Este es un producto interno en  $\mathbb{R}^n$  y se obtiene del producto interno usual ponderándolo con los pesos  $P_i$ .

Este producto interno determina una norma  $\| \cdot \|$ :

$$\|X\|^2 = \langle X, X \rangle = \sum x_i^2 P_i .$$

Se dice que  $X$  es una v.a. **normalizada** si  $EX = 0$  y  $\|X\| = 1$ .

**Covarianza, varianza y desviación.** La **covarianza**,  $Cov(X, Y)$ , de dos v.a.  $X$  y  $Y$ , con medias  $\mu$  y  $\nu$ , respectivamente, es el p.i. de  $\hat{X}$  y  $\hat{Y}$  :

$$Cov(X, Y) = \langle \hat{X}, \hat{Y} \rangle = \sum (x_i - \mu)(y_i - \nu) P_i .$$

Al número  $Cov(X, X)$  se le llama la **varianza** de  $X$  y se denota  $Var X$ :

$$Var X = \|\hat{X}\|^2 = \sum (x_i - \mu)^2 P_i .$$

La raíz cuadrada de  $Var X$ , es decir, la norma de  $\hat{X}$ , es la **desviación (estándar)** de  $X$  y se denota  $\sigma_X$ .

La varianza y la desviación miden el *grado de dispersión* de los valores de una v.a. con respecto a su media.

**Nota.** El producto interno y la norma que acaban de definirse son diferentes del producto interno y la norma usuales de  $\mathbb{R}^n$ . Para evitar confusiones puede escribirse  $\mathbf{b}^t \mathbf{a}$  ( $= \mathbf{a}^t \mathbf{b}$ ) para designar el producto interno usual de  $\mathbf{a}$  y  $\mathbf{b}$ .

**Correlación.** Sean  $X$  y  $Y$  dos v.a. con medias  $\mu$  y  $\nu$  respectivamente. El coseno del ángulo entre  $\hat{X}$  y  $\hat{Y}$  (en el sentido del p.i.  $\langle \cdot, \cdot \rangle$ ) se llama **coeficiente de correlación** de  $X$  y  $Y$  y se denota  $\rho(X, Y)$ :

$$\rho(X, Y) = \frac{\langle \hat{X}, \hat{Y} \rangle}{\|\hat{X}\| \|\hat{Y}\|} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{\sum (x_i - \mu)(y_i - \nu)P_i}{(\sum (x_i - \mu)^2 P_i)^{1/2} (\sum (y_i - \nu)^2 P_i)^{1/2}}.$$

Este número indica en qué medida las dos v.a. (vectores) tienen orientaciones similares. Cuando  $\rho(X, Y) = 0$ , es decir, cuando  $X \perp Y$ , se dice que las v.a.  $X$  y  $Y$  son *no correlacionadas*.

**Sistemas de variables aleatorias.** Sea  $\mathcal{X} = [X_1, \dots, X_k]$  un sistema de  $k$  variables aleatorias en  $\Omega$  (ver § 11). Los elementos de  $\langle \mathcal{X} \rangle$  son las combinaciones lineales de las  $X_i$ , es decir, los vectores  $U$  de la forma  $U = \mathcal{X}\mathbf{p}$  con  $\mathbf{p}$  en  $\mathbb{R}^k$ .

La matriz, simétrica y de orden  $k$ ,

$$C = [c_{ij}] \quad \text{donde} \quad c_{ij} = \text{Cov}(X_i, X_j),$$

es la **matriz de covarianza** de  $\mathcal{X}$ .

Si todas las  $X_i$  tienen media 0, entonces  $c_{ij} = \langle X_i, X_j \rangle$ . Si  $U = \mathcal{X}\mathbf{p}$  y  $V = \mathcal{X}\mathbf{q}$ , con  $\mathbf{p} = [\lambda_1, \dots, \lambda_k]^t$  y  $\mathbf{q} = [\mu_1, \dots, \mu_k]^t$ , entonces  $EU = EV = 0$  y, además,

$$\text{Cov}(U, V) = \left\langle \sum_i \lambda_i X_i, \sum_j \mu_j X_j \right\rangle = \sum_{i,j} \lambda_i \mu_j c_{ij} = \mathbf{q}^t C \mathbf{p}.$$

**Componentes principales.** Como la matriz  $C$  es simétrica, el teorema de los ejes principales implica que  $C$  tiene autovalores  $\lambda_1 \geq \dots \geq \lambda_k$ , y autovectores correspondientes  $\mathbf{p}_1, \dots, \mathbf{p}_k$  tales que  $P = [\mathbf{p}_1, \dots, \mathbf{p}_k]$  es ortogonal y

$$P^t C P = \Lambda \quad (= \Delta(\lambda_1, \dots, \lambda_k)).$$

Las variables aleatorias  $Y_i = X\mathbf{p}_i$ ,  $i = 1, \dots, k$ , se llaman las **componentes principales** de  $\mathcal{X}$ .



Como

$$\text{Cov}(Y_i, Y_j) = \mathbf{p}_j^t C \mathbf{p}_i = \begin{cases} 0 & \text{si } i \neq j \\ \lambda_i & \text{si } i = j \end{cases},$$

resulta que las  $Y_i$  son no correlacionadas y la varianza de  $Y_i$  es  $\lambda_i$ ; en particular,  $\lambda_i \geq 0$ , para todo  $i$ , es decir,  $C$  es positiva. Si  $\lambda_i = 0$  entonces  $\|Y_i\|^2 = 0$ , es decir,  $Y_i$  es cero, y recíprocamente.

Si  $Y_1, \dots, Y_r$  son las  $Y_i$  no nulas (i.e. tales que  $\lambda_i > 0$ ), y definimos

$$U_i = \frac{1}{\sqrt{\lambda_i}} Y_i \quad , \quad i = 1, \dots, r,$$

se ve que  $[U_1, \dots, U_r]$  es una base ortonormal del subespacio  $\langle \mathcal{X} \rangle$ .

Sea  $\mathbf{p}$  un vector unitario de  $\mathbb{R}^k$  (i.e.  $\mathbf{p}^t \mathbf{p} = 1$ ). Entonces  $\mathbf{p} = \sum x_i \mathbf{p}_i$  con  $\sum x_i^2 = 1$ , y

$$\text{Var}(\mathcal{X}\mathbf{p}) = \sum x_i^2 \lambda_i.$$

Como  $\sum x_i^2 \lambda_i \leq \sum x_i^2 \lambda_1$ , se deduce que esta varianza es máxima (y vale  $\lambda_1$ ) para  $\mathbf{p} = \mathbf{p}_1$ .

Las componentes principales son entonces combinaciones de las variables aleatorias originales, no correlacionadas entre sí y con varianza grande, en términos de las cuales pueden expresarse todas las otras combinaciones, disminuyendo así el número de variables, y concentrando la atención en los factores que presentan más variabilidad.

**Regresión.** Sea  $\mathcal{X} = [X_1, \dots, X_k]$  un sistema de variables aleatorias y supongamos que se desea conocer la mejor manera (en el sentido de los mínimos cuadrados) de expresar otra variable aleatoria  $Y$  como combinación de las  $X_i$ . Este problema, llamado de **regresión**, consiste en hallar la solución óptima del problema

$$\mathcal{X}\mathbf{p} = Y.$$

Usualmente  $\mathcal{X}$  es linealmente independiente y, por lo tanto,  $\mathcal{X}^t \mathcal{X}$  es invertible. Entonces (ver § 47, (3))

$$\mathbf{p} = (\mathcal{X}^t \mathcal{X})^{-1} \mathcal{X}^t Y.$$

... De allí la función de penitencia, de iniciación,  
que tienen las primeras cien páginas; (...)

¿Qué significa pensar en un lector capaz de superar  
el escollo penitencial de las primeras cien páginas?

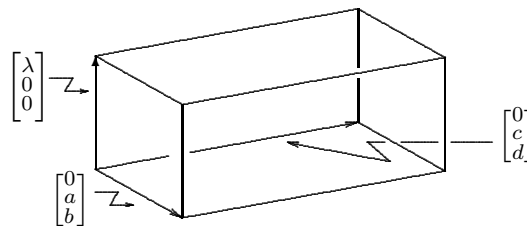
Significa exactamente escribir cien páginas con  
el objeto de construir un lector idóneo para las siguientes.

Umberto Eco, *Apostillas a El Nombre de la Rosa*

# Apéndice

## § 52 Determinantes

Un sistema  $[\mathbf{a}_1, \dots, \mathbf{a}_n]$  de  $n$  vectores de  $\mathbb{K}^n$ , es decir, una matriz  $n \times n$ , puede considerarse como un paralelepípedo  $n$ -dimensional orientado o  $n$ -bloque de “aristas”  $\mathbf{a}_i$  al cual se busca asignarle un *volumen* (orientado) que, en el caso real y para  $n = 1, 2, 3$ , corresponda a la *abscisa* de un punto, al *área* de un paralelogramo orientado y al *volumen* de un paralelepípedo orientado, respectivamente. Queremos, por supuesto, que el volumen de un bloque que tiene a  $B$  como base sea igual a su altura por el área de  $B$ :



$$\text{volumen de } \begin{bmatrix} \lambda & 0 & 0 \\ 0 & a & c \\ 0 & b & d \end{bmatrix} = \lambda \text{ área de } \begin{bmatrix} a & c \\ b & d \end{bmatrix} .$$

En particular, el volumen del  $n$ -cubo unidad  $I = [\mathbf{e}_1, \dots, \mathbf{e}_n]$  será 1 (y no  $-1$ , pues queremos llamar positiva la “orientación” de  $I$ ), y el volumen

del  $n$ -bloque  $[\lambda_1 e_1, \dots, \lambda_n e_n]$ , obtenido magnificando las aristas de  $I$  por factores  $\lambda_1, \dots, \lambda_n$ , respectivamente, será  $\lambda_1 \dots \lambda_n$ .

Si para cada operador  $\sigma$  en  $\mathbb{R}^n$  el cociente

$$d_\sigma = \text{vol } \sigma \langle M \rangle / \text{vol } M \quad , \quad \text{con } \text{vol } M \neq 0 \quad ,$$

no dependiera del  $n$ -bloque  $M$ , entonces  $d_\sigma$  sería un coeficiente de magnificación (de  $\sigma$ ), el cual no sería otra cosa que el volumen de la matriz  $A$  de  $\sigma$ . En efecto:

$$d_\sigma = \text{vol } \sigma \langle I \rangle / \text{vol } I = \text{vol } \sigma \langle I \rangle = \text{vol } A \quad . \quad (*)$$

Si  $\tau$  es otro operador, con matriz  $B$ , entonces

$$d_{\sigma\tau} = \frac{\text{vol } \sigma\tau \langle M \rangle}{\text{vol } M} = \frac{\text{vol } \sigma \langle \tau \langle M \rangle \rangle}{\text{vol } \tau \langle M \rangle} \frac{\text{vol } \tau \langle M \rangle}{\text{vol } M} = d_\sigma d_\tau \quad ,$$

y, como la matriz de  $\sigma\tau$  es  $AB$ , tendríamos, de acuerdo con (\*), que:

$$\text{vol } AB = \text{vol } A \text{ vol } B \quad .$$

La noción de *volumen* (orientado) de un  $n$ -bloque se formaliza mediante la noción de *determinante*.

**Nota.** En lo que sigue  $A$  y  $B$  denotarán matrices  $n \times n$  cualesquiera, a menos que se diga otra cosa.

1. DEFINICIÓN. Se llama **determinante**, de **orden**  $n$ , a una función  $\det_n$  que a cada matriz le asigna un número, de tal manera que:

$$(D.1) \quad \det_{n+1} \begin{bmatrix} \lambda & \mathbf{0} \\ \mathbf{0} & B \end{bmatrix} = \lambda \det_n B \quad , \quad (n = 1, 2, \dots) \quad ,$$

y, además,  $\det_1[\lambda] = \lambda$ .

$$(D.2) \quad \det_n AB = \det_n A \det_n B \quad .$$

**Nota.** Nuestra introducción heurística no garantiza de manera alguna que, para cada  $n$ , exista efectivamente una función que satisfaga (D.1) y (D.2). En principio puede no existir ninguna (o existir varias).

En lo que sigue asumiremos la existencia y unicidad de  $\det_n$ , para cada  $n$  y, como no hay riesgo de confusión, la denotaremos  $\det$ .

2. PROPIEDADES BÁSICAS. La propiedad (D.1) implica que

$$\det \Delta(\lambda_1, \dots, \lambda_n) = \lambda_1 \dots \lambda_n .$$

En particular, como  $E_i(\lambda) = \Delta(1, \dots, \lambda, \dots, 1)$ , entonces

$$\det E_i(\lambda) = \lambda . \tag{1}$$

Aplicando (D.2) a la relación (1) de la § 24, y teniendo en cuenta (1), se ve que  $\det E_{ij}(\lambda) = \det E_{ij}(1)$ , para todo  $\lambda \neq 0$ .

Análogamente, de las relaciones  $E_{ij}(1) E_{ij}(-1) = I$  y  $E_{ij}(1) E_{ij}(1) = E_{ij}(2)$  (ver § 24, (2), (c) ), y teniendo en cuenta que  $\det E_{ij}(2) = \det E_{ij}(1)$ , se deduce primero que  $\det E_{ij}(1) \neq 0$  y luego que  $\det E_{ij}(1) \det E_{ij}(1) = \det E_{ij}(1)$ , concluyéndose que  $\det E_{ij}(1) = 1$ . Por lo tanto:

$$\det E_{ij}(\lambda) = 1 \quad , \quad \text{para todo } \lambda . \tag{2}$$

Calculando  $\det A E_i(\lambda)$  y  $\det A E_{ij}(\lambda)$  se observa que:

[M] *Al multiplicar una columna de A por  $\lambda$  el determinante queda multiplicado por  $\lambda$ .*

En particular: si hay una columna nula el determinante es nulo.

[S] *Si a una columna de A se le suma un múltiplo de otra, el determinante no se altera.*

Por lo tanto: si a una columna se le suma una combinación lineal de las otras columnas el determinante no se altera.

Estas propiedades, aplicadas en la secuencia

$$\begin{aligned} A &= [ \dots, \mathbf{a}, \dots, \mathbf{b}, \dots ] \rightarrow [ \dots, \mathbf{a}, \dots, \mathbf{b} + \mathbf{a}, \dots ] \rightarrow \\ &\quad [ \dots, \mathbf{a} - (\mathbf{b} + \mathbf{a}), \dots, \mathbf{b} + \mathbf{a}, \dots ] \\ &= [ \dots, -\mathbf{b}, \dots, \mathbf{b} + \mathbf{a}, \dots ] \rightarrow [ \dots, -\mathbf{b}, \dots, (\mathbf{b} + \mathbf{a}) - \mathbf{b}, \dots ] \\ &= [ \dots, -\mathbf{b}, \dots, \mathbf{a}, \dots ] \rightarrow [ \dots, \mathbf{b}, \dots, \mathbf{a}, \dots ] , \end{aligned}$$

muestran que

[I] *Si se intercambian dos columnas de  $A$  el determinante cambia de signo.*

Se deduce que: si  $A$  tiene dos columnas iguales entonces  $\det A = 0$ .

Para referirse a la propiedad [I] se dice que la función  $\det$  es **anti-simétrica**. Como caso particular de esta propiedad tenemos que:

$$\det E_{ij} = -1 . \quad (3)$$

De (1), (2), (3) y de [5] (§ 24) se deduce que, para toda matriz elemental  $E$ ,

$$\det E^t = \det E . \quad (4)$$

Las propiedades [I], [M] y [S] muestran que las operaciones elementales (sobre las columnas) no alteran el carácter de nulidad del determinante i.e. si es nulo (resp. no nulo) continúa siendo nulo (resp. no nulo).

**Criterio de libertad.** Si el sistema  $A = [\mathbf{a}_1, \dots, \mathbf{a}_n]$  de  $\mathbb{K}^n$  es linealmente dependiente, es decir, si uno de los  $\mathbf{a}_i$  es c.l. de los restantes, entonces, según [S], dicho  $\mathbf{a}_i$  puede reducirse a  $\mathbf{0}$  sin alterar el determinante y entonces  $\det A = 0$ . Por otra parte, si  $A$  es libre entonces puede transformarse en  $I$  mediante operaciones elementales, y entonces  $\det A \neq 0$ . En resumen:

[1] *Una matriz  $A$  es regular si y sólo si  $\det A \neq 0$ .*

En otras palabras: *el problema  $A\mathbf{x} = \mathbf{0}$  tiene soluciones diferentes de la trivial si y sólo si  $\det A = 0$ .*

Aún más, *un sistema  $A$  de  $n$  vectores en  $\mathbb{K}^m$  es libre si y sólo si  $A^*A$  lo es* (ver [2], § 47). Por lo tanto:

[1'] *Un sistema  $A$  de vectores es libre si y sólo si  $\det A^*A \neq 0$ .*

**Determinante de la traspuesta.** Si  $A$  es regular entonces es producto de matrices elementales (ver § 24, [7]) y entonces (4) implica que

[2] *Para toda matriz  $A$ ,  $\det A^t = \det A$ .*

Cuando  $A$  es singular sus columnas (y también sus filas) son linealmente dependientes y entonces [2] se cumple trivialmente ( $0 = 0$ ).

Del mismo enunciado [2] se deduce que

[3] *Las propiedades del determinante con respecto a las filas son enteramente análogas a las propiedades con respecto a las columnas.*

3. REGLA DE CRAMER. Designaremos  $A\{\mathbf{b}\}_i$  la matriz obtenida poniendo  $\mathbf{b}$  en el lugar de  $\mathbf{a}_i$  (la  $i$ -ésima columna de  $A$ ).

Si  $\mathbf{s} = [s_1, s_2, \dots, s_n]^t$  es solución del problema lineal

$$A\mathbf{x} = \mathbf{b} ,$$

es decir, si  $A\mathbf{s} = \mathbf{b}$ , esto es

$$s_1 \mathbf{a}_1 + s_2 \mathbf{a}_2 + \dots + s_n \mathbf{a}_n = \mathbf{b} ,$$

la secuencia de operaciones

$$\begin{aligned} A &= [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n] \rightarrow [s_1 \mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n] \\ &\rightarrow [s_1 \mathbf{a}_1 + (s_2 \mathbf{a}_2 + \dots + s_n \mathbf{a}_n), \mathbf{a}_2, \dots, \mathbf{a}_n] = A\{\mathbf{b}\}_1 , \end{aligned}$$

muestra que

$$\det A\{\mathbf{b}\}_1 = s_1 \det A .$$

De la misma manera se muestra que, en general,

$$\det A\{\mathbf{b}\}_i = s_i \det A .$$

En consecuencia,

[4] *Si  $\det A \neq 0$ , el problema  $A\mathbf{x} = \mathbf{b}$  tiene una única solución  $\mathbf{s}$  cuyas componentes son*

$$s_i = \det A\{\mathbf{b}\}_i / \det A \quad ; \quad i = 1, 2, \dots, n .$$

4. MULTILINEALIDAD. Conservando fijas las columnas  $\mathbf{a}_2, \dots, \mathbf{a}_n$  puede definirse la función,

$$f_1(\mathbf{x}) = \det [\mathbf{x}, \mathbf{a}_2, \dots, \mathbf{a}_n] .$$

De acuerdo con [M]:

$$f_1(\lambda \mathbf{x}) = \lambda f_1(\mathbf{x}) .$$

En realidad  $f_1$  es lineal. Para verlo supongamos, en primer lugar, que  $[\mathbf{a}_2, \dots, \mathbf{a}_n]$  es libre. Entonces existe  $\mathbf{a}$  tal que  $[\mathbf{a}, \mathbf{a}_2, \dots, \mathbf{a}_n]$  es una base de  $\mathbb{K}^n$ . Dados  $\mathbf{x}, \mathbf{y}$  en  $\mathbb{K}^n$ , puede escribirse

$$\mathbf{x} = \lambda \mathbf{a} + \lambda_2 \mathbf{a}_2 + \dots + \lambda_n \mathbf{a}_n \quad , \quad \mathbf{y} = \mu \mathbf{a} + \mu_2 \mathbf{a}_2 + \dots + \mu_n \mathbf{a}_n .$$

La expresión correspondiente para  $\mathbf{x} + \mathbf{y}$  es entonces

$$\mathbf{x} + \mathbf{y} = (\lambda + \mu) \mathbf{a} + \nu_2 \mathbf{a}_2 + \dots + \nu_n \mathbf{a}_n \quad , \quad \text{donde} \quad \nu_i = \lambda_i + \mu_i .$$

Teniendo en cuenta la expresión de  $\mathbf{x}$ , y usando [S] y luego [M], se obtiene:

$$f_1(\mathbf{x}) = \det [\mathbf{x}, \mathbf{a}_2, \dots, \mathbf{a}_n] = \det [\lambda \mathbf{a}, \mathbf{a}_2, \dots, \mathbf{a}_n] = \lambda c ,$$

siendo  $c = \det [\mathbf{a}, \mathbf{a}_2, \dots, \mathbf{a}_n]$ . Similarmente  $f_1(\mathbf{y}) = \mu c$  y  $f_1(\mathbf{x} + \mathbf{y}) = (\lambda + \mu)c$ . Entonces

$$f_1(\mathbf{x} + \mathbf{y}) = f_1(\mathbf{x}) + f_1(\mathbf{y}) .$$

Cuando  $[\mathbf{a}_2, \dots, \mathbf{a}_n]$  es linealmente dependiente, esta igualdad se cumple trivialmente ( $0 = 0 + 0$ ). Análogamente se definen  $f_2, \dots, f_n$  y se demuestra que son lineales. Este hecho se expresa diciendo que

[5] *La función det es lineal en cada uno de sus  $n$  argumentos.*

Por esto se dice que det es  $n$ -lineal o *multilineal* de orden  $n$ .

5. RELACIÓN INDUCTIVA. El **menor**  $A_j^i$  de  $A$  es la matriz  $(n-1) \times (n-1)$  obtenida suprimiendo la  $i$ -ésima fila y la  $j$ -ésima columna de  $A$ . La matriz intermedia, de forma  $n \times (n-1)$ , obtenida suprimiendo la  $j$ -ésima columna de  $A$ , se denota  $A_j^{()}$ .

El **cofactor**  $\alpha_j^i$  de  $A$  es el determinante de  $A\{e_i\}_j$ . La matriz  $[\alpha_j^i]$  se denota  $\text{Cof } A$ .

Los conceptos anteriores están relacionados: efectuemos los  $j-1$  intercambios de columnas, seguidos de los  $i-1$  intercambios de filas y, finalmente, la sustracción de múltiplos de la 1a. columna a cada una de las otras, indicados en la secuencia siguiente:



$$\begin{aligned}
 A\{\mathbf{e}_i\}_j &= \begin{bmatrix} a_1^1 & \cdots & 0 & \cdots & a_n^1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_1^i & \cdots & 1 & \cdots & a_n^i \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_1^n & \cdots & 0 & \cdots & a_n^n \end{bmatrix} \rightarrow \begin{bmatrix} 0 & \cdots & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \cdots & A_j^{(j)} & \cdots \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & \cdots & \cdots \end{bmatrix} \\
 &\rightarrow \begin{bmatrix} 1 & a_1^i & \cdots & a_n^i \\ 0 & \cdots & A_j^i & \cdots \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & \cdots & \cdots \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & \cdots & A_j^i & \cdots \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & \cdots & \cdots \end{bmatrix}.
 \end{aligned}$$

Recordando la definición de  $\alpha_j^i$ , y teniendo en cuenta (D.1), se puede ver que

$$\alpha_j^i = (-1)^{i+j} \det A_j^i. \tag{5}$$

Fijemos ahora una columna, por ejemplo,  $\mathbf{a}_k$ . Por linealidad en el  $k$ -ésimo argumento (y recordando que  $\mathbf{a}_k = a_k^1 \mathbf{e}_1 + \cdots + a_k^n \mathbf{e}_n$ ) tenemos que

$$\det A = a_k^1 \det A\{\mathbf{e}_1\}_k + \cdots + a_k^n \det A\{\mathbf{e}_n\}_k,$$

es decir:

$$\det A = a_k^1 \alpha_k^1 + \cdots + a_k^n \alpha_k^n.$$

Una relación análoga es válida para filas. Entonces:

**[6]** *El determinante de una matriz es igual a la suma de los productos de los elementos de una línea (fila o columna) por los cofactores correspondientes.*

6. DESARROLLO POR UNA FILA Y UNA COLUMNA. Sea  $M$  una matriz de orden  $n + 1$ , la cual representaremos en la forma

$$M = \begin{bmatrix} \lambda & \mathbf{y}^t \\ \mathbf{x} & A \end{bmatrix}, \text{ donde } \lambda \in \mathbb{K} \ ; \ \mathbf{x}, \mathbf{y} \in \mathbb{K}^n \ \text{ y } \ A \in \mathcal{M}_n,$$

(las filas y columnas de  $M$  se numerarán de 0 hasta  $n$ ). Desarrollando

det  $M$  por la primera fila obtenemos:

$$\det M = \lambda \det A - \sum_{j=1}^n (-1)^{j+1} y^j \det M_j^0,$$

pero  $M_j^0 = [\mathbf{x}, \mathbf{a}_1, \dots, \mathbf{a}_{j-1}, \mathbf{a}_{j+1}, \dots, \mathbf{a}_n]$  y entonces

$$\det M_j^0 = (-1)^{j-1} \det [\mathbf{a}_1, \dots, \mathbf{a}_{j-1}, \mathbf{x}, \mathbf{a}_{j+1}, \dots, \mathbf{a}_n].$$

Desarrollando el último determinante por la  $j$ -ésima columna (en donde los cofactores coinciden con los de  $A$ ), llamando  $C$  a la matriz  $(\text{Cof } A)^t$  (i.e.  $c_j^i = \alpha_i^j$ ) y haciendo las sustituciones del caso, se tiene que

$$\det M = \lambda \det A - \sum (\mathbf{c}^j \mathbf{x}) y^j = \langle C \mathbf{x}, \mathbf{y} \rangle.$$

Este es el desarrollo de  $A$  por la 1a. fila y la 1a. columna. Explícitamente:

$$\det \begin{bmatrix} \lambda & \mathbf{y}^t \\ \mathbf{x} & A \end{bmatrix} = \lambda \det A - \langle (\text{Cof } A)^t \mathbf{x}, \mathbf{y} \rangle. \quad (6)$$

7. ORIENTACIÓN. Se dice que una base  $A = [\mathbf{a}_1, \dots, \mathbf{a}_n]$  de  $\mathbb{R}^n$  está **positivamente** (resp. **negativamente**) **orientada** si  $\det A > 0$  (resp.  $\det A < 0$ ). La base canónica tiene orientación positiva.

En  $\mathbb{R}^3$  una base  $[\mathbf{a}, \mathbf{b}, \mathbf{c}]$  tiene orientación positiva si  $\mathbf{c}$  señala en el sentido de avance de un tornillo derecho que gire de  $\mathbf{a}$  hacia  $\mathbf{b}$ .

8. FÓRMULA PARA  $A^{-1}$ . Si  $A$  es regular y  $B = A^{-1}$  entonces (ver §22), para todo  $k = 1, \dots, n$ :

$$A \mathbf{b}_k = \mathbf{e}_k,$$

es decir,  $\mathbf{b}_k$  es la solución del problema  $A \mathbf{x} = \mathbf{e}_k$  y entonces, por la regla de Cramer,

$$b_k^i = \det A \{ \mathbf{e}_k \}_i / \det A = \alpha_i^k / \det A.$$

En otros términos:

$$A^{-1} = \frac{1}{\det A} (\text{Cof } A)^t,$$

lo cual es equivalente a  $A(\text{Cof } A)^t = (\det A)I$ .

9. EXISTENCIA Y UNICIDAD. La relación inductiva [6] permite expresar  $\det_n$  en términos de  $\det_{n-1}$ . Como para  $n = 1$  necesariamente

$$\det [a] = a ,$$

se concluye que  $\det_n$  es *única*.

En realidad puede demostrarse que  $\det A$  es un polinomio en las componentes de  $A$ ; explícitamente:

$$\det A = \sum \pm a_{i_1}^1 \cdots a_{i_n}^n , \quad (7)$$

donde  $(i_1, \dots, i_n)$  es una permutación de  $(1, \dots, n)$  y el signo está dado por  $\det [e_{i_1}, \dots, e_{i_n}]$ .

La existencia puede demostrarse usando la misma relación para definir inductivamente  $\det_n$  y mostrar que satisface las propiedades (D.1) y (D.2).

## § 53 La forma de Jordan

Con el propósito de diagonalizar un operador  $T$  de  $\mathcal{U}$  se construye, para cada autovalor  $\lambda$ , una base del subespacio propio (§ 26) correspondiente. La reunión de tales bases es un sistema libre, de autovectores de  $T$ , tal que cualquier otro autovector es combinación lineal del sistema. El operador  $T$  es diagonalizable si y sólo si dicho sistema es una *base* de  $\mathcal{U}$ ; en el caso de una matriz  $A$ , el sistema constituye una matriz invertible  $B$  tal que  $B^{-1}AB$  es diagonal. Cuando  $A$  no es diagonalizable, el sistema obtenido no es generador, pero ¿podría ser completado para obtener  $B$  tal que  $B^{-1}AB$  sea aproximadamente diagonal?

Para examinar el asunto, supongamos que  $A$  es una matriz *no* diagonalizable, que tiene un solo autovalor  $\lambda$  (con multiplicidad algebraica *tres*) y que el subespacio propio correspondiente es unidimensional, con base  $[\mathbf{u}_1]$ . Queremos tener otros *dos* vectores,  $\mathbf{u}_1$  y  $\mathbf{u}_2$ , tales que  $B = [\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3]$  sea

una base y la matriz  $J = B^{-1}AB$  tenga la siguiente forma “casi” diagonal

$$J = \begin{bmatrix} \lambda & a & 0 \\ 0 & \lambda & b \\ 0 & 0 & \lambda \end{bmatrix},$$

en la cual, como sucedería si  $A$  fuese diagonalizable,  $\lambda$  aparece repetido *tres* veces en la diagonal. Teniendo en cuenta que (cfr. § 27)

$$\begin{aligned} AB &= A[\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3] = [A\mathbf{u}_1, A\mathbf{u}_2, A\mathbf{u}_3], \text{ y} \\ BJ &= B[\lambda\mathbf{e}_1, \lambda\mathbf{e}_2 + a\mathbf{e}_1, \lambda\mathbf{e}_3 + b\mathbf{e}_2] \\ &= [\lambda\mathbf{u}_1, \lambda\mathbf{u}_2 + a\mathbf{u}_1, \lambda\mathbf{u}_3 + b\mathbf{u}_2], \end{aligned}$$

se ve que la condición  $B^{-1}AB = J$ , es decir,  $AB = BJ$ , es equivalente a las siguientes:

1)  $A\mathbf{u}_1 = \lambda\mathbf{u}_1$ . Esto significa que  $\mathbf{u}_1$  es, como hemos dicho, un  $\lambda$ -autovector de  $A$ .

2)  $A\mathbf{u}_2 = \lambda\mathbf{u}_2 + a\mathbf{u}_1$ . Nótese que  $a$  no puede ser nulo (si lo fuera,  $\mathbf{u}_2$  sería otro autovector, independiente de  $\mathbf{u}_1$ ) y entonces podemos cambiar  $\mathbf{u}_2$  por  $a\mathbf{u}_2$ :  $A\mathbf{u}_2 = \lambda\mathbf{u}_2 + \mathbf{u}_1$ .

3) Como en el caso anterior, la condición queda así:  $A\mathbf{u}_3 = \lambda\mathbf{u}_3 + \mathbf{u}_2$ .

Si existen vectores  $\mathbf{u}_i$  que satisfagan estas condiciones, ellos forman una base que reduce la matriz  $A$  a la forma  $J$ , con  $a = b = 1$ .

1. BASES DE JORDAN. Las observaciones anteriores sugieren que, en el caso general, se busquen cadenas

$$\mathbf{u}_1, \dots, \mathbf{u}_m,$$

con  $m \geq 1$ , tales que:

1)  $T\mathbf{u}_1 = \lambda\mathbf{u}_1$  ( $\iff (T - \lambda)\mathbf{u}_1 = \mathbf{0}$ ), i.e.  $\mathbf{u}_1$  es un autovector, con valor propio  $\lambda$ .

2)  $T\mathbf{u}_j = \lambda\mathbf{u}_j + \mathbf{u}_{j-1}$  ( $\iff (T - \lambda)\mathbf{u}_j = \mathbf{u}_{j-1}$ ), para  $j = 2, \dots, m$ .

Diremos, en este caso, que  $\mathbf{u}_1, \dots, \mathbf{u}_m$  es una  $\lambda$ -**cadena de Jordan** para  $T$ . Una base constituida por tales cadenas es una **base de Jordan** (de  $\mathcal{U}$ ) para  $T$ . Vamos a ver que todo operador tiene una base de Jordan.

La afirmación es trivial en espacios de dimensión 1. Por otra parte hay, según veremos, una forma de extender una base de Jordan de  $\mathcal{R}_T$  para obtener una base de Jordan de  $\mathcal{U}$  (recuérdese que  $T$  puede ser considerado como un operador en  $\mathcal{R}_T$ , pues éste es  $T$ -invariante). Por consiguiente podemos proceder por inducción sobre  $n (= \dim \mathcal{U})$ . Para organizar el argumento necesitamos que  $m = \dim \mathcal{R}_T < n$  ( $\iff p = \dim \mathcal{N}_T \geq 1 \iff \mathcal{N}_T \neq \{\mathbf{0}\} \iff 0$  es un autovalor de  $T$ ). Pero esto puede suponerse pues  $T$  debe tener algún autovalor  $\lambda_0$  y, como toda base de Jordan para  $T - \lambda_0$  lo es también para  $T$  (toda  $\lambda$ -cadena para  $T - \lambda_0$  es una  $(\lambda + \lambda_0)$ -cadena para  $T$ ), el argumento se aplicaría a  $T - \lambda_0$ .

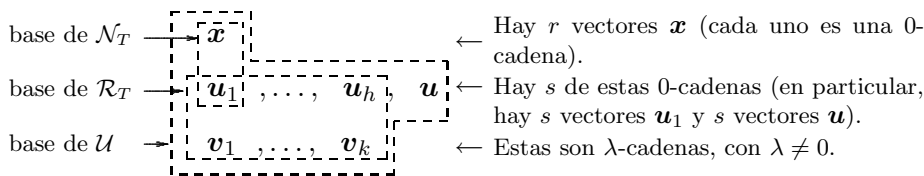
Supongamos entonces que  $m < n$  y que (hipótesis de inducción)  $\mathcal{R}_T$  tiene una base de Jordan para  $T$ . En esta base hay, en general, dos tipos de cadenas:

$$\begin{aligned} & \mathbf{u}_1, \dots, \mathbf{u}_h, & (m \text{ vectores, en total}) & (1) \\ & \mathbf{v}_1, \dots, \mathbf{v}_k. \end{aligned}$$

Las del primer tipo (digamos que hay  $s$  de ellas) son 0-cadenas (nótese que  $h$  no es fijo). Las del segundo tipo son  $\lambda$ -cadenas, con  $\lambda \neq 0$  (nótese que ni  $\lambda$  ni  $k$  son fijos). Entre los  $\mathbf{u}_i$  y los  $\mathbf{v}_j$  hay, en total,  $m$  elementos.

Los  $s$  vectores  $\mathbf{u}_1$ , siendo 0-autovectores, están en  $\mathcal{N}_T$  y, como son linealmente independientes, pueden completarse con  $r$  vectores, designados genéricamente con  $\mathbf{x}$ , para formar una base de  $\mathcal{N}_T$  (nótese que  $r + s = p$ ). Además, como cada  $\mathbf{u}_h$  está en  $\mathcal{R}_T$ , existe  $\mathbf{u}$ , en  $\mathcal{U}$ , tal que  $T\mathbf{u} = \mathbf{u}_h$ .

Afirmamos que la colección de vectores denotada abreviadamente por



es una base de Jordan, de  $\mathcal{U}$ , para  $T$ .

Observemos, en primer lugar, que a los  $m$  vectores originales se agregaron  $r$  vectores  $\mathbf{x}$  y  $s$  vectores  $\mathbf{u}$ . Total:  $m + r + s = m + p = n (= \dim \mathcal{U})$ .

En consecuencia, basta probar que son linealmente independientes. Para ello, supongamos que

$$\begin{aligned} \xi \mathbf{x} + \alpha_1 \mathbf{u}_1 + \alpha_2 \mathbf{u}_2 + \cdots + \alpha_{h-1} \mathbf{u}_{h-1} + \alpha_h \mathbf{u}_h + \alpha \mathbf{u} \\ + \beta_1 \mathbf{v}_1 + \beta_2 \mathbf{v}_2 + \cdots + \beta_{k-1} \mathbf{v}_{k-1} + \beta_k \mathbf{v}_k = \mathbf{0}, \end{aligned}$$

(recuérdese que  $\xi \mathbf{x}$  es algo de la forma  $\xi_1 \mathbf{x}_1 + \cdots + \xi_r \mathbf{x}_r$ , etc).

Aplicando  $T$  y reuniendo algunos términos se obtiene:

$$\begin{aligned} \mathbf{0} + \mathbf{0} + \alpha_2 \mathbf{u}_1 + \cdots + \alpha_{h-1} \mathbf{u}_{h-2} + \alpha_h \mathbf{u}_{h-1} + \alpha \mathbf{u}_h \\ + (\beta_1 \lambda + \beta_2) \mathbf{v}_1 + (\beta_2 \lambda + \beta_3) \mathbf{v}_2 + \cdots + (\beta_{k-1} \lambda + \beta_k) \mathbf{v}_{k-1} + \beta_k \lambda \mathbf{v}_k = \mathbf{0}. \end{aligned}$$

Como (1) es linealmente independiente se tiene, primero:  $\alpha_2 = \cdots = \alpha_h = \alpha = 0$ ; y luego (debido a que  $\lambda \neq 0$ ):  $\beta_k = 0$  ( $\implies \beta_{k-1} = 0 \implies \cdots \implies \beta_1 = 0$ ). Por lo tanto

$$\xi \mathbf{x} + \alpha_1 \mathbf{u}_1 = \mathbf{0} .$$

Pero, como los vectores  $\mathbf{x}$  junto con los vectores  $\mathbf{u}_1$  forman una base de  $\mathcal{N}_T$ , entonces también los escalares  $\xi$  y los escalares  $\alpha_1$  son nulos. Así concluye la demostración de existencia.

2. UNICIDAD. Demostraremos que el número de  $\lambda$ -cadenas de longitud dada está determinado por las nulidades de  $(T - \lambda)^k$ ,  $k = 1, 2, \dots$ . Como las  $\lambda$ -cadenas de  $T$  son 0-cadenas de  $T - \lambda$ , basta ver que el número  $\zeta_j$  de 0-cadenas de longitud  $j$  está determinado por los números  $\nu_k = \text{dimensión del núcleo } \mathcal{N}_k \text{ de } T^k$ . Nótese que  $\mu_k = n - \nu_k$  es la dimensión del recorrido  $\mathcal{R}_k$  de  $T^k$ .

Dada una base de Jordan para  $T$ , con  $n$  ( $= \text{dim } \mathcal{U}$ ) elementos, sea  $\mathcal{N}$  (resp.  $\mathcal{M}$ ) el subespacio generado por las 0-cadenas (resp.  $\lambda$ -cadenas, con  $\lambda \neq 0$ ). Entonces  $\mathcal{N}$  y  $\mathcal{M}$  son  $T$ -invariantes y  $\mathcal{U} = \mathcal{N} \oplus \mathcal{M}$ . Además  $T^k$  es invertible en  $\mathcal{M}$ , de tal modo que  $\mathcal{N}_k \subset \mathcal{N}$  y  $\mathcal{M} \subset \mathcal{R}_k$ , para todo  $k$ .

Formemos dos sistemas de vectores,  $\mathcal{A}_k$  y  $\mathcal{B}_k$ , así: i) Sea  $\mathbf{u}_1, \dots, \mathbf{u}_m$  una 0-cadena; si  $m \leq k$ , a todos los  $\mathbf{u}_i$  los ponemos en  $\mathcal{A}_k$ ; si  $m > k$  entonces  $\mathbf{u}_1, \dots, \mathbf{u}_k$  estarán en  $\mathcal{A}_k$  y  $\mathbf{u}_{k+1}, \dots, \mathbf{u}_m$  estarán en  $\mathcal{B}_k$ . ii) A todos los elementos de las  $\lambda$ -cadenas (con  $\lambda \neq 0$ ) los ponemos en  $\mathcal{B}_k$ .

Obsérvese que si  $n_k$  (resp.  $m_k$ ) es el número de elementos de  $\mathcal{A}_k$  (resp.

$\mathcal{B}_k$ ), entonces  $n_k + m_k = n$ . Además:

$$n_k = \zeta_1 + 2\zeta_2 + \cdots + k\zeta_k + k\zeta_{k+1} + \cdots + k\zeta_n, \quad k = 1, \dots, n.$$

También puede verse que  $\mathcal{A}_k \subset \mathcal{N}_k$  y  $\mathcal{B}_k \subset \mathcal{R}_k$ . Por lo tanto  $n_k \leq \nu_k$  y  $m_k \leq \mu_k$ , y entonces

$$n = n_k + m_k \leq \nu_k + \mu_k = n,$$

concluyéndose que  $n_k = \nu_k$ . Entonces los  $\zeta_j$  satisfacen las ecuaciones

$$\zeta_1 + 2\zeta_2 + \cdots + k\zeta_k + k\zeta_{k+1} + \cdots + k\zeta_n = \nu_k, \quad k = 1, \dots, n.$$

Si de cada ecuación, empezando por la última, se sustrae la anterior y, luego, de cada ecuación, empezando por la primera, se sustrae la siguiente, se ve rápidamente que

$$\zeta_k = -\nu_{k-1} + 2\nu_k - \nu_{k+1}, \quad k = 1, \dots, n,$$

donde hay que tomar  $\nu_0 = 0$  y  $\nu_{n+1} = \nu_n$ . Esto prueba que los  $\zeta_k$  no dependen de la base.

3. LA FORMA DE JORDAN. En la matriz  $J$  de un operador con respecto a una base de Jordan, cada  $\lambda_j$ -cadena de longitud  $k$  da origen al bloque  $k \times k$

$$J_k(\lambda_j) = \begin{bmatrix} \lambda_j & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_j \end{bmatrix}.$$

Puede haber varios bloques con el mismo  $\lambda_j$ . La matriz  $J$  consta de bloques como éste situados a lo largo de la diagonal mientras que en los demás lugares hay ceros. Esta es la **forma de Jordan**, la cual es única, salvo por el orden en que aparecen los bloques. Desarrollando  $\det(J - \lambda)$ , repetidamente, por la primera columna, se ve que los elementos que aparecen en la diagonal son los autovalores, repetidos según su multiplicidad algebraica.

## § 54 Soluciones óptimas

1. PROBLEMAS LINEALES GENERALES. Un **problema lineal** es una ecuación de la forma

$$T\mathbf{u} = \mathbf{v} , \quad (\text{C})$$

donde  $T : \mathcal{U} \longrightarrow \mathcal{V}$  es una transformación lineal y  $\mathbf{v}$  es un vector fijo de  $\mathcal{V}$ . Una **solución** de (C) es cualquier vector  $\mathbf{u}$  de  $\mathcal{U}$  que satisface la ecuación; las soluciones de (C) constituyen su **conjunto solución** (o **solución general**)  $\mathcal{S}$ . El problema es **consistente** si tiene, por lo menos, una solución; esto ocurre si y sólo si  $\mathbf{v} \in \mathcal{R}_T$ . En caso contrario se dice que es **inconsistente**.

Cuando  $\mathbf{v} = \mathbf{0}$  el problema es **homogéneo**:

$$T\mathbf{u} = \mathbf{0} . \quad (\text{H})$$

Por contraposición, (C) es un problema **completo**. Un problema homogéneo es siempre consistente: su conjunto solución es el núcleo  $\mathcal{N}$  de  $T$ , y éste, siendo un subespacio, contiene siempre al vector  $\mathbf{0}$  (la **solución trivial**). Si (H) tiene una solución  $\mathbf{u}_0 \neq \mathbf{0}$  entonces tiene *infinitas* (pues  $\lambda\mathbf{u}_0 \in \mathcal{N}$ , para todo  $\lambda$ ).

Si  $\mathbf{u}_*$  es una solución de (C), entonces

$$\mathcal{S} = \mathcal{N} + \mathbf{u}_* \quad (1)$$

lo cual quiere decir que las soluciones de (C) son precisamente los vectores de la forma  $\mathbf{u} + \mathbf{u}_*$ , con  $\mathbf{u}$  en  $\mathcal{N}$ . En otros términos:  $\mathcal{S}$  es la variedad lineal obtenida trasladando  $\mathcal{N}$  por medio de una solución particular de (C).

Cuando el problema (C) es *consistente* sólo caben *dos* alternativas: o bien tiene *una sola* solución, o bien tiene *infinitas*, y ello ocurre según que el problema homogéneo (H) tenga sólo la solución trivial o tenga soluciones diferentes de la trivial, respectivamente.

**Nota.** Sea  $T : \mathcal{U} \longrightarrow \mathcal{V}$  una t.l., *sobreyectiva*, cuyo núcleo es el subespacio  $\mathcal{N}$  de  $\mathcal{U}$ . Cada  $\mathbf{v}$  de  $\mathcal{V}$  determina una variedad lineal de  $\mathcal{U}$ , a saber, la solución general de  $T\mathbf{u} = \mathbf{v}$ . El espacio  $\mathcal{U}$  se fracciona en variedades



disyuntas dos a dos; una por cada elemento de  $\mathcal{V}$ . Queda establecida así una biyección entre  $\mathcal{V}$  y el conjunto  $\mathcal{U}/\mathcal{N}$  de todos los subconjuntos de  $\mathcal{U}$  de la forma  $\mathcal{N} + \mathbf{u}$ , por medio de la cual se puede trasladar a  $\mathcal{U}/\mathcal{N}$  toda la estructura de  $\mathcal{V}$  y definir una t.l.  $\Phi : \mathcal{U} \rightarrow \mathcal{U}/\mathcal{N}$ ,  $\Phi(\mathbf{u}) = \mathcal{N} + \mathbf{u}$ , cuyo núcleo es  $\mathcal{N}$ . Una vez reproducida, de esta manera, la situación original, se puede prescindir de  $\mathcal{V}$  y de  $T$ . El espacio vectorial  $\mathcal{U}/\mathcal{N}$  es el **cociente** de  $\mathcal{V}$  por  $\mathcal{N}$  (se obtiene *dividiendo* a  $\mathcal{U}$  en variedades “paralelas” a  $\mathcal{N}$ ) y  $\Phi$  es su transformación **canónica**. Puede decirse que  $\mathcal{U}/\mathcal{N}$  encarna la apariencia que toma  $\mathcal{U}$  cuando nos negamos a distinguir entre elementos de  $\mathcal{U}$  cuya diferencia es sólo un elemento de  $\mathcal{N}$ . Cuando  $\mathcal{U}$  es, por ejemplo, el plano y  $\mathcal{N}$  es una recta (por el origen),  $\mathcal{U}/\mathcal{N}$  es el plano, pero no como conjunto de puntos sino como conjunto de rectas (las paralelas a  $\mathcal{N}$ ).

2. INVERSA GENERALIZADA. Supongamos que el problema  $T\mathbf{u} = \mathbf{v}$  es inconsistente. Ante la imposibilidad de hallar un elemento de  $\mathcal{R}_T$  que sea igual a  $\mathbf{v}$ , nos preguntamos si, en presencia de un producto interno en  $\mathcal{V}$ , existe un elemento de  $\mathcal{R}_T$  cuya distancia a  $\mathbf{v}$  sea lo más pequeña posible. Se trata entonces de buscar un elemento  $\mathbf{u}$  de  $\mathcal{U}$  que minimice a  $\|T\mathbf{u} - \mathbf{v}\|$ . Es lo que se llama una **solución** aproximada, **óptima** en el sentido de los **mínimos cuadrados** (minimizar  $\| \ \|$ , es minimizar  $\| \ \|^2$  y  $\| \ \|^2$  es, en  $\mathbb{R}^n$ , una suma de cuadrados). Para hallar una de tales soluciones basta tomar  $\mathbf{w} =$  la proyección ortogonal de  $\mathbf{v}$  sobre  $\mathcal{R}_T$  (éste es el elemento de  $\mathcal{R}_T$  más próximo a  $\mathbf{v}$ ; ver § 32) y resolver el problema

$$T\mathbf{u} = \mathbf{w} ,$$

el cual es consistente, puesto que  $\mathbf{w} \in \mathcal{R}_T$ .

Supongamos ahora que el problema  $T\mathbf{u} = \mathbf{v}$  es consistente pero que su solución *no* es única (i.e. existen infinitas soluciones). Nos preguntamos si, en presencia de un producto interno en  $\mathcal{U}$ , existe una **solución mínima**. La respuesta es *sí*: el conjunto solución, siendo una variedad lineal, tiene un *único* vector de norma mínima (ver § 35, ejemplo). Se comprueba sin dificultad que la función que a cada  $\mathbf{w}$  de  $\mathcal{R}_T$  le asigna la solución mínima de  $T\mathbf{u} = \mathbf{w}$  es una transformación lineal  $T^\circ : \mathcal{R}_T \rightarrow \mathcal{U}$ .

Suponiendo que  $\mathcal{U}$  y  $\mathcal{V}$  son espacios con producto interno, pueden juntarse las dos observaciones anteriores: para *cada*  $\mathbf{v}$  de  $\mathcal{V}$  existe una *mínima* solución *óptima*, y una sola, del problema lineal  $T\mathbf{u} = \mathbf{v}$ . Para hallarla

se toma primero la proyección ortogonal  $\mathbf{w}$  de  $\mathbf{v}$  sobre  $\mathcal{R}_T$ . Se halla una solución particular  $\mathbf{u}_0$  del problema (consistente)  $T\mathbf{u} = \mathbf{w}$ . La proyección de  $\mathbf{u}_0$  sobre el ortogonal del núcleo de  $T$  es el  $\mathbf{u}$  buscado. Es claro que  $\mathbf{u} = T^\circ \mathbf{w}$ , de modo que, si  $P$  es la proyección (ortogonal) sobre  $\mathcal{R}_T$ , entonces  $\mathbf{u} = T^\circ P\mathbf{v}$ . La transformación lineal

$$T^- = T^\circ P : \mathcal{V} \longrightarrow \mathcal{U}$$

es la **seudoinversa** (o *inversa generalizada* de Moore–Penrose) de  $T$ .

Si  $A$  es una matriz  $m \times n$  y  $T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$  es su transformación lineal asociada, entonces la matriz  $A^-$  de  $T^- : \mathbb{R}^m \longrightarrow \mathbb{R}^n$  es la *seudoinversa* de  $A$ . Para cada  $\mathbf{b} \in \mathbb{R}^m$  existe un único vector  $\mathbf{x}$ , de norma mínima, que minimiza a  $\|A\mathbf{x} - \mathbf{b}\|$ , a saber,  $\mathbf{x} = A^-\mathbf{b}$ .



## Símbolos y abreviaturas

$\implies$	implica
$\iff$	si y sólo si
♠	fin de la demostración
t.l.	transformación lineal
c.l.	combinación lineal
l.d.	linealmente dependiente
l.i.	linealmente independiente, libre
p.i.	producto interno
$\perp$	ortogonal, ortogonalmente
v.a.	variable aleatoria
RA	relación de adjunción
$\mathbb{R}$	números reales
$\mathbb{R}^+$	números reales no negativos
$\mathbb{R}_0^+$	números reales (estrictamente) positivos
$\mathbb{C}$	números complejos
$\mathbb{C}_0$	números complejos no nulos
$\mathbb{U}$	números complejos de módulo 1
$\mathbb{K}$	números reales o números complejos
$x^i$	$i$ -ésima componente del vector columna $\mathbf{x}$
$a_j$	$j$ -ésima componente del vector fila $\mathbf{a}$
$\mathbf{a}^i$	$i$ -ésima fila de la matriz $A$
$\mathbf{a}_j$	$j$ -ésima columna de la matriz $A$
$a_j^i$	$i$ -ésima componente del vector columna $\mathbf{a}_j$
	$j$ -ésima componente del vector fila $\mathbf{a}^i$
	$(i, j)$ -componente de la matriz $A$
$[a_j^i]$	matriz de componentes $a_j^i$
$\delta_j^i$	delta de Kronecker ( $\delta_i^i = 1$ y $\delta_j^i = 0$ si $i \neq j$ )

$\langle \cdot, \cdot \rangle$ , 3, 60, 93	$[f_i \leftrightarrow f_j]$ , $[cf_i]$ , $[f_i + cf_j]$ , 13
$\  \cdot \ $ , 3, 61, 93	$[(i) \leftrightarrow (j)]$ , 13
1, $1_{\mathcal{U}}$ , 2, 26	$]f_i[$ , 14
$\mathcal{A}$ , 27	$\mathcal{F}(S, \mathcal{U})$ , 24, 39
$\tilde{\mathcal{A}}$ , 28	$I$ , 32, 42
$\langle \mathcal{A} \rangle$ , 29	(I), 13, 28, 48, 49
$A^*$ , 70	$J_k(\lambda_j)$ , 109
$\overline{A}$ , 56	$\mathcal{L}(\mathcal{U}, \mathcal{V})$ , 39
$A^{-1}$ , 46	$\mathcal{L}(\mathcal{U})$ , 40, 70
$A^-$ , 88, 112	$\lambda$ , 92
$A^t$ , 8	(M), 13, 28, 48, 49
$A > 0$ , 81	$\mathcal{M}_{m,n}$ , $\mathcal{M}_n$ , 8, 41, 43
$A \geq 0$ , 82	$\mathcal{M} + c$ , 25
$[\mathbf{a}_1, \dots, \mathbf{a}_n]$ , 7, 27	$\mathcal{M} \cap \mathcal{N}$ , 25
$\overline{\mathbf{ab}}$ , 85	$\mathcal{M} + \mathcal{N}$ , 25
$\mathbf{a} \cdot \mathbf{b}$ , 3	$\mathcal{N}_1 \cap \dots \cap \mathcal{N}_s$ , 25
$\mathbf{a} \wedge \mathbf{b}$ , 91	$\mathcal{N}_1 + \dots + \mathcal{N}_s$ , 25
$AB$ , 42	$\mathcal{N}_1 \oplus \dots \oplus \mathcal{N}_s$ , 31
$A \subset B$ , 34	$\mathcal{N}_\lambda$ , 53
$A \rightsquigarrow B$ , 28	$\mathcal{N}_T$ , 45
$A\{\mathbf{b}\}_i$ , 101	$\nu(T)$ , 45
$A_j^i$ , $A_j^0$ , 102	(O), 14
$a_{ij}$ , 7, 51	$\mathcal{P}_n$ , 25
$A_k$ , $A_K$ , 82	$\overrightarrow{PQ}$ , 84
$\mathbf{ax}$ , 8, 41	$\text{prl } A$ , 87
$\alpha_j^i$ , 102	$ \text{prl } A $ , 89, 91
(C), 13	$\mathbb{R}^n$ , $\mathbb{R}_n$ , 8, 24
$\mathbb{C}^n$ , 55, 61	$\mathbb{R}_{(n)}^k$ , 26
$\text{Cof } A$ , 102	$\mathcal{R}_T$ , 2, 45
$\text{Cov}, (X, Y)$ , 93	$\mathcal{R} \perp \mathcal{S}$ , 62
$\Delta(\lambda_1, \dots, \lambda_n)$ , 32	$\text{rg } A$ , 37
$\det A$ , $\det_n A$ , 98	$\rho(T)$ , 45
$\dim \mathcal{U}$ , 32	$\rho(X, Y)$ , 94
$\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$ , 4, 32, 70	(S), 13, 28, 48, 49
$E_{ij}$ , $E_i(\lambda)$ , $E_{ij}(\lambda)$ , 48	$S \circ T$ , $ST$ , 1, 40
$EX$ , 92	$S^\perp$ , $S^{\perp\perp}$ , 62
$\mathbf{f}_j$ , 70	$\text{Spec } T$ , 75

$\sigma_X$ , 93  
 $\langle T \rangle$ , 2  
 $T^*$ , 69  
 $T^\circ$ , 111  
 $T^{-1}$ , 2, 26  
 $T^-$ , 112  
 $T \geq 0, T > 0$ , 77  
 $T\langle \mathcal{A} \rangle$ , 29  
 $T\langle \mathcal{M} \rangle$ , 2  
 $T(u), Tu$ , 1  
 $[\mathcal{T}]$ , 6, 42  
 $[\mathbf{u}_1, \dots, \mathbf{u}_n]$ , 27  
 $\langle \mathbf{u}_1, \dots, \mathbf{u}_n \rangle$ , 29  
 $[\mathbf{u}]_{\mathcal{B}}$ , 33  
 $\mathcal{U}/\mathcal{N}$ , 111  
 $\mathcal{U} \longrightarrow \mathcal{V}$ , 1, 25  
 $\mathcal{R} \perp \mathcal{S}$ , 62  
 $Var X$ , 93  
 $\mathbf{w}^*$ , 61  
 $\mathcal{X}$ , 94  
 $\hat{X}$ , 93  
 $(x, y), (x_1, \dots, x_n)$ , 3, 7  
 $\mathbf{x}_{(n)}$ , 26  
 $\mathbf{x} + i\mathbf{y}$ , 55  
 $XY$ , 92  
 $\bar{\mathbf{z}}$ , 55

# Índice alfabético

- adición, 24
- adjunción, 68, 69, 71
- adjunta, 69
- álgebra, 40, 43
- algoritmo, 9
- ángulo, 89
- antisimétrica, 100
- autoadjunto, 76, 77
- autovalor, 53
- autovector, 53
  
- base, 32
  - (cambio de), 47
  - adecuada, 56
  - canónica, 32
  - de Jordan, 106
  - orientada, 104
  - ortonormal, 65
- biyectiva, 2
  
- cadena de Jordan, 106
- cambio de base, 47
- canónica, 111
- característica, 18
- cociente, 111
- coeficiente, 9, 94
  - de correlación, 94
- cofactor, 102
- combinación lineal, 27, 29
- completo (problema), 110
  
- componente(s), 7
  - principales, 94
- composición, 43
- compuesta, 1, 40
- conjugada (matriz), 56
- conjunto solución, 10, 29, 110
- consistente, 10
- coordenadas, 33
- coseno, 3
- covarianza, 93
- Cramer (regla de), 101
- criterio
  - de inyectividad, 100
  - de libertad, 100
  - de regularidad, 100
  
- definición en la base, 41
- definiciones, 1
- definida, 51
- delta de Kronecker, 33, 114
- demostraciones, 1
- dependencia lineal, 30
- desarrollo de un determinante, 103
- descomposición, 50
  - espectral, 58
- desigualdad de Schwarz, 63
- desigualdad del triángulo, 64
- desviación estándar, 93
- determinante, 98

- 
- diagonal, 8
  - diagonalizable, 56
  - diagonalización, 52
    - ortonormal, 67
    - simultánea, 58
  - dimensión, 32
  - directa
    - suma, 32
  - distancia, 61
  - dual, 39
  
  - Eco, Umberto, 96
  - ecuación
    - característica, 54
    - dimensional, 45
    - homogénea, 9
    - lineal, 9
  - elementales
    - (matrices), 48
    - (operaciones), 13, 28
  - eliminación, 9, 13
  - equivalentes
    - (problemas), 10, 29, 35
    - (sistemas), 29, 35
  - escalar, 24
  - escalonada, 18
  - espacio
    - complejo, 55
    - dual, 39
    - euclidiano, 7, 8
    - vectorial, 23, 24
  - espectral
    - (descomposición), 58
    - teorema, 76
  - espectro, 75
  - esperanza, 92
  - estrictamente positiva, 81, 82
  - evaluación, 51
  
  - extensión por linealidad, 41
  
  - forma
    - bilineal, 50
    - cuadrática, 51
    - de Jordan, 58
    - definida, 51
    - escalonada, 18
    - espectral, 58, 77
    - final reducida, 18
    - hermitiana, 72
    - lineal, 39
    - reducida, 18
    - sesquilineal, 72
    - simétrica, 51
  - función, 1
    - biyectiva, 2
    - compuesta, 1
    - identidad, 2
    - inversa, 2
    - inyectiva, 2
    - sobreyectiva, 2
  - funcional, 39
  
  - generado (subespacio), 29
  - generador, 29
  - Gram–Schmidt, 65
  
  - hermitiano(a), 72, 77
  - hiperplano, 86
  - homogéneo(a), 9, 10, 29, 110
  
  - identidad, 2
  - igualdad, 3, 7, 8
  - imagen, 1, 2
  - incógnita(s), 9
    - principales, 12
  - inconsistente, 10, 110
  - independencia lineal, 30



- inducida, 44
- invariante, 2
- inversa, 2, 46
  - (fórmula para la), 104
  - de una matriz elemental, 50
  - generalizada, 111, 112
- invertible, 46
- involución, 70
- inyectiva, 2
- isomorfismo, 26
- isomorfos, 26
- isomorfos(as), 26, 40, 42, 51
  
- Jordan, 58, 105, 109
  
- ley del paralelogramo, 3
- libre, 30
- lineal
  - (combinación), 27, 29
  - (ecuación), 9
  - (operador), 40
  - (problema), 10, 28, 110
  - (transformación), 5, 25
  - (variedad), 25, 67, 110, 111
  - forma, 39
- linealmente dependiente, 30
  - independiente, 30
- longitud, 3
  
- matriz(ces), 5, 7, 10
  - cuadrada, 8
  - diagonal, 33
  - elemental, 48
  - estrictamente positiva, 81, 82
  - inversa, 46
  - invertible, 46
  - normal, 79
  - ortogonal, 80
  - positiva, 81
  - regular, 46
  - semejantes, 48
  - simétrica, 79
  - singular, 47
  - traspuesta, 8
  - unidad, 42
- maximal, 34
- media, 92
- menor, 102
  - principal, 82
- método de eliminación, 13
- mínimos cuadrados, 87, 88, 111
- Moore–Penrose, 112
- multilineal, 101, 102
- multiplicación, 6, 8, 12, 24, 40, 43
- multiplicidad
  - algebraica, 54
  - geométrica, 53, 54
  
- $n$ -pla ordenada, 7
- norma, 3, 61
- normal, 74
- núcleo, 10, 45
- nulidad, 45
  
- operaciones elementales, 13, 28
- operador, 2
  - autoadjunto, 77
  - diagonalizable, 57
  - estrictamente positivo, 77
  - hermitiano, 77
  - invertible, 46
  - lineal, 40
  - normal, 74
  - ortogonal, 77
  - positivo, 77
  - simétrico, 77

- unitario, 77
- orden, 8, 98
- orientación, 97, 104
- ortogonal, 62, 64, 77, 80
- ortonormal, 62
- ortonormalización, 65
  
- parámetro, 12
- paralelepípedo, 87
- pareja ordenada, 3
- parte, 63, 64
- Pitágoras, 66
- plano, 85
- positiva(o), 51, 77, 79
- problema lineal, 10, 28, 110
  - completo, 110
  - consistente, 10, 110
  - homogéneo, 10, 110
  - inconsistente, 10, 110
- problemas equivalentes, 10, 29
- proceso de ortogonalización, 65
- producto, 3, 6, 8, 40, 42
  - externo, 91
  - externo (o vectorial), 91
  - interno, 3, 93
  - tensorial, 73
- proyección ortogonal, 63, 64, 66
- proyecciones asociadas, 40
  
- raíz cuadrada, 78
- rango, 45
- recorrido, 2, 45
- recta, 85
- reducida, 18
- regresión, 95
- regular, 46
- relación
  - de dependencia, 30
  - inductiva, 102
- relación de adjunción, 69, 71
- representación de funcionales, 68
- restricción, 2
- rigor, 1
  
- satisfacer, 9, 10
- Schwarz, 63
- segmento, 85
- semejantes, 48
- sesquilineal, 72
- seudoinversa, 88, 112
- simétrica
  - (forma bilineal), 51
  - (matriz), 79
- simétrico (operador), 77
- singular, 47
- sistema, 10, 27, 94
  - ortonormal, 62
- sistemas equivalentes, 35
- sobreyectiva, 2
- solución, 10, 110, 111
  - óptima, 88, 111
  - general, 20, 110
  - mínima, 88, 111
  - trivial, 110
- soluciones, 9
- subespacio, 25
  - generado, 29
  - propio, 53
- subespacios independientes, 31
- submatriz principal, 82
- suma, 3, 6, 8
  - directa, 32
  - externa, 73
- teorema, 21
  - de los ejes principales, 80

- 
- espectral, 75, 76
  - término independiente, 9
  - transformación
    - canónica, 111
    - de una matriz, 41, 42
    - inducida, 44
    - lineal, 5, 25
  - transformación asociada, 28
  - traspuesta, 8, 48, 100
  - trivial, 10, 30, 110
  
  - unidad, 42, 48
  - unitario, 77
  
  - valor, 1
    - propio, 53
  - variable
    - aleatoria, 92
    - normalizada, 93
    - normalizada, 93
  - varianza, 93
  - variedad lineal, 25, 110
  - vector(es), 24
    - cero, 3, 24
    - columna, 8
    - de coordenadas, 33
    - fila, 8
    - geométrico, 3
    - ortogonales, 61
    - propio, 53
    - unitario, 61
  - volumen, 89

