Métodos Matemáticos

para Estadística

Colección manuales uex - 58

(E.E.E.S.)

Ignacio Ojeda

Jesús Gago

58

MÉTODOS MATEMÁTICOS
PARA ESTADÍSTICA

MANUALES UEX

58

(E.E.E.S.)
Espacio
Europeo
Educación
Superior

IGNACIO OJEDA MARTÍNEZ DE CASTILLA
JESÚS GAGO VARGAS

MÉTODOS MATEMÁTICOS
PARA ESTADÍSTICA

UNIVERSIDAD DE EXTREMADURA
UEX
2008

IGNACIO OJEDA MARTÍNEZ DE CASTILLA / JESÚS GAGO VARGAS

Métodos Matemáticos para Estadística. / Ignacio Ojeda Martínez de Castilla, Jesús Gago Vargas. – Cáceres: Universidad de Extremadura, Servicio de Publicaciones, 2008

533 pp.; 27,8 x 19 cm (Manuales UEX, ISSN 1135-870-X; 58)

ISBN 978-84-691-6429-7

1. Álgebra Lineal. 2. Métodos Numéricos. 3. Análisis Funcional. I. Ojeda Martínez de Castilla, Ignacio. II. Métodos Matemáticos para Estadística. III. Universidad de Extremadura, Servicio de Publicaciones, ed. IV. Manuales UEX

512, 517, 519.6

La publicación del presente manual forma parte de las "Acciones para el Desarrollo del Espacio Europeo de Educación Superior en la Universidad de Extremadura Curso 2007/08" en el marco de la VI Convocatoria de Acciones para la Adaptación de la UEX al Espacio Europeo de Educación Superior (Proyectos Pilotos: modalidad A1) del Vicerrectorado de Calidad y Formación Continua y financiada por la Junta de Extremadura, el Ministerio de Educación y Ciencia y la Universidad de Extremadura. La elaboración del apéndice A se ha realizado en colaboración con Dña. Amelia Álvarez Sánchez.

UNIVERSIDAD DE EXTREMADURA

UNIÓN EUROPEA
Fondo Social Europeo

JUNTA DE EXTREMADURA

Edita

Universidad de Extremadura. Servicio de Publicaciones
C./ Caldereros, 2 - Planta 2ª - 10071 Cáceres (España)
Telf. 927 257 041 - Fax 927 257 046
publicac@unex.es
www.unex.es/publicaciones

ISSN 1135-870-X
ISBN 978-84-691-6429-7
Depósito Legal M-46.669-2008

Edición electrónica: Pedro Cid, S.A.
Teléf.: 914 786 125

Índice general

Introducción

El presente manual está concebido para servir de apoyo a la docencia de una asignatura de métodos matemáticos de un Grado en Estadística y se ha redactado a partir de los apuntes elaborados durante varios cursos para impartir las asignaturas Álgebra y Geometría y Análisis Matemático de la Licenciatura en Ciencias y Técnicas Estadísticas en la Universidad de Extremadura, y de la asignatura Métodos Matemáticos de dicha licenciatura en la Universidad de Sevilla. No obstante, dado su enfoque generalista, este manual puede ser también empleado en asignaturas de Matemáticas de otros grados de la Ramas de Ciencias e Ingeniería y Arquitectura.

El principal objetivo de este manual no es otro que el de proporcionar a los estudiantes de un Grado de Estadística las herramientas matemáticas necesarias para el manejo y comprensión de otras materias, habida cuenta del carácter instrumental de las Matemáticas en todos los procesos y métodos estadísticos.

Los contenidos seleccionados son sistemas lineales, álgebra matricial avanzada, inversas generalizadas, diferenciación matricial, técnicas y software numéricos y una breve introducción a los conceptos elementales del análisis funcional, exponiendo una materia de 12 ó 18 créditos ECTS dependiendo del nivel de conocimiento que tenga el estudiante de álgebra lineal básica. Esta materia podría desglosarse en varias asignaturas con distintas configuraciones. En todo caso, hemos procurado que la materia esté siempre vertebrada en torno dos temas transversales: sistema de ecuaciones lineales y ortogonalidad.

Al final de cada tema se incluye una relación de ejercicios con los que se pretende que el alumno reafirme y aplique los conocimientos adquiridos y se ejercite en el manejo de las técnicas y métodos aprendidos. También hemos considerado fundamental incluir una serie de prácticas con `MATLAB` con el doble objetivo de proporcionar cierta formación en el manejo de software numérico y de servir de ejemplo prácticos de los contenidos teóricos desarrollados en el manual.

Ambos autores quisieran agradecer la ayuda prestada por M. Ángeles Mulero Díaz, Juan Antonio Navarro González, Inés del Puerto García y Batildo Requejo Fernández quienes con sus comentarios y sugerencias han enriquecido notablemente el el manual.

Badajoz-Sevilla, julio de 2008.

TEMA I

Generalidades sobre matrices

Este tema es de carácter introductorio en el que esencialmente se establece gran parte de la notación y se introducen las definiciones de distintos tipos de matrices que se usarán a lo largo del manual.

En primer lugar definimos el concepto de matriz, matriz cuadrada, matriz columna, matriz fila y submatriz. A continuación, y a modo de ejemplo, se definen la matriz nula, las matrices diagonales (y, como caso particular de éstas, la matriz identidad) y las matrices triangulares. Luego, se muestran las operaciones aritméticas elementales de las matrices, aunque sin hacer mención a las distintas estructuras algebraicas determinadas por tales operaciones. Finalmente, se definen la matriz traspuesta, el concepto de matriz invertible y de matriz inversa, y el de matriz ortogonal. Así mismo, se tratan brevemente algunos tipos de matrices con entradas en los complejos (matriz traspuesta conjugada, matriz hermítica, matriz unitaria y matriz normal) aunque sólo serán usadas puntualmente en el manual, y generalmente para advertir de que ciertos resultados válidos para matrices reales no tienen su análogo si cambiamos reales por complejos. Hablando de cuerpos, conviene avisar que casi siempre (por no decir siempre) el cuerpo considerado será $\mathbb{R}$ ó $\mathbb{C}$.

En la segunda sección se definen y estudian la traza y el determinante de una matriz cuadrada. Hemos optado por la siguiente definición de determinante de una matriz A

$$|A| = \sum_{\sigma \in S_n} \operatorname{sign}(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)},$$

donde S_n denota al grupo simétrico, que requiere un cierto grado de abstracción, frente a una definición por recurrencia mediante la fórmula del desarrollo por una fila o columna; no obstante, se propone como ejercicio al lector la demostración de la equivalencia entre ambas definiciones, y de igual modo se propone como ejercicio la demostración de las propiedades habituales del determinante. A continuación, en esta misma sección se introduce el concepto de matriz adjunta y se demuestra la fórmula de la matriz de inversa.

La tercera sección de este tema es quizá la única parte realmente nueva para el estudiante; en ella se introducen y estudian las matrices divididas por bloques y algunas de sus operaciones aritméticas. Desde un punto vista conceptual, no se añade

nada nuevo más allá de una cuestión de notación; sin embargo, el uso de las matrices divididas (también hay quien dice particionadas) por bloques simplifica considerablemente la notación, por ejemplo, para definir la forma canónica de Jordan. Además, se introducen la suma directa y el producto de Kronecker de matrices como ejemplos de construcciones de matrices por bloques. Ambas construcciones serán utilizadas posteriormente, y en concreto, el producto de Kronecker será estudiado con más detalle en el tema VII. En esta última sección se muestran expresiones para la inversa y el determinante para las matrices divididas en la forma 2×2

$$A = \left(\begin{array}{c|c} A_{11} & A_{12} \\ \hline A_{21} & A_{22} \end{array} \right).$$

Las referencias bibliográficas básicas para las dos primeras secciones son el capítulo 1 de [**SV95**] y el capítulo 2 de [**CnR05**]. En el capítulo 3 de [**Mey00**] se pueden encontrar multitud de ejemplos del uso de las matrices en problemas concretos de Estadística y Probabilidad. Para un desarrollo más profundo de las matrices divididas por bloques véase el capítulo 7 de [**Sch05**].

1. Matrices. Definición y propiedades

En todo el manual, $\Bbbk$ denotará un cuerpo (véase la sección 2 del apéndice B) que por lo general será $\mathbb{R}$ ó $\mathbb{C}$.

Se denotará por $\bar{\lambda}$ el **conjugado** de un número complejo $\lambda \in \mathbb{C}$. Así, si $\lambda = \alpha + \beta\,\mathrm{i}$, donde α y β son número reales, será $\bar{\lambda} = \alpha - \beta\,\mathrm{i}$. Las propiedades más comunes de las conjugación compleja son las siguientes:

- $\bar{\bar{\lambda}} = \lambda$;
- $\overline{(\lambda + u)} = \bar{\lambda} + \bar{\mu}$;
- $\overline{\lambda\,\mu} = \bar{\lambda}\,\bar{\mu}$;
- $|\lambda| = \sqrt{\bar{\lambda}\,\lambda}$.

El número real positivo $|\lambda|$ se llama **módulo** de λ. Si λ es un número real, su módulo es su **valor absoluto**.

- $\bar{\lambda} = \lambda$ si, y sólo si, λ es real.

Definición I.1.1. Se llama **matriz** de orden $m \times n$ con coeficientes en $\Bbbk$ a un conjunto ordenado de escalares $a_{ij} \in \Bbbk, i = 1, \ldots, m$ y $j = 1, \ldots, n$, dispuestos en m filas y n columnas, formando un rectángulo. Se representa por

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

Las matrices de orden $n \times n$ con coeficientes en $\mathbb{k}$ se llaman **matrices cuadradas** de orden n con coeficientes en $\mathbb{k}$.

El conjunto de las matrices de orden $m \times n$ con coeficientes en $\mathbb{k}$ se designará por $\mathcal{M}_{m\times n}(\mathbb{k})$, y el conjunto de las matrices cuadradas de orden n con coeficientes en $\mathbb{k}$ se designará por $\mathcal{M}_n(\mathbb{k})$.

Definición I.1.2. Sea $A \in \mathcal{M}_{m\times n}(\mathbb{k})$. El escalar (por ejemplo, el número real o complejo) que se encuentra en la fila i-ésima y la columna j-ésima se llama **entrada** (i,j)-ésima de A; es usual denotarla por a_{ij}, y por tanto representar a la matriz A por (a_{ij}).

Definición I.1.3. Sea $A \in \mathcal{M}_{m\times n}(\mathbb{k})$. Dado $j \in \{1,\dots,n\}$ la matriz

$$\begin{pmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{pmatrix} \in \mathcal{M}_{m\times 1}(\mathbb{k})$$

se llama **columna j-ésima** de A, y dado $i \in \{1,\dots,m\}$ la matriz $(a_{i1} \ \dots \ a_{in}) \in \mathcal{M}_{1\times n}(\mathbb{k})$ se denomina **fila i-ésima** de A.

Definición I.1.4. Dos matrices son iguales si tienen el mismo orden y coinciden entrada a entrada; es decir, si (a_{ij}) y $(b_{ij}) \in \mathcal{M}_{m\times n}(\mathbb{k})$, entonces

$$(a_{ij}) = (b_{ij}) \Longleftrightarrow a_{ij} = b_{ij},\ i = 1,\dots,m,\ j = 1,\dots,n.$$

Definición I.1.5. Sea $A \in \mathcal{M}_{m\times n}(\mathbb{k})$. Llamaremos **submatriz** o **matriz extraída** de A a cualquier matriz obtenida a partir de A suprimiendo algunas de sus filas y/o columnas.

Ejemplos I.1.6. **Algunos tipos de matrices**

i) La **matriz nula** $0 \in \mathcal{M}_{m\times n}(\mathbb{k})$ es aquella con m filas y n columnas cuyas entradas son todas iguales a 0. En algunas ocasiones escribiremos $0_{m\times n}$ para denotar a la matriz nula de orden $m \times n$.

ii) Se dice que una matriz cuadrada $D = (d_{ij}) \in \mathcal{M}_n(\mathbb{k})$ es **diagonal** si $d_{ij} = 0$ para todo $i \neq j$.

En ocasiones, escribiremos

$$\operatorname{diag}(\lambda_1,\dots,\lambda_n),$$

con $\lambda_i \in \mathbb{k},\ i = 1,\dots,n$, para denotar la matriz de diagonal $D = (d_{ij}) \in \mathcal{M}_n(\mathbb{k})$ tal que $d_{ii} = \lambda_i,\ i = 1,\dots,n$.

iii) A la matriz diagonal tal que $d_{ii} = 1$ para todo $i = 1,\dots,n$, se la denomina **matriz identidad** (ó **matriz unidad**) **de orden** n, y se denota por I_n; es

decir,

$$I_n = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}.$$

Con la notación habitual de la *delta de Kronecker*

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

se tine que $I_n = (\delta_{ij}) \in \mathcal{M}_n(\Bbbk)$.

iii) Se dice que una matriz cuadrada $A = (a_{ij}) \in \mathcal{M}_n(\Bbbk)$ es **triangular superior** si $a_{ij} = 0$ cuando $i > j$, y se dice A es **triangular inferior** si $a_{ij} = 0$ cuando $i < j$.

Suma de matrices: En el conjunto $\mathcal{M}_{m\times n}(\Bbbk)$ se define la suma de matrices de la siguiente manera: si $A = (a_{ij})$ y $B = (b_{ij}) \in \mathcal{M}_{m\times n}(\Bbbk)$, entonces

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

La suma de matrices se define como la suma entrada a entrada.

Nota I.1.7. Nótese que la suma de matrices verifica las propiedades asociativa, conmutativa y además,

i) si $A \in \mathcal{M}_{m\times n}(\Bbbk)$ y $0 \in \mathcal{M}_{m\times n}(\Bbbk)$, entonces $A + 0 = 0 + A = A$.
ii) si $A = (a_{ij}) \in \mathcal{M}_{m\times n}(\Bbbk)$, entonces $-A = (-a_{ij})$, de tal forma que $A+(-A) = (-A) + A = 0 \in \mathcal{M}_{m\times n}(\Bbbk)$.

Producto de un escalar por una matriz: Si $A = (a_{ij}) \in \mathcal{M}_{m\times n}(\Bbbk)$ y $\lambda \in \Bbbk$, se define

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

esto es, el producto de un escalar por una matriz es la matriz que resulta al multiplicar cada una de las entradas de la matriz por el escalar.

Producto de matrices: Para que dos matrices puedan multiplicarse, el número de columnas del factor de la izquierda ha de coincidir con el número de filas del factor de la derecha. Sean $A = (a_{il}) \in \mathcal{M}_{m\times p}(\Bbbk)$ y $B = (b_{lj}) \in \mathcal{M}_{p\times n}(\Bbbk)$. Se llama matriz producto $A \cdot B$ a $C = (c_{ij}) \in \mathcal{M}_{m\times n}(\Bbbk)$, cuya entrada (i, j)-ésima es

$$c_{ij} = \sum_{l=1}^{p} a_{il} b_{lj}, \ i = 1, \dots, m, \ j = 1, \dots, n.$$

Definición I.1.8. Sea $A \in \mathcal{M}_{m\times n}(\Bbbk)$ llamamos **matriz traspuesta** de A a la matriz de $\mathcal{M}_{n\times m}(\Bbbk)$ que resulta de cambiar filas por columnas y columnas por filas en A. La matriz traspuesta de A siempre existe y se denota por A^{t}.

Definición I.1.9. Se dice que una matriz $A = (a_{ij}) \in \mathcal{M}_n(\Bbbk)$ es

(a) **Simétrica** si $A = A^{\mathrm{t}}$, es decir, $a_{ij} = a_{ji}$, para todo $i, j = 1, 2, \ldots, n$.
(b) **Antisimétrica** si $A = -A^{\mathrm{t}}$, es decir, $a_{ij} = -a_{ji}$, para todo $i, j = 1, 2, \ldots, n$.

Definición I.1.10. Diremos que una matriz $A \in \mathcal{M}_n(\Bbbk)$ es **invertible** (o **no singular**) si existe $B \in \mathcal{M}_n(\Bbbk)$ tal que $A \cdot B = B \cdot A = I_n$. La matriz B si existe es única[1] se denomina **matriz inversa** de A y la denotaremos por A^{-1}.

Más adelante daremos un criterio para saber si una matriz es invertible y, en este caso, una fórmula para calcular la matriz inversa.

Definición I.1.11. Diremos que una matriz $A \in \mathcal{M}_n(\mathbb{R})$ es **ortogonal** si $A^{\mathrm{t}} = A^{-1}$, es decir, $A\,A^{\mathrm{t}} = A^{\mathrm{t}}A = I_n$.

Definición I.1.12. Sea $A = (a_{ij}) \in \mathcal{M}_{m\times n}(\mathbb{C})$. La matriz $A^* = (\overline{a}_{ji}) \in \mathcal{M}_{n\times m}(\mathbb{C})$ se denomina **matriz traspuesta conjugada**[2]; siendo $\overline{a_{ji}}$ el conjugado complejo de $a_{ji},\ i = 1, \ldots, m,\ j = 1, \ldots, n$.

Claramente, $(A^*)^* = A$ y además, cuando A es real, se tiene que $A^* = A^{\mathrm{t}}$.

Nótese que si

$$\mathbf{v} = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \in \Bbbk^n,$$

entonces $\mathbf{v}^* = (\overline{v}_1, \ldots, \overline{v}_n)$.

Definición I.1.13. Se dice que una matriz $A = (a_{ij}) \in \mathcal{M}_n(\mathbb{C})$ es

(a) **Hermítica** si $A = A^*$, es decir, $a_{ij} = \overline{a_{ji}}$, para todo $i, j = 1, 2, \ldots, n$.
(b) **Unitaria** si $A^* = A^{-1}$, es decir, $A\,A^* = A^*A = I_n$.
(c) **Normal** si $A\,A^* = A^*A$.

Proposición I.1.14.

i) *Toda matriz hermítica o unitaria es normal.*
ii) *Si A es hermítica e invertible, entonces A^{-1} es también hermítica.*
iii) *Si A es normal e invertible, entonces A^{-1} es normal.*

[1] Si existen B y C tales que $AB = BA = I_n = AC = CA$, entonces

$$0 = A(B - C) \Rightarrow 0 = BA = BA(B - C) = B - C \Rightarrow B = C.$$

[2] Algunos autores llaman a esta matriz adjunta.

Demostración. La demostración de esta proposición se propone como ejercicio a lector (ejercicio 6). ∎

2. La traza y el determinante de una matriz

Definición I.2.1. Sea $A = (a_{ij}) \in \mathcal{M}_n(\Bbbk)$. Se denomina **traza** de A al escalar

$$\operatorname{tr}(A) = \sum_{i=1}^{n} a_{ii}.$$

La traza es invariante por transformaciones unitarias:

Proposición I.2.2. *Si $A = (a_{ij}) \in \mathcal{M}_n(\mathbb{C})$ y P es una matriz invertible, entonces* $\operatorname{tr}(A) = \operatorname{tr}(P^{-1}AP)$. *En particular si Q es una matriz unitaria* $\operatorname{tr}(A) = \operatorname{tr}(Q^*AQ)$.

Demostración. La demostración de esta proposición es una consecuencia del apartado 6 del ejercicio 9. ∎

Definición I.2.3. Sea $A = (a_{ij}) \in \mathcal{M}_n(\Bbbk)$. Se llama **determinante** de A, y se representa por $|A|$, al escalar definido por la expresión:

$$|A| = \sum_{\sigma \in S_n} \operatorname{sign}(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)},$$

donde S_n denota al grupo simétrico.[3]

Ejemplo I.2.4. Veamos las expresiones explícitas para los determinantes de las matrices cuadradas de ordenes 2 y 3.

i) Si $A = (a_{ij}) \in \mathcal{M}_2(\Bbbk)$, entonces

$$|A| = a_{11}a_{22} - a_{12}a_{22}$$

ya que $S_2 = \{1, (1\,2)\}$.

ii) Si $A = (a_{ij}) \in \mathcal{M}_3(\Bbbk)$, entonces

$$|A| = a_{11}a_{22}a_{33} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31} - a_{11}a_{23}a_{32} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32},$$

ya que $S_3 = \{1, (1\,2), (1\,3), (2\,3), (1\,2\,3), (3\,2\,1)\}$.

Definición I.2.5. Sea $A \in \mathcal{M}_{m\times n}(\Bbbk)$. Dado un entero positivo $p \leq \min(m, n)$, llamaremos **menores de orden** p de A a los determinantes de las submatrices cuadradas de orden p de A.

Si $m = n$, se llama **menor principal de orden** p al determinate de la submatriz de A que se obtiene al eliminar las últimas $n - p$ filas y columnas de A.

[3] Sea X un conjunto arbitrario con n entradas se llama **grupo simétrico** S_n al conjunto de las biyecciones de X con la composición de aplicaciones (véanse, por ejemplo, la sexta sección del segundo capítulo de [**Nav96**] o la sección décimoquinta de los preliminares de [**BCR07**]).

Nótese que si A es una matriz cuadrada de orden n, entonces tiene un sólo menor de orden n, que es precisamente el determinante de A.

Definición I.2.6. Sea $A = (a_{ij}) \in \mathcal{M}_n(\Bbbk)$. Llamaremos **menor adjunto de la entrada** a_{ij} de A al determinante de la submatriz de A que se obtiene al eliminar la fila i-ésima y la columna j-ésima de A, y lo denotaremos por $|A_{ij}|$.

Los menores adjuntos de una matriz $A \in \mathcal{M}_n(\Bbbk)$ proporcionan otra fórmula para el determinante de A.

Teorema I.2.7. *Sea* $A = (a_{ij}) \in \mathcal{M}_n(\Bbbk)$.

(a) *El determinante de una matriz es igual a la suma alternada de los productos de las entradas de una fila (o columna) cualquiera por sus adjuntos respectivos. Es decir, si elegimos la fila i-ésima, el determinante de la matriz A es:*

$$\begin{aligned}|A| &= (-1)^{i+1}a_{i1}|A_{i1}| + (-1)^{i+2}a_{i2}|A_{i2}| + \ldots + (-1)^{i+n}a_{in}|A_{in}| \\ &= \sum_{j=1}^{n}(-1)^{i+j}a_{ij}|A_{ij}|,\end{aligned}$$

o si elegimos la columna j-ésima, el determinante de la matriz A es:

$$\begin{aligned}|A| &= (-1)^{1+j}a_{1j}|A_{1j}| + (-1)^{2+j}a_{2j}|A_{2j}| + \ldots + (-1)^{n+j}a_{nj}|A_{nj}| \\ &= \sum_{i=1}^{n}(-1)^{i+j}a_{ij}|A_{ij}|.\end{aligned}$$

A la primera expresión se la llama ***desarrollo del determinante por la fila i-esima*** *y a la segunda* ***desarrollo del determinante por la columna j-esima****.*

(b) *La suma alternada de los productos de las entradas de una fila por los adjuntos de las entradas respectivas de otra es igual a cero, es decir:*

$$(-1)^{i+1}a_{i1}|A_{j1}| + (-1)^{i+2}a_{i2}|A_{j2}| + \ldots + (-1)^{i+n}a_{in}|A_{jn}| = 0,$$

para todo $i \neq j$. Obviamente, la afirmación anterior también es cierta por columnas.

Demostración. La demostración es un sencillo (aunque bastante tedioso) ejercicio que sigue de la propia definición de determinante de un matriz. ∎

Propiedades de los determinantes. *Sea* $A = (a_{ij}) \in \mathcal{M}_n(\Bbbk)$.

1. *Si B es la matriz traspuesta de A, entonces $|B| = |A|$, es decir, $|A^{\mathrm{t}}| = |A|$.*
2. *Si una fila (o columna) de A es combinación lineal de otras de sus filas (o columnas), es decir, es el resultado de sumar otras de sus filas (o columnas) multiplicadas por un escalar, entonces $|A| = 0$.*

Así, en particular, el determinante de una matriz A con dos filas (o columnas) iguales o proporcionales es nulo. Asimismo, si todos las entradas de una fila (o columna) de A son nulas, entonces $|A| = 0$.

3. *Si se intercambian entre sí dos filas (o columnas) de A, el determinante de la matriz B obtenida es el opuesto del determinante de A, es decir, $|B| = -|A|$.*
4. *Si se multiplica una fila (o columna) cualquiera de la matriz A por un escalar λ, el determinante de la matriz B obtenida es igual al producto de λ por el determinante de A, esto es, $|B| = \lambda|A|$.*
5. *Si cada entrada de una fila (o columna), por ejemplo la fila p, de la matriz A es de la forma $a_{pj} = a'_{pj} + a''_{pj}$, entonces el determinante de A es igual a la suma de los determinantes de dos matrices B y C, tales que la fila p de B está formada por las entradas a'_{pj} y la fila p de C está formada por las entradas a''_{pj}, y las restantes filas de ambas matrices son respectivamente iguales a las de A.*
6. *Si a la fila (o columna) p de A se le suma otra fila (columna) q multiplicada por un escalar λ, el determinante de la matriz obtenida es igual al determinante de A.*

Nota I.2.8. Es importante resaltar que $|A + B| \neq |A| + |B|$ y que $|\lambda\, A| \neq \lambda\, |A|$.

Fórmula de la matriz inversa.

Terminamos esta sección mostrando una fórmula para la matriz inversa de una matriz invertible dada. Comenzamos definiendo qué se entiende por matriz adjunta.

Definición I.2.9. Sea $A \in \mathcal{M}_n(\Bbbk)$. Llamaremos **matriz adjunta**[4] de A, y la denotaremos por $\mathrm{adj}(A)$, a la matriz

$$\mathrm{adj}(A) = ((-1)^{i+j}|A_{ji}|) \in \mathcal{M}_n(\Bbbk).$$

La matriz adjunta verifica la siguiente propiedad.

Lema I.2.10. *Sea $A \in \mathcal{M}_n(\Bbbk)$. Entonces se cumple que*

$$A \cdot \mathrm{adj}(A) = \mathrm{adj}(A) \cdot A = \begin{pmatrix} |A| & 0 & \dots & 0 \\ 0 & |A| & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & |A| \end{pmatrix} = |A| \cdot I_n,$$

donde I_n denota a la matriz identidad de orden n.

[4]No confundir con la matriz traspuesta conjugada.

Demostración. Sea $A \cdot \operatorname{adj}(A) = (c_{ij}) \in \mathcal{M}_n(\Bbbk)$. Dados dos índices $i, j \in \{1, \ldots, n\}$ tenemos que

$$c_{ij} = \sum_{h=1}^{n} a_{ih}((-1)^{h+j}|A_{jh}|);$$

luego, del teorema I.2.7 se sigue que $c_{ij} = |A|$ si $i = j$ y $c_{ij} = 0$ en otro caso. ∎

Fórmula de la matriz inversa. *La condición necesaria y suficiente para que una matriz cuadrada A tenga inversa es que su determinante sea distinto de cero. En cuyo caso,*

$$A^{-1} = \frac{1}{|A|} \operatorname{adj}(A).$$

Demostración. El resultado es una consecuencia inmediata del lema I.2.10 y de la unicidad de la matriz inversa. ∎

3. Matrices por bloques

A menudo es aconsejable dividir una matriz dada en submatrices. Por ejemplo, dada $A = (a_{ij}) \in \mathcal{M}_5(\mathbb{R})$, queremos dividirla en cuatro submatrices de la siguiente manera

$$A = \left(\begin{array}{cc|ccc} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ \hline a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} \end{array}\right) = \left(\begin{array}{c|c} A_{11} & A_{12} \\ \hline A_{21} & A_{22} \end{array}\right), \tag{I.3.1}$$

donde

$$A_{11} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \quad A_{21} = \begin{pmatrix} a_{31} & a_{32} \\ a_{41} & a_{42} \\ a_{51} & a_{52} \end{pmatrix}, \quad A_{12} = \begin{pmatrix} a_{13} & a_{14} & a_{15} \\ a_{23} & a_{24} & a_{25} \end{pmatrix},$$

y

$$A_{22} = \begin{pmatrix} a_{33} & a_{34} & a_{35} \\ a_{43} & a_{44} & a_{45} \\ a_{53} & a_{54} & a_{55} \end{pmatrix}.$$

En general, una matriz se puede descomponer de multitud de formas en submatrices con cualquier número de entradas, suponiendo, claro está, que el numero total de filas y columnas sea igual que el número de filas y columnas original. Una matriz descompuesta de esta forma se conoce como **matriz divida por bloques**. Habitualmente las matrices bloques se usan para enfatizar el papel de algunas de las entradas que ocupan filas y/o columnas adyacentes. Recíprocamente, podemos considerar que A

es una **matriz aumentada por bloques**, donde las matrices A_{11}, A_{21}, A_{12} y A_{22} se han combinado para construir una matriz mayor. Evidentemente, la aumentación se puede entender como el proceso opuestos al de la división.

Se pueden realizar **operaciones con matrices por bloques** de un modo muy parecido al que hicimos con la matrices en la primera sección. Sea A la matriz por bloques

$$A = \begin{pmatrix} A_{11} & A_{12} & \dots & A_{1m} \\ A_{21} & A_{22} & \dots & A_{2m} \\ \vdots & \vdots & & \vdots \\ A_{n1} & A_{n2} & \dots & A_{nm} \end{pmatrix}$$

donde las entradas A_{ij} son submatrices. Entonces, si otra B es otra matriz divida por bloques de la misma forma, es decir, tal que B_{ij} tiene el mismo orden que $A_{ij},\ i = 1, \dots, n; j = 1, \dots, m$, entonces

$$A + B = \begin{pmatrix} A_{11} + B_{11} & A_{12} + B_{12} & \dots & A_{1m} + B_{1m} \\ A_{21} + B_{21} & A_{22} + B_{22} & \dots & A_{2m} + B_{2m} \\ \vdots & \vdots & & \vdots \\ A_{n1} + B_{n1} & A_{n2} + B_{n2} & \dots & A_{nm} + B_{nm} \end{pmatrix}$$

también es una matriz divida por bloques. Análogamente si las dimensiones de las submatrices de dos matrices por bloques C y D son apropiadas para la multiplicación, entonces tenemos que

$$\begin{aligned} CD &= \begin{pmatrix} C_{11} & C_{12} & \dots & C_{1p} \\ C_{21} & C_{22} & \dots & C_{2p} \\ \vdots & \vdots & & \vdots \\ C_{m1} & C_{m2} & \dots & C_{mp} \end{pmatrix} \begin{pmatrix} D_{11} & D_{12} & \dots & D_{1m} \\ D_{21} & D_{22} & \dots & C_{2m} \\ \vdots & \vdots & & \vdots \\ D_{p1} & D_{p2} & \dots & D_{pm} \end{pmatrix} \\ &= \left(\sum_{l=1}^{p} C_{il} D_{lj} \right), \end{aligned}$$

donde C_{ij} y D_{ij} son submatrices de ordenes apropiados para que el producto tenga sentido. Como se puede observar tanto en la suma como en el producto podemos considerar que la submatrices juegan un papel análogo al de los escalares respecto a la suma y el producto de matrices estudiados en la primera sección.

Se pueden definir otros productos y sumas de matrices en términos de matrices aumentadas por bloques, si bien es cierto que de una forma completamente distinta a la anterior. Sean A y B dos matrices cuadradas de ordenes n y m, respectivamente. Entonces las **suma directa** se define como la siguiente matriz aumentada de orden

$(n+m)\times(m+n)$

$$A\oplus B := \left(\begin{array}{c|c} A & 0 \\ \hline 0 & B \end{array}\right).$$

Evidentemente, la suma directa se puede generalizar a cualquier cantidad finita de matrices cuadradas. El resultado de esta operación es lo que se conoce como una **matriz diagonal por bloques**. Es claro que la suma directa de matrices es asociativa, aunque no es conmutativa.

Proposición I.3.1. *Sean $A_1,\ldots,A_r$ matrices tales que $A_i\in\mathcal{M}_{m_i}(\mathbb{R})$, $i=1,\ldots,r$. Se cumple que*

(a) $\operatorname{tr}(A_1\oplus\ldots\oplus A_r)=\operatorname{tr}(A_1)+\ldots+\operatorname{tr}(A_r)$.
(b) $|A_1\oplus\ldots\oplus A_r|=|A_1|\cdots|A_r|$,
(c) *si cada A_i es invertible, entonces $A=A_1\oplus\ldots\oplus A_r$ también es invertible y $A^{-1}=A_1^{-1}\oplus\ldots\oplus A_r^{-1}$.*

Demostración. La demostración, que no es más una sencilla comprobación, se deja como ejercicio al lector. ∎

Sean ahora A y B dos matrices de ordenes $m\times n$ y $p\times q$, respectivamente. Se define el **producto de Kronecker** de A por B como la matriz por bloques de orden $mp\times nq$ tal que

$$A\otimes B := \begin{pmatrix} a_{11}B & a_{12}B & \ldots & a_{1n}B \\ a_{21}B & a_{22}B & \ldots & a_{2n}B \\ \vdots & \vdots & & \vdots \\ a_{m1}B & a_{m2}B & \ldots & a_{mn}B \end{pmatrix}.$$

También se pueden expresar funciones escalares de las matrices cuadradas tales como la traza o el determinante, así como la (única) matriz inversa, en términos de matrices divididas por bloques. Sea $A\in\mathcal{M}_n(\Bbbk)$ divida por bloques de la siguiente manera

$$A=\left(\begin{array}{c|c} A_{11} & A_{12} \\ \hline A_{21} & A_{22} \end{array}\right),$$

con A_{11} y A_{22} cuadradas. Entonces, se comprueba fácilmente que

$$\operatorname{tr}(A)=\operatorname{tr}(A_{11})+\operatorname{tr}(A_{22}),$$

puesto que en la definición de traza de una matriz sólo están involucrados las entradas de la diagonal principal. Además, cuando A_{11} es invertible, el determinante viene dado por

$$|A|=|A_{11}||A_{22}-A_{21}A_{11}^{-1}A_{12}|,$$

o por

$$|A|=|A_{22}||A_{11}-A_{12}A_{22}^{-1}A_{21}|$$

cuando A_{22} es invertible. En el caso especial en que las matrices A_{11}, A_{12}, A_{21} y A_{22} son cuadradas se tiene también que

$$\begin{aligned}
|A| &= |A_{11}A_{22} - A_{21}A_{12}| \quad \text{si} \quad A_{11}A_{21} = A_{21}A_{11},\\
|A| &= |A_{22}A_{11} - A_{21}A_{12}| \quad \text{si} \quad A_{11}A_{12} = A_{12}A_{11},\\
|A| &= |A_{11}A_{22} - A_{12}A_{21}| \quad \text{si} \quad A_{21}A_{22} = A_{22}A_{21},\\
|A| &= |A_{22}A_{11} - A_{12}A_{21}| \quad \text{si} \quad A_{12}A_{22} = A_{22}A_{12}.
\end{aligned}$$

Cuando ambas matrices A_{11} y A_{22} son invertibles, se puede comprobar mediante multiplicación de forma directa que la inversa de A se puede expresar como sigue

$$A^{-1} = \begin{pmatrix} B & -BA_{12}A_{22}^{-1} \\ -A_{22}^{-1}A_{21}B & A_{22}^{-1} - A_{22}^{-1}A_{21}BA_{12}A_{22}^{-1} \end{pmatrix},$$

donde B es $(A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1}$. Aunque parezca difícil de creer, a veces es más fácil invertir A usando la fórmula anterior.

Ejercicios del tema I

Ejercicio 1. Sean A y $B \in \mathcal{M}_{m\times n}(\Bbbk)$ y $\lambda \in \Bbbk$. Probar que el producto de un escalar por una matriz verifica las siguientes propiedades:

1. $\lambda \cdot (A + B) = \lambda \cdot A + \lambda \cdot B$.
2. $(\lambda + \mu) \cdot A = \lambda \cdot A + \mu \cdot A$.
3. $(\lambda \cdot \mu) \cdot A = \lambda \cdot (\mu \cdot A)$.
4. $1 \cdot A = A$.

Ejercicio 2. Probar las siguientes afirmaciones siempre que sea posible efectuar los productos indicados (por ejemplo si las matrices son cuadradas de orden n).

1. El producto de matrices es asociativo: $(A \cdot B) \cdot C = A \cdot (B \cdot C)$.
2. El producto de matrices no es conmutativo.
3. Dada una matriz A, no existe, en general, el elemento inverso de A.
4. El elemento unidad de $\mathcal{M}_n(\Bbbk)$ para el producto de matrices es I_n la matriz identidad de orden n, es decir, $A \cdot I_n = I_n \cdot A = A$.
5. El producto de matrices es distributivo respecto de la suma: $A \cdot (B + C) = A \cdot B + A \cdot C$ y $(B + C) \cdot A = B \cdot A + C \cdot A$.

Ejercicio 3. Sea $A \in \mathcal{M}_{m\times n}(\Bbbk)$. Probar las siguientes igualdades y afirmaciones

1. $(A^{\mathrm{t}})^{\mathrm{t}} = A$.
2. $(A + B)^{\mathrm{t}} = A^{\mathrm{t}} + B^{\mathrm{t}}$, para cualquier matriz $B \in \mathcal{M}_{m\times n}(\Bbbk)$.
3. $(A \cdot B)^{\mathrm{t}} = B^{\mathrm{t}} \cdot A^{\mathrm{t}}$, para cualquier matriz $B \in \mathcal{M}_{n\times p}(\Bbbk)$.
4. Si A es invertible, $(A^{-1})^{\mathrm{t}} = (A^{\mathrm{t}})^{-1}$.
5. Si A tiene coeficientes reales, entonces $A^{\mathrm{t}} \cdot A = 0$ si, sólo si, $A = 0$.

¿Son ciertas las igualdades y afirmaciones anteriores si se sustituye la traspuesta por la traspuesta conjugada?

Ejercicio 4. Sea $A \in \mathcal{M}_n(\mathbb{R})$. Probar que

1. $(A + A^{\mathrm{t}})$ es simétrica y $(A - A^{\mathrm{t}})$ es antisimétrica.
2. $A = \frac{1}{2}(A + A^{\mathrm{t}}) + \frac{1}{2}(A - A^{\mathrm{t}})$
3. A puede escribirse, de modo único,como suma de una matriz simétrica y otra antisimétrica.

Ejercicio 5. Sean a, b y c números reales tales que $a^2+b^2+c^2 = 1$ y consideramos la matriz:

$$A = \begin{pmatrix} 0 & a & -b \\ -a & 0 & c \\ b & -c & 0 \end{pmatrix}$$

1. Probar que la matriz $M = A^2 + I_3$ es simétrica, siendo I_3 la matriz identidad de orden tres.
2. Demostrar que la matriz A es antisimétrica (es decir, $A^{\mathrm{t}} = -A$).
4. Demostrar que la matriz M es **idempotente** (es decir, $M^2 = M$).

Ejercicio 6. Probar que

i) Toda matriz hermítica o unitaria es normal.
ii) Toda matriz triangular y unitaria es diagonal.
iii) Si $A \in \mathcal{M}_n(\mathbb{C})$ es hermítica e invertible, entonces A^{-1} es también hermítica.
iv) Si $A \in \mathcal{M}_n(\mathbb{C})$ es normal e invertible, entonces A^{-1} es normal.

[El ejercicio 3 será de utilidad.]

Ejercicio 7. Probar que

i) $|I_n| = 1$.
ii) $|\lambda\, A| = \lambda^n\, |A|$, para cualquier $A \in \mathcal{M}_n(\Bbbk)$ y $\lambda \in \Bbbk$.
iii) $|AB| = |A||B|$, para cualquier $A \in \mathcal{M}_n(\Bbbk)$ y $B \in \mathcal{M}_n(\Bbbk)$.

Ejercicio 8. Sea $A \in \mathcal{M}_n(\Bbbk)$. Probar que A es invertible si, y sólo si, $|A| \neq 0$, en cuyo caso,

$$|A^{-1}| = \frac{1}{|A|}.$$

Ejercicio 9. Si $A = (a_{ij}) \in \mathcal{M}_n(\Bbbk)$ es una matriz cuadrada de orden n, entonces se define la *traza* de A, que denotaremos por $\mathrm{tr}\,(A)$, como $\mathrm{tr}\,(A) = \sum_{i=1}^n a_{ii}$. Probar que si A y B son matrices cuadradas de orden n, entonces:

1. $\mathrm{tr}\,(A + B) = \mathrm{tr}\,(A) + \mathrm{tr}\,(B)$.
2. $\mathrm{tr}(A) = \mathrm{tr}(A^{\mathrm{t}})$.
3. $\mathrm{tr}(I_n) = n$.
4. $\mathrm{tr}\,(A \cdot B) = \mathrm{tr}\,(B \cdot A)$.
5. $\mathrm{tr}(ABC) = \mathrm{tr}(CAB) = \mathrm{tr}(BCA)$. Comprobar que dicho escalar no tiene por qué ser igual a $\mathrm{tr}(CBA)$.
6. $\mathrm{tr}(A) = \mathrm{tr}(PAP^{-1})$, para cualquier matriz invertible $P \in \mathcal{M}_n(\Bbbk)$.
7. $\mathrm{tr}(AA^{\mathrm{t}}) = \sum_{i,j} a_{ij}^2$.

Ejercicio 10. Se llama **determinante de Vandermonde** de unos ciertos escalares $(x_1, \ldots, x_n)$ al determinante definido por la igualdad

$$V(x_1, \ldots, x_n) = \begin{vmatrix} 1 & 1 & \ldots & 1 \\ x_1 & x_2 & \ldots & x_n \\ x_1^2 & x_2^2 & \ldots & x_n^2 \\ \vdots & \vdots & & \vdots \\ x_1^{n-1} & x_2^{n-1} & \ldots & x_n^{n-1} \end{vmatrix}.$$

Probar la siguiente relación de recurrencia:

$$V(x_1,\ldots,x_n)=(x_n-x_1)\cdot(x_{n-1}-x_1)\cdot\ldots\cdot(x_2-x_1)\cdot V(x_2,\ldots,x_n).$$

Concluir de lo anterior la siguiente igualdad: $V(x_1,\ldots,x_n)=\prod_{i<j}(x_j-x_i)$. Como consecuencia, el determinante de Vandermonde de unos escalares es igual a 0 si y sólo si entre dichos escalares hay dos iguales.

Como aplicación de lo anterior probar que se satisface la igualdad

$$\begin{vmatrix} 1 & 1 & 1 & \ldots & 1 \\ 1 & 2 & 2^2 & \ldots & 2^{n-1} \\ 1 & 3 & 3^2 & \ldots & 3^{n-1} \\ \ldots & \ldots & \ldots & \ddots & \ldots \\ 1 & n & n^2 & \ldots & n^{n-1} \end{vmatrix} = 1!\cdot 2!\cdot\ldots\cdot(n-1)!.$$

Ejercicio 11. Diremos que una matriz N cuadrada de orden n es **nilpotente** si existe un número natural $r\geq 1$ tal que $N^r=0_n$. Probar que si N es nilpotente, entonces la matriz I_n-N es invertible y, además:

$$(I-N)^{-1}=I_n+N+N^2+\ldots+N^{r-1}.$$

Como aplicación, calcular la matriz inversa de la matriz siguiente:

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 0 & 1 & 2 & 3 & 4 \\ 0 & 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Ejercicio 12. Suponiendo que las inversas existen. Probar que

1. $(I+A^{-1})^{-1}=A(A+I)^{-1}$.
2. $(A+BB^{\mathrm{t}})^{-1}B=A^{-1}B(I+B^{\mathrm{t}}A^{-1}B)^{-1}$.
3. $(A^{-1}+B^{-1})^{-1}=A(A+B)^{-1}B=B(A+B)^{-1}A$.
4. $(I+AB)^{-1}=I-A(I+BA)^{-1}B$.
5. $(I+AB)^{-1}A=A(I+BA)^{-1}$.
6. $(A+UBV)^{-1}=A^{-1}-A^{-1}UBV(I+A^{-1}UBV)^{-1}A^{-1}$.

Ejercicio 13. Probar que $\mathbf{v}\mathbf{v}^{\mathrm{t}}-\mathbf{v}^{\mathrm{t}}\mathbf{v}I$ no es invertible.

Ejercicio 14. Dados $A\in\mathcal{M}_n(\mathbb{R})$ invertible y $\mathbf{b}\in\mathbb{R}^n$ tales que $\mathbf{b}^{\mathrm{t}}A^{-1}\mathbf{b}\neq 1$, probar que $(A-\mathbf{b}\mathbf{b}^{\mathrm{t}})^{-1}=A^{-1}+(1-\mathbf{b}^{\mathrm{t}}A^{-1}\mathbf{b})^{-1}(A^{-1}\mathbf{b})(\mathbf{b}^{\mathrm{t}}A^{-1})$.

Ejercicio 15. Probar que

1. $(I+\mathbf{a}\mathbf{b}^{\mathrm{t}})^{-1}=I-\frac{1}{1+\mathbf{b}^{\mathrm{t}}\mathbf{a}}\mathbf{a}\mathbf{b}^{\mathrm{t}}$.
2. $(A+\mathbf{c}\mathbf{d}^{\mathrm{t}})^{-1}=A^{-1}-\frac{A^{-1}\mathbf{c}\mathbf{d}^{\mathrm{t}}A^{-1}}{1+\mathbf{d}^{t}A^{-1}\mathbf{c}}$.

Ejercicio 16. Si $\mathbf{u}, \mathbf{v} \in \mathbb{C}^n$, la matriz $A = I_n + \mathbf{u}\mathbf{v}^*$ se llama **perturbación de rango** 1 de la identidad. Demostrar que si A es invertible, entonces su inversa tiene la forma $A^{-1} = I + \alpha\mathbf{u}\mathbf{v}^*$, para algún escalar α. Deducir una expresión para α. ¿Para qué vectores $\mathbf{u}$ y $\mathbf{v} \in \mathbb{C}^n$ la matriz A no es invertible?

Ejercicio 17. Probar que A y B son invertibles si, y sólo si, $A \oplus B$ es invertible. En tal caso $(A \oplus B)^{-1} = A^{-1} \oplus B^{-1}$.

Ejercicio 18. Consideremos la matriz cuadrada

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix},$$

con A_{11} y A_{22} matrices cuadradas. Probar que si A_{11} es invertible, entonces

$$|A| = |A_{11}| \cdot |A_{22} - A_{21}A_{11}^{-1}A_{12}|.$$

Ejercicio 19. Sean A_{11}, A_{12}, A_{21} y A_{22} matrices de órdenes respectivos $m \times m, m \times n, n \times m$ y $n \times n$, con A_{11} invertible. Probar que

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

es invertible si, y sólo si, $B = A_{22} - A_{21}A_{11}^{-1}A_{12}$ es invertible. En cuyo caso,

$$A^{-1} = \begin{pmatrix} A_{11}^{-1}(A_{11} + A_{12}B^{-1}A_{21})A_{11}^{-1} & -A_{11}^{-1}A_{12}B^{-1} \\ -B^{-1}A_{21}A_{11}^{-1} & B^{-1} \end{pmatrix}.$$

La matriz B se denomina **complemento de Schur** de A_{11} en A.

Ejercicio 20. Dadas $A \in \mathcal{M}_{m\times n}(\Bbbk)$ y $B \in \mathcal{M}_{n\times m}$. Probar que la matriz por bloques

$$L = \begin{pmatrix} I_n - BA & B \\ 2A - ABA & AB - I_m \end{pmatrix}$$

tiene la propiedad $L^2 = I_{m+n}$.

Ejercicio 21. Sea $A \in \mathcal{M}_{m\times n}(\Bbbk)$. Probar que las matrices por bloques

$$\begin{pmatrix} I_n & 0 \\ A & I_m \end{pmatrix}$$

y

$$\begin{pmatrix} I_m & A \\ 0 & I_n \end{pmatrix}$$

son invertibles, y que

$$\begin{pmatrix} I_n & 0 \\ A & I_m \end{pmatrix}^{-1} = \begin{pmatrix} I_n & 0 \\ -A & I_m \end{pmatrix}.$$

Ejercicio 22. Sean A, B y C matrices de órdenes respectivos $m \times m$, $n \times m$ y $n \times n$. Probar que la matriz por bloques

$$\begin{pmatrix} A & 0 \\ B & C \end{pmatrix}$$

es invertible si, y sólo si, A y C son invertibles. En tal caso,

$$\begin{pmatrix} A & 0 \\ B & C \end{pmatrix}^{-1} = \begin{pmatrix} A^{-1} & 0 \\ -C^{-1}BA^{-1} & C^{-1} \end{pmatrix}.$$

Ejercicio 23. Dada la matriz

$$A = \begin{pmatrix} 1 & 0 & 0 & 1/3 & 1/3 & 1/3 \\ 0 & 1 & 0 & 1/3 & 1/3 & 1/3 \\ 0 & 0 & 1 & 1/3 & 1/3 & 1/3 \\ 0 & 0 & 0 & 1/3 & 1/3 & 1/3 \\ 0 & 0 & 0 & 1/3 & 1/3 & 1/3 \\ 0 & 0 & 0 & 1/3 & 1/3 & 1/3 \end{pmatrix}.$$

Calcular A^{300} mediante una división por bloques.

TEMA II

Matrices y aplicaciones lineales

El planteamiento inicial del tema consiste en introducir la equivalencia de matrices: diremos que dos matrices A y B son equivalentes, si existen P y Q invertibles, tales que $B = Q^{-1}AP$, y proponer el problema de decidir cuándo dos matrices son equivalentes; o lo que es lo mismo, determinar la clase de equivalencia de una matriz dada. Así, comenzamos definiendo las transformaciones elementales por filas y por columnas de una matriz, identificando las matrices elementales de paso en cada caso, mostrando de este modo que las transformaciones elementales producen matrices equivalentes. A continuación probamos que toda matriz es equivalente a su forma reducida por filas y a su forma reducida por columnas mediante el método de Gauss-Jordan, y comprobamos que la forma reducida por filas de la forma reducida por columnas y que la forma reducida por columnas de la forma reducida por filas de la matriz A dada, confluyen en una misma matriz

$$R = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}$$

que llamamos forma reducida de A. Usando que las formas reducidas por filas y por columnas de una matriz son únicas salvo permutación de algunas columnas y filas, respectivamente, concluimos que la forma reducida es única, y, por consiguiente, que toda matriz tiene asociado un invariante numérico por la equivalencia de matrices; concretamente, el orden de la matriz identidad que aparece en su forma reducida, al que llamaremos rango de la matriz. De esta forma se resuelve el problema planteado inicialmente, ya que podemos afirmar que dos matrices son equivalentes si, y sólo si, tienen el mismo rango; siendo además su forma reducida un representante canónico de su clase equivalencia.

Si bien nuestro problema inicial ya está resuelto, nos proponemos determinar la naturaleza geométrica del rango de una matriz. Para ello recurrimos a las aplicaciones lineales entre espacios vectoriales abstractos. Este es un buen momento para recordar que en todas las titulaciones que dan acceso a la Licenciatura en Ciencias y Técnicas Estadísticas se imparte Álgebra Lineal básica, por lo tanto, se entiende que los conceptos de espacio vectorial, dependencia e independencia lineal y base son conocidos. Por supuesto, todos los espacios vectoriales de esta asignatura serán de

dimensión finita a menos que diga lo contrario. En la segunda sección de este tema se parte de la definición de aplicación lineal entre espacios vectoriales abstractos, y se recuerdan las definiciones de monomorfismo, epimorfismo, isomorfismo, núcleo e imagen de una aplicación lineal. Asimismo, se recuerda qué se entiende por coordenadas de un vector respecto de una base, y se da la definición de matriz asociada a una aplicación lineal.

A modo de ejemplo se comenta que, por defecto, se entenderá que una matriz $A \in \mathcal{M}_{m\times n}(\mathbb{R})$ define una aplicación lineal de $\mathbb{R}^n$ en $\mathbb{R}^m$; concretamente la aplicación lineal cuya matriz respecto de las bases usuales de $\mathbb{R}^m$ y $\mathbb{R}^n$ es A. Esto nos permitirá hablar con libertad de A en términos de aplicaciones lineales. Así, por ejemplo, podremos afirmar que *si A tiene rango r y $R = Q^{-1}AP$ es su forma reducida, con $P \in \mathcal{M}_n(\mathbb{R})$ y $Q \in \mathcal{M}_m(\mathbb{R})$ invertibles, entonces las últimas $n-r$ columnas de P forman una base de* $\ker(A)$ *y las r primeras columnas de Q forman una base de* $\mathrm{im}(A)$. Entendiendo que núcleo e imagen lo son de la aplicación natural que define A. Destacamos este ejemplo por ser el que podríamos considerar ejemplo fundamental del tema, ya que pone de manifiesto la clave de la demostración del teorema del rango.

A continuación se enuncian y demuestran algunos resultados básicos de las aplicaciones lineales con los que el alumno debe estar familiarizado. A saber, las ecuaciones de una aplicación lineal, el isomorfismo entre el espacio vectorial de las aplicaciones lineales de V en V' y el correspondiente espacio vectorial de matrices para cada par de bases fijas de V y V', la correspondencia entre la composición de aplicaciones lineales y el producto de matrices, y, en el caso de los isomorfismos, su correspondencia con las matrices invertibles. Estos resultados sólo son enunciados en clase y, generalmente, usando transparencias.

La siguiente sección del tema está dedicada a los cambios de base, y cómo afectan éstos a las matrices asociadas a las aplicaciones lineales. Es decir, demostramos que *dos matrices son equivalentes si, y sólo si, están asociadas a una misma aplicación lineal respecto de bases distintas.* Este argumento nos permite afirmar que el rango de una matriz tiene carácter puramente geométrico (Teorema del rango).

Al final de este tema se comentan brevemente algunos aspectos relacionados con la resolución de sistemas de ecuaciones lineales como antesala a la resolución aproximada mínimo cuadrática de sistema de ecuaciones lineales que se estudiará en el tema VI.

La bibliografía básica utilizada en este tema ha sido [**SV95**] y [**MS06**] para la primera sección, y el tema 3 de [**BCR07**] para el resto de secciones. Para un desarrollo más geométrico de este tema se puede consultar [**Her85**]. El capítulo 6 de [**Sea82**] está completamente dedicado al rango, y cuenta con bastantes ejemplos relacionados con la Estadística. En el capítulo 4 de [**Mey00**] también se pueden encontrar aplicaciones y ejercicios aplicados a situaciones reales de los contenidos de este tema.

En el desarrollo de este tema, y en el del manual en general, se ha supuesto que el estudiante está familiarizado con los conceptos de espacio y subespacio vectorial, dependencia lineal, base y dimensión. En todo caso, con el ánimo de hacer este manual lo más autocontenido posible, en el apéndice C pueden encontrarse todos estos conceptos tratados con bastante profusión.

1. Matrices equivalentes

Definición II.1.1. Se dice que $A \in \mathcal{M}_{m\times n}(\Bbbk)$ **es equivalente a** $A' \in \mathcal{M}_{m\times n}(\Bbbk)$ si existen $P \in \mathcal{M}_n(\Bbbk)$ y $Q \in \mathcal{M}_m(\Bbbk)$ invertibles tales que

$$A' = Q^{-1}\, A\, P.$$

La relación anterior es de equivalencia, es decir, verifica las propiedades reflexiva, simétrica y transitiva (compruébese).

Definición II.1.2. Se llaman **operaciones elementales por filas** en una matriz $A \in \mathcal{M}_{m\times n}(\Bbbk)$ a las siguientes transformaciones:

(a) Tipo I: Intercambiar las filas i-ésima y l-ésima de A.
(b) Tipo II: Multiplicar la fila i-ésima de A por $\lambda \in \Bbbk \setminus \{0\}$.
(c) Tipo III: Sumar a la fila i-ésima de A su fila l-ésima multiplicada por $\lambda \in \Bbbk$.

Las operaciones elementales por filas en una matriz $A \in \mathcal{M}_{m\times n}(\Bbbk)$ producen matrices equivalentes a A. En efecto, a cada una de las operaciones elementales por filas le corresponden un par de matrices invertibles $P \in \mathcal{M}_n(\Bbbk)$ y $Q \in \mathcal{M}_m(\Bbbk)$ tales que el resultado de la operación elemental es $Q^{-1}AP$:

(a) Tipo I: Intercambiar las filas i-ésima y l-ésima de A se consigue tomando Q igual a la matriz T_{il} que se obtiene al permutar las filas i-ésima y l-ésima de la matriz identidad de orden m y P igual a la matriz identidad de orden n (compruébese usando el ejercicio 1 ajustado a la igualdad $I_n\, A = A$).
(b) Tipo II: Multiplicar la fila i-ésima de A por $\lambda \in \Bbbk \setminus \{0\}$ se consigue tomando Q igual a la matriz $M_i(\frac{1}{\lambda})$ que se obtiene al multiplicar la fila i-ésima de la matriz identidad de orden m por $1/\lambda$ y P igual a la matriz unida de orden n (compruébese usando el ejercicio 1 ajustado a la igualdad $I_n\, A = A$).
(c) Tipo III: Sustituir la fila i-ésima de A por ella misma más $\lambda \in \Bbbk$ veces su fila l-ésima se consigue tomando Q igual a la matriz $S_{il}(-\lambda)$ que se obtiene al sustituir por $-\lambda$ la entrada (i,l)-ésima de la matriz identidad de orden m y P igual a la matriz identidad de orden n (compruébese usando el ejercicio 1 ajustado a la igualdad $I_n\, A = A$).

Las matrices T_{il}, $M_i(\lambda)$ con $\lambda \in \Bbbk \setminus \{0\}$ y $S_{il}(\lambda)$ con $\lambda \in \Bbbk$ se llaman **matrices elementales**.

En el ejercicio 2 puedes encontrar algunas interesantes propiedades de las matrices elementales.

Nota II.1.3. Nótese que en las operaciones elementales por filas la matriz P siempre es la identidad del orden correspondiente.

Definición II.1.4. A las matrices que son producto de matrices de la forma T_{il} se les llama **matrices de permutación**.

Obsérvese que las matrices de permutación son ortogonales (véase el apartado 1. del ejercicio 2).

Al igual que hemos definido las operaciones elementales por filas en una matriz, se pueden definir **operaciones elementales por columnas** en una matriz de forma totalmente análoga, lo que proponemos como ejercicio al lector.

Teorema II.1.5. ***Forma reducida por filas.***
Sea $A \in \mathcal{M}_{m\times n}(\Bbbk)$ no nula. Mediante operaciones elementales por filas y, si es necesario, permutando las columnas de A, se puede obtener una matriz A' equivalente a A de la forma:

$$A' = \begin{pmatrix} 1 & 0 & \dots & 0 & a'_{1\,r+1} & \dots & a'_{1n} \\ 0 & 1 & \dots & 0 & a'_{2\,r+1} & \dots & a'_{2n} \\ \vdots & \vdots & \ddots & \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 1 & a'_{r\,r+1} & \dots & a'_{rn} \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 0 & 0 & 0\dots & 0 & 0 & \dots & 0 \end{pmatrix}, \tag{II.1.1}$$

*La matriz A' se llama **forma reducida por filas de** A y es única salvo permutación de las últimas $n-r$ columnas.*

Demostración. Si las entradas de la primera columna de A son todas 0, pasamos la primera columna al lugar n-ésimo En otro caso, hay alguna entrada no nula, que colocamos en lugar $(1,1)$ mediante una operación del tipo I. Con una operación del tipo II conseguimos que esta entrada sea 1 y con operaciones del tipo III se puede conseguir que las entradas $(i,1)$-ésimas sean 0, para cada $i = 2,\dots,m$. La primera columna queda, por tanto, en la forma buscada. Supongamos que tenemos h columnas en la forma deseada. Si en la columna $(h+1)$-ésima las entradas de las filas $h+1,\dots,m$ son 0, la situamos (mediante operación por columnas del tipo I) en el lugar n. En caso contrario, alguna de las entradas de las filas $h+1,\dots,m$ en la columna $h+1$-ésima es distinta de 0; haciendo una operación del tipo I lo emplazamos al lugar $(h+1,h+1)$; con una operación del tipo II conseguimos que esta entrada sea 1 y con

operaciones del tipo III hacemos ceros en las entradas $(i, h+1)$-ésimas, para cada $i = h+2, \ldots, m$. Observamos que las columnas anteriores no varían. Continuando con este mismo proceso conseguimos una matriz de la forma (II.1.1).

La unicidad es una consecuencia del siguiente resultado:

Lema II.1.6. *Sean A y $B \in \mathcal{M}_{m\times n}(\Bbbk)$ dos matrices en forma reducida por filas. Si existe $P \in \mathcal{M}_m(\Bbbk)$ invertible tal que $P^{-1}A = B$, entonces $A = B$.*

Demostración. Veámoslo por inducción sobre el número de columnas n. Para $n = 1$, si $A = 0$ entonces, al ser $P^{-1}A = B$, ha de ser forzosamente $B = 0$. Si A y B son no nulas, entonces

$$A = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = B.$$

Supongamos ahora que el enunciado es cierto para matrices de orden $m \times (n-1)$ y comprobémoslo para matrices de orden $m \times n$. Llamemos A_1 y $B_1 \in \mathcal{M}_{m\times(n-1)}(\Bbbk)$ a las submatrices de A y B que se obtienen al eliminar la última columna. Es claro, que las matrices A_1 y B_1 están en forma reducida por filas. Además, como $P^{-1}A = B$, se tiene que $P^{-1}A_1 = B_1$. Por tanto, aplicando la hipótesis de inducción se concluye que $A_1 = B_1$. Queda comprobar que también las últimas columnas de A y B son iguales.

Si la última columna de A es

$$\begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \leftarrow r\text{-ésimo}$$

y A_1 tiene sus $m - r + 1$ últimas filas nulas, entonces A y B son necesariamente iguales; de hecho, en este caso, se tiene que $r = n$ y

$$A = B = \left(\frac{I_n}{0}\right).$$

Supongamos, pues, que A_1 (y por lo tanto B_1) tiene sus r primeras filas no nulas y las $m-r$ últimas filas nulas, y que las últimas columnas de A y B son

$$\mathbf{a}_n = \begin{pmatrix} a_{1n} \\ \vdots \\ a_{rn} \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad \text{y} \quad \mathbf{b}_n = \begin{pmatrix} b_{1n} \\ \vdots \\ b_{rn} \\ b_{r+1\,n} \\ \vdots \\ b_{mn} \end{pmatrix},$$

respectivamente.

Teniendo ahora en cuenta que $P^{-1}(A_1|\mathbf{a}_n) = P^{-1}A = B = (B_1|\mathbf{b}_n)$ y que

$$A_1 = B_1 = \left(\begin{array}{c|c} I_r & C \\ \hline 0 & 0 \end{array}\right)$$

y que se sigue que $P^{-1}\mathbf{a}_n = \mathbf{b}_n$ y que

$$P^{-1} = \left(\begin{array}{c|c} I_r & P_1 \\ \hline 0 & P_2 \end{array}\right),$$

de donde se deduce fácilmente que $\mathbf{a}_n = \mathbf{b}_n$. ∎

Retornando ahora a la unicidad de la forma reducida por filas de A, basta tener en cuenta que si A'' es otra matrices en forma reducida obtenida a partir de A mediante operaciones elementales por filas y permutaciones de columnas, existen una matriz invertible $P \in \mathcal{M}_m(\Bbbk)$ y una matriz de permutación $Q \in \mathcal{M}_n(\Bbbk)$ tales que $P^{-1}A'Q = A''$. En primer lugar, observamos que $B = A'Q$ está en forma reducida por filas[1]. Por consiguiente, usando el lema anterior concluimos que $A'Q = B = A''$. Además, las permutaciones recogidas en Q sólo pueden afectar a las últimas $n-r$ columnas de A', al ser ésta y A'' matrices en forma reducida por filas. ∎

[1]Según hemos visto en la primera parte de la demostración se realizan permutaciones de columnas cuando la matriz no está en forma reducida y en la columna $(h+1)$-ésima las entradas de las filas $h+1, \ldots, m$ son cero.

Es claro que intercambiando filas por columnas y viceversa en el teorema anterior, se obtiene que la matriz A es equivalente a una de la forma

$$A'' = \begin{pmatrix} 1 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 1 & 0 & \dots & 0 \\ a''_{s+1\,1} & a''_{s+1,\,2} & \dots & a''_{s+1\,s} & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ a''_{m\,1} & a''_{m\,2} & \dots & a''_{m\,s} & 0 & \dots & 0 \end{pmatrix}, \tag{II.1.2}$$

que se llama **forma reducida por columnas de** A y es única salvo permutación de las últimas $m - s$ filas.

Nota II.1.7. Obsérvese que la demostración del teorema II.1.5 proporciona un procedimiento algorítmico para calcular la forma reducida por filas (o por columnas, con las modificaciones pertinentes) de una matriz dada. Este procedimiento se llama **método de Gauss-Jordan.**

Por otra parte, si en el teorema II.1.5 prescindimos de las permutaciones de las columnas, no se obtiene la forma reducida por filas (al menos como la nosotros la hemos definido); sin embargo, se obtiene una matriz en **forma escalonada por filas**. Y lo mismo ocurre si prescindimos de las permutaciones de filas cuando se construye la forma reducida por columnas; en cuyo caso, la matriz que se obtiene estará en **forma escalonada por columnas**.

Corolario II.1.8. *Sea $A \in \mathcal{M}_{m\times n}(\Bbbk)$. Si A' y $A'' \in \mathcal{M}_{m\times n}$ son las formas reducidas por filas y por columnas de A, respectivamente, entonces existe un único entero $r \geq 0$ tal que la forma reducida por columnas de A' y la forma reducida por filas de A'' coinciden con*

$$R = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix},$$

donde I_r es la matriz identidad de orden r y el resto son matrices nulas de los ordenes correspondientes. Esta matriz se llama ***forma reducida*** *de A.*

Del corolario anterior se deduce que el número de filas distintas de cero de la forma reducida por filas de una matriz dada es igual al número de columnas distintas de cero de la forma reducida por columnas de la misma matriz. Además, de la unicidad de las formas reducidas por filas y por columnas se sigue la unicidad de r.

Definición II.1.9. Sea $A \in \mathcal{M}_{m\times n}(\Bbbk)$. Se llama **rango de la matriz** A al número de filas (o columnas) distintas de cero en su forma reducida, y se denota $\mathrm{rg}(A)$.

Proposición II.1.10. *Dos matrices A y $B \in \mathcal{M}_{m\times n}(\Bbbk)$ son equivalentes si, y sólo si, tienen el mismo rango.*

Demostración. Si A y B son equivalentes, entonces tienen la misma forma reducida por filas, de donde se sigue que $\mathrm{rg}(A) = \mathrm{rg}(B)$.

Recíprocamente, si A y B tienen el mismo rango, existen P_1 y $P_2 \in \mathcal{M}_n(\Bbbk)$ y Q_1 y $Q_2 \in \mathcal{M}_m(\Bbbk)$ tales que

$$Q_1^{-1}A(P_1) = Q_2^{-1}B(P_2) = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}$$

(véase el corolario II.1.8), de donde se sigue que $B = Q_2(Q_1^{-1}A(P_1))P_2^{-1}$, es decir,

$$B = (Q_1Q_2^{-1})^{-1}A(P_1P_2^{-1}).$$

Luego, A y B son equivalentes. ∎

Nota II.1.11. **Cálculo de las matrices de paso para obtener la forma reducida:** Sea $A \in \mathcal{M}_{m\times n}(\Bbbk)$ tal que $\mathrm{rg}(A) = r$ y sean $P \in \mathcal{M}_n(\Bbbk)$ y $Q \in \mathcal{M}_m(\Bbbk)$ las matrices invertibles tales que

$$Q^{-1}AP = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix},$$

entonces:

i) Q^{-1} es la matriz que resulta de hacer en I_m (la matriz identidad de orden m) las mismas transformaciones elementales por filas que se hacen en A para llegar a la forma reducida,

$$Q^{-1} = \ldots \cdot (2^a \text{ t.f.}) \cdot (1^a \text{ t.f.}),$$

donde (1^a t.f.) denota a la matriz elemental de la primera transformación elemental por filas, (2^a t.f.) a la matriz elemental de la segunda transformación elemental por filas, ...

ii) P es la matriz que resulta de hacer en I_n (la matriz identidad de orden n) las mismas transformaciones elementales por columnas que se hacen en A para llegar a la forma reducida,

$$P = (1^a\text{t.c.}) \cdot (2^a\text{t.c.}) \cdot \ldots$$

donde (1^a t.c.) denota a la matriz elemental de la primera transformación elemental por columnas, (2^a t.c.) a la matriz elemental de la segunda transformación elemental por columnas, ...

2. Aplicaciones lineales

En lo que sigue, y a lo largo de todo esta sección, V y V' denotarán dos espacios vectoriales sobre un mismo cuerpo $\Bbbk$.

Definición II.2.1. Se dice que una aplicación $T : V \longrightarrow V'$ es un **morfismo de $\Bbbk$-espacios vectoriales** (ó **aplicación $\Bbbk$-lineal** ó **aplicación lineal** si es claro que el cuerpo es $\Bbbk$), si es un morfismo de grupos compatible con el producto por escalares, es decir, si verifica:

(a) $T(\mathbf{u}+\mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$ (morfismo de grupos);
(b) $T(\lambda\mathbf{u}) = \lambda T(\mathbf{u})$ (compatible con el producto por escalares),

para todo $\mathbf{u}$ y $\mathbf{v} \in V$ y $\lambda \in \Bbbk$.

Equivalentemente (compruébese), T es morfismo de $\Bbbk$-espacios vectoriales si, y sólo si, es compatible con combinaciones lineales, es decir, $T(\lambda\mathbf{u} + \mu\mathbf{v}) = \lambda T(\mathbf{u}) + \mu T(\mathbf{v})$, para todo $\mathbf{u}$ y $\mathbf{v} \in V$ y λ y $\mu \in \Bbbk$.

Nota II.2.2. Observése que, en general, se tiene que si $T : V \longrightarrow V'$ es aplicación lineal, entonces

$$T\left(\sum_{i=1}^{r} \lambda_i \mathbf{v}_i\right) = \sum_{i=1}^{r} \lambda_i T(\mathbf{v}_i),$$

para todo $\mathbf{v}_i \in V$ y $\lambda_i \in \Bbbk,\ i = 1, \ldots, r$.

Ejemplo II.2.3. Veamos los ejemplos más sencillos de aplicaciones lineales.

1. Sea $T : V \longrightarrow V'$ la aplicación definida por $T(\mathbf{v}) = \mathbf{0}_{V'}$, para todo $\mathbf{v} \in V$. Esta aplicación es lineal y se llama **aplicación trivial** o **nula**.
2. Si denotamos, como es usual, con $\mathbf{0}$ al $\Bbbk$-espacio vectorial cuyo único vector es el cero, entonces es claro que la única aplicación lineal de $\mathbf{0}$ a V es la aplicación nula, la cual, denotaremos por $\mathbf{0} \longrightarrow V$. Del mismo modo, la única aplicación lineal de V en $\mathbf{0}$ es la aplicación nula, que denotaremos por $V \longrightarrow \mathbf{0}$.
3. Si $L \subseteq V$ es un subespacio vectorial de V, entonces la aplicación $i : L \hookrightarrow V$ definida por $i(\mathbf{v}) = \mathbf{v}$, para todo $\mathbf{v} \in L$, es lineal y se llama **inclusión de L en V**. En el caso particular, en que $L = V$, la aplicación anterior se llama **identidad de V** y se denota por Id_V.

Definición II.2.4. Diremos que una aplicación lineal es un **monomorfismo** (**epimorfismo, isomorfismo**, respectivamente) cuando sea inyectiva (epiyectiva, biyectiva, respectivamente).

Cuando una T aplicación lineal está definida en V y valora también en V, esto es, $T : V \longrightarrow V$, se dice que es un **endomorfismo** (de V); los endomorfismos (de V) que son isomorfismos se denominan **automorfismos** (de V).

Dados dos espacios vectoriales V y V' sobre un mismo cuerpo $\Bbbk$, denotaremos por $\mathrm{Hom}_{\Bbbk}(V, V')$ al conjunto de todas aplicaciones $\Bbbk$-lineales de V en V'. El conjunto formado por las aplicaciones lineales de V en V, es decir, por los endomorfismos de V, se denota por $\mathrm{End}_{\Bbbk}(V)$. Es un sencillo ejercicio comprobar que $\mathrm{Hom}_{\Bbbk}(V, V')$ y $\mathrm{End}_{\Bbbk}(V)$ son espacios vectoriales sobre $\Bbbk$ con la suma y producto por escalares usuales de las aplicaciones, es decir, $f + g$ es la aplicación tal que $(f + g)(\mathbf{v}) = f(\mathbf{v}) + g(\mathbf{v})$ y (λf) es la aplicación tal que $(\lambda f)(\mathbf{v}) = \lambda f(\mathbf{v})$, para todo $\mathbf{v} \in V$.

Proposición II.2.5. *Si $T : V \longrightarrow V'$ es un isomorfismo, entonces $T^{-1} : V' \longrightarrow V'$ es un isomorfismo.*

Demostración. Como T es biyectiva, T^{-1} también es biyectiva, por tanto, sólo hay que probar que T^{-1} es lineal. Sean $\mathbf{u}'$ y $\mathbf{v}' \in V'$ y λ y $\mu \in \Bbbk$. Por ser T biyectiva, existen unos únicos $\mathbf{u}$ y $\mathbf{v} \in V$ tales que $T(\mathbf{u}) = \mathbf{u}'$ y $T(\mathbf{v}) = \mathbf{v}'$. Además, por ser T lineal, $T(\lambda\mathbf{u} + \mu\mathbf{v}) = \lambda T(\mathbf{u}) + \mu T(\mathbf{v}) = \lambda\mathbf{u}' + \mu\mathbf{v}'$. De ambos hechos se deduce que

$$T^{-1}(\lambda\mathbf{u}' + \mu\mathbf{v}') = \lambda\mathbf{u} + \mu\mathbf{v} = \lambda T^{-1}(\mathbf{u}') + \mu T^{-1}(\mathbf{v}'),$$

y por tanto que T^{-1} es lineal. ∎

Esta última proposición dota de sentido a la siguiente definición.

Definición II.2.6. Diremos que los espacios vectoriales V y V' son isomorfos si existe algún isomorfismo entre ellos, en cuyo caso escribiremos $V \cong V'$ (ó $V \xrightarrow{\sim} V'$).

Ejercicio II.2.7. Probar que la composición de aplicaciones es una aplicación lineal. Probar que "ser isomorfos", $\cong$, es una relación de equivalencia.

Como todo morfismo de $\Bbbk$-espacios vectoriales es, en particular, un morfismo de grupos, tenemos las siguientes propiedades elementales.

Proposición II.2.8. *Si $T : V \longrightarrow V'$ es una aplicación lineal, entonces se cumple que:*

(a) $T(\mathbf{0}_V) = \mathbf{0}_{V'}$;
(b) $T(-\mathbf{v}) = -T(\mathbf{v})$;
(c) $T(\mathbf{v} - \mathbf{u}) = T(\mathbf{v}) - T(\mathbf{u})$,

para todo $\mathbf{v}$ y $\mathbf{u} \in V$.

Demostración. (a) Sea $\mathbf{v} \in V$. Como $T(\mathbf{v}) = T(\mathbf{v} + \mathbf{0}_V) = T(\mathbf{v}) + T(\mathbf{0}_V)$, de la unicidad del elemento neutro en V' se sigue que $T(\mathbf{0}_V) = \mathbf{0}_{V'}$.

(b) Basta tomar $\lambda = 1$ en el apartado (b) de la definición de aplicación lineal (definición II.2.1).

(c) $T(\mathbf{u} - \mathbf{v}) = T(\mathbf{u}) + T(-\mathbf{v}) = T(\mathbf{u}) - T(\mathbf{v})$. ∎

Definición II.2.9. Sea $T : V \longrightarrow V'$ una aplicación lineal. Se llama **núcleo** de T al subconjunto $\ker(T) := \{\mathbf{v} \in V \mid T(\mathbf{v}) = \mathbf{0}_{V'}\} \subseteq V$. Se llama **imagen** de T al subconjunto $\mathrm{Im}(T) := \{T(\mathbf{v}) \mid \mathbf{v} \in V\} \subseteq V'$.

Nota II.2.10. Obsérvese que $\mathrm{Im}(T)$ coincide con el siguiente subconjunto de V',

$$\{\mathbf{v}' \in V' \mid \text{ existe } \mathbf{v} \in V \text{ con } T(\mathbf{v}) = \mathbf{v}'\}.$$

Ejemplo II.2.11. Calculemos el núcleo y la imagen para las aplicaciones lineales del ejemplo II.2.3

1. Si $T : V \longrightarrow V'$ es la aplicación nula, entonces $\ker(T) = V$ e $\mathrm{Im}(T) = \{\mathbf{0}_{V'}\}$.
2. El núcleo y la imagen de la aplicación $\mathbf{0} \longrightarrow V$ son, obviamente, $\{\mathbf{0}\}$ y $\{\mathbf{0}_V\}$, respectivamente. También es claro que el núcleo y la imagen de la aplicación $V \longrightarrow \mathbf{0}$ son V y $\{\mathbf{0}\}$, respectivamente.
3. Sean $L \subseteq V$ es un subespacio vectorial. Si $i : L \hookrightarrow V$ es la inclusión de L en V, entonces $\ker(i) = \{\mathbf{0}_V\}$ e $\mathrm{Im}(i) = L$, y si $\mathrm{Id}_V : V \longrightarrow V$ es la identidad de V, entonces $\ker(\mathrm{Id}_V) = \{\mathbf{0}_V\}$ e $\mathrm{Im}(\mathrm{Id}_V) = V$.
4. Sea $h_\lambda : V \longrightarrow V$ la homotecia lineal de razón $\lambda \in \Bbbk$. Si $\lambda = 0$, entonces h_λ es la aplicación nula, en otro caso, $\ker(h_\lambda) = \{\mathbf{0}_V\}$ e $\mathrm{Im}(h_\lambda) = V$.

Nótese que en los ejemplos anteriores tanto el núcleo como la imagen son subespacios vectoriales. Veamos que esto no es un hecho aislado y se cumple siempre.

Proposición II.2.12. *Si $T : V \longrightarrow V'$ es una aplicación lineal, entonces*

(a) $\ker(T)$ *es un subespacio vectorial de V.*
(b) $\mathrm{Im}(T)$ *es un subespacio vectorial de V'.*

Demostración. (a) Por la proposición II.2.8(a), tenemos que $T(\mathbf{0}_V) = \mathbf{0}_{V'}$, es decir, $\mathbf{0}_V \in \ker(T)$ y por tanto podemos asegurar que $\ker(T)$ es un subconjunto no vacío de V.

Si $\mathbf{u}$ y $\mathbf{v} \in \ker(T)$ y λ y $\mu \in \Bbbk$, entonces

$$T(\lambda\mathbf{u} + \mu\mathbf{v}) \overset{T \text{ lineal}}{=} \lambda T(\mathbf{u}) + \mu T(\mathbf{v}) \overset{\mathbf{u},\mathbf{v} \in \ker(T)}{=} \lambda\mathbf{0}_{V'} + \mu\mathbf{0}_{V'} = \mathbf{0}_{V'}.$$

Por la proposición C.2.3, $\ker(T)$ es subespacio vectorial de V.

(b) Por la proposición II.2.8(a), tenemos que $T(\mathbf{0}_V) = \mathbf{0}_{V'}$, es decir, $\mathbf{0}_{V'} \in \mathrm{Im}(T)$ y, por tanto, que $\mathrm{Im}(T)$ es un subconjunto no vacío de V'.

Si $\mathbf{u}'$ y $\mathbf{v}' \in \mathrm{Im}(T)$, entonces existen $\mathbf{u}$ y $\mathbf{v} \in V$ tales que $T(\mathbf{u}) = \mathbf{u}'$ y $T(\mathbf{v}) = \mathbf{v}'$. De tal forma que si λ y $\mu \in \Bbbk$, tenemos que

$$\lambda\mathbf{u}' + \mu\mathbf{v}' = \lambda T(\mathbf{u}) + \mu T(\mathbf{v}) \overset{T \text{ lineal}}{=} T(\lambda\mathbf{u} + \mu\mathbf{v}).$$

Luego $\lambda\mathbf{u}' + \mu\mathbf{v}' \in \mathrm{Im}(T)$ y, por consiguiente, $\mathrm{Im}(T)$ es subespacio vectorial de V. ∎

Es claro que, por definición, tenemos que una aplicación $T : V \longrightarrow V'$ es epiyectiva si, y sólo si, la imagen de T es V'. De modo que podemos determinar cuándo una aplicación es epimorfismo dependiendo de su imagen. Veamos que el núcleo caracteriza a los monomorfismos.

Proposición II.2.13. *Sea $T : V \longrightarrow V'$ una aplicación lineal. T es inyectiva, es decir, es un monomorfismo si, y sólo si,* $\ker(T) = \{\mathbf{0}_V\}$.

Demostración. $\boxed{\Rightarrow}$ Sea $\mathbf{v} \in \ker(T)$, entonces, por T inyectiva tenemos que $T(\mathbf{v}) = \mathbf{0}_{V'} = T(\mathbf{0}_V)$ implica $\mathbf{v} = \mathbf{0}_V$.

$\boxed{\Leftarrow}$ Si $\mathbf{u}$ y $\mathbf{v}$ son vectores de V tales que $T(\mathbf{u}) = T(\mathbf{v})$, entonces

$$\mathbf{0}_{V'} = T(\mathbf{u}) - T(\mathbf{v}) \overset{T \text{ lineal}}{=} T(\mathbf{u} - \mathbf{v}).$$

Luego $\mathbf{u} - \mathbf{v} \in \ker(T) = \{\mathbf{0}_V\}$, de donde se sigue que $\mathbf{u} - \mathbf{v} = \mathbf{0}_V$, es decir, $\mathbf{u} = \mathbf{v}$. ∎

De forma inmediata tenemos el siguiente:

Corolario II.2.14. *Sea $T : V \longrightarrow V'$ una aplicación lineal. T es isomorfismo si, y sólo si,* $\ker(T) = \{\mathbf{0}_V\}$ *e* $\mathrm{Im}(T) = V'$.

3. Matriz asociada a una aplicación lineal

Sea $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ es una base de un $\Bbbk$-espacio vectorial V de dimensión finita $n > 0$.

Sabemos, que todo vector $\mathbf{v} \in V$ se expresa de forma única como combinación lineal de los vectores de $\mathcal{B}$; es decir, existen unos únicos $\lambda_1, \ldots, \lambda_n \in \Bbbk$ tales que $\mathbf{v} = \lambda_1\mathbf{v}_1 + \ldots + \lambda_n\mathbf{v}_n$, llamados **coordenadas de $\mathbf{v} \in V$ respecto de $\mathcal{B}$**.

Por otra parte, existe una única aplicación lineal

$$\varphi_{\mathcal{B}} : V \longrightarrow \Bbbk^n; \quad \varphi_{\mathcal{B}}(\mathbf{v}_i) = \mathbf{e}_i := (0, \ldots, 0, \overset{i)}{1}, 0, \ldots, 0), \quad i = 1, \ldots, n.$$

De hecho esta aplicación es un isomorfismo de V en $\Bbbk^n$ que "manda" un vector $\mathbf{v} \in V$ de coordenadas $\lambda_1, \ldots, \lambda_n$ respecto de $\mathcal{B}$ a la n-upla $(\lambda_1, \ldots, \lambda_n) \in \Bbbk^n$. De aquí que, en lo sucesivo, denotaremos a las coordenadas de $\mathbf{v} \in V$ respecto $\mathcal{B}$ por la n-upla correspondiente en $\Bbbk^n$, es decir, escribiremos $(\lambda_1, \ldots, \lambda_n)$ (ó $(\lambda_1, \ldots, \lambda_n)_{\mathcal{B}}$ si queremos destacar la base) para expresar las coordenadas de $\mathbf{v}$ respecto de $\mathcal{B}$.

Nota II.3.1. Mediante el isomorfismo anterior podemos ver cualquier espacio vectorial V de dimensión n como un espacio vectorial numérico de dimensión n, esto es, $\Bbbk^n$. Sin embargo, es conveniente resaltar que esta identificación depende de la base de V elegida, y por lo tanto que, en algunos casos, se puede perder generalidad en los razonamientos.

Una vez fijada la notación que usaremos de esta sección en adelante, pasamos a definir la matriz asociada a una aplicación lineal.

En lo que sigue V y V' serán dos $\Bbbk$-espacios vectoriales de dimensiones finitas $n > 0$ y $m > 0$, respectivamente, $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ una base de V y $\mathcal{B}' = \{\mathbf{v}'_1, \ldots, \mathbf{v}'_m\}$ una base de V'.

Si $T \in \mathrm{Hom}_{\Bbbk}(V, V')$, entonces es claro que existen $a_{ij} \in \Bbbk$ con $i \in \{1, \ldots, m\}$ y $j \in \{1, \ldots, n\}$ tales que

$$T(\mathbf{v}_j) = \sum_{i=1}^{m} a_{ij}\mathbf{v}'_i,$$

es decir, tales que las coordenadas de $T(\mathbf{v}_j) \in V'$ respecto de $\mathcal{B}'$ son $(a_{1j}, \ldots, a_{mj})$, para cada $j = 1, \ldots, m$. Además, T está determinado por las imágenes de una base de V. Luego tenemos que T "está determinado por las coordenadas" de $T(\mathbf{v}_j)$, $j = 1, \ldots, n$, respecto de $\mathcal{B}'$, aunque obviamente estas coordenadas dependen de las bases $\mathcal{B}$ y $\mathcal{B}'$ elegidas.

Definición II.3.2. Dado $T \in \mathrm{Hom}_{\Bbbk}(V, V')$ se define la **matriz asociada a T respecto de la bases $\mathcal{B}$ y $\mathcal{B}'$**, $M_{\mathcal{B},\mathcal{B}'}(T)$, como la matriz $A = (a_{ij}) \in \mathcal{M}_{m\times n}(\Bbbk)$ cuya columna j-ésima son las coordenadas de $T(\mathbf{v}_j)$ respecto de $\mathcal{B}$, es decir,

$$M_{\mathcal{B},\mathcal{B}'}(T) = \begin{array}{c} T(\mathbf{v}_1)\ T(\mathbf{v}_2) \ldots T(\mathbf{v}_n) \\ \left(\begin{array}{cccc} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{array}\right) \end{array} \begin{array}{c} \\ \mathbf{v}'_1 \\ \mathbf{v}'_2 \\ \vdots \\ \mathbf{v}'_m \end{array}$$

Cuando $V' = V$ y $\mathcal{B}' = \mathcal{B}$, se dice que $M_{\mathcal{B},\mathcal{B}'}(T)$ es la **matriz de T respecto de $\mathcal{B}$** y se escribe $M_{\mathcal{B}}(T)$.

La matriz asociada a una aplicación lineal permite obtener una expresión matricial que relaciona las coordenadas de un vector de V respecto de $\mathcal{B}$ con las coordenadas de su imagen por T respecto de $\mathcal{B}'$.

Proposición II.3.3. *Sean $T \in \mathrm{Hom}_{\Bbbk}(V, V')$ y $A = (a_{ij}) \in \mathcal{M}_{m\times n}(\Bbbk)$ la matriz asociada a T respecto de las bases $\mathcal{B}$ y $\mathcal{B}'$. Si $(x_1, x_2, \ldots, x_n)$ son las coordenadas de un vector $\mathbf{v} \in V$, entonces se cumple que $(x'_1, x'_2, \ldots, x'_m)$ son las coordenadas de $T(\mathbf{v})$ respecto de $\mathcal{B}'$ si, y sólo si,*

$$\left(\begin{array}{cccc} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{array}\right)\left(\begin{array}{c} x_1 \\ x_2 \\ \vdots \\ x_n \end{array}\right) = \left(\begin{array}{c} x'_1 \\ x'_2 \\ \vdots \\ x'_n \end{array}\right). \tag{II.3.3}$$

A la expresión (II.3.3) se la llama **ecuaciones de T respecto de $\mathcal{B}$ y $\mathcal{B}'$.**

Demostración. Si $\mathbf{v}' = \sum_{i=1}^{m} x_i' \mathbf{v}_i' \in V'$, entonces $T(\mathbf{v}) = \mathbf{v}'$ si, y sólo si,

$$\sum_{i=1}^{m} x_i' \mathbf{v}_i' = T\Big(\sum_{j=1}^{n} x_j \mathbf{v}_j\Big) = \sum_{j=1}^{n} x_j T\Big(\mathbf{v}_j\Big) = \sum_{j=1}^{n} x_j \Big(\sum_{i=1}^{m} a_{ij} \mathbf{v}_i'\Big) = \sum_{j=1}^{n} \Big(\sum_{i=1}^{m} x_j a_{ij}\Big) \mathbf{v}_i'$$

si, y sólo si, $x_i' = \sum_{i=1}^{m} x_i a_{ij},\ i = 1, \ldots, n$ si, y sólo si,

$$A \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} x_1' \\ x_2' \\ \vdots \\ x_n' \end{pmatrix}$$

■

El hecho de que a cada aplicación lineal se le asocie una matriz permite definir una aplicación de $\mathrm{Hom}_{\Bbbk}(V, V')$ en $\mathcal{M}_{m\times n}(\Bbbk)$ tal que a cada $T \in \mathrm{Hom}_{\Bbbk}(V, V')$ le asigna la matriz asociada a T respecto de las bases $\mathcal{B}$ y $\mathcal{B}'$ de V y V', respectivamente. Veamos que esta aplicación es un isomorfismo de espacios vectoriales.

Nota II.3.4. Recordemos que el conjunto de matrices de orden $m \times n$ con coeficientes en $\Bbbk$ tiene estructura de $\Bbbk$-espacio vectorial con la suma y producto por escalares habituales de matrices: $A+B = (a_{ij})+(b_{ij}) = (a_{ij}+b_{ij})$ y $\lambda A = \lambda(a_{ij}) = (\lambda a_{ij})$ con $A = (a_{ij})$ y $B = (a_{ij}) \in \mathcal{M}_{m\times n}(\Bbbk)$ y $\lambda \in \Bbbk$ (veánse la nota I.1.7 y el ejercicio 1). Además, la dimensión de $M_{m\times n}(\Bbbk)$ como $\Bbbk$-espacio vectorial es $m \cdot n$; pues una base de $\mathcal{M}_{m\times n}(\Bbbk)$ la forman las matrices $E_{ij} \in \mathcal{M}_{m\times n}(\Bbbk)$ con un 1 en el lugar (i, j)-ésimo y ceros en el resto.

Teorema II.3.5. *La aplicación $\phi : \mathrm{Hom}_{\Bbbk}(V, V') \longrightarrow \mathcal{M}_{m\times n}(\Bbbk)$ que a cada aplicación lineal $T : V \longrightarrow V'$ le hace corresponder su matriz asociada respecto de las bases $\mathcal{B}$ y $\mathcal{B}'$ es un isomorfismo de $\Bbbk$-espacios vectoriales.*

Demostración. La aplicación ϕ es lineal. En efecto, dados T y $S \in \mathrm{Hom}_{\Bbbk}(V, V')$ tenemos que existen $A = (a_{ij})$ y $B = (b_{ij}) \in \mathcal{M}_{m\times n}(\Bbbk)$ tales que $\phi(T) = A$ y $\phi(S) = B$. Luego $T(\mathbf{v}_j) = \sum_{i=1}^{m} a_{ij} \mathbf{v}_i'$ y $S(\mathbf{v}_j) = \sum_{i=1}^{m} b_{ij} \mathbf{v}_i'$, para $j \in \{1, \ldots, n\}$. Por consiguiente, si λ y $\mu \in \Bbbk$,

$$\begin{aligned}(\lambda T + \mu S)(\mathbf{v}_j) &= \lambda(T(\mathbf{v}_j)) + \mu(S(\mathbf{v}_j)) = \lambda(\sum_{i=1}^{m} a_{ij} \mathbf{v}_i') + \mu(\sum_{i=1}^{m} b_{ij} \mathbf{v}_i') \\ &= \sum_{i=1}^{m} (\lambda a_{ij} + \mu b_{ij}) \mathbf{v}_i',\end{aligned}$$

para cada $j \in \{1, \ldots, m\}$. De donde se sigue que la matriz asociada a $\lambda T + \mu S$ es $\lambda A + \mu B = (\lambda a_{ij} + \mu b_{ij})$, y por lo tanto que $\phi(\lambda T + \mu S) = \lambda\phi(T) + \mu\phi(S)$.

Por último, veamos que ϕ es biyectiva. Sea $A = (a_{ij}) \in \mathcal{M}_{m\times n}(\Bbbk)$. Para cada $j \in \{1, \ldots, n\}$ definimos $\mathbf{u}_j = a_{1j}\mathbf{v}'_1 + \ldots + a_{mj}\mathbf{v}'_m \in \Bbbk$. Es claro que existe una única aplicación lineal $T \in \mathrm{Hom}_{\Bbbk}(V, V')$ tal que $T(\mathbf{v}_j) = \mathbf{u}_j,\ j = 1, \ldots, n$, y que $\phi(T) = A$. Esto prueba que ϕ es epiyectiva, y además, al ser T única, tenemos que ϕ es inyectiva. ∎

Probemos ahora que la composición de aplicaciones lineales (cuando tenga sentido) corresponde al producto de matrices. Para ello consideramos un tercer $\Bbbk$-espacio vectorial V'' de dimensión finita y una base $\mathcal{B}'' = \{\mathbf{v}''_1, \ldots, \mathbf{v}''_p\}$ de V''.

Proposición II.3.6. *Sean $T : V \longrightarrow V'$ y $S : V' \longrightarrow V''$ dos aplicaciones lineales. Si $A = (a_{ij}) \in \mathcal{M}_{m\times n}$ es la matriz asociada a T respecto de $\mathcal{B}$ y $\mathcal{B}'$ y $B = (b_{li}) \in \mathcal{M}_{p\times m}$ es la matriz S respecto de $\mathcal{B}'$ y $\mathcal{B}''$, entonces $C = B \cdot A$ es la matriz asociada a $S \circ T$ respecto de $\mathcal{B}$ y $\mathcal{B}''$.*

Demostración. Para cada $j \in \{1, \ldots, n\}$ tenemos que

$$\begin{aligned} S \circ T(\mathbf{v}_j) &= S(T(\mathbf{v}_j)) = S\big(\textstyle\sum_{i=1}^m a_{ij}\mathbf{v}'_i\big) = \sum_{i=1}^m a_{ij} S(\mathbf{v}'_i) \\ &= \textstyle\sum_{i=1}^m a_{ij}\big(\sum_{l=1}^p b_{li}\mathbf{v}''_l\big) = \sum_{l=1}^p \big(\sum_{i=1}^m b_{li}a_{ij}\big)\mathbf{v}''_l \end{aligned}.$$

De donde sigue que la matriz asociada a $S \circ T$ es $C = \sum_{i=1}^m b_{li}a_{ij} \in \mathcal{M}_{p\times n}(\Bbbk)$. Por la definición de producto de matrices, concluimos que $C = B \cdot A$. ∎

A continuamos veremos una caracterización de los automorfismos de un espacio vectorial de dimensión finita en términos de su matriz asociada.

Corolario II.3.7. *Sea V un $\Bbbk$-espacio vectorial de dimensión finita, $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ una base de V y $T \in \mathrm{End}_{\Bbbk}(V)$. Si A es la matriz asociada a T respecto de $\mathcal{B}$, entonces T es un automorfismo si, y sólo si, A es invertible, en cuyo caso, la matriz asociada a T^{-1} respecto de $\mathcal{B}$ es A^{-1}.*

Demostración. Basta tener en cuenta que $T \in \mathrm{End}_{\Bbbk}(V)$ es un automorfismo si, y sólo si, $T : V \longrightarrow V$ es una aplicación lineal biyectiva si, y sólo si, existe $T^{-1} \in \mathrm{End}_{\Bbbk}(V)$ tal que $T \circ T^{-1} = T^{-1} \circ T = \mathrm{Id}_V$ si, y sólo si, por la proposición II.3.6, $A \cdot B = B \cdot A = I_n$, donde $B \in \mathcal{M}_n(\Bbbk)$ es la matriz asociada a T^{-1} respecto de $\mathcal{B}$ si, y sólo si, A es invertible y $B = A^{-1}$ es la matriz asociada a T^{-1} respecto de $\mathcal{B}$. ∎

4. Cambios de bases. Teorema del rango

Sabemos que si V un $\Bbbk$-espacio vectorial de dimensión finita $n > 0$ y $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ es una base de V, para cada un vector $\mathbf{v} \in V$, existe un vector de $\Bbbk^n$ que llamamos coordenadas de $\mathbf{v}$ respecto de $\mathcal{B}$. Si $\mathcal{B}' = \{\mathbf{v}'_1, \ldots, \mathbf{v}'_n\}$ es otra base

de V nos preguntamos ahora qué relación existe entre las coordenadas de $\mathbf{v}$ respecto de $\mathcal{B}$ y su coordenadas respecto de $\mathcal{B}'$.

Definición II.4.1. Con la notación anterior, definimos la **matriz,** $M(\mathcal{B},\mathcal{B}')$, **del cambio de la base** $\mathcal{B}$ **a la base** $\mathcal{B}'$ como la matriz asociada al endomorfismo identidad de V respecto de las bases $\mathcal{B}$ y $\mathcal{B}'$, es decir, $M(\mathcal{B},\mathcal{B}') \in \mathcal{M}_n(\Bbbk)$ es la matriz cuya columna j-ésima corresponde a las coordenadas $\mathbf{v}_j$ respecto de $\mathcal{B}'$,

$$M(\mathcal{B},\mathcal{B}') = \begin{array}{c} \begin{array}{cccc} \mathbf{v}_1 & \mathbf{v}_2 & \dots & \mathbf{v}_n \end{array} \\ \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \end{array} \begin{array}{c} \\ \mathbf{v}'_1 \\ \mathbf{v}'_2 \\ \vdots \\ \mathbf{v}'_m \end{array}$$

Si convenimos que $\mathcal{B}$ es la "base antigua" y que $\mathcal{B}'$ es la "base nueva," entonces la matriz $M(\mathcal{B},\mathcal{B}')$ nos permite obtener las coordenadas de un vector $\mathbf{v} \in V$ respecto de la base nueva a partir de sus coordenadas respecto de la base antigua. Para ello, por la proposición II.3.3, basta considerar las ecuaciones de Id_V respecto de las bases $\mathcal{B}$ y $\mathcal{B}'$. Así, si las coordenadas de $\mathbf{v}$ respecto de $\mathcal{B}$ son $(\lambda_1,\dots,\lambda_n)$ y sus coordenadas respecto de $\mathcal{B}'$ son $(\lambda'_1,\dots,\lambda'_n)$, entonces

$$M(\mathcal{B},\mathcal{B}') \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix} = \begin{pmatrix} \lambda'_1 \\ \vdots \\ \lambda'_n \end{pmatrix}$$

Por otra parte, si consideramos la matriz $M(\mathcal{B}',\mathcal{B})$ del cambio de la base $\mathcal{B}'$ a la base $\mathcal{B}$, entonces, por la proposición II.3.6, $M(\mathcal{B}',\mathcal{B}) \cdot M(\mathcal{B},\mathcal{B}')$ ($M(\mathcal{B},\mathcal{B}') \cdot M(\mathcal{B}',\mathcal{B})$, respectivamente) es la matriz asociada al endomorfismo identidad de V respecto de la base $\mathcal{B}$ (respecto de la base $\mathcal{B}'$, respectivamente), es decir, $M(\mathcal{B}',\mathcal{B}) \cdot M(\mathcal{B},\mathcal{B}') = I_n$ ($M(\mathcal{B},\mathcal{B}') \cdot M(\mathcal{B}',\mathcal{B}) = I_n$), donde I_n es la matriz identidad de orden n. Resumiendo, la matriz $M(\mathcal{B},\mathcal{B}')$ es invertible y $M(\mathcal{B},\mathcal{B}')^{-1}$ es la matriz del cambio de la base $\mathcal{B}'$ a la base $\mathcal{B}$.

Una vez que hemos visto cómo afectan los cambios de bases a las coordenadas de un vector, nos interesa saber cómo cambia la matriz asociada a una aplicación lineal al cambiar las bases.

Si V y V' son dos $\Bbbk$-espacios vectoriales de dimensión finita, $\mathcal{B}_1$ es una base de V, $\mathcal{B}'_1$ es una base de V' y $T \in \mathrm{Hom}_{\Bbbk}(V,V')$, tenemos definida la matriz $M_{\mathcal{B}_1,\mathcal{B}'_1}(T)$ de T respecto de las bases $\mathcal{B}_1$ y $\mathcal{B}'_1$.

Consideremos ahora otras bases $\mathcal{B}_2$ y $\mathcal{B}'_2$ de V y V', respectivamente, y las matrices, $M(\mathcal{B}_2,\mathcal{B}_1)$ y $M(\mathcal{B}'_1,\mathcal{B}'_2)$, de cambio de la base $\mathcal{B}_2$ a la base $\mathcal{B}_1$ y de la base $\mathcal{B}'_1$ a la

base $\mathcal{B}'_2$, respectivamente. Teniendo en cuenta que $\mathrm{Id}_{V'} \circ T \circ \mathrm{Id}_V = T$, la proposición II.3.6 y el siguiente diagrama conmutativo,

$$\begin{array}{ccc} V & \xrightarrow{T} & V' \\ {\scriptstyle \mathrm{Id}_V}\downarrow & & \downarrow{\scriptstyle \mathrm{Id}_{V'}} \\ V & \xrightarrow{T} & V', \end{array}$$

se concluye que la matriz asociada a T respecto de las bases $\mathcal{B}_2$ y $\mathcal{B}'_2$ es

$$M_{\mathcal{B}_2,\mathcal{B}'_2}(T) = M(\mathcal{B}'_2,\mathcal{B}'_1)^{-1} \cdot M_{\mathcal{B}_1,\mathcal{B}'_1}(T) \cdot M(\mathcal{B}_2,\mathcal{B}_1). \tag{II.4.4}$$

Esta expresión se llama **fórmula del cambio de base**

Nota II.4.2. Si observamos detenidamente la fórmula (II.4.4) y la comparamos con la definición de matrices equivalentes (definición II.1.1) , podemos afirmar que las matrices $M_{\mathcal{B}_1,\mathcal{B}'_1}(T)$ y $M_{\mathcal{B}_2,\mathcal{B}'_2}(T)$ son equivalentes. Por consiguiente, *dos matrices asociadas a una misma aplicación lineal son equivalentes.* El recíproco de esta afirmación también es cierto, ya que si $B = Q^{-1}AP \in \mathcal{M}_{m\times n}(\Bbbk)$, con P y Q invertibles, entonces A y B definen la misma aplicación lineal de $\mathbb{R}^n$ en $\mathbb{R}^m$, siendo A la matriz asociada a la aplicación respecto de las bases usuales de $\mathbb{R}^n$ y $\mathbb{R}^m$, y B la matriz asociada respecto de las bases de $\mathbb{R}^n$ y $\mathbb{R}^m$ determinadas por las columnas de P y Q, respectivamente.

Ejemplo II.4.3. Sea $A \in \mathcal{M}_{m\times n}(\mathbb{R})$. La matriz A define una aplicación lineal de $\mathbb{R}^n$ en $\mathbb{R}^m$; en efecto, la aplicación $\mathbb{R}^n \longrightarrow \mathbb{R}^m;\ \mathbf{x} \mapsto A\mathbf{x} \in \mathbb{R}^m$ es lineal. De hecho, se trata de la aplicación lineal cuya matriz respecto de las bases usuales de $\mathbb{R}^n$ y $\mathbb{R}^m$ es A. De aquí que a menudo también se denote por A a la aplicación lineal, y se escriba $\mathrm{im}(A)$ y $\ker(A)$, es decir,

$$\mathrm{im}(A) = \{A\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n\} \quad \text{y} \quad \ker(A) = \{\mathbf{x} \in \mathbb{R}^n \mid A\mathbf{x} = \mathbf{0}\}.$$

Por otra parte, destacamos que si A tiene rango r y $R = Q^{-1}AP$ es su forma reducida, con $P \in \mathcal{M}_n(\mathbb{R})$ y $Q \in \mathcal{M}_m(\mathbb{R})$ invertibles, entonces las últimas $n - r$ columnas de P forman una base de $\ker(A)$ y las r primeras columnas de Q forman una base de $\mathrm{im}(A)$. Esta relación entre el rango de A y las dimensiones de su núcleo e imagen no es casual, y volveremos a ellas al final de la siguiente sección.

Finalizamos esta sección con un comentario sobre las transformaciones elementales por filas y explorando la relación que existe entre el rango de una aplicación lineal (esto es, la dimensión su imagen) y su matriz asociada.

Nota II.4.4. Con la misma notación que antes, las operaciones elementales por filas en $A = M_{\mathcal{B}_1,\mathcal{B}'_1}(T)$ (véase la definición II.1.2) no son más que cambios de bases en V'. En efecto:

Tipo I: La matriz que se consigue al intercambiar las filas i-ésima y l-ésima de A es la matriz asociada a T respecto de $\mathcal{B}_1$ y la base $\mathcal{B}'_2$ de V' que se obtiene al permutar el vector i-ésimo y l-ésimo de la base $\mathcal{B}'_1$ (compruébese).
Tipo II: La matriz que se consigue al multiplicar la fila i-ésima de A por $\lambda \in \Bbbk \setminus \{0\}$ es la matriz asociada a T respecto de las bases $\mathcal{B}_1$ y la base $\mathcal{B}'_2$ que se obtiene al sustituir el vector $\mathbf{v}'_i$ de $\mathcal{B}'_1$ por $\lambda^{-1}\mathbf{v}'_i$ (compruébese).
Tipo III: La matriz que se consigue al sumar a la fila i-ésima de A su fila l-ésima multiplicada por $\lambda \in \Bbbk$ es la asociada a T respecto de $\mathcal{B}_1$ y la base $\mathcal{B}'_2$ de V' que se obtiene al sustituir el vector $\mathbf{v}'_l$ de $\mathcal{B}'_2$ por $\mathbf{v}'_l - \lambda\mathbf{v}'_i$ con $i \neq l$ (compruébese).

Análogamente se puede comprobar que las operaciones elementales por columnas en A son cambios de base en V.

Teorema del rango. *Sean V y V' dos $\Bbbk$-espacios vectoriales de dimensiones finitas n y m, respectivamente, $\mathcal{B}_1$ y $\mathcal{B}'_1$ bases de V y V', respectivamente, y T una aplicación lineal de V en V'. Si $A \in \mathcal{M}_{m\times n}(\Bbbk)$ es la matriz asociada a T respecto de $\mathcal{B}$ y $\mathcal{B}'$, entonces*

1. $\operatorname{rg}(A) = \dim(\operatorname{Im}(T))$.
2. $\operatorname{rg}(A) = n - \dim(\ker(T))$.

Demostración. Sabemos que, si $r = \operatorname{rg}(A)$, existen unas matrices $P \in \mathcal{M}_n(\Bbbk)$ y $Q = \mathcal{M}_m(\Bbbk)$ invertibles tales que

$$Q^{-1}AP = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}$$

(véase el corolario II.1.8). Estas matrices son producto de las matrices elementales que se han ido obteniendo al realizar operaciones elementales por filas y por columnas en A. Luego, según lo explicado en la nota II.4.4, existen una base $\mathcal{B}_2$ de V y una base $\mathcal{B}'_2$ de V', tales que $P = M(\mathcal{B}_2, \mathcal{B}_1)$ y $Q = M(\mathcal{B}'_2, \mathcal{B}'_1)$, y por consiguiente, que

$$\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}$$

es la matriz de T respecto de $\mathcal{B}_2$ y $\mathcal{B}'_2$. De donde se sigue que los primeros r vectores de $\mathcal{B}'_2$ forman un base de $\operatorname{Im}(T)$ y que los últimos $n - r$ vectores de $\mathcal{B}_2$ forman una base de $\ker(T)$. ∎

5. Sistema de ecuaciones lineales (I)

A lo largo de esta sección V y V' serán dos $\Bbbk$-espacios vectoriales de dimensiones finitas $n > 0$ y $m > 0$, respectivamente, $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ una base de V y $\mathcal{B}' = \{\mathbf{v}'_1, \ldots, \mathbf{v}'_m\}$ una base de V'.

Las ecuaciones de la aplicación lineal T respecto de las bases $\mathcal{B}$ y $\mathcal{B}'$ (véase la expresión II.3.3) se pueden entender como un sistema lineal de ecuaciones, lo cual es no es sorprendente si tenemos en cuenta la siguiente definición.

Definición II.5.1. Llamaremos **sistema lineal de m ecuaciones y n incógnitas** a todo par $(T, \mathbf{b})$ donde $T \in \mathrm{Hom}(V, V')$ y $\mathbf{b} \in V'$; abreviadamente lo denotaremos por $T(\mathbf{x}) = \mathbf{b}$.

Un vector $\mathbf{v} \in V$ se dice que es **solución del sistema** $T(\mathbf{x}) = \mathbf{b}$ si $T(\mathbf{v}) = \mathbf{b}$; por lo tanto un sistema lineal de ecuaciones tiene solución si, y sólo si, $\mathbf{b} \in \mathrm{Im}(T)$. Un sistema se dice **compatible** si tienes soluciones, **incompatible** si no tiene soluciones, y **determinado** si tiene una única solución.

Un sistema lineal de ecuaciones $T(\mathbf{x}) = \mathbf{b}$ es **homogéneo** cuando $\mathbf{b} = \mathbf{0}_{V'}$. Es claro que un sistema homogéneo es siempre compatible, pues $\mathbf{0}_{V'} \in \mathrm{Im}(T)$, y que el conjunto de sus soluciones es $\ker(T)$. Cada sistema lineal de ecuaciones $T(\mathbf{x}) = \mathbf{b}$ tiene asociado un sistema homogéneo $T(\mathbf{x}) = \mathbf{0}_{V'}$.

Nota II.5.2. Sean $T \in \mathrm{Hom}_{\Bbbk}(V, V')$ y $A = (a_{ij}) \in \mathcal{M}_{m\times n}(\Bbbk)$ la matriz asociada a T respecto de las bases $\mathcal{B}$ y $\mathcal{B}'$. Sabemos que el núcleo de T son los vectores $\mathbf{x} \in V$ tales que $T(\mathbf{x}) = \mathbf{0}_{V'}$. Luego, se tiene que $\mathbf{v} \in \ker(T)$ si, y sólo si, sus coordenadas respecto de $\mathcal{B}$ son solución del sistema de ecuaciones lineales homogéneo $A\mathbf{x} = \mathbf{0}$.

Proposición II.5.3. *Sea $T(\mathbf{x}) = \mathbf{b}$ un sistema lineal de ecuaciones compatible. Si $\mathbf{v}_0 \in V$ es una solución particular de $T(\mathbf{x}) = \mathbf{b}$, entonces el conjunto de todas las soluciones del sistema es*

$$\mathbf{v}_0 + \ker(T) = \{\mathbf{v}_0 + \mathbf{v} \mid \mathbf{v} \in \ker(T)\}.$$

Demostración. La demostración es básicamente una comprobación y se deja como ejercicio al lector. ∎

Obsérvese que de la proposición anterior se deduce que un sistema lineal de ecuaciones $T(\mathbf{x}) = \mathbf{b}$ es compatible determinado si, y sólo si, $\mathbf{b} \in \mathrm{Im}(T)$ y $\ker(T) = \{\mathbf{0}_V\}$, es decir, si, y sólo si, $\mathbf{b} \in \mathrm{Im}(T)$ y T es inyectiva.

Este último hecho constituye la demostración del teorema de Rouché-Fröbenius que enunciaremos y probaremos a continuación, para lo cual es necesario definir un par de concepto previos.

Definición II.5.4. Sean $T \in \mathrm{Hom}_{\Bbbk}(V, V')$ y $\mathbf{b} \in V'$ un sistema de ecuaciones lineales. Si $A = (a_{ij}) \in \mathcal{M}_{m\times n}(\Bbbk)$ es la matriz asociada a T respecto de las bases $\mathcal{B}$ y $\mathcal{B}'$ y $(b_1, \ldots, b_m)$ son las coordenadas de $\mathbf{b}$ respecto de $\mathcal{B}'$, se llama **matriz ampliada asociada al sistema** $T(\mathbf{x}) = \mathbf{b}$ a la matriz $(A|\mathbf{b}) \in \mathcal{M}_{m\times(n+1)}(\Bbbk)$ definida de la

siguiente forma:

$$(A|\mathbf{b}) = \left(\begin{array}{cccc|c} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ a_{21} & a_{22} & \dots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \\ a_{m1} & a_{m2} & \dots & a_{mn} & b_m \end{array}\right).$$

Teorema de Rouché-Fröbenius. *Con la notación anterior, el sistema lineal de ecuaciones* $T(\mathbf{x}) = \mathbf{b}$ *es compatible si, y sólo si, las matrices* A *y* $(A|\mathbf{b})$ *tienen el mismo rango, y es compatible determinado si y sólo si las matrices* A *y* $(A|\mathbf{b})$ *tienen rango igual a* $\dim V$*, es decir, el rango es máximo.*

Demostración. $T(\mathbf{x}) = \mathbf{b}$ es compatible si, y sólo si, $\mathbf{b} \in \mathrm{Im}(T)$ si, y sólo si, $\mathbf{b}$ es combinación lineal de $\{T(\mathbf{v}_1), \dots, T(\mathbf{v}_n)\}$ si, y sólo si, las coordenadas de $\mathbf{b}$ respecto de $\mathcal{B}'$ son combinación lineal de las coordenadas de $\{T(\mathbf{v}_1), \dots, T(\mathbf{v}_n)\}$ respecto de $\mathcal{B}'$ si, y sólo si, $\mathrm{rg}(A) = \mathrm{rg}(A|\mathbf{b})$, por el ejercicio 4.

Para ver la segunda parte de la proposición basta tener en cuenta lo anterior y que T es inyectiva si, y sólo si, $\ker(T) = \{\mathbf{0}_V\}$, si, y sólo si, $\mathrm{rg}(A) = n$, por el Teorema del rango. ∎

Ejercicios del tema II

Ejercicio 1. Sean $A \in \mathcal{M}_{m\times p}(\Bbbk)$, $B \in \mathcal{M}_{p\times n}(\Bbbk)$ y $C = AB \in \mathcal{M}_{m\times n}(\Bbbk)$. Probar que si $A' = (a_{il}) \in \mathcal{M}_{m\times p}(\Bbbk)$ es la matriz obtenida al hacer una operación elemental por filas en A, entonces $C' = A'B$ es la matriz obtenida al hacer en C la misma operación elemental por filas. [Úsese la definición del producto de matrices.]

Ejercicio 2. Probar que

1. $T_{il}^{-1} = T_{li} = (T_{il})^{\mathrm{t}}$.
2. $(M_i(\lambda))^{\mathrm{t}} = M_i(\lambda)$ y $M_i(\lambda)^{-1} = M_i(1/\lambda)$, con $\lambda \in \Bbbk \setminus \{0\}$.
3. $(S_{il}(\lambda))^{\mathrm{t}} = S_{li}(\lambda)$ y $S_{il}(\lambda)^{-1} = S_{il}(-\lambda)$, con $\lambda \in \Bbbk$.

Ejercicio 3. A una matriz $A \in \mathcal{M}_{2\times 3}$ se le aplican, por el orden dado, las siguientes transformaciones elementales:

1. a la fila primera se suma la segunda.
2. a la fila tercera se le suma la primera y después la segunda.
3. la fila primera se multiplica por 2.

Determinar las matrices P y Q tales que la matriz obtenida después de realizar estas transformaciones sea $A' = QAP^{-1}$.

Si en lugar de aplicar las transformaciones elementales en el orden dado se aplican en el orden $1, 3$ y 2 ¿se obtiene el mismo resultado? ¿Y si se aplican en el orden $3, 2$ y 1?

Ejercicio 4. Sea $A \in \mathcal{M}_{m\times n}(\Bbbk)$. Probar que si la fila (o columna) i-ésima de la matriz A es combinación lineal del resto y A' es la submatriz de A que se obtiene eliminando la fila (o columna) i-ésima de A, entonces $\mathrm{rg}(A) = \mathrm{rg}(A')$.

Ejercicio 5. Sea $A \in \mathcal{M}_{m\times n}(\Bbbk)$.

1. Si $Q \in \mathcal{M}_n(\Bbbk)$ y $P \in \mathcal{M}_n(\Bbbk)$ son invertibles, entonces $\mathrm{rg}(Q^{-1}A) = \mathrm{rg}(AP) = \mathrm{rg}(A)$.
2. $\mathrm{rg}(A + B) \leq \mathrm{rg}(A) + \mathrm{rg}(B)$, para cualquier matriz $B \in \mathcal{M}_{m\times n}(\Bbbk)$.
3. $\mathrm{rg}(AB) \leq \mathrm{mín}(\mathrm{rg}(A), \mathrm{rg}(B))$, para cualquier matriz $B \in \mathcal{M}_{n\times p}(\Bbbk)$.
4. Si A y $B \in \mathcal{M}_n(\Bbbk)$, entonces $\mathrm{rg}(AB) \geq \mathrm{rg}(A) + \mathrm{rg}(B) - n$.

Ejercicio 6. Calcular el rango de la matriz

$$\begin{pmatrix} 2 & 2 & 2 & 1 & 1 & 4 \\ -1 & -1 & -3 & 0 & 2 & -1 \\ 1 & 2 & 1 & 1 & 1 & 3 \\ 3 & 1 & 2 & -2 & -1 & -1 \\ 4 & -2 & -2 & -6 & 0 & 8 \end{pmatrix}.$$

Definición. Se dice que una matriz $A \in \mathcal{M}_{m\times n}(\Bbbk)$ tiene **rango pleno por filas** si $\mathrm{rg}(A) = m$ y diremos que tiene **rango pleno por columnas** si $\mathrm{rg}(A) = n$.

Ejercicio 7. Sean $A \in \mathcal{M}_{n\times p}(\Bbbk)$ y $B \in \mathcal{M}_{p\times n}$. Si el producto de dos matrices $A\cdot B$ tiene determinante no nulo, ¿cuáles de las siguientes afirmaciones son necesariamente ciertas?

1. A tiene rango pleno por filas.
2. B tiene rango pleno por filas.
3. A tiene rango pleno por columnas.
4. B tiene rango pleno por columnas.

Ejercicio 8. Si una matriz B tiene rango pleno por columnas, ¿podemos concluir que $\mathrm{rg}(AB) = \mathrm{rg}(A)$? ¿y que $\mathrm{rg}(BA) = \mathrm{rg}(A)$?

Si C tiene rango pleno por filas, ¿podemos concluir que $\mathrm{rg}(AC) = \mathrm{rg}(A)$?¿y que $\mathrm{rg}(CA) = \mathrm{rg}(A)$?

Ejercicio 9. Probar que si una matriz A tiene rango pleno por columnas (respectivamente por filas), entonces la forma reducida de A puede obtenerse haciendo sólo transformaciones elementales en A por filas (respectivamente por columnas).

Ejercicio 10. Obtener la matriz asociada a la aplicación lineal $T : \mathbb{R}^2 \longrightarrow \mathbb{R}^3$ determinada por la igualdades $f(1,2) = (1,1,2),\ f(2,3) = (2,10,1)$ respecto de las bases $\mathcal{B} = \{(1,1),(1,3)\}$ de $\mathbb{R}^2$ y $\mathcal{B}' = \{(1,0,1),(1,1,0),(0,0,2)\}$ de $\mathbb{R}^3$.

Ejercicio 11. Sea $T : \mathbb{R}^2 \longrightarrow \mathbb{R}^3$ la aplicación lineal definida como $T(x,y) = (x+y, x+y, x+y)$.

1. Hallar la matriz asociada a T en las bases usuales.
2. Calcular bases de $\ker(T)$ e $\mathrm{Im}(T)$.

Ejercicio 12. Consideremos la aplicación lineal $T : \mathbb{R}^3 \to \mathbb{R}^4$ que respecto de las bases usuales de $\mathbb{R}^3$ y $\mathbb{R}^4$ viene dada por

$$T(x,y,z) = (x+z, y+z, x+z, y+z)$$

1. Calcular la matriz A de T respecto de las bases usuales de $\mathbb{R}^3$ y $\mathbb{R}^4$.
2. Calcular el rango r de A y determinar matrices P y Q tales que
$$Q^{-1}AP = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}.$$
3. Escribir una base de $\ker(T)$.
4. Escribir una base de $\mathrm{Im}(T)$.

Ejercicio 13. En $\mathbb{R}^3$ consideramos una base $\mathcal{B}$ fija. Sean T y $S \in \mathrm{End}_{\mathbb{R}}(\mathbb{R}^3)$ tales que sus matrices asociadas respecto de $\mathcal{B}$ son A y B, donde

$$A = \begin{pmatrix} 1 & 1 & 2 \\ 2 & 1 & 1 \\ 1 & 2 & 1 \end{pmatrix}, \qquad B = \begin{pmatrix} 0 & 2 & 1 \\ 1 & 3 & 1 \\ 1 & 1 & 0 \end{pmatrix}.$$

Calcular las matrices asociadas a las aplicaciones $S \circ T$ y $T \circ S$ respecto de $\mathcal{B}$.

Ejercicio 14. Calcular las coordenadas de un vector de $\mathbb{R}^3$ respecto de la base $\mathcal{B}_1 = \{(1,2,3),\ (3,4,0),\ (1,1,0)\}$ sabiendo que sus coordenadas respecto de la base $\mathcal{B}_2 = \{(1,1,0),\ (0,1,1),\ (1,0,1)\}$ son $(1,1,1)$.

Ejercicio 15. Sean $\mathcal{B}_1 = \{\mathbf{e}_1, \mathbf{e}_2\}$, $\mathcal{B}_2 = \{\mathbf{u}_1, \mathbf{u}_2\}$ y $\mathcal{B}_3 = \{\mathbf{v}_1, \mathbf{v}_2\}$ tres bases de $\mathbb{R}^2$ tales que $\mathbf{u}_1 = \mathbf{e}_1$, $\mathbf{u}_2 = 2\mathbf{e}_1 + \mathbf{e}_2$, $\mathbf{v}_1 = \mathbf{e}_1$ y $\mathbf{v}_2 = \mathbf{e}_1 + 4\mathbf{e}_2$. Usando las matrices de cambio de bases, calcular las coordenadas del vector $\mathbf{u} = 2\mathbf{u}_1 + 5\mathbf{u}_2$ respecto de la base $\mathcal{B}_3$.

Ejercicio 16. Dada la aplicación lineal $T : \mathbb{R}^3 \longrightarrow \mathbb{R}^2$ definida por $f(x,y,z) = (2x + y, y - z)$, calcular la matriz asociada a T respecto de:

1. las bases usuales de $\mathbb{R}^3$ y $\mathbb{R}^2$;
2. las bases $\mathcal{B} = \{(1,1,1),(0,1,2),(0,2,1)\}$ de $\mathbb{R}^3$ y $\mathcal{B}' = \{(2,1),(1,0)\}$ de $\mathbb{R}^2$.

Ejercicio 17. Sea $T : V \longrightarrow V'$ una aplicación lineal entre $\Bbbk$-espacios vectoriales de dimensión finita n. Probar que existen bases $\mathcal{B}$ y $\mathcal{B}'$ de V y V', respectivamente, tales que la matriz asociada a T respecto de $\mathcal{B}$ y $\mathcal{B}'$ es

$$\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix},$$

donde I_r es la matriz identidad de orden $r \leq n$. ¿Qué significado tiene r?

TEMA III

Matrices cuadradas y endomorfismos

En este tema vamos a estudiar los endomorfismos de un espacio vectorial desde el punto de vista de las matrices que los representan. En cualquier caso, dado que un endomorfismo no es más que un caso particular de aplicación lineal, siempre tendremos los resultados análogos a los del tema anterior adaptados a los endomorfismos. Por ejemplo,

Ejercicio. Sean V un $\Bbbk$ -espacio vectorial de dimensión finita, $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ y $T \in \mathrm{End}_{\Bbbk}(V)$. Probar que:

1. La matriz asociada a T respecto de $\mathcal{B}$ es una matriz $M_{\mathcal{B}}(T)$ cuadrada de orden n con coeficientes en $\Bbbk$.
2. Existe un isomorfismo $\phi : \mathrm{End}_{\Bbbk}(V) \longrightarrow \mathcal{M}_n(\Bbbk)$.
3. El $\Bbbk$ -espacio vectorial $\mathrm{End}_{\Bbbk}(V)$ es de dimensión finita y $\dim_k(\mathrm{End}_{\Bbbk}(V)) = n^2$.
4. La matriz del endomorfismo identidad de V respecto de $\mathcal{B}$ es I_n, es decir, la matriz identidad de orden n.

Buscando la analogía con el tema anterior, podemos preguntarnos si dos matrices cuadradas A y $B \in \mathcal{M}_n(\Bbbk)$ distintas representan un mismo endomorfismo aunque respecto de diferentes bases. En este caso, la fórmula del cambio de base determina una relación de equivalencia sobre las matrices cuadradas que llamaremos semejanza. Se demuestra que *dos matrices cuadradas son semejantes si, y sólo si, representan a un mismo endomorfismo,* y se plantea el problema de determinar de forma efectiva si dos matrices son semejantes. A diferencia de lo que ocurría en el caso de la equivalencia de matrices, el problema es mucho más complicado, ya que require un planteamiento teórico avanzado.

En la segunda sección del tema, se comienza definiendo el polinomio característico de una matriz, que nos da una condición necesaria (aunque no suficiente) para que dos matrices sean semejantes. A continuación, se muestra que el polinomio característico es un invariante asociado al endomorfismo, es decir, no depende de las bases elegidas. De este modo nos centramos en los endomorfismos como objeto geométrico asociado a las matrices cuadradas. Así, definimos los autovalores de un endomorfismo como las raíces de su polinomio característico, dando a continuación otras definiciones

equivalentes que nos permiten definir qué se entiende por autovector asociado a un autovalor de un endomorfismo.

La sección tercera está dedicada a la diagonalización; como es natural, lo primero que hacemos es definir qué entendemos por endomorfismo y matriz diagonalizable; así, diremos que un endomorfismo es diagonalizable si existe una base respecto de la cual su matriz es diagonal; y , por lo tanto, un matriz será diagonalizable si es semejante a una matriz diagonal. A continuación, se dan otras definiciones equivalentes de endomorfismo diagonalizable, y se demuestra que efectivamente son equivalentes. De donde se obtiene un primer criterio de diagonalización, y una condición suficiente para que un endomorfismo sea diagonalizable. Concretamente, *si un endomorfismo tiene tantos autovalores distintos como la dimensión del espacio vectorial, entonces es diagonalizable.*

Una condición necesaria y suficiente para que un endomorfismo sea diagonalizable nos la proporciona el llamado criterio de diagonalización por el polinomio característico. La clave de este otro criterio de diagonalización está en la acotación de las dimensiones de los subespacios propios asociados a los autovalores del endomorfismo, esta cota superior la proporciona lo que se conoce como multiplicidad del autovalor. De este modo, usando el concepto de multiplicidad, se obtiene un importante criterio de diagonalización.

La principal ventaja que presenta este criterio de diagonalización es que para probar que un endomorfismo no es diagonalizable basta encontrar un subespacio propio cuya dimensión sea distinta de la multiplicidad del autovalor correspondiente.

Si interpretamos los resultados obtenidos hasta el momento en términos de matrices, podemos afirmar que el problema de la semejanza está resuelto para las matrices diagonalizables. En efecto, dos matrices diagonalizables son semejantes si, y sólo si, tienen los mismos autovalores con idénticas multiplicidades. En resumen, los invariantes geométricos asociados a la semejanza de matrices diagonalizables son sus autovalores y las multiplicidades de éstos. Pero, ¿qué ocurre cuándo nos encontramos con una matriz no diagonalizable? Responderemos parcialmente a esta pregunta en la última sección.

En la sección cuarta, estudiamos con cierto detalle los subespacios invariantes por un endomorfismo. La relación con lo anterior es clara si tenemos en cuenta que el subespacio vectorial generado por los autovectores asociados a un autovalor de un endomorfismo es invariante por el endomorfismo. En cualquier caso, profundizamos en la teoría de subespacios invariantes por un endomorfismo con un segundo objetivo: justificar el interés práctico de la descomposición de un espacio vectorial en subespacios invariantes por un endomorfismo a la hora de estudiar el endomorfismo en cuestión (y en particular las matrices asociadas al mismo).

Para terminar el tema, abordamos el problema del cálculo de la forma canónica de Jordan de una endomorfismo (o una matriz cuadrada) cuando el polinomio característico tiene todas sus raíces en el cuerpo base. Para ello se comienza dando las definiciones de bloque y matriz de Jordan, de forma canónica de Jordan. A continuación se introducen los subespacios propios generalizados asociados a un autovalor, y entre otras cuestiones, se prueba que estos subespacios propios generalizados son invariantes por el endomorfismo, y que para cada autovalor existe una cadena creciente de subespacios propios generalizados que estabiliza en lo que denominamos subespacio propio máximo del autovalor. El primer resultado importante de esta sección es el teorema que afirma que

(a) La dimensión del subespacio propio máximo de autovalor coincide con su multiplicidad.
(b) Si todos los autovalores de un endomorfismo están en el cuerpo base, el espacio vectorial descompone en suma directa de los subespacios propios máximos asociados a los autovalores.

Veamos que los criterios de diagonalización estudiados en la tercera sección no son más que el caso particular del teorema anterior en el caso diagonalizable.

El teorema anterior permite fijar nuestra atención en cada uno de los subespacios propios máximos de forma individual mediante la restricción del endomorfismo a cada uno de ellos. Luego, a partir de este momento, para simplificar la notación, nos centraremos en el caso de los endomorfismos con un único autovalor de multiplicidad igual a la dimensión del espacio vectorial. A continuación, definimos qué se entiende por partición de la multiplicidad, y demostramos que la partición de la multiplicidad determina la forma canónica de Jordan del endomorfismo.

De este modo, concluimos que la forma canónica de Jordan queda determinada por los autovalores, en este caso λ, sus multiplicidades, en este caso n, y las particiones de multiplicidades, en este caso, $p_1 \geq p_2 \geq \ldots \geq p_s > 0$. Más concretamente, en nuestro caso, la forma canónica de Jordan consiste en

$$\begin{array}{cl} p_s & \text{bloques de orden } s \\ p_{s-1} - p_s & \text{bloques de orden } s-1 \\ \vdots & \\ p_1 - p_2 & \text{bloques de orden } 1 \end{array}$$

Nótese que estos números dependen exclusivamente del endomorfismo y no de la base elegida, por lo que podemos afirmar que *la forma canónica de Jordan es única salvo permutación de los bloques.* Lo importante de la forma canónica de Jordan es que se puede construir automáticamente a partir de los autovalores, sus multiplicidades y las particiones de multiplicidades.

Aunque todas las situaciones anteriores se han ido ilustrando con ejemplos, resaltamos aquí la necesidad de realizar un ejemplo para facilitar la compresión del cálculo de la forma canónica de Jordan.

En resumen, también podemos afirmar que el problema de la semejanza de matrices queda resuelto en este caso, si tenemos en cuenta que *dos matrices con todos sus autovalores en el cuerpo base son semejantes si, y sólo si, tienen los mismos autovalores con idénticas multiplicidades y particiones de multiplicidades.*

En este tema, hemos utilizado el capítulo 6 de [**BCR07**] y el capítulo 10 de [**Her85**] para las primeras secciones. Para la última sección hemos seguido principalmente el capítulo 5 de [**SV95**], aunque las secciones 1 y 2 del capítulo IV de [**MS06**] también han sido de utilidad.

1. Matrices semejantes

Nota III.1.1. Sean V un $\Bbbk$ -espacio vectorial de dimensión finita, $\mathcal{B}$ y $\mathcal{B}'$ dos bases de V y $T \in \mathrm{End}_{\Bbbk}(V)$. Si $M_{\mathcal{B}}(T)$ es la matriz asociada a T respecto $\mathcal{B}$, $M_{\mathcal{B}'}(T)$ es la matriz asociada a T respecto $\mathcal{B}'$ y $M(\mathcal{B}, \mathcal{B}')$ es del cambio de la base $\mathcal{B}$ a $\mathcal{B}'$, entonces la matriz asociada a T respecto $\mathcal{B}'$ es

$$M_{\mathcal{B}'}(T) = M(\mathcal{B}', \mathcal{B})^{-1} \cdot M_{\mathcal{B}}(T) \cdot M(\mathcal{B}', \mathcal{B}), \tag{III.1.1}$$

según la fórmula del cambio de base.

La fórmula (III.1.1) justifica en parte la siguiente definición.

Definición III.1.2. Sean A y $B \in \mathcal{M}_n(\Bbbk)$. Se dice que A y B son **semejantes** si existe una matriz invertible $P \in \mathcal{M}_n(\Bbbk)$ tal que $B = P^{-1}AP$.

La semejanza de matrices es una relación de equivalencia, es decir, verifica las propiedades reflexiva, simétrica y transitiva (compruébese).

Proposición III.1.3. *Dos matrices A y $B \in \mathcal{M}_n(\Bbbk)$ son semejantes si, y sólo si, A y $B \in \mathcal{M}_n(\Bbbk)$ son matrices asociadas a un mismo endomorfismo $T \in \mathrm{End}_{\Bbbk}(V)$ respecto de ciertas bases $\mathcal{B}$ y $\mathcal{B}'$ de V, respectivamente.*

Demostración. Sean $A = (a_{ij})$, $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ una base de V y T el endomorfismo de V definido por $T(\mathbf{v}_j) = a_{1j}\mathbf{v}_1 + \ldots + a_{nj}\mathbf{v}_n$, para cada $j = 1, \ldots, n$. Obsérvese que, por construcción, la matriz asociada T respecto de $\mathcal{B}$ es precisamente A.

Como A y $B \in \mathcal{M}_n(\Bbbk)$ son semejantes, existe una matriz invertible $P \in \mathcal{M}_n(\Bbbk)$ tal que $B = P^{-1}AP$. De modo que si $\mathcal{B}'$ es la familia de vectores cuyas coordenadas respecto de $\mathcal{B}$ son las columnas de P, entonces $\mathcal{B}'$ es una base de V y P^{-1} es la matriz del cambio de base de $\mathcal{B}'$ a $\mathcal{B}$ (pues P es invertible). Usando ahora que $B = P^{-1}AP$, por la fórmula del cambio de base para la matriz de asociada a un endomorfismo, se sigue que B es la matriz asociada a T respecto de $\mathcal{B}'$.

La otra implicación es una consecuencia directa de la fórmula del cambio de base. ■

Por consiguiente, según el resultado anterior, *dos matrices cuadradas son semejantes si, y sólo si, representan a un mismo endomorfismo.* No obstante, el ejercicio 3 pone de manifiesto que determinar de forma efectiva si dos matrices son semejantes es más difícil[1] que determinar si son equivalentes (donde bastaba calcular la forma reducida por filas).

El objetivo de este tema consistirá en dar condiciones necesarias y suficientes para que dos matrices A y B sean semejantes; en cuyo caso, calcularemos la matriz P tal que $B = P^{-1}AP$. Además, dada A determinaremos un representante especial de su clase de equivalencia que llamaremos forma canónica de Jordan de A.

Nota III.1.4. Obsérvese que el determinante y la traza se conservan por semejanza, es decir, si A y B son matrices semejantes, entonces $|A| = |B|$ y $\mathrm{tr}(A) = \mathrm{tr}(B)$. Luego, por la proposición anterior, podemos afirmar que *la traza y el determinante son invariantes por cambios de base*, lo que pone de manifiesto su naturaleza geométrica.

2. Polinomio característico. Autovalores y autovectores

A lo largo de esta sección V denotará un espacio vectorial sobre un cuerpo $\Bbbk$ (por ejemplo, $\Bbbk = \mathbb{R}$ ó $\mathbb{C}$) de dimensión finita $n > 0$.

Definición III.2.1. Sea $A = (a_{ij}) \in \mathcal{M}_n(\Bbbk)$. Se llama **polinomio característico de la matriz** A , y se denota por $\aleph_A(x)$, al determinante de la matriz $x\,I_n - A \in \mathcal{M}_n(k(x))$, donde I_n es la matriz identidad de orden n y $k(x)$ el cuerpo de las fracciones racionales en una indeterminada con coeficientes en $\Bbbk$. Es decir,

$$\aleph_A(x) = |x\,I_n - A| = \begin{vmatrix} x - a_{11} & -a_{12} & \dots & -a_{1n} \\ -a_{21} & x - a_{22} & \dots & -a_{2n} \\ \vdots & \vdots & & \vdots \\ -a_{n1} & -a_{n2} & \dots & x - a_{nn} \end{vmatrix}.$$

Obsérvese que el grado del polinomio característico coincide con el orden de la matriz y es unitario[2] (ejercicio 4).

Proposición III.2.2. *Sean $T \in \mathrm{End}_{\Bbbk}(V)$ y $\mathcal{B}$ y $\mathcal{B}'$ dos bases de V. Si A y $B \in \mathcal{M}_n(\Bbbk)$ son las matrices asociadas a T respecto de $\mathcal{B}$ y $\mathcal{B}'$, respectivamente, entonces*

[1] Para determinar si dos matrices A y $B \in \mathcal{M}_n(\Bbbk)$ son semejantes hay que determinar si el sistema de ecuaciones $XA - BX = 0$ tiene alguna solución invertible.

[2] Se dice que un polinomio es **unitario** (o **mónico**) si el coeficiente del término de mayor grado es uno.

el polinomio característico de A es igual al polinomio característico de B, es decir, $\aleph_A(x) = \aleph_B(x)$.

Demostración. Si $P \in \mathcal{M}_n(\Bbbk)$ es la matriz del cambio de base de $\mathcal{B}$ a $\mathcal{B}'$, entonces, por la fórmula del cambio de base, $B = P^{-1}\,A\,P$. Por lo tanto,

$$\begin{aligned}\aleph_B(x) &= |x\,I_n - B| = |x\,P^{-1}\,P - P^{-1}AP| = |P^{-1}xI_nP - P^{-1}AP| \\ &= |P^{-1}\,(xI_n - A)P| = |P^{-1}|\,|xI_n - A)|\,|P| = |P^{-1}|\,|P|\,|xI_n - A| \\ &= |xI_n - A| = \aleph_A(x).\end{aligned}$$

■

Corolario III.2.3. *Sean A y $B \in \mathcal{M}_n(\Bbbk)$. Si A y B son semejantes, entonces tienen el mismo polinomio característico.*

Demostración. Es una consecuencia inmediata de la proposición III.2.2 sin más que tener en cuenta la definición de matrices semejantes (véase la definición III.1.2). ■

El recíproco del resultado anterior no es cierto en general como se deduce del siguiente ejemplo.

Ejemplo III.2.4. Sea $V = \mathbb{R}^2$. Sabemos que la matriz asociada al endomorfismo nulo de $\mathbb{R}^2$ respecto de cualquier base de $\mathbb{R}^2$ es la matriz nula de orden 2. El polinomio característico del endomorfismo nulo es x^2.

Si consideramos la matriz

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix},$$

obtenemos que el polinomio característico de A también es x^2. Sin embargo, es del todo imposible que A sea la matriz del endomorfismo nulo respecto de ninguna base de $\mathbb{R}^2$, pues, por ejemplo $A(0,1)^{\mathrm{t}} \neq (0,0)^{\mathrm{t}}$.

La proposición III.2.2 asegura que los polinomios característicos de las distintas matrices asociadas a un mismo endomorfismo son iguales. Esto dota de sentido a la siguiente definición.

Definición III.2.5. Sea $T \in \mathrm{End}_{\Bbbk}(V)$. Se llama **polinomio característico del endomorfismo** T , y se denota por $\aleph_T(x)$, al polinomio característico de cualquiera de las matrices asociadas a T.

Autovalores y autovectores.

Definición III.2.6. Sea $T \in \mathrm{End}_{\Bbbk}(V)$. Se dice que $\lambda \in \Bbbk$ es un **autovalor** o **valor propio** de T si $\aleph_T(\lambda) = 0$, es decir, si es una raíz del polinomio característico de T.

Proposición III.2.7. *Sean $T \in \mathrm{End}_{\Bbbk}(V)$ y $\lambda \in \Bbbk$. Las afirmaciones siguientes son equivalentes*

(a) *λ es un autovalor de T.*
(b) *El endomorfismo $\lambda \mathrm{Id}_V - T$ de V no es inyectivo, es decir,* $\ker(\lambda \mathrm{Id}_V - T) \neq \{\mathbf{0}\}$.
(c) *Existe $\mathbf{v} \in V$ no nulo tal que $T(\mathbf{v}) = \lambda \mathbf{v}$.*

Demostración. $\boxed{\text{(a)} \Leftrightarrow \text{(b)}}$ Sea $A \in \mathcal{M}_n(\Bbbk)$ la matriz asociada a T respecto de alguna base $\mathcal{B}$ de V. Entonces, como la matriz asociada a $\lambda \mathrm{Id}_V - T$ respecto de $\mathcal{B}$ es $\lambda I_n - A, ,$ tenemos que $\lambda \in \Bbbk$ es un autovalor de T si, y sólo si, $\lambda \in \Bbbk$ es una raíz de $|\lambda I_n - A| = \aleph_T(x)$, si y sólo si, por el corolario II.3.7, $|\lambda I_n - A| = 0$, si, y sólo si, $\lambda \mathrm{Id}_V - T$ no es inyectivo.

La equivalencia $\boxed{\text{(b)} \Leftrightarrow \text{(c)}}$ es inmediata. ∎

Nótese que, como el grado del polinomio característico de un endomorfismo T de V es $n = \dim(V)$ (ejercicio 4), entonces, según el Teorema Fundamental del Álgebra (véase, por ejemplo, el teorema 2.1 de la página 86 de [**Nav96**]), el polinomio característico tiene, a lo sumo n raíces en $\Bbbk$. Luego, podemos afirmar que el número de autovalores de un endomorfismo de V es menor o igual que n.

Ejemplos III.2.8. Sea $V = \mathbb{R}^2$ y $T \in \mathrm{End}_{\mathbb{R}}(V)$.

i) Si $T(v_1, v_2) = (v_1, v_2)$, para todo $(v_1, v_2) \in \mathbb{R}^2$, entonces la matriz asociada a T respecto de la base usual de $\mathbb{R}^2$ es

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

luego el polinomio característico de T es

$$\aleph_T(x) = |xI_2 - A| = \begin{vmatrix} x-1 & 0 \\ 0 & x-1 \end{vmatrix} = (x-1)^2,$$

y por lo tanto el único autovalor de T es $\lambda = 1$.

ii) Si $T(v_1, v_2) = (v_1 - v_2, v_2)$, para todo $(v_1, v_2) \in \mathbb{R}^2$, entonces la matriz asociada a T respecto de la base usual de $\mathbb{R}^2$ es

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

luego el polinomio característico de T es

$$\aleph_T(x) = |xI_2 - A| = \begin{vmatrix} x-1 & -1 \\ 0 & x-1 \end{vmatrix} = (x-1)^2,$$

y por lo tanto el único autovalor de T es $\lambda = 1$.

iii) Si $T(v_1, v_2) = (-v_1, v_2)$, para todo $(v_1, v_2) \in \mathbb{R}^2$, entonces la matriz asociada a T respecto de la base usual de $\mathbb{R}^2$ es

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

luego el polinomio característico de T es

$$\aleph_T(x) = |xI_2 - A| = \begin{vmatrix} x+1 & 0 \\ 0 & x-1 \end{vmatrix} = (x+1)(x-1),$$

y por lo tanto los únicos autovalores de T son $\lambda = \pm 1$.

iv) Si $T(v_1, v_2) = (-v_2, v_1)$, para todo $(v_1, v_2) \in \mathbb{R}^2$, entonces la matriz asociada a T respecto de la base usual de $\mathbb{R}^2$ es

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

luego el polinomio característico de T es

$$\aleph_T(x) = |xI_2 - A| = \begin{vmatrix} x & 1 \\ -1 & x \end{vmatrix} = x^2 + 1,$$

y por lo tanto T no tiene autovalores. Obsérvese que si en vez de ser $V = \mathbb{R}^2$ fuese $V = \mathbb{C}^2$ (como espacio vectorial sobre $\mathbb{C}$), entonces T tendría dos autovalores distintos $\lambda_1 = \mathrm{i}$ y $\lambda_2 = -\mathrm{i}$.

Definición III.2.9. Sean $T \in \mathrm{End}_{\Bbbk}(V)$ y $\lambda \in \Bbbk$ un autovalor de T. El subespacio $\ker(\lambda \,\mathrm{Id}_V - T)$ se denomina **subespacio propio de T asociado a λ** . Los vectores no nulos de $\ker(\lambda \,\mathrm{Id}_V - T)$ se llaman **autovectores** o **vectores propios de T asociados a λ**.

Espectro de una matriz.

Teniendo en cuenta que una matriz $A \in \mathcal{M}_n(\Bbbk)$ define el endomorfismo

$$\Bbbk^n \longrightarrow \Bbbk^n; \mathbf{v} \mapsto A\mathbf{v}$$

de $\Bbbk^n$, que abusando la notación también denotaremos por A (nótese que se trata del endomorfismo de $\Bbbk^n$ cuya matriz respecto de la base usual de $\Bbbk^n$ es A) tiene perfecto sentido hablar de los autovalores y los autovectores A. En particular, por

el corolario III.2.3, se tiene que *si dos matrices son semejantes, entonces tienen los mismos autovalores.*

Obsérvese también que, por el Teorema Fundamental del Álgebra (véase, por ejemplo, el teorema 2.1 de la página 86 de [**Nav96**]), una matriz $A \in \mathcal{M}_n(\mathbb{R})$ tiene n autovalores complejos posiblemente repetidos.

Definición III.2.10. Sea $A \in \mathcal{M}_n(\mathbb{C})$.

(a) Llamaremos **espectro** de A al conjunto de todos los autovalores reales o complejos de la matriz A y lo representaremos por $\mathrm{sp}(A)$.

(b) El número real no negativo

$$\varrho(A) = \max\{|\lambda| : \lambda \in \mathrm{sp}(A)\}$$

es el **radio espectral** de A, donde $|\lambda|$ es el módulo de λ.

Como se observa, el radio espectral de una matriz es un número real, igual al radio del círculo más pequeño centrado en el origen que contiene a todos los autovalores de la matriz.

3. Diagonalización

Definición III.3.1. Sean V un $\Bbbk$ -espacio vectorial de dimensión finita y $T \in \mathrm{End}_{\Bbbk}(V)$. Se dice que T es **diagonalizable** si existe una base $\mathcal{B}$ de V tal que la matriz asociada a T respecto de $\mathcal{B}$ es diagonal.

Nota III.3.2. Obsérvese que si T es un endomorfismo diagonalizable de V y $D \in \mathcal{M}_n(\Bbbk)$ es una matriz diagonal, es decir,

$$D = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{pmatrix}$$

asociada a T, entonces $\lambda_i,\ i = 1, \dots, n$ son los autovalores (no necesariamente distintos) de T.

Se dice que una **matriz** $A \in \mathcal{M}_n(\Bbbk)$ es **diagonalizable** si es semejante a una matriz diagonal. De hecho, si A es diagonalizable, entonces es semejante a una matriz diagonal tal que las entradas de la diagonal son los autovalores de A. Es más, *dos matrices diagonalizables son semejantes si, y sólo si, son semejantes a la misma matriz diagonal.*

A continuación daremos condiciones necesarias y suficientes para que un endomorfismo (o una matriz) sea diagonalizable.

Lema III.3.3. *Si λ y $\mu \in \Bbbk$ son dos autovalores distintos de un endomorfismo T de V, entonces* $\ker(\lambda \operatorname{Id}_V - T)$ *y* $\ker(T - \mu \operatorname{Id}_V)$ *están en suma directa.*

Demostración. Si $\mathbf{v} \in \ker(T - \lambda \operatorname{Id}_V) \cap \ker(T - \mu \operatorname{Id}_V)$, entonces $T(\mathbf{v}) = \lambda \mathbf{v} = \mu \mathbf{v}$. De donde se sigue que $\mathbf{v} = \mathbf{0}$, por ser $\lambda \neq \mu$. ∎

Nótese que del resultado anterior se deduce que si $\mathbf{v}_1$ y $\mathbf{v}_2 \in V$ son autovectores asociados a distintos autovalores de un mismo endomorfismo de V, entonces $\{\mathbf{v}_1, \mathbf{v}_2\}$ es un conjunto linealmente independiente.

Teorema III.3.4. *Sean $\lambda_1, \ldots, \lambda_r \in \Bbbk$ los autovalores distintos de un endomorfismo T de V. Las siguientes afirmaciones son equivalentes*

(a) *T es diagonalizable.*
(b) *Existe una base de V formada por autovectores de T.*
(c) $V = \ker(T - \lambda_1 \operatorname{Id}_V) \oplus \ldots \oplus \ker(T - \lambda_r \operatorname{Id}_V)$.

Demostración. (a) ⇒ (b) Si T es diagonalizable, entonces existe una base $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ de V tal que la matriz asociada a T respecto de $\mathcal{B}$ es diagonal. Por tanto, $T(\mathbf{v}_i) = \mu \mathbf{v}_i$, $i = 1, \ldots, n$, para ciertos $\mu_1, \ldots, \mu_n \in \Bbbk$ no necesariamente distintos entre sí. Luego, $\mu_1, \ldots, \mu_n$ son autovalores (posiblemente repetidos) de T, y por lo tanto los vectores de $\mathcal{B}$ son autovectores de T.

(b) ⇒ (c) Sea $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ una base de V formada por autovectores de T. Para cada $i \in \{1, \ldots, n\}$, existe un autovalor λ_j, $j \in \{1, \ldots, r\}$ tal que $\mathbf{v}_i \in \ker(\lambda_j \operatorname{Id}_V - T)$. Luego,

$$V = \langle \mathbf{v}_1, \ldots, \mathbf{v}_n \rangle \subseteq \ker(T - \lambda_1 \operatorname{Id}_V) + \ldots + \ker(T - \lambda_r \operatorname{Id}_V) \subseteq V.$$

Por consiguiente, $V = \ker(T - \lambda_1 \operatorname{Id}_V) + \ldots + \ker(T - \lambda_r \operatorname{Id}_V)$. Finalmente, veamos que la suma es directa. Por el lema III.3.3, dos subespacios propios asociados a distintos autovalores están en suma directa. Luego, el resultado es cierto para $r \leq 2$. Supongamos, pues, que r es mayor o igual que tres. De este modo, si λ_1, λ_2 y λ_3 son autovalores distintos de T (sin pérdida de generalidad podemos suponer $\lambda_3 \neq 0$,) y $\mathbf{v} \in (\ker(T - \lambda_1 \operatorname{Id}_V) + \ker(T - \lambda_2 \operatorname{Id}_V)) \cap \ker(T - \lambda_3 \operatorname{Id}_V)$ es no nulo, existen unos únicos $\mathbf{v}_1 \in \ker(T - \lambda_1 \operatorname{Id}_V)$ y $\mathbf{v}_2 \in \ker(T - \lambda_2 \operatorname{Id}_V)$ no nulos tales que $\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2$. De donde se sigue que $\lambda_3 \mathbf{v} = T(\mathbf{v}) = \lambda_1 \mathbf{v}_1 + \lambda_2 \mathbf{v}_2$, y por lo tanto que $\mathbf{v} = \lambda_1/\lambda_3 \mathbf{v}_1 + \lambda_2/\lambda_3 \mathbf{v}_2$; luego, $\lambda_1 = \lambda_2 = \lambda_3$, lo que no es posible por hipótesis. Repitiendo un razonamiento análogo tantas veces como sea necesario se concluye el resultado buscado.

(c) ⇒ (a) Tomando una base de cada uno de los subespacio propios $\ker(T - \lambda_1 \operatorname{Id}_V)$, …, $\ker(T - \lambda_r \operatorname{Id}_V)$ obtenemos una base de V respecto de la cual la matriz asociada a T es diagonal. ∎

Realmente, en la demostración de la implicación (a) $\Rightarrow$ (b), no sólo hemos probado que existe una base formada por autovectores sino que toda base respecto de la cual T es diagonal está formada por autovectores.

Del teorema III.3.4, se deduce el siguiente **criterio de diagonalización**.

Corolario III.3.5. *Un endomorfismo $T \in \mathrm{End}_{\Bbbk}(V)$ es diagonalizable si, y sólo si, la suma de las dimensiones de los subespacios propios asociados a cada autovalor de T es igual a $n = \dim(V)$.*

Demostración. Si T es diagonalizable, por el teorema III.3.4, tenemos que la suma de las dimensiones de los subespacios invariantes asociados a cada autovalor de T es igual a $n = \dim(V)$.

Recíprocamente, si $\lambda_1, \ldots, \lambda_r$ los distintos autovalores de T, entonces

$$n \geq \dim(\ker(\lambda_1 \mathrm{Id}_V - T) \oplus \ldots \oplus \ker(\lambda_r \mathrm{Id}_V - T)) = \sum_{i=1}^{r} \dim(\ker(\lambda_i \mathrm{Id}_V - T)) = n,$$

de donde se sigue que $\ker(\lambda_1 \mathrm{Id}_V - T) \oplus \ldots \oplus \ker(\lambda_n \mathrm{Id}_V - T) = V$. Luego, por el teorema III.3.4, concluimos que T es diagonalizable.

■

Corolario III.3.6. *Sea $T \in \mathrm{End}_{\Bbbk}(V)$. Si T posee $n = \dim(V)$ autovalores distintos en $\Bbbk$, entonces T es diagonalizable.*

Demostración. Es una consecuencia inmediata del corolario III.3.5. ■

Nótese que el recíproco del teorema anterior no es cierto en general; tómese por ejemplo T igual a la identidad de V, que es diagonalizable y tiene todos sus autovalores iguales.

Ejemplo III.3.7. Sea $V = \mathbb{R}^3$ y T el endomorfismo de $\mathbb{R}^3$ cuya matriz asociada respecto de la base usual de $\mathbb{R}^3$ es

$$A = \begin{pmatrix} -3 & 2 & -2 \\ 0 & -2 & -1 \\ 0 & -5 & 2 \end{pmatrix}.$$

El polinomio característico de T es $\aleph_T(x) = (x+3)^2(x-3)$, por lo tanto los autovalores de T son $\lambda_1 = 3$ y $\lambda_2 = -3$. Calculemos el subespacio invariante asociado a cada autovalor.

Para el autovalor $\lambda_1 = 3$, la matriz asociada a $\lambda_1 \mathrm{Id}_V - T$ respecto de la base usual de $\mathbb{R}^3$ es

$$3\, I_n - A = \begin{pmatrix} 6 & -2 & 2 \\ 0 & 5 & 1 \\ 0 & 5 & 1 \end{pmatrix}.$$

Luego, $\dim(\mathrm{Im}(\lambda_1 \mathrm{Id}_V - T)) = r(3\,I_n - A) = 2$, y por tanto

$$\dim(\ker(\lambda_1 \mathrm{Id}_V - T)) = \dim(V) - \dim(\mathrm{Im}(\lambda_1 \mathrm{Id}_V - T)) = 3 - 2 = 1.$$

Sabemos que los vectores de $\ker(\lambda_1 \mathrm{Id}_V - T)$ son las soluciones del sistema lineal homogéneo $(\lambda_1 \mathrm{Id}_V - T)\mathbf{x} = \mathbf{0}$, es decir, los vectores de coordenadas $(x,\ y,\ z)$ respecto de la base usual de $\mathbb{R}^3$ que satisfacen

$$\begin{pmatrix} 6 & -2 & 2 \\ 0 & 5 & 1 \\ 0 & 5 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Resolviendo este sistema obtenemos que $x = 2t,\ y = t,\ z = -5t$ con $t \in \mathbb{R}$. Luego, los vectores de $\ker(\lambda_1 \mathrm{Id}_V - T)$ son los que tienen coordenadas respecto de la base usual de $\mathbb{R}^3$ de la forma $(2t, t, -5t)$ para algún $t \in \mathbb{R}$. Así, obtenemos que una base de $\ker(\lambda_1 \mathrm{Id}_V - T)$ la forma, por ejemplo, el vector de coordenadas $(2, 1, -5)$ respecto de la base usual de $\mathbb{R}^3$.

Para el autovalor $\lambda_2 = -3$ la matriz asociada a T respecto de la base usual de $\mathbb{R}^3$ es

$$(-3)I_n - A = \begin{pmatrix} 0 & -2 & 2 \\ 0 & -1 & 1 \\ 0 & 5 & -5 \end{pmatrix}.$$

Luego $\dim(\mathrm{Im}(\lambda_2 \mathrm{Id}_V - T)) = r(3I_n - A) = 1$, y por tanto

$$\dim(\ker(\lambda_2 \mathrm{Id}_V - T)) = \dim(V) - \dim(\mathrm{Im}(\lambda_2 \mathrm{Id}_V - T)) = 3 - 1 = 2.$$

Sabemos que los vectores de $\ker(\lambda_2 \mathrm{Id}_V - T)$ son las soluciones del sistema lineal homogéneo $(\lambda \mathrm{Id}_V - T)\mathbf{x} = \mathbf{0}$, es decir, los vectores de coordenadas (x, y, z) respecto de la base usual de $\mathbb{R}^3$ que satisfacen

$$\begin{pmatrix} 0 & -2 & 2 \\ 0 & -1 & 1 \\ 0 & 5 & -5 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Resolviendo este sistema obtenemos que $x = t, y = s, z = s$ con t y $s \in \mathbb{R}$. Luego, los vectores de $\ker(\lambda_1 \mathrm{Id}_V - T)$ son los que tienen coordenadas respecto de la base usual de $\mathbb{R}^3$ de la forma (t, s, s) para algunos t y $s \in \mathbb{R}$. Así, obtenemos que una base de $\ker(\lambda_1 \mathrm{Id}_V - T)$ la forman, por ejemplo, los vectores de coordenadas $(1, 0, 0)$ y $(0, 1, 1)$ respecto de la base usual de $\mathbb{R}^3$.

Finalmente, como la suma de las dimensiones de los subespacios invariantes asociados a los autovalores es $1 + 2 = 3$, y coincide con la dimensión de V, concluimos que T es diagonalizable y que una base de V respecto de la cual la matriz asociada a T es diagonal la forman los vectores de coordenadas $(2, 1, -5), (1, 0, 0)$ y $(0, 1, 1)$ respecto de la base usual de $\mathbb{R}^3$. En este caso, por tratarse de coordenadas respecto

de la base usual, se tiene que una base de $V = \mathbb{R}^3$ respecto de la cual la matriz de T es diagonal es $\mathcal{B}' = \{(2,1,-5),(1,0,0),(0,1,1)\}$.

Observamos que si $P \in \mathcal{M}_3(\mathbb{R})$ es la matriz cuyas columnas son las coordenadas respecto de la base usual de los vectores de la base $\mathcal{B}'$, es decir,

$$P = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 0 & 1 \\ -5 & 0 & 1 \end{pmatrix},$$

por la fórmula del cambio de base se tiene que

$$P^{-1}AP = D = \begin{pmatrix} 3 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & -3 \end{pmatrix}.$$

El proceso anterior se puede acortar considerablemente en caso de que el endomorfismo no resulte ser diagonalizable; esta será la principal aportación del *criterio de diagonalización por el polinomio característico* que veremos en breve. La clave de este otro criterio de diagonalización está en la acotación de las dimensiones de los subespacios propios asociados a los autovalores del endomorfismo (esta cota superior la proporciona lo que se conoce como multiplicidad del autovalor).

Si observamos el ejemplo anterior, el autovalor 3 correspondía al factor $(x-3)$ del polinomio característico y el autovalor -3 al factor $(x+3)^2$ del polinomio característico. Es decir, que en cierto sentido podríamos decir que el autovalor -3 "aparece dos veces" si consideramos $(x+3)^2 = (x+3)(x+3)$. La multiplicidad de un autovalor nos permite distinguir "el número de veces que se repite un mismo autovalor".

Definición III.3.8. Sea $T \in \mathrm{End}_{\Bbbk}(V)$. Llamaremos **multiplicidad de un autovalor λ de** T a la mayor potencia de $(x-\lambda)$ que divide al polinomio característico de T.

A la vista de la definición anterior, decir que un autovalor λ de T tiene multiplicidad m_λ significa que $(x-\lambda)^{m_\lambda}$ divide a $\aleph_T(x)$ y que $(x-\lambda)^{m_\lambda+1}$ no lo divide; equivalentemente, que en la descomposición en potencias de factores irreducibles de $\aleph_T(x)$ aparece $(x-\lambda)^{m_\lambda}$ como factor. Es claro que, al ser λ una raíz de $\aleph_T(x)$, su multiplicidad siempre es mayor o igual que 1. De hecho siempre es mayor o igual que la dimensión del subespacio propio asociado a λ, como asegura el siguiente lema.

Lema III.3.9. *Sea $T \in \mathrm{End}_{\Bbbk}(V)$. Si $\lambda \in \Bbbk$ es un autovalor de T, entonces la dimensión del subespacio propio* $\ker(\lambda \mathrm{Id}_V - T)$ *asociado a λ es menor o igual que la multiplicidad de λ.*

Demostración. Sean $L = \ker(\lambda \operatorname{Id}_V - T)$ el subespacio invariante asociado a un autovalor λ de T y $\mathcal{B}_L$ una base de L. Si ampliamos la base $\mathcal{B}_L$ a una base $\mathcal{B}$ de V, entonces la matriz asociada a T respecto de $\mathcal{B}$ es

$$A = \begin{pmatrix} \lambda I_r & A_1 \\ 0 & A_2 \end{pmatrix},$$

donde I_r es la matriz identidad de orden $r = \dim(L)$, $A_1 \in \mathcal{M}_{r\times(n-r)}(\Bbbk)$ y $A_2 \in \mathcal{M}_{n-r}(\Bbbk)$. De modo que

$$\aleph_T(x) = |xI_n - A| = \begin{vmatrix} \lambda(x-\lambda)I_r & -A_1 \\ 0 & xI_{n-r} - A_2 \end{vmatrix} = (x-\lambda)^r \aleph_{A_2}(x),$$

es decir, $(x-\lambda)^r$ divide a al polinomio característico de T. De donde se sigue que la multiplicidad de λ es mayor o igual que $r = \dim(L)$. ∎

Veamos que la acotación superior de la dimensión del subespacio invariante asociado a un autovalor λ por su multiplicidad puede ser estricta, es decir, existen casos donde no se cumple la igualdad.

Ejemplo III.3.10. Sean $V = \mathbb{R}^2$ y $T \in \operatorname{End}_{\mathbb{R}}(V)$ tal que $T(v_1, v_2) = (v_1 - v_2, v_2)$, para todo $(v_1, v_2) \in \mathbb{R}^2$. Anteriormente vimos que $\aleph_T(x) = (x-1)^2$, luego T tiene un sólo autovalor $\lambda = 1$ de multiplicidad $m_\lambda = 2$. El subespacio propio $\ker(\lambda \operatorname{Id}_V - T)$ asociado a λ es $\langle(1,0)\rangle$. Luego, se cumple que $\dim(\ker(\lambda \operatorname{Id}_V - T)) = 1 \leq 2 = m_\lambda$, pero no se da la igualdad.

Sea $T \in \operatorname{End}_{\Bbbk}(V)$. Si $\lambda \in \Bbbk$ es autovalor de T de multiplicidad m_λ, entonces solamente podemos asegurar la igualdad a priori cuando $m_\lambda = 1$ ya que en este caso tenemos que

$$1 \leq \dim(\ker(\lambda \operatorname{Id}_V - T)) \leq m_\lambda = 1,$$

lo que obviamente implica que $\dim(\ker(\lambda \operatorname{Id}_V - T)) = 1$.

Criterio de diagonalización por el polinomio característico. *Sean $T \in \operatorname{End}_{\Bbbk}(V)$. Si $\lambda_1, \ldots, \lambda_r \in \Bbbk$ son los distintos autovalores de T y sus multiplicidades son $m_1, \ldots, m_r$, respectivamente, entonces T es diagonalizable si, y sólo si,*

(a) $\dim(\ker(\lambda_i \operatorname{Id}_V - T)) = m_i,\ i = 1, \ldots, r.$

(b) $m_1 + \ldots + m_r = n.$

Demostración. Si T es diagonalizable, por el teorema III.3.4, tenemos que

$$V = \ker(\lambda_1 \operatorname{Id}_V - T) \oplus \ldots \oplus \ker(\lambda_r \operatorname{Id}_V - T).$$

Además, por el lema III.3.9, $\dim(\ker(\lambda_i \operatorname{Id}_V - T)) \leq m_i$, para cada $i = 1, \ldots, r$. De ambos hechos se deduce que

$$n = \dim(\ker(\lambda_1 \operatorname{Id}_V - T)) + \ldots + \dim(\ker(\lambda_r \operatorname{Id}_V - T)) \leq m_1 + \ldots + m_r \leq n.$$

Por lo tanto, $\dim(\ker(\lambda_1 \operatorname{Id}_V - T)) + \ldots + \dim(\ker(\lambda_r \operatorname{Id}_V - T)) = m_1 + \ldots + m_r = n$, y, como consecuencia (usando de nuevo el lema III.3.9) $m_i = \dim(\ker(\lambda_i \operatorname{Id}_V - T))$, para cada $i = 1, \ldots, r$.

Recíprocamente, como

$$\begin{aligned}\dim(ker(\lambda_1 \operatorname{Id}_V - T) \oplus \ldots \oplus \ker(\lambda_r \operatorname{Id}_V - T)) &= \sum_{i=1}^{r} \dim(ker(\lambda_i \operatorname{Id}_V - T)) \\ &= m_1 + \ldots + m_r = n = \dim(V),\end{aligned}$$

del teorema III.3.4, se sigue que T es diagonalizable. ∎

Nota III.3.11. Obsérvese que el teorema anterior dice que un endomorfismo T de V es diagonalizable si, y sólo si, $\aleph_T(x)$ tiene todas sus raíces en $\Bbbk$ y la multiplicidad de cada autovalor coincide con la dimensión del subespacio propio correspondiente.

La principal ventaja que presenta el criterio de diagonalización por el polinomio característico es que para probar que un endomorfismo no es diagonalizable basta encontrar un subespacio propio cuya dimensión sea distinta de la multiplicidad del autovalor correspondiente.

Ejemplo III.3.12. En el ejemplo III.3.10, vimos que $\dim(\ker(\lambda \operatorname{Id}_V - T)) = 1 \neq 2 = m_\lambda$. Luego T no es diagonalizable.

Nota III.3.13. Si interpretamos esta teoría en términos de matrices cuadradas, observamos que hemos determinado cuándo una matriz $A \in \mathcal{M}_n(\Bbbk)$ es diagonalizable; en tal caso, se tiene que si $P \in \mathcal{M}_n(\Bbbk)$ es la matriz cuyas columnas forman una base de $\Bbbk^n$ formada por autovectores de A (que existe por el teorema III.3.4), entonces $P^{-1}AP$ es diagonal. De modo que podemos afirmar que *dos matrices diagonalizables son semejantes si, y sólo si, tienen los mismos autovalores con idénticas multiplicidades.*

Pero, ¿qué ocurre cuándo nos encontramos con una matriz no diagonalizable? Responderemos parcialmente a esta pregunta en la última sección.

4. Subespacios invariantes

Definición III.4.1. Dado $T \in \operatorname{End}_{\Bbbk}(V)$. Diremos que un subespacio L de V es **invariante** por T cuando $T(L) \subseteq L$, es decir, la restricción de T a L, que se suele denotar por $T_{|L}$, es un endomorfismo de L.

Nótese que los subespacios trivial y total de V son invariantes para cualquier endomorfismo $T \in \operatorname{End}_{\Bbbk}(V)$.

Lema III.4.2. *Sean $T \in \operatorname{End}_{\Bbbk}(V)$. Si L_1 y L_2 son dos dos subespacios de V invariantes por T, entonces se verifica que $L_1 + L_2$ es invariante por T.*

Demostración. Si $\mathbf{v} \in L_1 + L_2$, entonces existen $\mathbf{v}_1 \in L_1$ y $\mathbf{v}_2 \in L_2$ tales que $\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2$. Además, $T(\mathbf{v}_1) \in L_1$ y $T(\mathbf{v}_2) \in L_2$, pues L_1 y L_2 son invariantes por T. Por consiguiente $T(\mathbf{v}) = T(\mathbf{v}_1 + \mathbf{v}_2) = T(\mathbf{v}_1) + T(\mathbf{v}_2) \in L_1 + L_2$. ∎

A continuación veremos que, además de los casos triviales, existen muchos otros subespacios invariantes por T; pero antes introduciremos la siguiente notación: dado un endomorfismo T de V y un polinomio $p(x) = a_m x^m + \ldots + a_1 x + a_0 \in \Bbbk[x]$ denotaremos por $p(T)$ al siguiente endomorfismo de V

$$a_0 \mathrm{Id}_V + a_1 T + \ldots + a_m T^m,$$

donde $\mathrm{Id}_V = T^0$ es la identidad de V y $T^r = \overbrace{T \circ \cdots \circ T}^{r \text{ veces}}$, para cada $r = 1, \ldots, m$.

Nota III.4.3. El lector con cierto conocimientos de algebra conmutativa básica puede observar que $p(T)$ no es más que la imagen de $p(x)$ por el morfismo de anillos $\Phi_T : \Bbbk[x] \longrightarrow \mathrm{End}_{\Bbbk}(V)$. Así, como $\Bbbk[x]$ es un anillo conmutativo, se sigue que

$$\begin{aligned} p(T) \circ q(T) &= \Phi_T(p(x)) \cdot \Phi_T(q(x)) = \Phi_T(p(x)q(x)) \\ &= \Phi_T(q(x)p(x)) = \Phi_T(q(x)) \cdot \Phi_T(p(x)) = q(T) \circ p(T). \end{aligned}$$

Usaremos esta igualdad en la demostración del siguiente resultado.

Proposición III.4.4. *Sea $T \in \mathrm{End}_{\Bbbk}(V)$. Para todo $p(x) \in \Bbbk[x]$, se cumple que:*

(a) $\ker(p(T))$ *es invariante por* T;
(b) $\mathrm{Im}(p(T))$ *es invariante por* T.

Demostración. (a) Sea $p(x) \in \Bbbk[x]$. Para demostrar que $T(\ker(p(T))) \subset \ker(p(T))$, basta probar que $T(\mathbf{v}) \in \ker(p(T))$, para todo $\mathbf{v} \in \ker(p(T))$, es decir, $p(T)(T(\mathbf{v})) = \mathbf{0}$, para todo $\mathbf{v} \in \ker(p(T))$. Lo cual es inmediato, tomando $q(x) = x \in \Bbbk[x]$ y teniendo en cuenta que, según la nota III.4.3, $p(T)$ y $q(T)$ conmutan entre sí, ya que

$$p(T)(T(\mathbf{v})) = p(T)(q(T)(\mathbf{v})) = q(T)(p(T)(\mathbf{v})) = q(T)(\mathbf{0}) = T(\mathbf{0}) = \mathbf{0},$$

como queríamos probar.

(b) Sean $p(x) \in \Bbbk[x]$ y $\mathbf{v}' \in \mathrm{Im}(p(T))$. Queremos probar que $T(\mathbf{v}') \in \mathrm{Im}(p(T))$. Por estar $\mathbf{v}'$ en la imagen de $p(T)$, se tiene que existe $\mathbf{v}$ tal que $\mathbf{v}' = p(T)(\mathbf{v})$, tomando $q(x) := x \in \Bbbk[x]$ y teniendo en cuenta que $p(T)$ y $q(T)$ conmutan entre sí, se sigue que

$$T(\mathbf{v}') = T(p(T)(\mathbf{v})) = q(T)(p(T)(\mathbf{v})) = p(T)(q(T)(\mathbf{v})) = p(T)(T(\mathbf{v})) \in \mathrm{Im}(p(T)).$$

∎

Ejemplo III.4.5. Sea T el endomorfismo identidad de V. Si $p(x) = a - x \in \Bbbk[x]$ con $a \neq 1$, entonces $p(T) = a\mathrm{Id}_V - T = a\mathrm{Id}_V - \mathrm{Id}_V = (a-1)\mathrm{Id}_V$ que la homotecia de razón $(a-1) \neq 0$ y por consiguiente automorfismo de V, luego $\ker(p(T)) = 0$ e $\mathrm{Im}(p(T)) = V$.

Ejemplo III.4.6. Sea T el endomorfismo de $V = \mathbb{R}^2$ tal que $T(x,y) = (x,-y)$ Si $p(x) = 1 - x$, entonces $p(T)(x,y) = (\mathrm{Id}_V - T)(x,y) = (x,y) - (x,-y) = (0,2y)$. Luego $\ker(p(T)) = \langle(1,0)\rangle$ e $\mathrm{Im}(p(T)) = \langle(0,1)\rangle$, son subespacios de $\mathbb{R}^2$ invariantes por T. De hecho, no es difícil comprobar que son los únicos subespacios propios de $\mathbb{R}^2$ invariantes por T distintos del trivial.

Ejemplo III.4.7. El subespacio vectorial $\ker(\lambda\,\mathrm{Id}_V - T)$ de V es invariante por T ; en efecto, si $\mathbf{v} \in \ker(\lambda\,\mathrm{Id}_V - T)$, entonces

$$\begin{aligned}(\lambda\,\mathrm{Id}_V - T)(T(\mathbf{v})) &= (\lambda\mathrm{Id}_V - T)(T(\mathbf{v}) - \lambda\mathbf{v} + \lambda\mathbf{v}) \\ &= -(\lambda\,\mathrm{Id}_V - T)((\lambda\,\mathrm{Id}_V - T)(\mathbf{v})) + \lambda(\lambda\,\mathrm{Id}_V - T)(\mathbf{v}) \\ &= -(\lambda\,\mathrm{Id}_V - T)(\mathbf{0}) + \mathbf{0} = \mathbf{0}.\end{aligned}$$

En realidad, habría bastado observar que $\ker(\lambda\,\mathrm{Id}_V - T)$ es $\ker(p(T))$ para $p(x) = \lambda - x \in \Bbbk[x]$ y entonces usar la proposición III.4.4 para concluir.

Terminemos esta sección viendo una serie de interesantes resultados sobre subespacios invariantes que serán de suma utilidad posteriormente.

Proposición III.4.8. *Sean V un $\Bbbk$ -espacio vectorial, $T \in \mathrm{End}_{\Bbbk}(V)$ y $p(x) \in \Bbbk[x]$ un polinomio distinto de cero tal que $\ker(p(T))$ es no nulo. Si $p(x) = q_1(x)q_2(x)$ tal que $q_1(x)$ y $q_2(x)$ son unitarios y primos entre sí* [3] *entonces*

$$\ker(p(T)) = \ker(q_1(T)) \oplus \ker(q_2(T)).$$

Demostración. En primer lugar, como $q_1(x)$ y $q_2(x)$ son primos entre sí, entonces, según la Identidad de Bezout (véase la página 66 de [**Nav96**]) , existen $h_1(x)$ y $h_2(x) \in \Bbbk[x]$ tales que $1 = h_1(x)q_1(x) + h_2(x)q_2(x)$, Luego, tenemos que $I = h_1(T) \circ q_1(T) + h_2(T) \circ q_2(T)$, es decir

$$\mathbf{v} = (h_1(T) \circ q_1(T))(\mathbf{v}) + (h_2(T) \circ q_2(T))(\mathbf{v}), \tag{III.4.2}$$

para todo $\mathbf{v} \in V$.

Si $\mathbf{v} \in \ker(p(T))$, entonces $p(T)(\mathbf{v}) = q_1(T) \circ q_2(T)(\mathbf{v}) = q_2(T) \circ q_1(T)(\mathbf{v}) = \mathbf{0}$. Por consiguiente,

$$q_2(T)((h_1(T) \circ q_1(T))(\mathbf{v})) = h_1(T)((q_2(T) \circ q_1(T))\mathbf{v}) = h_1(T)(\mathbf{0}) = \mathbf{0},$$

[3] Dos polinomios "son primos entre sí" si no tienen factores comunes, es decir, $\mathrm{mcd}(q_1(x), q_2(x)) = 1$.

para todo $\mathbf{v} \in \ker(p(T))$, de donde se sigue que $h_1(T) \circ q_1(T))(\mathbf{v}) \in \ker(q_2(T)))$. para todo $\mathbf{v} \in \ker(p(T))$. Análogamente, se prueba que $h_2(T) \circ q_2(T))(\mathbf{v}) \in \ker(q_1(T)))$. para todo $\mathbf{v} \in \ker(p(T))$. De ambas afirmaciones, junto con la expresión (III.4.2), se deduce que $\ker(p(T)) \subseteq \ker(q_1(T)) + \ker(q_2(T))$.

Recíprocamente, si $\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2 \in \ker(q_1(T)) + \ker(q_2(T))$, con $\mathbf{v}_i \in \ker(q_i(T))$, $i = 1, 2$, entonces

$$\begin{aligned} p(T)(\mathbf{v}) &= (q_1(T) \circ q_2(T))(\mathbf{v}) = (q_1(T) \circ q_2(T))(\mathbf{v}_1 + \mathbf{v}_2) \\ &= (q_1(T) \circ q_2(T))(\mathbf{v}_1) + (q_1(T) \circ q_2(T))(\mathbf{v}_2) \\ &= (q_2(T) \circ q_1(T))(\mathbf{v}_1) + (q_1(T) \circ q_2(T))(\mathbf{v}_2) \\ &= q_2(T)(q_1(T)(\mathbf{v}_1)) + q_1(T)(q_2(T)(\mathbf{v}_2)) = q_2(T)(\mathbf{0}) + q_1(T)(\mathbf{0}) = \mathbf{0}. \end{aligned}$$

Hemos probado que $\ker(p(T)) \subseteq \ker(q_1(T)) + \ker(q_2(T))$. Nos queda ver que $\ker(q_1(T)) \cap \ker(q_2(T)) = \{\mathbf{0}\}$. Sea $\mathbf{v} \in \ker(q_1(T)) \cap \ker(q_2(T))$, entonces sigue que

$$\mathbf{v} = (h_1(T) \circ q_1(T))(\mathbf{v}) + (h_2(T) \circ q_2(T))(\mathbf{v}) = \mathbf{0} + \mathbf{0} = \mathbf{0},$$

y por consiguiente el único vector de $\ker(q_1(T)) \cap \ker(q_2(T))$ es el cero. ∎

Proposición III.4.9. *Sea $T \in \mathrm{End}_{\Bbbk}(V)$. Las condiciones siguientes son equivalentes:*

(a) *V es suma directa $V = L_1 \oplus \cdots \oplus L_r$ de subespacios invariantes por T.*

(b) *Existe una base $\mathcal{B}$ de V tal que la matriz de T respecto de ella es*[4]

$$A_1 \oplus \ldots \oplus A_r,$$

donde las A_i son matrices cuadradas.

Demostración. Supongamos que se verifica la condición primera y procedamos por inducción sobre r. Si $r = 1$, evidentemente no hay nada que demostrar. Supongamos, pues, que $r > 1$ y que el resultado es cierto para un espacio vectorial que descompone en suma directa de $r - 1$ subespacios invariantes. En particular, la matriz de $T_{|L}$, con $L = L_2 \oplus \ldots \oplus L_r$, respecto de $\mathcal{B} = \cup_{i \geq 2} \mathcal{B}_i$ es $A = A_2 \oplus \ldots \oplus A_r$. Nótese que, por el lema III.4.2, L es un subespacio invariante por T.

Por consiguiente, queda ver que la matriz de T respecto de $\mathcal{B}_1 \cup \mathcal{B}$ es $A_1 \oplus A$; para lo cual, es suficiente observar que $T(\mathbf{v})$ es combinación de lineal de elementos de $\mathcal{B}_1$ si $\mathbf{v} \in \mathcal{B}_1$ y $T(\mathbf{v})$ es combinación lineal de elementos de $\mathcal{B}$ si $\mathbf{v} \in \mathcal{B}$, por ser L_1 y L subespacios invariantes por T y $\mathcal{B}_1$ y $\mathcal{B}$ bases de aquellos, respectivamente.

Recíprocamente, supongamos que se verifica la condición segunda y que $A_i \in \mathcal{M}_{n_i}(\Bbbk)$, $i = 1, \ldots, r$. Dividamos $\mathcal{B}$ en subconjuntos $\mathcal{B}_i$, $i = 1, \ldots, r$, de forma consistente con la bloques de A. Sea L_i el subespacio vectorial generado por $\mathcal{B}_i$; por la

[4] Véase la definición de suma directa de matrices en la sección 3 del tema I.

forma de A es claro que $T(L_i) \subseteq L_i,\ i = 1, \ldots, r$, y naturalmente $V = L_1 \oplus \cdots \oplus L_s$. ■

Observando ahora las proposiciones anteriores conjuntamente podemos llegar a la siguiente conclusión: si somos capaces de hallar un polinomio $p(x) \in \Bbbk[x]$ tal que $p(T) = 0 \in \mathrm{End}_{\Bbbk}(V)$ y $\prod_{i=1}^{r} q_i(x)^{m_i}$ es su descomposición en potencias de factores irreducibles en $\Bbbk[x]$, por la proposición III.4.8, obtenemos que

$$V = \ker(p(T)) = \ker(q_1(T)^{m_1}) \oplus \ldots \oplus \ker(q_r(T)^{m_r}),$$

esto es, una descomposición de V en subespacios invariantes. De tal modo que, usando la proposiciones III.4.8 y III.4.9, podemos reducir el estudio de la matriz de T al de las matrices de las restricción de T a cada uno de los subespacios invariantes $\ker(q_i(T)^{m_i}),\ i = 1, \ldots, r$.

5. Forma canónica de Jordan

A lo largo de esta sección V será un espacio vectorial de dimensión finita $n > 0$ sobre un cuerpo $\Bbbk$ y T un endomorfismo de V cuyos autovalores distintos son $\lambda_1, \ldots, \lambda_r \in \Bbbk$ de multiplicidades $m_1, \ldots, m_r$, respectivamente.

En la sección anterior vimos que no todos los endomorfismos de V son diagonalizables, es decir, que en general no se puede encontrar una base de V tal que la matriz de T sea diagonal. Por lo que el objetivo de hallar una base de V tal que la matriz de T sea "lo más sencilla posible" nos obliga a determinar en primer lugar qué entendemos por "lo más sencilla posible".

Definición III.5.1. Un **bloque de Jordan** de orden s es una matriz cuadrada con s filas y s columnas que tiene todos las entradas de la diagonal principal idénticos, la diagonal por encima de ésta está formada por 1 y las restantes entradas son cero, es decir, $B = (b_{ij}) \in \mathcal{M}_s(\Bbbk)$ es un bloque de Jordan si

$$b_{ij} = \begin{cases} \lambda & \text{si } i = j; \\ 1 & \text{si } i + 1 = j; \\ 0 & \text{en otro caso.} \end{cases}$$

para algún $\lambda \in \Bbbk$.

Obsérvese que un bloque de Jordan de orden s no es otra cosa que la suma de una matriz diagonal $D_\lambda \in \mathcal{M}_s(\Bbbk)$ y una matriz nilpotente

$$N = \begin{pmatrix} 0 & 1 & \ldots & 0 & 0 \\ 0 & 0 & \ddots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \ldots & 0 & 1 \\ 0 & 0 & \ldots & 0 & 0 \end{pmatrix} \in \mathcal{M}_s(\Bbbk)$$

tal que $N^{s-1} \neq 0$ y $N^s = 0$.

Ejemplo III.5.2. Un bloque de Jordan de orden 1 es un escalar λ. Un bloque de Jordan de orden 2 es de la forma

$$\begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}$$

y uno de orden 3 es

$$\begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix}.$$

Definición III.5.3. Una **matriz de Jordan** es una matriz diagonal por bloques de manera que cada bloque es de Jordan, esto es, $J \in \mathcal{M}_n(\Bbbk)$ es de Jordan si

$$J = \begin{pmatrix} B_1 & 0 & \dots & 0 \\ 0 & B_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & B_r \end{pmatrix},$$

donde cada $B_i,\ i = 1, \dots, r$, es un bloque de Jordan.

En esta sección demostraremos que, si todos los autovalores de T están en $\Bbbk$, existe una base $\mathcal{B}$ de V tal que la matriz asociada a T respecto de $\mathcal{B}$ es de Jordan. La base $\mathcal{B}$ se llama **base de Jordan** y la matriz de T respecto de $\mathcal{B}$ se llama **forma canónica de Jordan de** T , que veremos que es única salvo permutación de los bloques de Jordan.

Dicho de otro modo, demostraremos que *toda matriz cuadrada con coeficientes en* $\Bbbk$ *tal que todos sus autovalores están en* $\Bbbk$, *es semejante a una matriz de Jordan; en particular, a una matriz triangular superior.*

Ejemplo III.5.4. Si T es diagonalizable entonces su forma canónica de Jordan es

$$J = \begin{pmatrix} \mu_1 & 0 & \dots & 0 \\ 0 & \mu_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mu_n \end{pmatrix},$$

donde $\mu_i,\ i = 1, \dots, n$ son los autovalores de T repetidos tantas veces como indique su multiplicidad. Dicho de otro modo, T *es diagonalizable si, y sólo si, los bloques de Jordan en su forma canónica tienen orden* 1.

A continuación vamos a introducir una serie de subespacios invariantes que necesitamos para construir la base de Jordan y veremos sus propiedades más relevantes.

Definición III.5.5. Para cada $i \in \{1, \ldots, r\}$ y $j \geq 0$, llamaremos **subespacios propios generalizados asociados al autovalor** λ_i a

$$L_{i,j} = \ker((\lambda_i \operatorname{Id}_V - T)^j).$$

Nótese que $L_{i\,0} = \ker((\lambda_i \operatorname{Id}_V - T)^0) = \ker(\operatorname{Id}_V) = \{\mathbf{0}\}$, para todo $i = 1, \ldots, r$.

Nota III.5.6. Obsérvese que para cada $i \in \{1, \ldots, r\}$ se tiene que

1. $L_{i,1} = \ker(\lambda_i \operatorname{Id}_V - T)$, esto es, el subespacio propio asociado a λ_i.
2. $L_{i,1} \subseteq L_{i,2} \subseteq \ldots \subseteq L_{i,s} \subseteq \ldots$ En efecto, si $\mathbf{v} \in \ker((\lambda_i \operatorname{Id}_V - T)^j)$, entonces

$$\begin{aligned}(\lambda_i \operatorname{Id}_V - T)^{j+1}(\mathbf{v}) &= \big((\lambda_i \operatorname{Id}_V - T) \circ (\lambda_i \operatorname{Id}_V - T)^j\big)(\mathbf{v}) \\ &= \big(\lambda_i \operatorname{Id}_V - T\big)\big((\lambda_i \operatorname{Id}_V - T)^j(\mathbf{v})\big) = (\lambda_i \operatorname{Id}_V - T)(\mathbf{0}) = \mathbf{0}.\end{aligned}$$

3. L_{ij} es un subespacio invariante por T. En efecto, si $\mathbf{v} \in L_{ij}$, entonces $T(\mathbf{v}) \in L_{ij}$, ya que

$$\begin{aligned}(\lambda_i \operatorname{Id}_V - T)^j(T(\mathbf{v})) &= (\lambda_i \operatorname{Id}_V - T)^j(T(\mathbf{v}) - \lambda_i \mathbf{v} + \lambda_i \mathbf{v}) \\ &= -(\lambda_i \operatorname{Id}_V - T)^{j+1}(\mathbf{v}) + \lambda_i (\lambda_i \operatorname{Id}_V - T)^j(\mathbf{v}) \\ &= \mathbf{0} + \mathbf{0} = \mathbf{0}\end{aligned}$$

Como V es dimensión finita, para cada $i = 1, \ldots, r$, la cadena de subespacios L_{ij} se estabiliza; más aún veremos que se estabiliza definitivamente a partir del momento en que $L_{is_i} = L_{i\,s_i+1}$ para algún $j \geq 1$. Es decir, las inclusiones del apartado 1 de la nota III.5.6 son igualdades a partir un cierto $s_i \geq 1$, que depende de i.

Lema III.5.7. *Si* $L_{is} = L_{i\,s+1}$, *entonces* $L_{ij} = L_{is}$, *para todo* $j \geq s$.

Demostración. Basta demostrar que $L_{i\,s+2} = L_{i\,s+1}$. Una inclusión la hemos visto anteriormente. Para la otra, sea $\mathbf{v} \in L_{i\,s+2}$. Entonces,

$$\mathbf{0} = (\lambda_i \operatorname{Id}_V - T)^{s+2}(\mathbf{v}) = (\lambda_i \operatorname{Id}_V - T)^{s+1}\big((\lambda_i \operatorname{Id}_V - T)(\mathbf{v})\big),$$

por lo que $(\lambda_i \operatorname{Id}_V - T)(\mathbf{v}) \in L_{i\,s+1} = L_{is}$, de donde se sigue que

$$(\lambda_i \operatorname{Id}_V - T)^s((\lambda_i \operatorname{Id}_V - T)(\mathbf{v})) = \mathbf{0}$$

y tenemos que $\mathbf{v} \in \ker(\lambda_i \operatorname{Id}_V - T)^{s+1} = L_{i\,s+1}$. ∎

Nótese que, según el lema anterior, la cadena de inclusiones del apartado 1 de la nota III.5.6 queda de la forma

$$L_{i1} \subsetneq L_{i2} \subsetneq \ldots \subsetneq L_{i\,s_i} = L_{i\,s_i+1} = \ldots,$$

para cada $i = 1, \ldots, r$. El subespacio $L_{i\,s_i}$ se llama **subespacio propio máximo del autovalor** λ_i.

A continuación demostremos que la dimensión del subespacio propio máximo de un autovalor coincide con su multiplicidad y que, si todos los autovalores de T están en $\Bbbk$, entonces V descompone en suma directa de los subespacios propios máximos.

Lema III.5.8. *El único autovalor de la restricción de T a $L_{i\,s_i}$ es λ_i.*

Demostración. Sea $\mu \in \overline{\Bbbk}$ un autovalor de T (es decir, μ es un autovalor de T en el cierre algebraico[5] $\overline{\Bbbk}$ de $\Bbbk$) y $\mathbf{v} \in L_{i\,s_i}$ un autovector de T asociado a μ. Como $T(\mathbf{v}) = \mu\mathbf{v}$ y $(\lambda_i \operatorname{Id}_{L_{i\,s_i}} - T)^{s_i}(\mathbf{v}) = \mathbf{0}$, se tiene que

$$\begin{aligned}\mathbf{0} &= (\lambda_i \operatorname{Id}_{L_{i\,s_i}} - T)^{s_i}(\mathbf{v}) = (\lambda_i \operatorname{Id}_{L_{i\,s_i}} - T)^{s_i-1}\big((\lambda_i \operatorname{Id}_{L_{i\,s_i}} - T)(\mathbf{v})\big)\\ &= (\lambda_i - \mu)(\lambda_i \operatorname{Id}_{L_{i\,s_i}} - T)^{s_i-1}(\mathbf{v}) = \ldots = (\lambda_i - \mu)^{s_i}\,\mathbf{v},\end{aligned}$$

de donde se sigue que $\mu = \lambda_i$. ∎

Lema III.5.9. *Sea $\mathbf{v} \in L_{i\,j} \setminus L_{i,j-1}$, para algún $j \in \{1, \ldots, s_i\}$. Para todo $\alpha \in \Bbbk$ distinto de λ_i se cumple que $(\alpha \operatorname{Id}_V - T)^s(\mathbf{v}) \in L_{i\,j} \setminus L_{i,j-1}$, para todo $s \geq 0$. En particular, $(\alpha \operatorname{Id}_V - T)^s(\mathbf{v}) \neq \mathbf{0}$, para todo $s \geq 0$.*

Demostración. Basta probar el enunciado para $s = 1$. Se tiene que

$$\begin{aligned}(\alpha \operatorname{Id}_V - T)(\mathbf{v}) &= \big((\lambda_i \operatorname{Id}_V - T) + (\alpha - \lambda_i)\operatorname{Id}_V\big)(\mathbf{v})\\ &= (\lambda_i \operatorname{Id}_V - T)(\mathbf{v}) + (\alpha - \lambda_i)\mathbf{v}.\end{aligned}$$

Como $(\alpha - \lambda_i)\mathbf{v} \in L_{i\,j} \setminus L_{i\,j-1}$ y $(\lambda_i \operatorname{Id}_V - T)(\mathbf{v}) \in L_{i\,j-1}$, necesariamente es $(\alpha \operatorname{Id}_V - T)(\mathbf{v}) \in L_{i\,j} \setminus L_{i\,j-1}$. ∎

Teorema III.5.10. *Con la notación anterior. Se verifica que*

(a) $\dim(L_{i\,s_i}) = m_i$, $i = 1, \ldots, r$, *es decir, la dimensión del subespacio propio máximo de cada autovalor coincide con su multiplicidad.*

(b) *Si todos los autovalores de T están en $\Bbbk$, entonces $V = L_{1\,s_1} \oplus \ldots \oplus L_{r\,s_r}$*

Demostración. (a) Fijemos un índice i, $1 \leq i \leq r$. Sea $\mathcal{B}_i$ una base de V que sea ampliación de una de base $L_{i\,s_i}$. Como $L_{i\,s_i}$ es un subespacio invariante por T (véase el apartado 3. de la nota III.5.6), la matriz de T respecto de $\mathcal{B}_i$ es del tipo

$$\begin{pmatrix} A_i & N_i \\ 0 & M_i \end{pmatrix}.$$

Pongamos $n_i = \dim(L_{i\,s_i})$. El polinomio característico de T es, pues, igual a

$$\aleph_T(x) = \begin{vmatrix} xI_{n_i} - A_i & -N_i \\ 0 & xI_{n-n_i} - M_i \end{vmatrix} = |xI_{n_i} - A_i|\,|xI_{n-n_i} - M_i|.$$

[5]Por ejemplo, si $\Bbbk = \mathbb{R}$, entonces su cierre algebraico es $\overline{\Bbbk} = \mathbb{C}$. El lector interesado en conocer más sobre el cierre algebraico puede consultar el Apéndice I de [**Nav96**].

Además, por el lema III.5.8, $|xI_{n_i} - A_i| = (x-\lambda_i)^{n_i}$. De modo que

$$\aleph_T(x) = (x-\lambda_i)^{n_i}|xI_{n-n_i} - M_i|.$$

Supongamos que λ_i es uno de los autovalores de M_i y elijamos un vector no nulo

$$\mathbf{v} = (0, \ldots, 0, v_{n_i+1}, \ldots, v_n)$$

tal que $(\lambda_i I_{n-n_i} - M_i)(v_{n_i+1}, \ldots, v_n)^{\mathrm{t}} = \mathbf{0}$; es claro que $\mathbf{v} \notin L_{i\,s_i}$, además, el vector $(\lambda_i \mathrm{Id}_V - T)(\mathbf{v})$ tiene coordenadas

$$\begin{pmatrix} \lambda_i I_{n_i} - A_i & -N_i \\ 0 & \lambda_i I_{n-n_i} - M_i \end{pmatrix} \mathbf{v} = \begin{pmatrix} v_1' \\ \vdots \\ v_{n_i}' \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

respecto de $\mathcal{B}$. Luego, $(\lambda_i \mathrm{Id}_V - T)(\mathbf{v}) \in L_{i\,s_i}$, de donde se sigue que $(\lambda_i \mathrm{Id}_V - T)^{s_i+1}(\mathbf{v}) = \mathbf{0}$ y entonces $\mathbf{v} \in L_{i\,s_{i+1}} = L_{i\,s_i}$, lo que supone una contradicción. Esto demuestra que λ_i no es un autovalor de M_i, luego todos los factores de $(x-\lambda_i)$ en el polinomio característico de T están en $|xI_{n_i} - A_i| = (x-\lambda_i)^{n_i}$. Por consiguiente, $n_i = m_i$.

(b) Si todos los autovalores de T están en $\Bbbk$, entonces $\sum_{i=1}^r m_i = n = \dim(V)$; de donde se sigue que $V = L_{1\,s_1} \oplus \ldots \oplus L_{r\,s_r}$, si, y sólo si, los subespacios propios máximos están en suma directa. Para demostrar que la suma es directa tenemos que ver si $\mathbf{v}_i \in L_{i\,s_i}$ son tales que $\sum_{i=1}^r \mathbf{v}_i = \mathbf{0}$, entonces $\mathbf{v}_i = \mathbf{0},\ i = 1, \ldots, r$. Probémoslo por reducción al absurdo. Supongamos, por ejemplo, que $\mathbf{v}_1 \neq 0$. Entonces existe un índice $j \in \{1, \ldots, s_1\}$, tal que $\mathbf{v}_1 \in L_{1\,j} \setminus L_{1\,j-1}$. Se tiene por tanto,

$$\mathbf{0} = \left(\prod_{i=2}^r (\lambda_i \mathrm{Id}_V - T)^{s_i}\right)\left(\sum_{j=1}^r \mathbf{v}_j\right) = \left(\prod_{i=2}^r (\lambda_i \mathrm{Id}_V - T)^{s_i}\right)(\mathbf{v}_1),$$

que pertenece a $L_{1\,j} \setminus L_{1\,j-1}$ por el lema III.5.9; lo que supone una contradicción pues $\mathbf{0} \in L_{1\,j-1}$. ∎

Obsérvese que el criterio de diagonalización por el polinomio característico es el caso particular del teorema anterior en el caso diagonalizable.

A partir de ahora, y a lo largo de toda esta sección, supondremos que T tiene todos sus autovalores en $\Bbbk$ (esto ocurre, por ejemplo, si $\Bbbk = \mathbb{C}$ independiente del endormofismo T).

Nota III.5.11. Sin pérdida de generalidad, por el teorema III.5.10, podemos suponer que T tiene un sólo autovalor $\lambda \in \Bbbk$ de multiplicidad $n = \dim(V)$. Denotaremos por L_s al subespacio propio máximo de λ, de tal forma que tenemos la siguiente sucesión de subespacios invariantes por T
(III.5.3)
$L_0 = \{\mathbf{0}\} \subsetneq L_1 = \ker(\lambda \operatorname{Id}_V - T) \subsetneq L_2 = \ker(\lambda \operatorname{Id}_V - T)^2 \subsetneq \ldots \subsetneq L_s = \ker(\lambda \operatorname{Id}_V - T)^s$,
con $\dim(L_s) = n$, es decir, $L_s = V$, por el teorema III.5.10 de nuevo.

Vamos a construir la base canónica de Jordan para el subespacio propio máximo $L_s = V$ de λ.

Definición III.5.12. Sean $H_1 \subsetneq H_2$ subespacios vectoriales de V. Diremos que $\mathbf{v}_1, \ldots, \mathbf{v}_t \in H_2$ son **linealmente independientes módulo** H_1 si $\alpha_1, \ldots, \alpha_q \in \Bbbk$ son tales que $\alpha_1\mathbf{v}_1 + \ldots + \alpha_q\mathbf{v}_q \in H_1$, entonces $\alpha_1 = \ldots = \alpha_q = 0$.

Lema III.5.13. *Sea $H_0 \subsetneq H_1 \subsetneq \ldots \subsetneq H_s$ una cadena estrictamente creciente de subespacios vectoriales de V. Si H es un conjunto finito de vectores de V,*

$$H = \{\mathbf{v}_{ij} \mid 1 \leq j \leq t_i,\ 1 \leq i \leq s\}$$

tal que para todo $i = 1, \ldots, s$ los vectores $\{\mathbf{v}_{ij} \mid 1 \leq j \leq t_i\}$ pertenecen a H_i y son independientes módulo H_{i-1}, entonces H es un sistema de vectores linealmente independiente.

Demostración. Sean $\alpha_{ij} \in \Bbbk$ tales que $\sum \alpha_{ij}\mathbf{v}_{ij} = \mathbf{0}$. Entonces

$$\sum_{j=1}^{t_s} \alpha_{sj}\mathbf{v}_{sj} = -\left(\sum_{1\leq i<s, 1\leq j\leq t_i} \alpha_{ij}\mathbf{v}_{ij}\right) \in H_{s-1}.$$

Como $\{\mathbf{v}_{sj} \mid 1 \leq j \leq t_i\}$ pertenecen a H_s y son independientes módulo H_{s-1}, se tiene que $\alpha_{s1} = \ldots = \alpha_{s\,t_s} = 0$. Repitiendo el razonamiento agrupando los vectores de H_{s-2}, luego los de H_{s-3} y así sucesivamente, vemos que todos los α_{ij} deben ser cero. ∎

Lema III.5.14. *Si $\mathbf{v}_1, \ldots, \mathbf{v}_q \in L_j$ son linealmente independientes módulo L_{j-1}, entonces*

$$(\lambda \operatorname{Id}_V - T)(\mathbf{v}_1), \ldots, (\lambda \operatorname{Id}_V - T)(\mathbf{v}_q) \in L_{j-1}$$

son linealmente independientes módulo L_{j-2}.

Demostración. Sean $\alpha_1, \ldots, \alpha_q \in \Bbbk$ tales que $\sum_{l=1}^{q} \alpha_l\big((\lambda_l \operatorname{Id}_V - T)(\mathbf{v}_l)\big) \in L_{j-2}$. Así $(\lambda_i \operatorname{Id}_V - T)\big(\sum_{l=1}^{q} \alpha_l\mathbf{v}_l\big) \in L_{j-2}$, luego $\sum_{l=1}^{q} \alpha_l\mathbf{v}_l \in L_{j-1}$, de donde se sigue que $\alpha_1 = \ldots = \alpha_q = 0$. ∎

Proposición III.5.15. *Sean* $n_j = \dim(L_j)$ *y* $p_j = n_j - n_{j-1}$, *para cada* $j = 1, \ldots, s$. *Entonces,*

(a) *El número máximo de vectores de* L_j *que son linealmente independientes módulo* L_{j-1} *es* p_j.
(b) *Se cumple que* $p_1 \geq p_2 \geq \ldots \geq p_s > 0$.

Teniendo en cuenta que $n = \sum_{j=1}^{s} p_j$ (compruébese) y que n es la multiplicidad de λ, a los $p_i,\ i = 1, \ldots, s$, se les llama **partición de la multiplicidad del autovalor** λ.

Demostración. (a) Sea $\mathcal{B}_j = \{\mathbf{v}_1, \ldots, \mathbf{v}_{p_j}, \mathbf{u}_1, \ \ldots, \ \mathbf{u}_{n_{j-1}}\}$ una base de L_j tal que $\mathcal{B}_{j-1} = \{\mathbf{u}_1, \ldots, \mathbf{u}_{n_{j-1}}\}$ sea una base L_{j-1} ($\mathcal{B}_j$ siempre existe, pues podemos tomar una base de L_{j-1} y ampliarla a una de L_j). Por un lado, es claro que los vectores $\mathbf{v}_1, \ldots, \mathbf{v}_{p_j}$ son linealmente independientes módulo L_{j-1}; en efecto, si existen $\alpha_1, \ldots, \alpha_{p_j} \in \Bbbk$ tales que $\sum_{l=1}^{p_j} \alpha_l \mathbf{v}_l \in L_{j-1}$, entonces $\sum_{l=1}^{p_j} \alpha_l \mathbf{v}_l = \mathbf{0}$, pues en otro caso $\mathcal{B}_j$ no sería una base, y como $\mathbf{v}_1, \ldots, \mathbf{v}_{p_j}$ son linealmente independientes se sigue que $\alpha_1 = \ldots = \alpha_{p_j} = 0$. Por otra parte, si $\mathbf{w}_1, \ldots, \mathbf{w}_q \in L_j$ son linealmente independientes módulo L_{j-1}, entonces, por el lema III.5.13 aplicado a la cadena $\{\mathbf{0}\} \subsetneq L_{j-1} \subsetneq L_j$, los vectores $\mathbf{w}_1, \ldots, \mathbf{w}_q, \mathbf{u}_1, \ldots, \mathbf{u}_{n_{j-1}}$ de L_j son linealmente independientes; de donde se sigue que $q + n_{j-1} \leq n_j$ y por lo tanto que $q \leq n_j - n_{j-1} = p_j$.

(b) Ahora, usando el lema III.5.14, concluimos que $p_{j-1} \geq p_j$, para cada $i = 2, \ldots, s$. ∎

Lema III.5.16. *Sean* $\mathbf{u}_j \in L_j \backslash L_{j-1}$ *y* $\mathbf{u}_{j-l} = -(\lambda \operatorname{Id}_V - T)(\mathbf{u}_{j-l+1}),\ l = 1, \ldots, j-1$. *Entonces,*

(a) $\{\mathbf{u}_1, \ldots, \mathbf{u}_j\}$ *es un conjunto de vectores de* L_j *linealmente independiente.*
(b) *Si* $L = \langle \mathbf{u}_1, \ldots, \mathbf{u}_j \rangle$, *entonces* L *es un subespacio invariante por* T *y la matriz* A_L *de la restricción de* T *a* L *respecto de* $\{\mathbf{u}_1, \ldots, \mathbf{u}_j\}$ *es una matriz de Jordan, concretamente,*

$$A_L = \begin{pmatrix} \lambda & 1 & \ldots & 0 & 0 \\ 0 & \lambda & & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda & 1 \\ 0 & 0 & \ldots & 0 & \lambda \end{pmatrix}$$

Demostración. (a) Los vectores $\{\mathbf{u}_1, \ldots, \mathbf{u}_j\}$ son linealmente independientes por los Lemas III.5.14 y III.5.13.

(b) De las relaciones

$$\begin{array}{rcl}(\lambda\,\mathrm{Id}_V - T)(\mathbf{u}_j) & = & -\mathbf{u}_{j-1}\\ (\lambda\,\mathrm{Id}_V - T)(\mathbf{u}_{j-1}) & = & -\mathbf{u}_{j-2}\\ & \vdots & \\ (\lambda\,\mathrm{Id}_V - T)(\mathbf{u}_2) & = & -\mathbf{u}_1\\ (\lambda\,\mathrm{Id}_V - T)(\mathbf{u}_1) & = & \mathbf{0}\end{array}$$

se obtiene que

$$\begin{array}{rcl}T(\mathbf{u}_1) & = & \lambda\mathbf{u}_1\\ T(\mathbf{u}_2) & = & \mathbf{u}_1 + \lambda\mathbf{u}_2\\ & \vdots & \ddots\\ T(\mathbf{u}_{j-1}) & = & \mathbf{u}_{j-2} + \lambda\mathbf{u}_{j-1}\\ T(\mathbf{u}_j) & = & \mathbf{u}_{j-1} + \lambda\mathbf{u}_j,\end{array}$$

de donde se sigue que L es un subespacio invariante por T (luego, la restricción de T a L está bien definida) y que la matriz A_L de la restricción de T a L respecto de $\{\mathbf{u}_1, \ldots, \mathbf{u}_j\}$ es una matriz de Jordan. ∎

Teorema III.5.17. *Con la notación anterior. Existe una base $\mathcal{B}$ de L_s tal que la matriz de T respecto de $\mathcal{B}$ es una matriz de Jordan.*

Demostración. En primer lugar tomamos unos vectores $\{\mathbf{v}_1, \ldots, \mathbf{v}_{p_s}\}$ de L_s que sean linealmente independientes módulo L_{s-1} y a partir de ellos se construye, usando el lema III.5.16, la base de Jordan correspondiente. La simple unión conjuntista de los vectores obtenidos es un conjunto de vectores linealmente independientes de L_s, por los lemas III.5.14 y III.5.13. Si el número de vectores es igual a $\dim(L_s)$, ya hemos terminado. Supongamos que no, y sea $j < s$ el mayor índice tal que los vectores que están en $L_j \setminus L_{j-1}$ no alcanzan el máximo número de vectores linealmente independientes módulo L_{j-1}, es decir, j es el mayor índice tal que $p_j > p_s$ (véase la proposición III.5.15). Ampliando este conjunto de vectores hasta alcanzar el número máximo, se obtiene un nuevo conjunto de vectores $\{\mathbf{v}'_1, \ldots, \mathbf{v}'_{p_j - p_s}\}$, con el que repetimos lo anterior, y así sucesivamente. El final de este proceso es una base $\mathcal{B}$ de L_s tal que la matriz de T respecto de $\mathcal{B}$ está formada por bloques de Jordan colocados diagonalmente (véase el lema III.5.16). ∎

Nota III.5.18. La forma canónica de Jordan queda determinada por los autovalores, en este caso λ, sus multiplicidades, en este caso n, y las particiones de multiplicidades, en este caso, $p_1 \geq p_2 \geq \ldots \geq p_s > 0$. Más concretamente, en nuestro caso,

la forma canónica de Jordan consiste en

$$\begin{array}{cc} p_s & \text{bloques de orden } s \\ p_{s-1}-p_s & \text{bloques de orden } s-1 \\ \vdots & \\ p_1-p_2 & \text{bloques de orden } 1 \end{array}$$

Nótese que estos números dependen exclusivamente de T, y no de la base elegida. Por lo que podemos afirmar que *la forma canónica de Jordan es única salvo permutación de los bloques.* **Lo importante de la forma canónica de Jordan es que se puede construir automáticamente a partir de los autovalores, sus multiplicidades y las particiones de multiplicidades**.

Ejemplo III.5.19. Sean V un espacio vectorial sobre $\mathbb{R}$ de dimensión 4 y $\mathcal{B} = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4\}$ una base V. Definimos el endomorfismo T de V por

$$\begin{array}{rcrcrcrcr} T(\mathbf{u}_1) & = & \mathbf{u}_1 & + & -\mathbf{u}_2 & + & -\mathbf{u}_3 & & \\ T(\mathbf{u}_2) & = & -\mathbf{u}_1 & & & + & 3\,\mathbf{u}_3 & + & 4\,\mathbf{u}_4 \\ T(\mathbf{u}_3) & = & \mathbf{u}_1 & + & \mathbf{u}_2 & + & -10\,\mathbf{u}_3 & + & -12\,\mathbf{u}_4 \\ T(\mathbf{u}_4) & = & -\mathbf{u}_1 & + & -\mathbf{u}_2 & + & 9\,\mathbf{u}_3 & + & 11\,\mathbf{u}_4 \end{array}$$

En tal caso, la matriz del endomorfismo T respecto de la base $\mathcal{B}$ es

$$A = \begin{pmatrix} 1 & -1 & 1 & -1 \\ -1 & 0 & 1 & -1 \\ -1 & 3 & -10 & 9 \\ 0 & 4 & -12 & 11 \end{pmatrix}.$$

El polinomio característico de T es

$$\aleph_T(x) = |xI_n - A| = (x-1)^3(x+1),$$

luego T tiene dos autovalores distintos en $\mathbb{R}$, $\lambda_1 = 1$ de multiplicidad $m_1 = 3$ y $\lambda_2 = -1$ de multiplicidad $m_2 = 1$.

Como T tiene todos sus autovalores en $\mathbb{R}$, podemos calcular una base de V tal que la matriz de T respecto de ella es de Jordan.

- Tenemos que

$$\lambda_1 I_4 - A = \begin{pmatrix} 0 & 1 & -1 & 1 \\ 1 & 1 & -1 & 1 \\ 1 & -3 & 11 & -9 \\ 0 & -4 & 12 & -10 \end{pmatrix},$$

entonces $\operatorname{rg}(\lambda_1 \mathrm{Id}_V - T) = 3$, por lo que

$$n_{1,1} = \dim(L_{1,1}) = \dim(\ker(\lambda_1 \mathrm{Id}_V - T)) = 4 - \operatorname{rg}(\lambda_1 \mathrm{Id}_V - T) = 1 < 3 = m_1.$$

Nótese que, según el criterio de diagonalización por el polinomio característico, T no es diagonalizable.

Calculemos, pues, los subespacios propios generalizados del autovalor λ_1 :

- En primer lugar calculamos una base de $L_{1,1}$. Para ello resolvemos el sistema de ecuaciones lineales $(\lambda_1 I_4 - A)\mathbf{x} = \mathbf{0}$ y obtenemos que una base de $L_{1,1}$ expresada en coordenadas respecto de $\mathcal{B}$ es $\{(0,-1,3,4)\}$.
- Para el cálculo de $L_{1,2} = \ker\big((\lambda_1 \mathrm{Id}_V - T)^2\big)$ necesitamos obtener $(\lambda_1 I_4 - A)^2$.

$$(\lambda_1 I_4 - A)^2 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 1 \\ 8 & 1 & 15 & -11 \\ 8 & 0 & 16 & -12 \end{pmatrix},$$

entonces $\mathrm{rg}\big((\lambda_1 \mathrm{Id}_V - T)^2\big) = 2$, por lo que

$$\begin{aligned} n_{1,2} &= \dim(L_{1,2}) = \dim\big(\ker\big((\lambda_1 \mathrm{Id}_V - T)^2\big)\big) \\ &= 4 - \mathrm{rg}\big((\lambda_1 \mathrm{Id}_V - T)^2\big) = 2 < 3 = m_1. \end{aligned}$$

Luego, $L_{1,2}$ no es el subespacio propio máximo de λ_1.
A continuación ampliamos la base de $L_{1,1}$ a una base de $L_{1,2}$. Para ello resolvemos el sistema lineal de ecuaciones $(\lambda_1 I_4 - A)^2\mathbf{x} = \mathbf{0}$ y obtenemos que una base de $L_{1,2}$ expresada en coordenadas respecto de $\mathcal{B}$ es $\{(0,-1,3,4),(3,-2,\,0,2)\}$.

- Para el cálculo de $L_{1,3} = \ker\big((\lambda_1 \mathrm{Id}_V - T)^3\big)$ necesitamos obtener $(\lambda_1 I_4 - A)^3$.

$$(\lambda_1 I_4 - A)^3 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 16 & 8 & 24 & -16 \\ 16 & 8 & 24 & -16 \end{pmatrix},$$

entonces $\mathrm{rg}\big((\lambda_1 \mathrm{Id}_V - T)^2\big) = 1$, por lo que

$$\begin{aligned} n_{1,3} &= \dim(L_{1,3}) = \dim\big(\ker\big((\lambda_1 \mathrm{Id}_V - T)^3\big)\big) \\ &= 4 - \mathrm{rg}\big((\lambda_2 \mathrm{Id}_V - T)^3\big) = 3 = 3 = m_1. \end{aligned}$$

Luego, el subespacio propio máximo de λ_1 es $L_{1,3}$.
A continuación ampliamos la base de $L_{1,2}$ a una base de $L_{1,3}$. Para ello resolvemos el sistema lineal de ecuaciones $(\lambda_1 I_4 - A)^3\mathbf{x} = \mathbf{0}$ y obtenemos que una base de $L_{1,3}$ expresada en coordenadas respecto de $\mathcal{B}$ es $\{(0,-1,3,4),(3,-2,\,0,2),(1,0,0,1)\}$.

La partición de la multiplicidad del autovalor λ_1 es

$$p_{13} = n_{13} - n_{12} = 1,\ p_{12} = n_{12} - n_{11} = 1,\ p_{11} = n_{11} - 0 = 1.$$

Luego, el bloque de Jordan del autovalor λ_1 consiste en

$$\begin{array}{rl} p_{13} = 1 & \text{bloques de orden 3} \\ p_{12} - p_{13} = 0 & \text{bloques de orden 2} \\ p_{11} - p_{12} = 0 & \text{bloques de orden 1} \end{array},$$

esto es

$$\begin{pmatrix} \lambda_1 & 1 & 0 \\ 0 & \lambda_1 & 1 \\ 0 & 0 & \lambda_1 \end{pmatrix}$$

Para calcular la base canónica de Jordan de L_{13}, elegimos $p_{13} = 1$ vectores de L_{13} que sean linealmente independientes módulo L_{12}, por ejemplo, el vector $\mathbf{v}_{13}$ de coordenadas $(1, 0, 0, 1)$ respecto de $\mathcal{B}$, y calculamos los vectores $\mathbf{v}_{12} = -(\lambda_1 \mathrm{Id}_V - T)(\mathbf{v}_{13})$ y $\mathbf{v}_{11} = -(\lambda_1 \mathrm{Id}_V - T)(\mathbf{v}_{12})$; en nuestro caso $\mathbf{v}_{12}$ y $\mathbf{v}_{11}$ son los vectores de coordenadas $(-1, -2, 8, 10)$ y $(0, 1, -3, -4)$, respectivamente, respecto de $\mathcal{B}$. Finalmente, como $\{\mathbf{v}_{11},\ \mathbf{v}_{12}, \mathbf{v}_{13}\}$ es ya una base de L_{13}, por el teorema III.5.17, concluimos que es la base de Jordan del bloque asociado al autovalor λ_1.

- Por otra parte, tenemos que Tenemos que

$$\lambda_2 I_4 - A = \begin{pmatrix} -2 & 1 & -1 & 1 \\ 1 & -1 & -1 & 1 \\ 1 & -3 & 9 & -9 \\ 0 & -4 & 12 & -12 \end{pmatrix},$$

entonces $\mathrm{rg}(\lambda_2 \mathrm{Id}_V - T) = 3$, por lo que

$$n_{21} = \dim(L_{21}) = \dim(\ker(\lambda_2 \mathrm{Id}_V - T)) = 4 - \mathrm{rg}(\lambda_2 \mathrm{Id}_V - T) = 1 = m_2.$$

En este caso, L_{21} es el subespacio propio máximo del autovalor λ_2. Luego, $p_{21} = n_{21} - 0 = 1$, de tal forma que sólo hay 1 bloque de Jordan de orden 1 para el autovalor λ_2 y una base de Jordan la forma cualquier vector no nulo de L_{21}, por ejemplo, el vector $\mathbf{v}_{21}$ cuyas coordenadas respecto de $\mathcal{B}$ son $(0, 0, 1, 1)$.

Finalmente, por el teorema III.5.10, tenemos que $V = L_{13} \oplus L_{21}$; de donde se sigue que la base de Jordan de V es $\mathcal{B}' = \{\mathbf{v}_{11}, \mathbf{v}_{12}, \mathbf{v}_{13}, \mathbf{v}_{21}\}$ y que la matriz de Jordan de

T es

$$J = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}.$$

Además, si P es la matriz cuyas columnas son las coordenadas de los vectores de $\mathcal{B}'$ respecto de $\mathcal{B}$, es decir,

$$P = \begin{pmatrix} 0 & -1 & 1 & 0 \\ 1 & -2 & 0 & 0 \\ -3 & 8 & 0 & 1 \\ -4 & 10 & 1 & 1 \end{pmatrix},$$

se cumple que

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

Terminamos con una condición necesaria y suficiente para que dos matrices cuadradas sean semejantes.

Proposición III.5.20. *Dos matrices cuadradas A y $B \in \mathcal{M}_n(\Bbbk)$ con todos sus autovalores en $\Bbbk$ son semejantes si, y sólo si, tienen la misma forma canónica de Jordan.*

Demostración. Es claro que si A y B tienen la misma forma canónica de Jordan, son semejantes. Recíprocamente, si A y B son semejantes, entonces, por la proposición III.1.3, existen ciertas bases $\mathcal{B}$ y $\mathcal{B}'$ de V tales que $A = M_{\mathcal{B}}(T)$ y $B = M_{\mathcal{B}'}(T)$, para algún endomorfismo T de $\Bbbk^n$ (por ejemplo, $\Bbbk^n \to \Bbbk^n; \mathbf{v} \mapsto A\mathbf{v}$). Sabemos que la forma canónica de Jordan de T está determinada por sus autovalores, sus multiplicidades y las particiones de sus multiplicidades, que dependen exclusivamente de T, y no de la base elegida. Entonces A y B tienen la misma forma canónica de Jordan, la del endomorfismo T. ∎

En resumen, dos matrices cuadradas A y $B \in \mathcal{M}_n(\Bbbk)$ con todos sus autovalores en $\Bbbk$ son semejantes si, y sólo si, tienen los mismos los autovalores con idénticas multiplicidades y particiones de multiplicidades.

Definición III.5.21. Sean $A \in \mathcal{M}_n(\Bbbk)$ y $J = P^{-1}AP$ su forma canónica de Jordan. Se llama **descomposición espectral de** A a

$$A = PJP^{-1}.$$

En el siguiente tema veremos algunas aplicaciones concretas de la descomposición espectral de una matriz.

Ejercicios del tema III

Ejercicio 1. Dado el endomorfismo $T : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ definido por $T(x, y) = (x + y, x - y)$, obtener su matriz respecto de la base usual de $\mathbb{R}^2$. Obtener también las matrices de los endomorfismos $T^2 - \mathrm{Id}_{\mathbb{R}^2}$ y $T^3 = T \circ T \circ T$.

Ejercicio 2. Sea V un espacio vectorial de dimensión 2 y sea T un endomorfismo de V no nulo y **nilpotente** (se dice que un endomorfismo es nilpotente si existe un número natural $p > 1$ tal que $T^p = 0$, donde T^p es $T \circ \cdots \circ T$ p veces). Probar que existe una base de V respecto de la cual la matriz asociada a T es $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$. Aplicar lo anterior al endomorfismo del $\mathbb{C}$-espacio vectorial $\mathbb{C}^2$ cuya matriz asociada respecto cierta base es $\begin{pmatrix} \mathrm{i} & 1 \\ 1 & -\mathrm{i} \end{pmatrix}$.

Ejercicio 3. Dadas las matrices

$$A = \begin{pmatrix} 1 & -1 & 3 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix} \quad \text{y} \quad C = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 2 & 3 \\ 0 & 0 & 2 \end{pmatrix},$$

¿representan todas al mismo endomorfismo?

Ejercicio 4. Probar que el polinomio característico de una matriz $A \in \mathcal{M}_n(\Bbbk)$ está en $\Bbbk[x]$, es decir, en el anillo de polinomios en una indeterminada con coeficientes en $\Bbbk$, tiene grado n y es unitario (esto es, el coeficiente del término de grado más alto es 1).

Ejercicio 5. Sean $A_1, \ldots, A_r$ matrices tales que $A_i \in \mathcal{M}_{m_i}(\mathbb{R})$, $i = 1, \ldots, r$. Probar que si los autovalores de A_i son $\lambda_{i,1}, \ldots, \lambda_{i,s_i}$, $i = 1, \ldots, r$, entonces los autovalores de $A_1 \oplus \ldots \oplus A_r$ son $\{\lambda_{ij} \mid i = 1, \ldots, r;\ j = 1, \ldots, s_i\}$.

Ejercicio 6. Sea $\aleph_T(x) = a_0 + a_1 x + \ldots + a_{n-1} x^{n-1} + x^n$ el polinomio característico de un endomorfismo T de un $\Bbbk$ -espacio vectorial V de dimensión finita $n > 0$. Probar que el determinante de T es igual a $(-1)^n a_0$.

Ejercicio 7. Sea V un $\Bbbk$ -espacio vectorial de dimensión finita $n > 0$ y $T \in \mathrm{End}_{\Bbbk}(V)$ tal que $I_n + T^2 = 0$. Probar que T no tiene autovalores reales.

Ejercicio 8. Sean T y T' dos endomorfismos de un $\mathbb{C}$-espacio vectorial V de dimensión finita. Probar que si T y T' conmutan, entonces T y T' tienen autovectores comunes.

Ejercicio 9. Sea V un $\Bbbk$ -espacio vectorial de dimensión n y $T \in \mathrm{End}_{\Bbbk}(V)$ nilpotente. Probar que $\aleph_T(x) = x^n$. Concluir que los valores propios de un endomorfismo nilpotente son todos nulos. ¿Es cierto el recíproco?

Ejercicio 10. Sea V un $\Bbbk$ -espacio vectorial de dimensión finita $n > 0$ y $T \in \mathrm{End}_{\Bbbk}(V)$ tal que la suma de las entradas de cada una de las filas de su matriz asociada $A \in \mathcal{M}_n(\Bbbk)$ respecto de alguna base de V es igual 1 (es decir, A es una **matriz estocástica**). Probar que 1 es un autovalor de T.

Ejercicio 11. Sean V un $\Bbbk$ -espacio vectorial de dimensión finita y $T \in \mathrm{End}_{\Bbbk}(V)$ biyectivo, es decir, T es un automorfismo de V. Probar que λ es un autovalor de T si y sólo si $\lambda \neq 0$ y λ^{-1} es autovalor de T^{-1}.

Ejercicio 12. Comprobar que si $\{\lambda_1, \ldots, \lambda_r\}$ son los autovalores de una matriz A, entonces

1. Los autovalores de αA (siendo $\alpha \neq 0$) son $\{\alpha\lambda_1, \ldots, \alpha\lambda_r\}$. Un vector $\mathbf{v}$ es autovector de A asociado a λ_i si, y sólo si $\mathbf{v}$ es autovector de αA asociado a $\alpha\lambda_i$.
2. A es invertible si, y solo si, $0 \notin \{\lambda_1, \ldots, \lambda_r\}$ y en este caso, los autovalores de A^{-1} son $\{(\lambda_1)^{-1}, \ldots, (\lambda_r)^{-1}\}$. Un vector $\mathbf{v}$ es autovector de A asociado a λ_i si, y sólo si $\mathbf{v}$ es autovector de A^{-1} asociado a $(\lambda_i)^{-1}$.

Ejercicio 13. Probar que si $\lambda_1, \ldots, \lambda_n \in \Bbbk$ son autovalores (no necesariamente distintos) de una matriz $A \in \mathcal{M}_n(\Bbbk)$, entonces

1. $|A| = \lambda_1 \cdots \lambda_n$.
2. $\mathrm{tr}(A) = \lambda_1 + \ldots + \lambda_n$.

Ejercicio 14. Sean $V = \mathbb{R}^4$ y $T \in \mathrm{End}_{\Bbbk}(V)$ tal su matriz asociada respecto de la base usual de $\mathbb{R}^4$ es

$$(a)\ \begin{pmatrix} -1 & -2 & 3 & 2 \\ 0 & 1 & 1 & 0 \\ -2 & -2 & 4 & 2 \\ 0 & 0 & 0 & 2 \end{pmatrix}, \qquad (b)\ \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 3 & 5 \end{pmatrix}.$$

Estudiar si T es diagonalizable.

Ejercicio 15. Sean $V = \Bbbk^3$ y $T \in \mathrm{End}_{\Bbbk}(V)$ tal que su matriz asociada respecto de alguna base de V es

$$(a)\ \begin{pmatrix} a & -1 & 1 \\ 0 & 1 & 3 \\ 0 & 2 & 2 \end{pmatrix}, \qquad (b)\ \begin{pmatrix} 1 & a & b \\ 0 & 2 & c \\ 0 & 0 & 1 \end{pmatrix}, \qquad (c)\ \begin{pmatrix} 5 & 0 & 0 \\ 0 & -1 & b \\ 3 & 0 & a \end{pmatrix}$$

con a, b y $c \in \Bbbk$. Estudiar (según los valores de a, b y $c \in \Bbbk$), primero sobre $\Bbbk = \mathbb{R}$ y luego sobre $\Bbbk = \mathbb{C}$, si T es diagonalizable.

Ejercicio 16. Sean $V = \mathbb{R}^4$ y $T \in \mathrm{End}_{\Bbbk}(V)$ tal que su matriz asociada respecto de alguna base de V es

$$\begin{pmatrix} 1 & -1 & 0 & 0 \\ -4 & 1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 0 & a & 1 & 3 \end{pmatrix}.$$

Estudiar, según el valor de $a \in \mathbb{R}$, si T es diagonalizable, y calcular, cuando sea posible, una base de V respecto de cual la matriz de T sea diagonal.

Ejercicio 17. Sean $V = \mathbb{R}^3$ y $T \in \mathrm{End}_{\Bbbk}(V)$ tal que su matriz asociada respecto de la base usual de $\mathbb{R}^3$ es cada una de las matrices del ejercicio 15 para las cuales T es diagonalizable. Hallar una base de V respecto de cual la matriz de T sea diagonal.

Ejercicio 18. Sean $V = \mathbb{R}^3$ y T y $T' \in \mathrm{End}_{\Bbbk}(V)$ tales que $T(v_1, v_2, v_3) = (v_1 + v_2 + v_3, 2v_1 + 5v_2 + 2v_3, -2v_1 - 5v_2 - 2v_3)$ y $T'(v_1, v_2, v_3) = (-2v_2 - 2v_3, 0, 2v_2 + 2v_3)$, para cada $\mathbf{v} = (v_1, v_2, v_3) \in \mathbb{R}^3$. Hallar, si es posible, sendas bases de V respecto de las cuales las matrices de T y T' sean diagonales.

Ejercicio 19. Sean V un espacio vectorial de dimensión finita sobre un cuerpo $\Bbbk$ y T un endomorfismo de V. Probar que

1. Si $\Bbbk = \mathbb{C}$ y V no tiene subespacios invariantes por T distintos del cero y el total, entonces la dimensión de V es 1.
2. Si $\Bbbk = \mathbb{R}$ y V no tiene subespacios invariantes por T distintos del cero y el total, entonces la dimensión de V es menor o igual que dos.

Ejercicio 20. Sean T y S dos endomorfismos de un $\Bbbk$ -espacio vectorial V de dimensión finita. Probar:

(a) Si T es diagonalizable, entonces para todo subespacio L de V que es invariante por T el endomorfismo $T|_L$ también es diagonalizable.
(b) Si T y S conmutan, entonces los subespacios invariantes asociados a los autovalores de T son los subespacios invariantes asociados a los autovalores de S, y recíprocamente.
(c) Los endomorfismos T y S son simultáneamente diagonalizables (esto es, existe una base de V formada por autovectores de los dos endomorfismos) si y sólo si T y S son diagonalizables y conmutan.

Ejercicio 21. Clasificar los endomorfismos de un espacio vectorial sobre $\mathbb{R}$ de dimensión 4 que:

1. Tienen un único autovalor real.
2. No tienen ningún autovalor real.
3. Tienen dos autovalores reales distintos.

4. Tienen al menos un autovalor real.
5. Tienen al menos tres autovalores reales.
6. Tienen un único factor invariante.

Ejercicio 22.

Calcular la forma canónica y la base de Jordan de los siguientes endomorfismos cuyas matrices respecto de la base canónica del correspondiente $\mathbb{C}$-espacio vectorial son:

$$\text{(a)} \begin{pmatrix} 3 & -2 & 0 \\ -2 & 3 & 0 \\ 0 & 0 & 5 \end{pmatrix}, \quad \text{(b)} \begin{pmatrix} -14 & 1 & 12 \\ -13 & 0 & 12 \\ -17 & 1 & 15 \end{pmatrix}, \quad \text{(c)} \begin{pmatrix} -1 & 2 & -1 \\ -2 & 3 & -2 \\ -2 & 2 & -1 \end{pmatrix}$$

$$\text{(d)} \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 3 & 5 \end{pmatrix}, \quad \text{(e)} \begin{pmatrix} 3 & 45 & 37 & -9 \\ 2 & 12 & 8 & -5 \\ -2 & -4 & -1 & 4 \\ 3 & 33 & 26 & -8 \end{pmatrix},$$

$$\text{(f)} \begin{pmatrix} 3 & 67 & 59 & -9 \\ 2 & -16 & -20 & -5 \\ -2 & 28 & 31 & 4 \\ 3 & 31 & 24 & -8 \end{pmatrix}, \quad \text{(g)} \begin{pmatrix} 3 & 17 & 9 & -9 \\ 2 & 16 & 12 & -5 \\ -2 & -12 & -9 & 4 \\ 3 & 17 & 10 & -8 \end{pmatrix},$$

$$\text{(h)} \begin{pmatrix} 3 & 45 & 37 & -9 \\ 2 & 10 & 6 & -5 \\ -2 & -2 & 1 & 4 \\ 3 & 32 & 25 & -8 \end{pmatrix}, \quad \text{(i)} \begin{pmatrix} 3 & 31 & 23 & -9 \\ 2 & 7 & 3 & -5 \\ -2 & -1 & 2 & 4 \\ 3 & 21 & 14 & -8 \end{pmatrix},$$

$$\text{(j)} \begin{pmatrix} 3 & 42 & 34 & -9 \\ 2 & -29 & -33 & -5 \\ -2 & 38 & 41 & 4 \\ 3 & 7 & 0 & -8 \end{pmatrix},$$

Ejercicio 23. Sean V un espacio vectorial de dimensión 25 y f un endomorfismo de V. Si $\aleph_T(x) = (x-1)^{25}$ $\dim(\ker(f-1)) = 11$, $\dim(\ker((f-1)^2)) = 16$, $\dim(\ker((f-1)^3)) = 19$, $\dim(\ker((f-1)^4)) = 22$ y $\dim(\ker((f-1)^5)) = 25$, escribir la forma canónica de Jordan de f.

TEMA IV

Potencias de matrices. Matrices no negativas

Este tema bien se podría denominar "algunas aplicaciones de la forma canónica de Jordan" , ya que vamos a usar la forma canónica de Jordan como herramienta de resolución de problemas concretos.

Así, la primera sección está dedicada a la obtención de un expresión general para la potencia m-ésima de una matriz A de la que conocemos su forma canónica de Jordan J y una matriz de invertible P tal que $P^{-1}AP = J$. Esta fórmula se aplica, por ejemplo, para calcular el término general de la solución de una ecuación lineal homogénea en diferencias finitas con coeficientes constantes con condición inicial dada; dedicamos la segunda parte de esta sección a la resolución de este tipo de ecuaciones. En primer lugar, transformamos la ecuación en diferencias en un sistema de ecuaciones en diferencias, escribimos el sistema matricialmente y concluimos que el término general x_{n+p} de la solución de la ecuación en diferencias se obtiene a partir de la fórmula de la potencia n-ésima de la matriz del sistema. Cabe destacar que A es una matriz de coeficientes reales; luego, en principio podría parecer que necesitamos la forma canónica real de A, que no ha sido estudiada en el tema anterior. Sin embargo, podemos prescindir de ella (al menos formalmente), tratando el problema sobre los complejos habida cuenta de que $A^n = PJ^nP^{-1}$ tiene que tener coeficientes reales, aún cuando la forma de Jordan, J, y la matriz de paso, P, tengan coeficientes complejos; tal y como queda reflejado en el teorema IV.2.3.

La segunda sección lleva por título matrices no negativas. Una matriz no negativa es aquella cuyas entradas son números reales positivos o nulos. Nótese que la matrices no negativas son fundamentales en Estadística y Probabilidad, ya que las matrices estocásticas, las matrices de Leontieff y de Leslie son no negativas. En realidad, nosotros nos centraremos en las matrices no negativas irreducibles y posteriormente en las primitivas por sus buenas e interesantes propiedades espectrales.

Las matrices no negativas e irreducibles tienen la particularidad de poseer un autovalor real positivo ρ de multiplicidad 1 con un autovector positivo asociado tal que $|\lambda| \leq \rho$, para todo autovalor (real o complejo) λ de A. Este es el resultado principal de esta parte del tema, y se denomina Teorema de Perron-Fröbenius. El autovalor ρ de una matriz no negativa e irreducible A se llama autovalor de Perron

de A y el autovector positivo asociado a ρ cuyas entradas suman 1 se llama autovector de Perron.

Una matriz no negativa A tal que $A^m > 0$ para algún m, se dice que es primitiva. Las matrices primitivas son irreducibles, y además cumplen que su autovalor de Perron es estrictamente mayor en módulo que cualquier otro de sus autovalores.

Terminamos esta sección mostrando un interesante ejemplo sobre un modelo poblacional basado en matrices irreducibles no negativas: el llamado modelo matricial de Leslie. Este ejemplo ilustra a la perfección el interés práctico de las matrices irreducibles no negativas, y por añadidura, el estudio de los autovalores y autovectores de una matriz.

La última sección del tema lleva por nombre "cadenas de Markov homogéneas y finitas" y sirve como introducción teórica para la práctica 7.

Las ecuaciones en diferencias estudiadas en este tema aparecen en la asignatura *Series Temporales* en el estudio de los modelos autorregresivos (véase el capítulo 15 de [**dR87**]); más concretamente para el cálculo de las funciones de autocorrelación simple de los modelos mixtos autorregresivos-media móvil. Prescindiendo de los nombres, basta decir que estos modelos están definidos por una ecuación lineal homogénea en diferencias finitas con coeficientes constantes.

Para la elaboración de la primera parte de este tema, hemos usado los capítulos 9 y 10 de [**FVV03**] pero con la vista puesta en la sección quinta del capítulo 10 de [**Her85**]. En los capítulos citados de [**FVV03**] se pueden encontrar multitud de ejemplos interesantes del uso práctico de las ecuaciones en diferencias estudiadas en este tema. Para las dos últimas secciones hemos seguido el capítulo 8 de [**Mey00**], aunque también hemos utilizado parcialmente la sección 8 del capítulo 8 de [**Sch05**], del capítulo 7 de [**Sea82**] y del capítulo 1 de [**Sen81**].

1. Potencias de matrices

En la primera parte de este tema vamos calcular una expresión general para la potencia m-ésima de una matriz $A \in \mathcal{M}_n(\Bbbk)$, a partir de su forma canónica de Jordan.

Teorema IV.1.1. *Sean $A \in \mathcal{M}_n(\Bbbk)$. Si $J = P^{-1}AP$ es la forma canónica de Jordan de A, entonces*

$$A^m = PJ^mP^{-1}.$$

Demostración. Basta tener en cuenta que si $J = P^{-1}AP$, entonces $A = PJP^{-1}$, de donde se sigue que $A^m = (PJP^{-1})^m = PJ^mP^{-1}$. ∎

El teorema anterior reduce el cálculo de la potencia m-ésima de A al del cálculo de la potencia m-ésima de su forma canónica de Jordan, que como sabemos es una matriz diagonal por bloques (de Jordan). Teniendo en cuenta que el producto de matrices diagonales por bloques se calcula efectuando los correspondientes productos

de los bloques, para obtener una expresión general de la potencia m-ésima de una matriz de Jordan basta determinar cuál es la potencia m-ésima de un bloque de Jordan.

Proposición IV.1.2. *Sea $B \in \mathcal{M}_s(\Bbbk)$ un bloque Jordan de orden s. Si $\lambda \in \Bbbk$ es una entrada de la diagonal principal de B, entonces*

$$B^m = \begin{pmatrix} \lambda^m & \binom{m}{1}\lambda^{m-1} & \binom{m}{2}\lambda^{m-2} & \dots & \binom{m}{s-1}\lambda^{m-s+1} \\ 0 & \lambda^m & \binom{m}{1}\lambda^{m-1} & \dots & \binom{m}{s-2}\lambda^{m-s+2} \\ 0 & 0 & \lambda^m & \dots & \binom{m}{s-3}\lambda^{m-s+3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \lambda^m \end{pmatrix}, \tag{IV.1.1}$$

entendiendo que $\binom{m}{r} = 0$ si $m < r$.

Demostración. Sabemos que B es la suma de la matriz diagonal $D_\lambda \in \mathcal{M}_s(\Bbbk)$ y la matriz nilpotente

$$N = \begin{pmatrix} 0 & 1 & \dots & 0 & 0 \\ 0 & 0 & \ddots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & \dots & 0 & 0 \end{pmatrix} \in \mathcal{M}_s(\Bbbk).$$

Como D_λ conmuta con cualquier matriz cuadra de orden s y $N^{s-1} \neq 0$ y $N^m = 0,\ m \geq s$, se tiene que

$$\begin{aligned} B^m &= (D_\lambda + N)^m \\ &= (D_\lambda)^m + \binom{m}{1}(D_\lambda)^{m-1}N + \binom{m}{2}(D_\lambda)^{m-2}N^2 + \ldots + \binom{m}{s-1}(D_\lambda)^{m-s+1}N^{s-1} \\ &= \lambda^m I_s + \binom{m}{1}\lambda^{m-1}N + \binom{m}{2}\lambda^{m-2}N^2 + \ldots + \binom{m}{s-1}\lambda^{m-s+1}N^{s-1}, \end{aligned}$$

de donde se sigue la expresión buscada. ∎

Por consiguiente, la expresión general de la potencia m-ésima de $A \in \mathcal{M}_n(\Bbbk)$ es

$$A^m = P \begin{pmatrix} B_1^m & 0 & \dots & 0 \\ 0 & B_2^m & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & B_t^m \end{pmatrix} P^{-1},$$

donde $P^{-1}AP$ es la forma canónica de Jordan de A y cada B_j^m es la potencia m-ésima de un bloque Jordan, esto es, una matriz de la forma (IV.1.1).

Ejemplo IV.1.3. La matriz

$$A = \begin{pmatrix} 7/2 & -6 \\ 1/2 & 0 \end{pmatrix}$$

es claramente diagonalizable, pues tiene dos autovalores distintos $\lambda_1 = 2$ y $\lambda_2 = 3/2$. Su forma canónica de Jordan es

$$J = \begin{pmatrix} 2 & 0 \\ 0 & 3/2 \end{pmatrix}$$

y una matriz de paso es

$$P = \begin{pmatrix} 1 & -3 \\ -1/2 & 2 \end{pmatrix}.$$

Por consiguiente, la expresión general de la potencia m-ésima de A es

$$A^m = PJ^mP^{-1} = \begin{pmatrix} 4 & 6 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 2^m & 0 \\ 0 & (3/2)^m \end{pmatrix} \begin{pmatrix} 1 & -3 \\ -1/2 & 2 \end{pmatrix}.$$

Obsérvese que la expresión anterior para potencia m-ésima de A se puede obtener siempre (independientemente de que A tenga todos sus autovalores en $\Bbbk$ o no), ya que si bien la matriz de Jordan de A puede tener sus entradas en una extensión del cuerpo $\Bbbk$ (por ejemplo, si $\Bbbk = \mathbb{R}$ y alguno de los autovalores de A está en $\mathbb{C}$), el resultado final A^m pertenece claramente a $\mathcal{M}_n(\Bbbk)$.

Ejemplo IV.1.4. La matriz

$$A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \in \mathcal{M}_2(\mathbb{R})$$

tiene dos autovalores complejos $\lambda = \mathrm{i}$ y $\bar{\lambda} = -\mathrm{i}$. Su forma canónica compleja es

$$J = \begin{pmatrix} \mathrm{i} & 0 \\ 0 & -\mathrm{i} \end{pmatrix}$$

y una matriz de paso es

$$P = \begin{pmatrix} 1 & 1 \\ -\mathrm{i} & \mathrm{i} \end{pmatrix}.$$

Por consiguiente,

$$\begin{aligned} A^m = PJ^mP^{-1} &= \frac{1}{2} \begin{pmatrix} 1 & 1 \\ -\mathrm{i} & \mathrm{i} \end{pmatrix} \begin{pmatrix} \mathrm{i}^m & 0 \\ 0 & (-\mathrm{i})^m \end{pmatrix} \begin{pmatrix} 1 & \mathrm{i} \\ 1 & -\mathrm{i} \end{pmatrix} \\ &= 1/2 \begin{pmatrix} \mathrm{i}^m + (-\mathrm{i})^m & \mathrm{i}^{m+1} + (-\mathrm{i})^{m+1} \\ -\mathrm{i}^{m+1} - (-\mathrm{i})^{m+1} & -(\mathrm{i}^{m+2} + (-\mathrm{i})^{m+2}) \end{pmatrix}, \end{aligned}$$

que, aunque no lo parezca, es una matriz real.

2. Ecuaciones en diferencias finitas

Definición IV.2.1. Dados $a_1, \ldots, a_p \in \mathbb{R}$, con $a_p \neq 0$, se llama **ecuación lineal en diferencias finitas con coeficientes constantes** de orden p a una relación de recurrencia del tipo

$$x_{n+p} - a_1 x_{n+p-1} - \ldots - a_p x_n = \varphi(n), \quad \text{para todo } n \geq 1 \tag{IV.2.2}$$

donde $\varphi : \mathbb{N} \to \mathbb{R}$ es una función.

Si $\varphi(n) = 0$, para todo $n \in \mathbb{N}$, se dice que la ecuación lineal en diferencias con coeficientes constantes (IV.2.2) es **homogénea**.

Una solución para la ecuación (IV.2.2) es una sucesión $\{x_n\}_{n\geq 1}$ que la satisfaga.

Ejemplo IV.2.2. La ecuación $x_{n+2} = x_{n+1} + x_n,\ n \geq 1$, es una ecuación lineal en diferencias con coeficientes constantes homogénea. Mas adelante (en el ejemplo IV.2.5) veremos que una de sus soluciones es la sucesión de Fibonacci.

A continuación, vamos a hallar una expresión explícita de x_n en función de n tal que la sucesión $\{x_n\}_{n\geq 1}$ sea solución de la ecuación (IV.2.2) en el caso homogéneo. El caso no homogéneo puede consultarse en [**FVV03**] por ejemplo.

Consideremos la ecuación lineal en diferencias con coeficientes constantes homogénea de orden p

$$x_{n+p} - a_1 x_{n+p-1} - \ldots - a_p x_n = 0, \quad \text{para todo } n \geq 1. \tag{IV.2.3}$$

Para cada $n \geq 1$, se tiene el siguiente sistema de ecuaciones lineales (en diferencias con coeficientes constantes)

$$\left.\begin{array}{rcccccccc} x_{n+p} & = & a_1\, x_{n+p-1} & + & \ldots & + & a_{p-1}\, x_{n+1} & + & a_p\, x_n \\ x_{n+p-1} & = & x_{n+p-1} & & & & & & \\ & \vdots & & & & & & & \\ x_{n+1} & = & & & & & x_{n+1} & & \end{array}\right\}$$

cuya matriz es

$$A = \begin{pmatrix} a_1 & a_2 & \ldots & a_{p-1} & a_p \\ 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & 0 \end{pmatrix}, \tag{IV.2.4}$$

que llamaremos **matriz asociada a la ecuación en diferencias**[1]. De tal forma que, si, para cada $n \geq 1$, denotamos $\mathbf{x}_n = (x_{n+p}, x_{n+p-1}, \ldots, x_{n+1})^{\mathrm{t}} \in \Bbbk^p$, entonces

$$\mathbf{x}_n = A\mathbf{x}_{n-1} = A^2\mathbf{x}_{n-2} = \ldots = A^n\mathbf{x}_0.$$

De donde se sigue que *el término general x_{n+p} de cualquier solución de la ecuación en diferencias (IV.2.3) es una combinación lineal de las entradas de la primera fila de A^n.*

Dado que sabemos cómo calcular una expresión general para las potencias de cualquier matriz cuadrada, vamos a tratar de afinar un poco más la afirmación anterior.

Teorema IV.2.3. *Sean $a_1, \ldots, a_p \in \mathbb{R}$, con $a_p \neq 0$. El término general de la solución de la ecuación en diferencias $x_{n+p} = a_1 x_{n+p-1} + \ldots + a_p x_n$, para todo $n \geq 1$, es una combinación lineal con coeficientes reales de*

$$\lambda^n, n\lambda^n, \ldots, n^{m-1}\lambda^n,$$

para cada autovalor real λ de multiplicidad m de la matriz de la ecuación en diferencias y de

$$\rho^n\cos(n\theta), n\rho^n\cos(n\theta), \ldots, n^{m-1}\rho^n\cos(n\theta),$$
$$\rho^n\mathrm{sen}(n\theta), n\rho^n\mathrm{sen}(n\theta), \ldots, n^{m-1}\rho^n\mathrm{sen}(n\theta),$$

para cada autovalor complejo $\lambda = \rho(\cos(\theta) + \mathrm{i}\,\mathrm{sen}(\theta))$ de multiplicidad m de la matriz de la ecuación en diferencias.

Demostración. Sea $A \in \mathcal{M}_p(\mathbb{R})$ la matriz de la ecuación en diferencias.

Sabemos que el término general x_{n+p} de cualquier solución de la ecuación en diferencias es una combinación lineal con coeficientes en $\mathbb{R}$ de las entradas de la primera fila de A^n. Por consiguiente, si $J = P^{-1}AP$ es la forma canónica de Jordan de A, entonces, por el teorema IV.1.1, $A^n = PJ^nP^{-1}$, de donde se sigue que las entradas de la primera fila de A serán combinaciones lineales de las entradas de J^n; estas entradas son, en virtud de la proposición IV.1.2,

$$\lambda^n, \binom{n}{1}\lambda^{n-1}, \binom{n}{2}\lambda^{n-2}, \ldots, \binom{n}{m-1}\lambda^{n-m+1},$$

para cada autovalor λ de A, siendo m su multiplicidad (pues los bloques de Jordan son a lo sumo de orden m).

[1] Asimismo, se llama **polinomio característico de la ecuación en diferencias**. al polinomio característico de A. Se comprueba fácilmente, por inducción en p, que $\aleph_A(x) = x^p - a_1x^{p-1} - \ldots - a_{p-1}x - a_p$. No obstante, es tradición en la teoría de series temporales denominar polinomio característico de la ecuación en diferencias a $p(y) = 1 - a_1y - \ldots - a_py^p$, esto es, $-a_p\aleph_{A^{-1}}(y)$ (véase el apéndice 15A de [**dR87**])

Teniendo ahora en cuenta que, para cada $s = 1, \ldots, m-1$,

$$\binom{n}{s}\lambda^{n-s} = \frac{\lambda^{-s}}{s!}\big(n(n-1)\cdots(n-s+1)\big)\lambda^n$$
$$= \frac{\lambda^{-s}}{s!}\big(n^s + b_{1s}n^{s-1} + \ldots + b_{s-1,s}n\big)\lambda^n,$$

para ciertos $b_{s1}, \ldots, b_{s-1,s} \in \mathbb{R}$, concluimos que las entradas de PJ^nP^{-1} son combinaciones lineales de

$$\lambda^n, n\lambda^n, \ldots, n^{m-1}\lambda^n,$$

para cada autovalor λ de A, siendo m su multiplicidad.

Finalmente, si λ es un autovalor complejo de A, entonces $\bar{\lambda}$ también es un autovalor (complejo) de A. Dado que $\lambda = \rho(\cos(\theta) + \mathrm{i}\,\mathrm{sen}(\theta))$ y, consecuentemente, $\bar{\lambda} = \rho(\cos(\theta) - \mathrm{i}\,\mathrm{sen}(\theta))$, se sigue que las combinaciones lineales de

$$\lambda^n, n\lambda^n, \ldots, n^{m-1}\lambda^n, \bar{\lambda}^n, n\bar{\lambda}^n, \ldots, n^{m-1}\bar{\lambda}^n,$$

son combinaciones con coeficientes reales de

$$\rho^n\cos(n\theta), n\rho^n\cos(n\theta), \ldots, n^{m-1}\rho^n\cos(n\theta),$$

$$\rho^n\mathrm{sen}(n\theta), n\rho^n\mathrm{sen}(n\theta), \ldots, n^{m-1}\rho^n\mathrm{sen}(n\theta),$$

para cada autovalor complejo $\lambda = \rho(\cos(\theta)+\mathrm{i}\,\mathrm{sen}(\theta))$ de A, siendo m su multiplicidad, habida cuenta que $\lambda^n = \rho^n(\cos(n\theta) + \mathrm{i}\,\mathrm{sen}(n\theta))$. ∎

Corolario IV.2.4. *Sean $a_1, \ldots, a_p \in \mathbb{R}$, con $a_p \neq 0$. Si la matriz $A \in \mathcal{M}_p(\mathbb{R})$ de la ecuación en diferencias $x_{n+p} - a_1x_{n+p-1} - \ldots - a_px_n = 0$, para todo $n \geq 1$ es diagonalizable y $\lambda_1, \ldots, \lambda_r \in \mathbb{R}$ son los autovalores distintos de A (en particular, si los autovalores de A son reales y distintos, véase el corolario III.3.6), entonces el término general de la solución general de la ecuación en diferencias es*

$$x_{n+p} = c_1\lambda_1^n + c_2\lambda_2^n + \ldots + c_r\lambda_r^n,$$

donde $c_1, c_2, \ldots, c_r \in \mathbb{R}$ son constantes arbitrarias.

Demostración. Si A es diagonalizable y $J = P^{-1}AP$ es la forma canónica de Jordan de A, entonces, por el teorema IV.1.1, $A^n = PJ^nP^{-1}$, de donde se sigue que las entradas de la primera fila de A serán combinaciones lineales de las entradas de J^n; es decir, de $\lambda_1^n, \ldots, \lambda_r^n$, ya que al ser A diagonalizable, se tiene que J es una matriz diagonal y las entradas de su diagonal principal son precisamente los autovalores de A repetidos tantas veces como indique su multiplicidad (véase la nota III.3.2). ∎

El *término general de la solución* de una ecuación lineal en diferencias con coeficientes constantes de orden p depende de p constantes arbitrarias. Si en la solución

general se dan valores particulares a las p constantes, se obtiene una *solución particular*. En general, las p constantes se determinan a partir de p condiciones adicionales llamadas *condiciones iniciales*.

Ejemplo IV.2.5. **La sucesión de Fibonacci.**

Leonardo Fibonacci (o Leonardo de Pisa, 1175-1230) planteó en su *Liber abaci* el siguiente problema: *Un hombre pone una pareja de conejos en un lugar cercado por todos lados. ¿Cuántos conejos tendrá al cabo de un año si se supone que cada pareja engendra cada mes una nueva pareja que, a su vez, es fértil a partir del segundo mes de vida?*

Se supones además que no muere ningún conejo. Sea F_n el número de parejas existentes al cabo del mes n-ésimo; se comienza con una pareja recién nacida: $F_1 = 1$; al final del primer mes esa pareja todavía no es fértil, así que sigue teniéndose $F_2 = 1$; al final del segundo mes la pareja anterior, ya fértil, da origen a una nueva pareja: $F_3 = 1 + 1 = F_2 + F_1$. Y en general, se tendrá

$$F_{n+2} = F_{n+1} + F_n,\ n \geq 1 \tag{IV.2.5}$$

pues por la ley supuesta, cada mes nacen tantas parejas como parejas había dos meses antes.

Empezando con $F_0 = 1$ y $F_1 = 1$, se tiene la sucesión

$$1, 1, 2, 3, 5, 8, 13, 21, 34, 55, \ldots$$

Esta es la sucesión de Fibonacci; aparece en una variedad increíble de contextos y está relacionada con la *sección áurea* de los griegos (véase [**FVV03**] pp. 543-548).

La ecuación característica de (IV.2.5) es

$$x^2 - x - 1$$

con lo que los autovalores son

$$\lambda_1 = \frac{1+\sqrt{5}}{2},\ \lambda_2 = \frac{1-\sqrt{5}}{2}.$$

La solución general de (IV.2.5) es, por el corolario IV.2.4, será pues

$$F_{n+2} = c_1 \left(\frac{1+\sqrt{5}}{2}\right)^n + c_2 \left(\frac{1+\sqrt{5}}{2}\right)^n,\ n \geq 1,\ c_1, c_2 \in \mathbb{R}. \tag{IV.2.6}$$

La *sucesión de Fibonacci* corresponde a los datos $F_1 = 1$ y $F_2 = 1$; imponiendo estas condiciones iniciales en la fórmula (IV.2.6) se obtienen los valores

$$c_1 = \frac{1}{\sqrt{5}}\frac{1+\sqrt{5}}{2}, c_2 = -\frac{1}{\sqrt{5}}\frac{1-\sqrt{5}}{2}$$

con lo que la expresión de su término general es

$$F_{n+2} = \frac{1}{\sqrt{5}} \left[\left(\frac{1+\sqrt{5}}{2} \right)^{n+1} - \left(\frac{1-\sqrt{5}}{2} \right)^{n+1} \right].$$

Nótese que esta fórmula genera números naturales a pesar de contener expresiones irracionales.

3. Matrices no negativas

Definición IV.3.1. Una matriz $A = (a_{ij}) \in \mathcal{M}_n(\mathbb{R})$ es **no negativa**, lo que denotaremos por $A \geq 0$, si $a_{ij} \geq 0$, para todo $i, j \in \{1, \ldots, n\}$. Si $a_{ij} > 0$, para todo $i, j \in \{1, \ldots, n\}$, diremos que la matriz A es **positiva** y lo denotaremos por $A > 0$.

Definición IV.3.2. Sea $n \geq 2$. Se dice que una matriz $A \in \mathcal{M}_n(\mathbb{R})$ es **irreducible** si no existe ninguna matriz de permutación[2] $P \in \mathcal{M}_n(\Bbbk)$ tal que

$$PAP^{\mathrm{t}} = \begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix},$$

donde A_{11} (y A_{22}) es cuadrada de orden menor que n; en otro caso, se dice que A **reducible**.

Nótese que si T es el endomorfismo de $\mathbb{R}^n$ cuya matriz asociada respecto de una base $\mathcal{B}$ (por ejemplo la base usual) de $\mathbb{R}^n$ es A, la condición necesaria y suficiente para que A sea irreducible es que no exista ningún subconjunto de vectores de $\mathcal{B}$ que genere un subespacio de $\mathbb{R}^n$ invariante por T.

Proposición IV.3.3. *Sea $A \in \mathcal{M}_n(\mathbb{R})$. Si A es no negativa e irreducible, entonces*

$$(I_n + A)^{n-1}\mathbf{v} > 0,$$

para todo $\mathbf{v} \in V$ no nulo; en particular, $(I_n + A)^{n-1} > 0$.

Demostración. Consideremos un vector $\mathbf{v} \in \mathbb{R}^n$ no nulo tal que $\mathbf{v} \geq 0$ y escribamos

$$\mathbf{w} = (I_n + A)\mathbf{v} = \mathbf{v} + A\mathbf{v}.$$

Como $A \geq 0$, el producto $A\mathbf{v} \geq 0$, por lo que $\mathbf{w}$ tiene, al menos, tantas entradas no nulas, y por tanto positivas, como $\mathbf{v}$. Vamos a probar que si $\mathbf{v}$ no es ya positivo, entonces el vector $\mathbf{w}$ tiene al menos una entrada no nula más que $\mathbf{v}$. Si $P \in \mathcal{M}_n(\Bbbk)$ es una matriz de permutación tal que

$$P\mathbf{v} = \begin{pmatrix} \mathbf{u} \\ \mathbf{0} \end{pmatrix}$$

[2]Una **matriz de permutación** es un producto de matrices correspondientes a transformaciones elementales de tipo I. Recuérdese que si P es una matriz permutación, entonces $P^{-1} = P^{\mathrm{t}}$.

y $\mathbf{u} > 0$, entonces

$$P\mathbf{w} = P(I_n + A)\mathbf{v} = P(I_n + A)P^{\mathrm{t}} \begin{pmatrix} \mathbf{u} \\ \mathbf{0} \end{pmatrix} = \begin{pmatrix} \mathbf{u} \\ \mathbf{0} \end{pmatrix} + PAP^{\mathrm{t}} \begin{pmatrix} \mathbf{u} \\ \mathbf{0} \end{pmatrix}, \tag{IV.3.7}$$

ya que $PP^{\mathrm{t}} = I_n$. Si agrupamos las entradas de $P\mathbf{w}$ y de PAP^{t} de forma consistente con la de $P\mathbf{v}$

$$P\mathbf{w} = \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \quad \text{y} \quad PAP^{\mathrm{t}} = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix},$$

entonces, de (IV.3.7) se sigue que

$$\mathbf{x} = \mathbf{u} + A_{11}\mathbf{u} \quad \text{e} \quad \mathbf{y} = A_{21}\mathbf{u}.$$

Como A es no negativa e irreducible, se tiene que $A_{11} \geq 0, A_{21} \geq 0$ y $A_{21} \neq 0$, por lo que $\mathbf{x} > 0$ y $\mathbf{y} \geq 0$; además, como $\mathbf{u} > 0$, se tiene que $\mathbf{y} \neq 0$. Así, concluimos que $\mathbf{w}$ tiene al menos una componente no nula más que $\mathbf{v}$.

Si $\mathbf{w} = (I_n + A)\mathbf{v}$ no es ya un vector positivo, repetimos el argumento anterior con $\mathbf{w}$, y entonces $(I_n + A)^2\mathbf{v}$ tiene, al menos, dos componentes positivas más que $\mathbf{v}$. De este modo, después de a lo más $n - 1$ pasos encontramos que

$$(I_n + A)^{n-1}\mathbf{v} > 0,$$

para cualquier vector no nulo $\mathbf{v} \geq 0$.

Finalmente, tomando $\mathbf{v} = \mathbf{e}_i,\ i = 1, 2, \ldots, n$, donde $\mathbf{e}_i$ es el vector i-ésimo de la base usual de $\mathbb{R}^n$, concluimos que $(I_n + A)^{n-1} > 0$. ∎

Veamos ahora un **criterio práctico para determinar si una matriz** $A \in \mathcal{M}_n(\mathbb{R})$ **es irreducible:**

El concepto de matriz irreducible no está asociado con las magnitudes o con los signos, sino con la disposición de las entradas nulas y no nulas en la matriz. De modo que, para estudiar si una matriz dada es irreducible, podemos pensar que todos las entradas no nulas son unos, obteniéndose de este modo la matriz de adyacencia de un grafo dirigido.

Más concretamente, sean $A = (a_{ij}) \in \mathcal{M}_n(\mathbb{R})$ una matriz cualquiera y $\mathcal{G}_A = (V, E)$ es el grafo dirigido cuyo conjunto de vertices es $V = \{1, \ldots, n\}$ tal que $(i, j) \in E$ si, y sólo si, $a_{ij} \neq 0$ (obsérvese que la matriz de adyacencias de $\mathcal{G}$ es $\overline{A} = (\bar{a}_{ij}) \in \mathcal{M}_n(\mathbb{R})$ con $\bar{a}_{ij} = 1$ si $a_{ij} \neq 0$ y cero en otro caso.

Definición IV.3.4. Sea dice que un **grafo dirigido** $\mathcal{G}_A = (V, E)$ es **fuertemente conexo** si para cada par de vértices $i, j \in V$ existe un camino dirigido $(i, i_1), (i_1, i_2), \ldots, (i_s, j) \in E$ que conecta i con j.

Obsérvese que podría haber un camino dirigido de i a j pero no de j a i.

Lema IV.3.5. *Sea $A = (a_{ij}) \in \mathcal{M}_n(\mathbb{R})$. Si existe i ó j tal que $a_{ij} = 0$, para todo $i \neq j$, entonces $\mathcal{G}_A$ no es fuertemente conexo.*

Demostración. Por simplicidad supongamos que $a_{12} = \ldots = a_{1n} = 0$. Entonces, no hay ninguna flecha que comience en el vértice i. Luego, no hay conexión desde el vértice i hacía ningún otro. ∎

Lema IV.3.6. *Sea $A \in \mathcal{M}_n(\mathbb{R})$. Si*

$$A = \begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix}$$

con A_{11} (y A_{22}) cuadrada de orden $r < n$, entonces $\mathcal{G}_A$ no e fuertemente conexo.

Demostración. Basta observar que no se puede conectar el vértice $r+1$ con el vértice r, ya que cualquier camino dirigido que comience en $r+1$ sólo conecta vértices mayores o iguales que $r+1$ y cualquier camino dirigido que finalice en r sólo conecta vértices menores o iguales que r. De modo que para que existiese un camino dirigido de $r+1$ a r tendría que haber una flecha (i, j) con $i \geq r+1$ y $j \leq r$, lo que no es posible pues $a_{ij} = 0$ si $i \geq r+1$ y $j \leq r$, por hipótesis. ∎

Lema IV.3.7. *Sean $A \in \mathcal{M}_n(\mathbb{R})$ y $P \in \mathcal{M}_n(\mathbb{R})$ una matriz de permutación. El grafo $\mathcal{G}_A$ es fuertemente conexo si, y sólo si, el grafo $\mathcal{G}_{P^{\mathrm{t}}AP}$ es fuertemente conexo.*

Demostración. Basta observar que el grafo dirigido asociado a $P^{\mathrm{t}}AP$ se obtiene del de A mediante una reordenación de sus vértices, y esto no afecta al carácter fuertemente conexo. ∎

Teorema IV.3.8. *Sea $A \in \mathcal{M}_n(\mathbb{R})$. Si $\mathcal{G}_A$ es fuertemente conexo, entonces A es irreducible.*

Demostración. Si A es reducible, el grafo $G_{P^{\mathrm{t}}AP}$ no es fuertemente conexo para alguna matriz de permutación $P \in \mathcal{M}_n(\mathbb{R})$, lo cual es equivalente a que $\mathcal{G}_A$ tampoco lo sea. ∎

Teorema de Perron-Fröbenius.

A continuación vamos a demostrar que *toda matriz cuadrada no negativa e irreducible posee un autovalor real de multiplicidad* 1 *y módulo máximo.*

Sean $A = (a_{ij}) \in \mathcal{M}_n(\mathbb{R})$ una matriz no negativa e irreducible y $\varrho : \mathcal{L} \subset \mathbb{R}^n \to \mathbb{R}$, con $\mathcal{L} = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{x} \geq 0 \text{ con } \mathbf{x} \neq 0\}$, la función definida por

$$\varrho(\mathbf{x}) = \min_{\mathbf{x} \in \mathcal{L}} \left\{ \frac{\sum_{j=1}^{n} a_{ij} x_j}{x_i} \mid x_i \neq 0,\ i = 1, \ldots, n \right\}.$$

Lema IV.3.9. *Con la notación anterior, para todo* $\mathbf{x} \in \mathcal{L}$ *se cumple que*

i) $\varrho(\mathbf{x}) \geq 0$.
ii) $\varrho(\mathbf{x})x_i \leq \sum_{j=1}^{n} a_{ij}x_j$, *para todo* $i = 1, \ldots, n$.
iii) $A\mathbf{x} - \varrho(\mathbf{x})\mathbf{x} \geq 0$, *además* $\varrho(\mathbf{x})$ *es el mayor número con esta propiedad.*
iv) *Si* $\mathbf{x} = (1, 1, \ldots, 1)^{\mathrm{t}} \in \mathbb{R}^n$, *entonces* $\varrho(\mathbf{x}) = \min\{\sum_{j=1}^{n} a_{ij} \mid i = 1, \ldots, n\}$.

Demostración. La demostración es un sencillo ejercicio que se propone al lector. ∎

Veamos que ϱ alcanza su valor máximo en el interior de $\mathcal{L}$.

Lema IV.3.10. *Con la notación anterior, existe* $\mathbf{v} > 0$ *tal que* $\varrho(\mathbf{v}) = \max\{\varrho(\mathbf{x}) \mid \mathbf{x} \in \mathcal{L}\}$.

Demostración.[3] En primer lugar, observamos que $\varrho(\alpha\mathbf{x}) = \varrho(\mathbf{x})$, para todo $\mathbf{x} \in \mathcal{L}$ y $\alpha > 0$; por tanto, a la hora de calcular el supremo de $\{\varrho(\mathbf{x}) \mid \mathbf{x} \in \mathcal{L}\}$ podemos restringirnos al conjunto $\mathcal{M} = \{\mathbf{x} = (x_1, \ldots, x_n) \in \mathcal{L} \mid x_1^2 + \ldots + x_n^2 = 1\}$ que es un subconjunto cerrado y acotado de $\mathbb{R}^n$. De tal forma que si ϱ fuese continua en $\mathcal{M}$ entonces se alcanzaría el supremo; sin embargo, puede ocurrir que ϱ no sea continua en $\mathcal{M}$.

Consideremos entonces $\mathcal{N} = \{(I_n + A)^{n-1}\mathbf{x} \mid \mathbf{x} \in \mathcal{M}\}$. Por la proposición IV.3.3, todo elemento de $\mathcal{N}$ es un vector positivo, por lo que $\mathcal{N} \subset \mathcal{L}$. Además, $\mathcal{N}$ es una imagen continua de $\mathcal{M}$, por lo que es cerrado y acotado, y ϱ es continua en $\mathcal{N}$ porque no hay denominadores nulos. Por consiguiente, ϱ alcanza un máximo en $\mathcal{N}$ (véase el teorema A.4.9); y como $\mathcal{N} \subset \mathcal{L}$, se tiene que

$$\max\{\varrho(\mathbf{x}) \mid \mathbf{x} \in \mathcal{N}\} \leq \sup\{\varrho(\mathbf{x}) \mid \mathbf{x} \in \mathcal{L}\}.$$

Dado $\mathbf{x} \in \mathcal{M}$, sea $\mathbf{y} \in \mathcal{N}$ tal que $\mathbf{y} = (I_n + A)^{n-1}\mathbf{x}$; veamos que $\varrho(\mathbf{x}) \leq \varrho(\mathbf{y})$. Como $A\mathbf{x} - \varrho(\mathbf{x})\mathbf{x} \geq 0$ (véase el apartado iii) del lema IV.3.9), se tiene que

$$0 \leq (I_n + A)^{n-1}(A\mathbf{x} - \varrho(\mathbf{x})\mathbf{x}) = A(I_n + A)^{n-1}\mathbf{x} - \varrho(\mathbf{x})(I_n + A)^{n-1}\mathbf{x} = A\mathbf{y} - \varrho(\mathbf{x})\mathbf{y};$$

pues A y $(I_n + A)^{n-1}$ conmutan.

Teniendo ahora en cuenta que $\varrho(\mathbf{y})$ es el mayor número tal que $A\mathbf{y} - \varrho(\mathbf{y})\mathbf{y} \geq 0$, obtenemos que $\varrho(\mathbf{x}) \leq \varrho(\mathbf{y})$; luego,

$$\sup\{\varrho(\mathbf{x}) \mid \mathbf{x} \in \mathcal{L}\} = \sup\{\varrho(\mathbf{x}) \mid \mathbf{x} \in \mathcal{M}\} \leq \max\{\varrho(\mathbf{y}) \mid \mathbf{y} \in \mathcal{N}\}.$$

En conclusión

$$\sup\{\varrho(\mathbf{x}) \mid \mathbf{x} \in \mathcal{L}\} = \max\{\varrho(\mathbf{x}) \mid \mathbf{x} \in \mathcal{N}\}$$

y existe $\mathbf{y} > 0$ tal que $\rho(\mathbf{y}) = \sup\{\varrho(\mathbf{x}) \mid \mathbf{x} \in \mathcal{L}\}$. ∎

[3] La demostración hace uso de algunos resultados básicos sobre funciones en el espacio euclídeo $\mathbb{R}^n$, véase, por ejemplo, el capítulo 1 de [**Spi88**].

Puede existir más de un vector positivo en $\mathcal{L}$ donde la función ϱ alcance su valor máximo; tales vectores se denominan **vectores extremales** de A.

Lema IV.3.11. *Sean $A \in \mathcal{M}_n(\mathbb{R})$ irreducible y no negativa, $\mathbf{v} \in \mathbb{R}^n$ un vector extremal de A y $\rho = \varrho(\mathbf{v}) \in \mathbb{R}_{\geq 0}$.*

(a) *Si $A\mathbf{u} - \rho\mathbf{u} \geq 0$, para algún $\mathbf{u} \geq 0$ no nulo, entonces $A\mathbf{u} = \rho\mathbf{u}$.*

(b) *Cualquier autovector de A asociado a ρ tiene todas sus entradas no nulas.*

Demostración. (a) Sea $\mathbf{u} \geq 0$ no nulo tal que $A\mathbf{u} - \rho\mathbf{u} \geq 0$. Si $A\mathbf{u} - \rho\mathbf{u} \neq 0$, entonces, por la proposición IV.3.3,

$$(I + A)^{n-1}(A\mathbf{u} - \rho\mathbf{u}) > 0.$$

Luego, si $\mathbf{w} = (I + A)^{n-1}\mathbf{u}$, entonces $A\mathbf{w} - \rho\mathbf{w} > 0$, es decir,

$$\rho < \frac{\sum_{j=1}^{n} a_{ij}w_j}{w_i}, \text{ para todo } i = 1, \ldots, n.$$

De donde se sigue que $\rho < \varrho(\mathbf{w})$, lo que supone una clara contradicción con el hecho de que $\mathbf{v}$ sea extremal. Por consiguiente, $A\mathbf{u} - \rho\mathbf{u} = 0$, esto es, ρ es un autovalor de A y $\mathbf{u}$ un autovector de A asociado a ρ.

(b) Sea $\mathbf{u}$ un autovector de A asociado a ρ. Entonces $A\mathbf{u} = \rho\mathbf{u}$ y $\mathbf{u} \neq 0$, por lo que

$$\rho|\mathbf{u}| = |\rho\mathbf{u}| = |A\mathbf{u}| \leq^4 A|\mathbf{u}|,$$

donde $|A\mathbf{u}|$ y $|\mathbf{u}|$ son los vectores de $\mathbb{R}^n$ cuyas entradas son los valores absolutos de las entradas de $A\mathbf{u}$ y $\mathbf{u}$, respectivamente. Luego, $A|\mathbf{u}| - \rho|\mathbf{u}| \geq 0$; de donde se sigue, usando el apartado anterior, que $|\mathbf{u}|$ es un autovector de A asociado a ρ. Por otra parte, por la proposición IV.3.3, tenemos $\mathbf{w} = (I_n + A)^{n-1}|\mathbf{u}| > 0$, de modo que

$$0 < \mathbf{w} = (I_n + A)^{n-1}|\mathbf{u}| = (1 + \rho)^{n-1}|\mathbf{u}|,$$

por ser $|\mathbf{u}|$ un autovector de A asociado a ρ. De donde se deduce que $|\mathbf{u}| > 0$ y, por lo tanto, que $\mathbf{u}$ no tiene ninguna de sus entradas nula. ∎

Teorema de Perron-Fröbenius. *Sea $A \in \mathcal{M}_n(\mathbb{R})$ irreducible y no negativa. Entonces*

(a) *A tiene, al menos, un autovalor ρ real y positivo con un autovector asociado $\mathbf{v} > 0$.*

(b) *el autovalor ρ tiene multiplicidad 1.*

(c) *$|\lambda| \leq \rho$, para todo autovalor λ (real o complejo) de A, es decir, ρ es el radio espectral[5] de A.*

[4]Recuérdese que $|z_1 + z_2| \leq |z_1| + |z_2|$, para todo $z_1, z_2 \in \mathbb{C}$.

[5]Recuérdese que el **radio espectral** de un matriz es el mayor de los módulos de sus autovalores reales y complejos.

Demostración. Sean $\mathbf{v} \in \mathbb{R}^n$ un vector extremal y $\rho = \varrho(\mathbf{v}) \in \mathbb{R}_{\geq 0}$.

(a) Por el apartado iii) de lema IV.3.9, $A\mathbf{v} - \rho\mathbf{v} \geq 0$, luego del lema IV.3.11(a) se sigue que $\rho \in \mathbb{R}_{\geq 0}$ es un autovalor de A y $\mathbf{v} > 0$ es un autovector de A asociado a ρ.

(b) Supongamos que existen dos autovectores linealmente independientes de A, $\mathbf{u} = (u_1, \ldots, u_n)$ y $\mathbf{w} = (w_1, \ldots, w_n)$, asociados a ρ; según el lema IV.3.11(b) ningún autovector de A asociado a ρ tiene componentes nulas, por lo que cualquier combinación lineal de $\mathbf{u}$ y $\mathbf{w}$ no las tendrá. Sin embargo,

$$w_1\mathbf{u} - u_1\mathbf{w} = (0, w_1u_2 - u_1w_2, \ldots, w_1u_n - u_1w_n)$$

lo que supone una contradicción. Por consiguiente, no existen dos autovectores linealmente independientes de A asociados a ρ, es decir, el subespacio propio $L_1 = \ker(\rho I_n - A)$ asociado a ρ tiene dimensión 1. Luego, L_1 está generado por el vector extremal $\mathbf{v}$.

Veamos ahora que $L_1 = L_2 = \ker\big((\rho I_n - A)^2\big)$. La inclusión $L_1 \subseteq L_2$ se da siempre, por lo que basta demostrar la inclusión $L_1 \supseteq L_2$. Si $\mathbf{u} \in L_2$, es claro que $(\rho I_n - A)\mathbf{u} \in L_1$ por lo que existe $\alpha \in \mathbb{R}$ tal que $(\rho I_n - A)\mathbf{u} = \alpha\mathbf{v}$, si α es cero, entonces $\mathbf{u} \in L_1$. Supongamos, pues, que $\alpha \neq 0$ y consideremos un autovector $\mathbf{w}$ de A^{t} asociado a ρ, que, por los argumentos anteriores, podemos tomar positivo; de tal modo que, como $\mathbf{w}^{\mathrm{t}}(\rho I_n - A) = \mathbf{0}$, se tiene que

$$\mathbf{0} = \mathbf{w}^{\mathrm{t}}(\rho I_n - A)\mathbf{u} = \mathbf{w}^{\mathrm{t}}(\alpha\mathbf{v}) = \alpha\mathbf{w}^{\mathrm{t}}\mathbf{v},$$

lo que contradice el carácter positivo de los vectores.

De todo esto se deduce que la multiplicidad de ρ es igual a 1.

(c) Sea λ un autovalor de A. Entonces para algún $\mathbf{u} \neq 0$ (que puede tener coordenadas complejas) se tiene que

$$\sum_j a_{ij}u_j = \lambda u_i,$$

de donde se sigue que

$$|\lambda u_i| = \big|\sum_j a_{ij}u_j\big| \leq \sum_j a_{ij}|u_j|.$$

Luego,

$$|\lambda| \leq \frac{\sum_j a_{ij}|u_i|}{|u_i|},$$

para todo u_i no nulo. De modo que si $|\mathbf{u}|$ es el vector de $\mathbb{R}^n$ cuyas entradas son los módulos de las entradas de $\mathbf{u}$, concluimos que

$$|\lambda| \leq \varrho(|\mathbf{u}|) \leq \rho,$$

por la maximalidad de ρ. ∎

Definición IV.3.12. Sea $A \in \mathcal{M}_n(\mathbb{R})$ no negativa e irreducible. El autovalor ρ cuya existencia demuestra el Teorema de Perron-Fröbenius se llama **autovalor de Perron** de A, el autovector $\mathbf{v} > 0$ de A asociado a ρ cuyas entradas suman 1 se llama **autovector de Perron**.

Corolario IV.3.13. *Sean $A \in \mathcal{M}_n(\mathbb{R})$ no negativa e irreducible y ρ su autovalor de Perron. Si A tiene una fila de entradas no nulas, entonces $|\lambda| < \rho$, para todo autovalor λ de A distinto de ρ.*

Demostración. Supongamos que todas las entradas de la primera fila de A son no nulas. Sea λ un autovalor de A tal que $|\lambda| = \rho$ y $\mathbf{u}$ un autovector de A asociado (que puede tener coordenadas complejas). Entonces,

$$\rho|\mathbf{u}| = |\lambda\mathbf{u}| = |A\mathbf{u}| \leq A|\mathbf{u}|,$$

donde $|A\mathbf{u}|$ y $|\mathbf{u}|$ son los vectores de $\mathbb{R}^n$ cuyas entradas son los valores absolutos de los entradas de $A\mathbf{u}$ y $\mathbf{u}$, respectivamente. Como $A|\mathbf{u}| - \rho|\mathbf{u}| \geq 0$, por el lema IV.3.11, $|\mathbf{u}|$ es un autovector de A asociado a ρ. Por consiguiente,

$$|A\mathbf{u}| = |\lambda||\mathbf{u}| = \rho|\mathbf{u}| = A|\mathbf{u}|.$$

Si nos fijamos en la primera fila nos queda que

$$\Big|\sum_{j=1}^{n} a_{1j}u_j\Big| = \sum_{j=1}^{n} a_{1j}|u_j|,$$

y como $a_{1j} \neq 0,\ j = 1, \ldots, n$, se sigue que todas las entradas de $\mathbf{u}$ son reales[6] y simultáneamente no positivos o no negativos[7] es decir, $\mathbf{u}$ es un múltiplo de un vector no negativo $\mathbf{w}$. Entonces $\mathbf{u} = \alpha\mathbf{w}$, con $\mathbf{w} \geq 0$. Por tanto, $|\mathbf{u}| = |\alpha|\mathbf{w}$, luego $\mathbf{w}$ es un autovector de A asociado a ρ, y concluimos que $\mathbf{u}$ también lo es y que $\lambda = \rho$. ∎

Matrices primitivas.

Definición IV.3.14. Se dice que una matriz $A \in \mathcal{M}_n(\mathbb{R})$ no negativa es **primitiva** si existe $m > 0$ tal que $A^m > 0$.

Nota IV.3.15. *Toda matriz primitiva es irreducible.* En efecto, sea A una matriz primitiva y supongamos que existe una matriz de permutación $P \in \mathcal{M}_n(\mathbb{R})$ tal que

$$PAP^{\mathrm{t}} = \begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix},$$

[6]Basta tener en cuenta que $|z_1 + z_2| = |z_1| + |z_2|$ si, y sólo si, z_1 y z_2 son números reales positivos o negativos simultáneamente.

[7]Nótese que si $x \in \mathbb{R}$ es positivo e $y \in \mathbb{R}$ negativo, entonces $|x + y| < \text{máx}(|x|, |y|) < |x| + |y|$,

con A_{11} y A_{22} matrices cuadradas de orden menor que n. Entonces

$$A^m = P^{\mathrm{t}} \begin{pmatrix} A_{11}^m & A_{12}' \\ 0 & A_{22}^m \end{pmatrix} P,$$

para todo $m > 1$, lo que es del todo imposible, pues A es primitiva y existe $m > 0$ tal que $A^m > 0$.

Sin embargo, no toda matriz irreducible es primitiva, considérese por ejemplo

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

Teorema IV.3.16. *Sea $A \in \mathcal{M}_n(\mathbb{R})$ primitiva. Existe un único autovalor real positivo ρ de A de multiplicidad 1 con un autovector asociado $\mathbf{v} > 0$ tal que*

$$|\lambda| < \rho,$$

para todo autovalor λ de A distinto de ρ.

Demostración. Como A es primitiva, es no negativa e irreducible; luego, por el Teorema de Perron-Fröbenius existe autovalor real ρ de A de multiplicidad 1 con un autovector asociado $\mathbf{v} > 0$ tal que

$$|\lambda| \leq \rho,$$

para todo autovalor λ de A. Por otra parte, existe $m > 0$ tal que $A^m > 0$. La matriz A^m es obviamente primitiva, por lo que es no negativa e irreducible, y además tiene todas sus filas de entradas no nulas. Por consiguiente, del corolario IV.3.13 se sigue que el autovalor de Perron ρ' de A^m verifica que

$$|\lambda'| < \rho'$$

para todo autovalor λ' de A^m distinto de ρ'.

Teniendo ahora en cuenta que los autovalores de A^m son las potencias m-ésimas de los autovalores de A, de las desigualdades anteriores se deduce que $\rho' = \rho^m$, y por lo tanto que en la desigualdad $|\lambda| \leq \rho$, para todo autovalor λ de A, sólo se da la igualdad cuando $\lambda = \rho$. ∎

Modelo matricial de Leslie.

Dividamos la población de hembras de una misma especie en distintos grupos de edad $G_1, G_2, \ldots, G_n$, donde cada grupo tiene la misma amplitud. Así, si la vida más larga se estima en L años, la amplitud de cada grupo de edades es de L/n años. El grupo G_1 está formado por los individuos cuya edad está en el intervalo $[0, L/n)$ es decir, que tienen menos de L/n años. El siguiente grupo por edades G_1, lo forman los individuos cuya edad está en el intervalo $[L/n, 2L/n)$. El siguiente grupo lo forman

los individuos con edad en $[2L/n, 3L/n)$, y así, hasta llegar al último grupo formado por los individuos cuya edad está comprendida en el intervalo $[(n-1)L/n, L]$.

Supongamos que los censos de población se realizan en intervalos de tiempo iguales a la amplitud de los grupos de edades, y consideremos las tasas de fecundidad y supervivencia: denotamos por f_i el número promedio de hijas de cada hembra del grupo G_i (esto es la tasa de fecundidad específica del grupo G_i). Llamamos s_i a la fracción de individuos del grupo G_i que sobreviven al intervalo entre censos y pasan a formar parte del grupo G_{i+1}.

Si $p_i(m)$ es el número de hembras de G_i en el instante m, entonces se sigue que

$$\text{(IV.3.8)} \qquad \begin{array}{rcl} p_1(m+1) & = & p_1(m)f_1 + p_2(m)f_1 + \ldots + p_n(m)f_n \\ p_i(m+1) & = & p_{i-1}(m)s_{i-1}; \quad \text{para } i = 2, \ldots, n. \end{array}$$

Además,

$$P_i(m) = \frac{p_i(m)}{p_0(m) + p_1(m) + \ldots + p_n(mj)}$$

es la proporción de población en G_i en el instante m.

El vector $\mathbf{P}(m) = (P_1(m), P_2(m), \ldots, P_n(m))^{\mathrm{t}}$ representa a la *distribución de edades de la población* en el instante m, y, suponiendo que existe, $\mathbf{P}^* = \lim_{m\to\infty} \mathbf{P}(m)$ es la *distribución de edades de la población a largo plazo.*

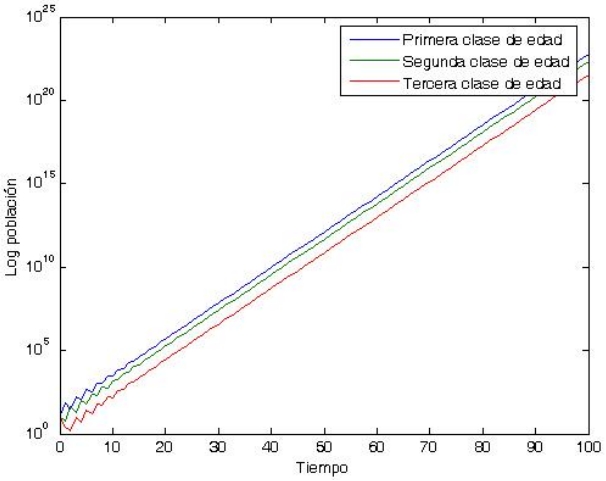


FIGURA 1. Distribución de edades de una población divida en tres grupos edad a lo largo del tiempo.

Las ecuaciones (IV.3.8) constituyen un sistema de ecuaciones lineales en diferencias homogéneo que se puede escribir en forma matricial como

$$\text{(IV.3.9)}\quad \mathbf{p}(m) = A\,\mathbf{p}(m-1), \quad \text{donde} \quad A = \begin{pmatrix} f_1 & f_2 & \dots & f_{n-1} & f_n \\ s_1 & 0 & \dots & 0 & 0 \\ 0 & s_2 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & s_{n-1} & 0 \end{pmatrix} \in \mathcal{M}_n(\mathbb{R})$$

y $\mathbf{p}(m) = (p_1(m), \dots, p_n(m))^{\mathrm{t}}$, para todo $m \geq 0$. De modo que $\mathbf{p}(m) = A^m\mathbf{p}(0)$ para todo $m > 0$.

La matriz A se llama **Matriz de Leslie** en honor de P.H. Leslie que introdujo este modelo en 1945.

La matriz A es una matriz no negativa, pues $s_i > 0,\ i = 1, \dots, n-1$ y $f_i \geq 0,\ i = 1, \dots, n$. Además, si $n > 2$ y f_{n-1}, f_n son positivos, entonces A es primitiva (ejercicio 14), en cuyo caso existirá $\mathbf{P}^*$ y podremos determinar su valor.

Supongamos, pues, que f_{n-1}, f_n son positivos; de este modo, el teorema IV.3.16 garantiza la existencia de un autovalor real positivo ρ de A de multiplicidad 1 con un autovector asociado $\mathbf{v} > 0$ tal que

$$|\lambda| < \rho,$$

para todo autovalor λ de A distinto de ρ. De tal forma que el límite de A^m/ρ^m cuando j tiende a infinito es una matriz no nula cuyas columnas son proporcionales a $\mathbf{v}$ es decir,

$$\lim_{m\to\infty} \frac{A^m}{\rho^m} = \mathbf{v}\mathbf{w}^{\mathrm{t}},$$

para algún $\mathbf{w} \in \mathbb{R}^n$. Por otra parte, tenemos que

$$\begin{aligned} \mathbf{P}^* &= \lim_{m\to\infty} \mathbf{P}(m) = \lim_{m\to\infty} \frac{\mathbf{p}(m)}{(1,1,\dots,1)\mathbf{p}(m)} = \lim_{m\to\infty} \frac{A^m\mathbf{p}(0)}{(1,1,\dots,1)A^m\mathbf{p}(0)} \\ &= \lim_{m\to\infty} \frac{(A^m\mathbf{p}(0))/\rho^m}{(1,1,\dots,1)(A^m\mathbf{p}(0))/\rho^m} = \frac{\lim_{m\to\infty}(A^m/\rho^m)\mathbf{p}(0)}{\lim_{m\to\infty}(1,1,\dots,1)(A^m/\rho^m)\mathbf{p}(0)} \\ &= \frac{(\mathbf{v}\mathbf{w}^{\mathrm{t}})\mathbf{p}(0)}{(1,1,\dots,1)(\mathbf{v}\mathbf{w}^{\mathrm{t}})\mathbf{p}(0)} = \frac{\mathbf{v}(\mathbf{w}^{\mathrm{t}}\mathbf{p}(0))}{(1,1,\dots,1)\mathbf{v}(\mathbf{w}^{\mathrm{t}}\mathbf{p}(0))} \\ &= \frac{\mathbf{v}}{v_1 + \dots + v_n}. \end{aligned}$$

En resumen, $\mathbf{P}^*$ es el autovector de Perron de A, es decir, el autovector de Perron es la distribución de edades de la población a largo plazo.

Ejemplo IV.3.17. Las hembras de cierta especie animal viven tres años. Supongamos que la tasa de supervivencia de hembras en sus primero y segundo años es del 60 % y 25 %, respectivamente. Cada hembra del segundo grupo de edad tiene 4 hijas al año de media, y cada hembra del tercer grupo tiene una media de 3 hijas por año.

La figura 1 muestra la distribución de los tres grupos edades a lo largo tiempo en escala semilogarítmica. Observamos que si bien la población de hembras crece indefinidamente, cuando el tiempo es suficientemente alto, la proporción de hembras de cada grupo de edad se mantiene estable, según el autovector de Perron de la correspondiente matriz de Leslie. En la práctica 6 estudiaremos este y otros ejemplos con más detalle.

4. Cadenas de Markov homogéneas y finitas

Definición IV.4.1. Sea $P = (p_{ij}) \in \mathcal{M}_n(\mathbb{R})$ tal que $p_{ij} \in [0,1],\ i,j = 1,\dots,n$. Se dice que P es una **matriz estocástica** cuando sus columnas o filas suman 1. Diremos que es **doblemente estocástica** cuando sus columnas y filas suman 1.

Nos centraremos en el caso en que las columnas suman 1. No es raro encontrar textos donde esta condición se supone sobre las filas, pero los resultados son semejantes.

Definición IV.4.2. Un **vector** no negativo $\mathbf{p} = (p_1,\dots,p_n)^{\mathrm{t}} \in \mathbb{R}^m$ se dice que es **de probabilidad** si $\|\mathbf{p}\|_1 := \sum_{i=1}^n p_i = 1$.

De esta forma una matriz estocástica tiene como columnas a vectores de probabilidad. Nótese que las matrices estocásticas son no negativas.

Supongamos que estamos observando algún fenómeno aleatorio a lo largo del tiempo, y que en cualquier punto concreto del tiempo nuestra observación puede tomar uno de los n valores, a veces llamados estados, $1,\dots,n$. En otras palabras, tenemos una sucesión de variables aleatorias X_m, para periodos de tiempo $m = 0,1,\dots$, donde cada variable puede ser igual a de los números, $1,\dots,n$. Si la probabilidad de que X_m se encuentre en el estado i sólo depende del estado en que se hallase X_{m-1} y no en los estados de periodos anteriores de tiempo, entonces el proceso se dice que es una **cadena de Markov**. Si la probabilidad tampoco depende del valor de m, entonces la cadenas de Markov se dice que es **homogénea**, y si el número de estados es finito, como es nuestro caso, la cadena de Markov se dice **finita**.

En el caso de las cadenas de Markov homogéneas y finitas, la probabilidades de cualquier periodo de tiempo se pueden calcular a partir de la probabilidades iniciales

de los estados y lo que se conoce como probabilidades de transición. Denotaremos

$$\mathbf{p}_0 = \begin{pmatrix} p_1^{(0)} \\ \vdots \\ p_n^{(0)} \end{pmatrix}$$

al vector de probabilidades iniciales, donde $p_i^{(0)}$ es la probabilidad de que el proceso comience en el estado i. La **matriz de transición de probabilidades** es la matriz $P = \mathcal{M}_n(\mathbb{R})$ cuya entrada (i,j)-ésima, p_{ij}, da la probabilidad de que X_m se halle en el estado i supuesto que X_{m-1} se hallaba en el estado j. Por consiguiente, si

$$\mathbf{p}_m = \begin{pmatrix} p_1^{(m)} \\ \vdots \\ p_n^{(m)} \end{pmatrix}$$

siendo $p_i^{(m)}$ la probabilidad de que el sistema se encuentre en el estado i en el instante m, entonces, por el teorema de la probabilidad total se tiene que

$$\begin{aligned} \mathbf{p}_1 &= P\,\mathbf{p}_0, \\ \mathbf{p}_2 &= P\,\mathbf{p}_1 = P\,P\,\mathbf{p}_0 = P^2\mathbf{p}_0, \end{aligned}$$

y en general,

$$\mathbf{p}_m = P^m\mathbf{p}_0.$$

Nótese que P es una matriz estocástica pues su columna j-ésima nos indica la probabilidad de los posibles estados en un determinado instante cuando en el instante inmediatamente anterior el estado sea j.

Si tenemos una población considerable de individuos sujetos a este proceso aleatorio, entonces $p_i^{(m)}$ se puede describir como la proporción de individuos en el estado i al instante m, mientras que $p_i^{(0)}$ sería la proporción de individuos que comienzan en el estado i. De modo natural nos podemos hacer las siguientes preguntas ¿qué ocurre con estas proporciones cuando m aumenta? Es decir, ¿podemos determinar el comportamiento límite de $\mathbf{p}_m$? Nótese que la respuesta depende del comportamiento asintótico de P^m, y que P es una matriz no negativa ya que cada una de sus entradas es una probabilidad. Por consiguiente, si P es primitiva, podemos garantizar que existe un único autovalor real ρ dominante. Se comprueba fácilmente que $\rho = 1$; en efecto, basta tener en cuenta que los autovalores de P son los mismos que los de su traspuesta P^{t} y que

$$|\lambda| \leq \frac{|\sum_{i=1}^n p_{ij}x_i|}{|x_j|} \leq \frac{\sum_{i=1}^n |p_{ij}|\ |x_i|}{|x_j|} \leq \sum_{i=1}^n p_{ij} = 1,$$

siendo λ un autovalor (real o complejo) de P, $(x_1, \ldots, x_n)^{\mathrm{t}}$ un autovector de P^{t} asociado a λ y $x_j = \max\{x_i \mid i = 1, \ldots, n\}$. En consecuencia, si P es primitiva existe un único un autovector $\mathbf{p} > 0$ asociado al autovalor $\rho = 1$ tal que $\sum_{i=1} p_i = 1$. Entonces,

$$\lim_{m\to\infty} (\rho^{-1}P)^m = \lim_{m\to\infty} P^m = \mathbf{p}\mathbf{1}_n^{\mathrm{t}},$$

donde $\mathbf{1}_n^{\mathrm{t}} = (1, \ldots, 1) \in \mathcal{M}_{1\times n}(\mathbb{R})$. Usando la igualdad anterior, obtenemos que

$$\lim_{m\to\infty} \mathbf{p}_m = \lim_{t\to\infty} P^m\mathbf{p}_0 = \mathbf{p}\mathbf{1}_n^{\mathrm{t}}\mathbf{p}_0 = \mathbf{p},$$

donde el último paso se sigue de que $\mathbf{1}_n^{\mathrm{t}}\mathbf{p}_0 = 1$. Por tanto, el sistema se aproxima a un punto de equilibrio en que las proporciones de los distintos estados vienen dadas por las entradas de $\mathbf{p}$. Además, el comportamiento límite no depende de las proporciones iniciales.

Ejercicios del tema IV

Ejercicio 1. Comprobar que si $\{\lambda_1, \ldots, \lambda_r\}$ son los autovalores de una matriz A, entonces los autovalores de A^m son $\{(\lambda_1)^m, \ldots, (\lambda_r)^m\}$. Si $\mathbf{v}$ es un autovector de A asociado a λ_i, entonces $\mathbf{v}$ es autovector de A^m asociado a $(\lambda_i)^m$. Poner un ejemplo que muestre que el recíproco no es cierto.

Ejercicio 2. Sea

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

Los autovalores de esta matriz son $\lambda_1 = 1 + \sqrt{3}$, $\lambda_2 = 1 - \sqrt{3}$ y $\lambda_3 = 0$. El autovalor de mayor módulo es λ_1. Asociado a este autovalor tenemos el autovector $\mathbf{v} = (\sqrt{3} - 1, 1, 1)$ de componentes estrictamente positivas.

Para un vector cualquiera $\mathbf{b}$, comprobar que el producto $B^m\mathbf{b}$ se aproxima, para valores grandes de m a $c\lambda_1^m\mathbf{v}_1$, donde c es una cierta constante y $\mathbf{v_1}$ es un autovector asociado a λ_1.

Ejercicio 3. Sean V un $\Bbbk$-espacio vectorial de dimensión $n > 0$ y $T \in \mathrm{End}_{\Bbbk}(V)$ diagonalizable. Dado $r \in \mathbb{Z}_+$, diremos que $S \in \mathrm{End}_{\Bbbk}(V)$ es una **raíz r-ésima de** T si $S^r = T$. Encontrar condiciones necesarias y suficientes para que existan raíces r-ésimas de T.

Sean $V = \mathbb{R}^3$ y $T \in \mathrm{End}_{\Bbbk}(V)$ tal que su matriz asociada respecto de la base usual de $\mathbb{R}^3$ es

$$\begin{pmatrix} 8 & -6 & 4 \\ -6 & 9 & -2 \\ 4 & -2 & 4 \end{pmatrix}.$$

Hallar, si es posible, la matriz asociada a la raíz cuadrada de T respecto de la base usual de $\mathbb{R}^3$.

Ejercicio 4. Sean $V = \mathbb{R}^3$ y $T \in \mathrm{End}_{\Bbbk}(V)$ tal que su matriz asociada respecto de la base usual de $\mathbb{R}^3$ es

$$(a)\ \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \qquad (b)\ \begin{pmatrix} 5 & 0 & 0 \\ 0 & -1 & 1 \\ 3 & 0 & 2 \end{pmatrix}.$$

Hallar la matriz asociada T^m respecto de la base usual de $\mathbb{R}^3$

Ejercicio 5. Resolver la ecuación en diferencias $x_{n+2} - 3x_{n+1} + 2x_n = 0$ dados $x_1 = 1,\ x_2 = 0$ y $x_3 = 1$.

Ejercicio 6. Dado el sistema de ecuaciones en diferencias $\mathbf{u}_n = A\mathbf{u}_{n-1}$, siendo

$$A = \begin{pmatrix} 0 & a^2 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & a^2 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

1. Obtener la expresión general de $\mathbf{u}_n$.
2. Calcular $\mathbf{u}_{10}$, dado el vector inicial $\mathbf{u}_0 = (0, 2, 0, 2)$.

Ejercicio 7. Sean $A \in \mathcal{M}_n(\mathbb{R})$ y $\varepsilon > 0$. Probar que si A es no negativa e irreducible, entonces $(\varepsilon I_n + A)^{n-1} > 0$.

Ejercicio 8. Sea $A = (a_{ij}) \in \mathcal{M}_n(\mathbb{R})$ una matriz no negativa e irreducible. Si $a_{ii} \neq 0$, para todo $i = 1, \ldots, n$, entonces A es primitiva. [Tómese $\varepsilon = \text{mín}\{a_{ii} \mid i = 1, \ldots, n\}$, compruébese que $B = A - \varepsilon I_n$ es no negativa e irreducible, y úsese el ejercicio 7 para concluir que $A = I_n + B$ es primitiva.

Ejercicio 9. Sea $A \in \mathcal{M}_n(\mathbb{R})$ una matriz positiva e irreducible. Probar que si la suma de las entradas de cualquier fila (o columna) es ρ, entonces el autovalor de Perron de A es ρ.

Ejercicio 10. Comprobar el teorema de Perron-Fröbenius calculando los autovalores y autovectores de la matriz

$$A = \begin{pmatrix} 7 & 2 & 3 \\ 1 & 8 & 3 \\ 1 & 2 & 9 \end{pmatrix}.$$

Encontrar el autovalor y el autovector de Perron de A.

Ejercicio 11. Calcular el autovalor y el autovector de Perron de la matriz

$$A = \begin{pmatrix} 1-\alpha & \beta \\ \alpha & 1-\beta \end{pmatrix},$$

donde $\alpha + \beta = 1$ con α y $\beta > 0$.

Ejercicio 12. Sea

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

1. Probar que A es irreducible.
2. Hallar el autovalor y el autovector de Perron de A.

Ejercicio 13. Demuestre que el polinomio característico de la matriz

$$A = \begin{pmatrix} f_1 & f_2 & f_3 \\ s_1 & 0 & 0 \\ 0 & s_2 & 0 \end{pmatrix}$$

es igual a

$$\aleph_A(x) = \det(xI - A) = x^3 - f_1x^2 - f_2s_1x - f_3s_1s_2.$$

Demuestre que el polinomio característico de la matriz

$$A = \begin{pmatrix} f_1 & f_2 & f_3 & f_4 \\ s_1 & 0 & 0 & 0 \\ 0 & s_2 & 0 & 0 \\ 0 & 0 & s_3 & 0 \end{pmatrix}$$

es igual a

$$\aleph_A(x) = \det(xI - A) = x^4 - f_1x^3 - f_2s_1x^2 - f_3s_1s_2x - f_4s_1s_2s_3.$$

Dada la matriz de Leslie

$$A = \begin{pmatrix} f_1 & f_2 & \dots & f_{n-1} & f_n \\ s_1 & 0 & \dots & 0 & 0 \\ 0 & s_2 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \dots & \dots \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & s_{n-1} & 0 \end{pmatrix}$$

intente deducir una fórmula para su polinomio característico.

Ejercicio 14. Sea $A \in \mathcal{M}_n(\mathbb{R})$ una matriz de Leslie tal que $f_{n-1} \cdot f_n \neq 0$. Probar que

1. A es irreducible.
2. Si $f_1 = \ldots = f_{n-2} = 0$, entonces

   $$A^n = s_1 \cdot s_{n-2}f_{n-1}A + s_1 \cdot s_{n-1}f_nI_n.$$

   Usando esta igualdad concluir que es no negativa e irreducible y, por el ejercicio 8, que es primitiva.
3. En general $A^n = s_1 \cdot s_{n-2}f_{n-1}A + s_1 \cdot s_{n-1}f_nI_n + B$ para cierta matriz B no negativa. Usando esta igualdad concluir que es no negativa e irreducible y, por el ejercicio 8, que es primitiva.

Ejercicio 15. Un estudio ha determinado que el sector de ocupación de un niño, cuando sea adulto, depende del sector en que trabaje su padre, y está dada por la

siguiente matriz de transición, con los sectores de producción P = sector primario, S = sector secundario, T = sector terciario.

$$\begin{array}{cc} & \text{Sector del padre} \\ & \begin{array}{ccc} \text{T} & \text{S} & \text{P} \end{array} \\ \text{Sector del hijo} \begin{array}{c} \text{T} \\ \text{S} \\ \text{P} \end{array} & \begin{pmatrix} 0{,}8 & 0{,}3 & 0{,}2 \\ 0{,}1 & 0{,}5 & 0{,}2 \\ 0{,}1 & 0{,}2 & 0{,}6 \end{pmatrix} \end{array}$$

Así, la probabilidad de que el hijo de alguien que trabaja en el sector terciario también lo haga en ese sector es 0,8.

1. ¿Cuál es la probabilidad de que el nieto de un trabajador del sector terciario trabaje en ese sector?
2. A largo plazo, ¿qué proporción de la población trabajará en el sector secundario?

Ejercicio 16. Para la matriz de transición

$$P = \begin{pmatrix} 0{,}4 & 0{,}5 \\ 0{,}6 & 0{,}5 \end{pmatrix},$$

1. calcular $\mathbf{x}^{(m)}$ para $n = 1, 2, 3, 4, 5$, si $\mathbf{x}^{(0)} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$;
2. probar que P es una matriz primitiva y calcular el vector de estado estacionario.

Ejercicio 17. Consideremos la matriz de transición

$$P = \begin{pmatrix} 0{,}5 & 0 \\ 0{,}5 & 1 \end{pmatrix}.$$

1. Probar que P no es primitiva.
2. Probar que cuando $m \to \infty$, $P^m\mathbf{x}^{(0)}$ se aproxima a $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$, para cualquier vector inicial $\mathbf{x}^{(0)}$.

Ejercicio 18. Verificar que si P es una matriz de transición primitiva de orden n, cuyas filas suman todas uno, entonces su vector de estado estacionario tiene todas sus componentes iguales a $1/n$.

Ejercicio 19. Probar que la matriz de transición

$$P = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}$$

es primitiva, y aplicar el ejercicio 17 para calcular su vector de estado estacionario.

Ejercicio 20. Consideremos la sucesión de matrices de transición $\{P_2, P_3, P_4, \ldots\}$, con

$$P_2 = \begin{pmatrix} 0 & \frac{1}{2} \\ 1 & \frac{1}{2} \end{pmatrix}, \qquad P_3 = \begin{pmatrix} 0 & 0 & \frac{1}{3} \\ 0 & \frac{1}{2} & \frac{1}{3} \\ 1 & \frac{1}{2} & \frac{1}{3} \end{pmatrix},$$

$$P_4 = \begin{pmatrix} 0 & 0 & 0 & \frac{1}{4} \\ 0 & 0 & \frac{1}{3} & \frac{1}{4} \\ 0 & \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ 1 & \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \end{pmatrix}, \quad P_5 = \begin{pmatrix} 0 & 0 & 0 & 0 & \frac{1}{5} \\ 0 & 0 & 0 & \frac{1}{4} & \frac{1}{5} \\ 0 & 0 & \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \\ 0 & \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \\ 1 & \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \end{pmatrix}.$$

y sucesivamente. Probar que estas matrices de transición son regulares, y determinar los vectores de estado estacionarios $\mathbf{x}_m$ tales que $P_m\mathbf{x}_m = \mathbf{x}_m$, para $m = 2, 3, \ldots, n$.

TEMA V

Matrices simétricas y formas cuadráticas

En este tema volvemos a ocuparnos de cuestiones teóricas relacionadas con las matrices más en la línea de un curso clásico de Álgebra Lineal. El planteamiento inicial es similar al de los de temas II y III. Tras introducir el concepto de forma bilineal y forma bilineal simétrica, se fija una base y se determina la matriz asociada a una forma bilineal. A continuación, se demuestra la fórmula del cambio de base para las matrices asociadas a una forma bilineal, y a la relación de equivalencia que determinada por esta forma se le da el nombre de congruencia de matrices. Sin embargo, a diferencia de los temas anteriores, en este tema la congruencia de matrices no juega el mismo papel de hilo conductor que desempeñaban la equivalencia y semejanza de matrices en los temas anteriores, ya que este papel lo asumen el producto escalar y la proyección, que son las verdaderas estrellas del tema, así como las matrices simétricas reales.

En la segunda sección se definen el producto escalar y los espacios vectoriales euclídeos. Se hace una especial mención al espacio vectorial $\mathbb{R}^n$ con la estructura euclídea determinada por el producto escalar usual, aunque se muestran otros ejemplos de espacios vectoriales euclídeos. A continuación, tratamos el concepto de norma en un espacio vectorial euclídeo. Estos conceptos se estudiarán con mayor profundidad en los temas VIII y XII.

Nuestra siguiente sección se dedica a la ortogonalidad, al método ortogonalización de Gram-Schmidt y, consecuentemente, a la existencia de bases ortonormales en un espacio vectorial euclídeo. Ya en la sección cuarta, se define qué entendemos por subespacio ortogonal y se enuncian y demuestran algunas de sus propiedades; entre otras, la descomposición de un espacio vectorial euclídeo como suma directa de un subespacio y su ortogonal, lo que nos permite definir la proyección ortogonal sobre un subespacio vectorial. El significado geométrico de la proyección ortogonal es fundamental en esta asignatura como se podrá ver en el siguiente tema. Por tanto, demostramos que la proyección ortogonal de un vector $\mathbf{v}$ sobre un subespacio vectorial L consiste en calcular el vector de L más próximo a $\mathbf{v}$. Asimismo, se describe la matriz de la aplicación proyección ortogonal sobre un subespacio L respecto de una base $\mathcal{B}$ del espacio vectorial euclídeo V en términos de la matriz del producto escalar de V y la matriz cuyas columnas son las coordenadas de una base de L respecto de $\mathcal{B}$.

La sección quinta está dedica a las matrices simétricas reales; en primer lugar se enuncia y demuestra que *toda matriz simétrica real diagonaliza a través de una matriz ortogonal*. En particular, toda matriz simétrica real es semejante y congruente con una matriz diagonal. Este resultado tiene interés en Estadística y Probabilidad, si tenemos en cuenta que las matrices de covarianza y correlación son simétricas y reales. La segunda parte de esta sección se centra en las matrices simétricas (semi)definidas positivas, mostrándose condiciones necesarias y suficientes para que una matriz simétrica sea (semi)definida positiva en términos de sus autovalores. Toda esta sección está plagada de resultados relacionados con las matrices simétricas y las matrices simétricas (semi)definidas positivas que serán utilizados posteriormente en la asignatura *Modelos Lineales*. Estos resultados son en su mayoría sencillos ejercicios tales como la existencia de raíces cuadradas de matrices simétricas semidefinidas positivas (que será usada en el próximo tema para definir la descomposición en valores singulares) o la factorización $A = QQ^{\mathrm{t}}$ de una matriz simétrica semidefinida positiva, pudiéndose elegir Q triangular superior. Al final de la sección trataremos algunas cuestiones relacionadas con matrices hermíticas.

La última sección del tema trata sobre las formas cuadráticas. Así, se define qué entenderemos por forma cuadrática y se demuestra la relación de éstas con las matrices simétricas. Lo que nos permite escribir cualquier forma cuadrática en la forma $\sum_{i=1}^{n} d_i i x_i^2$ mediante un cambio de base, siendo $d_i i,\ i = 1, \ldots, n$ los autovalores de la matriz simétrica asociada a la forma cuadrática. Al final de la sección y del tema se hace una breve mención a la relación entre las formas cuadráticas y las métricas simétricas.

La mayor parte de los contenidos teóricos de este tema tienen aplicación directa en otras asignaturas de la Licenciatura; por ejemplo, la proyección ortogonal es fundamental en las asignaturas *Modelos Lineales* y *Análisis Multivariante*. Téngase en cuenta que un modelo lineal normal consiste en considerar un subespacio vectorial propio L de $\mathbb{R}^m$ y un vector aleatorio $\mathbf{y} = \mu + \varepsilon$ con $\varepsilon \sim N_n(0, \sigma^2 I_n),\ \mu \in L$ y $\sigma^2 > 0$. De este modo, resulta natural tomar $\hat{\mu} = \pi_L(\mathbf{y})$ como estimador de μ, siendo π_L la proyección ortogonal de $\mathbf{y}$ sobre L, y $\hat{\sigma}^2 = \|\mathbf{y} - \pi_(\mathbf{y})\|^2$ como estimador de la varianza; y esto sólo es el principio de la historia.

En este tema, hemos seguido el capítulo 2 de [**Sch05**] y el capítulo 5 de [**MS06**], si bien hemos tenido en cuenta el capítulo 8 de [**BCR07**].

1. Formas bilineales

Mientras no se diga lo contrario, a lo largo de este tema V denotará a un espacio vectorial sobre $\mathbb{R}$ de dimensión finita $n > 0$.

Definición V.1.1. Diremos que una aplicación $T_2 : V \times V \longrightarrow \mathbb{R}$ es una **forma bilineal**, o **métrica**, sobre V si satisface

(a) $T_2(\mathbf{u}_1 + \mathbf{u}_2, \mathbf{v}) = T_2(\mathbf{u}_1, \mathbf{v}) + T_2(\mathbf{u}_2, \mathbf{v})$;
(b) $T_2(\mathbf{u}, \mathbf{v}_1 + \mathbf{v}_2) = T_2(\mathbf{u}, \mathbf{v}_1) + T_2(\mathbf{u}, \mathbf{v}_2)$;
(c) $T_2(\lambda\mathbf{u}, \mathbf{v}) = \lambda\, T_2(\mathbf{u}, \mathbf{v})$;
(d) $T_2(\mathbf{u}, \mu\mathbf{v}) = \mu\, T_2(\mathbf{u}, \mathbf{v})$,

para todo $\mathbf{u}_1, \mathbf{u}_2, \mathbf{v}_1$ y $\mathbf{v}_2 \in V$ y λ y $\mu \in \mathbb{R}$.

Definición V.1.2. Sea T_2 una forma bilineal sobre V. Se dice que T_2 es **simétrica** si $T(\mathbf{u}, \mathbf{v}) = T(\mathbf{v}, \mathbf{u})$, para todo $\mathbf{u}, \mathbf{v} \in V$. Se dice que T_2 es **antisimétrica** si $T(\mathbf{u}, \mathbf{v}) = -T(\mathbf{v}, \mathbf{u})$, para todo $\mathbf{u}, \mathbf{v} \in V$.

Ejemplo V.1.3. Sean $V = \mathbb{R}^2$ y $T_2 : V \times V \longrightarrow \mathbb{R}$ tal que $T_2((x_1, x_2), (y_1, y_2)) = x_1 y_2$. La aplicación T_2 es una forma bilineal que no es simétrica, pues $T_2((1, 0), (0, 1)) = 1 \neq 0 = T_2((0, 1), (1, 0))$.

Matriz asociada a una forma bilineal.

Definición V.1.4. Sean T_2 una forma bilineal sobre V y $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ una base de V. Se llama **matriz asociada a** T_2 **respecto de** $\mathcal{B}$ a la matriz $A = (a_{ij}) \in \mathcal{M}_n(\mathbb{R})$ determinada por las igualdades $a_{ij} = T_2(\mathbf{v}_i, \mathbf{v}_j)$, para cada $i, j \in \{1, \ldots, n\}$.

Conocida la matriz asociada a una forma bilineal respecto de una base podemos determinar las imágenes por la forma bilineal de cualquier par de vectores de V.

Proposición V.1.5. *Sean T_2 una forma bilineal sobre V y $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ una base de V. Dados $\mathbf{x}$ e $\mathbf{y} \in V$ de coordenadas $(x_1, \ldots, x_n)$ y $(y_1, \ldots, y_n)$ respecto de $\mathcal{B}$ se cumple que*

$$T_2(\mathbf{x}, \mathbf{y}) = (x_1 \ \ldots \ x_n)\, A \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix},$$

donde A es la matriz asociada a T_2 respecto de $\mathcal{B}$.

Demostración. Teniendo en cuenta que T_2 es bilineal y la propia definición de A se sigue que

$$\begin{aligned} T_2(\mathbf{x}, \mathbf{y}) &= \sum_{i=1}^{n} x_i T_2(\mathbf{v}_i, \mathbf{y}) = \sum_{i,j=1}^{n} x_i y_j T_2(\mathbf{v}_i, \mathbf{v}_j) \\ &= \sum_{i,j=1}^{n} x_i a_{ij} y_j = (x_1 \ \ldots \ x_n)\, A \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}. \end{aligned}$$

■

Ejemplo V.1.6. Sobre $\mathbb{R}^n$ consideramos la aplicación $T_2 : \mathbb{R}^n \times \mathbb{R}^n \longrightarrow \mathbb{R}$ tal que

$$T_2(\mathbf{x}, \mathbf{y}) = x_1 y_1 + \ldots + x_n y_n = \sum_{i=1}^{n} x_i y_i,$$

para todo $\mathbf{x} = (x_1, \ldots, x_n)^{\mathrm{t}}$ y $\mathbf{y} = (y_1, \ldots, y_n)^{\mathrm{t}} \in \mathbb{R}^n$. La aplicación T_2 es una forma bilineal simétrica.

i) Si $\mathcal{B}$ es la base usual de $\mathbb{R}^n$, entonces la matriz asociada a T_2 respecto de $\mathcal{B}$ es la matriz identidad de orden n.

ii) Si $\mathcal{B}' = \{(1, 0, 0 \ldots, 0, 0), (1, 1, 0 \ldots, 0, 0), \ldots, (1, 1, 1, \ldots, 1, 0), (1, 1, 1, \ldots, 1, 1)\}$, entonces la matriz asociada a T_2 respecto de $\mathcal{B}'$ es $A = (a_{ij}) \in \mathcal{M}_n(\mathbb{R})$ donde $a_{ij} = \min(i, j)$, para cada $i, j \in \{1, \ldots, n\}$, es decir,

$$A = \begin{pmatrix} 1 & 1 & \ldots & 1 \\ 1 & 2 & \ldots & 2 \\ \vdots & \vdots & & \vdots \\ 1 & 2 & \ldots & n \end{pmatrix}.$$

Obsérvese que, como era de esperar, una misma forma bilineal tiene distintas matrices respecto de diferentes bases.

Corolario V.1.7. *Sean T_2 una forma bilineal sobre V, $\mathcal{B}$ una base de V, $A \in \mathcal{M}_n(\mathbb{R})$ la matriz asociada a T_2 respecto de $\mathcal{B}$. La forma bilineal T_2 es simétrica si, y sólo si, la matriz A es simétrica (es decir, $A = A^{\mathrm{t}}$).*

Demostración. Dados $\mathbf{x}$ y $\mathbf{y} \in V$ de coordenadas $(x_1, \ldots, x_n)$ e $(y_1, \ldots, y_n)$ respecto de $\mathcal{B}$, respectivamente, se tiene que

$$T(\mathbf{x}, \mathbf{y}) = (x_1 \ \ldots \ x_n)\, A \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

y que

$$T(\mathbf{y}, \mathbf{x}) = (y_1 \ \ldots \ y_n)\, A \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = (x_1 \ \ldots \ x_n)\, A^{\mathrm{t}} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix};$$

de donde se deduce el resultado buscado. ■

Terminamos esta sección estudiando cómo afectan los **cambios de base** en la matriz de una forma bilineal sobre V.

Proposición V.1.8. *Sean T_2 una forma bilineal sobre V y $\mathcal{B}$ y $\mathcal{B}'$ dos bases de V. Si $A = (a_{ij}) \in \mathcal{M}_n(\mathbb{R})$ y $A' = (a_{ij}) \in \mathcal{M}_n(\mathbb{R})$ son las matrices asociadas a T_2 respecto de $\mathcal{B}$ y $\mathcal{B}'$, respectivamente, y $P = (p_{hi}) \in \mathcal{M}_n(\mathbb{R})$ es la matriz del cambio de la base $\mathcal{B}'$ a la base $\mathcal{B}$ entonces*

$$A' = P^{\mathrm{t}}AP.$$

Demostración. Basta tener en cuenta que, por las definiciones de forma bilineal, matriz asociada a una forma bilineal y de producto de matrices se tiene, se tiene que

$$a'_{ij} = \sum_{h,l=1}^{n} p_{hi}p_{lj}a_{hl} = \sum_{h,l=1}^{n} p_{hi}a_{hl}p_{lj},$$

es decir, $A' = P^{\mathrm{t}}AP$. ∎

Definición V.1.9. Dadas A y $A' \in \mathcal{M}_n(\mathbb{R})$, se dice que A' es **congruente** con A si existe una matriz invertible $P \in \mathcal{M}_n(\mathbb{R})$ tal que $A' = P^{\mathrm{t}}AP$.

Es claro que la relación "ser congruente con" es de equivalencia (es decir, verifica las propiedades reflexiva, simétrica y transitiva).

Nota V.1.10. Obsérvese que, según la proposición V.1.8, dos matrices A *y* $A' \in \mathcal{M}_n(\mathbb{R})$ *son congruentes si, y sólo si, representan a una misma forma bilineal expresada respecto de distintas bases.*

2. Producto escalar. Espacios vectoriales euclídeos

Definición V.2.1. Sea T_2 una forma bilineal sobre V. Se dice que T_2 es **definida positiva** si $T_2(\mathbf{u}, \mathbf{u}) > 0$, para todo $\mathbf{u} \in V$ no nulo.

Nótese que si T_2 es una forma bilineal definida positiva sobre V, entonces $T_2(\mathbf{v}, \mathbf{v}) = 0$ si y sólo si $\mathbf{v} = \mathbf{0}$. En particular, se tiene que la matriz, A, de T_2 respecto de cualquier base de V es invertible; en otro caso, existiría $\mathbf{v} \in \ker(A)$ no nulo y se tendría que $T_2(\mathbf{v}, \mathbf{v}) = \mathbf{v}^{\mathrm{t}}A\mathbf{v} = \mathbf{v}^{\mathrm{t}}\mathbf{0} = 0$.

Ejemplo V.2.2. Sean $V = \mathbb{R}^2$.

(a) $T_2((x_1, x_2), (y_1, y_2)) = x_1y_1 + x_2y_2$, es una forma bilineal simétrica (compruébese) que es definida positiva pues $T_2((x_1, x_2), (x_1, x_2)) = x_1^2 + x_2^2 > 0$ para todo $(x_1, x_2) \in \mathbb{R}^2$ no nulo.

(b) $T_2((x_1, x_2), (y_1, y_2)) = x_1y_1 - x_2y_2$, es una forma bilineal simétrica (compruébese) que no es definida positiva pues $T_2((0, 1), (0, 1)) = -1 < 0$.

Definición V.2.3. Llamaremos **espacio vectorial euclídeo** a todo par (V, T_2) donde V es un $\mathbb{R}$-espacio vectorial y T_2 es una forma bilineal simétrica definida positiva.

Las formas bilineales simétricas definidas positivas son **productos escalares**. Así, no es de extrañar que, dado un espacio vectorial euclídeo (V, T_2), se use la notación multiplicativa $\cdot$ y se escriba $(V, \cdot)$ (ó simplemente V) en lugar de (V, T_2) y $\mathbf{u} \cdot \mathbf{v}$ en vez de $T_2(\mathbf{u}, \mathbf{v})$.

Ejemplo V.2.4. Sobre $\mathbb{R}^n$ consideramos la aplicación $\cdot : \mathbb{R}^n \times \mathbb{R}^n \longrightarrow \mathbb{R}$ tal que

$$\mathbf{u} \cdot \mathbf{v} = u_1 v_1 + \ldots + u_n v_n = \sum_{i=1}^{n} u_i v_i,$$

para todo $\mathbf{u} = (u_1, \ldots, u_n)^{\mathrm{t}}$ y $\mathbf{v} = (v_1, \ldots, v_n)^{\mathrm{t}} \in \mathbb{R}^n$. La aplicación $\cdot$ es una forma bilineal simétrica y definida positiva. Luego, $\cdot$ es un producto escalar sobre $\mathbb{R}^n$, y por tanto dota a $\mathbb{R}^n$ de estructura de espacio vectorial euclídeo, es decir, el par $(\mathbb{R}^n, \cdot)$ es un espacio vectorial euclídeo.

El producto escalar definido anteriormente se llama **producto escalar usual**, de aquí que a $(\mathbb{R}^n, \cdot)$ se le llame **espacio vectorial euclídeo usual**. Nótese que la matriz asociada a la forma bilineal T_2 respecto de la base usual de $\mathbb{R}^n$ es la matriz identidad de orden n (véase el ejemplo V.1.6(a)).

Conviene resaltar que se pueden definir infinidad de formas bilineales sobre un mismo $\mathbb{R}$-espacio vectorial. La forma bilineal usual no es más que una de ellas.

Ejemplo V.2.5. Sobre $\mathbb{R}^3$ consideramos una forma bilineal $T_2 : \mathbb{R}^3 \times \mathbb{R}^3 \longrightarrow \mathbb{R}$ cuya matriz asociada respecto de una base $\mathcal{B}$ de $\mathbb{R}$ es

$$A = \begin{pmatrix} 1 & 1 & -1 \\ 1 & 2 & 1 \\ -1 & 1 & 6 \end{pmatrix}.$$

Como la forma bilineal T_2 es simétrica (véase el corolario V.1.7) y definida positiva[1], T_2 dota a $\mathbb{R}^3$ de estructura de espacio vectorial euclídeo. Además, si $\mathbf{x}$ e $\mathbf{y}$ son vectores de $\mathbb{R}^3$ de coordenadas (x_1, x_2, x_3) e (y_1, y_2, y_3) respecto de $\mathcal{B}$, respectivamente, entonces, por la proposición V.1.5, tenemos que

$$\mathbf{x} \cdot \mathbf{y} = x_1 y_1 + x_2 y_1 - x_3 y_1 + x_1 y_2 + 2x_2 y_2 + x_3 y_2 - x_1 y_3 + x_2 y_3 + 6x_3 y_3.$$

Módulo de un vector. Distancia.

Si $\mathbf{u}$ y $\mathbf{v}$ dos vectores no nulos de V linealmente dependientes, entonces sabemos que existe $\alpha \in \mathbb{R}$ tal que $\mathbf{v} = \alpha\mathbf{u}$. En este caso podemos decir que "$\mathbf{v}$ es α veces $\mathbf{u}$", y ampliar esta comparación a todos los vectores de $\langle \mathbf{u} \rangle$. Sin embargo, cuando $\mathbf{u}$ y $\mathbf{v}$ son linealmente independientes esta comparación no tiene ningún sentido.

[1] Más adelante veremos que una forma bilineal simétrica es definida positiva si y sólo si los menores principales de su matriz asociada respecto alguna base de V son estrictamente positivos.

Una de las principales aportaciones del producto escalar en un espacio vectorial euclídeo es que nos permite "comparar" dos vectores no necesariamente linealmente dependientes.

Definición V.2.6. Sea V un espacio vectorial euclídeo. Se llama **norma** (o **módulo**) **de un vector** $\mathbf{v} \in V$ al único número real no negativo, que denotamos por $||\mathbf{v}||$ tal que $\mathbf{v} \cdot \mathbf{v} = ||\mathbf{v}||^2$. Así mismo, se define la **distancia**[2] entre $\mathbf{u}$ y $\mathbf{v} \in V$ como el número real $\mathrm{d}(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|$.

Nótese que, como el producto escalar valora en $\mathbb{R}$ y $\mathbf{v} \cdot \mathbf{v} > 0$ para todo $\mathbf{v} \in V$ no nulo, tiene perfecto sentido considerar $||\mathbf{v}|| = (\mathbf{v} \cdot \mathbf{v})^{1/2}$. Asimismo destacamos que la norma del vector $\mathbf{0}$ es 0; de hecho, es el único vector de norma cero, por ser $\cdot$ una forma bilineal definida positiva.

Nota V.2.7. En los temas VIII y XII se estudiarán los espacios vectoriales (arbitrarios) dotados de una norma (véase la definición VIII.1.1) y de un producto escalar (véase la definición XII.1.1), respectivamente, entre lo que se encontrarán los espacios vectoriales euclídeos como ejemplo notable en ambos casos.

3. Ortogonalidad. Bases ortogonales y ortonormales

Definición V.3.1. Diremos que dos vectores $\mathbf{u}$ y $\mathbf{v} \in V$ son **ortogonales** si $\mathbf{u} \cdot \mathbf{v} = 0$.

Definición V.3.2. Diremos que los vectores de un conjunto $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ de V, con $\mathbf{v}_i \neq \mathbf{0},\ i = 1, \ldots, r$, son **ortogonales entre sí** si $\mathbf{v}_i \cdot \mathbf{v}_j = 0$ para todo $i \neq j$. En este caso diremos que $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ es un **conjunto ortogonal**.

Proposición V.3.3. *Si $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\} \subseteq V$ es un conjunto ortogonal, entonces es un conjunto linealmente independiente.*

Demostración. Si $\lambda_1 \mathbf{v}_1 + \ldots + \lambda_r \mathbf{v}_r = \mathbf{0}$, para ciertos $\lambda \in \mathbb{R},\ i = 1, \ldots, r$, entonces

$$0 = (\lambda_1 \mathbf{v}_1 + \ldots + \lambda_r \mathbf{v}_r) \cdot \mathbf{v}_i = \lambda_i \mathbf{v}_i \cdot \mathbf{v}_i,$$

para cada $i = 1, \ldots, r$. Teniendo en cuenta que todo producto escalar es una forma bilineal definida positiva y que $\mathbf{v}_i \neq \mathbf{0}$, para todo $i \in \{1, \ldots, r\}$, se sigue que $\mathbf{v}_i \cdot \mathbf{v}_i \neq 0$ y por lo tanto que $\lambda_i = 0$, para todo $i = 1, \ldots, r$. ∎

Obsérvese que cualquier conjunto ortogonal tiene, a lo más, n vectores; en otro caso, no sería linealmente independiente.

[2]El lector interesado puede comprobar que efectivamente se trata de una distancia (véase la definición A.1.1). Así, podemos afirmar que todo espacio vectorial euclídeo es un espacio métrico.

Ejemplo V.3.4. Es claro que el recíproco de la proposición anterior no es cierto en general. Por ejemplo, en $\mathbb{R}^2$ con el producto escalar usual, se tiene que $\{(1,1),(0,1)\}$ es un conjunto linealmente independiente que no es conjunto ortogonal; $(1,1)\cdot(0,1) = 1 \neq 0$.

El hecho de que todo conjunto ortogonal sea linealmente independiente implica que cualquier conjunto ortogonal que genere al espacio vectorial euclídeo V es base de V.

Definición V.3.5. Diremos que un conjunto de vectores $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ de V es una **base ortogonal** si es conjunto ortogonal que genera a V.

Nótese que $\mathcal{B}$ es una base ortogonal de V si, sólo si, la matriz asociada al producto escalar definido sobre V es diagonal.

Definición V.3.6. Se dice que un vector $\mathbf{v} \in V$ es **unitario** si $\|\mathbf{v}\| = 1$.

Teniendo en cuenta la definición de norma de un vector se tiene que un vector $\mathbf{v} \in V$ es unitario si y sólo si $\mathbf{v} \cdot \mathbf{v} = 1$.

Definición V.3.7. Diremos que $\mathcal{B} = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\} \subseteq V$ es una **base ortonormal** de V si es base ortogonal formada por vectores unitarios, es decir, si $\mathbf{u}_i \cdot \mathbf{u}_j = \delta_{ij}$, donde δ_{ij} es la función Delta de Kronecker.

Ejemplo V.3.8. Veamos algunos ejemplos de bases ortonormales.

(a) La base usual de $\mathbb{R}^n$ es una base ortonormal para el producto escalar usual de $\mathbb{R}^n$.

(b) Sobre $\mathbb{R}^3$ consideramos el producto escalar $\cdot$ cuya matriz respecto de la base usual de $\mathbb{R}^3$ es

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

La base $\mathcal{B} = \{(1,1,0),(0,1,-1),(0,0,1)\}$ del espacio vectorial euclídeo $(\mathbb{R}^3, \cdot)$ es ortonormal.

Método de ortonormalización de Gram-Schmidt (caso finito).

Sea $\mathcal{B} = \{\mathbf{w}_1, \ldots, \mathbf{w}_n\}$ una base de V. Vamos a describir un procedimiento para construir, a partir de $\mathcal{B}$, una base ortonormal de V.

- Definimos $\mathbf{v}_1 = \mathbf{w}_1$ y $\mathbf{v}_2 = \mathbf{w}_2 + \mu_{12}\mathbf{v}_1$, donde $\mu_{12} \in \mathbb{R}$ se elige de modo que $\mathbf{v}_1$ y $\mathbf{v}_2$ sean ortogonales. Es decir, como queremos que

$$0 = \mathbf{v}_1 \cdot \mathbf{v}_2 = \mathbf{v}_1 \cdot (\mathbf{w}_2 + \mu_{12}\mathbf{v}_1) = \mathbf{v}_1 \cdot \mathbf{w}_2 + \mu_{12}(\mathbf{v}_1 \cdot \mathbf{v}_1) = \mathbf{v}_1 \cdot \mathbf{w}_2 + \mu_{12}\|\mathbf{v}_1\|^2,$$

tomamos $\mu_{12} = -(\mathbf{v}_1 \cdot \mathbf{w}_2)/\|\mathbf{v}_1\|^2$ y por lo tanto

$$\mathbf{v}_2 = \mathbf{w}_2 - \frac{\mathbf{v}_1 \cdot \mathbf{w}_2}{\|\mathbf{v}_1\|^2}\,\mathbf{v}_1.$$

- Definimos a continuación $\mathbf{v}_3 = \mathbf{w}_3 + \mu_{13}\mathbf{v}_1 + \mu_{23}\mathbf{v}_2$ eligiendo μ_{13} y $\mu_{23} \in \mathbb{R}$ tales que $\mathbf{v}_1 \cdot \mathbf{v}_3 = 0$ y $\mathbf{v}_2 \cdot \mathbf{v}_3 = 0$. Es decir, como queremos que

$$\begin{aligned} 0 &= \mathbf{v}_1 \cdot \mathbf{v}_3 = \mathbf{v}_1 \cdot (\mathbf{w}_3 + \mu_{13}\mathbf{v}_1 + \mu_{23}\mathbf{v}_2) = \mathbf{v}_1 \cdot \mathbf{w}_3 + \mu_{13}\mathbf{v}_1 \cdot \mathbf{v}_1 + \mu_{23}\mathbf{v}_1 \cdot \mathbf{v}_2 \\ &= \mathbf{v}_1 \cdot \mathbf{w}_3 + \mu_{13}\|\mathbf{v}_1\|^2 \end{aligned}$$

y que

$$\begin{aligned} 0 &= \mathbf{v}_2 \cdot \mathbf{v}_3 = \mathbf{v}_2 \cdot (\mathbf{w}_3 + \mu_{13}\mathbf{v}_1 + \mu_{23}\mathbf{v}_2) = \mathbf{v}_2 \cdot \mathbf{w}_3 + \mu_{13}\mathbf{v}_2 \cdot \mathbf{v}_1 + \mu_{23}\mathbf{v}_2 \cdot \mathbf{v}_2 \\ &= \mathbf{v}_2 \cdot \mathbf{w}_3 + \mu_{23}\|\mathbf{v}_2\|^2 \end{aligned},$$

tomamos $\mu_{13} = -(\mathbf{v}_1 \cdot \mathbf{w}_3)/\|\mathbf{v}_1\|^2$ y $\mu_{23} = -(\mathbf{v}_2 \cdot \mathbf{w}_3)/\|\mathbf{v}_2\|^2$ y por lo tanto

$$\mathbf{v}_3 = \mathbf{w}_3 - \frac{\mathbf{v}_1 \cdot \mathbf{w}_3}{\|\mathbf{v}_1\|^2}\,\mathbf{v}_1 - \frac{\mathbf{v}_2 \cdot \mathbf{w}_3}{\|\mathbf{v}_2\|^2}\,\mathbf{v}_2.$$

- Repitiendo el proceso anterior definimos $\mathbf{v}_j = \mathbf{w}_j + \mu_{1j}\mathbf{v}_1 + \mu_{2j}\mathbf{v}_2 + \ldots + \mu_{j-1\,j}\mathbf{v}_{j-1}$, tomando $\mu_{ij} \in \mathbb{R}$ tal que $\mathbf{v}_j \cdot \mathbf{v}_i = 0$, para cada $i < j$ e $j = 4, \ldots, n$. Se comprueba fácilmente que

$$\mathbf{v}_j = \mathbf{w}_j - \sum_{i=1}^{j-1} \frac{\mathbf{v}_i \cdot \mathbf{w}_j}{\|\mathbf{v}_i\|^2}\mathbf{v}_i,$$

para cada $j = 4, \ldots, n$.

En resumen mediante el proceso anterior hemos obtenido un conjunto ortogonal de vectores $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$, donde

$$\begin{aligned} \mathbf{v}_1 &= \mathbf{w}_1 \\ \mathbf{v}_j &= \mathbf{w}_j - \textstyle\sum_{i=1}^{j-1} \frac{\mathbf{v}_i \cdot \mathbf{w}_j}{\|\mathbf{v}_i\|^2}\mathbf{v}_i, \qquad j = 2, \ldots, n, \end{aligned}$$

que forma una base de V, pues, por la proposición V.3.3, $\mathcal{B}' = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ es un conjunto linealmente independiente y $\dim V = n$. Luego $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ es una base ortogonal de V.

Finalmente, sin más que tomar $\mathbf{u}_j = \|\mathbf{v}_j\|^{-1}\mathbf{v}_j,\ j = 1, \ldots, n$, obtenemos que $\mathcal{B}'' = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ es una base ortonormal de V.

Veamos ahora algunas consecuencias inmediatas del método de ortonormalización de Gram-Schmidt.

Corolario V.3.9. *En todo espacio vectorial euclídeo existen bases ortonormales.*

Nota V.3.10. Siguiendo con la misma notación que en el método de Gram-Schmidt, si elegimos $p_{ii} = 1,\ i = 1, \ldots, n,\ p_{ij} = \frac{\mathbf{v}_i \cdot \mathbf{w}_j}{\|\mathbf{v}_i\|^2}$ para todo $j > i$ y $p'_{ij} = 0$ para todo $j < i$, entonces la matriz del cambio de la base $\mathcal{B}$ a $\mathcal{B}'$ es la matriz triangular

superior $P = (p_{ij}) \in \mathcal{M}_n(\mathbb{R})$. Además, si tomamos $r_{ij} = p_{ij}/\|\mathbf{v}_j\|$, entonces la matriz del cambio de base de $\mathcal{B}$ a $\mathcal{B}''$ es la matriz triangular superior $R = (r_{ij}) \in \mathcal{M}_n(\mathbb{R})$.

Corolario V.3.11. *Si $A \in \mathcal{M}_n(\mathbb{R})$ es invertible, existen $Q \in \mathcal{M}_n(\mathbb{R})$ ortogonal y $R \in \mathcal{M}_n(\mathbb{R})$ triangular superior e invertible tales que*

$$A = Q\,R.$$

Esta descomposición se conoce como ***factorización QR*** *de A.*

Demostración. Como A es invertible, sus columnas forman una base $\mathcal{B}$ de $\mathbb{R}^n$. Considerando el producto escalar usual de $\mathbb{R}^n$ y aplicando el método de Gram-Schmidt a $\mathcal{B}$ obtenemos una base ortonormal $\mathcal{B}'$ de $\mathbb{R}^n$. Por tanto, basta tomar Q como la matriz cuyas columnas son los vectores de $\mathcal{B}'$ y R como la matriz del cambio de base de $\mathcal{B}$ a $\mathcal{B}'$, para obtener el resultado buscado, pues Q es claramente ortonormal y R es triangular superior e invertible por la nota V.3.10 y por ser la matriz de un cambio de base, respectivamente. ∎

Ejemplo V.3.12. Sobre $\mathbb{R}^3$ consideramos el producto escalar $\cdot$ cuya matriz respecto de la base usual de $\mathbb{R}^3$ es

$$A = \begin{pmatrix} 6 & 3 & -1 \\ 3 & 2 & -1 \\ -1 & -1 & 1 \end{pmatrix}.$$

Como la matriz viene dada respecto de la base usual, partimos de $\mathcal{B} = \{\mathbf{e}_1 = (1,0,0), \mathbf{e}_2 = (0,1,0), \mathbf{e}_3 = (0,0,1)\}$.

En primer lugar tomamos $\mathbf{v}_1 = \mathbf{e}_1 = (1,0,0)$ y definimos $\mathbf{v}_2 = \mathbf{e}_2 + \mu_{12}\mathbf{v}_1$, eligiendo $\mu_{21} \in \mathbb{R}$ tal que $\mathbf{v}_1$ y $\mathbf{v}_2$ sean ortogonales. Según el método de Gram-Schmidt debemos tomar

$$\mu_{12} = -\frac{\mathbf{v}_1 \cdot \mathbf{e}_2}{\|\mathbf{v}_1\|^2} = -\frac{1}{2},$$

y por lo tanto $\mathbf{v}_2 = \mathbf{e}_2 - \frac{1}{2}\mathbf{v}_1 = (-1/2, 1, 0)$. Definimos ahora $\mathbf{v}_3 = \mathbf{e}_3 + \mu_{13}\mathbf{e}_1 + \mu_{23}\mathbf{e}_2$, eligiendo μ_{13} y $\mu_{23} \in \mathbb{R}$ tales que $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ sea un conjunto ortogonal. Según el método de Gram-Schmidt debemos tomar

$$\mu_{13} = -\frac{\mathbf{v}_1 \cdot \mathbf{e}_3}{\|\mathbf{v}_1\|^2} = \frac{1}{6} \quad \text{y} \quad \mu_{23} = -\frac{\mathbf{v}_2 \cdot \mathbf{e}_3}{\|\mathbf{v}_2\|^2} = 1,$$

y por lo tanto $\mathbf{v}_3 = \mathbf{e}_3 + \frac{1}{6}\mathbf{v}_1 + \mathbf{v}_2 = (-1/6, 1, 1)$.

Así obtenemos que $\mathcal{B}' = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ con $\mathbf{v}_1 = \mathbf{e}_1 = (1,0,0)$, $\mathbf{v}_2 = \mathbf{e}_2 - \frac{1}{2}\mathbf{v}_1 = (-1/2, 1, 0)$ y $\mathbf{v}_3 = \mathbf{e}_3 + \frac{1}{6}\mathbf{v}_1 + \mathbf{v}_2 = (-1/3, 1, 1)$, es una base ortogonal de $\mathbb{R}^3$. Y una base ortonormal de $\mathbb{R}^3$ es $\mathcal{B}'' = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ con $\mathbf{u}_1 = \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|} = (\sqrt{6}/6, 0, 0)$, $\mathbf{u}_2 = \frac{\mathbf{v}_2}{\|\mathbf{v}_2\|} = (-\sqrt{2}/2, \sqrt{2}, 0)$ y $\mathbf{u}_3 = \frac{\mathbf{v}_3}{\|\mathbf{v}_3\|} = (-\sqrt{3}/3, \sqrt{3}, \sqrt{3})$.

Proposición V.3.13. *Sean V un espacio vectorial euclídeo y $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ una base ortogonal de V. Dado $\mathbf{v} \in V$, se cumple que*

$$\mathbf{v} = \frac{\mathbf{v} \cdot \mathbf{v}_1}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 + \ldots + \frac{\mathbf{v} \cdot \mathbf{v}_n}{\|\mathbf{v}_n\|^2}\mathbf{v}_n.$$

Además, si $\mathcal{B}$ es ortonormal, entonces $\mathbf{v} = (\mathbf{v} \cdot \mathbf{v}_1)\mathbf{v}_1 + \ldots + (\mathbf{v} \cdot \mathbf{v}_n)\mathbf{v}_n$.

Demostración. Como $\mathcal{B}$ es una base de V, existen $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$ tales que $\mathbf{v} = \sum_{i=1}^n \alpha_i \mathbf{v}_i$. Como $\mathcal{B}$ es ortogonal, se tiene que

$$\mathbf{v} \cdot \mathbf{v}_j = \big(\sum_{i=1}^n \alpha_i \mathbf{v}_i\big) \cdot \mathbf{v}_j = \sum_{i=1}^n \alpha_i (\mathbf{v}_i \cdot \mathbf{v}_j) = \alpha_j \, (\mathbf{v}_j \cdot \mathbf{v}_j),$$

de donde se sigue que $\alpha_j = (\mathbf{v} \cdot \mathbf{v}_j)/\|\mathbf{v}_j\|$, para cada $j = 1, \ldots, n$.

Finalmente, si $\mathcal{B}$ es además ortonormal, entonces $\|\mathbf{v}_j\| = 1$, para todo $j = 1, \ldots, n$; luego, $\alpha_j = \mathbf{v} \cdot \mathbf{v}_j$, para cada $j = 1, \ldots, n$. ∎

De la proposición anterior se deduce que las coordenadas de un vector $\mathbf{v}$ de un espacio vectorial euclídeo V respecto de una base ortonormal $\mathcal{B} = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ de V, son $(\mathbf{v} \cdot \mathbf{u}_1, \ldots, \mathbf{v} \cdot \mathbf{u}_n)$.

Nota V.3.14. Destacamos que $\mathcal{B} = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ es una base ortonormal de V si, sólo si, la matriz asociada al producto escalar definido sobre V es la matriz identidad de orden n. Este hecho permite obtener una expresión en coordenadas del producto escalar respecto de $\mathcal{B}$ realmente sencilla: Sean V espacio vectorial euclídeo y $\mathcal{B} = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ una base ortonormal de V. En virtud de la proposición V.1.5, si $\mathbf{x}$ e $\mathbf{y} \in V$ tienen coordenadas $(x_1, \ldots, x_n)$ e $(y_1, \ldots, y_n)$ respecto de $\mathcal{B}$, entonces

$$\mathbf{x} \cdot \mathbf{y} = x_1 y_1 + \ldots + x_n y_n.$$

Luego a la vista de lo anterior, siempre que podamos asegurar la existencia de bases ortonormales en cualquier espacio vectorial euclídeo, podremos realizar un cambio de base de forma que la expresión en coordenadas del producto escalar sea "lo más sencilla posible".

Otro hecho a tener en cuenta es el siguiente:

Nota V.3.15. Sean $\mathcal{B} = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ y $\mathcal{B}' = \{\mathbf{u}'_1, \ldots, \mathbf{u}'_n\}$ dos bases ortonormales de V. Si $P = (p_{ij}) \in \mathcal{M}_n(\mathbb{R})$ es la matriz de cambio de la base $\mathcal{B}'$ a la base $\mathcal{B}$, entonces la matriz $P^{\mathrm{t}} I_n P = P^{\mathrm{t}} P$ es igual a la matriz identidad I_n, es decir, $P^{-1} = P^{\mathrm{t}}$. En efecto: por una parte, por ser $\mathcal{B}$ ortonormal, tenemos

$$(p_{1i}, \ldots, p_{ni}) \begin{pmatrix} p_{1j} \\ \vdots \\ p_{nj} \end{pmatrix} = \mathbf{u}'_i \cdot \mathbf{u}'_j,$$

y por otra parte, al ser $\mathcal{B}'$ ortonormal, obtenemos $\mathbf{u}'_i \cdot \mathbf{u}'_j = \delta_{ij}$.

Como consecuencia de lo anterior se sigue que la matriz de cambio de una base ortonormal a otra base ortonormal tiene determinante igual a ± 1 :

$$|P|^2 = |P||P| = |P^{\mathrm{t}}||P| = |P^{\mathrm{t}}P| = |I_n| = 1.$$

Recuérdese que un matriz $P \in \mathcal{M}_n(\mathbb{R})$ se dice ortogonal cuando $P^{\mathrm{t}} = P^{-1}$. Por tanto, según lo anterior, podemos afirmar que *las matrices de cambio de base ortonormales son las matrices ortogonales.*

4. Subespacio ortogonal. Proyección ortogonal

Veamos que el conjunto de todos los vectores que son ortogonales a los vectores de un subespacio L de V es un subespacio vectorial de V. Este subespacio se llama **subespacio ortogonal a** L y se denota por $L^{\perp}$.

Proposición V.4.1. *Sea L un subespacio de V. El conjunto*

$$L^{\perp} = \{\mathbf{v} \in V \ \mid \ \mathbf{v} \cdot \mathbf{u} = 0, \ \textit{para todo } \mathbf{u} \in L\}$$

es un subespacio vectorial de V.

Demostración. Basta tener en cuenta que, como el producto escalar es una forma bilineal sobre V, se tiene que $(\alpha\mathbf{v} + \beta\mathbf{w}) \cdot \mathbf{u} = \alpha(\mathbf{v} \cdot \mathbf{u}) + \beta(\mathbf{w} \cdot \mathbf{u}) = 0$, para todo $\mathbf{v}, \mathbf{w} \in L^{\perp}$, $\mathbf{u} \in L$ y α y $\beta \in \mathbb{R}$. ∎

Proposición V.4.2. *Sean L y L' dos subespacios vectoriales de V. Se cumple que:*

(a) $V^{\perp} = \{\mathbf{0}\}$ *y* $\{\mathbf{0}\}^{\perp} = V$;
(b) *Si* $L \subseteq L'$, *entonces* $(L')^{\perp} \subseteq L^{\perp}$;
(c) $(L + L')^{\perp} = L^{\perp} \cap (L')^{\perp}$ *y* $(L \cap L')^{\perp} = L^{\perp} + (L')^{\perp}$;
(d) $L^{\perp} \cap L = \{\mathbf{0}\}$.
(e) $\dim(L) + \dim(L^{\perp}) = \dim(V)$;
(f) $V = L \oplus L^{\perp}$.
(g) $(L^{\perp})^{\perp} = L$.

Demostración. (a) Si $\mathbf{v} \in V^{\perp}$, entonces $\mathbf{v} \cdot \mathbf{u} = 0$ para todo $\mathbf{u} \in V$, en particular, para $\mathbf{u} = \mathbf{v}$, se tiene que $\mathbf{v} \cdot \mathbf{v} = 0$; de donde se sigue que $\mathbf{v} = \mathbf{0}$, es decir, $V^{\perp} = \{\mathbf{0}\}$. Por otra parte, se tiene que $\mathbf{0} \cdot \mathbf{v} = 0$, para todo $\mathbf{v} \in V$, es decir, $\{\mathbf{0}\}^{\perp} = V$.

(b) Supongamos que $L \subseteq L'$ y sea $\mathbf{v} \in (L')^{\perp}$, entonces $\mathbf{v} \cdot \mathbf{u} = 0$, para todo $\mathbf{u} \in L'$, y como $L \subseteq L'$, se tiene que $\mathbf{v} \cdot \mathbf{u} = 0$, para todo $\mathbf{u} \in L$; de donde se sigue que $\mathbf{v} \in L^{\perp}$.

(c) Por el apartado (b), tomar ortogonales invierte las inclusiones. Luego, por un lado se tiene que el ortogonal del menor subespacio vectorial de V que contiene a L y a L', esto es el ortogonal de $L+L'$, es el mayor subespacio vectorial de V contenido en

$L^\perp$ y en $(L')^\perp$, esto es $L^\perp \cap (L')^\perp$. Y por otra parte, el ortogonal del mayor subespacio vectorial de V contenido en L y en L', esto es, el ortogonal de $L \cap L'$, es el menor subespacio vectorial de V que contiene a $L^\perp$ y a $(L')^\perp$, esto es, $L^\perp + (L')^\perp$.

(d) Si $\mathbf{v} \in L^\perp \cap L$, entonces $\mathbf{v} \cdot \mathbf{v} = 0$, de donde se sigue que $\mathbf{v} = \mathbf{0}$, es decir, $L^\perp \cap L = \{\mathbf{0}\}$.

(e) Supongamos que $\dim(L) = r \leq n$ y sea $\{\mathbf{u}_1, \ldots, \mathbf{u}_r, \mathbf{u}_{r-1}, \ldots, \mathbf{u}_n\}$ una base ortonormal de V tal que $\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$ es un base ortonormal L (lo que siempre se puede conseguir aplicando el método de Gram-Schmidt a la ampliación a V de una base de L). Es claro que, por construcción, $\mathbf{u}_{r-1}, \ldots, \mathbf{u}_n \subseteq L^\perp$ y como, por el apartado (d), $L^\perp \cap L = \{\mathbf{0}\}$, se sigue que $\langle \mathbf{u}_{r-1}, \ldots, \mathbf{u}_n \rangle = L^\perp$, es decir, $\dim(L^\perp) = n - r$.

(f) Es consecuencia directa de los apartados (d) y (e).

(g) Si $\mathbf{v} \in L$, entonces $\mathbf{v} \cdot \mathbf{u} = 0$, para todo $\mathbf{u} \in L^\perp$; luego, $L \subseteq (L^\perp)^\perp$. Teniendo ahora en cuenta que $\dim(L) = \dim((L^\perp)^\perp)$, pues, por el apartado (e), $\dim(L) + \dim(L^\perp) = \dim(V)$ y $\dim(L^\perp) + \dim((L^\perp)^\perp) = \dim(V)$, concluimos que $L = (L^\perp)^\perp$. ■

Proyección ortogonal de un vector sobre un subespacio.

Dado un subespacio vectorial L de V, por el apartado (f) de la proposición anterior tenemos que $V = L \oplus L^\perp$. Entonces, para cada $\mathbf{v} \in V$, existe unos únicos $\mathbf{v}_1 \in L$ y $\mathbf{v}_2 \in L^\perp$ tales que $\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2$. Dicho de otro modo, existe un único $\mathbf{v}_1 \in L$ tal que $\mathbf{v} - \mathbf{v}_1 \in L^\perp$.

Definición V.4.3. Sea L un subespacio vectorial de V. Dado $\mathbf{v} \in V$, se llama **proyección ortogonal de v sobre** L al único vector $\mathbf{v}_1 \in L$ tal que $\mathbf{v} - \mathbf{v}_1 \in L^\perp$.

Ejemplo V.4.4. Sea $\mathbf{u}$ un vector no nulo de un espacio vectorial euclídeo V. Veamos cómo es la proyección ortogonal sobre $L = \langle \mathbf{u} \rangle$, lo que se conoce por **proyección ortogonal sobre el vector u** : dado $\mathbf{v} \in V$, si $\mathbf{v}_1$ es la proyección ortogonal de $\mathbf{v}$ sobre L entonces $\mathbf{v}_1 \in \langle \mathbf{u} \rangle$ y $\mathbf{v} - \mathbf{v}_1 \in \langle \mathbf{u} \rangle^\perp$, es decir, existe $\alpha \in \mathbb{R}$ tal que $\mathbf{v}_1 = \alpha \mathbf{u}$ y $(\mathbf{v} - \alpha \mathbf{u}) \cdot \mathbf{u} = 0$; por lo tanto,

$$\mathbf{v}_1 = \frac{\mathbf{v} \cdot \mathbf{u}}{\|\mathbf{u}\|^2} \mathbf{u}.$$

Proposición V.4.5. *Sean L un subespacio vectorial de V y $\mathcal{B}_L = \{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ una base ortogonal de L. Si $\mathbf{v} \in V$, entonces su proyección ortogonal sobre L, es*

$$\frac{\mathbf{v} \cdot \mathbf{v}_1}{\|\mathbf{v}_1\|^2} \mathbf{v}_1 + \ldots + \frac{\mathbf{v} \cdot \mathbf{v}_r}{\|\mathbf{v}_r\|^2} \mathbf{v}_r.$$

Demostración. Basta comprobar que $\mathbf{v} - \frac{\mathbf{v}\cdot\mathbf{v}_1}{\|\mathbf{v}_1\|^2}\,\mathbf{v}_1 - \ldots - \frac{\mathbf{v}\cdot\mathbf{v}_r}{\|\mathbf{u}_r\|^2}\,\mathbf{v}_r \in L^\perp$. ∎

Nótese que si en la proposición anterior consideramos una base ortonormal $\mathcal{B}_L = \{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$ de L, entonces la proyección ortogonal de $\mathbf{v} \in V$ sobre L es $(\mathbf{v}\cdot\mathbf{u}_1)\,\mathbf{u}_1 + \ldots + (\mathbf{v}\cdot\mathbf{u}_r)\,\mathbf{u}_r$.

Teorema V.4.6. *Sean L un subespacio vectorial de un espacio vectorial euclídeo V y $\mathbf{v} \in V$. Si $\mathbf{v}_1$ es la proyección ortogonal de $\mathbf{v}$ sobre L, entonces*

$$\mathrm{d}(\mathbf{v}, \mathbf{v}_1) \leq \mathrm{d}(\mathbf{v}, \mathbf{u}),$$

para todo $\mathbf{u} \in L$.

Demostración. Sea $\mathbf{u} \in L$ distinto de $\mathbf{v}_1$, entonces $\mathbf{v} - \mathbf{u} = \mathbf{v} - \mathbf{v}_1 + \mathbf{v}_1 - \mathbf{u}$, con $\mathbf{v} - \mathbf{v}_1 \in L^\perp$ y $\mathbf{v}_1 - \mathbf{u} \in L$, es decir, $(\mathbf{v} - \mathbf{v}_1)\cdot(\mathbf{v}_1 - \mathbf{u}) = 0$. Entonces,

$$\|\mathbf{v} - \mathbf{u}\|^2 = \|\mathbf{v} - \mathbf{v}_1\|^2 + \|\mathbf{v}_1 - \mathbf{u}\|^2;$$

y se sigue que $\|\mathbf{v} - \mathbf{u}\| \geq \|\mathbf{v} - \mathbf{v}_1\|$ y se da la igualdad si, sólo si, $\mathbf{v} = \mathbf{v}_1$. Luego, $\mathrm{d}(\mathbf{v}, \mathbf{v}_1) = \|\mathbf{v} - \mathbf{v}_1\| \leq \|\mathbf{v} - \mathbf{u}\| = \mathrm{d}(\mathbf{v}, \mathbf{u})$, para todo $\mathbf{u} \in L$. ∎

El teorema anterior afirma que *la distancia de $\mathbf{v} \in V$ a L es igual a la distancia de $\mathbf{v}$ a su proyección ortogonal sobre L.*

Proyección ortogonal sobre un subespacio.

Sea V un espacio vectorial euclídeo de dimensión $n > 0$. Dado un subespacio vectorial L de V, se define la **proyección ortogonal sobre** L como la aplicación π_L que asigna a cada vector $\mathbf{v} \in V$ su proyección ortogonal sobre L, es decir, el único vector $\mathbf{v}_1 \in L$ tal que $\mathbf{v} - \mathbf{v}_1 \in L^\perp$, o dicho de otro modo, el vector de L más próximo a $\mathbf{v}$.

Lema V.4.7. *La proyección ortogonal π_L es un endomorfismo de V de imagen L y núcleo $L^\perp$; en particular,* $\mathrm{rg}(\pi_L) = \dim(L)$.

Demostración. La demostración es un sencillo ejercicio que se propone al lector. ∎

Sean ahora $\mathcal{B}$ una base de V y $A \in \mathcal{M}_n(\mathbb{R})$ la matriz del producto escalar de V respecto de $\mathcal{B}$. Si $\dim(L) = r$, las columnas de $B \in \mathcal{M}_{n\times r}(\mathbb{R})$ son las coordenadas respecto de $\mathcal{B}$ de los vectores de una base de L y $C = AB$, entonces se cumple que

Proposición V.4.8. *La matriz de π_L respecto de $\mathcal{B}$ es*

$$P = C(C^{\mathrm{t}}C)^{-1}C^{\mathrm{t}}.$$

Demostración. En primer lugar, como A es invertible (véase el comentario posterior a la definición V.2.1), se tiene que $\mathrm{rg}(C) = \mathrm{rg}(AB) = \mathrm{rg}(B) = r$. Por otra parte, se tiene que $C^{\mathrm{t}}C$ es simétrica e invertible[3]. Así pues, dado $\mathbf{v} \in \mathbb{R}^n$ se tiene que $P\mathbf{v} = C(C^{\mathrm{t}}C)^{-1}C^{\mathrm{t}}\mathbf{v} \in L$. Además, dado cualquier $\mathbf{u} \in \mathbb{R}^r$, se tiene que

$$\begin{aligned}(\mathbf{v} - P\mathbf{v})B\mathbf{u} &= (\mathbf{v} - P\mathbf{v})^{\mathrm{t}}AB\mathbf{u} = (\mathbf{v} - C(C^{\mathrm{t}}C)^{-1}C^{\mathrm{t}})\mathbf{v})^{\mathrm{t}}AB\mathbf{u}\\ &= \mathbf{v}^{\mathrm{t}}AB\mathbf{u} - \mathbf{v}^{\mathrm{t}}C(C^{\mathrm{t}}C)^{-1}C^{\mathrm{t}})AB\mathbf{u}\\ &= \mathbf{v}^{\mathrm{t}}AB\mathbf{u} - \mathbf{v}^{\mathrm{t}}(AB((AB)^{\mathrm{t}}(AB))^{-1}(AB)^{\mathrm{t}})AB\mathbf{u}\\ &= \mathbf{v}^{\mathrm{t}}AB\mathbf{u} - \mathbf{v}^{\mathrm{t}}AB\mathbf{u} = \mathbf{0},\end{aligned}$$

es decir, $\mathbf{v} - P\mathbf{v} \in L^{\perp}$. ∎

Obsérvese que de la proposición anterior se deduce que la matriz de una proyección ortogonal es simétrica e idempotente. Además, el recíproco de esta afirmación es cierto en el siguiente sentido: *si $P \in \mathcal{M}_n(\mathbb{R})$ es una matriz simétrica e idempotente, entonces la aplicación lineal $\mathbb{R}^n \to \mathbb{R}^n; \mathbf{x} \mapsto P\mathbf{x}$ es la proyección ortogonal sobre* $\mathrm{im}(P)$ (compruébese).

Proposición V.4.9. *Si L tiene rango r, existe una base ortonormal $\mathcal{B}'$ de V tal que la matriz de π_L respecto de $\mathcal{B}'$ es*

$$\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}.$$

Demostración. Basta tomar $\mathcal{B}'$ igual a la union de una base ortonormal de L con una base ortonormal de $L^{\perp}$. ∎

La proposición anterior no es más que un caso particular de una propiedad que estudiaremos con más detalle en la siguiente sección.

5. Matrices simétricas reales (y matrices hérmiticas)

A lo largo de esta sección consideraremos el espacio vectorial $\mathbb{R}^n$ con el producto escalar usual

$$\mathbf{x} \cdot \mathbf{y} = \sum_{i=1}^{n} x_i y_i,$$

donde $\mathbf{x} = (x_1, \ldots, x_n)^{\mathrm{t}}$ e $\mathbf{y} = (y_1, \ldots, y_n)^{\mathrm{t}} \in \mathbb{R}^n$; sabemos que, en este caso, la base usual $\mathcal{B} = \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ de $\mathbb{R}^n$ es ortonormal.

[3]La comprobación de que es simétrica es elemental. Para ver que es invertible, basta observar que $\mathbf{x}^{\mathrm{t}}C^{\mathrm{t}}C\mathbf{x} > 0$, para todo $\mathbf{x} \in \mathbb{R}^r$, por ser $\mathbf{x}^{\mathrm{t}}C^{\mathrm{t}}C\mathbf{x}$ el cuadrado de la norma de $C\mathbf{x}$ para el producto escalar usual de $\mathbb{R}^n$.

Diagonalización de matrices simétricas reales.

Lema V.5.1. *Si A es simétrica, para todo $\mathbf{x}$ e $\mathbf{y} \in \mathbb{R}^n$, se cumple que*

(a) $\mathbf{x} \cdot (A\mathbf{y}) = (A\mathbf{x}) \cdot \mathbf{y}$.
(b) $\mathbf{x} \cdot (A^m\mathbf{y}) = (A^m\mathbf{x}) \cdot \mathbf{y}$, *para cualquier* $m \in \mathbb{N}$.
(c) $\mathbf{x} \cdot (p(A)\mathbf{y}) = (p(A)\mathbf{x}) \cdot \mathbf{y}$, *para cualquier* $p(x) \in \mathbb{R}[x]$.

Demostración. (a) Si $\mathbf{x} = (x_1, \ldots, x_n)^{\mathrm{t}}$ e $\mathbf{y} = (y_1, \ldots, y_n)^{\mathrm{t}} \in \mathbb{R}^n$, entonces

$$\mathbf{x} \cdot (A\mathbf{y}) = (x_1, \ldots, x_n)A \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix},$$

y como $A = A^{\mathrm{t}}$,

$$\begin{aligned}(x_1, \ldots, x_n)A \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} &= \left(A^{\mathrm{t}} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}\right)^{\mathrm{t}} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \\ &= \left(A \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}\right)^{\mathrm{t}} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = (A\mathbf{x}) \cdot \mathbf{y}.\end{aligned}$$

(b) Sea $m \in \mathbb{N}$. Si A es simétrica, entonces A^m es simétrica; por consiguiente, basta aplicar el apartado (a) a la matriz A^m.

(c) Sea $p(x) = c_m x^m + \ldots + c_1 x + c_0 \in \mathbb{R}[x]$. Si A es simétrica, entonces $p(A) = c_m A^m + \ldots + c_1 A + c_0 I_n$ es simétrica, por consiguiente, basta aplicar el apartado (a) a la matriz $p(A)$. ∎

Proposición V.5.2. *Si $A \in \mathcal{M}_n(\mathbb{R})$ es simétrica, entonces dos autovectores asociados a autovalores distintos de A son ortogonales.*

Demostración. Sean λ y μ dos autovalores distintos de A y $\mathbf{u}$ y $\mathbf{v}$ autovectores de A asociados a λ y a μ, respectivamente. Entonces,

$$\lambda(\mathbf{u} \cdot \mathbf{v}) = (\lambda\mathbf{u})\mathbf{v} = (A\mathbf{u}) \cdot \mathbf{v} = \mathbf{u} \cdot (A\mathbf{v}) = \mathbf{u} \cdot (\mu\mathbf{v}) = \mu(\mathbf{u} \cdot \mathbf{v}),$$

y como λ y μ son distintos se concluye que $\mathbf{u} \cdot \mathbf{v} = 0$. ∎

Teorema V.5.3. *Si $A \in \mathcal{M}_n(\mathbb{R})$ es simétrica, entonces existe $P \in \mathcal{M}_n(\mathbb{R})$ ortogonal tal que $P^{\mathrm{t}}AP$ es diagonal; en particular, toda matriz simétrica es congruente con una matriz diagonal.*

Demostración. En primer lugar vamos a probar que todas las raíces del polinomio característico de A son reales, es decir que $\aleph_A(x)$ no tiene factores irreducibles de segundo grado.

Supongamos que un factor irreducible de $\aleph_A(x)$ es $p(x) = (x-\alpha)(x-\bar{\alpha}) = (x-a)^2 + b^2$, donde $\alpha = a + b\mathrm{i} \in \mathbb{C} \setminus \mathbb{R}$. Tomemos un vector no nulo[4] $\mathbf{v} \in \ker((A - aI_n)^2 + b^2 I_n)$. Entonces,

$$\begin{aligned}\mathbf{0} &= \big((A-aI_n)^2(\mathbf{v}) + b^2\mathbf{v}\big)\cdot\mathbf{v} = (A-aI_n)^2(\mathbf{v})\cdot\mathbf{v} + b^2\mathbf{v}\cdot\mathbf{v}\\ &= (A-aI_n)(\mathbf{v})\cdot(A-aI_n)(\mathbf{v}) + b^2(\mathbf{v}\cdot\mathbf{v}).\end{aligned}$$

donde la igualdad

$$(A-aI_n)^2(\mathbf{v})\cdot\mathbf{v} = (A-aI_n)(\mathbf{v})\cdot(A-aI_n)(\mathbf{v})$$

se debe a la simetría $A - aI_n$ (véase el lema V.5.1(a)). Además, si $(A-aI_n)\mathbf{v} = \mathbf{0}$, entonces $(A-aI_n)^2(\mathbf{v}) = \mathbf{0}$, y $b^2\mathbf{v} = \mathbf{0}$, lo que es contradictorio con $b \neq 0$.

Por tanto, como los vectores $\mathbf{v}$ y $(A-aI_n)\mathbf{v}$ son no nulos, tenemos que

$$(A-aI_n)(\mathbf{v})\cdot(A-aI_n)(\mathbf{v}) + b^2(\mathbf{v}\cdot\mathbf{v}) > 0,$$

con lo que, al suponer que el polinomio característico $\aleph_A(x)$ tiene algún factor irreducible de segundo grado, llegamos a una contradicción.

Probemos ahora que si λ es una raíz de $\aleph_A(x)$ con multiplicidad m, entonces $\ker(A-\lambda I_n) = \ker(A-\lambda I_n)^2$, en cuyo caso, tendremos que $\dim(\ker(A-\lambda I_n)) = m$ (véase el teorema III.5.10(a)). Si $\mathbf{v} \in \ker(A-\lambda I_n)^2$, entonces

$$\mathbf{0} = (A-\lambda I_n)^2\mathbf{v}\cdot\mathbf{v} = (A-\lambda I_n)\mathbf{v}\cdot(A-\lambda I_n)\mathbf{v},$$

luego, $(A-\lambda I_n)\mathbf{v} = \mathbf{0}$, es decir, $\mathbf{v} \in \ker(A-\lambda I_n)$.

Con esto queda probado que $\mathbb{R}^n = \ker(A-\lambda_1 I_n) \oplus \ldots \oplus \ker(A-\lambda_r I_n)$, es decir, que la matriz A es diagonalizable. Para obtener una base ortonormal de autovectores, tomamos una base ortonormal $\mathcal{B}'_i$ en cada uno de los subespacios $\ker(A-\lambda_i I)$. Por la proposición V.5.2, $\mathcal{B}' = \cup\mathcal{B}'_i$ es una base ortonormal de autovectores. ∎

Corolario V.5.4. *Sean $A \in \mathcal{M}_n(\mathbb{R})$ simétrica, $\lambda_1 \geq \ldots \geq \lambda_n$ los autovalores (posiblemente repetidos) de A y $P \in \mathcal{M}_n(\mathbb{R})$ una matriz ortogonal tal que $P^{\mathrm{t}}AP = D = (d_{ij}) \in \mathcal{M}_n(\mathbb{R})$ es diagonal con $d_{ii} = \lambda_i$, $i = 1, \ldots, n$. Si $\mathbf{u}_i$ denota a la columna i-ésima de P, entonces*

$$\lambda_i = \mathbf{u}_i \cdot A\mathbf{u}_i = \text{máx}\left\{\frac{\mathbf{v}\cdot A\mathbf{v}}{\|\mathbf{v}\|^2} \;\middle|\; \mathbf{v} \in \langle\mathbf{u}_i, \ldots, \mathbf{u}_n\rangle \setminus \{\mathbf{0}\}\right\},$$

para cada $i = 1, \ldots, n$.

[4] Si $\alpha \in \mathbb{C}\setminus\mathbb{R}$ es un autovalor de A y $\mathbf{z} \in \mathbb{C}^n$ es un autovector de A asociado a α, entonces $\bar{\mathbf{z}} \in \mathbb{C}^n$ es un autovector de A asociado a $\bar{\alpha}$ y $\mathbf{v} = \mathbf{z} - \bar{\mathbf{z}} \in \mathbb{R}^n$ es un vector no nulo de $\ker((A-aI_n)^2 + b^2 I_n)$.

Demostración. En primer lugar, observamos que si $\mathbf{u}_i$ es un autovector ortonormal asociado a λ_i, entonces $\lambda_i = \mathbf{u}_i \cdot A\mathbf{u}_i,\ i = 1, \ldots, n$, puesto que $P^{\mathrm{t}}AP = D$ y $d_{ii} = \lambda_i,\ i = 1, \ldots, n$.

Por otra parte, como

$$\frac{\mathbf{v} \cdot A\mathbf{v}}{\|\mathbf{v}\|^2} = \frac{(\alpha\mathbf{v}) \cdot A(\alpha\mathbf{v})}{\|\alpha\mathbf{v}\|^2},$$

para todo $\alpha \in \mathbb{R}$ y $\mathbf{v} \in \mathbb{R}^n$ no nulo, basta demostrar que

$$\lambda_i = \max \{\mathbf{v} \cdot A\mathbf{v} \ \mid \ \mathbf{v} \in \langle \mathbf{u}_i, \ldots, \mathbf{u}_n \rangle \text{ con } \|\mathbf{v}\| = 1\},$$

para cada $i = 1, \ldots, n$. Sea, pues, $\mathbf{v} \in \langle \mathbf{u}_i, \ldots, \mathbf{u}_n \rangle$ con $\|\mathbf{v}\| = 1$, es decir, $\mathbf{v} = \sum_{j=i}^{n} \alpha_j \mathbf{u}_j$, con $\sum_{j=i}^{n} \alpha_j^2 = 1$, entonces

$$\begin{aligned} \mathbf{v} \cdot A\mathbf{v} &= \big(\sum_{j=i}^{n} \alpha_j \mathbf{u}_j\big) \cdot \big(A(\sum_{j=i}^{n} \alpha_j \mathbf{u}_j)\big) = \big(\sum_{j=i}^{n} \alpha_j \mathbf{u}_j\big) \cdot \big(\sum_{j=i}^{n} \alpha_j (A\mathbf{u}_j)\big) \\ &= \big(\sum_{j=i}^{n} \alpha_j \mathbf{u}_j\big) \cdot \big(\sum_{j=i}^{n} \alpha_j (\lambda_j \mathbf{u}_j)\big) = \sum_{j=i}^{n} \lambda_j \alpha_j^2 \leq \lambda_i \sum_{j=i}^{n} \alpha_j^2 = \lambda_i, \end{aligned}$$

y la igualdad se alcanza en $\mathbf{v} = \mathbf{u}_i$. ∎

Corolario V.5.5. *Si $A \in \mathcal{M}_n(\mathbb{R})$ es simétrica de rango r, entonces existe $Q \in \mathcal{M}_n(\mathbb{R})$ invertible tal que*

$$Q^{\mathrm{t}}AQ = \begin{pmatrix} I_p & 0 & 0 \\ 0 & -I_q & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

donde I_p e I_q son las matrices identidad de ordenes p y q, respectivamente, con $p+q = r$.

Demostración. Según el teorema V.5.3, existe una matriz ortogonal $P \in \mathcal{M}_n(\mathbb{R})$ tal que $P^{\mathrm{t}}AP = D = (d_{ij}) \in \mathcal{M}_n(\mathbb{R})$ es diagonal. Sea $R = (r_{ij}) \in \mathcal{M}_n(\mathbb{R})$ la matriz diagonal tal que

$$r_{ii} = \begin{cases} \frac{1}{\sqrt{|d_{ii}|}}, & \text{si} \quad d_{ii} \neq 0; \\ 1 & \text{si} \quad d_{ii} = 0 \end{cases}, i = 1, \ldots, n.$$

Tomando $Q = PR$, y ordenando debidamente las entradas de la diagonal de $Q^{\mathrm{t}}AQ$, se obtiene el resultado buscado, ∎

Matrices simétricas (semi)definidas positivas.

Definición V.5.6. Diremos que $A \in \mathcal{M}_n(\mathbb{R})$ es **semidefinida positiva**, si $\mathbf{v}^{\mathrm{t}} A\mathbf{v} \geq 0$, para todo $\mathbf{v} \in \mathbb{R}^n$. Si además, $\mathbf{v}^{\mathrm{t}} A\mathbf{v} > 0$, para todo $\mathbf{v} \in \mathbb{R}^n$ no nulo, diremos que A es **definida positiva**.

Obsérvese que la definición de matriz (semi)definida positiva es consistente con la definición de forma bilineal (semi)definida positiva. En efecto, $T_2 : V \times V \to \mathbb{R}$ *es una forma bilineal (semi)definida positiva si, y sólo si, la matriz de* T_2 *respecto de cualquier base de* V *es (semi)definida positiva* (compruébese).

Proposición V.5.7. *Sea* $A \in \mathcal{M}_n(\mathbb{R})$. *Si* A *es semidefinida positiva, entonces todos sus autovalores reales son no negativos. Si* A *es definida positiva, entonces todos sus autovalores reales son positivos.*

Demostración. Sean $\lambda \in \mathbb{R}$ un autovalor de A y $\mathbf{v} \in \mathbb{R}^n$ un autovalor de A asociado a λ. Entonces,

$$\mathbf{v}^{\mathrm{t}} A\mathbf{v} = \mathbf{v}^{\mathrm{t}}(A\mathbf{v}) = \mathbf{v}^{\mathrm{t}}(\lambda\mathbf{v}) = \lambda(\mathbf{v}^{\mathrm{t}}\mathbf{v}) = \lambda\|\mathbf{v}\|^2;$$

de donde se sigue que $\lambda \geq 0$ si A es semidefinida positiva y $\lambda > 0$ si A definida positiva. ∎

También se puede definir los conceptos de matriz semidefinida y definida negativa de la forma obvia. No obstante, nosotros solamente consideraremos matrices semidefinidas y definidas positivas; de hecho solo nos van a interesar la **matrices simétricas (semi)definidas positivas** y sus propiedades.

Corolario V.5.8. *Sea* $A \in \mathcal{M}_n(\mathbb{R})$ *una matriz simétrica.* A *es semidefinida positiva si, y sólo si, todos sus autovalores son no negativos.* A *es definida positiva si, y sólo si, todos sus autovalores son positivos.*

Demostración. Como A es simétrica, por el teorema V.5.3, existe una matriz $P \in \mathcal{M}_n(\mathbb{R})$ ortogonal tal que $P^{\mathrm{t}}AP$ es diagonal; en particular, tiene todos sus autovalores en $\mathbb{R}$; luego, la proposición V.5.7 permite concluir que todos los autovalores de A son no negativos, si A es semidefinida positiva, y positivos, si A es definida positiva.

Recíprocamente, sea $\mathbf{v} = (v_1, \ldots, v_n)^{\mathrm{t}} \in \mathbb{R}^n$. Como P es invertible, existe un único $\mathbf{w} = (w_1, \ldots, w_n)^{\mathrm{t}} \in \mathbb{R}^n$ tal que $P\mathbf{w} = \mathbf{v}$. Luego,

$$\mathbf{v}^{\mathrm{t}} A\mathbf{v} = (P\mathbf{w})^{\mathrm{t}} A(P\mathbf{w}) = \mathbf{w}^{\mathrm{t}}(P^{\mathrm{t}}AP)\mathbf{w} = \sum_{i=1}^{n} \lambda_i w_i^2,$$

donde $\lambda_i,\ i = 1, \ldots, n$, son los autovalores (posiblemente repetidos) de A. Por consiguiente, $\mathbf{v}^{\mathrm{t}} A\mathbf{v}$ es no negativo si $\lambda_i \geq 0,\ i = 1, \ldots, n$ y positivo si $\lambda_i > 0,\ i = 1, \ldots, n$. ∎

Corolario V.5.9. *Sea $A \in \mathcal{M}_n(\mathbb{R})$ simétrica. Si A es semidefinida positiva, entonces existe una matriz simétrica $A^{1/2}$ tal que $A = A^{1/2}A^{1/2}$. Si A es definida positiva existe una matriz $A^{-1/2}$ tal que $A^{-1} = A^{-1/2}A^{-1/2}$.*

Demostración. Según el teorema V.5.3, existe una matriz ortogonal $P \in \mathcal{M}_n(\mathbb{R})$ tal que $P^{\mathrm{t}}AP = D = (d_{ij}) \in \mathcal{M}_n(\mathbb{R})$ es diagonal; además, por el corolario V.5.8, todas las entradas de la diagonal de D son no negativos.

Sea $R = (r_{ij}) \in \mathcal{M}_n(\mathbb{R})$ la matriz diagonal tal que

$$r_{ii} = \begin{cases} \sqrt{d_{ii}}, & \text{si} \quad d_{ii} \neq 0; \\ 0 & \text{si} \quad d_{ii} = 0 \end{cases}, i = 1, \ldots, n.$$

Tomando $A^{1/2} = PRP^{\mathrm{t}}$ se obtiene el resultado buscado. En efecto,

$$A^{1/2}A^{1/2} = (PRP^{\mathrm{t}})(PRP^{\mathrm{t}}) = PR^2P^{\mathrm{t}} = PDP^{\mathrm{t}} = A.$$

Finalmente, si A es definida positiva, entonces, por el corolario V.5.8, todas las entradas de la diagonal de D son no positivos, por lo que R es invertible. Tomando $A^{-1/2} = PR^{-1}P^{\mathrm{t}}$ se obtiene el resultado buscado. En efecto,

$$A^{-1/2}A^{-1/2} = (PR^{-1}P^{\mathrm{t}})(PR^{-1}P^{\mathrm{t}}) = P(R^2)^{-1}P^{\mathrm{t}} = PD^{-1}P^{\mathrm{t}} = A^{-1}.$$

■

Corolario V.5.10. *Sea $A \in \mathcal{M}_n(\mathbb{R})$. Si A es simétrica y semidefinida positiva, existe $Q \in \mathcal{M}_n(\mathbb{R})$ tal que $A = QQ^{\mathrm{t}}$.*

Demostración. Según el teorema V.5.3, existe una matriz ortogonal $P \in \mathcal{M}_n(\mathbb{R})$ tal que $P^{\mathrm{t}}AP = D = (d_{ij}) \in \mathcal{M}_n(\mathbb{R})$ es diagonal; además, por el corolario V.5.8, todos las entradas de la diagonal de D son no negativos.

Sea $R = (r_{ij}) \in \mathcal{M}_n(\mathbb{R})$ la matriz diagonal tal que

$$r_{ii} = \begin{cases} \sqrt{d_{ii}}, & \text{si} \quad d_{ii} \neq 0; \\ 0 & \text{si} \quad d_{ii} = 0 \end{cases}, i = 1, \ldots, n.$$

Tomando $Q = PRP$ se obtiene el resultado buscado; en efecto,

$$QQ^{\mathrm{t}} = (PRP)(PRP)^{\mathrm{t}} = PRPP^{\mathrm{t}}RP^{\mathrm{t}} = PR^2P^{\mathrm{t}} = PDP^{\mathrm{t}} = A.$$

■

Nota V.5.11. A menudo, el corolario anterior se suele redactar en los siguientes términos: *sea $A \in \mathcal{M}_n(\mathbb{R})$. Si A es simétrica, semidefinida positiva y tiene rango r, existe $Q \in \mathcal{M}_{r\times n}(\mathbb{R})$ tal que $A = QQ^{\mathrm{t}}$*. Lo cual se demuestra exactamente igual que antes tomando $R = (r_{ij}) \in \mathcal{M}_{r\times n}(\mathbb{R})$ tal que $r_{ii} = \sqrt{d_{ii}}, i = 1, \ldots, r$, y $r_{ij} = 0$, si $i \neq j$.

Corolario V.5.12. *Sea $A \in \mathcal{M}_n(\mathbb{R})$. Si A es simétrica y definida positiva, existe una única matriz $Q \in \mathcal{M}_n(\mathbb{R})$ triangular inferior tal que*

$$A = QQ^{\mathrm{t}}.$$

Esta descomposición se conoce como ***factorización de Cholesky*** *de A.*

Demostración. Por el corolario V.5.10, sabemos que existe $B \in \mathcal{M}_n(\mathbb{R})$ tal que $A = BB^{\mathrm{t}}$. Además, como A es simétrica y definida positiva, es invertible; por lo que B también es invertible. Luego, las filas de B son linealmente independientes.

Para cada matriz ortogonal $P \in \mathcal{M}_n(\mathbb{R})$ se tiene que $A = (BP)(BP)^{\mathrm{t}}$. Luego, basta probar que, para cada $B \in \mathcal{M}_n(\mathbb{R})$ existe P ortogonal tal que BP es triangular inferior. Si $\mathbf{b}_1, \ldots, \mathbf{b}_n \in \mathcal{M}_{1\times n}(\mathbb{R})$ son las filas de B, construimos P de tal manera que sus columnas $\mathbf{p}_1, \ldots, \mathbf{p}_n \in \mathbb{R}^n$ sean de norma 1 y satisfagan que

$$\mathbf{p}_n \in \langle \mathbf{b}_1^{\mathrm{t}}, \ldots, \mathbf{b}_{n-1}^{\mathrm{t}} \rangle^{\perp}$$

y

$$\mathbf{p}_{n-i} \in \langle \mathbf{b}_1^{\mathrm{t}}, \ldots, \mathbf{b}_{n-i-1}^{\mathrm{t}}, \mathbf{p}_{n-i+1}, \ldots, \mathbf{p}_n \rangle^{\perp},\ i = 1, \ldots, n-1.$$

Obsérvese que P está unívocamente determinada y puede comprobarse fácilmente que P es ortogonal y que BP es triangular inferior. ∎

Terminamos esta sección mostrando otra condición necesaria y suficiente para que una matriz simétrica sea (semi)definida positiva.

Proposición V.5.13. *Sea $A \in \mathcal{M}_n(\mathbb{R})$ simétrica. A es semidefinida positiva si, y sólo si, todos sus menores principales son no negativos. A es definida positiva si, y sólo si, todos sus menores principales son positivos.*

Demostración. Sea

$$A_i = \begin{pmatrix} a_{11} & \ldots & a_{1i} \\ \vdots & & \vdots \\ a_{i1} & \ldots & a_{ii} \end{pmatrix} \in \mathcal{M}_i(\mathbb{R}),$$

es decir, A_i es la submatriz de A que se obtiene al eliminar las últimas $n-i$ filas y columnas. Por ser A_i una matriz simétrica, existe una matriz ortogonal $P \in \mathcal{M}_i(\mathbb{R})$ tal que

$$P^{\mathrm{t}} A_i P = \begin{pmatrix} \lambda_1 & \ldots & 0 \\ \vdots & & \vdots \\ 0 & \ldots & \lambda_i \end{pmatrix}.$$

Si $\mathbf{u}_j = \begin{pmatrix} \mathbf{p}_j \\ \mathbf{0} \end{pmatrix} \in \mathbb{R}^n$, donde $\mathbf{p}_j$ denota a la columna j-ésima de P, entonces $\lambda_j = \mathbf{u}_j^{\mathrm{t}} A \mathbf{u}_j \geq 0,\ j = 1, \ldots, n$; de donde se sigue que $|A_i| = \lambda_1 \cdots \lambda_i$ es no negativo si A es semidefinida positiva y es positivo si A es definida positiva.

Para probar la implicación contraría procederemos por inducción en n. Para $n = 1$, el resultado es evidentemente cierto. Sea $n > 1$ y supongamos que el resultado es cierto para toda matriz simétrica de orden menor que $n-1$ cuyos menores principales sean no negativos o positivos.

Sea $A_{n-1} \in \mathcal{M}_{n-1}(\mathbb{R})$ la matriz obtenida eliminando la última fila y la última columna de A. Como A_{n-1} es definida positiva, por hipótesis de inducción, sabemos que sus autovalores $\lambda_1, \ldots, \lambda_{n-1}$ son todos estrictamente positivos. Sean $P \in \mathcal{M}_{n-1}(\mathbb{R})$ una matriz ortogonal tal que $P^{\mathrm{t}} A_{n-1} P$ es diagonal y

$$\mathbf{u}_j = \begin{pmatrix} \mathbf{p}_j \\ 0 \end{pmatrix} \in \mathbb{R}^n, \ j = 1, \ldots, n-1,$$

donde $\mathbf{p}_j$ denota a la j-ésima columna de P; es claro que $\{\mathbf{u}_1, \ldots, \mathbf{u}_{n-1}\}$ es una base ortonormal de $\langle \mathbf{e}_1, \ldots, \mathbf{e}_{n-1} \rangle$, siendo $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ la base usual de $\mathbb{R}^n$.

Consideremos el vector

$$\mathbf{u}_n = \mathbf{e}_n - \sum_{i=1}^{n-1} \frac{\mathbf{e}_n^{\mathrm{t}} A \mathbf{u}_i}{\lambda_i} \mathbf{u}_i.$$

Por ser

$$\mathbf{u}_n^{\mathrm{t}} A \mathbf{u}_i = \mathbf{e}_n^{\mathrm{t}} A \mathbf{u}_i - \frac{\mathbf{e}_n^{\mathrm{t}} A \mathbf{u}_i}{\lambda_i} \lambda_i = 0,$$

tenemos que si Q es la matriz del cambio de la base $\{\mathbf{u}_1, \ldots, \mathbf{u}_{n-1}, \mathbf{u}_n\}$ a la base usual de $\mathbb{R}^n$,

$$Q^{\mathrm{t}} A Q = \begin{pmatrix} \lambda_1 & \ldots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \ldots & \lambda_{n-1} & 0 \\ 0 & 0 & \ldots & \mathbf{u}_n^{\mathrm{t}} A \mathbf{u}_n \end{pmatrix}.$$

De donde se sigue que

$$|D| = \lambda_1 \cdot \ldots \cdot \lambda_{n-1} \cdot (\mathbf{u}_n^{\mathrm{t}} A \mathbf{u}_n) = |Q|^2 |A| \geq 0 \ (> 0, \text{ respectivamente}),$$

luego, $\mathbf{u}_n^{\mathrm{t}} A \mathbf{u}_n \geq 0$ (> 0, respectivamente). Finalmente, si $\mathbf{v} = \sum_{j=1}^{n-1} \alpha_j \mathbf{u}_j + \alpha_n \mathbf{u}_n$, entonces

$$\mathbf{v}^{\mathrm{t}} A \mathbf{v} = \sum_{j=1}^{n-1} \lambda_j \alpha_j^2 + \alpha_n^2 (\mathbf{u}_n^{\mathrm{t}} A \mathbf{u}) \geq 0 \ (> 0, \text{ respectivamente}),$$

es decir, A es semidefinida positiva (definida positiva, respectivamente). ■

Matrices hermíticas.

El concepto análogo a matriz simétrica para las matrices con coeficientes complejos es el de matriz hermítica. Veamos a continuación una serie de resultados sobre diagonalización de matrices hermíticas, normales y unitarias. La mayoría de las demostraciones de estos resultados son similares o consecuencias directas de las realizadas con anterioridad, por lo se proponen como ejercicio al lector; no obstante, hemos preferido añadir referencias de las mismas para facilitar la tarea si fuese necesario.

Es conveniente advertir que en el espacio vectorial $\mathbb{C}^n$ también podemos definir un "producto escalar usual": la aplicación bilineal

$$\mathbb{C}^n \times \mathbb{C}^n \longrightarrow \mathbb{C};\ (\mathbf{u}, \mathbf{v}) \mapsto \mathbf{u}^*\mathbf{v}$$

es simétrica y definida positiva (compruébese). También se comprueba fácilmente que el método de Gram-Schmidt tiene perfecto sentido en $\mathbb{C}^n$, donde se deduce la existencia de bases ortonormales y la factorización QR de matrices complejas invertibles, sólo que ahora Q es unitaria en vez de ortogonal (véase el ejercicio 2.12 de [**IR99**] p. 94).

Proposición V.5.14. *Sea $A \in \mathcal{M}_n(\mathbb{C})$.*

(a) *Si A es hermítica, entonces todos sus autovalores son reales.*

(b) *Si A es unitaria, entonces $|\lambda| = 1$, para todo autovalor λ de A.*

Demostración. Proposición 2.5 de [**IR99**] p. 61. ∎

Teorema V.5.15. *Sea $A \in \mathcal{M}_n(\mathbb{C})$.*

(a) *Existe una matriz $Q \in \mathcal{M}_n(\mathbb{C})$ unitaria tal que $Q^*AQ = T$ es triangular*[5]

(b) *A es normal si, y sólo si, existe Q unitaria tal que Q^*AQ es diagonal.*

Demostración. (a) Como $A \in \mathcal{M}_n(\mathbb{C})$, sabemos que su forma canónica de Jordan, J, es una matriz triangular superior. Sea $P \in \mathcal{M}_n(\mathbb{C})$ tal que $P^{-1}AP = J$. Por otra parte, como P es invertible existen Q unitaria y R triangular superior e invertible tales que $P = QR$. Combinando ambas igualdades se sigue que

$$J = P^{-1}AP = (QR)^{-1}A(QR) = R^{-1}Q^*AQR,$$

y por consiguiente que $T = Q^*AQ = RJR^{-1}$, que es triangular superior.

En realidad no es imprescindible usar la forma canónica de Jordan para demostrar este apartado: véanse la sección la sección 6.4 de [**BCR07**] o la demostración del Teorema 2.1 de [**IR99**] p. 62 donde también se demuestra (b) que nosotros proponemos como ejercicio. ∎

[5]La descomposición $A = QTQ^*$ se conoce como **factorización de Schur** de A.

Definición V.5.16. Una matriz hermítica $A \in \mathcal{M}_n(\mathbb{C})$ es

(a) **definida positiva** si $\mathbf{v}^* A\mathbf{v} > 0$, para todo $\mathbf{v} \in V \setminus \{0\}$.

(b) **semidefinida positiva** si $\mathbf{v}^* A\mathbf{v} \geq 0$, para todo $\mathbf{v} \in V$.

Proposición V.5.17. *Si $A \in \mathcal{M}_n(\mathbb{C})$ es una matriz hermítica, se verifica:*

(a) *A es definida positiva si, y sólo si, todos sus autovalores son reales positivos*

(b) *A es semidefinida positiva si, y sólo si, son reales no negativos.*

Demostración. Proposición 2.7 de [**IR99**] p. 66. ∎

Proposición V.5.18. *Dada una matriz $A \in \mathcal{M}_n(\mathbb{C})$ se verifica que A^*A es una matriz hemítica y semidefinida positiva. Además, cuando A es invertible la matriz A^*A es, de hecho, definida positiva.*

Demostración. Proposición 2.8 de [**IR99**] p. 67. ∎

6. Formas cuadráticas

Definición V.6.1. Una **forma cuadrática** en V es una aplicación

$$q : V \to \mathbb{R} \quad \text{tal que} \quad q(\mathbf{x}) = \sum_{i,j=1}^{n} a_{ij} x_i x_j,$$

donde $a_{ij} \in \mathbb{R}$, $i, j \in \{1, \ldots, n\}$ y $(x_1, \ldots, x_n)$ son las coordenadas de $\mathbf{x} \in \mathbb{R}^n$ respecto de un base $\mathcal{B}$ de V.

Obsérvese que una forma cuadrática sobre V no es más que un polinomio homogéneo de grado 2 en n variables con coeficientes en $\mathbb{R}$.

Sea $\mathcal{B}$ y $\mathcal{B}'$ bases de V. Si $A = (a_{ij}) \in \mathcal{M}_n(\mathbb{R})$, la forma cuadrática $q(\mathbf{x}) = \sum_{i,j=1}^{n} a_{ij} x_i x_j$ se escribe

$$q(\mathbf{x}) = q(x_1, \ldots, x_n) = (x_1, \ldots, x_n) A \begin{pmatrix} x_1 \\ \ldots \\ x_n \end{pmatrix},$$

donde $(x_1, \ldots, x_n)$ son las coordenadas de $\mathbf{x} \in \mathbb{R}^n$ respecto de $\mathcal{B}$. Por otra parte, si $\mathcal{B}'$ es otra base de V y $(x'_1, \ldots, x'_n)$ son las coordenadas de $\mathbf{x}$ respecto de $\mathcal{B}'$, entonces

$$q(\mathbf{x}) = q(x_1, \ldots, x_n) = (x'_1, \ldots, x'_n) P^{\mathrm{t}} A P \begin{pmatrix} x'_1 \\ \ldots \\ x'_n \end{pmatrix},$$

donde $P \in \mathcal{M}_n(\mathbb{R})$ es la matriz del cambio de la base $\mathcal{B}'$ a la base $\mathcal{B}$.

Observemos que la matriz de una forma cuadrática q de V no es única.

Ejemplo V.6.2. Sean $V = \mathbb{R}^3$ y $\mathcal{B}$ su base usual. La forma cuadrática

$$q(x_1, x_2, x_3) = x_1^2 + 3x_1x_2 + 6x_2^2 - x_2x_1 + x_2x_3 + x_3^2 + 3x_3x_2$$

se puede escribir

$$(x_1, x_2, x_3) \begin{pmatrix} 1 & 3 & 0 \\ -1 & 6 & 1 \\ 0 & 3 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

Como

$$\begin{aligned} q(x_1, x_2, x_3) &= x_1^2 + 3x_1x_2 + 6x_2^2 - x_2x_1 + x_2x_3 + x_3^2 + 3x_3x_2 \\ &= x_1^2 + 2x_1x_2 + 6x_2^2 + 4x_2x_3 + x_3^2, \end{aligned}$$

también se puede escribir

$$q(x_1, x_2, x_3) = (x_1, x_2, x_3) \begin{pmatrix} 1 & 2 & 0 \\ 0 & 6 & 4 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix};$$

o también,

$$q(x_1, x_2, x_3) = (x_1, x_2, x_3) \begin{pmatrix} 1 & 1 & 0 \\ 1 & 6 & 2 \\ 0 & 2 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

Proposición V.6.3. *Sean q una forma cuadrática de V y $\mathcal{B}$ una base de V. Existe una única matriz simétrica S tal que*

$$q(\mathbf{x}) = (x_1, \ldots, x_n) S \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix},$$

donde $(x_1, \ldots, x_n)$ son las coordenadas de $\mathbf{x} \in V$ respecto de $\mathcal{B}$; es decir, existe una matriz simétrica asociada a q respecto de $\mathcal{B}$.

Demostración. Sea $A \in \mathcal{M}_n(\mathbb{R})$ una de las matrices de q respecto de $\mathcal{B}$. Sabemos que A puede escribirse, de forma única, como la suma de una matriz simétrica y otra antisimétrica (ejercicio 4):

$$A = \frac{1}{2}(A + A^{\mathrm{t}}) + \frac{1}{2}(A - A^{\mathrm{t}}).$$

Por otra parte, si $H \in \mathcal{M}_n(\mathbb{R})$ es antisimétrica, entonces

$$(x_1,\dots,x_n)H\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \left((x_1,\dots,x_n)H\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}\right)^{\mathrm{t}}$$
$$= (x_1,\dots,x_n)H^{\mathrm{t}}\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = -(x_1,\dots,x_n)H\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix},$$

luego

$$(x_1,\dots,x_n)H\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = 0,$$

donde $(x_1,\dots,x_n)$ son las coordenadas respecto de $\mathcal{B}$ de $\mathbf{x} \in \mathbb{R}^n$. Por consiguiente, si $S = \frac{1}{2}(A + A^{\mathrm{t}})$, entonces

$$q(\mathbf{x}) = (x_1,\dots,x_n)A\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = (x_1,\dots,x_n)S\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix},$$

donde $(x_1,\dots,x_n)$ son las coordenadas respecto de $\mathcal{B}$ de $\mathbf{x} \in \mathbb{R}^n$.

La unicidad de S se sigue de la unicidad de la descomposición de A como suma de una matriz simétrica y otra antisimétrica. ∎

Definición V.6.4. Sea $\mathcal{B}$ una base de V. Llamaremos **matriz de la forma cuadrática** q **de** V **respecto de** $\mathcal{B}$ a la única matriz simétrica $S \in \mathcal{M}_n(\mathbb{R})$ tal que

$$q(\mathbf{x}) = (x_1,\dots,x_n)S\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

Recordemos ahora que para cualquier matriz simétrica $A \in \mathcal{M}_n(\mathbb{R})$ existe una matriz ortogonal $P \in \mathcal{M}_n(\mathbb{R})$ tal que $P^{\mathrm{t}}AP = D = (d_{ij}) \in \mathcal{M}_n(\mathbb{R})$ es diagonal. Por tanto, si A es la matriz (simétrica) de la forma cuadrática q respecto de $\mathcal{B}$, entonces existe una base $\mathcal{B}'$ de V, concretamente aquella tal que la matriz del cambio de base de $\mathcal{B}'$ a $\mathcal{B}$ es P, de tal manera que q se puede escribir también como

$$q(\mathbf{x}) = (x_1,\dots,x_n)D\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \sum_{i=1}^{n} d_{ii}x_i^2, \tag{V.6.1}$$

donde $(x_1, \ldots, x_n)$ son las coordenadas de $\mathbf{x} \in V$ respecto de la base $\mathcal{B}'$. La expresión (V.6.1) se conoce como **forma canónica** de q.

Definición V.6.5. Una forma cuadrática q sobre $\mathbb{R}^n$ es semidefinida positiva si $q(\mathbf{x}) \geq 0$, para todo $\mathbf{x} \in \mathbb{R}^n$. Una forma cuadrática q es definida positiva si $q(\mathbf{x}) > 0$, para todo $\mathbf{x} \in \mathbb{R}^n$ no nulo.

De manera análoga se definen las formas cuadráticas definidas negativas y semidefinidas negativas.

Formas cuadráticas y métricas simétricas.

Si $T_2 : V \times V \to \mathbb{R}$ es una forma bilineal simétrica, entonces la aplicación $q : V \to \mathbb{R}$ definida por $q(x) = T_2(x, x)$ es una forma cuadrática. Si A es la matriz de T_2 respecto de $\mathcal{B}$, entonces

$$q(\mathbf{x}) = q(x_1, \ldots, x_n) = (x_1, \ldots, x_n)A \begin{pmatrix} x_1 \\ \ldots \\ x_n \end{pmatrix},$$

donde $(x_1, \ldots, x_n)$ son las coordenadas de $\mathbf{x} \in \mathbb{R}^n$ respecto de $\mathcal{B}$.

Recíprocamente, si $q : V \to \mathbb{R}$ es una forma cuadrática,

$$q(\mathbf{x}) = q(x_1, \ldots, x_n) = (x_1, \ldots, x_n)A \begin{pmatrix} x_1 \\ \ldots \\ x_n \end{pmatrix},$$

donde $(x_1, \ldots, x_n)$ son las coordenadas de $\mathbf{x} \in \mathbb{R}^n$ respecto de $\mathcal{B}$, entonces la aplicación $T_2 : V \times V \to \mathbb{R}$ definida por

$$T_2(\mathbf{x}, \mathbf{y}) = \frac{1}{4}(q(\mathbf{x} + \mathbf{y}) - q(\mathbf{x} - \mathbf{y}))$$

es bilineal y simétrica. A T_2 se le denomina **forma bilineal simétrica asociada a la forma cuadrática** q. Observemos que si $A = (a_{ij}) \in \mathcal{M}_n(\mathbb{R})$ es la matriz simétrica de q respecto de $\mathcal{B}$, entonces A es la matriz de T_2 respecto de $\mathcal{B}$.

Es inmediato comprobar que las anteriores correspondencias establecen una biyección (de hecho, un isomorfismo lineal) entre el espacio de las formas cuadráticas de V y el de las forma bilineales simétricas sobre V.

Ejercicios del tema V

Ejercicio 1. Sobre $\mathbb{R}^3$ consideramos una forma bilineal $T_2 : \mathbb{R}^3 \times \mathbb{R}^3 \longrightarrow \mathbb{R}$ cuya matriz asociada respecto de la base usual de $\mathbb{R}$ es $A \in \mathcal{M}_3(\mathbb{R})$. Determinar si T_2 es simétrica cuando A es:

$$(a)\ \begin{pmatrix} -1 & 2 & 3 \\ 2 & 4 & 1 \\ 3 & 1 & 5 \end{pmatrix}, \quad (b)\ \begin{pmatrix} 1 & 1 & -1 \\ 1 & 2 & 1 \\ -1 & 1 & 6 \end{pmatrix}, \quad (c)\ \begin{pmatrix} 1 & 2 & 3 \\ 2 & 5 & 6 \\ 1 & 1 & 4 \end{pmatrix}.$$

Ejercicio 2. Comprobar la fórmula del cambio de base en el ejemplo V.1.6 para $n = 3$.

Ejercicio 3. Hallar la matriz respecto de la base usual $\mathbb{R}^3$ de la métrica simétrica $T_2 : \mathbb{R}^3 \times \mathbb{R}^3 \longrightarrow \mathbb{R}$ definida por

$$\begin{array}{lllllll} T_2(\mathbf{u}_1, \mathbf{u}_1) & = & 5; & T_2(\mathbf{u}_1, \mathbf{u}_2) & = & 0; & T_2(\mathbf{u}_1, \mathbf{u}_3) = -1; \\ & & & T_2(\mathbf{u}_2, \mathbf{u}_2) & = & 1; & T_2(\mathbf{u}_2, \mathbf{u}_3) = 4; \\ & & & & & & T_2(\mathbf{u}_3, \mathbf{u}_3) = 0; \end{array}$$

donde $\mathbf{u}_1 = (1, 2, 1)$, $\mathbf{u}_2 = (-1, 2, 0)$ y $\mathbf{u}_3 = (1, 0, 1)$.

Ejercicio 4. Sean V un $\mathbb{R}$-espacio vectorial de dimensión $n > 0$ y T_2 una forma bilineal sobre V. Probar que si $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ es una base de V tal que $T_2(\mathbf{v}_i, \mathbf{v}_j) = \delta_{ij}$, donde δ_{ij} es la función Delta de Kronecker, entonces T_2 es un producto escalar sobre V.

Ejercicio 5. Sean V un espacio vectorial euclídeo y $L \subseteq V$ un subespacio de V. Probar que la restricción del producto escalar de V a L dota a este último de estructura de espacio vectorial euclídeo; es decir, *todo subespacio vectorial de un espacio vectorial euclídeo hereda una estructura natural de espacio vectorial euclídeo.*

Ejercicio 6. Aplicar el método de Gram-Schmidt para calcular bases ortonormales, a partir de la bases dadas en los siguientes espacios vectoriales euclídeos:

1. $\{(1, 1, 1), (0, 1 - 1)\,(0, 2, 0)\}$ en $\mathbb{R}^3$, con el producto escalar usual.
2. $\{1, x, x^2\}$ en el espacio V de los polinomios de $\mathbb{R}[x]$ con grado menor o igual que 2, y el producto escalar $T_2(P, Q) = P(0)Q(0) + P(1)Q(1) + P(2)Q(2)$.
3. $\{1, x, x^2\}$ en el espacio V de los polinomios de $\mathbb{R}[x]$ con grado menor o igual que 2, y el producto escalar $\bar{T}_2(P, Q) = \int_0^1 P(x)Q(x)dx$.

Ejercicio 7. En el $\mathbb{R}$-espacio vectorial $\mathbb{R}^2$ consideramos la aplicación

$$\begin{array}{rcl} T_2 : \mathbb{R}^2 \times \mathbb{R}^2 & \longrightarrow & \mathbb{R} \\ ((x_1, y_1), (x_2, y_2)) & \mapsto & T_2((x_1, y_1), (x_2, y_2)) = x_1x_2 + x_1y_2 + x_2y_1 + 2y_1y_2. \end{array}$$

1. Probar que T_2 es un producto escalar.

2. Obtener una base ortonormal del espacio vectorial euclídeo $(\mathbb{R}^2, T_2)$.

Ejercicio 8. Sean V un $\mathbb{R}$-espacio vectorial de dimensión $3, \mathcal{B}$ una base de V y T_2 la forma bilineal sobre V cuya matriz respecto de $\mathcal{B}$ es

$$\begin{pmatrix} 2 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 4 & 6 \end{pmatrix}.$$

Hallar una base de V respecto de la cual la matriz de T_2 sea diagonal.

Ejercicio 9. Sobre $\mathbb{R}^3$ se considera la forma bilineal T_2 cuya matriz en la base usual es

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

1. Probar que T_2 es un producto escalar.
2. Hallar una base ortonormal del espacio vectorial euclídeo $(\mathbb{R}^3, T_2)$.
3. Calcular el módulo de $\mathbf{v} = (1, 3, -2)$ y el ángulo que forman los vectores $\mathbf{u}_1 = (1, -2 - 2)$ y $\mathbf{u}_2 = (2, 1, 0)$ para el producto escalar T_2.

Ejercicio 10. Sobre $\mathbb{R}^3$ consideramos la forma bilineal T_2 cuya matriz en la base $\mathcal{B}$ usual es

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

1. Probar que T_2 es producto escalar.
2. Hallar una base ortonormal del espacio vectorial euclídeo $(\mathbb{R}^3, T_2)$.
3. Calcular el módulo del vector $\mathbf{v} \in \mathbb{R}^3$ de coordenadas $(1, 0, 2)$ respecto de $\mathcal{B}$. Calcular el ángulo que forman el vector $\mathbf{v}$ con el vector $\mathbf{u}$ de coordenadas $(1, 0, 0)$ respecto de $\mathcal{B}$.

Ejercicio 11. Sea

$$A = \begin{pmatrix} 4 & -4 & 4 \\ -4 & 9 & -4 \\ 4 & -4 & 10 \end{pmatrix}$$

la matriz respecto de la base usual de $\mathbb{R}^3$ de un producto escalar que dota de estructura de espacio vectorial euclídeo a $\mathbb{R}^3$.

1. Encontrar una base de $\mathbb{R}^3$ respecto de la cual la matriz del producto escalar sea diagonal.
2. Hallar una base ortonormal $\mathbb{R}^3$.
3. Usar el apartado anterior para calcular A^{-1}.

Ejercicio 12. Sea V el espacio vectorial de las matrices simétricas reales de orden dos.

1. Hallar una base de V.
2. Probar que la aplicación

$$\begin{array}{rcl} V \times V & \longrightarrow & \mathbb{R} \\ (A, B) & \longmapsto & A \cdot B = \operatorname{tr}(AB) \end{array}$$

   es un producto escalar sobre V y obtener su matriz en la base hallada en el apartado (a).
3. Calcular una base ortonormal de V para el producto escalar anterior.

Ejercicio 13. Consideramos el espacio vectorial euclídeo $\mathbb{R}_2[x]$ de los polinomios de grado de menor o igual que 2 con el producto escalar

$$\begin{array}{rcl} \mathbb{R}_2[x] \times \mathbb{R}_2[x] & \longrightarrow & \mathbb{R} \\ (p(x), q(x)) & \longmapsto & p(x) \cdot q(x) = p(0)q(0) + p(1)q(1) + p(2)q(2). \end{array}$$

1. Calcular la matriz del producto escalar en respecto de la base $\mathcal{B} = \{1, x, x^2\}$.
2. Calcular los módulos de los vectores de la base $\mathcal{B}$, así como los ángulos que forman dichos vectores entre sí.
3. Hallar una base ortonormal de $\mathbb{R}_2[x]$.

Ejercicio 14. Consideremos en $\mathbb{R}^3$ el producto escalar T_2 que en la base $B = \{\mathbf{v}_1 = (1, 1, 0), \mathbf{v}_2 = (1, 0, 1), \mathbf{v}_3 = (0, 1, 1)\}$ tiene matriz

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

1. Calcular una base ortonormal para T_2.
2. Escribir la matriz de T_2 en la base usual de $\mathbb{R}^3$.
3. Calcular las ecuaciones, en la base B, del subespacio ortogonal al plano π que en la base usual tiene ecuación $z = 0$.
4. Calcular la distancia de $\mathbf{v}_1$ a π.

Ejercicio 15. Consideremos en $\mathbb{R}^4$ el producto escalar euclídeo

$$T_2(x, y) = 4x_1y_1 + x_1y_2 + x_2y_1 + 2x_2y_2 + x_3y_3.$$

Calcular la proyección ortogonal del vector $\mathbf{v} = (0, 1, 0)$ sobre el subespacio $L = \langle (1, 0, 0), (0, 0, 1) \rangle$ y determinar la distancia de $\mathbf{v}$ a L.

Ejercicio 16. Sean V un $\mathbb{R}$-espacio vectorial de dimensión 4, $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$ una base de V. Consideramos el producto escalar definido por

$$\begin{array}{llllllll} \mathbf{v}_1 \cdot \mathbf{v}_1 = 7; & \mathbf{v}_1 \cdot \mathbf{v}_2 = 3; & \mathbf{v}_1 \cdot \mathbf{v}_3 = 3; & \mathbf{v}_1 \cdot \mathbf{v}_4 = -1; \\ & \mathbf{v}_2 \cdot \mathbf{v}_2 = 2; & \mathbf{v}_2 \cdot \mathbf{v}_3 = 1; & \mathbf{v}_2 \cdot \mathbf{v}_4 = 0; \\ & & \mathbf{v}_3 \cdot \mathbf{v}_3 = 2; & \mathbf{v}_3 \cdot \mathbf{v}_4 = -1; \\ & & & \mathbf{v}_4 \cdot \mathbf{v}_4 = 1, \end{array}$$

que dota a V de estructura de espacio vectorial euclídeo. Dado el subespacio L de V generado por los vectores $\mathbf{u}_1 = \mathbf{v}_2 + \mathbf{v}_4$, $\mathbf{u}_2 = 2\mathbf{v}_1 + \mathbf{v}_2 + \mathbf{v}_3$ y $\mathbf{u}_3 = \mathbf{v}_3 - 2\mathbf{v}_4 - \mathbf{v}_5$, obtener una base de $L^\perp$.

Ejercicio 17. Sean $V = \mathbb{R}^4$, $\mathcal{B}$ la base usual de $\mathbb{R}^4$ y T_2 la forma bilineal simétrica cuya matriz respecto de $\mathcal{B}$ es

$$\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}.$$

Si L es un subespacio de V, definimos

$$L^\perp = \{\mathbf{v} \in V \mid T_2(\mathbf{v}, \mathbf{u}) = 0, \ \forall\, \mathbf{u} \in L\}.$$

1. Probar $L^\perp$ es un subespacio de V.
2. Hallar una base de $V^\perp$.
3. Sea L el subespacio de V definido por $\{(x_1, x_2, x_3, x_4) \mid x_1 - x_4 = x_2 - x_3 = 0\}$. Comprobar que $(L^\perp)^\perp \neq L$.
4. ¿Contradicen los apartados anteriores a las propiedades vistas para del subespacio ortogonal de un subespacio de un espacio vectorial euclídeo? Justificar la respuesta.

Ejercicio 18. Sea $\mathcal{B} = \{\mathbf{v}_1 = (1,1,0), \mathbf{v}_2 = (1,0,1), \mathbf{v}_3 = (0,1,1)\}$ una base de $\mathbb{R}^3$. Sobre $\mathbb{R}^3$ consideramos el producto escalar cuya matriz respecto de $\mathcal{B}$ es

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

que lo dota de estructura de espacio vectorial euclídeo.

1. Calcular una base ortonormal de $\mathbb{R}^3$.
2. Calcular la matriz del producto escalar respecto de la base usual de $\mathbb{R}^3$.
3. Dado el subespacio $L = \{(x, y, z) \in \mathbb{R}^3 \mid z = 0\}$, calcular $L^\perp$.

Ejercicio 19. Sobre $V = \mathcal{M}_2(\mathbb{R})$, esto es, el espacio vectorial de las matrices reales de orden 2, se considera el producto escalar dado por la igualdad $A \cdot B := \mathrm{tr}(A^{\mathrm{t}}B)$, para cada A y $B \in \mathcal{M}_2(\mathbb{R})$.

1. Calcular el ortogonal del subespacio L formado por las matrices diagonales de $\mathcal{M}_2(\mathbb{R})$.
2. Determinar la proyección ortogonal de cada matriz $C \in \mathcal{M}_2(\mathbb{R})$ sobre L.

Ejercicio 20. Sean $\mathcal{B} = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ una base ortonormal de un espacio vectorial euclídeo V, L un subespacio de V, $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ una base de L y $A \in \mathcal{M}_{n\times r}(\mathbb{R})$ la matriz cuyas columnas son las coordenadas de $\mathbf{v}_1, \ldots, \mathbf{v}_r$ respecto de $\mathcal{B}$.

1. Probar que la matriz $A^{\mathrm{t}}A$ es invertible.
2. Dado un vector $\mathbf{v} = \lambda_1\mathbf{u}_1 + \cdots + \lambda_n\mathbf{u}_n$, demostrar que las coordenadas de la proyección ortogonal de $\mathbf{v}$ sobre L respecto de $\mathcal{B}$ son
$$A(A^{\mathrm{t}}A)^{-1}A^{\mathrm{t}} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}.$$
3. Aplicar lo anterior para calcular, en $\mathbb{R}^4$ con su producto escalar usual, la proyección ortogonal de $(-1, 2, -3, -1)$ sobre $L = \langle (1, 3, -2, 0), (3, 2, 0, 0)\rangle$.

Ejercicio 21. Dada $A \in \mathcal{M}_n(\mathbb{R})$, consideremos la matriz $B = A^{\mathrm{t}}A$. Probar que $\ker(A) = \ker(B)$ y deducir de ello que $\mathrm{rg}(A) = \mathrm{rg}(B)$.

Ejercicio 22. Sea $A \in \mathcal{M}_n(\mathbb{R})$. Probar que $\mathrm{rg}(A^{\mathrm{t}}\,A) = \mathrm{rg}(A\,A^{\mathrm{t}}) = \mathrm{rg}(A) = \mathrm{rg}(A^{\mathrm{t}})$. Dar un ejemplo de una matriz con coeficientes complejos tal que $\mathrm{rg}(A^{\mathrm{t}}\,A) \neq \mathrm{rg}(A)$.

Ejercicio 23. Probar las siguientes afirmaciones:

1. Si $A \in \mathcal{M}_n(\mathbb{R})$ es simétrica y P es una matriz invertible, entonces A es (semi)definida positiva si, y sólo si, lo es $P^{\mathrm{t}}AP$.
2. Si Si $A \in \mathcal{M}_n(\mathbb{R})$ es simétrica, entonces A es definida positiva si, y sólo si, existe una matriz P invertible tal que $P^{\mathrm{t}}AP = I_n$.
3. Si $A \in \mathcal{M}_n(\mathbb{R})$ es simétrica, entonces A es definida positiva si, y sólo si, existe una matriz Q invertible tal que $A = Q^{\mathrm{t}}Q$.
4. Si $A \in \mathcal{M}_{m\times n}(\mathbb{R})$, las matrices $A^{\mathrm{t}}A$ y AA^{t} son semidefinidas positivas.
5. Si $A \in \mathcal{M}_{m\times n}(\mathbb{R})$, entonces el rango de A es m si, y sólo si, la matriz AA^{t} es definida positiva.
6. Si $A \in \mathcal{M}_{m\times n}(\mathbb{R})$, entonces el rango de A es n si, y sólo si, la matriz $A^{\mathrm{t}}A$ es definida positiva.

7. Si $A \in \mathcal{M}_n(\mathbb{R})$ es simétrica de rango r, entonces existe una matriz $B \in \mathcal{M}_{n\times r}(\mathbb{C})$ de rango r tal que $A = BB^{\mathrm{t}}$. Además, si A es semidefinida positiva, entonces B puede tomarse real.

Ejercicio 24. Consideremos la matriz cuadrada

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix},$$

con A_{11} y A_{22} matrices cuadradas. Probar que si A es simétrica y definida positiva, y la inversa de A es

$$B = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix},$$

entonces $B_{11}^{-1} = A_{11} - A_{12}A_{22}^{-1}A_{21}$.

Ejercicio 25. Aplicar los distintos criterios para determinar si las siguientes formas cuadráticas son definidas positivas (negativas) o semidefinidas positivas o negativas. Escribir también la forma reducida de cada una de ellas.

1. $q_1(x, y, z) = 3x^2 + 16y^2 + 139z^2 + 12xy + 30xz + 92yz$.
2. $q_2(x, y, z) = -4x^2 - 5y^2 - 2z^2 + 4xz$.
3. $q_3(x, y, z) = x^2 + 4y^2 - 4xy$.
4. $q_4(x, y, z, t) = -4x^2 + 4xy - y^2 - 9z^2 + 6zt - t^2$.
5. $q_5(x, y) = xy$.
6. $q_6(x, y, z, t) = 2xt + 2yz$.

Ejercicio 26. Dada la matriz $A = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}$,

1. Escribir una matriz ortogonal P tal que $P^{-1}AP$ sea una matriz diagonal D.
2. Escribir una matriz Q, que pueda expresarse como producto de matrices de transformaciones elementales del tipo T_{ij} y $S_{ij}(\lambda)$, tal que $Q^{\mathrm{t}}AQ$ sea una matriz diagonal $\bar{D}$
3. Escribir, si es posible, una matriz R, que pueda expresarse como producto de matrices de transformaciones elementales, tal que $R^{\mathrm{t}}AR = I_3$.

Sea T_2 la forma bilineal simétrica que, en la base usual de $\mathbb{R}^3$ tiene matriz A y sea q la forma cuadrática asociada a T_2.

4. Comprobar que T_2 es un producto escalar.
5. Las columnas de P forman una base ortonormal para el producto escalar usual de $\mathbb{R}^3$. Comprobar que dichas columnas forman una base ortogonal para T_2.
6. Comprobar que las columnas de Q forman una base ortogonal para T_2 y que las de R forman una base ortonormal para T_2.

7. Escribir la expresión de q en coordenadas para las bases dadas por las columnas de P, de Q y de R.

TEMA VI

Inversas generalizadas. Mínimos cuadrados

La inversa de una matriz está definida para todas las matrices cuadradas que no son singulares, es decir, aquellas que tienen determinante no nulo. Sin embargo, hay muchas situaciones en las que podemos encontrarnos con una matriz rectangular (no cuadrada) o singular, y aún así sea necesario calcular otra matriz que de alguna manera se comporte como una matriz inversa. Una de estas situaciones, que aparece a menudo en Estadística y Probabilidad así como en otros campos de la Matemática Aplicada, es la relacionada con el cálculo de soluciones de sistemas de ecuaciones lineales. Un sistema de ecuaciones lineales se puede escribir matricialmente como

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

con $A \in \mathcal{M}_{m\times n}(\mathbb{R})$ y $\mathbf{b} \in \mathbb{R}^M$, siendo $\mathbf{x} \in \mathbb{R}^n$ el vector que queremos calcular. Si A es cuadrada e invertible, entonces $\mathbf{x} = A^{-1}\mathbf{b}$. Pero ¿qué ocurre cuando A^{-1} no existe? ¿Cómo podemos determinar si el sistema tiene alguna solución, y en este caso, cuántas hay y cómo podemos calcularlas? El teorema de Rouche-Fröbenius responde parcialmente la última pregunta, pues da un criterio para determinar si un sistema es compatible, pero no nos indica cómo calcular las soluciones en caso de existir.

Existen diversas generalizaciones del concepto de matriz inversa. La mayoría de estas generalizaciones surgen al exigir a la *inversa generalizada* o *seudoinversa* G de una matriz dada $A \in \mathcal{M}_{m\times n}(\mathbb{R})$ que cumpla una, dos, tres o cuatro de las siguientes condiciones:

$$\begin{array}{ll} \text{(G1)} & A\,G\,A = A, \\ \text{(G2)} & G\,A\,G = G, \\ \text{(G3)} & A\,G \text{ es simétrica}, \\ \text{(G4)} & G\,A \text{ es simétrica}. \end{array}$$

Al final de este tema veremos que las respuestas a todas la preguntas anteriores se pueden expresar en términos de inversas generalizadas.

En este tema nos centraremos en el estudio de la inversa de Moore-Penrose, que es la cumple las cuatro condiciones, la {1}-inversa, que es la que cumple la primera de las cuatro condiciones y, por último, la inversa mínimo cuadrática, que es la que cumple la primera y la tercera de las cuatro condiciones. La {1}-inversa se aplica

para determinar la compatibilidad de los sistemas de ecuaciones lineales y caracterizar todas las soluciones. La inversa mínima cuadrática resuelve el problema de la aproximación mínimo cuadrática de la solución de un sistema de ecuaciones lineales. Finalmente, veremos que la inversa de Moore-Penrose permite calcular la solución aproximada mínimo cuadrática de norma mínima de un sistema incompatible.

Introduciremos la inversa de Moore-Penrose de una matriz $A \in \mathcal{M}_{m\times n}(\mathbb{R})$ desde una perspectiva álgebro-geométrica y la calcularemos usando la llamada *descomposición en valores singulares* de A. En la práctica 9 veremos otro método para calcularla.

La primera sección del tema está dedicada a la descomposición en valores singulares de una matriz. En primer lugar, se estudian las matrices $A^{\mathrm{t}}A$ y $A\,A^{\mathrm{t}}$ con $A \in \mathcal{M}_{m\times n}(\mathbb{R})$. Ambas matrices son simétricas semidefinidas positivas y tiene el mismo rango que A, por tanto sus autovalores son reales no negativos y el número de autovalores positivos coincide con el rango de A. Estos resultados darán sentido y serán la clave para definir la descomposición en valores singulares de A.

Tras la definición de la descomposición en valores singulares, el resto de la sección se dedica a mostrar sus propiedades. Quizá lo más interesante, por ser un aspecto poco tratado en los libros sobre inversas generalizadas, es la interpretación geométrica que damos al final de la sección

La siguiente sección trata exclusivamente sobre la inversa (generalizada) de Moore-Penrose. Al principio de la sección damos dos definiciones de inversa de Moore-Penrose, y demostramos que son equivalentes. A continuación demostramos que toda matriz tiene inversa de Moore-Penrose y que ésta es única. La demostración de la existencia consiste en comprobar que la inversa de Moore-Penrose es $A^+ := Q\Delta^{-1}P^{\mathrm{t}}$, siendo $P\Delta Q^{\mathrm{t}}$ la descomposición en valores singulares de A, lo que claramente pone de manifiesto la relación entre las dos primeras secciones del tema.

La otra definición de inversa de Moore-Penrose tiene un sabor más geométrico, y la mostramos en el teorema VI.2.4.

A continuación, usando la interpretación geométrica de la descomposición en valores singulares, damos la interpretación geométrica de la inversa de Moore-Penrose. Finalmente, mostramos algunas de las propiedades de la inversa de Moore-Penrose, y con ellas concluimos la sección. Es interesante destacar que si la matriz A tiene inversa a izquierda y/o a derecha, entonces la inversa de Moore-Penrose es una inversa a izquierda y/o derecha; en particular, si A es invertible $A^+ = A^{-1}$.

En la tercera sección nos ocupamos de otras inversas generalizadas. Tal y como se apuntó al principio, la mayoría de las inversas generalizas surgen al exigir que una cierta matriz cumpla una, dos, tres o cuatro de las condiciones (G1)-(G4). En esta sección estudiamos las inversas generalizadas que cumplen (G1) y aquellas que cumplen (G1) y (G3). A las primeras las llamaremos inversas generalizadas a secas, pues todas las inversas que estudiamos en esta asignatura cumplen, al menos, (G1);

a las segundas las llamaremos inversas mínimo, cuyo nombre se justifica en la última sección del tema.

A modo de ejemplo, ilustramos la relación de las inversas generalizadas con la forma reducida estudiada en el tema II. Además, mostramos su expresión general y estudiamos sus propiedades. En concreto, mostramos que si A tiene inversa a izquierda, entonces las inversas generalizadas son inversas a izquierda y lo mismo ocurre cuando A tiene inversa a derecha. Finalmente, damos una expresión general para todas las inversas generalizadas de una matriz a partir de una inversa generalizada dada.

A continuación, se muestran algunas propiedades de las inversas generalizadas de $A^{\mathrm{t}}A$. Estas propiedades son de suma utilidad en la obtención de inversas generalizadas mínimo cuadráticas; concretamente, si $(A^{\mathrm{t}}A)^{-}$ es una inversa generalizada de $A^{\mathrm{t}}A$, entonces $(A^{\mathrm{t}}A)^{-}A$ es una inversa mínimo cuadrática de A. Es interesante resaltar que para cualquier inversa mínimo cuadrática, $A^{\square}$, de A, se cumple que $AA^{\square} = AA^{+}$; luego, podemos definir las inversas mínimo cuadráticas como las matrices B tales que AB es la matriz de la proyección ortogonal sobre la imagen A respecto de la base usual correspondiente.

En la última sección del tema, retomamos los sistemas de ecuaciones lineales, usamos las inversas generalizadas para estudiar su compatibilidad y damos una fórmula que describe todas las soluciones en términos de una inversa generalizada A^{-} de A. Para los sistemas incompatibles recurrimos a las inversas mínimo cuadráticas. En este caso, el sistema de ecuaciones $A\mathbf{x} = \mathbf{b}$ no tiene solución, por lo que buscamos los vectores $\bar{\mathbf{x}}$ tales que $\|A\bar{\mathbf{x}} - \mathbf{b}\|^2$ es mínima. Usando lo estudiado en el tema 5 sobre proyecciones ortogonales, concluimos que los vectores buscados son las soluciones del sistema $A\mathbf{x} = \mathbf{b}_1$, siendo $\mathbf{b}_1$ la proyección ortogonal de $\mathbf{b}$ sobre la imagen de A. Por consiguiente, teniendo en cuenta la relación de las inversas mínimo cuadráticas con la proyección ortogonal, utilizamos estas inversas generalizadas para resolver el problema de una forma similar a como se hizo en el caso compatible.

Como hemos dicho en repetidas ocasiones, este es un tema completamente nuevo y con el que alumno ni siquiera suele estar familiarizado. Sin embargo, este tema tiene multitud de utilidades en Estadística y Probabilidad, véase, por ejemplo, el capítulo 6 de [**Bas83**], el capítulo 5 de [**Sch05**] (que es el que hemos usado principalmente para la elaboración del tema), el capítulo 8 de [**Sea82**], y por supuesto, [**RM71**] que es un libro de referencia básica sobre inversas generalizadas. Por citar un ejemplo de uso de la inversa generalizada en Estadística, retomemos los modelos lineales normales comentados anteriormente; para ello supongamos que estamos en las condiciones del modelo lineal normal, pero en este caso consideramos un sistema de n generadores de L, esto es, una matriz $A \in \mathcal{M}_{m\times n}(\mathbb{R})$ de rango $\dim(L)$. Así, podemos expresar la media μ mediante su vector de coordenadas β respecto de las columnas de A, es decir,

β una solución del sistema de ecuaciones $A\beta = \mu$. El parámetro β se puede expresar en términos de μ como $\beta = A^-\mu$, siendo A^- una inversa generalizada de A. Es más, sabemos cómo son todas las soluciones del sistema $A\beta = \mu$, en términos de A^- y μ. No obstante, en general μ es desconocido, por lo que interesarán las soluciones aproximadas mínimo cuadráticas de $A\beta = \mathbf{y}$, y generalmente la de norma mínima, que según se ve en este tema está complemente determinadas por las inversas mínimo cuadrática y la inversa de Moore-Penrose.

En los capítulos 10 y 12 de [**CnR05**] se pueden encontrar multitud de ejercicios sobre mínimos cuadrados e inversas generalizadas, respectivamente. También en [**MS06**], hay todo un capítulo dedicado a estos temas.

1. Descomposición en valores singulares (SVD)

Comenzamos esta sección estudiando algunas de las **propiedades de $A^{\mathrm{t}}A$ y de $A\,A^{\mathrm{t}}$** con $A \in \mathcal{M}_n(\mathbb{R})$.

Proposición VI.1.1. *Sea $A \in \mathcal{M}_{m\times n}(\mathbb{R})$. Se cumple que:*

(a) $\ker(A) = \ker(A^{\mathrm{t}}A)$ *y* $\ker(A^{\mathrm{t}}) = \ker(A\,A^{\mathrm{t}})$; *luego* $\mathrm{rg}(A) = \mathrm{rg}(A^{\mathrm{t}}A) = \mathrm{rg}(A\,A^{\mathrm{t}})$.

(b) *$A^{\mathrm{t}}A$ y $A\,A^{\mathrm{t}}$ son simétricas y semidefinidas positivas. En particular, $A^{\mathrm{t}}A$ es definida positiva si, y sólo si,* $\mathrm{rg}(A) = n$, *y $A\,A^{\mathrm{t}}$ es definida positiva si, y sólo si,* $\mathrm{rg}(A) = m$.

Demostración. (a) En primer lugar recordamos que $\ker(A) = \{\mathbf{v} \in \mathbb{R}^n \mid A\mathbf{v} = \mathbf{0}\}$; luego es claro que $\ker(A) \subseteq \ker(A^{\mathrm{t}}A)$. Recíprocamente, si $\mathbf{v} \in \ker(A^{\mathrm{t}}A)$, se tiene que $(A^{\mathrm{t}}A)\mathbf{v} = \mathbf{0}$, de modo que

$$0 = \mathbf{v}^{\mathrm{t}}\mathbf{0} = \mathbf{v}^{\mathrm{t}}(A^{\mathrm{t}}A)\mathbf{v} = (\mathbf{v}^{\mathrm{t}}A^{\mathrm{t}})(A\mathbf{v}) = (A\mathbf{v})^{\mathrm{t}}(A\mathbf{v}),$$

de donde se sigue que $A\mathbf{v} = \mathbf{0}$, como queríamos demostrar.

La demostración de la igualdad $\ker(A^{\mathrm{t}}) = \ker(A\,A^{\mathrm{t}})$ se hace de forma completamente análoga por lo que deja como ejercicio al lector.

Finalmente, por el teorema del rango, se tiene que

$$\mathrm{rg}(A) = n - \dim(\ker(A)) = n - \dim(\ker(A^{\mathrm{t}}\,A)) = \mathrm{rg}(A^{\mathrm{t}}A);$$
$$\mathrm{rg}(A^{\mathrm{t}}) = m - \dim(\ker(A^{\mathrm{t}})) = m - \dim(\ker(A\,A^{\mathrm{t}})) = \mathrm{rg}(A\,A^{\mathrm{t}}).$$

Usando ahora que $\mathrm{rg}(A) = \mathrm{rg}(A^{\mathrm{t}})$ se obtiene el resultado buscado.

(b) Es claro que $A^{\mathrm{t}}A$ y $A\,A^{\mathrm{t}}$ son simétricas, pues $(A^{\mathrm{t}}A)^{\mathrm{t}} = A^{\mathrm{t}}(A^{\mathrm{t}})^{\mathrm{t}} = A^{\mathrm{t}}A$ y $(A\,A^{\mathrm{t}})^{\mathrm{t}} = (A^{\mathrm{t}})^{\mathrm{t}}A^{\mathrm{t}} = A\,A^{\mathrm{t}}$. Al ser ambas matrices simétricas podemos garantizar que todos sus autovalores son reales, de tal que forma que para demostrar que son semidefinidas positivas basta ver que todos sus autovalores son no negativos. Sea,

pues, $\lambda \in \mathbb{R}$ un autovalor de $A^{\mathrm{t}}A$ y $\mathbf{v} \in \mathbb{R}^n$ un autovector de $A^{\mathrm{t}}A$ asociado a λ. Entonces,

$$0 \leq \|A\mathbf{v}\|^2 = (A\mathbf{v})^{\mathrm{t}}(A\mathbf{v}) = \mathbf{v}^{\mathrm{t}}(A^{\mathrm{t}}A)\mathbf{v} = \mathbf{v}^{\mathrm{t}}(\lambda\mathbf{v}) = \lambda(\mathbf{v}^{\mathrm{t}}\mathbf{v}),$$

de donde se sigue que $\lambda \geq 0$. La demostración de que todos los autovalores $A\,A^{\mathrm{t}}$ son no negativos es totalmente análoga; basta cambiar A por A^{t}.

Finalmente, $A^{\mathrm{t}}A \in \mathcal{M}_n(\mathbb{R})$ es definida positiva si, y sólo si, todos los autovalores son positivos, esto es equivalente a que sea invertible, y por lo tanto a que tenga rango n, que coincide con el rango de A, por el apartado (a). La demostración de que la condición necesaria y suficiente para que $A\,A^{\mathrm{t}}$ sea definida positiva es que A tenga rango m es similar. ∎

Teniendo en cuenta que las matrices $A^{\mathrm{t}}A$ y $A\,A^{\mathrm{t}}$ son simétricas, semidefinidas positivas y tienen el mismo rango que A, r, según la proposición anterior, se sigue que ambas diagonalizan mediante una matriz de paso ortogonal y tienen r autovalores estrictamente positivos (no necesariamente distintos) y el resto nulos. Veamos que además tienen los mismos autovalores.

Proposición VI.1.2. *Sea $A \in \mathcal{M}_{m\times n}(\mathbb{R})$. Se cumple que:*

(a) *$A^{\mathrm{t}}A$ y $A\,A^{\mathrm{t}}$ tienen los mismos autovalores no nulos.*
(b) *Si $\mathbf{v}$ es un autovector de $A^{\mathrm{t}}A$ asociado a $\sigma_i^2 \neq 0$, entonces $A\mathbf{v}$ es un autovector de $A\,A^{\mathrm{t}}$ asociado a σ_i^2.*
(c) *Si $\mathbf{u}$ es un autovector de $A\,A^{\mathrm{t}}$ asociado a $\sigma_i^2 \neq 0$, entonces $A^{\mathrm{t}}\mathbf{u}$ es un autovector de $A^{\mathrm{t}}A$ asociado a σ_i^2.*
(d) *La multiplicidad de los autovalores no nulos de $A^{\mathrm{t}}A$ coincide con la de los de $A\,A^{\mathrm{t}}$.*

Demostración. Sea λ un autovalor no nulo de $A^{\mathrm{t}}A$ y $\mathbf{v}$ un autovector de $A^{\mathrm{t}}A$ asociado a λ. Entonces

$$(A\,A^{\mathrm{t}})A\mathbf{v} = A\,(A^{\mathrm{t}}A)\mathbf{v} = A(\lambda\mathbf{v}) = \lambda(A\mathbf{v});$$

luego, λ es un autovalor de $A\,A^{\mathrm{t}}$ y $A\mathbf{v}$ es autovector de $A\,A^{\mathrm{t}}$ asociado a λ. Nótese que $A\mathbf{v} \neq \mathbf{0}$, en otro caso $\lambda\mathbf{v} = A^{\mathrm{t}}A\mathbf{v} = \mathbf{0}$, es decir, $\lambda = 0$, lo que no es posible por hipótesis. El recíproco es similar y se deja como ejercicio al lector.

Sea ahora λ un autovalor no nulo de $A^{\mathrm{t}}A$. Si $\mathbf{u}$ y $\mathbf{v}$ son dos autovectores linealmente independientes de $A^{\mathrm{t}}A$ asociados a λ, entonces $A\mathbf{u}$ y $A\mathbf{v}$ también son linealmente independientes. En efecto, si $\alpha A\mathbf{u} + \beta A\mathbf{v} = \mathbf{0}$, entonces

$$\mathbf{0} = A^{\mathrm{t}}(\alpha A\mathbf{u} + \beta A\mathbf{v}) = \alpha(A^{\mathrm{t}}A)\mathbf{u} + \beta(A^{\mathrm{t}}A)\mathbf{v} = \lambda(\alpha\mathbf{v} + \beta\mathbf{v});$$

de donde se sigue que $\mathbf{0} = \alpha\mathbf{v} + \beta\mathbf{v}$ y por lo tanto que $\alpha = \beta = 0$. Al igual que antes, el recíproco es similar y se deja como ejercicio al lector.

Finalmente, como $A^{\mathrm{t}}A$ y $A\,A^{\mathrm{t}}$ son diagonalizables, se tiene que la multiplicidad de λ coincide con la dimensión del subespacio propio correspondiente. Luego, por el argumento anterior, concluimos que los autovalores no nulos de $A^{\mathrm{t}}A$ y de $A\,A^{\mathrm{t}}$ tienen la misma multiplicidad. ∎

Nótese que los autovalores de no nulos de $A^{\mathrm{t}}A$ (y los de $A\,A^{\mathrm{t}}$) son positivos, puesto que $A^{\mathrm{t}}A$ es definida positiva. De aquí que los denotemos $\sigma_1^2, \ldots, \sigma_r^2$ siendo $r = \mathrm{rg}(A) = \mathrm{rg}(A^{\mathrm{t}}A) = \mathrm{rg}(A\,A^{\mathrm{t}})$.

Teorema VI.1.3. ***Forma reducida ortogonal.*** *Sea $A \in \mathcal{M}_{m\times n}(\mathbb{R})$. Si A tiene rango $r > 0$, existen $P \in \mathcal{M}_m(\mathbb{R})$ y $Q \in \mathcal{M}_n(\mathbb{R})$ ortogonales, tales que $P^{\mathrm{t}}AQ = D$, donde la matriz $D \in \mathcal{M}_{m\times n}(\mathbb{R})$ es una matriz de la forma*

$$\left(\begin{array}{c|c} \Delta & \mathbf{0}_{r\times(n-r)} \\ \hline \mathbf{0}_{(m-r)\times r} & \mathbf{0}_{(m-r)\times(n-r)} \end{array}\right)$$

y Δ es una matriz diagonal con entradas positivas en su diagonal. Las entradas diagonales de Δ^2 son los autovalores positivos de $A^{\mathrm{t}}A$ (que coinciden con los de $A\,A^{\mathrm{t}}$).

Nota VI.1.4. De ahora en adelante, por simplicidad en la notación, escribiremos 0 para denotar a cualquier matriz nula, y sólo especificaremos su orden cuando exista posibilidad de confusión.

Demostración. Sea $\Delta^2 \in \mathcal{M}_r(\mathbb{R})$ la matriz diagonal cuyas entradas en la diagonal son los r autovalores positivos de $A^{\mathrm{t}}A$ (que son los mismos que los autovalores positivos de $A\,A^{\mathrm{t}}$). Sea Δ la matriz diagonal cuyas entradas en la diagonal son las raíces cuadradas positivas de las correspondientes entradas en la diagonal de Δ^2. Como $A^{\mathrm{t}}A$ es una matriz simétrica de orden n, podemos encontrar una matriz ortogonal $Q \in \mathcal{M}_n(\mathbb{R})$ tal que

$$Q^{\mathrm{t}}A^{\mathrm{t}}AQ = \begin{pmatrix} \Delta^2 & 0 \\ 0 & 0 \end{pmatrix}.$$

Partiendo Q como $Q = (Q_1|Q_2)$, donde Q_1 es una matriz $n \times r$, la identidad anterior implica que

$$Q_1^{\mathrm{t}}A^{\mathrm{t}}AQ_1 = \Delta^2 \tag{VI.1.1}$$

y

$$(AQ_2)^{\mathrm{t}}(AQ_2) = Q_2^{\mathrm{t}}A^{\mathrm{t}}AQ_2 = \mathbf{0}_{(n-r)\times(n-r)}, \tag{VI.1.2}$$

de donde se sigue que

$$AQ_2 = \mathbf{0}_{n\times(n-r)}. \tag{VI.1.3}$$

Sea $P_1 = AQ_1\Delta^{-1} \in \mathcal{M}_{m\times r}(\mathbb{R})$. En primer lugar observamos que las columnas de P_1 son ortogonales; en efecto,

$$P_1^{\mathrm{t}}P_1 = (AQ_1\Delta^{-1})^{\mathrm{t}}(AQ_1\Delta^{-1}) = (\Delta^{-1})^{\mathrm{t}}Q_1^{\mathrm{t}}A^{\mathrm{t}}AQ_1\Delta^{-1} = \Delta^{-1}\Delta^2\Delta^{-1} = I_r.$$

Sea ahora $P = (P_1|P_2)$ una matriz ortogonal de orden m, donde $P_2 \in \mathcal{M}_{m\times(m-r)}(\mathbb{R})$ es cualquier matriz que la haga ortogonal. Por consiguiente, se tiene que $P_2^{\mathrm{t}}P_1 = P_2^{\mathrm{t}}AQ_1\Delta^{-1} = \mathbf{0}_{(m-r)\times r}$ ó, equivalentemente,

$$P_2^{\mathrm{t}}AQ_1 = \mathbf{0}_{(m-r)\times r} \tag{VI.1.4}$$

Usando ahora (VI.1.1), (VI.1.3) y (VI.1.4), obtenemos que

$$\begin{aligned} P^{\mathrm{t}}AQ &= \begin{pmatrix} P_1^{\mathrm{t}}AQ_1 & P_1^{\mathrm{t}}AQ_2 \\ P_2^{\mathrm{t}}AQ_1 & P_2^{\mathrm{t}}AQ_2 \end{pmatrix} = \begin{pmatrix} \Delta^{-1}Q_1^{\mathrm{t}}A^{\mathrm{t}}AQ_1 & \Delta^{-1}Q_1^{\mathrm{t}}A^{\mathrm{t}}AQ_2 \\ P_2^{\mathrm{t}}AQ_1 & P_2^{\mathrm{t}}AQ_2 \end{pmatrix} \\ &= \begin{pmatrix} \Delta^{-1}\Delta^2 & \Delta^{-1}Q_1^{\mathrm{t}}A^{\mathrm{t}}\mathbf{0}_{n\times(n-r)} \\ 0 & P_2^{\mathrm{t}}\mathbf{0}_{n\times(n-r)} \end{pmatrix} = \begin{pmatrix} \Delta & 0 \\ 0 & 0 \end{pmatrix} \end{aligned}$$

■

Definición VI.1.5. Sea $A \in \mathcal{M}_{m\times n}(\mathbb{R})$. Las raíces cuadradas positivas de los autovalores de $A^{\mathrm{t}}A$ (y de $A\,A^{\mathrm{t}}$), se llaman **valores singulares de** A. La descomposición $A = PDQ^{\mathrm{t}}$ dada en el teorema VI.1.3 se llama **descomposición en valores singulares** o **SVD** de A.

Nota VI.1.6. Los valores singulares se denotan como $\sigma_1, \sigma_2, \ldots, \sigma_r$ con la ordenación $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > 0$.

Siguiendo la notación del teorema VI.1.3, los valores singulares de A son las entradas de la diagonal de Δ. Por la demostración del teorema VI.1.3, es obvio que las columnas de Q forman una base ortonormal de autovectores $A^{\mathrm{t}}A$, y por lo tanto

$$A^{\mathrm{t}}A = QD^{\mathrm{t}}DQ^{\mathrm{t}}. \tag{VI.1.5}$$

También es importante destacar que las columnas de P forman una base ortonormal de autovectores de $A\,A^{\mathrm{t}}$ ya que

$$A\,A^{\mathrm{t}} = PDQ^{\mathrm{t}}QDP^{\mathrm{t}}. \tag{VI.1.6}$$

Si volvemos a considerar particiones P y Q como $P = (P_1|P_2)$ y $Q = (Q_1|Q_2)$, con $P_1 \in \mathcal{M}_{m\times r}(\mathbb{R})$ y $Q_1 \in \mathcal{M}_{n\times r}(\mathbb{R})$, entonces la descomposición en valores singulares de A se puede reescribir como sigue.

Corolario VI.1.7. *Sea $A \in \mathcal{M}_{m\times n}(\mathbb{R})$. Si A tiene rango $r > 0$, entonces existen $P_1 \in \mathcal{M}_{m\times r}(\mathbb{R})$ y $Q_1 \in \mathcal{M}_{n\times r}(\mathbb{R})$ tales que $P_1^{\mathrm{t}}P_1 = Q_1^{\mathrm{t}}Q_1 = I_r$, y*

$$A = P_1\Delta Q_1^{\mathrm{t}}, \tag{VI.1.7}$$

donde $\Delta \in \mathcal{M}_r(\mathbb{R})$ es diagonal con entradas positivas en su diagonal.

La expresión (VI.1.7) se llama **descomposición en valores singulares corta** o **SVD corta** de A.

Se sigue de (VI.1.5) y de (VI.1.6) que P_1 y Q_1 son matrices *semiortogonales*, es decir, matrices cuyas columnas son mutuamente ortogonales y de norma 1, verificando

$$P_1^{\mathrm{t}} A A^{\mathrm{t}} P_1 = Q_1^{\mathrm{t}} A^{\mathrm{t}} A Q_1 = \Delta^2. \tag{VI.1.8}$$

Sin embargo, en la descomposición $A = P_1 \Delta Q_1^{\mathrm{t}}$, la elección de la matriz *semiortogonal* P_1 verificando (VI.1.8) depende de la elección de la matriz Q_1. Téngase en cuenta que en la demostración del teorema VI.1.3 se elige una matriz *semiortogonal* Q_1 verificando (VI.1.8), pero P_1 viene dada por $P_1 = AQ_1\Delta^{-1}$. Alternativamente, se podría haber seleccionado primero P_1 verificando (VI.1.8) y tomar posteriormente $Q_1 = A^{\mathrm{t}} P_1 \Delta^{-1}$.

De esta descomposición en valores singulares se puede obtener gran cantidad de información sobre la estructura de la matriz A. El número de valores singulares es el rango de A, mientras que las columnas de P_1 y Q_1 son bases ortogonales de $\mathrm{im}(A)$ e $\mathrm{im}(A^{\mathrm{t}})$, respectivamente. Análogamente, las columnas de P_2 generan $\ker(A^{\mathrm{t}})$ y las columnas de Q_2 generan $\ker(A)$.

Ejemplo VI.1.8. Hallemos la descomposición en valores singulares corta de la siguiente matriz

$$A = \begin{pmatrix} 2 & 0 & 1 \\ 3 & -1 & 1 \\ -2 & 4 & 1 \\ 1 & 1 & 1 \end{pmatrix}.$$

En primer lugar, calculamos los autovalores y autovectores normalizados de la matriz

$$A^{\mathrm{t}} A = \begin{pmatrix} 18 & -10 & 4 \\ -10 & 18 & 4 \\ 4 & 4 & 4 \end{pmatrix}.$$

Los autovalores son $\sigma_1^2 = 28, \sigma_2^2 = 12$ y $\sigma_3^2 = 0$, y sus respectivos autovectores normalizados son $(1/\sqrt{2}, -1/\sqrt{2}, 0)^{\mathrm{t}}$, $(1/\sqrt{3}, 1/\sqrt{3}, 1/\sqrt{3})^{\mathrm{t}}$ y $(1/\sqrt{6}, 1/\sqrt{6}, -2/\sqrt{6})^{\mathrm{t}}$. Es claro, que el rango de A es 2 y que los dos valores singulares de A son $\sigma_1 = \sqrt{28}$ y $\sigma_2 = \sqrt{12}$. Por tanto,

$$\Delta = \mathrm{diag}(\sigma_1, \sigma_2) = \begin{pmatrix} \sqrt{28} & 0 \\ 0 & \sqrt{12} \end{pmatrix}.$$

Sean $Q_1 \in \mathcal{M}_{3\times 2}(\mathbb{R})$ la matriz cuyas columnas son los dos primeros autovectores, $Q_2 \in \mathcal{M}_{3\times 1}(\mathbb{R})$ y $Q = (Q_1|Q_2) \in \mathcal{M}_3(\mathbb{R})$. Por tanto la matriz $P_1 \in \mathcal{M}_{4\times 2}(\mathbb{R})$ es

$$P_1 = AQ_1\Delta^{-1} = \begin{pmatrix} 2 & 0 & -1 \\ 3 & -1 & 1 \\ -2 & 4 & 1 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{3} \\ -1/\sqrt{2} & 1/\sqrt{3} \\ 0 & 1/\sqrt{3} \end{pmatrix} \begin{pmatrix} 1/\sqrt{28} & 0 \\ 0 & 1/\sqrt{12} \end{pmatrix}$$

$$= \begin{pmatrix} 1/\sqrt{14} & 1/2 \\ 2/\sqrt{14} & 1/2 \\ -3/\sqrt{14} & 1/2 \\ 0 & 1/2 \end{pmatrix}.$$

Por consiguiente, la descomposición en valores singulares corta de A es

$$\begin{pmatrix} 1/\sqrt{14} & 1/2 \\ 2/\sqrt{14} & 1/2 \\ -3/\sqrt{14} & 1/2 \\ 0 & 1/2 \end{pmatrix} \begin{pmatrix} \sqrt{28} & 0 \\ 0 & \sqrt{12} \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} & 0 \\ 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \end{pmatrix}.$$

Nota VI.1.9. La descomposición en valores singulares de un vector es muy fácil de construir. En efecto, si $\mathbf{v} \in \mathcal{M}_{m\times 1}(\mathbb{R})$ es un vector no nulo de $\mathbb{R}^m$, su descomposición en valores singulares es de la forma

$$\mathbf{v} = \mathbf{p}_1 \delta q_1,$$

con $\delta = \sqrt{\mathbf{v}^{\mathrm{t}}\mathbf{v}}$, $\mathbf{p}_1 = \delta^{-1}\mathbf{v}$ y $q_1 = 1$.

Cuando la matriz A es simétrica, los valores singulares de A están directamente relacionados con sus autovalores. En efecto, si A es simétrica, entonces $A\,A^{\mathrm{t}} = A^2$, y los autovalores de A^2 son los cuadrados de los autovalores de A. Por consiguiente, los valores singulares de A serán los valores absolutos de los autovalores de A. Si P es una matriz cuyas columnas forman una base ortonormal de autovectores de A, entonces la matriz Q del teorema VI.1.3 será idéntica a P excepto para aquella columnas asociadas a autovalores negativos de A que serán -1 veces la correspondiente columna de P. Si A es semidefinida positiva, entonces la descomposición de valores singulares de A es precisamente la descomposición $A = PDP^{\mathrm{t}}$ estudiada en el tema V. Esta bonita relación entre los autovalores y los valores singulares no ocurre en general.

Ejemplo VI.1.10. Consideremos la matriz

$$A = \begin{pmatrix} 6 & 6 \\ -1 & 1 \end{pmatrix}.$$

Como

$$A\,A^{\mathrm{t}} = \begin{pmatrix} 72 & 0 \\ 0 & 2 \end{pmatrix},$$

los valores singulares de A son $\sqrt{72} = 6\sqrt{2}$ y $\sqrt{2}$, mientras que los autovalores de A son 4 y 3.

Veamos ahora algunas aplicaciones inmediatas de las descomposición en valores singulares.

Corolario VI.1.11. *Sean A y $B \in \mathcal{M}_{m\times n}(\mathbb{R})$. Si $A^{\mathrm{t}}A = B^{\mathrm{t}}B$, entonces existe una matriz ortogonal $U \in \mathcal{M}_m(\mathbb{R})$ tal que $B = UA$.*

Demostración. Si la descomposición en valores singulares de A es $A = P_1\Delta Q_1^{\mathrm{t}}$, entonces la descomposición en valores singulares de B es $B = P_1'\Delta Q_1^{\mathrm{t}}$ con $P_1' = BQ_1\Delta^{-1}$. Luego, $B = (P_1'P_1^{\mathrm{t}})A$. La comprobación de que $U = P_1'P_1^{\mathrm{t}} \in \mathcal{M}_m(\mathbb{R})$ es ortogonal se deja como ejercicio al lector. ∎

Corolario VI.1.12. *Sean X y $Y \in \mathcal{M}_{m\times n}(\mathbb{R})$ y B y $C \in \mathcal{M}_m(\mathbb{R})$ simétricas definidas positivas. Si $X^{\mathrm{t}}B^{-1}X = YC^{-1}Y$, entonces existe una matriz invertible $A \in \mathcal{M}_m(\mathbb{R})$ tal que $Y = AX$ y $C = A\,B\,A^{\mathrm{t}}$.*

Demostración. Por ser B y C simétricas y definidas positivas existen $B^{1/2}$ y $C^{1/2}$ simétricas tales que $B = B^{1/2}B^{1/2}$ y $C = C^{1/2}C^{1/2}$, y también existen $B^{-1/2}$ y $C^{-1/2}$ simétricas tales que $B^{-1} = B^{-1/2}B^{-1/2}$ y $C^{-1} = C^{-1/2}C^{-1/2}$ (véase el corolario V.5.9).

Sean $X_1 = B^{-1/2}X$ y $X_2 = C^{-1/2}Y$. Como

$$X_1^{\mathrm{t}}X_1 = X^{\mathrm{t}}B^{-1/2}B^{-1/2}X = X^{\mathrm{t}}B^{-1}X = YC^{-1}Y^{-1} = Y^{\mathrm{t}}C^{-1/2}C^{-1/2}Y = X_2^{\mathrm{t}}X_2,$$

por el corolario VI.1.11, obtenemos que existe una matriz U ortogonal tal que $X_2 = UX_1$, es decir, $C^{-1/2}Y = UB^{-1/2}X$, luego, $Y = C^{1/2}UB^{-1/2}X$. De modo que basta tomar $A = B^{1/2}UC^{-1/2}$ para concluir que $Y = AX$ y que

$$ABA^{\mathrm{t}} = C^{1/2}UB^{-1/2}BB^{-1/2}U^{\mathrm{t}}C^{1/2} = C.$$

∎

Interpretación geométrica de la descomposición en valores singulares.

Sean $A \in \mathcal{M}_{m\times n}(\mathbb{R})$ y $T : \mathbb{R}^n \to \mathbb{R}^m$ la aplicación lineal cuya matriz respecto de las bases usuales de $\mathbb{R}^n$ y $\mathbb{R}^m$, respectivamente, es A. Consideremos las descomposiciones $\mathbb{R}^n = \ker(T)^{\perp} \oplus \ker(T)$ y $\mathbb{R}^m = \mathrm{im}(T) \oplus \mathrm{im}(T)^{\perp}$.

Obsérvese que $\phi = T_{|_{\ker(T)^{\perp}}}$ es inyectiva. Además,

$$\dim(\ker(T)^{\perp}) = n - \dim(\ker(T)) = \mathrm{rg}(A) = \dim(\mathrm{im}(T)).$$

Por lo tanto, ϕ establece un isomorfismo de $\ker(T)^{\perp}$ con $\mathrm{im}(T)$.

Supongamos que $A = P_1 \Delta Q_1^{\mathrm{t}}$ es una descomposición en valores singulares de A. Entonces la matriz de ϕ respecto de la base ortonormal de $\ker(T)^{\perp}$, que forman las columnas de Q_1 y la base ortonromal de $\mathrm{im}(T)$ que forman las columnas de P_1, es Δ.

Para conseguir un punto de vista más visual de los valores singulares y de la descomposición en valores singulares, considérese la esfera S de radio uno en $\mathbb{R}^n$. La aplicación lineal T envía esta esfera a un elipsoide de $\mathbb{R}^m$. Los valores singulares son simplemente las longitudes de los semiejes del elipsoide.

2. La inversa de Moore-Penrose

Una inversa generalizada de utilidad en aplicaciones estadísticas es la desarrollada por E.H. Moore[1] y R. Penrose[2].

Definición VI.2.1. La **inversa de Moore-Penrose** de un matriz $A \in \mathcal{M}_{m\times n}(\mathbb{R})$ es la matriz de orden $n \times m$, que denotaremos por A^+, que verifica las siguientes condiciones.

(G1)	$A\,A^+A = A$,
(G2)	$A^+A\,A^+ = A^+$,
(G3)	$(A\,A^+)^{\mathrm{t}} = A\,A^+$, es decir, $A\,A^+$ es simétrica,
(G4)	$(A^+A)^{\mathrm{t}} = A^+A$, es decir, A^+A es simétrica.

Uno de las particularidades más importantes de la inversa de Moore-Penrose que la distingue de otras inversas generalizadas, es que está unívocamente definida. Este hecho, junto con su existencia, se establece en el siguiente resultado.

Teorema VI.2.2. *Dada una matriz $A \in \mathcal{M}_{m\times n}(\mathbb{R})$, existe una única matriz $A^+ \in \mathcal{M}_{n\times m}(\mathbb{R})$ verificando las condiciones (G1)-(G4) de la definición VI.2.1*

Demostración. En primer lugar probamos la existencia de A^+. Si A es la matriz nula, entonces las cuatro condiciones de la definición VI.2.1 se cumplen trivialmente para $A^+ = \mathbf{0}_{n\times m}$. Si A no es nula, entonces tiene rango $r > 0$. De modo que, por el corolario VI.1.7, sabemos que existen $P_1 \in \mathcal{M}_{m\times r}(\mathbb{R})$ y $Q_1 \in \mathcal{M}_{n\times r}(\mathbb{R})$ tales que $P_1^{\mathrm{t}}P_1 = Q_1^{\mathrm{t}}Q_1 = I_r$, y

$$A = P_1 \Delta Q_1^{\mathrm{t}},$$

[1]Moore, E. H. (1920). *On the reciprocal of the general algebraic matrix.* Bulletin of the American Mathematical Society **26**: 394-395.

[2]Penrose, R. (1955). *A generalized inverse for matrices.* Proceedings of the Cambridge Philosophical Society **51**: 406-413.

donde $\Delta \in \mathcal{M}_r(\mathbb{R})$ es diagonal con entradas positivas en su diagonal. Nótese que si definimos $A^+ = Q_1\Delta^{-1}P_1^{\mathrm{t}}$, entonces

$$\begin{aligned} A\,A^+A &= P_1\Delta Q_1^{\mathrm{t}}Q_1\Delta^{-1}P_1^{\mathrm{t}}P_1\Delta Q_1^{\mathrm{t}} = P_1\Delta\Delta^{-1}\Delta Q_1^{\mathrm{t}} = P_1\Delta Q_1^{\mathrm{t}} = A,\\ A^+A\,A^+ &= Q_1\Delta^{-1}P_1^{\mathrm{t}}P_1\Delta Q_1^{\mathrm{t}}Q_1\Delta^{-1}P_1^{\mathrm{t}} = Q_1\Delta^{-1}\Delta\Delta^{-1}P_1^{\mathrm{t}} = Q_1\Delta^{-1}P_1^{\mathrm{t}} = A^+,\\ A\,A^+ &= P_1\Delta Q_1^{\mathrm{t}}Q_1\Delta^{-1}P_1^{\mathrm{t}} = P_1\,P_1^{\mathrm{t}} \text{ es simétrica},\\ A^+A &= Q_1\Delta^{-1}P_1^{\mathrm{t}}P_1\Delta Q_1^{\mathrm{t}} = Q_1\,Q_1^{\mathrm{t}} \text{ es simétrica}. \end{aligned}$$

Por consiguiente, $A^+ = Q_1\Delta^{-1}P_1^{\mathrm{t}}$ es una inversa de Moore-Penrose de A, y por lo tanto se demuestra la existencia de inversas de Moore-Penrose.

Ahora, supongamos que B y C son dos inversas de Moore-Penrose, es decir, dos matrices de orden $n \times m$ que verifican las condiciones (G1)-(G4) de la definición VI.2.1. Usando estas condiciones, encontramos que

$$AB = (AB)^{\mathrm{t}} = B^{\mathrm{t}}A^{\mathrm{t}} = B^{\mathrm{t}}(ACA)^{\mathrm{t}} = B^{\mathrm{t}}A^{\mathrm{t}}(AC)^{\mathrm{t}} = (AB)^{\mathrm{t}}AC = ABAC = AC$$

y

$$BA = (BA)^{\mathrm{t}} = A^{\mathrm{t}}B^{\mathrm{t}} = (ACA)^{\mathrm{t}}B^{\mathrm{t}} = (CA)^{\mathrm{t}}A^{\mathrm{t}}B^{\mathrm{t}} = CA(BA)^{\mathrm{t}} = CABA = CA.$$

Usando estas dos identidades y la condición (G2) de la definición VI.2.1, vemos que

$$B = BAB = BAC = CAC = C.$$

De modo que, como B y C son idénticas, la inversa de Moore-Penrose es única. ∎

Como acabamos de ver en la demostración del teorema VI.2.2 la inversa de Moore-Penrose de una matriz A está relacionada explícitamente con la descomposición en valores singulares de A; es decir, podemos considerarla como una función de las matrices que componen la descomposición en valores singulares de A.

Ejemplo VI.2.3. La inversa de Moore-Penrose de

$$A = \begin{pmatrix} 2 & 0 & 1\\ 3 & -1 & 1\\ -2 & 4 & 1\\ 1 & 1 & 1 \end{pmatrix}$$

del ejemplo VI.1.8 es

$$\begin{aligned} A^+ = &\begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{3}\\ -1/\sqrt{2} & 1/\sqrt{3}\\ 0 & 1/\sqrt{3} \end{pmatrix} \begin{pmatrix} 1/\sqrt{28} & 0\\ 0 & 1/\sqrt{12} \end{pmatrix} \cdot\\ &\cdot \begin{pmatrix} 1/\sqrt{14} & 2/\sqrt{14} & -3/\sqrt{14} & 0\\ 1/2 & 1/2 & 1/2 & 1/2 \end{pmatrix} = \frac{1}{84}\begin{pmatrix} 10 & 13 & -2 & 7\\ 4 & 1 & 16 & 7\\ 7 & 7 & 7 & 7 \end{pmatrix}. \end{aligned}$$

Nótese que, como hemos apuntando antes, lo único que necesitamos para calcular la inversa de Moore-Penrose es conocer su descomposición en valores singulares.

La definición VI.2.1 es la definición de inversa generalizada dada por Penrose. La siguiente definición alternativa, que es más útil en determinadas ocasiones, es la definición original de Moore. Esta definición aplica el concepto de matrices de proyecciones ortogonales. Recuérdese que si L es un subespacio vectorial de $\mathbb{R}^m$, la proyección ortogonal sobre L es la aplicación lineal

$$\pi_L : \mathbb{R}^m \longrightarrow \mathbb{R}^m; \mathbf{v} \mapsto \mathbf{v}_1 \in L,$$

donde $\mathbf{v}_1$ es el único vector de $\mathbb{R}^m$ tal que $\mathbf{v} - \mathbf{v}_1 \in L^\perp$. Además, si $\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$ es una base ortonormal de L la matriz de π_L respecto de la base usual de $\mathbb{R}^m$ es

$$\mathbf{u}_1\,\mathbf{u}_1^{\mathrm{t}} + \ldots + \mathbf{u}_r\,\mathbf{u}_r^{\mathrm{t}}.$$

Teorema VI.2.4. *Sea $A \in \mathcal{M}_{m\times n}(\mathbb{R})$. La inversa de Moore-Penrose es la única matriz $A^+ \in \mathcal{M}_{n\times m}(\mathbb{R})$ tal que*

(a) *$A\,A^+$ es la matriz de la proyección ortogonal de $\mathbb{R}^m$ sobre $\mathrm{im}(A) \subseteq \mathbb{R}^m$ respecto de la base usual de $\mathbb{R}^m$.*

(b) *A^+A es la matriz de la proyección ortogonal de $\mathbb{R}^n$ sobre $\mathrm{im}(A^+) \subseteq \mathbb{R}^n$ respecto de la base usual de $\mathbb{R}^n$.*

Demostración. Sea A^+ la inversa de Moore-Penrose de A. Entonces, de (G1) y de (G3) se sigue que

$$(\mathbf{v} - A\,A^+\mathbf{v})^{\mathrm{t}}A\mathbf{u} = \mathbf{v}^{\mathrm{t}}A\mathbf{u} - \mathbf{v}^{\mathrm{t}}(A\,A^+)^{\mathrm{t}}A\mathbf{u} = \mathbf{v}^{\mathrm{t}}A\mathbf{u} - \mathbf{v}^{\mathrm{t}}A\,A^+A\mathbf{u} = \mathbf{v}^{\mathrm{t}}A\mathbf{u} - \mathbf{v}^{\mathrm{t}}A\mathbf{u} = \mathbf{0}.$$

De donde se sigue que $(\mathbf{v} - A\,A^+\mathbf{v}) \in \mathrm{im}(A)^\perp$, para todo $\mathbf{u}$ y $\mathbf{v} \in \mathbb{R}^n$.

Por otra parte, como las columnas de P_1 forman una base ortonormal de $\mathrm{im}(A)$, se sigue que $AA^+ = P_1P_1^{\mathrm{t}} = P_1(P_1^{\mathrm{t}}P_1)P_1^{\mathrm{t}}$. Luego, por la proposición V.4.8 se sigue que AA^+ es la matriz de la proyección ortogonal sobre $\mathrm{im}(A)$ respecto de la base usual de R^m.

La demostración de que A^+A es la matriz de las proyección ortogonal sobre $\mathrm{im}(A^+) \subseteq \mathbb{R}^n$ respecto de la base usual de $\mathbb{R}^n$ se obtiene de igual modo usando (G2) y (G4), es decir, intercambiando los papeles de A y A^+.

En cuanto la unicidad, veamos que una matriz B verificando (a) y (b) debe también satisfacer la definición VI.2.1. Las condiciones (G3) y (G4) son inmediatas ya que las matrices de las proyecciones ortogonales son simétricas (véase la proposición V.4.8), mientras que las condiciones (G1) y (G2) siguen del hecho de que las columnas de A están en $\mathrm{im}(A)$, y por lo tanto

$$A\,BA = (A\,B)A = A,$$

y de que las columnas de B están en $\mathrm{im}(B)$, y por lo tanto

$$BA\,B = (BA)B = B.$$

Ahora, la unicidad de la inversa de Moore-Penrose implica que $B = A^+$. ∎

Interpretación geométrica de la inversa de Moore-Penrose.

Sean $A \in \mathcal{M}_{m\times n}(\mathbb{R})$ una matriz de rango r y $T : \mathbb{R}^n \to \mathbb{R}^m$ la aplicación lineal cuya matriz respecto de las bases usuales de $\mathbb{R}^n$ y $\mathbb{R}^m$ es A. Según vimos en la interpretación geométrica de la descomposición en valores singulares, la restricción ϕ de T a $\ker(T)^\perp$ establece un isomorfismo de $\ker(T)^\perp$ en $\mathrm{im}(T)$. Luego, existe $\phi^{-1} : \mathrm{im}(T) \to \ker(T)^\perp$.

Sea $T^+ : \mathbb{R}^m \to \mathbb{R}^n$ la aplicación lineal definida de la siguiente manera,

$$T^+(\mathbf{v}) = \phi^{-1}(\mathbf{v}_1)$$

donde $\mathbf{v}_1$ es la proyección ortogonal de $\mathbf{v}$ sobre $\mathrm{im}(T)$.

Proposición VI.2.5. *Con la notación anterior, la matriz de T^+ respecto de las bases usuales de $\mathbb{R}^m$ y $\mathbb{R}^n$ es A^+, es decir, la inversa de Moore-Penrose de A.*

Demostración. Si $\mathbf{v}_1$ es la proyección ortogonal de $\mathbf{v}$ sobre $\mathrm{im}(T)$, se tiene que

$$T \circ T^+(\mathbf{v}) = T(\phi^{-1}(\mathbf{v}_1)) = \phi(\phi^{-1}(\mathbf{v}_1)) = \mathbf{v}_1,$$

para todo $\mathbf{v} \in \mathbb{R}^n$, es decir, la composición $T\circ T^+$ es la aplicación proyección ortogonal de $\mathbb{R}^m$ en $\mathrm{im}(T) \subseteq \mathbb{R}^m$. Por otro lado,

$$T^+ \circ T(\mathbf{u}) = \phi^{-1}(T(\mathbf{u})) = \mathbf{u}_1,$$

donde $\mathbf{u}_1$ es la proyección ortogonal de $\mathbf{u}$ sobre $\ker(T)^\perp = \mathrm{im}(T^+)$. Luego, la composición $T^+ \circ T$ es la proyección ortogonal de $\mathbb{R}^n$ en $\mathrm{im}(T^+) \subseteq \mathbb{R}^n$.

Tomando ahora las bases usuales de $\mathbb{R}^m$ y $\mathbb{R}^n$ en cada uno de los casos, respectivamente; por el teorema VI.2.4, se obtiene que A^+ es la inversa de Moore-Penrose de A. ∎

Obsérvese que, por definición, se cumplen las siguientes igualdades

$$\mathrm{im}(T \circ T^+) = \mathrm{im}(T^+) = \mathrm{im}(\phi^{-1}) = \ker(T)^\perp$$

y

$$\mathrm{im}(T \circ T^+) = \mathrm{im}(T).$$

Luego, se cumple que

$$\mathrm{rg}(A) = \mathrm{rg}(A^+) = \mathrm{rg}(A\,A^+) = \mathrm{rg}(A^+A). \tag{VI.2.9}$$

Algunas propiedades básicas de la inversa de Moore-Penrose.

Proposición VI.2.6. *Sea $A \in \mathcal{M}_{m\times n}(\mathbb{R})$. Entonces,*

(a) $(\alpha A)^+ = \alpha^{-1}A^+$, *para todo* $\alpha \in \mathbb{R}$, *no nulo.*
(b) $(A^+)^{\mathrm{t}} = (A^{\mathrm{t}})^+$.
(c) $(A^+)^+ = A$.
(d) $A^+ = A^{-1}$, *si A es cuadrada e invertible.*
(e) $(A^{\mathrm{t}}A)^+ = A^+(A^+)^{\mathrm{t}}$ *y* $(A\,A^{\mathrm{t}})^+ = (A^+)^{\mathrm{t}}A^+$.
(f) $(A\,A^+)^+ = A\,A^+$ *y* $(A^+A)^+ = A^+A$.
(g) $A^+ = (A^{\mathrm{t}}A)^+A^{\mathrm{t}} = A^{\mathrm{t}}(A\,A^{\mathrm{t}})^+$.
(h) $A^+ = (A^{\mathrm{t}}A)^{-1}A^{\mathrm{t}}$ *y* $A^+A = I_n$, *si, y sólo si,* $\mathrm{rg}(A) = n$.
(i) $A^+ = A^{\mathrm{t}}(A\,A^{\mathrm{t}})^{-1}$ *y* $A\,A^+ = I_m$, *si, y sólo si,* $\mathrm{rg}(A) = m$.
(j) $A^+ = A^{\mathrm{t}}$, *si las columnas de A son ortogonales, es decir, si* $A^{\mathrm{t}}A = I_n$.

Demostración. Cada uno de los apartados se demuestra usando simplemente las condiciones (G1)-(G4) o la interpretación geométrica de la inversa de Moore-Penrose. Aquí, solamente verificaremos la igualdad $(A^{\mathrm{t}}A)^+ = A^+(A^+)^{\mathrm{t}}$, dada en el apartado (e), dejando los restantes apartados como ejercicios para lector.

(e) Como A^+ verifica las condiciones (G1)-(G4), tenemos que

$$\begin{aligned} A^{\mathrm{t}}A\,A^+(A^+)^{\mathrm{t}}A^{\mathrm{t}}A &= A^{\mathrm{t}}A\,A^+(A\,A^+)^{\mathrm{t}}A = A^{\mathrm{t}}A\,A^+A\,A^+A \\ &= A^{\mathrm{t}}A\,A^+A = A^{\mathrm{t}}A, \\ A^+(A^+)^{\mathrm{t}}A^{\mathrm{t}}A\,A^+(A^+)^{\mathrm{t}} &= A^+(A\,A^+)^{\mathrm{t}}A\,A^+(A^+)^{\mathrm{t}} = A^+A\,A^+A\,A^+(A^+)^{\mathrm{t}} \\ &= A^+A\,A^+(A^+)^{\mathrm{t}} = A^+(A^+)^{\mathrm{t}} = (A^{\mathrm{t}}A)^+. \end{aligned}$$

Luego, $A^+(A^+)^{\mathrm{t}}$ verifica las condiciones (G1) y (G2) de la inversa de Moore-Penrose de $(A^{\mathrm{t}}A)^+$. Además, nótese que

$$\begin{gathered} (A^{\mathrm{t}}A)(A^+(A^+)^{\mathrm{t}}) = A^{\mathrm{t}}A(A^{\mathrm{t}}A)^+ = A^{\mathrm{t}}AA^+(A^+)^{\mathrm{t}} = A^{\mathrm{t}}(A^+(A\,A^+)^{\mathrm{t}})^{\mathrm{t}} \\ A^{\mathrm{t}}(A^+A\,A^+)^{\mathrm{t}} = A^{\mathrm{t}}(A^+)^{\mathrm{t}} = (A^+A)^{\mathrm{t}}, \end{gathered}$$

y como A^+A es simétrica por definición, se sigue que la condición (G3) se cumple para $(A^{\mathrm{t}}A)^+ = A^+(A^+)^{\mathrm{t}}$. Análogamente, la condición (G4) también se cumple, pues

$$\begin{aligned} (A^+(A^+)^{\mathrm{t}})(A^{\mathrm{t}}A) &= (A^{\mathrm{t}}A)^+A^{\mathrm{t}}A = A^+(A^+)^{\mathrm{t}}A^{\mathrm{t}}A = A^+(A\,A^+)^{\mathrm{t}}A \\ &= A^+A\,A^+A = A^+A. \end{aligned}$$

Esto demuestra que $(A^{\mathrm{t}}A)^+ = A^+(A^+)^{\mathrm{t}}$. ∎

Las propiedades (h) e (i) de la proposición VI.2.6 proporcionan fórmulas para calcular la inversa de Moore-Penrose de matrices que tienen rango pleno por columnas o por filas[3], respectivamente. Ilustremos su utilidad con un ejemplo.

Ejemplo VI.2.7. Sea

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

Como $\mathrm{rg}(A) = 2$, podemos usar la propiedad (i). Si calculamos $A\,A^{\mathrm{t}}$ y luego $(A\,A^{\mathrm{t}})^{-1}$, obtenemos que

$$A\,A^{\mathrm{t}} = \begin{pmatrix} 6 & 4 \\ 4 & 5 \end{pmatrix} \quad \text{y} \quad (A\,A^{\mathrm{t}})^{-1} = \frac{1}{14}\begin{pmatrix} 5 & -4 \\ -4 & 6 \end{pmatrix},$$

y por tanto

$$A^{+} = A^{\mathrm{t}}(A\,A^{\mathrm{t}})^{-1} = \frac{1}{14}\begin{pmatrix} 1 & 2 \\ 2 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 5 & -4 \\ -4 & 6 \end{pmatrix} = \frac{1}{14}\begin{pmatrix} -3 & 8 \\ 6 & -2 \\ 5 & -4 \end{pmatrix};$$

y podemos comprobar que $A\,A^{+} = I_2$; en efecto,

$$A\,A^{+} = \frac{1}{14}\begin{pmatrix} 1 & 2 & 1 \\ 2 & 1 & 0 \end{pmatrix}\begin{pmatrix} -3 & 8 \\ 6 & -2 \\ 5 & -4 \end{pmatrix} = \frac{1}{14}\begin{pmatrix} 14 & 0 \\ 0 & 14 \end{pmatrix} = I_2.$$

Sin embargo, $A^{+}A \neq I_3$ como podemos comprobar

$$A^{+}A = \frac{1}{14}\begin{pmatrix} -3 & 8 \\ 6 & -2 \\ 5 & -4 \end{pmatrix}\begin{pmatrix} 1 & 2 & 1 \\ 2 & 1 & 0 \end{pmatrix} = \frac{1}{14}\begin{pmatrix} 13 & 2 & -3 \\ 2 & 10 & 6 \\ -3 & 6 & 5 \end{pmatrix}.$$

De hecho las propiedades (h) e (i) de la proposición VI.2.6 dan una condición necesaria y suficiente para que una matriz $A \in \mathcal{M}_{m\times n}(\mathbb{R})$ tenga inversa a izquierda y/o a derecha. La inversa a izquierda (a derecha, respectivamente) si existe no tiene por qué ser única; es decir, pueden existir multitud de inversas a izquierda (a derecha, respectivamente).

3. Otras inversas generalizadas

La inversa de Moore-Penrose sólo es una de las muchas inversas generalizadas que han sido desarrolladas en los últimos años. En esta sección, trataremos brevemente otras dos inversas generalizas que tienen aplicación en estadística. Ambas se pueden definir usando las condiciones (G1)-(G4) ó, por simplicidad, 1-4, de la inversa de

[3]Se dice que una matriz $A \in \mathcal{M}_{m\times n}(\Bbbk)$ tiene rango pleno por filas si $\mathrm{rg}(A) = m$ y diremos que tiene rango pleno por columnas si $\mathrm{rg}(A) = n$.

Moore-Penrose. De hecho, podemos definir diferentes clases de inversas generalizadas, según el subconjunto de las condiciones 1-4 que la inversa generalizada ha de cumplir.

Definición VI.3.1. Sea $A \in \mathcal{M}_{m\times n}(\mathbb{R})$. Denotaremos por $A^{(i_1,\ldots,i_r)}$ a cualquier matriz que cumpla las condiciones $i_1,\ldots,i_r$ entre las condiciones 1-4; se dirá que la $A^{(i_1,\ldots,i_r)}$ es una $\{i_1,\ldots,i_r\}$**-inversa**.

Según la definición anterior, la inversa de Moore-Penrose de A es una $\{1,2,3,4\}$-inversa de A; es decir, $A^+ = A^{(1,2,3,4)}$. Nótese que para cualquier subconjunto propio $\{i_1,\ldots,i_r\}$ de $\{1,2,3,4\}$, A^+ también será una $\{i_1,\ldots,i_r\}$-inversa de A, pero no será la única. Como en muchos casos, hay muchas $\{i_1,\ldots,i_r\}$-inversas de A, puede ser más fácil calcular una $\{i_1,\ldots,i_r\}$-inversa de A que la inversa de Moore-Penrose.

Ejemplo VI.3.2. Sea $A \in \mathcal{M}_{m\times n}(\mathbb{R})$. Si A tiene rango r y $P \in \mathcal{M}_m(\mathbb{R})$ y $Q \in \mathcal{M}_n(\mathbb{R})$ son matrices invertibles tales que $R = P^{-1}AQ$ es la forma reducida por filas de A, entonces

$$B = QR^{\mathrm{t}}P^{-1}$$

es una $\{1,2\}$-inversa de A. En efecto,

$$ABA = PRQQ^{-1}R^{\mathrm{t}}PP^{-1}RQ = A,$$
$$BAB = QR^{\mathrm{t}}P^{-1}PRQ^{-1}QR^{\mathrm{t}}P^{-1} = B.$$

Obsérvese que la inversa de Moore-Penrose de R es R^{t}.

Veamos un caso concreto: sea A la matriz

$$\begin{pmatrix} 1 & 1 & 1 & 2 \\ 1 & 0 & 1 & 0 \\ 2 & 1 & 2 & 2 \end{pmatrix}.$$

La forma reducida de A es

$$PAQ^{-1} = R = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

con

$$P = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 2 & 1 & 0 \end{pmatrix} \quad \text{y} \quad Q = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & -1 & 0 \\ 0 & 1/2 & 0 & -1/2 \end{pmatrix}.$$

Entonces, una inversa generalizada de A es

$$B = QR^{\mathrm{t}}P^{-1} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 1/2 \end{pmatrix}.$$

El resto de esta sección está dedicado a la $\{1\}$-inversa y a la $\{1,3\}$-inversa de A, cuyas aplicaciones serán discutidas en la última sección de este tema. En la siguiente sección veremos que para resolver sistemas de ecuaciones lineales, solamente necesitaremos matrices que verifiquen la primera condición de las definición de inversa de Moore-Penrose. Nos referiremos a tales $\{1\}$-inversas de A simplemente como inversas generalizadas de A, y escribiremos A^- en vez de $A^{(1)}$.

Sabemos que otra forma de calcular una inversa generalizada de una matriz consiste en conocer su descomposición en valores singulares. Veamos que la descomposición en valores singulares permite determinar todas las inversas generalizadas de una matriz dada.

Proposición VI.3.3. *Sea $A \in \mathcal{M}_{m\times n}(\mathbb{R})$. Si A tiene rango $r > 0$ y*

$$A = P \begin{pmatrix} \Delta & 0 \\ 0 & 0 \end{pmatrix} Q^{\mathrm{t}}$$

es una descomposición en valores singulares de A, entonces para cada $E \in \mathcal{M}_{r\times(m-r)}$, $F \in \mathcal{M}_{(n-r)\times r}(\mathbb{R})$ y $G \in \mathcal{M}_{(n-r)\times(m-r)}(\mathbb{R})$, la matriz

$$B = Q \begin{pmatrix} \Delta^{-1} & E \\ F & G \end{pmatrix} P^{\mathrm{t}}$$

es una inversa generalizada de A, y cualquier inversa generalizada de A se puede expresar en la forma de B para ciertas E, F y G.

Demostración. Nótese que

$$\begin{aligned} A\,B\,A &= P \begin{pmatrix} \Delta & 0 \\ 0 & 0 \end{pmatrix} Q^{\mathrm{t}} Q \begin{pmatrix} \Delta^{-1} & E \\ F & G \end{pmatrix} P^{\mathrm{t}} P \begin{pmatrix} \Delta & 0 \\ 0 & 0 \end{pmatrix} Q^{\mathrm{t}} = P \begin{pmatrix} \Delta\,\Delta^{-1}\Delta & 0 \\ 0 & 0 \end{pmatrix} Q^{\mathrm{t}} \\ &= P \begin{pmatrix} \Delta & 0 \\ 0 & 0 \end{pmatrix} Q^{\mathrm{t}} = A, \end{aligned}$$

y por lo tanto B es una inversa generalizada de A independientemente de la elección de E, F y G. Por otra parte, si escribimos $Q = (Q_1|Q_2)$ y $P = (P_1|P_2)$, con $Q_1 \in \mathcal{M}_{n\times r}(\mathbb{R})$ y $P \in \mathcal{M}_{m\times r}(\mathbb{R})$, entonces, como $P\,P^{\mathrm{t}} = I_m$ y $Q\,Q^{\mathrm{t}} = I_n$, cualquier inversa generalizada B, de A, se puede expresar como

$$B = Q\,Q^{\mathrm{t}} B\,P\,P^{\mathrm{t}} = Q \begin{pmatrix} Q_1^{\mathrm{t}} \\ Q_2^{\mathrm{t}} \end{pmatrix} B\,(P_1|P_2)\,P^{\mathrm{t}} = Q \begin{pmatrix} Q_1^{\mathrm{t}} B\,P_1 & Q_1^{\mathrm{t}} B\,P_2 \\ Q_2^{\mathrm{t}} B\,P_1 & Q_2^{\mathrm{t}} B\,P_2 \end{pmatrix} P^{\mathrm{t}},$$

que tendrá la forma requerida si somos capaces de probar que $Q_1^{\mathrm{t}} B\,P_1 = \Delta^{-1}$. Como B es una inversa generalizada de A, $A\,B\,A = A$, o equivalentemente,

$$(P^{\mathrm{t}} A\,Q)(Q^{\mathrm{t}} B\,P)(P^{\mathrm{t}} A\,Q) = P^{\mathrm{t}} A\,Q.$$

Escribiendo esta igualdad en forma divida por bloques e igualando las matrices superiores izquierdas de ambos lados, obtenemos que

$$\Delta Q_1^{\mathrm{t}} B P_1 \Delta = \Delta$$

de donde se sigue que $Q_1^{\mathrm{t}} B P_1 = \Delta^{-1}$. ∎

Cuando $A \in \mathcal{M}_m(\mathbb{R})$ es invertible, la matriz B de la proposición VI.3.3 es $B = Q\Delta^{-1}P^{\mathrm{t}}$, esto es, la inversa de A. Por tanto, si A es invertible, la única inversa generalizada de A es A^{-1}.

Ejemplo VI.3.4. La matriz

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

tiene rango $r = 2$ y su descomposición en valores singulares (larga) es

$$A = \frac{1}{2}\begin{pmatrix} 1 & 1 & 1 & -1 \\ 1 & 1 & -1 & 1 \\ 1 & -1 & 1 & 1 \\ 1 & -1 & -1 & -1 \end{pmatrix}\begin{pmatrix} \sqrt{2} & 0 & 0 \\ 0 & \sqrt{3} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} & 0 \\ 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \\ 1/\sqrt{6} & 1/\sqrt{6} & -2/\sqrt{6} \end{pmatrix}.$$

Si tomamos E, F y G iguales a matrices nulas y usamos la ecuación de B dada en la proposición VI.3.3 obtenemos que una inversa generalizada de A es

$$\frac{1}{12}\begin{pmatrix} 5 & 5 & 1 & 1 \\ -1 & -1 & -5 & -5 \\ 2 & 2 & -2 & -2 \end{pmatrix}.$$

De hecho, según la demostración del teorema VI.2.2, sabemos que la matriz anterior es la inversa de Moore-Penrose de A. Se pueden construir otras inversas generalizadas de A mediante distintas elecciones de E, F y G; por ejemplo, si tomamos otra vez E y F nulas pero

$$G = \begin{pmatrix} 1/\sqrt{6} & 0 \end{pmatrix},$$

entonces obtenemos la inversa generalizada

$$\frac{1}{6}\begin{pmatrix} 3 & 2 & 1 & 0 \\ 0 & -1 & -2 & -3 \\ 0 & 2 & -2 & 0 \end{pmatrix}.$$

Obsérvese que esta matriz tiene rango 3, mientras que la inversa de Moore-Penrose tiene el mismo rango que A que, en este caso, es 2.

Veamos ahora algunas propiedades de las $\{1\}$-inversas.

Proposición VI.3.5. *Sea $A \in \mathcal{M}_{m\times n}(\mathbb{R})$, y sea $A^- \in \mathcal{M}_{n\times m}$ una inversa generalizada de A. Entonces*

(a) $(A^-)^{\mathrm{t}}$ *es una inversa generalizada de* A^{t}.
(b) *si* $\alpha \in \mathbb{R}$ *es no nulo,* $\alpha^{-1}A^-$ *es una inversa generalizada de* αA.
(c) *si* A *es cuadrada e invertible,* $A^- = A^{-1}$ *de forma única.*
(d) *si* B *y* C *son invertibles,* $C^{-1}A^-B^{-1}$ *es una inversa generalizada de* BAC.
(e) $\mathrm{rg}(A) = \mathrm{rg}(A\,A^-) = \mathrm{rg}(A^-A) \leq \mathrm{rg}(A^-)$.
(f) $\mathrm{rg}(A) = m$ *si, y sólo si,* $A\,A^- = I_m$.
(g) $\mathrm{rg}(A) = n$ *si, y sólo si,* $A^-A = I_n$.

Demostración. Las propiedades (a)-(d) se comprueban fácilmente, sin más que verificar que se cumple la condición (G1). Para demostrar (e), nótese que como $A = A\,A^-A$, podemos usar el ejercicio 5, para obtener que

$$\mathrm{rg}(A) = \mathrm{rg}(A\,A^-A) \leq \mathrm{rg}(A\,A^-) \leq \mathrm{rg}(A)$$

y

$$\mathrm{rg}(A) = \mathrm{rg}(A\,A^-A) \leq \mathrm{rg}(A^-A) \leq \mathrm{rg}(A),$$

por tanto $\mathrm{rg}(A) = \mathrm{rg}(A\,A^-) = \mathrm{rg}(A^-A)$. Además,

$$\mathrm{rg}(A) = \mathrm{rg}(A\,A^-A) \leq \mathrm{rg}(A^-A) \leq \mathrm{rg}(A^-).$$

De (e) se sigue que $\mathrm{rg}(A) = m$ si, y sólo si, $A\,A^-$ es invertible. Multiplicando a izquierda por $(A\,A^-)^{-1}$ la expresión

$$(A\,A^-)^2 = (A\,A^-A)A^- = A\,A^-$$

implica (f). La demostración de (g) es análoga y se deja como ejercicio al lector. ∎

Ejemplo VI.3.6. Algunas de las propiedades de la inversa de Moore-Penrose no se cumplen para las $\{1\}$-inversas. Por ejemplo, sabemos que la inversa de Moore-Penrose de A^+ es A; es decir, $(A^+)^+ = A$. Sin embargo, en general, no está garantizado que A sea la inversa generalizada de A^-, cuando A^- es una inversa generalizada arbitraria. Considérese, por ejemplo, la matriz $A = \mathrm{diag}(0, 2, 4)$. Una elección de inversa generalizada para A es $A^- = \mathrm{diag}(1, 1/2, 1/4)$. Aquí, A^- es invertible, por lo tanto su única inversa generalizada es $A^{-1} = \mathrm{diag}(1, 2, 4)$.

Todas las inversas generalizadas de una matriz A se pueden expresar en términos de cualquier inversa generalizada particular.

Teorema VI.3.7. *Sea $A^- \in \mathcal{M}_{n\times m}(\mathbb{R})$ una inversa generalizada de $A \in \mathcal{M}_{m\times n}$. Entonces para cualquier matriz $C \in \mathcal{M}_{n\times m}(\mathbb{R})$, se cumple que*

$$A^- + C - A^-ACAA^-$$

es una inversa generalizada de A, y cada inversa generalizada B de A se puede escribir de esta forma para $C = B - A^-$.

Demostración. Como $A\,A^-A = A$, se tiene que

$$A(A^-+C-A^-ACAA^-)A = AA^-A+ACA-AA^-ACAA^-A = A+ACA-ACA = A;$$

por tanto, $A^-+C-A^-ACAA^-$ es una inversa generalizada de A, independientemente de la elección de A^- y C.

Por otra parte, sea B una inversa generalizada de A y $C = B - A^-$. Entonces, como $ABA = A$, se tiene que

$$\begin{aligned} A^- + C - A^-ACAA^- &= A^- + (B - A^-) - A^-A(B - A^-)AA^- \\ &= B - A^-ABAA^- + A^-AA^-AA^- \\ &= B - A^-AA^- + A^-AA^- = B. \end{aligned}$$

■

Veamos ahora algunas propiedades de las inversas generalizadas de A^tA.

Proposición VI.3.8. *Sean $A \in \mathcal{M}_{m\times n}(\mathbb{R})$. Si $(A^tA)^-$ es una inversa generalizada cualquiera de A^tA, entonces*

(a) $((A^tA)^-)^t$ *es una inversa generalizada de A^tA.*

(b) *La matriz $A(A^tA)^-A^t$ no depende de la elección de inversa generalizada $(A^tA)^-$.*

(c) $A(A^tA)^-A^t$ *es simétrica, aún en el caso de que $(A^tA)^-$ no lo sea.*

Demostración. Trasponiendo la expresión $A^tA(A^tA)^-A^tA = A^tA$ se obtiene

$$A^tA((A^tA)^-)^tA^tA = A^tA,$$

de donde se sigue (a). Para probar (b) y (c), observamos primer lo siguiente

$$\begin{aligned} A(A^tA)^-A^tA &= AA^+A(A^tA)^-A^tA = (AA^+)^tA(A^tA)^-A^tA \\ &= (A^+)^tA^tA(A^tA)^-A^tA = (A^+)^tA^tA \\ &= (AA^+)^tA = AA^+A = A. \end{aligned}$$

Entonces,

$$\begin{aligned} A(A^tA)^-A^t &= A(A^tA)^-A^t(A^+)^tA^t = A(A^tA)^-A^t(AA^+)^t \\ &= A(A^tA)^-A^tAA^+ = AA^+, \end{aligned} \tag{VI.3.10}$$

donde la igualdad se sigue de la identidad $A(A^tA)^-A^tA = A$ probada más arriba; (b) sigue de (VI.3.10) ya que A^+, y por tanto AA^+, es única. La simetría de $A(A^tA)^-A^t$ se sigue de la simetría de AA^+. ■

En la siguiente sección veremos que la $\{1,3\}$-inversa es útil para hallar soluciones aproximadas mínimo cuadráticas de sistemas de ecuaciones lineales incompatibles.

Consecuentemente, estas inversas generalizadas se suelen llamar **inversas mínimo cuadráticas**, y las denotaremos $A^{\square}$ en vez de $A^{(1,3)}$. Como las inversas mínimo cuadráticas de A son también $\{1\}$-inversas de A, entonces las propiedades dadas en la proposición VI.3.5 también se aplican a $A^{\square}$ (en el contexto de las $\{1\}$-inversas, ¡claro!). Veamos algunas propiedad más de las inversas mínimo cuadráticas.

Proposición VI.3.9. *Sea $A \in \mathcal{M}_{m\times n}(\mathbb{R})$. Entonces,*

(a) *para cualquier inversa mínimo cuadrática, $A^{\square}$, de A, se cumple que $AA^{\square} = AA^{+}$,*

(b) *$(A^{\mathrm{t}}A)^{-}A^{\mathrm{t}}$ es una inversa mínimo cuadrática de A para cualquier inversa generalizada, $(A^{\mathrm{t}}A)^{-}$, de $A^{\mathrm{t}}A$.*

Demostración. Como $AA^{\square}A = A$ y $(AA^{\square})^{\mathrm{t}} = AA^{\square}$, podemos probar que

$$\begin{aligned} AA^{\square} &= AA^{+}AA^{\square} = (AA^{+})^{\mathrm{t}}(AA^{\square})^{\mathrm{t}} = (A^{+})^{\mathrm{t}}A^{\mathrm{t}}(A^{\square})^{\mathrm{t}}A^{\mathrm{t}} \\ &= (A^{+})^{\mathrm{t}}(AA^{\square}A)^{\mathrm{t}} = (A^{+})^{\mathrm{t}}A^{\mathrm{t}} = (AA^{+})^{\mathrm{t}} = AA^{+}. \end{aligned}$$

El apartado (b) se sigue de la demostración de la proposición VI.3.8 donde ya demostramos las igualdades

$$A\left((A^{\mathrm{t}}A)^{-}A^{\mathrm{t}}\right)A = A$$

y

$$A\left((A^{\mathrm{t}}A)^{-}A^{\mathrm{t}}\right) = AA^{+},$$

es decir, que $(A^{\mathrm{t}}A)^{-}A^{\mathrm{t}}$ es una inversa mínimo cuadrática. ∎

Corolario VI.3.10. *Sea $A \in \mathcal{M}_{m\times n}(\mathbb{R})$. Si A tiene rango $r > 0$ y*

$$A = P\begin{pmatrix} \Delta & 0 \\ 0 & 0 \end{pmatrix}Q^{\mathrm{t}}$$

es una descomposición en valores singulares de A, entonces para cada $F \in \mathcal{M}_{(n-r)\times r}(\mathbb{R})$ y $G \in \mathcal{M}_{(n-r)\times(m-r)}(\mathbb{R})$, la matriz

$$B = Q\begin{pmatrix} \Delta^{-1} & 0 \\ F & G \end{pmatrix}P^{\mathrm{t}}$$

es una mínimo cuadrática de A, y cualquier inversa mínimo cuadrática de A se puede expresar en la forma de B para ciertas F y G.

Demostración. La demostración es consecuencia directa de la proposición VI.3.3. ∎

4. Sistemas de ecuaciones lineales (II). Mínimos cuadrados.

Dados $A \in \mathcal{M}_{m\times n}(\mathbb{R})$ y $\mathbf{b} \in \mathbb{R}^m$, consideramos el sistema de ecuaciones lineales

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

con m ecuaciones y n incógnitas.

El teorema de Rouche-Fröbenius es útil para determinar si el sistema de ecuaciones lineales $A\mathbf{x} = \mathbf{b}$ es compatible, pero no nos dice cómo calcular una solución del sistema cuando es compatible. El siguiente resultado proporciona una condición necesaria y suficiente alternativa de compatibilidad usando una inversa generalizada, A^-, de A. Una consecuencia obvia de este resultado es que cuando el sistema $A\mathbf{x} = \mathbf{b}$ sea compatible, entonces una solución suya será $\mathbf{x} = A^-\mathbf{b}$.

Proposición VI.4.1. *El sistema de ecuaciones $A\mathbf{x} = \mathbf{b}$ es compatible si, y sólo si, para alguna inversa generalizada, A^-, de A se cumple que*

$$AA^-\mathbf{b} = \mathbf{b};$$

en cuyo caso, $\hat{\mathbf{x}} = A^-\mathbf{b}$ es una solución particular.

Demostración. En primer lugar, supongamos que el sistema es compatible y sea $\hat{\mathbf{x}}$ una solución, es decir, $\mathbf{b} = A\hat{\mathbf{x}}$. Multiplicando esta igualdad a izquierda por AA^-, donde A^- es una inversa generalizada de A, se obtiene que

$$AA^-\mathbf{b} = AA^-A\hat{\mathbf{x}} = A\hat{\mathbf{x}} = \mathbf{b},$$

como queríamos probar. Recíprocamente, supongamos que para una inversa generalizada, A^-, de A se tiene que $AA^-\mathbf{b} = \mathbf{b}$. Si $\hat{\mathbf{x}} = A^-\mathbf{b}$, entonces

$$A\hat{\mathbf{x}} = AA^-\mathbf{b} = \mathbf{b};$$

por tanto, $\hat{\mathbf{x}} = A^-\mathbf{b}$, es una solución, y el sistema es compatible. ∎

Nota VI.4.2. Supongamos que B y C son inversas generalizadas de A, por lo tanto $ABA = ACA = A$. Además, supongamos que B verifica la condición de compatibilidad de la proposición VI.4.1, es decir, $AB\mathbf{b} = \mathbf{b}$. Entonces, C verifica la misma condición ya que

$$AC\mathbf{b} = AC(AB\mathbf{b}) = (ACA)B\mathbf{b} = AB\mathbf{b} = \mathbf{b}.$$

Por tanto, para usar la proposición VI.4.1, solamente hay que verificar la condición para una inversa generalizada de A, sin importar qué inversa generalizada estemos usando.

Ejemplo VI.4.3. Consideremos el sistema de ecuaciones $A\mathbf{x} = \mathbf{b}$, donde

$$A = \begin{pmatrix} 1 & 1 & 1 & 2 \\ 1 & 0 & 1 & 0 \\ 2 & 1 & 2 & 2 \end{pmatrix} \quad \text{y} \quad \mathbf{b} = \begin{pmatrix} 3 \\ 2 \\ 5 \end{pmatrix}.$$

Según vimos en el ejemplo VI.3.2, una inversa generalizada de A es

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

Usando esta inversa generalizada observamos que

$$AA^-\mathbf{b} = \begin{pmatrix} 1 & 1 & 1 & 2 \\ 1 & 0 & 1 & 0 \\ 2 & 1 & 2 & 2 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 1/2 \end{pmatrix} \begin{pmatrix} 3 \\ 2 \\ 5 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} 3 \\ 2 \\ 5 \end{pmatrix} = \begin{pmatrix} 3 \\ 2 \\ 5 \end{pmatrix}.$$

Por tanto, una solución particular del sistema de ecuaciones $A\mathbf{x} = \mathbf{b}$ es

$$A^-\mathbf{b} = \begin{pmatrix} 0 \\ 0 \\ 2 \\ 1/2 \end{pmatrix}$$

No obstante, esta no es la única inversa generalizada de A. Por ejemplo, la inversa de Moore-Penrose de A es

$$A^+ = \begin{pmatrix} -1/6 & 1/3 & 1/6 \\ 1/5 & -1/5 & 0 \\ -1/6 & 1/3 & 1/6 \\ 2/5 & -2/5 & 0 \end{pmatrix}.$$

Según lo expresado en la nota VI.4.2, si una inversa generalizada de A verifica la condición de compatibilidad de la proposición VI.4.1, todas las inversas generalizadas de A la verifican. Por consiguiente,

$$A^+\mathbf{b} = \begin{pmatrix} -1/6 & 1/3 & 1/6 \\ 1/5 & -1/5 & 0 \\ -1/6 & 1/3 & 1/6 \\ 2/5 & -2/5 & 0 \end{pmatrix} \begin{pmatrix} 3 \\ 2 \\ 5 \end{pmatrix} = \begin{pmatrix} 1 \\ 1/5 \\ 1 \\ 2/5 \end{pmatrix}$$

es otra solución del sistema de ecuaciones.

Podemos considerar que el sistema de ecuaciones $A\mathbf{x} = \mathbf{b}$ es un caso particular de sistemas de ecuaciones lineales de la forma $AXC = B$ con $B \in \mathcal{M}_{m\times q}(\mathbb{R})$, $C \in \mathcal{M}_{p\times q}(\mathbb{R})$ y, por tanto, X será una matriz de incógnitas de orden $n \times p$. El siguiente resultado da una condición necesaria y suficiente para que exista una solución X.

Proposición VI.4.4. *Sea $A \in \mathcal{M}_{m\times n}(\mathbb{R})$, $B \in \mathcal{M}_{m\times q}(\mathbb{R})$ y $C \in \mathcal{M}_{p\times q}(\mathbb{R})$. El sistema de ecuaciones*

$$AXC = B,$$

es compatible si, y sólo si, para algunas inversas generalizadas A^- y C^-, se cumple que

$$AA^-BC^-C = B, \tag{VI.4.11}$$

en cuyo caso, $\hat{X} = A^-BC^-$ es una solución particular.

Demostración. Supongamos que el sistema es compatible y que la matriz $\hat{X}$ es una de sus soluciones, por tanto $B = A\hat{X}C$. Multiplicando a izquierda por AA^- y a derecha por C^-C, donde A^- y C^- son inversas generalizadas de A y C, obtenemos que

$$AA^-BC^-C = AA^-A\hat{X}CC^-C = A\hat{X}C = B.$$

Recíprocamente, si A^- y C^- cumplen la condición de compatibilidad, definimos $\hat{X} = A^-BC^-$, y observamos que $\hat{X}$ es una solución del sistema. ∎

Usando un argumento similar al de la nota VI.4.2, podemos comprobar que si la condición de compatibilidad (VI.4.11) se verifica para una elección particular de A^- y B^-, entonces se cumple para todas las inversas generalizadas de A y B. En consecuencia, la condición de compatibilidad (VI.4.11) es independiente de la elección de las inversas generalizadas de A y B.

Hemos visto que si un sistema de ecuaciones $A\mathbf{x} = \mathbf{b}$ es compatible, entonces $\mathbf{x} = A^-\mathbf{b}$ es una solución, independientemente de la elección de la inversa generalizada A^-. Por tanto, si A^- varía según la elección de A, entonces nuestro sistema tiene más de una solución (véase el ejemplo VI.4.3). El siguiente resultado da una expresión general para todas las soluciones de un sistema de ecuaciones.

Teorema VI.4.5. *Sean $A \in \mathcal{M}_{m\times n}(\mathbb{R})$ y $\mathbf{b} \in \mathcal{M}_{m\times 1}(\mathbb{R})$ tales que el sistema de ecuaciones $A\mathbf{x} = \mathbf{b}$ es compatible, y A^- una inversa generalizada de A. Entonces, para cada $\mathbf{y} \in \mathbb{R}^n$,*

$$\mathbf{x}_\mathbf{y} = A^-\mathbf{b} + (I_n - A^-A)\mathbf{y} \tag{VI.4.12}$$

es una solución del sistema, y para cualquier solución, $\hat{\mathbf{x}}$, existe $\mathbf{y} \in \mathbb{R}^n$ tal que $\hat{\mathbf{x}} = \mathbf{x}_\mathbf{y}$.

Demostración. Como $A\mathbf{x} = \mathbf{b}$ es compatible, por la proposición VI.4.1, $AA^-\mathbf{b} = \mathbf{b}$, entonces

$$A\mathbf{x}_\mathbf{y} = AA^-\mathbf{b} + A(I_n - A^-A)\mathbf{y} = \mathbf{b} + (A - AA^-A)\mathbf{y} = \mathbf{b},$$

pues $AA^-A = A$. Luego, $\mathbf{x}_\mathbf{y}$ es una solución independientemente de la elección de $\mathbf{y} \in \mathbb{R}^n$. Por otra parte, si $\hat{\mathbf{x}}$ es una solución arbitraria, entonces $A^-A\hat{\mathbf{x}} = A^-\mathbf{b}$, pues $A\hat{\mathbf{x}} = \mathbf{b}$. Consecuentemente,

$$A^-\mathbf{b} + (I_n - A^-A)\hat{\mathbf{x}} = A^-\mathbf{b} + \hat{\mathbf{x}} - A^-A\hat{\mathbf{x}} = \hat{\mathbf{x}},$$

luego $\hat{\mathbf{x}} = \mathbf{x}_{\hat{\mathbf{x}}}$. ∎

Ejemplo VI.4.6. Para el sistema de ecuaciones estudiado en el ejemplo VI.4.3, tenemos que

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

usando la primera de las dos inversas generalizadas dadas en el ejemplo. Consecuentemente, una solución del sistema de ecuaciones es

$$\mathbf{x}_\mathbf{y} = A^-\mathbf{b} + (I_4 - A^-A)\mathbf{y} = \begin{pmatrix} 0 \\ 0 \\ 2 \\ 1/2 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & -1/2 & 0 & 0 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ 2 - y_1 \\ 1/2 - y2/2 \end{pmatrix},$$

donde $\mathbf{y} = (y_1, y_2, y_3, y_4)^\mathrm{t}$ es un vector arbitrario.

Una consecuencia inmediata del teorema VI.4.5 es la siguiente:

Corolario VI.4.7. *Sean $A \in \mathcal{M}_{m\times n}(\mathbb{R})$ y $\mathbf{b} \in \mathcal{M}_{m\times 1}(\mathbb{R})$ tales que el sistema de ecuaciones $A\mathbf{x} = \mathbf{b}$ es compatible. El sistema tiene solución única si, y sólo si, $A^-A = I_n$, para cualquier inversa generalizada A^- de A.*

Demostración. Nótese que $\hat{\mathbf{x}} = A^-\mathbf{b}$ es la única solución del sistema $A\mathbf{x} = \mathbf{b}$ si, y sólo si, $\hat{\mathbf{x}} = \mathbf{x}_\mathbf{y}$, para todo $\mathbf{y} \in \mathbb{R}^n$, con $\mathbf{x}_\mathbf{y}$ definido como en (VI.4.12). En otras palabras, la solución es única si, y sólo si, $(I_n - A^-A)\mathbf{y} = \mathbf{0}$, para todo $\mathbf{y} \in \mathbb{R}^n$, es decir, si, y sólo si, $I_n - A^-A = 0$. ∎

Corolario VI.4.8. *Sean $A \in \mathcal{M}_{m\times n}(\mathbb{R})$ y $\mathbf{b} \in \mathcal{M}_{m\times 1}(\mathbb{R})$ tales que el sistema de ecuaciones $A\mathbf{x} = \mathbf{b}$ es compatible. El sistema tiene solución única si, y sólo si,* $\mathrm{rg}(A) = n$.

Demostración. Basta tener en cuenta la proposición VI.3.5(g) y el corolario VI.4.7. ∎

Soluciones aproximadas mínimo cuadráticas de sistemas de ecuaciones lineales.

Sean $A \in \mathcal{M}_{m\times n}(\mathbb{R})$ y $\mathbf{b} \in \mathbb{R}^m$ tales que $\mathbf{b} \notin \mathrm{im}(A)$. Según vimos en el tema III, el sistema de ecuaciones $A\mathbf{x} = \mathbf{b}$ es incompatible. Sin embargo, en algunas situaciones puede ser interesante conocer algún vector o un conjunto de vectores que estén "cerca" de verificar las ecuaciones. Si $\bar{\mathbf{x}} \in \mathbb{R}^n$ fuese una ellas, entonces $\bar{\mathbf{x}}$ verificará aproximadamente las ecuaciones de nuestro sistema si $A\bar{\mathbf{x}} - \mathbf{b}$ es próximo a $\mathbf{0}$. Si usamos la distancia para el producto escalar usual de $\mathbb{R}^m$, entonces la distancia al cuadrado de $A\bar{\mathbf{x}} - \mathbf{b}$ a $\mathbf{0}$ es la suma al cuadrado de sus componentes, esto es, $(A\bar{\mathbf{x}} - \mathbf{b})^{\mathrm{t}}(A\bar{\mathbf{x}} - \mathbf{b})$ en coordenadas respecto de la base usual de $\mathbb{R}^m$. Cualquier vector que minimice esta suma de cuadrados se llama solución aproximada mínimo cuadrática.

Definición VI.4.9. Sean $A \in \mathcal{M}_{m\times n}(\mathbb{R})$ y $\mathbf{b} \in \mathbb{R}^m$. Se dice que $\bar{\mathbf{x}} \in \mathbb{R}^n$ es una **solución (aproximada) mínimo cuadrática** del sistema de ecuaciones $A\mathbf{x} = \mathbf{b}$ si cumple la desigualdad

$$(A\bar{\mathbf{x}} - \mathbf{b})^{\mathrm{t}}(A\bar{\mathbf{x}} - \mathbf{b}) \leq (A\mathbf{x} - \mathbf{b})^{\mathrm{t}}(A\mathbf{x} - \mathbf{b}), \tag{VI.4.13}$$

para todo $\mathbf{x} \in \mathbb{R}^n$.

Nota VI.4.10. Obsérvese que si $\bar{\mathbf{x}} \in \mathbb{R}^n$ es una solución aproximada mínimo cuadrática del sistema de ecuaciones $A\mathbf{x} = \mathbf{b}$, entonces

$$\begin{aligned}\mathrm{d}(\mathbf{b}, \mathrm{im}(A))^2 &= \min\{\mathrm{d}(\mathbf{b}, A\mathbf{x})^2 \mid \mathbf{x} \in \mathbb{R}^n\} = \min\{\|A\mathbf{x} - \mathbf{b}\|^2 \mid \mathbf{x} \in \mathbb{R}^n\} \\ &\overset{coord.}{=} \min\{(A\mathbf{x} - \mathbf{b})^{\mathrm{t}}(A\mathbf{x} - \mathbf{b}) \mid \mathbf{x} \in \mathbb{R}^n\} = (A\bar{\mathbf{x}} - \mathbf{b})^{\mathrm{t}}(A\bar{\mathbf{x}} - \mathbf{b}) \\ &\overset{coord.}{=} \|A\bar{\mathbf{x}} - \mathbf{b}\|^2 = \mathrm{d}(\mathbf{b}, A\bar{\mathbf{x}})^2,\end{aligned}$$

donde las igualdades indicadas lo son en coordenadas respecto de la base usual de $\mathbb{R}^m$.

Según vimos en el tema V, la distancia de un vector $\mathbf{v} \in V$ a un subespacio vectorial L de V se alcanza en la proyección ortogonal de $\mathbf{v}$ sobre L, esto es, en el único vector $\mathbf{v}_1 \in L$ tal que $\mathbf{v} - \mathbf{v}_1 \in L^{\perp}$. Así, volviendo al problema que nos ocupa, si $\mathbf{b}_1$ es la proyección ortogonal de $\mathbf{b}$ sobre $\mathrm{im}(A)$, las soluciones aproximadas mínimo cuadráticas son las del sistema de ecuaciones lineales $A\mathbf{x} = \mathbf{b}_1$.

Proposición VI.4.11. *Sean $A \in \mathcal{M}_{m\times n}(\mathbb{R})$ y $\mathbf{b} \in \mathbb{R}^m$. Las soluciones aproximadas mínimo cuadráticas del sistema $A\mathbf{x} = \mathbf{b}$ son precisamente las soluciones del sistema $A\mathbf{x} = AA^+\mathbf{b}$.*

Demostración. Según la nota VI.4.10, las soluciones aproximadas mínimo cuadráticas de $A\mathbf{x} = \mathbf{b}$ son las soluciones del sistema de ecuaciones $A\mathbf{x} = \mathbf{b}_1$ donde $\mathbf{b}_1$ es la proyección ortogonal de $\mathbf{b}$ sobre $\mathrm{im}(A)$. Como, por el teorema VI.2.4, $AA^+\mathbf{b} = \mathbf{b}_1$, tenemos que $\bar{\mathbf{x}}$ es solución aproximada mínimo cuadrática del sistema $A\mathbf{x} = \mathbf{b}$ si, y sólo si, es solución del sistema $A\mathbf{x} = AA^+\mathbf{b}$. ∎

Corolario VI.4.12. *Sean $A^\square \in \mathcal{M}_{n\times m}(\mathbb{R})$ una inversa mínimo cuadrática de $A \in \mathcal{M}_{m\times n}(\mathbb{R})$, y $\mathbf{b} \in \mathbb{R}^m$. Entonces $\bar{\mathbf{x}} = A^\square\mathbf{b}$ es una solución aproximada mínimo cuadrática del sistema $A\mathbf{x} = \mathbf{b}$.*

Demostración. Es una consecuencia inmediata de la proposición VI.4.11, sin mas que tener en cuenta que, por la proposición VI.3.9, $AA^+ = AA^\square$ para cualquier inversa mínimo cuadrática $A^\square$ de A. ∎

Ejemplo VI.4.13. Consideremos el sistema de ecuaciones $A\mathbf{x} = \mathbf{b}$ con

$$A = \begin{pmatrix} 1 & 1 & 2 \\ 1 & 0 & 1 \\ 1 & 1 & 2 \\ 2 & 0 & 2 \end{pmatrix} \quad \text{y} \quad \mathbf{b} = \begin{pmatrix} 4 \\ 1 \\ 6 \\ 5 \end{pmatrix}.$$

Una inversa mínimo cuadrática de A es

$$A^\square = \frac{1}{10}\begin{pmatrix} -1/6 & 1/5 & -1/6 & 2/5 \\ 1/3 & -1/5 & 1/3 & -2/5 \\ 1/6 & 0 & 1/6 & 0 \end{pmatrix}.$$

Como

$$A\,A^\square\mathbf{b} = \frac{1}{10}\begin{pmatrix} 5 & 0 & 5 & 0 \\ 0 & 2 & 0 & 4 \\ 5 & 0 & 5 & 0 \\ 0 & 4 & 0 & 8 \end{pmatrix}\begin{pmatrix} 4 \\ 1 \\ 6 \\ 5 \end{pmatrix} = \frac{1}{5}\begin{pmatrix} 1 \\ 1 \\ 1 \\ 22 \end{pmatrix} \neq \mathbf{b},$$

de la proposición VI.4.1 se sigue que el sistema es incompatible. Una solución aproximada mínimo cuadrática es

$$A^\square\mathbf{b} = \begin{pmatrix} -1/6 & 1/5 & -1/6 & 2/5 \\ 1/3 & -1/5 & 1/3 & -2/5 \\ 1/6 & 0 & 1/6 & 0 \end{pmatrix}\begin{pmatrix} 4 \\ 1 \\ 6 \\ 5 \end{pmatrix} = \frac{1}{15}\begin{pmatrix} 8 \\ 17 \\ 25 \end{pmatrix}.$$

Veamos ahora que el recíproco del corolario VI.4.12 también es cierto.

Corolario VI.4.14. *Sean $A \in \mathcal{M}_{m\times n}(\mathbb{R})$ y $\mathbf{b} \in \mathbb{R}^m$. Si $\bar{\mathbf{x}}$ es una solución aproximada mínimo cuadrática del sistema $A\mathbf{x} = \mathbf{b}$, entonces existe una inversa mínimo cuadrática $A^\square$ de A tal que $\bar{\mathbf{x}} = A^\square\mathbf{b}$.*

Demostración. Por la proposición VI.4.11, $\bar{\mathbf{x}}$ es una solución del sistema de ecuaciones $A\mathbf{x} = AA^+\mathbf{b}$. Luego, por la proposición VI.4.1, existe una inversa generalizada A^- de A tal que $\bar{\mathbf{x}} = A^-AA^+\mathbf{b}$. Una simple comprobación demuestra que A^-AA^+ es una inversa mínimo cuadrática de A. ∎

Nótese que de los corolarios VI.4.12 y VI.4.14 se sigue que $\bar{\mathbf{x}}$ es solución aproximada mínimo cuadrática de $A\mathbf{x} = \mathbf{b}$ si, y sólo si,

$$A\bar{\mathbf{x}} = AA^\square\mathbf{b}, \tag{VI.4.14}$$

para alguna inversa mínimo cuadrática de A. Sin embargo, como, por la proposición VI.3.9, $AA^\square = AA^+$, para toda inversa mínimo cuadrática $A^\square$ de A, se sigue que la igualdad es independiente de la inversa mínimo cuadrática que elijamos.

Teorema VI.4.15. *Sean $A \in \mathcal{M}_{m\times n}(\mathbb{R})$, $\mathbf{b} \in \mathcal{M}_{m\times 1}(\mathbb{R})$ y $A^\square$ una inversa mínimo cuadrática de A. Entonces, para cada $\mathbf{y} \in \mathbb{R}^n$,*

$$\bar{\mathbf{x}}_\mathbf{y} = A^\square\mathbf{b} + (I_n - A^\square A)\mathbf{y} \tag{VI.4.15}$$

es una solución aproximada mínimo cuadrática del sistema, y para cualquier solución aproximada mínimo cuadrática, $\bar{\mathbf{x}}$, existe $\mathbf{y} \in \mathbb{R}^n$ tal que $\bar{\mathbf{x}} = \bar{\mathbf{x}}_\mathbf{y}$.

Demostración. Usando que, por la proposición VI.3.9, $AA^+ = AA^\square$, se comprueba fácilmente que $\bar{\mathbf{x}}_\mathbf{y}$ es una solución aproximada mínimo cuadrática de $A\mathbf{x} = \mathbf{b}$. Recíprocamente, si $\bar{\mathbf{x}}$ es una solución aproximada mínimo cuadrática de $A\mathbf{x} = \mathbf{b}$, entonces

$$A\bar{\mathbf{x}} = AA^\square\mathbf{b},$$

siendo $A^\square$ una inversa generalizada (cualquiera) de A. Ahora, basta tomar $\mathbf{y} = \bar{\mathbf{x}} - A^\square\mathbf{b}$ y comprobar que $\bar{\mathbf{x}} = \bar{\mathbf{x}}_\mathbf{y}$. ∎

Ejemplo VI.4.16. Calculemos todas las soluciones aproximadas mínimo cuadrática del sistema de ecuaciones del ejemplo VI.4.13. En primer lugar, observamos que

$$A^\square A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix},$$

de tal forma que

$$\begin{aligned}\bar{\mathbf{x}}_{\mathbf{y}} &= A^{\square}\mathbf{b} + (I_3 - A^{\square}A)\mathbf{y} \\ &= \frac{1}{10}\begin{pmatrix} 0 & 2 & 0 & 4 \\ 5 & -2 & 5 & -4 \\ 0 & 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} 4 \\ 1 \\ 6 \\ 5 \end{pmatrix} + \begin{pmatrix} 0 & 0 & -1 \\ 0 & 0 & -1 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} \\ &= \begin{pmatrix} 2{,}2 - y_3 \\ 2{,}8 - y_3 \\ y_3 \end{pmatrix}\end{aligned}$$

es una solución aproximada mínimo cuadrática para cada $y_3 \in \mathbb{R}$.

Terminamos esta sección calculando la **solución óptima mínimo cuadrática** de un sistema de ecuaciones lineales, que no es otra cosa que la solución aproximada mínimo cuadrática de norma (euclídea) mínima.

Corolario VI.4.17. *Sean $A \in \mathcal{M}_{m\times n}(\mathbb{R})$ y $\mathbf{b} \in \mathcal{M}_{m\times 1}(\mathbb{R})$. La solución óptima mínimo cuadrática del sistema de ecuaciones $A\mathbf{x} = \mathbf{b}$ es*

$$\mathbf{x}^+ = A^+\mathbf{b}.$$

Demostración. Como A^+ es, en particular, una inversa mínimo cuadrática de A, por el teorema VI.4.15, se tiene que todas las soluciones aproximadas mínimo cuadráticas de $A\mathbf{x} = \mathbf{b}$ son de la forma $\bar{\mathbf{x}}_{\mathbf{y}} = A^+\mathbf{b} + (I_n - A^+A)\mathbf{y}$, para algún $\mathbf{y} \in \mathbb{R}^n$. Por otra parte, al ser $A^+\mathbf{b}$ ortogonal a $(I_n - A^+A)\mathbf{y}$, del teorema de Pitágoras se sigue que

$$\|\bar{\mathbf{x}}_{\mathbf{y}}\|^2 = \|A^+\mathbf{b}\|^2 + \|(I_n - A^+A)\mathbf{y}\|^2 \geq \|A^+\mathbf{b}\|^2 = \|\mathbf{x}^+\|^2$$

y la igualdad se alcanza si, y sólo si $(I_n - A^+A)\mathbf{y} = \mathbf{0}$, luego $\mathbf{x}^+ = A^+\mathbf{b}$. ∎

Ejercicios del tema VI

Ejercicio 1. Calcular la descomposición en valores singulares (larga y corta) de la matriz

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

Ejercicio 2. Sea $A \in \mathcal{M}_{m\times n}(\mathbb{R})$.

1. Probar que los valores singulares de A son los mismos que los de A^{t}.
2. Probar que los valores singulares de A son los mismos que los de UAV, si $U \in \mathcal{M}_m(\mathbb{R})$ y $V \in \mathcal{M}_n(\mathbb{R})$ son matrices ortogonales.
3. Si $\alpha \neq 0$ es un escalar, ¿cómo son los valores singulares de αA en comparación con los de A?

Ejercicio 3. Sea $A \in \mathcal{M}_{m\times n}(\mathbb{R})$. Si A tiene rango r y su descomposición en valores singulares (larga) es

$$A = P \begin{pmatrix} \Delta & 0 \\ 0 & 0 \end{pmatrix} Q^{\mathrm{t}},$$

probar que, si $\mathbf{v}_i$ y $\mathbf{u}_i$ denotan, respectivamente, la columna i-ésima de P y Q, entonces, $\mathbf{v}_i = (1/\sigma_i)A^{\mathrm{t}}\mathbf{u}_i,\ i = 1, \ldots, r$.

Ejercicio 4. Usar la proposición VI.2.6(h) para calcular la inversa de Moore-Penrose de

$$\begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ 2 & 0 & 1 \end{pmatrix}.$$

Ejercicio 5. Probar que si $A^+ \in \mathcal{M}_{n\times m}(\mathbb{R})$ es la inversa de Moore-Penrose de $A \in \mathcal{M}_{m\times n}(\mathbb{R})$, entonces $(A^+)^2$ es la inversa de Moore-Penrose de A^2.

Ejercicio 6. Consideremos la matriz

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

1. Calcular la inversa generalizada de Moore-Penrose de AA^{t}, y usar la proposición VI.2.6(g) para hallar A^+.
2. Usar A^+ para calcular la matriz de proyección ortogonal de $\mathbb{R}^n$ sobre $\mathrm{im}(A)$ y de $\mathbb{R}^m$ sobre $\mathrm{im}(A^{\mathrm{t}})$.

Ejercicio 7. Sea $A \in \mathcal{M}_n(\mathbb{R})$. Probar que si A es simétrica, entonces

1. A^+ es simétrica.
2. $AA^+ = A^+A$.
3. $A^+ = A$, si A es idempotente.

Demostrar que el recíproco de 3. no es cierto en general. Es decir, encontrar una matriz simétrica A tal que $A^+ = A$ que no sea idempotente.

Ejercicio 8. Sea $A \in \mathcal{M}_{m\times n}(\mathbb{R})$. Probar que si $\mathrm{rg}(A) = 1$, entonces $A^+ = \alpha^{-1}A^{\mathrm{t}}$, donde $\alpha = \mathrm{tr}(A^+A)$.

Ejercicio 9. Sean $A \in \mathcal{M}_{m\times n}(\mathbb{R})$ y $B \in \mathcal{M}_{n\times m}(\mathbb{R})$. Probar que si A y B son definidas positivas, entonces

$$ABA^{\mathrm{t}}(ABA^{\mathrm{t}})^+A = A.$$

Ejercicio 10. Sea $A \in \mathcal{M}_{m\times n}(\mathbb{R})$. Probar que

1. $AB = \mathbf{0}$ si, y sólo si, $B^+A^+ = \mathbf{0}$, con $B \in \mathcal{M}_{n\times p}(\mathbb{R})$.
2. $A^+B = \mathbf{0}$ si, y sólo si, $A^{\mathrm{t}}B = \mathbf{0}$, con $B \in \mathcal{M}_{m\times p}(\mathbb{R})$.

Ejercicio 11. Sea $A \in \mathcal{M}_{m\times n}(\mathbb{R})$ simétrica y de rango r. Probar que si A tiene un autovalor λ no nulo de multiplicidad r, entonces $A^+ = \lambda^{-2}A$.

Ejercicio 12. Sean $A \in \mathcal{M}_{m\times n}(\mathbb{R})$ y $B \in \mathcal{M}_{n\times p}(\mathbb{R})$. Probar que si B tiene rango pleno por filas (es decir, $\mathrm{rg}(B) = n$), entonces

$$AB(AB)^+ = AA^+.$$

Ejercicio 13. Sean $A \in \mathcal{M}_{m\times n}(\mathbb{R})$ y $B \in \mathcal{M}_{m\times n}(\mathbb{R})$ simétricas y semidefinidas positivas tales que $A - B$ también es semidefinida positiva. Probar que $B^+ - A^+$ es semidefinida positiva si, y sólo si, $\mathrm{rg}(A) = \mathrm{rg}(B)$.

Ejercicio 14. Sean $A \in \mathcal{M}_{m\times n}(\mathbb{R})$ y $B \in \mathcal{M}_{n\times m}(\mathbb{R})$. Probar que $(AB)^+ = B^+A^+$ si $A^{\mathrm{t}}ABB^{\mathrm{t}} = BB^{\mathrm{t}}A^{\mathrm{t}}A$.

Ejercicio 15. Calcular la inversa de Moore-Penrose de

$$\begin{pmatrix} 2 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 2 & 0 \\ 0 & 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 4 \end{pmatrix}.$$

Ejercicio 16. Consideremos la matriz diagonal $A = \mathrm{diag}(0, 2, 3)$.

1. Hallar una inversa generalizada de A de rango 2.
2. Hallar una inversa generalizada de A de rango 3 y que sea diagonal.
3. Hallar una inversa generalizada de A que no sea diagonal.

Ejercicio 17. Sea $A \in \mathcal{M}_n(\mathbb{R})$ una matriz divida por bloques de la siguiente manera

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix},$$

con $A_{11} \in \mathcal{M}_r(\mathbb{R})$. Probar que si $\mathrm{rg}(A_{11}) = \mathrm{rg}(A) = r$, entonces

$$\begin{pmatrix} A_{11}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$$

es una inversa generalizada de A.

Ejercicio 18. Sean $A \in \mathcal{M}_{m\times n}(\mathbb{R})$ y A^- una inversa generalizada de A. Probar que:

1. $AA^-,\ A^-A,\ I_n - A^-A$ e $I_m - AA^-$ son idempotentes.
2. $\mathrm{rg}(I_n - A^-A) = n - \mathrm{rg}(A)$ y $\mathrm{rg}(I_m - AA^-) = m - \mathrm{rg}(A)$.

Ejercicio 19. Sean $A \in \mathcal{M}_{m\times n}(\mathbb{R})$ y $B \in \mathcal{M}_{n\times p}(\mathbb{R})$. Probar que B^-A^- será una inversa generalizada de AB para cualquier elección de A^- y B^- si $\mathrm{rg}(B) = n$.

Ejercicio 20. Sean $A \in \mathcal{M}_{m\times n}(\mathbb{R})$ y $B \in \mathcal{M}_{n\times p}(\mathbb{R})$. Probar que para cualquier elección de A^- y B^-, B^-A^- es una inversa generalizada de AB si, y sólo si, A^-BB^- es idempotente.

Ejercicio 21. Probar que la matriz B es una inversa generalizada de A si, y sólo si, AB es idempotente y $\mathrm{rg}(A) = \mathrm{rg}(AB)$.

Ejercicio 22. Sean $A \in \mathcal{M}_{m\times n}(\mathbb{R})$ y $B \in \mathcal{M}_{n\times m}(\mathbb{R})$. Probar que B es la inversa de Moore-Penrose de A si, y sólo si, B es una inversa mínimo cuadrática de A y A es una inversa mínimo cuadrática de B.

Ejercicio 23. Sea $A \in \mathcal{M}_{m\times n}(\mathbb{R})$. Si A tiene rango $r > 0$ y

$$A = P \begin{pmatrix} \Delta & 0 \\ 0 & 0 \end{pmatrix} Q^{\mathrm{t}}$$

es una descomposición en valores singulares de A, entonces para cada $F \in \mathcal{M}_{(n-r)\times r}(\mathbb{R})$ la matriz

$$B = Q \begin{pmatrix} \Delta^{-1} & 0 \\ F & 0 \end{pmatrix} P^{\mathrm{t}}$$

es una mínimo cuadrática de A de la forma $(A^{\mathrm{t}}A)^-A^{\mathrm{t}}$ y cualquier inversa mínimo cuadrática de A de la forma $(A^{\mathrm{t}}A)^-A^{\mathrm{t}}$ se puede expresar en la forma de B para cierta F.

Ejercicio 24. Sean $A \in \mathcal{M}_{m\times n}(\mathbb{R})$ y $(AA^{\mathrm{t}})^-$ y $(A^{\mathrm{t}}A)^-$ inversas generalizadas arbitrarias de AA^{t} y $A^{\mathrm{t}}A$, respectivamente. Probar que

$$A^+ = A^{\mathrm{t}}(AA^{\mathrm{t}})^-A(A^{\mathrm{t}}A)^-A^{\mathrm{t}}.$$

Ejercicio 25. Sea $A\mathbf{x} = \mathbf{b}$ un sistema de ecuaciones compatible. Probar que si B es una inversa generalizada de A, entonces $\mathbf{x} = B\mathbf{b}$ es una solución, y para cualquier solución $\hat{\mathbf{x}}$, existe una inversa generalizada B de A, tal que $\hat{\mathbf{x}} = B\mathbf{b}$.

Ejercicio 26. Sea $AXC = B$ un sistema de ecuaciones compatible, con $A \in \mathcal{M}_{m\times n}(\mathbb{R})$, $B \in \mathcal{M}_{m\times q}(\mathbb{R})$ y $C \in \mathcal{M}_{p\times q}(\mathbb{R})$. Probar que para cualesquiera inversas generalizadas A^- y C^-, y una matriz arbitraria $Y \in \mathcal{M}_{n\times p}(\mathbb{R})$,

$$X_Y = A^-BC^- + Y - A^-AYCC^-$$

es una solución, y para cualquier solución, $\hat{X}$, existe una matriz Y tal que $\hat{X} = X_Y$.

Ejercicio 27. Consideremos el sistema de ecuaciones $A\mathbf{x} = \mathbf{b}$, donde $A \in \mathcal{M}_{4\times 3}(\mathbb{R})$ es la matriz de ejercicio 4 y

$$\mathbf{b} = \begin{pmatrix} 1 \\ 3 \\ -1 \\ 0 \end{pmatrix}.$$

1. Probar que el sistema es compatible.
2. Hallar una solución de este sistema de ecuaciones.
3. ¿Cuántas soluciones linealmente independientes hay?

Ejercicio 28. Consideremos el sistema de ecuaciones $A\mathbf{x} = \mathbf{b}$, donde $A \in \mathcal{M}_{3\times 4}(\mathbb{R})$ es la matriz de ejercicio 3 y

$$\begin{pmatrix} 1 \\ 1 \\ 4 \end{pmatrix}.$$

1. Probar que el sistema de ecuaciones es compatible.
2. Dar la expresión para solución general.
3. Hallar el número r de soluciones linealmente independientes.
4. Dar un conjunto de r soluciones linealmente independientes.

Ejercicio 29. Consideremos el sistema de ecuaciones $AXC = B$, donde $X \in \mathcal{M}_3(\mathbb{R})$ es una matriz de incógnitas y

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

1. Probar que el sistema de ecuaciones es compatible.
2. Hallar la expresión de la solución general de este sistema.

Ejercicio 30. Calcular la solución óptima mínimo cuadrática del siguiente sistema de ecuaciones para todos los valores de $\alpha \in \mathbb{R}$:

$$\begin{pmatrix} 1 & 1 \\ 1 & \alpha \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}.$$

Ejercicio 31. Sea $A \in \mathcal{M}_{m\times n}(\mathbb{R})$ y $\mathbf{b} \in \mathbb{R}^m$. Probar que $\hat{\mathbf{x}}$ es una solución aproximada mínimo cuadrática del sistema $A\mathbf{x} = \mathbf{b}$ si, y sólo si, $\hat{\mathbf{x}}$ forma parte de una solución del sistema ampliado

$$\begin{pmatrix} I_m & A \\ A^{\mathrm{t}} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{y} \\ \hat{\mathbf{x}} \end{pmatrix} = \begin{pmatrix} \mathbf{b} \\ \mathbf{0} \end{pmatrix}$$

No es extraño encontrar problemas de mínimos cuadrados en los que la matriz A es muy grande pero contiene muchos ceros. Para esta situación, el anterior sistema ampliado contendrá menos entradas no nulas que el sistema de ecuaciones normales, y evitará los problemas de memoria que suelen dar los algoritmos de resolución. Además, se evita el cálculo de $A^{\mathrm{t}}A$ que puede producir problemas de mal condicionamiento. (véase la sección 3 del tema VIII).

Ejercicio 32. Consideremos el problema de calcular la solución de norma mínima del problema de mínimos cuadrados mín $\|A\mathbf{x} - b\|^2$, donde

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad \text{y} \quad \mathbf{b} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

Probar que

1. la solución $\hat{\mathbf{x}} = (1, 0)^{\mathrm{t}}$.
2. Consideremos la perturbación de A

   $$E_1 = \begin{pmatrix} 0 & \delta \\ 0 & 0 \end{pmatrix}$$

   donde δ es un número positivo pequeño. Resolver la versión perturbada del problema anterior mín $\|A_1\mathbf{y} - b\|^2$, donde $A_1 = A + E_1$. ¿Qué le ocurre a $\|\hat{\mathbf{x}} - \mathbf{y}\|$ cuando δ se aproxima a cero?
3. Ahora consideremos la perturbación de A

   $$E_2 = \begin{pmatrix} 0 & 0 \\ 0 & \delta \end{pmatrix}$$

   donde δ es un número positivo pequeño. Resolver la versión perturbada del problema anterior mín $\|A_2\mathbf{z} - b\|^2$, donde $A_2 = A + E_2$. ¿Qué le ocurre a $\|\hat{\mathbf{x}} - \mathbf{z}\|$ cuando δ se aproxima a cero?

TEMA VII

Derivación matricial

El cálculo diferencial tiene multitud de aplicaciones en Estadística. Por ejemplo, los procesos de estimación tales como el método de máxima verosimilitud o el método de mínimos cuadrados usan las propiedades de optimización de las derivadas, mientras que el llamado método delta para obtener la distribución asintótica de una función de variables aleatorias usa la primera derivada para obtener una aproximación por una serie de Taylor de primer orden. Estas y otras aplicaciones del cálculo diferencial involucran a menudo vectores y matrices. En este tema, mostraremos algunas de las derivadas matriciales más comúnmente utilizadas en Estadística.

En la primera sección de este tema, introduciremos brevemente algunos operadores matriciales especiales y estudiaremos algunas de sus propiedades. En particular, echaremos un vistazo a un producto de matrices que es diferente del usual. Este producto de matrices, llamado producto de Kronecker, produce una matriz divida por bloques tal que cada bloque es igual a un elemento de la primera matriz por la segunda (este producto ya fue definido a modo de ejemplo en el primer tema). Estrechamente relacionado con el producto Kronecker se halla el operador vec, o vectorización, que transforma matrices en vectores apilando las columnas una encima de otra. En muchas ocasiones, una matriz con una expresión aparentemente complicada se puede escribir de una forma realmente simple sin más que aplicar uno o más de estos operadores matriciales.

Ni que decir tiene que existen otros operadores matriciales, algunos ya conocidos como la suma directa de matrices (véase la sección 3 del tema III), y otros también importantes pero que no estudiaremos en esta asignatura, como por ejemplo el producto de Hadamard de dos matrices que no es más que el producto entrada a entrada de cada una de ellas (véase el capítulo 8 de [**Sch05**]).

El primero de los operadores que estudiamos en esta sección es el producto de Kronecker de matrices. Posteriormente mostramos sus propiedades básicas y su relación con la traza, la inversa, las inversas generalizas y el determinante. La elección de estas propiedades no es casual, ya que serán las que utilicemos para calcular las diferenciales de la funciones matriciales usuales. A continuación estudiamos el operador vec. La vectorización de una matriz consiste en construir un vector apilando las columnas de la matriz una encima de otra, conviene destacar que vec no es más

que una aplicación lineal de $\mathcal{M}_{m\times n}(\mathbb{R})$ en $\mathbb{R}^{mn}$. Las propiedades estudiadas de la vectorización son las que relacionan el operador vec con la traza y el producto de matrices. Terminamos esta sección introduciendo las matrices de conmutación que permiten relacionar la vectorización de una matriz y la de su traspuesta, y establecer la propiedad que relaciona la vectorización con el producto de Kronecker.

La segunda sección es la que da nombre al tema, en ella definimos y estudiamos las primeras propiedades del diferencial de una función matricial de variable matricial. La clave de la definición de diferencial es la vectorización de la función matricial y de la matriz de variables. Así, definimos la diferencial de $F(X)$ en A como la única aplicación lineal $\mathrm{d}F(A)$ tal que $\mathrm{vec}(\mathrm{d}F(A)) = \mathrm{dvec}(F(A))$. Esta estrategia permite reducir el estudio de la diferencial de una función matricial de variable matricial, al estudio de la diferencial de una función vectorial de variable vectorial, y definir la derivada de una función matricial de variable matricial como la derivada de $\mathrm{vec}(F(X))$ respecto de $\mathrm{vec}(X)^{\mathrm{t}}$, es decir, aquella que tiene como entrada (i,j)-ésima a la derivada parcial del entrada i-ésima de $\mathrm{vec}(F(X))$ con respecto a la entrada j-ésima de $\mathrm{vec}(X)$. Conviene advertir que existen otras definiciones de derivada matricial (véanse, por ejemplo, las secciones 3 y 4 de [**MN07**] y la sección 5.4 de [**BS98**]). Nuestra elección resulta útil cuando se está interesado fundamentalmente en aplicar a funciones matriciales resultados matemáticos relativos a funciones vectoriales, como es nuestro caso. El resto de la sección se dedica a las propiedades básica de la diferencial y su relación con algunas de las operaciones matriciales tales como la trasposición, el producto de Kronecker y la traza.

En tema finaliza con el cálculo de las diferenciales y derivadas de algunas funciones escalares y matriciales de variable matricial, por ejemplo, las funciones que a cada matriz le asignan su traza o su determinante, y las funciones que a cada matriz le asignan su inversa o su inversa de Moore-Penrose. Todas las que aparecen en esta sección las diferenciales y derivadas son calculadas con detalle, a excepción de la diferencial de la inversa de Moore-Penrose de la que solamente se muestran sus expresiones.

La bibliografía utilizada para este tema ha sido [**Sch05**] y [**MN07**], principalmente la teoría de los capítulos 8 y 9 del segundo, para la parte correspondiente a la diferenciación matricial y el capítulo 8 de [**Sch05**] para la sección sobre los operadores matriciales.

1. Algunos operadores matriciales

El producto de Kronecker.

Definición VII.1.1. Sean $A = (a_{ij}) \in \mathcal{M}_{m\times n}(\mathbb{R})$ y $B \in \mathcal{M}_{p\times q}(\mathbb{R})$. Se llama **producto de Kronecker** [1] de A por B, y se denota $A \otimes B$, a la matriz por bloques

$$\begin{pmatrix} a_{11}B & a_{12}B & \dots & a_{1n}B \\ a_{21}B & a_{22}B & \dots & a_{2n}B \\ \vdots & \vdots & & \vdots \\ a_{m1}B & a_{m2}B & \dots & a_{mn}B \end{pmatrix} \in \mathcal{M}_{mp\times nq}(\mathbb{R}). \tag{VII.1.1}$$

Este producto es conocido más concretamente como producto de Kronecker a derecha, siendo esta la definición más común del producto de Kronecker.

A diferencia de la multiplicación de matrices el producto de Kronecker $A \otimes B$ se puede definir independientemente de los órdenes de A y B. Sin embargo, al igual que la multiplicación, el producto de Kronecker no es, general, conmutativo.

Ejemplo VII.1.2. Sean

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

Por un lado se tiene que

$$A \otimes B = \begin{pmatrix} 0B & 1B & 2B \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 & 2 & 2 & 4 \\ 0 & 0 & 3 & 4 & 6 & 8 \end{pmatrix};$$

mientras que por otro

$$B \otimes A = \begin{pmatrix} 1A & 2A \\ 3A & 4A \end{pmatrix} = \begin{pmatrix} 0 & 1 & 2 & 0 & 2 & 4 \\ 0 & 3 & 6 & 0 & 4 & 8 \end{pmatrix};$$

A pesar de que el producto de Kronecker no es conmutativo, se puede demostrar que existen matrices de permutación P y Q tales que $P^{\mathrm{t}}(A \otimes B)Q = B \otimes A$; tal y como demostraremos en la proposición VII.1.20.

A continuación enunciamos algunas propiedades básicas del producto de Kronecker.

Proposición VII.1.3. *Sea A, B y C matrices cualesquiera con coeficientes en $\mathbb{R}$ y $\mathbf{a} \in \mathbb{R}^m$ y $\mathbf{b} \in \mathbb{R}^n$.*

[1] Sea $T : \mathbb{R}^n \to \mathbb{R}^m$ la aplicación lineal cuya matriz respecto de las bases usuales $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ y $\{\mathbf{e}'_1, \dots, \mathbf{e}'_m\}$ de $\mathbb{R}^n$ y $\mathbb{R}^m$, respectivamente, es A, y sea $S : \mathbb{R}^p \to \mathbb{R}^q$ la aplicación lineal cuya matriz respecto de las bases usuales $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ y $\{\mathbf{u}'_1, \dots, \mathbf{u}'_q\}$ de $\mathbb{R}^q$ y $\mathbb{R}^p$, respectivamente, es B. El lector familiarizado con el producto tensorial puede apreciar que el producto de Kronecker de A y B no es más que la matriz de la aplicación

$$T \otimes S : \mathbb{R}^n \otimes \mathbb{R}^p \to \mathbb{R}^m \otimes \mathbb{R}^q,$$

respecto de la bases $\{\mathbf{e}_1 \otimes \mathbf{u}_1, \dots, \mathbf{e}_1 \otimes \mathbf{u}_p, \dots, \mathbf{e}_n \otimes \mathbf{u}_1, \dots, \mathbf{e}_n \otimes \mathbf{u}_p\}$ y $\{\mathbf{e}'_1 \otimes \mathbf{u}'_1, \dots, \mathbf{e}'_1 \otimes \mathbf{u}'_q, \dots, \mathbf{e}'_m \otimes \mathbf{u}'_1, \dots, \mathbf{e}'_m \otimes \mathbf{u}_q\}$ de $\mathbb{R}^n \otimes \mathbb{R}^p$ y $\mathbb{R}^m \otimes \mathbb{R}^q$, respectivamente.

(a) $\alpha \otimes A = A \otimes \alpha = \alpha A$, *para todo* $\alpha \in \mathbb{R}$.
(b) $(\alpha A) \otimes (\beta B) = \alpha\beta(A \otimes B)$, *para todo* α *y* $\beta \in \mathbb{R}$.
(c) $(A \otimes B) \otimes C = A \otimes (B \otimes C)$.
(d) $(A + B) \otimes C = (A \otimes C) + (B \otimes C)$, *si* A *y* B *tienen el mismo orden.*
(e) $A \otimes (B + C) = (A \otimes B) + (A \otimes C)$, *si* B *y* C *tienen el mismo orden.*
(f) $(A \otimes B)^{\mathrm{t}} = A^{\mathrm{t}} \otimes B^{\mathrm{t}}$.
(g) $\mathbf{a}\mathbf{b}^{\mathrm{t}} = \mathbf{a} \otimes \mathbf{b}^{\mathrm{t}} = \mathbf{b}^{\mathrm{t}} \otimes \mathbf{a}$.

Demostración. Las demostraciones son consecuencia directa de la definición de producto de Kronecker por lo que se dejan como ejercicio al lector. En el capítulo 5 de [**BS98**] se puede encontrar una demostración completa de cada una de ellas. ∎

Veamos ahora una interesante propiedad que involucra tanto al producto de Kronecker como al producto usual de matrices.

Teorema VII.1.4. *Sean* $A = (a_{ij}) \in \mathcal{M}_{m\times r}(\mathbb{R}), B \in \mathcal{M}_{p\times s}(\mathbb{R}), C = (c_{jl}) \in \mathcal{M}_{r\times n}(\mathbb{R})$ *y* $D \in \mathcal{M}_{s\times q}$. *Entonces*

$$(A \otimes B)(C \otimes D) = AC \otimes BD. \tag{VII.1.2}$$

Demostración. El miembro de la izquierda de (VII.1.2) es

$$\begin{pmatrix} a_{11}B & \dots & a_{1r}B \\ \vdots & & \vdots \\ a_{m1}B & \dots & a_{mr}B \end{pmatrix} \begin{pmatrix} c_{11}D & \dots & c_{1n}D \\ \vdots & & \vdots \\ c_{r1}D & \dots & c_{rn}D \end{pmatrix} = \begin{pmatrix} F_{11} & \dots & F_{1n} \\ \vdots & & \vdots \\ F_{m1} & \dots & F_{mn} \end{pmatrix},$$

donde

$$F_{ij} = \sum_{j=1}^{r} a_{ij}c_{jl}BD = (AC)_{ij}BD.$$

El miembro de la derecha de (VII.1.2) es

$$AC \otimes BD = \begin{pmatrix} (AC)_{11}BD & \dots & (AC)_{1n}BD \\ \vdots & & \vdots \\ (AC)_{m1}BD & \dots & (AC)_{mn}BD \end{pmatrix},$$

y por tanto se sigue el resultado buscado. ∎

Nuestro siguiente resultado demuestra que la traza del producto de Kronecker $A \otimes B$ se puede expresar fácilmente en términos de la traza de A y de la traza B cuando ambas son matrices cuadradas.

Proposición VII.1.5. *Sean* $A = (a_{ij}) \in \mathcal{M}_m(\mathbb{R})$ *y* $B \in \mathcal{M}_p(\mathbb{R})$. *Entonces*

$$\mathrm{tr}(A \otimes B) = \mathrm{tr}(A)\mathrm{tr}(B).$$

Demostración. Usando expresión (VII.1.1) cuando $n = m$, vemos que

$$\mathrm{tr}(A \otimes B) = \sum_{i=1}^{m} a_{ii}\mathrm{tr}(B) = \left(\sum_{i=1}^{m} a_{ii}\right)\mathrm{tr}(B) = \mathrm{tr}(A)\mathrm{tr}(B).$$

■

La proposición VII.1.5 da una expresión simplificada para la traza de un producto de Kronecker. Existe un resultado análogo para el determinante; sin embargo, necesitamos estudiar primero la inversa del producto de Kronecker.

Proposición VII.1.6. *Sea $A \in \mathcal{M}_{m\times n}(\mathbb{R})$ y $B \in \mathcal{M}_{p\times q}(\mathbb{R})$. Se cumple que*

(a) *si $m = n$ y $p = q$, y $A \otimes B$ es invertible, entonces $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$,*
(b) $(A \otimes B)^{+} = A^{+} \otimes B^{+}$.
(c) $(A \otimes B)^{-} = A^{-} \otimes B^{-}$, *para cualquier inversa generalizada, A^{-} y B^{-}, de A y B, respectivamente.*

Demostración. Usando el teorema VII.1.4 se tiene que

$$(A^{-1} \otimes B^{-1})(A \otimes B) = (A^{-1}A \otimes B^{-1}B) = I_m \otimes I_q = I_{mp},$$

luego se cumple (a). La verificación de (b) y (c) se deja como ejercicio al lector. ■

Proposición VII.1.7. *Sean $A \in \mathcal{M}_m(\mathbb{R})$ y $B \in \mathcal{M}_n(\mathbb{R})$. Se cumple que*

$$|A \otimes B| = |A|^n |B|^m.$$

Demostración. Sean $A = PD_1Q^{\mathrm{t}}$ y $B = P'D_2(Q')^{\mathrm{t}}$ las descomposiciones en valores singulares (largas) de A y B, respectivamente. Como P, P', Q y Q' son ortogonales, se tiene que $|A| = |D_1|$ y $|B| = |D_2|$. Además, se comprueba fácilmente que D_1 y D_2 verifican la proposición, es decir, $|D_1 \otimes D_2| = |D_1|^n |D_2|^m$ por ser D_1 y D_2 matrices diagonales. Por lo tanto, tenemos que

$$|D_1 \otimes D_2| = |A|^n |B|^m.$$

Ahora, basta observar que

$$\begin{aligned} |A \otimes B| &= |(PD_1Q^{\mathrm{t}}) \otimes (P'D_2(Q')^{\mathrm{t}})| = |(P \otimes P')(D_1 \otimes D_2)(Q^{\mathrm{t}} \otimes (Q')^{\mathrm{t}})| \\ &= |(P \otimes P')||(D_1 \otimes D_2)||(Q^{\mathrm{t}} \otimes (Q')^{\mathrm{t}})| = |(D_1 \otimes D_2)| = |A|^n |B|^m, \end{aligned}$$

sin más que tener en cuenta que $P \otimes P'$ y $Q^{\mathrm{t}} \otimes (Q')^{\mathrm{t}} = (Q \otimes Q')^{\mathrm{t}}$ también son matrices ortogonales. En efecto, $(P \otimes P')^{\mathrm{t}}(P \otimes P') = (P^{\mathrm{t}} \otimes (P')^{\mathrm{t}})(P \otimes P') = (P^{\mathrm{t}}P) \otimes ((P')^{\mathrm{t}}P') = (I_m) \otimes (I_n) = I_{mn}$, y análogamente con $(Q \otimes Q')^{\mathrm{t}}$. ■

Nuestro último resultado sobre el producto de Kronecker identifica la relación entre el rango de $A \otimes B$ y los rangos de A y B.

Corolario VII.1.8. *Sean $A \in \mathcal{M}_{m\otimes n}(\mathbb{R})$ y $B \in \mathcal{M}_{p\times q}(\mathbb{R})$. Se cumple que*

$$\mathrm{rg}(A \otimes B) = \mathrm{rg}(A)\mathrm{rg}(B)$$

Demostración. La demostración es completamente análoga a la de la proposición VII.1.7 por lo que se deja como ejercicio al lector. ∎

Nota VII.1.9. Sin comparamos las propiedades del producto ordinario de matrices y del producto de Kronecker se tiene

$$\begin{array}{rcl} (AB)^{\mathrm{t}} & = & B^{\mathrm{t}}A^{\mathrm{t}} \\ (AB)^{-1} & = & B^{-1}A^{-1} \\ \mathrm{tr}(AB) & \neq & \mathrm{tr}(A)\mathrm{tr}(B) \\ |AB| & = & |A|\,|B| \\ \mathrm{rg}(AB) & \leq & \mín\{\mathrm{rg}(A), \mathrm{rg}(B)\} \end{array} \qquad \begin{array}{rcl} (A \otimes B)^{\mathrm{t}} & = & A^{\mathrm{t}} \otimes B^{\mathrm{t}} \\ (A \otimes B)^{-1} & = & A^{-1} \otimes B^{-1} \\ \mathrm{tr}(A \otimes B) & = & \mathrm{tr}(A)\mathrm{tr}(B) \\ |A \otimes B| & = & |A|^m\,|B|^n \\ \mathrm{rg}(A \otimes B) & = & \mathrm{rg}(A)\mathrm{rg}(B) \end{array}$$

entendiendo que, en cada caso, la matrices tienen los órdenes apropiados para que las fórmulas tengan sentido.

El operador vec.

El operador que transforma una matriz en un vector apilando sus columnas una encima de otra se conoce como el **operador vec**. Si la matriz $A \in \mathcal{M}_{m\times n}(\mathbb{R})$ tiene como i-ésima columna a $\mathbf{a}_i \in \mathbb{R}^m$, entonces $\mathrm{vec}(A)$ es el vector de $\mathbb{R}^{mn}$ definido por

$$\mathrm{vec}(A) = \begin{pmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_n \end{pmatrix}.$$

Obsérvese que

$$\mathrm{vec}(\mathbf{a}) = \mathrm{vec}(\mathbf{a}^{\mathrm{t}}) = \mathbf{a},$$

para todo $\mathbf{a} \in \mathbb{R}^m$.

Ejemplo VII.1.10. Si A es la matriz

$$\begin{pmatrix} 2 & 0 & 5 \\ 8 & 1 & 3 \end{pmatrix},$$

entonces $\mathrm{vec}(A)$ es el vector

$$\begin{pmatrix} 2 \\ 8 \\ 0 \\ 1 \\ 5 \\ 3 \end{pmatrix}.$$

Nota VII.1.11. Obsérvese que, si E_{ij} es la matriz de orden $m \times n$ cuya entrada (i,j)-ésima es 1 y el resto de sus entradas son ceros y $\mathbf{e}_k$ es el vector k-ésimo de la base usual de $\mathbb{R}^{mn}$, entonces vec es la aplicación lineal

$$\mathcal{M}_{m\times n}(\mathbb{R}) \longrightarrow \mathbb{R}^{mn};\ E_{ij} \mapsto \mathbf{e}_{m(j-1)+i}.$$

Se comprueba fácilmente que esta aplicación es un isomorfismo de espacios vectoriales, y que su inversa es

$$\mathbb{R}^{mn} \longrightarrow \mathcal{M}_{m\times n}(\mathbb{R});\ \mathbf{e}_k \mapsto E_{c+1\, r},$$

donde c y r son el cociente y el resto de la división euclídea de k entre m, respectivamente.

En esta sección, desarrollaremos algunos propiedades básicas asociadas a este operador. Por ejemplo, si $\mathbf{a} \in \mathbb{R}^m$ y $\mathbf{b} = (b_1, \ldots, b_n)^{\mathrm{t}} \in \mathbb{R}^n$, entonces $\mathbf{a}\mathbf{b}^{\mathrm{t}} \in \mathcal{M}_{m\times n}(\mathbb{R})$ y

$$\mathrm{vec}(\mathbf{a}\mathbf{b}^{\mathrm{t}}) = \mathrm{vec}\left((b_1\mathbf{a}, \ldots, b_n\mathbf{a})\right) = \begin{pmatrix} b_1\mathbf{a} \\ \vdots \\ b_n\mathbf{a} \end{pmatrix} = \mathbf{b} \otimes \mathbf{a}.$$

El siguiente resultado nos da este y otros resultados que se siguen de forma inmediata de la definición del operador vec.

Proposición VII.1.12. *Sean* $\mathbf{a} \in \mathbb{R}^m$, $\mathbf{b} \in \mathbb{R}^n$ *y* A *y* B *dos matrices del mismo orden con coeficientes en* $\mathbb{R}$. *Se cumple que:*

(a) $\mathrm{vec}(\mathbf{a}\mathbf{b}^{\mathrm{t}}) = \mathbf{b} \otimes \mathbf{a}$.
(b) $\mathrm{vec}(\alpha A + \beta B) = \alpha\mathrm{vec}(A) + \beta\mathrm{vec}(B)$, *con* α *y* $\beta \in \mathbb{R}$.

Demostración. La demostración es un sencillo ejercicio que proponemos al lector. ∎

La traza del producto de dos matrices se puede expresar en términos de sus vectorizaciones.

Proposición VII.1.13. *Sean* A *y* $B \in \mathcal{M}_{m\times n}(\mathbb{R})$. *Se cumple que*

$$\mathrm{tr}(A^{\mathrm{t}}B) = \mathrm{vec}(A)^{\mathrm{t}}\mathrm{vec}(B).$$

Demostración. Como es habitual denotemos $\mathbf{a}_1, \ldots, \mathbf{a}_n$ las columnas de A y $\mathbf{b}_1, \ldots, \mathbf{b}_n$ las columnas de B. Entonces

$$\mathrm{tr}(A^{\mathrm{t}}B) = \sum_{i=1}^{n}(A^{\mathrm{t}}B)_{ii} = \sum_{i=1}^{n}\mathbf{a}_i^{\mathrm{t}}\mathbf{b}_i = \left(\mathbf{a}_i^{\mathrm{t}}, \ldots, \mathbf{a}_n^{\mathrm{t}}\right)\begin{pmatrix} \mathbf{b}_1 \\ \vdots \\ \mathbf{b}_n \end{pmatrix} = \mathrm{vec}(A)^{\mathrm{t}}\mathrm{vec}(B).$$

∎

Teorema VII.1.14. *Sean* $A \in \mathcal{M}_{m\times n}(\mathbb{R})$, $B \in \mathcal{M}_{n\times p}(\mathbb{R})$ *y* $C \in \mathcal{M}_{p\times q}(\mathbb{R})$. *Se cumple que*

$$\operatorname{vec}(ABC) = (C^{\mathrm{t}} \otimes A)\operatorname{vec}(B).$$

Demostración. En primer lugar observamos que si $\mathbf{b}_1, \ldots, \mathbf{b}_p$ son las columnas de B, entonces B se puede escribir como

$$B = \sum_{i=1}^{p} \mathbf{b}_i \mathbf{e}_i^{\mathrm{t}},$$

donde $\mathbf{e}_i$ es el elemento i-ésimo de la base canónica de $\mathbb{R}^p$. Así, se tiene que

$$\begin{aligned}\operatorname{vec}(ABC) &= \operatorname{vec}\left(A\left(\sum_{i=1}^{p} \mathbf{b}_i \mathbf{e}_i^{\mathrm{t}}\right) C\right) = \sum_{i=1}^{p} \operatorname{vec}(A\mathbf{b}_i \mathbf{e}_i^{\mathrm{t}} C) = \sum_{i=1}^{p} \operatorname{vec}\left((A\mathbf{b}_i)(C^{\mathrm{t}}\mathbf{e}_i)^{\mathrm{t}}\right) \\ &= \sum_{i=1}^{p} C^{\mathrm{t}}\mathbf{e}_i \otimes A\mathbf{b}_i = (C^{\mathrm{t}} \otimes A)\sum_{i=1}^{p}(\mathbf{e}_i \otimes \mathbf{b}_i),\end{aligned}$$

donde la segunda y la última igualdad siguen de la proposición VII.1.12(a). Usando de nuevo la proposición VII.1.12(a), obtenemos que

$$\sum_{i=1}^{p}(\mathbf{e}_i \otimes \mathbf{b}_i) = \sum_{i=1}^{p} \operatorname{vec}(\mathbf{b}_i \mathbf{e}_i^{\mathrm{t}}) = \operatorname{vec}\left(\sum_{i=1}^{p} \mathbf{b}_i \mathbf{e}_i^{\mathrm{t}}\right) = \operatorname{vec}(B),$$

lo que, junto con lo anterior, implica el resultado buscado. ∎

Ejemplo VII.1.15. En el tema VI, estudiamos los sistemas de ecuaciones lineales de la forma $A\mathbf{x} = \mathbf{b}$, así como los sistemas de la forma $AXC = B$. Usando el operador vec y el teorema VII.1.14, este segundo sistema de ecuaciones se puede expresar de forma equivalente como

$$\operatorname{vec}(AXC) = (C^{\mathrm{t}} \otimes A)\operatorname{vec}(X) = \operatorname{vec}(B);$$

es decir, en un sistema de la forma $A\mathbf{x} = \mathbf{b}$, donde en lugar de A, $\mathbf{x}$ y $\mathbf{b}$, tenemos $(C^{\mathrm{t}} \otimes A)$, $\operatorname{vec}(X)$ y $\operatorname{vec}(B)$, respectivamente. Como consecuencia, el teorema VI.4.5 del tema VI, que da la forma general de la solución de $A\mathbf{x} = \mathbf{b}$, se puede usar para realizar el ejercicio 26 del tema VI, donde se mostraba una expresión general de la solución de $AXC = B$.

La proposición VII.1.13 se puede generalizar fácilmente al caso del producto de más de dos matrices.

Corolario VII.1.16. *Sean* $A \in \mathcal{M}_{m\times n}(\mathbb{R})$, $B \in \mathcal{M}_{n\times p}(\mathbb{R})$, $C \in \mathcal{M}_{p\times q}(\mathbb{R})$ *y* $D \in \mathcal{M}_{q\times m}$. *Se cumple que*

$$\operatorname{tr}(ABCD) = \operatorname{vec}(A^{\mathrm{t}})^{\mathrm{t}}(D^{\mathrm{t}} \otimes B)\operatorname{vec}(C).$$

Demostración. Usando la proposición VII.1.13 se sigue que

$$\mathrm{tr}(ABCD) = \mathrm{tr}(A(BCD)) = \mathrm{vec}(A^{\mathrm{t}})^{\mathrm{t}}\mathrm{vec}(BCD).$$

Sin embargo, por el teorema VII.1.14, sabemos que $\mathrm{vec}(BCD) = (D^{\mathrm{t}} \otimes B)\mathrm{vec}(C)$, lo que completa la demostración. ∎

Corolario VII.1.17. *Sean $A \in \mathcal{M}_{m\times n}(\mathbb{R})$ y $C \in \mathcal{M}_{n\times m}(\mathbb{R})$, y B y $D \in \mathcal{M}_n(\mathbb{R})$. Se cumple que:*

(a) $\mathrm{tr}(ABC) = \mathrm{vec}(A^{\mathrm{t}})^{\mathrm{t}}(I_m \otimes B)\mathrm{vec}(C)$.
(b) $\mathrm{tr}(AD^{\mathrm{t}}BDC) = (\mathrm{vec}(D))^{\mathrm{t}}(A^{\mathrm{t}}C^{\mathrm{t}} \otimes B)\mathrm{vec}(D)$.

Demostración. La demostración de esta otra consecuencia del teorema VII.1.14 se deja como ejercicio al lector. ∎

Existen otras transformaciones de una matriz, $A \in \mathcal{M}_m(\mathbb{R})$, en un vector que son útiles cuando A tiene una estructura particular. Una de estas transformaciones de A, que se denota $\mathrm{v}(A)$, consiste en construir el vector de $\mathbb{R}^{m(m+1)/2}$ que se obtiene al eliminar de $\mathrm{vec}(A)$ las entradas correspondientes a los elementos de A que están por encima de la diagonal principal de A. De este modo, si A es triangular inferior, $\mathrm{v}(A)$ contiene todos los elementos de A excepto los ceros de la parte triangular superior de A. Asimismo, otra transformación de A en un vector, que se denota $\tilde{\mathrm{v}}(A)$, consiste en construir el vector de $\mathbb{R}^{m(m-1)/2}$ que se obtiene al eliminar de $\mathrm{v}(A)$ las entradas correspondientes a la diagonal de A; es decir, $\tilde{\mathrm{v}}(A)$ es el vector que se obtiene apilando las porciones de columnas de A que están por debajo de la diagonal de A.

Ejemplo VII.1.18. Los operadores v y $\tilde{\mathrm{v}}$ son particularmente útiles cuando estamos manipulando matrices de covarianza y de correlación. Por ejemplo, supongamos que estamos interesados en la distribución de la matriz de covarianza muestral o en la distribución de la matriz de correlación muestral calculadas a partir de una muestra de observaciones de tres variables diferentes. Las matrices de covarianza y correlación resultantes son de la forma

$$S = \begin{pmatrix} s_{11} & s_{12} & s_{13} \\ s_{12} & s_{22} & s_{23} \\ s_{13} & s_{23} & s_{33} \end{pmatrix} \quad \text{y} \quad R = \begin{pmatrix} 1 & r_{12} & r_{13} \\ r_{12} & 1 & r_{23} \\ r_{13} & r_{23} & 1 \end{pmatrix},$$

respectivamente; de tal modo que

$$\mathrm{vec}(S) = (s_{11}, s_{12}, s_{13}, s_{12}, s_{22}, s_{23}, s_{13}, s_{23}, s_{33})^{\mathrm{t}},$$
$$\mathrm{vec}(R) = (1, r_{12}, r_{13}, r_{12}, 1, r_{23}, r_{13}, r_{23}, 1)^{\mathrm{t}}.$$

Como S y R son simétricas, hay elementos redundantes en $\mathrm{vec}(S)$ y en $\mathrm{vec}(R)$. La eliminación de estos elementos se puede obtener usando $\mathrm{v}(S)$ y $\mathrm{v}(R)$

$$\mathrm{v}(S) = (s_{11}, s_{12}, s_{13}, s_{22}, s_{23}, s_{33})^{\mathrm{t}},$$
$$\mathrm{v}(R) = (1, r_{12}, r_{13}, 1, r_{23}, 1)^{\mathrm{t}}.$$

Finalmente, eliminando los unos no aleatorios de $\mathrm{v}(R)$, obtenemos

$$\tilde{\mathrm{v}}(R) = (r_{12}, r_{13}, r_{23})^{\mathrm{t}}$$

que contiene todas las variables aleatorias de R.

Terminaremos esta sección mostrando una interesante propiedad que nos permite transformar el vec de un producto de Kronecker en el producto de Kronecker de los operadores vec. Esta propiedad es crucial para la diferenciación de productos de Kronecker. Pero antes, necesitamos introducir la siguiente notación.

Notación VII.1.19. Sea A una matriz arbitraria de orden $m \times n$. Denotaremos por K_{mn} la única matriz de orden $mn \times mn$ tal que

$$K_{mn}\mathrm{vec}(A) = \mathrm{vec}(A^{\mathrm{t}}). \tag{VII.1.3}$$

Si $m = n$, se escribe K_n en vez de K_{nn}. Obsérvese que K_{mn} es una matriz de permutación que no depende de A.

Las matrices K_{mn} se llama **matrices de conmutación**, este nombre está justificado por el siguiente resultado:

Proposición VII.1.20. *Sea $A \in \mathcal{M}_{m\times n}(\mathbb{R})$ y $B \in \mathcal{M}_{p\times q}(\mathbb{R})$. Entonces*

$$K_{pm}(A \otimes B) = (B \otimes A)K_{qn}.$$

Demostración. Sea $C \in \mathcal{M}_{q\times n}(\mathbb{R})$. Entonces, usando repetidas veces la expresión (VII.1.3) y el teorema VII.1.14, se tiene que

$$\begin{aligned} K_{pm}(A \otimes B)\mathrm{vec}(C) &= K_{pm}\mathrm{vec}(BCA^{\mathrm{t}}) = \mathrm{vec}(AC^{\mathrm{t}}B^{\mathrm{t}}) = (B \otimes A)\mathrm{vec}(C^{\mathrm{t}}) \\ &= (B \otimes A)K_{qn}\mathrm{vec}(C). \end{aligned}$$

Como C es arbitrario se sigue el resultado buscado. ∎

Ahora ya estamos en disposición de enunciar y demostrar el teorema anteriormente anunciado.

Teorema VII.1.21. *Sea $A \in \mathcal{M}_{m\times n}(\mathbb{R})$ y $B \in \mathcal{M}_{p\times q}(\mathbb{R})$. Entonces,*

$$\mathrm{vec}(A \otimes B) = (I_n \otimes K_{qm} \otimes I_p)(\mathrm{vec}(A) \otimes \mathrm{vec}(B)).$$

Demostración. Sean $\mathbf{a}_i,\ i = 1, \ldots, n$, y $\mathbf{b}_j = 1, \ldots, q$, las columnas de A y B, respectivamente. Asimismo, sean $\mathbf{e}_i,\ i = 1, \ldots, n$, y $\mathbf{e}'_j,\ j = 1, \ldots, q$, columnas de I_n e I_q, respectivamente. Con esta notación, podemos escribir A y B como sigue

$$A = \sum_{i=1}^{n} \mathbf{a}_i \mathbf{e}_i^{\mathrm{t}} \quad \text{y} \quad B = \sum_{j=1}^{q} \mathbf{b}_j (\mathbf{e}'_j)^{\mathrm{t}};$$

de este modo obtenemos que

$$\begin{aligned}
\operatorname{vec}(A \otimes B) &= \sum_{i=1}^{n}\sum_{j=1}^{q} \operatorname{vec}(\mathbf{a}_i \mathbf{e}_i^{\mathrm{t}} \otimes \mathbf{b}_j (\mathbf{e}'_j)^{\mathrm{t}}) = \sum_{i=1}^{n}\sum_{j=1}^{q} \operatorname{vec}((\mathbf{a}_i \otimes \mathbf{b}_j)(\mathbf{e}_i \otimes \mathbf{e}'_j)^{\mathrm{t}}) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{q} (\mathbf{e}_i \otimes \mathbf{e}'_j \otimes \mathbf{a}_i \otimes \mathbf{b}_j) = \sum_{i=1}^{n}\sum_{j=1}^{q} (\mathbf{e}_i \otimes K_{qm}(\mathbf{a}_i \otimes \mathbf{e}'_j) \otimes \mathbf{b}_j) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{q} (I_n \otimes K_{qm} \otimes I_p)(\mathbf{e}_i \otimes \mathbf{a}_i \otimes \mathbf{e}'_j \otimes \mathbf{b}_j) \\
&= (I_n \otimes K_{qm} \otimes I_p) \left(\left(\sum_{i=1}^{n} \operatorname{vec}(\mathbf{a}_i \mathbf{e}_i^{\mathrm{t}}) \right) \left(\sum_{j=1}^{q} \operatorname{vec}(\mathbf{b}_j (\mathbf{e}'_i)^{\mathrm{t}}) \right) \right) \\
&= (I_n \otimes K_{qm} \otimes I_p)(\operatorname{vec}(A) \otimes \operatorname{vec}(B)),
\end{aligned}$$

lo que completa la demostración. ∎

2. Diferenciación matricial

Comenzamos recordando algunos conceptos básicos sobre funciones en el espacio euclídeo con el único objetivo de fijar la notación que se usará a lo largo de la sección. Un desarrollo riguroso sobre este tema puede encontrarse, por ejemplo, en [**Spi88**].

Supongamos que $f_1, \ldots, f_m$ son funciones de $\mathbb{R}^n$ en $\mathbb{R}$. Estas m funciones determinan la función $f : \mathbb{R}^n \to \mathbb{R}^m$ con m componentes definida por

$$f(\mathbf{x}) = \begin{pmatrix} f_1(\mathbf{x}) \\ \vdots \\ f_m(\mathbf{x}) \end{pmatrix},$$

con $\mathbf{x} = (x_1, \ldots, x_n)^{\mathrm{t}}$; esto es, una función vectorial con variable vectorial.

La función f es diferenciable en $\mathbf{a} \in \mathbb{R}^n$ si, y sólo si, cada una de las componentes f_i es diferenciable en $\mathbf{a} \in \mathbb{R}^n$; equivalentemente, si existe una aplicación lineal $T : \mathbb{R}^n \to \mathbb{R}^m$ tal que

$$\lim_{\mathbf{u} \to \mathbf{0}} \frac{\|f(\mathbf{a} + \mathbf{u}) - f(\mathbf{a}) - T(\mathbf{u})\|}{\|\mathbf{u}\|} = 0.$$

Nótese que $\mathbf{u} \in \mathbb{R}^n$ y $f(\mathbf{a}+\mathbf{u}) - f(\mathbf{a}) - T(\mathbf{u}) \in \mathbb{R}^m$, por lo que en el numerador estamos usando la norma en $\mathbb{R}^m$ y en el denominador la norma en $\mathbb{R}^n$, ambas para el producto escalar usual. La aplicación lineal T cuando existe es única, se suele designar por $\mathrm{d}f(\mathbf{a})$ y se denomina diferencial de f en $\mathbf{a}$.

En muchas ocasiones es conveniente utilizar la matriz de $\mathrm{d}f(\mathbf{a})$ respecto de las bases usuales de $\mathbb{R}^n$ y $\mathbb{R}^m$. Esta matriz de orden $m \times n$ se suele llamar primera derivada de f en $\mathbf{a}$ o matriz Jacobiana de f en $\mathbf{a}$, y responde a la siguiente expresión:

$$\frac{\partial}{\partial \mathbf{x}^{\mathrm{t}}} f(\mathbf{a}) := \begin{pmatrix} \frac{\partial}{\partial x_1} f_1(\mathbf{a}) & \dots & \frac{\partial}{\partial x_n} f_1(\mathbf{a}) \\ \vdots & & \vdots \\ \frac{\partial}{\partial x_1} f_m(\mathbf{a}) & \dots & \frac{\partial}{\partial x_n} f_m(\mathbf{a}) \end{pmatrix}.$$

En algunas situaciones concretas, las funciones f_j y las variables x_i se ordenan en una matriz en vez de en un vector. Así, el caso más general lo engloba una función matricial de orden $m \times q$

$$F(X) = \begin{pmatrix} f_{11}(X) & \dots & f_{1q}(X) \\ \vdots & & \vdots \\ f_{m1}(X) & \dots & f_{mq}(X) \end{pmatrix}$$

de variable matricial X de orden $n \times p$. Es decir, F es una función de $\mathcal{M}_{n\times p}(\mathbb{R})$ en $\mathcal{M}_{m\times q}(\mathbb{R})$.

Los conceptos para funciones vectoriales de variable vectorial se pueden extender fácilmente a la función matricial $F(X)$ usando el operador vec; basta considerar la función $f : \mathbb{R}^{np} \to \mathbb{R}^{mq}$ tal que

$$f(\mathrm{vec}(X)) = \mathrm{vec}(F(X)).$$

De este modo, se define la **diferencial de** F **en** $A \in \mathcal{M}_{n\times p}(\mathbb{R})$ como la única aplicación lineal $\mathrm{d}F(A)$ que hace conmutativo el siguiente diagrama:

(VII.2.4)
$$\begin{array}{ccc} \mathcal{M}_{n\times p}(\mathbb{R}) & \xrightarrow{\mathrm{d}F(A)} & \mathcal{M}_{m\times q}(\mathbb{R}) \\ \Big\downarrow{\scriptstyle \mathrm{vec}} & & \Big\downarrow{\scriptstyle \mathrm{vec}} \\ \mathbb{R}^{np} & \xrightarrow{\mathrm{d}f(\mathrm{vec}(A))} & \mathbb{R}^{mq}, \end{array}$$

Es decir, por definición, $\mathrm{vec}(\mathrm{d}F(A)) = \mathrm{d}\,\mathrm{vec}(F(A))$.

Ahora, si consideramos las bases usuales de $\mathbb{R}^{np}$ y $\mathbb{R}^{mq}$, se tiene que la matriz Jacobiana de f en $\mathrm{vec}(A) \in \mathbb{R}^{np}$ es la matriz de orden $mq \times np$

(VII.2.5)
$$\frac{\partial}{\partial \mathrm{vec}(X)^{\mathrm{t}}} f(\mathrm{vec}(A)) = \frac{\partial}{\partial \mathrm{vec}(X)^{\mathrm{t}}} \mathrm{vec}(F(A)),$$

es decir, aquella que tiene como entrada (i,j)-ésima a la derivada parcial de la entrada i-ésima de $\mathrm{vec}(F(X))$ con respecto a la entrada j-ésima de $\mathrm{vec}(X)$.

Definición VII.2.1. A la matriz (VII.2.5) la llamaremos **derivada de F en A respecto de X**.

Ejemplo VII.2.2. La matriz de variables independientes X de orden $m \times p$ define una aplicación de $\mathcal{M}_{m\times p}(\mathbb{R}) \to \mathcal{M}_{m\times p}(\mathbb{R})$ cuya derivada respecto de X en cualquier punto es la matriz identidad de orden mp.

Existen otras definiciones de derivada matricial (véanse, por ejemplo, las secciones 3 y 4 de [**MN07**] y la sección 5.4 de [**BS98**]). La elección de la definición VII.2.1 resulta útil cuando se está interesado fundamentalmente en aplicar a funciones matriciales resultados matemáticos relativos a funciones vectoriales, como es nuestro caso.

Propiedades de la diferencial.

En lo que sigue, X denotará una matriz de orden $n\times p$ de variables independientes. Además, si F es una función matricial de X, escribiremos $\mathrm{d}F$ en vez de $\mathrm{d}F(A)$ con objeto de aligerar un poco la notación.

En la siguiente proposición se indican algunas de las reglas de derivación para las operaciones más usuales entre expresiones matriciales. Ni que decir tiene que todas las propiedades que veremos a continuación sólo tendrán sentido allí donde exista la diferencial.

Proposición VII.2.3.

(a) Derivada de la función constante. *Sea A una matriz de orden $m \times q$ cuyo elementos no dependen de los X. Entonces,*

$$\mathrm{d}A = 0.$$

(b) Derivada del producto por un escalar. *Sea F una matriz de orden $m\times q$ cuyos elementos son funciones de X. Entonces, para cualquier $\alpha \in \mathbb{R}$ se verifica que*

$$\mathrm{d}(\alpha F) = \alpha(\mathrm{d}F).$$

(c) Derivada de la suma. *Sean F y G dos matrices de orden $m\times q$ cuyos elementos son funciones de X. Entonces,*

$$\mathrm{d}(F+G) = \mathrm{d}F + \mathrm{d}G.$$

(d) Derivada del producto. *Sean F y G dos matrices de ordenes $m \times q$ y $q \times r$, respectivamente, cuyos elementos son funciones de X. Entonces,*

$$\mathrm{d}(FG) = (\mathrm{d}F)G + F\,(\mathrm{d}G).$$

Demostración. Los apartados (a), (b) y (c) se siguen de la definición de diferencial de una matriz en un punto.

(d) Sabemos que las funciones vectoriales de variable vectorial cumplen que $\mathrm{d}(fg) = (\mathrm{d}f)g + f\mathrm{d}g$. Usando esta igualdad se comprueba fácilmente que $(\mathrm{d}F)G + F(\mathrm{d}G)$ hace conmutativo el diagrama (VII.2.4), y se concluye que $(\mathrm{d}F)G + F(\mathrm{d}G) = \mathrm{d}(F\,G)$, por la unicidad de la diferencial. ∎

Obsérvese que de (a) y (d) se sigue que $\mathrm{d}(AF) = A\mathrm{d}F$.

Veamos ahora otras propiedades de la diferencial de una función matricial de variable X relacionadas con las operaciones específicas de las matrices.

Proposición VII.2.4. *Sean F una matriz de orden $m \times q$ y G una matriz de orden $r \times s$ cuyos elementos son funciones de una matriz X de orden $n \times p$ de variables independientes. Se cumple que*

(a) $\mathrm{d}F^{\mathrm{t}} = (\mathrm{d}F)^{\mathrm{t}}$.
(b) $\mathrm{d}(F \otimes G) = (\mathrm{d}\,F) \otimes G + F \otimes \mathrm{d}\,G$.
(c) *Si* $q = m$, *entonces* $\mathrm{d}(\mathrm{tr}(F)) = \mathrm{tr}(\mathrm{d}\,F)$.

Demostración. (a) Como

$$\begin{aligned}\mathrm{vec}(\mathrm{d}(F^{\mathrm{t}})) &= \mathrm{d}(\mathrm{vec}(F^{\mathrm{t}})) = \mathrm{d}(K_{mq}\mathrm{vec}(F)) = K_{mq}\mathrm{d}(\mathrm{vec}(F)) = K_{mq}\mathrm{vec}(\mathrm{d}F)\\ &= \mathrm{vec}((\mathrm{d}F)^{\mathrm{t}}),\end{aligned}$$

se concluye la igualdad buscada.

(b) Veamos en primer lugar que

$$\begin{aligned}\mathrm{d}(\mathrm{vec}(F) \otimes \mathrm{vec}(G)) &= \mathrm{d}(\mathrm{vec}(\mathrm{vec}(G)\mathrm{vec}(F)^{\mathrm{t}}))\\ &= \mathrm{vec}((\mathrm{d}\,\mathrm{vec}(G))\mathrm{vec}(F)^{\mathrm{t}} + \mathrm{vec}(G)\mathrm{d}(\mathrm{vec}(F)^{\mathrm{t}}))\\ &= \mathrm{vec}((\mathrm{d}\,\mathrm{vec}(G))\mathrm{vec}(F)^{\mathrm{t}}) + \mathrm{vec}(\mathrm{vec}(G)(\mathrm{d}\,\mathrm{vec}(F))^{\mathrm{t}})\\ &= \mathrm{vec}(F) \otimes (\mathrm{d}\,\mathrm{vec}(G)) + (\mathrm{d}\,\mathrm{vec}(F)) \otimes \mathrm{vec}(G)\\ &= \mathrm{vec}(F) \otimes \mathrm{vec}(\mathrm{d}G) + \mathrm{vec}(\mathrm{d}\,F) \otimes \mathrm{vec}(G)\\ &= \mathrm{vec}(\mathrm{d}\,F) \otimes \mathrm{vec}(G) + \mathrm{vec}(F) \otimes \mathrm{vec}(\mathrm{d}G)\end{aligned}$$

De modo que, como

$$\mathrm{vec}(F \otimes G) = (I_q \otimes K_{sm} \otimes I_r)(\mathrm{vec}(F) \otimes \mathrm{vec}(G))$$

y

$$\mathrm{vec}((\mathrm{d}\,F) \otimes G + F \otimes \mathrm{d}\,G) = (I_q \otimes K_{sm} \otimes I_r)(\mathrm{vec}(\mathrm{d}\,F) \otimes \mathrm{vec}(G) + \mathrm{vec}(F) \otimes \mathrm{vec}(\mathrm{d}G)),$$

concluimos que

$$\begin{aligned}\mathrm{vec}(\mathrm{d}((F\otimes G))) &= \mathrm{d}(\mathrm{vec}(F\otimes G))\\ &= (I_q\otimes K_{sm}\otimes I_r)\mathrm{d}(\mathrm{vec}(F)\otimes\mathrm{vec}(G)) = \mathrm{vec}((\mathrm{d}\,F)\otimes G + F(\mathrm{d}\,G)),\end{aligned}$$

de donde se sigue el resultado buscado, por ser vec un isomorfismo.

(c) Basta usar la proposición VII.1.13 para obtener la igualdad buscada; en efecto,

$$\mathrm{d}(\mathrm{tr}(F)) = \mathrm{d}(\mathrm{vec}(I_m)^{\mathrm{t}}\mathrm{vec}(F)) = \mathrm{vec}(I_m)^{\mathrm{t}}\mathrm{d}(\mathrm{vec}(F)) = \mathrm{vec}(I_m)^{\mathrm{t}}\mathrm{vec}(\mathrm{d}\,F) = \mathrm{tr}(\mathrm{d}\,F).$$

■

Ejemplo VII.2.5. Sea X una matriz de orden $n\times q$ de variables independientes. Si $F(X) = XX^{\mathrm{t}}$, entonces

$$\begin{aligned}\mathrm{vec}(\mathrm{d}F(X)) &= \mathrm{vec}(\mathrm{d}(XX^{\mathrm{t}})) = \mathrm{vec}((\mathrm{d}X)X^{\mathrm{t}} + X(\mathrm{d}X)^{\mathrm{t}})\\ &= \mathrm{vec}(I_n(\mathrm{d}X)X^{\mathrm{t}}) + \mathrm{vec}(X(\mathrm{d}X)^{\mathrm{t}}I_n)\\ &= (X\otimes I_n)\mathrm{d}\,\mathrm{vec}(X) + (I_n\otimes X)K_{nq}\mathrm{d}\,\mathrm{vec}(X)\\ &= ((X\otimes I_n) + K_n(X\otimes I_n))\mathrm{d}\,\mathrm{vec}(X)\\ &= (I_{n^2} + K_n)(X\otimes I_n)\mathrm{d}\,\mathrm{vec}(X)\end{aligned}$$

luego,

$$\frac{\partial}{\partial\mathrm{vec}(X)^{\mathrm{t}}}F = (I_{n^2} + K_n)(X\otimes I_n).$$

3. Algunas derivadas matriciales de interés

En la sección anterior ya hemos mostrado algunas diferenciales y derivadas de funciones escalares y matriciales de variable matricial; en esta última sección veremos algunas más. En el capítulo 9 de [**MN07**] y en el capítulo 5 de [**BS98**] se pueden encontrar muchas más diferenciales y derivadas de funciones escalares y matriciales de variable matricial.

A partir de ahora, cuando consideremos funciones de la forma $f(X)$ o $F(X)$, supondremos que X es una matriz de orden $m\times n$ de variable independientes; es decir, no consideraremos que X tenga ninguna estructura particular como pueden ser simetría, triangularidad, ... Comencemos viendo algunas funciones escalares de X.

Ejemplo VII.3.1. Sea $\mathbf{x}$ un vector de m variables independientes, y definimos la función

$$f(\mathbf{x}) = \mathbf{a}^{\mathrm{t}}\mathbf{x},$$

con $\mathbf{a}\in\mathbb{R}^m$. De

$$\mathrm{d}(f(\mathbf{x})) = \mathrm{d}(\mathbf{a}^{\mathrm{t}}\mathbf{x}) = \mathbf{a}^{\mathrm{t}}\mathrm{d}\mathbf{x},$$

concluimos que

$$\frac{\partial}{\partial \mathbf{x}^{\mathrm{t}}} f = \mathbf{a}^{\mathrm{t}}.$$

Ejemplo VII.3.2. Sea $\mathbf{x}$ un vector de m variables independientes, y definimos la función

$$g(\mathbf{x}) = \mathbf{x}^{\mathrm{t}} A \mathbf{x},$$

con $A \in \mathcal{M}_m(\mathbb{R})$ simétrica. Usando que

$$\begin{aligned} \mathrm{d}(g(\mathbf{x})) = \mathrm{d}(\mathbf{x}^{\mathrm{t}} A \mathbf{x}) &= \mathrm{d}(\mathbf{x}^{\mathrm{t}}) A \mathbf{x} + \mathbf{x}^{\mathrm{t}} A \mathrm{d}\mathbf{x} \\ &= (\mathrm{d}\mathbf{x})^{\mathrm{t}} A \mathbf{x} + \mathbf{x}^{\mathrm{t}} A \mathrm{d}\mathbf{x} = ((\mathrm{d}\mathbf{x})^{\mathrm{t}} A \mathbf{x})^{\mathrm{t}} + \mathbf{x}^{\mathrm{t}} A \mathrm{d}\mathbf{x} \\ &= 2\mathbf{x}^{\mathrm{t}} A \mathrm{d}\mathbf{x}, \end{aligned}$$

se sigue que

$$\frac{\partial}{\partial \mathbf{x}^{\mathrm{t}}} g = 2\mathbf{x}^{\mathrm{t}} A.$$

La traza y el determinante.

Proposición VII.3.3. *Sean X una matriz de orden m y* $\mathrm{adj}(X)$ *su matriz adjunta*[2]*. Entonces,*

(a) $\mathrm{d}(\mathrm{tr}(X)) = \mathrm{vec}(I_m)^{\mathrm{t}} \mathrm{d}(\mathrm{vec}(X))$ *y*

$$\frac{\partial}{\partial \mathrm{vec}(X)^{\mathrm{t}}} \mathrm{tr}(X) = \mathrm{vec}(I_m)^{\mathrm{t}}.$$

(b) $\mathrm{d}|X| = \mathrm{tr}(\mathrm{adj}(X)\mathrm{d}X)$ *y*

$$\frac{\partial}{\partial \mathrm{vec}(X)^{\mathrm{t}}} |X| = \mathrm{vec}(\mathrm{adj}(X)^{\mathrm{t}})^{\mathrm{t}}.$$

(c) *si X es invertible,* $\mathrm{d}|X| = |X|\mathrm{tr}(X^{-1}\mathrm{d}X)$ *y*

$$\frac{\partial}{\partial \mathrm{vec}(X)^{\mathrm{t}}} |X| = |X| \mathrm{vec}((X^{-1})^{\mathrm{t}})^{\mathrm{t}}.$$

Demostración. Teniendo en cuenta que $\mathrm{vec}(\mathrm{tr}(X)) = \mathrm{tr}(X)$ y $\mathrm{vec}(|X|) = |X|$, en el apartado (a) la relación entre la diferencial y la deriva es directa; mientras que en los apartados (b) y (c) la relación entre la diferencial y la derivada es consecuencia directa de la proposición VII.1.13.

(a) $\mathrm{d}(\mathrm{tr}(X)) = \mathrm{tr}(\mathrm{d}X) = \mathrm{vec}(I_m)^{\mathrm{t}} \mathrm{vec}(\mathrm{d}X) = \mathrm{vec}(I_m)^{\mathrm{t}} \mathrm{d}(\mathrm{vec}(X))$. Ahora, usando la relación entre la diferencial y la derivada se obtiene la expresión para la derivada buscada.

[2]Definición I.2.9 del tema III.

(b) Sabemos que $|X| = \sum_{k=1}^{m}(-1)^{i+k}x_{ik}|X_{ik}|$, donde X_{ik} es la submatriz de X que se obtiene eliminando la fila i-ésima y la columna k-ésima. Por tanto,

$$\frac{\partial}{\partial x_{ij}}|X| = (-1)^{i+j}|X_{ij}|,$$

pues $|X_{ik}|$ no depende de la variable x_{ij}, si $k \neq j$. De donde se sigue que

$$\frac{\partial}{\partial \mathrm{vec}(X)^{\mathrm{t}}}|X| = \mathrm{vec}(\mathrm{adj}(X)^{\mathrm{t}})^{\mathrm{t}},$$

y usando la relación entre la diferencial y la derivada se obtiene la diferencial buscada.

El apartado (c) sigue directamente del (b), sin más que tener en cuenta que si X es invertible, entonces $X^{-1} = |X|^{-1}\mathrm{adj}(X)$. ∎

Una consecuencia inmediata del apartado (c) de la proposición anterior es el siguiente resultado.

Corolario VII.3.4. *Sea X una matriz invertible de orden m. Entonces,*

$$\mathrm{d}(\log(|X|)) = \mathrm{tr}(X^{-1}\mathrm{d}X)$$

y

$$\frac{\partial}{\partial \mathrm{vec}(X)^{\mathrm{t}}}\log(|X|) = \mathrm{vec}((X^{-1})^{\mathrm{t}})^{\mathrm{t}}.$$

Demostración. Usando la regla de la cadena de las funciones vectoriales de variable vectorial se tiene que

$$\frac{\partial}{\partial \mathrm{vec}(X)^{\mathrm{t}}}\log(|X|) = \frac{1}{|X|}\frac{\partial}{\partial \mathrm{vec}(X)^{\mathrm{t}}}|X| = \mathrm{vec}((X^{-1})^{\mathrm{t}})^{\mathrm{t}},$$

usando ahora la relación entre la diferencial y la derivada se concluye el resultado buscado. ∎

Ejemplo VII.3.5. Si $F(X) = \mathrm{tr}(X^{\mathrm{t}}X) = \mathrm{tr}(X\,X^{\mathrm{t}})$, entonces

$$\begin{aligned}\mathrm{d}F(X) &= \mathrm{d}(\mathrm{tr}(X^{\mathrm{t}}X)) = \mathrm{tr}(\mathrm{d}(X^{\mathrm{t}}X)) = \mathrm{tr}((\mathrm{d}X)^{\mathrm{t}}X + X^{\mathrm{t}}\mathrm{d}X)\\ &= \mathrm{tr}((\mathrm{d}X)^{\mathrm{t}}X) + \mathrm{tr}(X^{\mathrm{t}}\mathrm{d}X) = 2\mathrm{tr}(X^{\mathrm{t}}\mathrm{d}X)\\ &= 2\,\mathrm{vec}(X)^{\mathrm{t}}\mathrm{vec}(\mathrm{d}X),\end{aligned}$$

luego,

$$\frac{\partial}{\partial \mathrm{vec}(X)^{\mathrm{t}}}F = 2\mathrm{vec}(X^{\mathrm{t}})^{\mathrm{t}}.$$

Ejemplo VII.3.6. Si $X\,X^{\mathrm{t}}$ es invertible y $F(X) = |X\,X^{\mathrm{t}}|$, entonces

$$\begin{aligned}
\mathrm{d}F(X) &= |X\,X^{\mathrm{t}}|\mathrm{tr}((X\,X^{\mathrm{t}})^{-1}\mathrm{d}(X\,X^{\mathrm{t}})) \\
&= |X\,X^{\mathrm{t}}|\mathrm{tr}((X\,X^{\mathrm{t}})^{-1}((\mathrm{d}X)X^{\mathrm{t}} + X(\mathrm{d}X)^{\mathrm{t}})) \\
&= |X\,X^{\mathrm{t}}|\mathrm{tr}((X\,X^{\mathrm{t}})^{-1}(\mathrm{d}X)X^{\mathrm{t}}) + \mathrm{tr}((X\,X^{\mathrm{t}})^{-1}X(\mathrm{d}X)^{\mathrm{t}}) \\
&= 2\,|X\,X^{\mathrm{t}}|\mathrm{tr}(X^{\mathrm{t}}(XX^{\mathrm{t}})^{-1}\mathrm{d}X) \\
&= 2\,|X\,X^{\mathrm{t}}|\mathrm{vec}((X^{\mathrm{t}}(XX^{\mathrm{t}})^{-1})^{\mathrm{t}})^{\mathrm{t}}\mathrm{vec}(\mathrm{d}X) \\
&= 2\,|X\,X^{\mathrm{t}}|\mathrm{vec}((XX^{\mathrm{t}})^{-1}X)^{\mathrm{t}}\mathrm{vec}(\mathrm{d}X)
\end{aligned}$$

luego,

$$\frac{\partial}{\partial \mathrm{vec}(X)^{\mathrm{t}}}F = 2\,|X\,X^{\mathrm{t}}|\mathrm{vec}((XX^{\mathrm{t}})^{-1}X)^{\mathrm{t}}.$$

La inversa y la inversa de Moore-Penrose.

El próximo resultado nos da la diferencial y la derivada de la inversa de una matriz invertible.

Proposición VII.3.7. *Si X es una matriz invertible de orden m, entonces*

$$\mathrm{d}(X^{-1}) = -X^{-1}(\mathrm{d}X)X^{-1}$$

y

$$\frac{\partial}{\partial \mathrm{vec}(X)^{\mathrm{t}}}\mathrm{vec}(X^{-1}) = -((X^{-1})^{\mathrm{t}} \otimes X^{-1}).$$

Demostración. Calculando la diferencial de ambos lados de la igualdad $I_m = XX^{-1}$, obtenemos que

$$0 = \mathrm{d}I_m = \mathrm{d}(XX^{-1}) = (\mathrm{d}X)X^{-1} + X(\mathrm{d}X^{-1}).$$

Multiplicando a izquierda por X^{-1} y despejando $\mathrm{d}(X^{-1})$, se tiene que

$$\mathrm{d}(X^{-1}) = -X^{-1}(\mathrm{d}X)X^{-1},$$

de donde sigue que

$$\begin{aligned}
\mathrm{d}(\mathrm{vec}(X^{-1})) &= \mathrm{vec}(\mathrm{d}(X^{-1})) = -\mathrm{vec}(X^{-1}(\mathrm{d}X)X^{-1}) \\
&= -((X^{-1})^{\mathrm{t}} \otimes X^{-1})\mathrm{vec}(\mathrm{d}X) = -((X^{-1})^{\mathrm{t}} \otimes X^{-1})\mathrm{d}(\mathrm{vec}(X))
\end{aligned}$$

lo que completa la demostración. ∎

Una generalización natural de la proposición VII.3.7 es el resultado que nos describe la diferencial y la derivada de la inversa de Moore-Penrose de una matriz.

Teorema VII.3.8. *Si X es una matriz $m \times n$ y X^+ es su inversa de Moore-Penrose, entonces*

$$\mathrm{d}X^+ = (I_n - X^+X)(\mathrm{d}X^{\mathrm{t}})(X^+)^{\mathrm{t}}(X^+)^{\mathrm{t}} + X^+(X^+)^{\mathrm{t}}\mathrm{d}(X^{\mathrm{t}})(I_m - XX^+) - X^+(dX)X^+$$

y

$$\frac{\partial}{\partial \mathrm{vec}(X)^{\mathrm{t}}} = \left((X^+)^{\mathrm{t}}X^+ \otimes (I_n - X^+X) + (I_m - XX^+) \otimes X^+(X^+)^{\mathrm{t}}\right) K_{mn} - ((X^+)\mathrm{t} \otimes X^+).$$

La demostración de este teorema no es difícil aunque sí muy extensa. El lector interesado puede encontrarla en la página 362 de [**Sch05**].

Ejercicios del tema VII

Ejercicio 1. Dadas las matrices

$$A = \begin{pmatrix} 2 & 3 \\ 1 & 2 \end{pmatrix} \quad \text{y} \quad B = \begin{pmatrix} 5 & 3 \\ 3 & 2 \end{pmatrix}$$

Calcular $A \otimes B,\ B \otimes A,\ \mathrm{tr}(A \otimes B),\ |A \otimes B|$, los autovalores de $A \otimes B$ y $(A \otimes B)^{-1}$.

Ejercicio 2. Sean $A \in \mathcal{M}_{m\times n}(\mathbb{R}),\ B \in \mathcal{M}_{p\times q}(\mathbb{R})$ y $\mathbf{c} \in \mathbb{R}^r$. Probar que

1. $A(I_n \otimes \mathbf{c}) = A \otimes \mathbf{c}^{\mathrm{t}}$.
2. $(\mathbf{c} \otimes I_p)B = \mathbf{c} \otimes B$.

Ejercicio 3. Probar que

1. Si A y B son simétricas, entonces $A \otimes B$ también es simétrica.
2. Si A y B son invertibles, entonces $A \otimes B$ también es invertible.
3. $A \otimes B = 0$ si, y sólo si, $A = 0$ ó $B = 0$.

Ejercicio 4. Hallar el rango de $A \otimes B$ donde

$$A = \begin{pmatrix} 2 & 6 \\ 1 & 4 \\ 3 & 1 \end{pmatrix} \quad \text{y} \quad B = \begin{pmatrix} 5 & 2 & 4 \\ 2 & 1 & 1 \\ 1 & 0 & 2 \end{pmatrix}.$$

Ejercicio 5. Sean $A \in \mathcal{M}_{m\times n}(\mathbb{R}),\ B \in \mathcal{M}_{n\times p}(\mathbb{R}),\ \mathbf{c} \in \mathbb{R}^p$ y $\mathbf{d} \in \mathbb{R}^n$. Probar que

1. $AB\mathbf{c} = (\mathbf{c}^{\mathrm{t}} \otimes A)\mathbf{vec}(B) = (A \otimes \mathbf{c}^{\mathrm{t}})\mathrm{vec}(B^{\mathrm{t}})$.
2. $\mathbf{d}^{\mathrm{t}}B\mathbf{c} = (\mathbf{c}^{\mathrm{t}} \otimes \mathbf{d}^{\mathrm{t}})\mathrm{vec}(B)$.

Ejercicio 6. Sean A, B y C matrices cuadradas de orden m. Probar que si C es simétrica, entonces

$$(\mathrm{vec}(C))^{\mathrm{t}}(A \otimes B)\mathrm{vec}(C) = (\mathrm{vec}(C))^{\mathrm{t}}(B \otimes A)\mathrm{vec}(C).$$

Ejercicio 7. Sean $A \in \mathcal{M}_{m\times n}(\mathbb{R})$ y $\mathbf{b} \in \mathbb{R}^p$. Probar que

$$\mathrm{vec}(A \otimes \mathbf{b}) = \mathrm{vec}(A) \otimes \mathbf{b}.$$

Ejercicio 8. Sean $A \in \mathcal{M}_{m\times n}(\mathbb{R})$ y $B \in \mathcal{M}_{n\times p}(\mathbb{R})$. Probar que

$$\mathrm{vec}(AB) = (I_p \otimes A)\mathrm{vec}(B) = (B^{\mathrm{t}} \otimes I_m)\mathrm{vec}(A) = (B^{\mathrm{t}} \otimes A)\mathrm{vec}(I_n).$$

Ejercicio 9. Sean $A \in \mathcal{M}_m(\mathbb{R}),\ B \in \mathcal{M}_n(\mathbb{R})$ y $C \in \mathcal{M}_{m\times n}(\mathbb{R})$. Probar que

$$\mathrm{vec}(AC + CB) = \big((I_n \otimes A) + (B^{\mathrm{t}} \otimes I_n)\big)\,\mathrm{vec}(C).$$

Ejercicio 10. Probar que la matriz de conmutación K_{mn} se puede escribir como

$$K_{mn} = \sum_{i=1}^{m} (\mathbf{e}_i \otimes I_n \otimes \mathbf{e}_i^{\mathrm{t}}),$$

donde $\mathbf{e}_i$ es el i-ésimo vector de la base canónica de I_m. Usar que si $A \in \mathcal{M}_{n\times m}(\mathbb{R})$, $\mathbf{x} \in \mathbb{R}^m$ y $\mathbf{y} \in \mathbb{R}^p m$ entonces

$$(K_{mn})^{\mathrm{t}}(\mathbf{x} \otimes A \otimes \mathbf{y}^{\mathrm{t}}) = A \otimes \mathbf{x}\mathbf{y}^{\mathrm{t}}.$$

Ejercicio 11. Sean $A \in \mathcal{M}_{m\times n}(\mathbb{R})$ de rango r y $\lambda_1, \ldots, \lambda_r$ los autovalores no nulos de $A^{\mathrm{t}}A$. Si definimos

$$P = K_{mn}(A^{\mathrm{t}} \otimes A),$$

probar que

1. P es simétrica.
2. $\mathrm{rg}(P) = r^2$.
3. $\mathrm{tr}(P) = \mathrm{tr}(A^{\mathrm{t}}A)$.
4. $P^2 = (AA^{\mathrm{t}}) \otimes (A^{\mathrm{t}}A)$.
5. los autovalores no nulos de P son $\lambda_1, \ldots, \lambda_r$ y $\pm(\lambda_i\lambda_j)^{1/2}$, para todo $i < j$.

Ejercicio 12. Sean $A \in \mathcal{M}_{m\times n}(\mathbb{R})$ y $B \in \mathcal{M}_{p\times q}(\mathbb{R})$. Probar que

1. $\mathrm{vec}(A^{\mathrm{t}} \otimes B) = (K_{mq,n} \otimes I_q)\,(\mathrm{vec}(A) \otimes \mathrm{vec}(B))\,.$
2. $\mathrm{vec}(A \otimes B^{\mathrm{t}}) = (I_n \otimes K_{p,mq}\,(\mathrm{vec}(A) \otimes \mathrm{vec}(B))\,.$

Ejercicio 13. Sean $A \in \mathcal{M}_{m\times n}(\mathbb{R})$ y $B \in \mathcal{M}_{p\times q}(\mathbb{R})$ con $mp = nq$. Probar que

$$\mathrm{tr}(A \otimes B) = (\mathrm{vec}(I_n) \otimes \mathrm{vec}(I_q))^{\mathrm{t}}\left(\mathrm{vec}(A) \otimes \mathrm{vec}(B^{\mathrm{t}})\right).$$

Ejercicio 14. Calcular la diferencial y la derivada de $f(\mathbf{x}) = A\mathbf{x}$ y de $g(x) = X\mathbf{a}$.

Ejercicio 15. Sea A y $B \in \mathcal{M}_m(\mathbb{R})$ y $\mathbf{x}$ un vector de m variables independientes. Hallar la diferencial y la derivada la función

$$f(x) = \frac{\mathbf{x}^{\mathrm{t}}A\mathbf{x}}{\mathbf{x}^{\mathrm{t}}B\mathbf{x}}.$$

Ejercicio 16. Sea X una matriz de orden m de variables independientes. Calcular la diferencial y la derivada de

1. $F(X) = \mathrm{tr}(X^2)$.
2. $F(X) = |X^2|$.

Ejercicio 17. Sean X una matriz invertible orden m de variables independientes, $A \in \mathcal{M}_m(\mathbb{R})$ y $\mathbf{a} \in \mathbb{R}^m$. Hallar la diferencial y la derivada de

1. $\mathrm{tr}(AX^{-1})$.
2. $\mathbf{a}^{\mathrm{t}}X^{-1}\mathbf{a}$.

Ejercicio 18. Sea X una matriz de orden $m \times n$ de variables independientes con rango n. Probar que

$$\frac{\partial}{\partial \mathrm{vec}(X)^{\mathrm{t}}}|X^{\mathrm{t}}X| = 2|X^{\mathrm{t}}X|\mathrm{vec}(X(X^{\mathrm{t}}X)^{-1})^{\mathrm{t}}.$$

Ejercicio 19. Sea $A \in \mathcal{M}_m(\mathbb{R})$ y X una matriz de orden m de variables independientes. Calcular las diferenciales y las derivadas de XAX^{t}, $X^{\mathrm{t}}AX$, XAX y $X^{\mathrm{t}}AX^{\mathrm{t}}$.

Ejercicio 20. Sean X una matriz de orden m de variables independientes y n un entero positivo. Probar que

$$\frac{\partial}{\partial \mathrm{vec}(X)^{\mathrm{t}}}\mathrm{vec}(X^n) = \sum_{i=1}^{n}\left((X^{n-i})^{\mathrm{t}} \otimes X^{i-1}\right).$$

Ejercicio 21. Sean $A \in \mathcal{M}_{n\times m}(\mathbb{R})$ y $B \in \mathcal{M}_{m\times n}(\mathbb{R})$. Si X es una matriz invertible de orden m de variables independientes, hallar la derivadas de

1. $\mathrm{vec}(AXB)$.
2. $\mathrm{vec}(AX^{-1}B)$.

Ejercicio 22. Sea X una matriz de orden $m \times n$ de variables independientes. Probar que

$$\frac{\partial}{\partial \mathrm{vec}(X)^{\mathrm{t}}}(X \otimes X) = (I_n \otimes K_{nm} \otimes I_m)\,(I_{mn} \otimes \mathrm{vec}(X) + \mathrm{vec}(X) \otimes I_{mn})$$

Ejercicio 23. Sean X una matriz invertible de orden m y $\mathrm{adj}(X)$ su matriz adjunta. Probar que

$$\frac{\partial}{\partial \mathrm{vec}(X)^{\mathrm{t}}}\mathrm{vec}(\mathrm{adj}(X)) = |X|\left(\mathrm{vec}(X^{-1})\mathrm{vec}((X^{-1})^{\mathrm{t}})^{\mathrm{t}} - ((X^{-1})^{\mathrm{t}} \otimes X^{-1})\right).$$

TEMA VIII

Normas vectoriales y matriciales

En el tema V estudiamos el concepto de norma en los espacios vectoriales euclídeos, nos proponemos ahora estudiar este mismo concepto con mayor generalidad. Para ello comenzaremos definiendo el concepto de norma de forma axiomática en cualquier espacio vectorial real o complejo de dimensión arbitraria. Evidentemente, un ejemplo destacado será el caso de las normas definidas a partir de un producto escalar en un espacio vectorial real de dimensión finita.

El par formado por un espacio vectorial V y una norma se conoce como espacio normado, estos espacios serán nuestro ambiente de trabajo en primera sección del tema. Estudiaremos algunas de sus propiedades elementales.

La introducción de una norma en un espacio vectorial nos permitirá definir la noción de convergencia para sucesiones de vectores, lo que a su vez nos permitirá hablar de límites y continuidad en los espacios normados. Tras estudiar algunos resultados elementales sobre convergencia y funciones continuas en espacios normados, introduciremos el concepto de normas equivalentes: diremos que dos normas son equivalentes si determinan la misma noción de convergencia; es decir, un sucesión es convergente para de las normas si, y sólo si, lo es para la otra. Es claro, por tanto, que las normas equivalentes también conservarán la noción de continuidad en idéntico sentido.

Terminamos esta primera sección del tema, mostrando que en los espacios de vectoriales de dimensión finita todas las normas son equivalentes, y concluiremos que las aplicaciones lineales entre espacios vectoriales de dimensión finita son funciones continuas.

La segunda sección del tema se dedica al estudio de las normas matriciales. La noción de norma matricial es una particularización de la noción de normas en los espacios vectoriales de las matrices cuadradas añadiendo una condición de compatibilidad con el producto de matrices. El primer caso de norma matricial estudiado es el de las normas matriciales subordinadas a una norma vectorial. Esto es, dada una norma $\|\cdot\|$ en $\mathbb{k}^n$ se puede definir una norma matricial $|||\cdot|||$ en $\mathcal{M}_n(\mathbb{k})$ tal que $\|A\mathbf{v}\| \leq |||A||| \, \|\mathbf{v}\|$, para todo $\mathbf{v} \in \mathbb{k}^n$, siendo $|||A|||$ el menor número real que verifica la desigualdad para todo $\mathbf{v} \in \mathbb{k}^n$. A continuación se muestran los ejemplos de

normas matriciales subordinadas más comunes y se dan sus expresiones expresiones explícitas.

Tal vez la norma matricial subordinada más importante es la que proviene de la norma usual de $\mathbb{k}^n$, es por esto por lo que dedicamos gran parte de nuestros esfuerzos a estudiar sus propiedades más interesantes; principalmente, aquellas que guardan relación con el radio espectral de la matriz. Es conveniente recordar ahora que gran parte de los resultados estudiados en los temas III y V serán fundamentales para alcanzar nuestro objetivo de relación las normas matriciales (y en particular la subordinada a la norma usual de $\mathbb{k}^n$) con el radio espectral. Esta relación pondrá de manifiesto (de nuevo, pues ya se vislumbró en el tema IV) que el mayor autovalor en módulo de una matriz cuadrada rige el comportamiento asintótico de las sucesiones de potencias de matrices, tal y como estudiaremos al final de la sección.

En esta segunda sección no todas las normas consideras serán subordinadas, se mostrarán ejemplos de normas no subordinadas y en todo momento se especificará qué resultados son sólo válidos para normas subordinadas y cuáles son válidos en general.

La última sección del tema se dedica el estudio del condicionamiento de sistemas de ecuaciones lineales $A\mathbf{x} = \mathbf{b}$ con $A \in \mathcal{M}_n(\mathbb{k})$ invertible y $\mathbf{b} \in \mathbb{k}^n$. Se dirá que un sistema está mal condicionado si pequeñas modificaciones en la matriz o en el término independientes producen grandes cambios en la solución del sistema. La herramienta clave para la detección de un buen o mal condicionamiento será las normas matriciales.

Para la elaboración de este tema hemos seguido esencialmente las secciones 2.3, 2.4 y el capítulo de 3 de [**IR99**] y las secciones 1.4 y 1.5 y el capítulo 3 de [**Cia82**]

1. Normas vectoriales. Espacios normados

A lo largo de este tema $\mathbb{k}$ denotará $\mathbb{R}$ ó $\mathbb{C}$, indistintamente, y en esta sección V y W serán espacios vectoriales sobre $\mathbb{k}$ de dimensión arbitraria, mientras no se indique lo contrario.

Definición VIII.1.1. Una **norma** sobre V es una aplicación $V \to \mathbb{R}$; $\mathbf{v} \mapsto \|\mathbf{v}\|$ tal que:

(a) $\|\mathbf{v}\| = 0$ si, y sólo si, $\mathbf{v} = \mathbf{0}$.
(b) $\|\lambda\mathbf{v}\| = |\lambda|\,\|\mathbf{v}\|$, para todo $\lambda \in \mathbb{k}$ y $\mathbf{v} \in V$.
(c) $\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$, para todo $\mathbf{u}$ y $\mathbf{v} \in V$.

La condición (c) se suele denominar *desigualdad triangular*. Por otra parte, como

$$0 = \|\mathbf{0}\| = \|\mathbf{v} - \mathbf{v}\| \leq \|\mathbf{v}\| + \|-\mathbf{v}\| = 2\|\mathbf{v}\|,$$

se tiene que $\|\mathbf{v}\| \geq 0$, para todo $\mathbf{v} \in V$.

Ejemplos VIII.1.2.

i) La función $\mathbb{R}^n \to \mathbb{R};\ \mathbf{v} = (v_1, \ldots, v_n) \mapsto \|\mathbf{v}\| = \sqrt{v_1^2 + \ldots + v_n^2}$ es una norma sobre $\mathbb{R}^n$. Esta norma se suele denominar **norma usual** de $\mathbb{R}^n$ y se denota $\|\cdot\|_2$. Obsérvese que, en este caso, se tiene que $\mathbf{v}^{\mathrm{t}}\mathbf{v} = \|\mathbf{v}\|_2^2$, para todo $\mathbf{v} \in \mathbb{R}^n$.

Obsérvese que la norma usual de $\mathbb{R}^n$ es la norma del espacio vectorial euclídeo $\mathbb{R}^n$ para el producto escalar usual estudiada en el tema V.

También son normas sobre $\mathbb{R}^n$ las dos siguientes:

$$\begin{array}{rcl} \|\mathbf{v}\|_1 & = & |v_1| + \ldots + |v_n|, \\ \|\mathbf{v}\|_\infty & = & \text{máx}\{|v_1|, \ldots, |v_n|\}. \end{array}$$

ii) La función $\mathbb{C}^n \to \mathbb{R};\ \mathbf{v} = (v_1, \ldots, v_n) \mapsto \|\mathbf{v}\|_2 = \sqrt{|v_1|^2 + \ldots + |v_n|^2}$ es una norma sobre $\mathbb{C}^n$, que se llama **norma usual** de $\mathbb{C}^n$. Nótese que, en este caso, se cumple que $\mathbf{v}^*\,\mathbf{v} = \|\mathbf{v}\|_2^2$ para todo $\mathbf{v} \in \mathbb{C}^n$.

También son normas sobre $\mathbb{C}^n$ las dos siguientes:

$$\begin{array}{rcl} \|\mathbf{v}\|_1 & = & |v_1| + \ldots + |v_n|, \\ \|\mathbf{v}\|_\infty & = & \text{máx}\{|v_1|, \ldots, |v_n|\}. \end{array}$$

Nota VIII.1.3. La desigualdad triangular de la norma determina, para todo par de vectores $\mathbf{u}$ y $\mathbf{v} \in V$ las desigualdades

$$\left\{ \begin{array}{rcl} \|\mathbf{u}\| & = & \|\mathbf{v} + (\mathbf{u} - \mathbf{v})\| \leq \|\mathbf{v}\| + \|\mathbf{u} - \mathbf{v}\|, \\ \|\mathbf{v}\| & = & \|\mathbf{u} + (\mathbf{v} - \mathbf{u})\| \leq \|\mathbf{u}\| + \|\mathbf{v} - \mathbf{u}\|. \end{array} \right.$$

Como $\|\mathbf{v} - \mathbf{u}\| = \|\mathbf{u} - \mathbf{v}\|$ se deduce la desigualdad

$$\Big|\|\mathbf{u}\| - \|\mathbf{v}\|\Big| \leq \|\mathbf{u} - \mathbf{v}\|, \tag{VIII.1.1}$$

para todo $\mathbf{u}, \mathbf{v} \in V$.

Definición VIII.1.4. Un espacio vectorial con una norma se llama **espacio normado**.

Nótese que un subespacio vectorial de un espacio normado es un espacio normado con la misma norma restringida al subespacio.

Ejemplos de espacios normados son los del ejemplo VIII.1.2 y los siguientes. Otros espacios normados se verán en el ejemplo XII.1.8.

Ejemplos VIII.1.5.

i) En el espacio vectorial de los polinomios con coeficientes reales de grado menor o igual que n, $\mathbb{R}[x]_{\leq n}$, la aplicación

$$\mathbb{R}[x]_{\leq n} \longrightarrow \mathbb{R}; p(x) \longmapsto ||p(x)|| = \left(\sum_{i=0}^{n}(p(i))^2\right)^{1/2}$$

es una norma.

ii) Sea $[a,b]$ un intervalo cerrado de $\mathbb{R}$. En el espacio vectorial de las funciones continuas reales de $[a,b]$, $\mathcal{C}([a,b];\mathbb{R})$, las siguientes aplicaciones de $\mathcal{C}([a,b];\mathbb{R})$ en $\mathbb{R}$ son normas:

$$f \longmapsto \|f\|_1 = \int_b^a f(x)\mathrm{d}x$$
$$f \longmapsto \|f\|_2 = \left(\int_b^a f(x)^2\mathrm{d}x\right)^{1/2}$$
$$f \longmapsto \|f\|_\infty = \sup_{x\in[a,b]}|f(x)|$$

Obsérvese que esta última aplicación está bien definida por el teorema A.4.9.

Evidentemente, es posible definir diferentes normas sobre el mismo espacio vectorial (véase el ejemplo VIII.1.2.i)). Por consiguiente, para definir un espacio normado necesitamos especificar tanto el espacio vectorial como la norma. Podemos decir pues que un espacio normado es un par $(V, \|\cdot\|)$, donde V es un espacio vectorial y $\|\cdot\|$ es una norma sobre V. No obstante, algunos espacios vectoriales están tradicionalmente equipados de una norma *usual*. Por ejemplo, cuando digamos *el espacio normado* $\Bbbk^n$ entenderemos que la norma es

$$\|\mathbf{v}\|_2 = \sqrt{|x_1|^2 + \ldots + |x_n|^2}.$$

Análogamente, las normas definidas en los ejemplos VIII.1.2.iii)-iv) son las usuales. De modo que cuando queramos considerar normas distintas a la usual en estos espacios diremos algo como "consideremos el espacio ... con la norma definida por ..." .

Proposición VIII.1.6. *Sea* $(V, \|\cdot\|)$ *un espacio normado.*

(a) $\|\mathbf{u}-\mathbf{v}\| \geq 0$, *para todo* $\mathbf{u}$ *y* $\mathbf{v} \in V$; *además,* $\|\mathbf{u}-\mathbf{v}\| = 0$ *si, y sólo si,* $\mathbf{u} = \mathbf{v}$.
(b) $\|\mathbf{u}-\mathbf{v}\| = \|\mathbf{v}-\mathbf{u}\|$, *para todo* $\mathbf{u}$ *y* $\mathbf{v} \in V$.
(c) $\|\mathbf{u}-\mathbf{w}\| \leq \|\mathbf{u}-\mathbf{v}\| + \|\mathbf{v}-\mathbf{w}\|$.

Demostración. La demostración de esta proposición se deja como ejercicio al lector. ∎

Sea $(V, \|\cdot\|)$ un espacio normado. De la proposición anterior se deduce que la aplicación $\mathrm{d} : V \times V \to \mathbb{R}$; $(\mathbf{u},\mathbf{v}) \mapsto \mathrm{d}(\mathbf{u},\mathbf{v}) := \|\mathbf{u}-\mathbf{v}\|$ es una métrica sobre V. Por consiguiente,

Corolario VIII.1.7. *Todo espacio normado* $(V, \|\cdot\|)$ *tiene una estructura natural de espacio métrico determinada por la métrica*

$$\mathrm{d}(\mathbf{u}, \mathbf{v}) := \|\mathbf{u} - \mathbf{v}\|.$$

Además, esta métrica es

(a) *invariante por traslaciones, es decir,*

$$\mathrm{d}(\mathbf{u} + \mathbf{w}, \mathbf{v} + \mathbf{w}) = \mathrm{d}(\mathbf{u}, \mathbf{v}),$$

para todo $\mathbf{u}, \mathbf{v}$ *y* $\mathbf{w} \in V$.

(b) *absolutamente homogénea por homotecias, es decir,*

$$\mathrm{d}(\lambda\mathbf{u}, \lambda\mathbf{v}) = |\lambda|\,\mathrm{d}(\mathbf{u}, \mathbf{v}),$$

para todo $\mathbf{u}$ *y* $\mathbf{v} \in V$ *y* $\lambda \in \Bbbk$.

Demostración. La primera parte es consecuencia directa de la proposición VIII.1.6. La demostración de la segunda parte del corolario se deja como ejercicio al lector. ■

Según el corolario anterior, siempre que tengamos un espacio normado, tenemos un espacio métrico con todas sus propiedades, definiciones, topología, etc.

Convergencia en espacios normados.

El valor absoluto es un norma en $\mathbb{R}$, y se usa para definir el concepto de convergencia, en pocas palabras el valor absoluto de la diferencia de dos números reales es la distancia entre éstos y la convergencia trata sobre "acercarse tanto como se desee al punto límite". En general, la norma sobre un espacio vectorial juega un papel similar. Mientras que $\|\mathbf{v}\|$ se puede interpretar como la magnitud de $\mathbf{v}$, $\|\mathbf{u} - \mathbf{v}\|$ proporciona una medida de la distancia entre $\mathbf{u}$ y $\mathbf{v}$. De modo que podemos recuperar la noción de convergencia de los espacios métricos.

Definición VIII.1.8. Sea $(V, \|\cdot\|)$ un espacio normado. Diremos que una sucesión $(\mathbf{v}_n)_{n\in\mathbb{N}}$ de elementos de V **converge** a $\mathbf{v} \in V$, si para todo $\varepsilon > 0$ existe un número N tal que para todo $n \geq N$ se tiene que $\|\mathbf{v}_n - \mathbf{v}\| < \varepsilon$. En este caso se escribe $\lim_{n\to\infty} \mathbf{v}_n = \mathbf{v}$ o simplemente $\mathbf{v}_n \to \mathbf{v}$.

La definición anterior es bastante más simple si recurrimos al concepto de convergencia de los números reales: $\mathbf{v}_n \to \mathbf{v}$ en V significa que $\|\mathbf{v}_n - \mathbf{v}\| \to 0$ en $\mathbb{R}$. La convergencia en un espacio normado tiene las propiedades básicas de la convergencia en $\mathbb{R}$:

- *Una sucesión convergente tiene un único límite.*
- *Si* $\mathbf{v}_n \to \mathbf{v}$ *y* $\lambda_n \to \lambda$, *entonces* $\lambda_n\mathbf{v}_n \to \lambda\mathbf{v}$, *siendo* $(\lambda_n)_{n\in\mathbb{N}}$ *una sucesión de escalares y* λ *un escalar.*
- *Si* $\mathbf{u}_n \to \mathbf{u}$ *y* $\mathbf{v}_n \to \mathbf{v}$, *entonces* $\mathbf{u}_n + \mathbf{v}_n \to \mathbf{u} + \mathbf{v}$.

Todas estas propiedades se demuestran de la misma manera que se hacia en el caso de la convergencia en $\mathbb{R}$, por lo que su comprobación de deja como ejercicio al lector.

Ejemplo VIII.1.9. La sucesión de vectores $(\mathbf{v}_n)_{n\in\mathbb{N}}$ de $\mathbb{R}^3$ con $\mathbf{v}_n = \left(2/n^3, 1 - 1/n^2, \mathrm{e}^{1/n}\right)^{\mathrm{t}} \in \mathbb{R}^3$ es convergente al vector $\mathbf{v} = \lim_{n\to\infty} \mathbf{v}_n = (0,1,1)^{\mathrm{t}}$.

Al igual que ocurre con el concepto de convergencia, la continuidad en espacios métricos tiene su traducción inmediata a los espacios normados.

Definición VIII.1.10. Sean $(V, \|\cdot\|_V)$ y $(W, \|\cdot\|_W)$ dos espacios normados. Se dice que una aplicación $f : V \to W$ es **continua en** $\mathbf{v}_0$ si para cada $\varepsilon > 0$, existe $\delta > 0$ tal que $\|\mathbf{v}_0 - \mathbf{v}\|_V < \delta$ implica que $\|f(\mathbf{v}_0) - f(\mathbf{v})\|_W < \varepsilon$.

Si f es continua en todo $\mathbf{v} \in V$, se dice que es **continua** en V.

Proposición VIII.1.11. *Sea $\|\cdot\|$ una norma sobre V. La aplicación $\|\cdot\| : V \to \mathbb{R};\ \mathbf{v} \mapsto \|\mathbf{v}\|$ es continua.*

Demostración. Dados $\mathbf{u} \in V$ y $\varepsilon > 0$ cualesquiera basta tomar $\delta = \varepsilon$ y $\mathbf{v} \in V$ con $\|\mathbf{u}-\mathbf{v}\| < \delta$ para que, aplicando la desigualdad (VIII.1.1), se verifique $\big|\|\mathbf{u}\|-\|\mathbf{v}\|\big| < \varepsilon$. ∎

Proposición VIII.1.12. *Sean $(V, \|\cdot\|_V)$ y $(W, \|\cdot\|_W)$ dos espacios normados y $f : V \to W$ es una aplicación lineal. Las siguientes afirmaciones son equivalentes:*

(a) *f es continua en un punto.*
(b) *f es continua.*
(c) *f es acotada en $B[\mathbf{0}, 1]$.*
(d) *Existe $M > 0$ tal que $\|f(\mathbf{v})\|_W \leq M\|\mathbf{v}\|_V$, para todo $\mathbf{v} \in V$.*

Demostración. $\boxed{\text{(a)} \Rightarrow \text{(b)}}$ Basta comprobar que f es continua en $\mathbf{v}_0 \in V$ si, y sólo si, lo es en $\mathbf{0}$. Lo cual es evidente si tenemos en cuenta que si para cada $\varepsilon > 0$ existe $\delta > 0$ tal que $\|\mathbf{v}_0 - \mathbf{v}\|_V < \delta$ implica $\|f(\mathbf{v}_0) - f(\mathbf{v})\|_W < \varepsilon$, entonces $\|(\mathbf{v}_0 - \mathbf{v}) - \mathbf{0}\|_V < \delta$ implica $\|f(\mathbf{v}_0 - \mathbf{v}) - f(\mathbf{0})\|_W < \varepsilon$, y recíprocamente.

$\boxed{\text{(b)} \Rightarrow \text{(c)}}$ Como f es continua en $\mathbf{0}$ se tiene que existe $\delta > 0$ tal que $\|\mathbf{0} - \mathbf{v}\|_V = \|\mathbf{v}\|_V < \delta$ implica que $\|f(\mathbf{0}) - f(\mathbf{v})\|_W = \|f(\mathbf{v})\|_W < 1$. Por tanto, si $\mathbf{u} \in B(\mathbf{0}, 1)$, es decir, $\|\mathbf{u}\|_V < 1$, se tiene que $\mathbf{v} = \delta\mathbf{u}$ cumple que $\|\mathbf{v}\|_V < \delta$, luego $\|f(\mathbf{v})\|_W < 1$. De este modo concluimos que $\|f(\mathbf{u})\|_W = \|f(\mathbf{v}/\delta)\|_W = \|f(\mathbf{v})/\delta\|_W = \|f(\mathbf{v})\|_W/\delta < 1/\delta$.

$\boxed{\text{(c)} \Rightarrow \text{(d)}}$ Si M es la cota de f en $B[0,1]$, entonces $\|f(\mathbf{v}/\|\mathbf{v}\|_V)\|_W < M$; de donde se sigue que $\|f(\mathbf{v})\|_W < M\|\mathbf{v}\|_V$, para todo $\mathbf{v} \in V$.

$\boxed{\text{(d)} \Rightarrow \text{(a)}}$ Por hipótesis, existe $M > 0$ tal que $\|f(\mathbf{v})\|_W < M\|\mathbf{v}\|_V$, para todo $\mathbf{v} \in V$. Ahora, dado $\varepsilon > 0$, basta tomar $\delta = \varepsilon/M$ para concluir que f es continua en $\mathbf{0}$. ∎

Definición VIII.1.13. Dos normas sobre el mismo espacio vectorial se dicen **equivalentes** si definen la misma convergencia. Más concretamente, dos normas $\|\cdot\|$ y $\|\cdot\|'$ sobre un espacio vectorial V son equivalentes si para cualquier sucesión $(\mathbf{v}_n)_{n\in\mathbb{N}}$ en V y $\mathbf{v}\in V$,

$$\|\mathbf{v}_n-\mathbf{v}\|\to 0 \text{ si, y sólo si, } \|\mathbf{v}_n-\mathbf{v}\|'\to 0.$$

El siguiente teorema proporciona un criterio práctico para la equivalencia de normas. La condición del teorema es usada a menudo como definición de equivalencia de normas.

Teorema VIII.1.14. *Sean $\|\cdot\|$ y $\|\cdot\|'$ dos normas sobre un espacio vectorial V. Las normas $\|\cdot\|$ y $\|\cdot\|'$ son equivalentes si, y sólo si, existen dos números positivos m y M tales que*

$$m\,\|\mathbf{v}\|\le\|\mathbf{v}\|'\le M\,\|\mathbf{v}\|,$$

para todo $\mathbf{v}\in V$.

Demostración. Es claro que la condición implica la equivalencia de las normas $\|\cdot\|$ y $\|\cdot\|'$. Supongamos pues que las normas son equivalentes, esto es $\|\mathbf{v}_n-\mathbf{v}\|\to 0$ si, y sólo si, $\|\mathbf{v}_n-\mathbf{v}\|'\to 0$. Si no existe $m>0$ tal que $m\|\mathbf{v}\|\le\|\mathbf{v}\|'$ para todo $\mathbf{v}\in V$, entonces para cada $n\in\mathbb{N}$ existe $\mathbf{v}_n\in V$ tal que

$$\frac{1}{n}\|\mathbf{v}_n\|>\|\mathbf{v}_n\|'.$$

Definamos

$$\mathbf{w}_n=\frac{1}{\sqrt{n}}\frac{\mathbf{v}_n}{\|\mathbf{v}_n\|'}.$$

Entonces $\|\mathbf{w}_n\|'=1/\sqrt{n}\to 0$. Por otra parte, $\|\mathbf{w}_n\|>n\,\|\mathbf{w}_n\|'=\sqrt{n}$. Esta contradicción demuestra que el número m con la propiedad requerida ha de existir. La existencia del número M se demuestra análogamente. ∎

Terminamos esta sección mostrando algunos resultados sobre espacios normados de dimensión finita.

Teorema VIII.1.15. *Sea V un espacio vectorial de dimensión finita $n>0$. Todas las normas sobre V son equivalentes.*

Demostración. Como V es isomorfo a $\mathbb{R}^n$, cualquier norma $\|\cdot\|$ sobre V induce una norma en $\mathbb{R}^n$; en efecto, si $(\lambda_1,\dots,\lambda_n)\in\mathbb{R}^n$ son las coordenadas de $\mathbf{v}\in V$ respecto de alguna base de V, entonces $\|(\lambda_1,\dots,\lambda_n)\|:=\|\mathbf{v}\|$ es una norma sobre $\mathbb{R}^n$. De modo que basta demostrar que todas las normas sobre $\mathbb{R}^n$ son equivalentes. De hecho vamos a probar que cualquier norma $\|\cdot\|$ sobre $\mathbb{R}^n$ es equivalente a la norma $\|(\lambda_1,\dots,\lambda_n)\|_\infty:=\max\{|\lambda_1|,\dots,|\lambda_n|\}$.

Sea $\mathbf{e}_1, \ldots, \mathbf{e}_n$ la base estándar de $\mathbb{R}^n$, donde $\mathbf{e}_1 = (1, 0, \ldots, 0), \mathbf{e}_2 = (0, 1, \ldots, 0)$, etcetera. Entonces, dado $\mathbf{v} = (\lambda_1, \ldots, \lambda_n) \in \mathbb{R}^n$ se tiene que

$$\|\mathbf{v}\| = \|\sum_{i=1}^{n} \lambda_i \mathbf{e}_i\| \leq \sum_{i=1}^{n} |\lambda_i| \, \|\mathbf{e}_i\| \leq n \cdot \left(\max_i |\lambda_i|\right) \left(\max_i \|\mathbf{e}_i\|\right) = M \, \|\mathbf{v}\|_\infty,$$

donde $M := n \cdot \max_i \|\mathbf{e}_i\|$.

Definamos ahora la función $f : \mathbb{R}^n \to \mathbb{R}; \mathbf{v} \mapsto f(\mathbf{v}) := \|\mathbf{v}\|$. La función f es continua respecto de la norma $\|\cdot\|_\infty$, pues $|f(\mathbf{u}) - f(\mathbf{v})| = \big|\|\mathbf{u}\| - \|\mathbf{v}\|\big| \leq \|\mathbf{u} - \mathbf{v}\| \leq M \, \|\mathbf{u} - \mathbf{v}\|_\infty$.

Sea $S := \{\mathbf{u} \in \mathbb{R}^n \mid \|\mathbf{u}\|_\infty = 1\}$, el conjunto S es compacto para la norma $\|\cdot\|_\infty$. Luego, f alcanza un máximo y un mínimo en S. Sea $m := f(\mathbf{w})$ tal que $\mathbf{w} \in S$ con $f(\mathbf{w}) \leq f(\mathbf{u})$, para todo $\mathbf{u} \in S$. Es decir, $\|\mathbf{u}\| \geq m$ para todo $\mathbf{u} \in S$. Nótese que $m \neq 0$; en otro caso, $0 = f(\mathbf{w}) = \|\mathbf{w}\|$ implica que $\mathbf{w} = \mathbf{0}$, pero $\mathbf{0} \notin S$.

Finalmente, dado $\mathbf{v} \in \mathbb{R}^n$, se tiene que $\mathbf{u} := \mathbf{v}/\|\mathbf{v}\|_\infty$ pertenece a S. De donde se sigue que $\|\mathbf{v}\|/\|\mathbf{v}\|_\infty = \|\mathbf{u}\| \geq m$ y por consiguiente que

$$m \, \|\mathbf{v}\|_\infty \leq \|\mathbf{v}\|.$$

■

El resultado anterior no es generalizable a espacios vectorial de dimensión arbitraria. Por ejemplo, en el espacio vectorial de las funciones reales continuas en el intervalo $[0, 1]$, $\mathcal{C}([0, 1]; \mathbb{R})$, las normas $\|\cdot\|_\infty$ y $\|\cdot\|_1$ definidas por

$$\|f\|_\infty = \sup_{x \in [0,1]} |f(x)| \quad \text{y} \quad \|f\|_1 = \int_0^1 f(x) \mathrm{d}x$$

no son equivalentes.

Corolario VIII.1.16. *Sean $(V, \|\cdot\|_V)$ y $(W, \|\cdot\|_W)$ dos espacios normados dimensión finita sobre $\Bbbk$. Cualquier aplicación lineal $f : V \to W$ es continua.*

Demostración. Supongamos que $\dim(V) = n$ y sea $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ una base de V. Dado $\mathbf{v} \in V$, existen $\lambda_i \in \Bbbk$, $i = 1, \ldots, n$ tales que $\mathbf{v} = \sum_{i=1}^n \lambda_i \mathbf{v}_i$. Luego, $f(\mathbf{v}) = \sum_{i=1}^n \lambda_i f(\mathbf{v}_i)$; sea $M_1 := \max_{1 \leq i \leq n} \|f(\mathbf{v}_i)\|_W$. Por otra parte, la aplicación $\|\cdot\|_1 : V \to \Bbbk; \ \mathbf{v} \mapsto \|\mathbf{v}\|_1 = \sum_{i=1}^n \lambda_i$ es una norma sobre V, y como todas las normas sobre V son equivalentes, existe un número positivo M_2 tal que $\|\cdot\|_1 \leq M_2 \|\cdot\|_V$.

De tal forma,

$$\begin{aligned} \|f(\mathbf{v})\|_W &= \left\| \sum_{i=1}^n \lambda_i f(\mathbf{v}_i) \right\|_W \leq \sum_{i=1}^n \lambda_i \|f(\mathbf{v}_i)\|_W \leq \sum_{i=1}^n \lambda_i \max_{1 \leq i \leq n} \|f(\mathbf{v}_i)\|_W \\ &= M_1 \|\mathbf{v}\|_1 \leq (M_1 M_2) \, \|\mathbf{v}\|_V \end{aligned}$$

De donde se sigue el resultado buscado. ■

2. Normas matriciales

Definición VIII.2.1. Una **norma matricial** es una aplicación $|\!|\!| \cdot |\!|\!| : \mathcal{M}_n(\Bbbk) \to \mathbb{R}$ verificando las siguientes propiedades:

(a) $|\!|\!| A |\!|\!| = 0$ si, y sólo si, $A = \mathbf{0}$.
(b) $|\!|\!| \lambda A |\!|\!| = |\lambda| \, |\!|\!| A |\!|\!|$, para todo $A \in \mathcal{M}_n(\Bbbk)$ y $\lambda \in \Bbbk$.
(c) $|\!|\!| A + B |\!|\!| \leq |\!|\!| A |\!|\!| + |\!|\!| B |\!|\!|$, para todo A y $B \in \mathcal{M}_n(\Bbbk)$.
(d) $|\!|\!| A\,B |\!|\!| \leq |\!|\!| A |\!|\!| \, |\!|\!| B |\!|\!|$, para todo A y $B \in \mathcal{M}_n(\Bbbk)$.

Las propiedades (a)-(c) aseguran que toda norma matricial es una norma sobre el espacio vectorial $\mathcal{M}_n(\Bbbk)$ y la propiedad (d) proporciona la "compatibilidad" de la norma con el producto de matrices.

Es claro que, al tratarse de una norma, se cumple que $|\!|\!| A |\!|\!| \geq 0$ para todo $A \in \mathcal{M}_n(\Bbbk)$; en efecto, $0 = |\!|\!| \mathbf{0} |\!|\!| = |\!|\!| A + (-A) |\!|\!| \leq |\!|\!| A |\!|\!| + |\!|\!| - A |\!|\!| = 2 |\!|\!| A |\!|\!|$.

Antes de mostrar algún ejemplo de norma matricial, veamos que toda norma vectorial tiene asociada una norma matricial.

Proposición VIII.2.2. *Sea $\| \cdot \|$ una norma vectorial sobre $V = \Bbbk^n$. La aplicación*

$$|\!|\!| \cdot |\!|\!| : \mathcal{M}_n(\Bbbk) \longrightarrow \mathbb{R}, \quad A \longmapsto |\!|\!| A |\!|\!| := \sup_{\mathbf{v} \neq 0} \frac{\|A\mathbf{v}\|}{\|\mathbf{v}\|} = \sup_{\|\mathbf{u}\|=1} \|A\mathbf{u}\|$$

es una norma matricial.

Demostración. Dado $\mathbf{v} \neq 0$, podemos considerar $\mathbf{u} := \mathbf{v}/\|\mathbf{v}\|$, de donde se sigue la igualdad de los dos supremos.

La aplicación $|\!|\!| \cdot |\!|\!|$ está bien definida debido a la continuidad de la aplicación $\mathbf{u} \mapsto \|A\mathbf{u}\|$ (que podemos entender como la composición de las aplicaciones continuas $\mathbf{u} \mapsto A\mathbf{u} \mapsto \|A\mathbf{u}\|$) sobre la esfera unidad, $\{\mathbf{u} \in V : \|\mathbf{u}\| = 1\}$, que es un compacto de V; luego, por el teorema A.4.9, tenemos garantizado que $\sup\{\|A\mathbf{u}\| : \mathbf{u} = 1\} < \infty$.

Veamos ahora que se trata de una norma matricial. La primera propiedad es trivial; en efecto, si $\|A\mathbf{v}\| = 0$, para todo $\mathbf{v} \in V$ no nulo, entonces $A\mathbf{v} = \mathbf{0}$ para todo $\mathbf{v} \in V$ de donde se sigue que A es la matriz nula. Por otra parte, tenemos que

$$|\!|\!| \lambda A |\!|\!| = \sup_{\|\mathbf{u}\|=1} \|\lambda A\mathbf{u}\| = \sup_{\|\mathbf{u}\|=1} |\lambda| \, \|A\mathbf{u}\| = |\lambda| \sup_{\|\mathbf{u}\|=1} \|A\mathbf{u}\| = |\lambda| \, |\!|\!| A |\!|\!|.$$

Para la siguiente propiedad

$$|\!|\!| A + B |\!|\!| = \sup_{\|\mathbf{u}\|=1} \|(A+B)\mathbf{u}\| \leq \sup_{\|\mathbf{u}\|=1} \|A\mathbf{u}\| + \sup_{\|\mathbf{u}\|=1} \|B\mathbf{u}\| = |\!|\!| A |\!|\!| + |\!|\!| B |\!|\!|.$$

Finalmente, sea $\mathbf{u} \in V$ tal que $\|\mathbf{u}\| = 1$ y llamemos $\mathbf{v} = B\mathbf{u}$. Si $\mathbf{v} = \mathbf{0}$, entonces $\|AB\mathbf{u}\| = 0 \leq |\!|\!| A |\!|\!| \, |\!|\!| B |\!|\!|$; en otro caso,

$$\begin{aligned} \|AB\mathbf{u}\| = \|A\mathbf{v}\| = \left\| A \frac{\mathbf{v}}{\|\mathbf{v}\|} \, \|\mathbf{v}\| \right\| = \|\mathbf{v}\| \left\| A \frac{\mathbf{v}}{\|\mathbf{v}\|} \right\| \leq \|\mathbf{v}\| \, |\!|\!| A |\!|\!| \\ = |\!|\!| A |\!|\!| \, \|B\mathbf{u}\| \leq |\!|\!| A |\!|\!| \, |\!|\!| B |\!|\!|. \end{aligned}$$

Por consiguiente, $\|AB\mathbf{u}\| \leq |\!|\!| A |\!|\!| \, |\!|\!| B |\!|\!|$, para todo $\mathbf{u}$ en la esfera unidad; en particular,

$$|\!|\!| A\, B |\!|\!| = \sup_{\|\mathbf{u}\|=1} \|(A\, B)\mathbf{u}\| \leq |\!|\!| A |\!|\!| \, |\!|\!| B |\!|\!|.$$

■

Definición VIII.2.3. La norma $|\!|\!| \cdot |\!|\!|$ dada en la proposición VIII.2.2 se denomina **norma matricial subordinada** a la norma vectorial $\| \cdot \|$.

Ejemplo VIII.2.4. De forma habitual utilizaremos las siguientes normas matriciales subordinadas:

$$|\!|\!| A |\!|\!|_1 := \sup_{\|\mathbf{u}\|=1} \|A\mathbf{u}\|_1, \quad |\!|\!| A |\!|\!|_2 := \sup_{\|\mathbf{u}\|=1} \|A\mathbf{u}\|_2 \quad \text{y} \quad |\!|\!| A |\!|\!|_\infty := \sup_{\|\mathbf{u}\|=1} \|A\mathbf{u}\|_\infty.$$

No obstante, conviene advertir que existen normas matriciales que no están subordinadas a ninguna norma vectorial (véase la proposición VIII.2.14).

Veamos ahora algunas propiedades importantes de las normas matriciales subordinadas.

Proposición VIII.2.5. *Sea $|\!|\!| \cdot |\!|\!|$ una norma matricial subordinada a una norma vectorial $\| \cdot \|$ sobre $V = \mathbb{k}^n$. Se cumple que:*

(a) $\|A\mathbf{v}\| \leq |\!|\!| A |\!|\!| \, \|\mathbf{v}\|$, *para todo* $A \in \mathcal{M}_n(\mathbb{k})$ *y* $\mathbf{v} \in V$.
(b) $|\!|\!| A |\!|\!| = \inf\{\lambda \geq 0 : \|A\mathbf{v}\| \leq \lambda \, \|\mathbf{v}\|, \mathbf{v} \in V\}$.
(c) *Existe* $\mathbf{u} \in V$ *tal que* $\|A\mathbf{u}\| = |\!|\!| A |\!|\!| \, \|\mathbf{u}\|$.
(d) $|\!|\!| I_n |\!|\!| = 1$.

Demostración. Los apartados (a), (b) y (d) se obtienen directamente de la proposición VIII.2.2. Para demostrar (c) basta tener en cuenta la continuidad de la aplicación $\|\mathbf{v}\| \mapsto \|A\mathbf{v}\|$ sobre la esfera unidad (que es compacta) para concluir que el supremo de la proposición VIII.2.2 se alcanza (véase el teorema A.4.9). De este modo, si $\mathbf{u} \in V$ con $\|\mathbf{u}\| = 1$ verifica $|\!|\!| A |\!|\!| = \|A\mathbf{u}\|$, entonces $\|A\mathbf{u}\| = |\!|\!| A |\!|\!| \, \|\mathbf{u}\|$. ■

Nota VIII.2.6. Dada $A \in \mathcal{M}_n(\mathbb{k})$, a la vista del apartado (b) de la proposición VIII.2.5, si existe una constante $M \geq 0$ tal que para una norma matricial subordinada $|\!|\!| \cdot |\!|\!|$ a una norma vectorial $\| \cdot \|$ sobre $V = \mathbb{k}^n$, se verifica

(a) $\|A\mathbf{v}\| \leq M \, \|\mathbf{v}\|$, para todo $\mathbf{v} \in V$;

(b) Existe $\mathbf{u} \in V$ tal que $\|A\mathbf{u}\| = M\|\mathbf{u}\|$,

entonces $M = |||A|||$.

A continuación mostremos expresiones explícitas para las normas matriciales subordinadas del ejemplo VIII.2.4. Para facilitar su compresión conviene recordar la definición III.2.10 donde se introdujeron los conceptos de espectro y de radio espectral de una matriz.

Teorema VIII.2.7. *Sea $A = (a_{ij}) \in \mathcal{M}_n(\Bbbk)$.*

(a) $|||A|||_1 = \max_{1\leq j\leq n} \sum_{i=1}^n |a_{ij}|$, *es decir, la norma $|||\cdot|||_1$ viene dada por la mayor de todas las cantidades que se obtienen al sumar los módulos de los elementos de cada columna.*

(b) $|||A|||_2 = \sqrt{\varrho(A^*A)} = \sqrt{\varrho(A\,A^*)} = |||A^*|||_2$.

(c) $|||A|||_\infty = \max_{1\leq i\leq n} \sum_{j=1}^n |a_{ij}|$, *es decir, la norma $|||\cdot|||_\infty$ viene dada por la mayor de todas las cantidades que se obtienen al sumar los módulos de los elementos de cada fila.*

Demostración. Como es habitual denotaremos $V = \Bbbk^n$.

(a) Para todo $\mathbf{v} \in V$ se verifica que

$$\begin{aligned}\|A\mathbf{v}\|_1 &= \sum_{i=1}^n |(A\,\mathbf{v})_i| = \sum_{i=1}^n \left|\sum_{j=1}^n a_{ij}v_j\right| \leq \sum_{i=1}^n \left(\sum_{j=1}^n |a_{ij}||v_j|\right)\\ &= \sum_{j=1}^n \left(\sum_{i=1}^n |a_{ij}|\,|v_j|\right) = \sum_{j=1}^n |v_j| \sum_{i=1}^n |a_{ij}| \leq \left(\max_{1\leq j\leq n} \sum_{i=1}^n |a_{ij}|\right) \|\mathbf{v}\|_1.\end{aligned}$$

Consideremos el vector $\mathbf{u} \in V$ de coordenadas

$$u_i = \delta_{i\,j_0} = \begin{cases} 1 & \text{si} \quad i = j_0;\\ 0 & \text{si} \quad i \neq j_0,\end{cases}$$

donde j_0 es un subíndice que verifica

$$\max_{i\leq j\leq n} \sum_{i=1}^n |a_{ij}| = \sum_{i=1}^n |a_{i\,j_0}|.$$

Como para este vector se tiene que $\|\mathbf{u}\|_1 = 1$ y

$$\begin{aligned}\|A\mathbf{u}\|_1 &= \sum_{i=1}^n |(A\mathbf{u})_i| = \sum_{i=1}^n |a_{i\,j_0}u_{j_0}| = \sum_{i=1}^n |a_{i\,j_0}|\\ &= \left(\max_{1\leq j\leq n} \sum_{i=1}^n |a_{ij}|\right) \|\mathbf{u}\|_1,\end{aligned}$$

de donde se sigue que

$$\|\!|A|\!\|_1 = \max_{1\leq j\leq n} \sum_{i=1}^{n} |a_{ij}|.$$

(b) Por un lado, como la matrices $A\,A^*$ y A^*A son hermíticas tienen todos sus autovalores reales (véanse las proposiciones V.5.18 y V.5.14). Además, usando los mismos argumentos que en la demostración de la proposición VI.1.2, se comprueba que $\mathrm{sp}(A\,A^*) = \mathrm{sp}(A^*A)$, de donde se sigue que $\varrho(A^*A) = \varrho(A\,A^*)$.

Por otra parte, como la matriz A^*A es hermítica, es normal y lo por tanto diagonalizable por una matriz de paso unitaria (véase el teorema V.5.15), es decir,

$$Q^*A^*AQ = D = \mathrm{diag}(\lambda_i(A^*A)),$$

lo que hace que se tenga que

$$A^*A = QDQ^*.$$

Por tanto, como $\|A\mathbf{v}\|_2 = \sqrt{(A\mathbf{v})^*(A\mathbf{v})}$, para todo $\mathbf{v} \in V$, se sigue que

$$\|A\mathbf{v}\|_2^2 = (A\mathbf{v})^*A\mathbf{v} = \mathbf{v}^*A^*A\mathbf{v} = \mathbf{v}^*QDQ^*\mathbf{v} = (Q^*\mathbf{v})^*D(\mathbf{Q}^*\mathbf{v}) = \sum_{i=1}^{n} \lambda_i(A^*A)|w_i|^2,$$

siendo $Q^*\mathbf{v} = (w_1, \ldots, w_n)^{\mathrm{t}}$. Consecuentemente,

$$\begin{aligned}\|A\mathbf{v}\|_2^2 \leq \varrho(A^*A) \sum_{i=1}^{n} |w_i|^2 = \varrho(A^*A)\,((Q^*\mathbf{v})^*Q^*\mathbf{v}) = \varrho(A^*A)\,(\mathbf{v}^*QQ^*\mathbf{v})\\ = \varrho(A^*A)\,(\mathbf{v}^*\mathbf{v}) = \varrho(A^*A)\,\|\mathbf{v}\|_2^2.\end{aligned}$$

Por otra parte, como los autovalores de A^*A son números reales no negativos (véanse la proposiciones V.5.17 y V.5.18), se cumple que

$$\lambda := \max\nolimits_{1\leq j\leq n}\lambda_j(A^*A) = \varrho(A^*A).$$

Por consiguiente, si $\mathbf{v} \in V \setminus \{0\}$ es un autovector de A^*A asociado a λ (es decir, $A^*A\mathbf{v} = \lambda\mathbf{v}$), entonces

$$\|A\mathbf{v}\|_2^2 = (A\mathbf{v})^*A\mathbf{v} = \mathbf{v}^*A^*A\mathbf{v} = \lambda\,\mathbf{v}^*\mathbf{v} = \lambda\,\|\mathbf{v}\|_2^2 = \varrho(A^*A)\|\mathbf{v}\|_2^2;$$

de donde se sigue que

$$\|\!|A|\!\|_2 = \sqrt{\varrho(A^*A)},$$

como queríamos probar.

(c) Para todo $\mathbf{v} \in V$ se verifica que

$$\|A\mathbf{v}\|_\infty = \max_{1\le i\le n} |(A\mathbf{v})_i| = \max_{1\le i\le n} \left|\sum_{j=1}^n a_{ij}v_j\right| \le \max_{1\le i\le n} \left(\sum_{j=1}^n |a_{ij}|\,|v_j|\right)$$
$$\le \left(\max_{1\le i\le n} \sum_{j=1}^n |a_{ij}|\right) \|\mathbf{v}\|_\infty.$$

Consideremos ahora el vector $\mathbf{u} \in V$ de componentes

$$u_j = \begin{cases} \frac{\overline{a_{i_0j}}}{|a_{i_0j}|} & \text{si} \quad a_{i_0j} \neq 0; \\ 1 & \text{si} \quad a_{i_0j} = 0, \end{cases}$$

siendo i_0 un subíndice tal que

$$\max_{1\le i\le n} \sum_{j=1}^n |a_{ij}| = \sum_{j=1}^n |a_{i_0j}|.$$

Como $|u_j| = 1,\ j = 1, 2, \ldots, n$, entonces $\|\mathbf{u}\|_\infty = 1$ y

$$|(A\mathbf{u})_i| = \left|\sum_{j=1}^n a_{ij}u_j\right| \le \sum_{j=1}^n |a_{ij}|\,|u_j| = \sum_{j=1}^n |a_{ij}| \le \sum_{j=1}^n |a_{i_0j}|$$

para todo $i = 1, 2, \ldots, n$, lo que hace que se tenga que

$$\max_{1\le i\le n} |(A\mathbf{u})_i| \le \sum_{j=1}^n |a_{i_0j}|.$$

Por otra parte, como

$$|(A\mathbf{u})_{i_0}| = \left|\sum_{j=1}^n a_{i_0j}u_j\right| = \left|\sum_{\substack{j=1 \\ a_{i_0j}\neq 0}}^n a_{i_0j}u_j\right| = \left|\sum_{\substack{j=1 \\ a_{i_0j}\neq 0}}^n a_{i_0j}\frac{\overline{a_{i_0j}}}{|a_{i_0j}|}\right| = \sum_{\substack{j=1 \\ a_{i_0j}\neq 0}}^n \frac{|a_{i_0j}|^2}{|a_{i_0j}|}$$
$$= \sum_{\substack{j=1 \\ a_{i_0j}\neq 0}}^n |a_{i_0j}| = \sum_{j=1}^n |a_{i_0j}|,$$

entonces

$$\|A\mathbf{u}\|_\infty = \max_{1\le i\le n} |(A\mathbf{u})_i| = |(A\mathbf{u})_{i_0}| = \sum_{j=1}^n |a_{i_0j}| = \max_{1\le i\le n} \sum_{j=1}^n |a_{ij}|$$
$$= \left(\max_{1\le i\le n} \sum_{j=1}^n |a_{ij}|\right) \|\mathbf{u}\|_\infty.$$

Así, se concluye que

$$\|A\|_\infty = \max_{1\leq i\leq n} \sum_{j=1}^{n} |a_{ij}|.$$

■

Nota VIII.2.8. De los apartados (a) y (c) del teorema VIII.2.7 se deduce que

$$\|A^*\|_1 = \|A\|_\infty.$$

Como se ha visto en el teorema VIII.2.7 las normas $\|\cdot\|_1$ y $\|\cdot\|_\infty$ son fácilmente calculables a partir de los elementos de la matriz, a diferencia de lo que ocurre con la norma $\|\cdot\|_2$. No obstante, esta norma tiene buenas propiedades desde el punto de vista teórico; veamos algunas:

Proposición VIII.2.9. *Sea $A \in \mathcal{M}_n(\mathbb{k})$.*

(a) *La norma $\|\cdot\|_2$ es invariante por transformaciones unitarias, es decir, dada $Q \in \mathcal{M}_n(\mathbb{k})$ tal que $Q\,Q^* = I_n$ se cumple que*

$$\|A\|_2 = \|AQ\|_2 = \|QA\|_2 = \|Q^*AQ\|_2$$

(b) *Si A es normal, entonces*

$$\|A\|_2 = \varrho(A).$$

Demostración. (a) Según se ha visto en el apartado (b) del teorema VIII.2.7,

$$\|A\|_2^2 = \varrho(A^*A) = \varrho(A^*Q\,Q^*A) = \varrho((Q^*A)^*(Q^*A)) = \|Q^*A\|_2^2,$$
$$\|A\|_2^2 = \varrho(AA^*) = \varrho(AQQ^*A^*) = \varrho((AQ)(AQ)^*) = \|AQ\|_2^2,$$

luego,

$$\|Q^*AQ\|_2 = \|AQ\|_2 = \|A\|_2.$$

(b) Si A es normal, por el teorema V.5.15, existe una matriz Q unitaria tal que

$$Q^*AQ = D = \operatorname{diag}(\lambda_i(A)).$$

Por tanto, el apartado anterior nos asegura que

$$\|A\|_2^2 = \|Q^*AQ\|_2^2 = \|D\|_2^2 = \varrho(D^*D).$$

Por otra parte, si $\operatorname{sp}(A) = \{\lambda_1, \ldots, \lambda_n\}$, entonces $D^* = \operatorname{diag}(\overline{\lambda_i})$ y $D^*D = \operatorname{diag}(|\lambda_i|^2)$; luego,

$$\operatorname{sp}(D^*D) = \left\{|\lambda_1|^2, |\lambda_2|^2, \ldots, |\lambda_n|^2\right\}$$

De esta forma, se concluye que

$$\|A\|_2^2 = \varrho(D^*D) = \max_{1\leq i\leq n} |\lambda_i|^2 = \left(\max_{1\leq i\leq n} |\lambda_i|\right)^2 = \varrho(A)^2.$$

■

Nota VIII.2.10. Sea $A \in \mathcal{M}_n(\Bbbk)$.

(a) Si A es hermítica, entonces $\|A\|_2 = \varrho(A)$.
(b) Si A es unitaria, entonces $\|A\|_2 = \sqrt{\varrho(A^*A)} = \sqrt{\varrho(I_n)} = 1$.

Como ya hemos dicho, existen normas matriciales que no están subordinadas a ninguna norma vectorial. Vamos a construir una de ellas (que, por otra parte, no es otra que la norma usual de $\mathcal{M}_n(\Bbbk)$ considerado como espacio vectorial de dimensión n^2 sobre $\Bbbk$) que servirá como complemento práctico a la norma $\|\cdot\|_2$.

Lema VIII.2.11. *Sea* $A = (a_{ij}) \in \mathcal{M}_n(\Bbbk)$. *Entonces* $\operatorname{tr}(A^*A) = \sum_{i,j=1}^n |a_{ij}|^2$.

Demostración. Como $A = (a_{ij})$, entonces $A^* = (\overline{a_{ji}})$, por lo que $A^*A = (\alpha_{ij})$ siendo $\alpha_{ij} = \sum_{k=1}^n \overline{a_{ki}}\, a_{kj}$ para $i, j = 1, 2, \ldots, n$. En particular, los elementos diagonales son de la forma

$$\alpha_{ii} = \sum_{k=1}^n \overline{a_{ki}}\, a_{ki} = \sum_{k=1}^n |a_{ki}|^2$$

para $i = 1, 2, \ldots, n$; consecuente

$$\operatorname{tr}(A^*A) = \sum_{i=1}^n \alpha_{ii} = \sum_{i,k=1}^n |a_{ki}|^2.$$

■

Proposición VIII.2.12. *La aplicación* $\|\cdot\|_F : \mathcal{M}_n(\Bbbk) \to \mathbb{R}$ *dada por*

$$\|A\|_F := \sqrt{\sum_{i,j=1}^n |a_{ij}|^2} = \sqrt{\operatorname{tr}(A^*A)} = \sqrt{\operatorname{tr}(A\,A^*)}$$

es una norma matricial.

Demostración. La aplicación $\|\cdot\|_F$ es la norma usual de $\mathcal{M}_n(\Bbbk)$ considerado como espacio vectorial de dimensión n^2 sobre $\Bbbk$, por lo que:

(a) $\|A\|_F = 0$ si, y sólo si, $A = \mathbf{0}$.
(b) $\|\lambda A\|_F = |\lambda|\, \|A\|_F$, para todo $A \in \mathcal{M}_n(\Bbbk)$ y $\lambda \in \Bbbk$.
(c) $\|A + B\|_F \leq \|A\|_F + \|B\|_F$, para todo A y $B \in \mathcal{M}_n(\Bbbk)$.

Para la cuarta propiedad aplicamos la desigualdad de Cauchy-Schwarz[1] a los vectores

$$\mathbf{a}_i = (a_{i1}, a_{i2}, \ldots, a_{in})^{\mathrm{t}} \quad \text{y} \quad \mathbf{b}_j = (b_{1j}, b_{2j}, \ldots, b_{nj})^{\mathrm{t}},$$

[1]Desigualdad de Cauchy-Schwarz: para todo $\mathbf{u}$ y $\mathbf{v} \in \Bbbk^n$ se cumple que $|\mathbf{u}^*\mathbf{v}| \leq \|\mathbf{u}\|\,\|\mathbf{v}\|$ para todo $\mathbf{u}$, y se da la igualdad cuando $\mathbf{u} = \alpha\mathbf{v}$, para $\alpha = \mathbf{v}^*\mathbf{u}/\mathbf{v}^*\mathbf{v}$.

obteniendo

$$\begin{aligned}\|AB\|_F^2 &= \sum_{i,j=1}^{n}\left|\sum_{k=1}^{n} a_{ik}b_{kj}\right|^2 \leq \sum_{i,j=1}^{n}\left(\sum_{k=1}^{n}|a_{ik}|^2\right)\left(\sum_{l=1}^{b}|b_{lj}|^2\right)\\ &= \left(\sum_{i,k=1}^{n}|a_{ik}|^2\right)\left(\sum_{j,l=1}^{n}|b_{lj}|^2\right) = \|A\|_F^2\|B\|_F^2.\end{aligned}$$

■

Definición VIII.2.13. La norma $\|\cdot\|_F$ dada en la proposición VIII.2.12 se denomina **norma de Fröbenius**.

Entre las principales propiedades de la norma de Fröbenius destacamos:

Proposición VIII.2.14. *La norma de Fröbenius $\|\cdot\|_F$ es una norma matricial no subordinada, si $n \geq 2$, invariante por transformaciones unitarias. Además,*

$$\|A\|_2 \leq \|A\|_F \leq \sqrt{n}\,\|A\|_2,$$

para todo $A \in \mathcal{M}_n(\Bbbk)$.

Demostración. Como

$$\|I_n\|_F = \sqrt{n} \neq 1 \text{ si } n \geq 2,$$

por la proposición VIII.2.5.(d) se obtiene que $\|\cdot\|_F$ no está subordinada si $n \geq 2$. Por otra parte, si Q es una matriz unitaria, se verifica que

$$\|A\|_F^2 = \operatorname{tr}(A^*A) = \operatorname{tr}(A^*Q\,Q^*A) = \operatorname{tr}((Q^*A)^*(Q^*A)) = \|Q^*A\|_F^2$$

$$\|A\|_F^2 = \operatorname{tr}(A\,A^*) = \operatorname{tr}(AQ\,Q^*A^*) = \operatorname{tr}(AQ(AQ)^*) = \|AQ\|_F^2$$

y

$$\|Q^*AQ\|_F^2 = \|AQ\|_F^2 = \|A\|_F^2.$$

Finalmente, como los autovalores de A^*A son números reales no negativos (véanse la proposiciones V.5.17 y V.5.18), entonces

$$\varrho(A^*A) \leq \sum_{i=1}^{n}\lambda_i \leq n\,\varrho(A^*A),$$

donde $\operatorname{sp}(A^*A) = \{\lambda_1, \ldots, \lambda_n\}$. Así, por el teorema VIII.2.7, se tiene que

$$\|A\|_2^2 = \varrho(A^*A) \leq \sum_{i=1}^{n}\lambda_i = \operatorname{tr}(A^*A) = \|A\|_F^2 \leq n\,\varrho(A^*A) = n\,\|A\|_2^2.$$

■

Nota VIII.2.15. Ya se ha comentado que el teorema VIII.2.7 proporciona la manera de calcular la norma $\|\!|\cdot|\!|_1$ y la norma $\|\!|\cdot|\!|_\infty$ de una matriz $A \in \mathcal{M}_n(\Bbbk)$ a partir de los elementos que la componen y que no ocurre así con la norma $\|\!|\cdot|\!|_2$. El interés por la norma de Fröbenius es que también se calcula directamente a partir de los elementos de la matriz y, según la última parte de la proposición VIII.2.14, puede usarse para obtener cotas de la norma $\|\!|\cdot|\!|_2$.

Sabemos que las matrices normales verifican que su norma $\|\!|\cdot|\!|_2$ coincide con su radio espectral. En el caso general (es decir, en el caso de una matriz y norma matricial cualquiera, subordinada o no, con coeficientes complejos) el resultado se convierte en desigualdad: el radio espectral es siempre menor o igual que la norma de la matriz.

Teorema VIII.2.16. *Sea $A \in \mathcal{M}_n(\Bbbk)$.*

(a) *Para toda norma matricial (subordinada o no) se verifica que*

$$\varrho(A) \leq |\!|A|\!|.$$

(b) *Para todo $\varepsilon > 0$ existe una norma matricial $\|\!|\cdot|\!|_{A,\varepsilon}$ (que se puede tomar subordinada) tal que*

$$|\!|A|\!|_{A,\varepsilon} \leq \varrho(A) + \varepsilon.$$

Demostración. (a) Sean $\mathbf{v} \in V = \mathbb{C}^n$ un autovector asociado al autovalor λ de $A \in \mathcal{M}_n(\Bbbk) \hookrightarrow \mathcal{M}_n(\mathbb{C})$ de módulo máximo, es decir, $A\mathbf{v} = \lambda\mathbf{v}$ con $|\lambda| = \varrho(A)$ y $\mathbf{w} \in V$ tal que la matriz $\mathbf{v}\mathbf{w}^{\mathrm{t}} \in \mathcal{M}_n(\mathbb{C})$ es no nula. Entonces

$$\varrho(A)\,|\!|\mathbf{v}\mathbf{w}^{\mathrm{t}}|\!| = |\lambda|\,|\!|\mathbf{v}\mathbf{w}^{\mathrm{t}}|\!| = |\!|\lambda\,\mathbf{v}\mathbf{w}^{\mathrm{t}}|\!| = |\!|A\mathbf{v}\mathbf{w}^{\mathrm{t}}|\!| \leq |\!|A|\!|\,|\!|\mathbf{v}\mathbf{w}^{\mathrm{t}}|\!|,$$

de donde se sigue el resultado buscado al ser $|\!|\mathbf{v}\mathbf{w}^{\mathrm{t}}|\!| > 0$.

(b) Considerando de nuevo la inmersión natural $A \in \mathcal{M}_n(\Bbbk) \hookrightarrow \mathcal{M}_n(\mathbb{C})$, por el teorema V.5.15(a), existen una matriz triangular superior $T = (t_{ij}) \in \mathcal{M}_n(\mathbb{C})$ y una matriz unitaria $Q \in \mathcal{M}_n(\mathbb{C})$ tales que $Q^*AQ = T$. Sabemos que los elementos de la diagonal de T son los autovalores de A que denotaremos $\lambda_1, \ldots, \lambda_n$.

Si para cada $\delta > 0$ consideramos la matriz diagonal

$$D_\delta = \mathrm{diag}(1, \delta, \delta^2, \ldots, \delta^{n-1}),$$

entonces el elemento (i,j)-ésimo de la matriz

$$D_\delta^{-1}Q^{-1}AQD_\delta = (QD_\delta)^{-1}AQD_\delta$$

es $\delta^{j-i}\,t_{ij}$ si $i < j$, λ_i si $j = i$ y cero en otro caso.

Dado $\varepsilon > 0$ tomamos $\delta > 0$ suficientemente pequeño para que

$$\sum_{j=i+1}^{n} \delta^{j-i}|t_{ij}| < \varepsilon$$

para $i = 1, \ldots, n-1$, y consideramos la aplicación $\|\!| \cdot |\!|\!|_{A,\varepsilon} : \mathcal{M}_n(\mathbb{C}) \to \mathbb{R}$ dada por

$$|\!|\!| B |\!|\!|_{A,\varepsilon} = |\!|\!| (QD_\delta)^{-1} B (QD_\delta) |\!|\!|_\infty.$$

Nótese que $\|\!| \cdot |\!|\!|_{A,\varepsilon}$ depende de la matriz A y de ε. Claramente, $\|\!| \cdot |\!|\!|_{A,\varepsilon}$ es una norma matricial subordinada a la norma vectorial

$$\mathbf{v} \mapsto \|(QD_\delta)^{-1}\mathbf{v}\|_\infty.$$

Además,

$$\begin{aligned} |\!|\!| A |\!|\!|_{A,\varepsilon} &= |\!|\!| (QD_\delta)^{-1} A (QD_\delta) |\!|\!|_\infty = \max_{1\le i\le n} \left(\sum_{j=i+1}^{n} \delta^{j-i} |t_{ij}| + |\lambda_i| \right) \\ &= \max_{1\le i\le n} \sum_{j=i+1}^{n} \delta^{j-i} |t_{ij}| + \max_{1\le i\le n} |\lambda_i| < \varepsilon + \varrho(A). \end{aligned}$$

■

Convergencia de las iteraciones de una matriz.

La noción de convergencia de una sucesión de vectores (véase la definición VI-II.1.8) incluye el caso de las matrices, basta considerar $\mathcal{M}_n(\Bbbk)$ como espacio vectorial de dimensión n^2. Concretamente,

Definición VIII.2.17. Sea $\|\!| \cdot |\!|\!|$ una norma matricial sobre $\mathcal{M}_n(\Bbbk)$. Diremos que una sucesión de matrices $(A_m)_{m\in\mathbb{N}}$ de $\mathcal{M}_n(\Bbbk)$ **converge** a un matriz $A \in \mathcal{M}_n(\Bbbk)$, y lo denotaremos $A = \lim_{m\to\infty} A_m$, si

$$\lim_{m\to\infty} |\!|\!| A_m - A |\!|\!| = 0.$$

Ejemplo VIII.2.18. La sucesión de matrices

$$A_m = \begin{pmatrix} 1 + \frac{m}{m^2+3} & \frac{4}{m} \\ \frac{1}{m} + \frac{2}{m^2} & 1 - \mathrm{e}^{-\frac{3}{m^4}} \end{pmatrix} \in \mathcal{M}_2(\mathbb{R})$$

converge a la matriz

$$A = \lim_{m\to\infty} A_m = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

El siguiente resultado caracteriza la convergencia a cero de las potencias sucesivas A^m de una matriz cuadrada A.

Teorema VIII.2.19. *Sea $A \in \mathcal{M}_n(\Bbbk)$. Son equivalentes:*

(a) $\lim_{m\to\infty} A^m = \mathbf{0}$.
(b) $\lim_{m\to\infty} A^m\mathbf{v} = \mathbf{0}$, *para todo* $\mathbf{v} \in V = \Bbbk^n$.
(c) $\varrho(A) < 1$.

(d) *Existe una norma matricial* $|||\cdot|||$ *(que se puede tomar subordinada) tal que* $|||A||| < 1$.

Demostración. (a) $\Rightarrow$ (b) Sea $|||\cdot|||$ la norma matricial subordinada a una norma vectorial $\|\cdot\|$. Por definición,

$$\lim_{m\to\infty} A^m = \mathbf{0} \iff \lim_{m\to\infty} |||A^m||| = 0.$$

Por tanto, como para todo $\mathbf{v} \in V$ se verifica que $\|A^m\mathbf{v}\| \leq |||A^m|||\,\|\mathbf{v}\|$, para todo $m \in \mathbb{N}$, entonces $\lim_{m\to\infty} \|A^m\mathbf{v}\| = 0$ y, así, $\lim_{m\to\infty} A^m\mathbf{v} = 0$.

(b) $\Rightarrow$ (c) Procedemos por reducción al absurdo. Si $\varrho(A) \geq 1$, entonces existe un autovalor (complejo) $\lambda = \lambda(A) \in \mathrm{sp}(A)$ con $|\lambda| \geq 1$; basta considerar un autovalor $\mathbf{v} \in \mathbb{C}^n \setminus \{\mathbf{0}\}$ asociado a λ para llegar a contradicción. En efecto, como $A\mathbf{v} = \lambda\mathbf{v}$ entonces

$$A^m\mathbf{v} = \lambda^m\mathbf{v}$$

para todo $m \in \mathbb{N}$ y, por tanto,

$$\lim_{m\to\infty} \|A^m\mathbf{v}\| = \lim_{m\to\infty} |\lambda|^m\|\mathbf{v}\| \neq 0.$$

(c) $\Rightarrow$ (d) Por el teorema VIII.2.16, dado $\varepsilon > 0$ existe una norma matricial $|||\cdot|||_{A,\varepsilon}$ tal que $|||A|||_{A,\varepsilon} \leq \varrho(A) + \varepsilon$. Tomando

$$0 < \varepsilon < 1 - \varrho(A)$$

se obtiene que

$$|||A|||_{A,\varepsilon} \leq \varrho(A) + (1 - \varrho(A)) = 1.$$

(d) $\Rightarrow$ (a) Claramente,

$$|||A^m||| = |||A^{m-1}A||| \leq |||A^{m-1}|||\,|||A||| \leq \ldots \leq |||A|||^m.$$

Por tanto, la hipótesis $|||A||| < 1$ implica

$$\lim_{m\to\infty} |||A^m||| = 0,$$

es decir, $\lim_{m\to\infty} A^m = 0$. ∎

En la práctica, el resultado anterior se utiliza del siguiente modo: si se quiere demostrar que las potencias sucesivas de una matriz A convergen a cero, bastará probar que todos los autovalores (complejos) de A tienen módulo menor que uno, o bien encontrar una norma matricial para la que $|||A||| < 1$. Volveremos a estas cuestiones en el siguiente tema.

El siguiente resultado muestra que la norma de las sucesivas potencias de una matriz se comporta asintóticamente como las sucesivas potencias de su radio espectral:

Teorema VIII.2.20. *Si $A \in \mathcal{M}_n(\mathbb{R})$ y $\|\!|\cdot|\!\|$ es una norma matricial (subordinada o no) entonces*

$$\lim_{m\to+\infty} \|\!|A^m|\!\|^{1/m} = \varrho(A).$$

Demostración. Como $\varrho(A)^m = \varrho(A^m)$, para todo $m \in \mathbb{N}$, el teorema VIII.2.16(a) asegura que $\varrho(A)^m = \varrho(A^m) \leq \|\!|A^m|\!\|$, para todo $m \in \mathbb{N}$ y, por consiguiente, que

$$\varrho(A) \leq \|\!|A^m|\!\|^{1/m},$$

para todo $m \in \mathbb{N}$. Para demostrar que, tomando límite, se da la igualdad, basta probar que para cada $\varepsilon > 0$ existe $m_0 \in \mathbb{N}$ tal que

$$\|\!|A^m|\!\|^{1/m} < \varrho(A) + \varepsilon,$$

para todo $m \geq m_0$. Para ello, dado $\varepsilon > 0$ definimos la matriz

$$A_\varepsilon = \frac{A}{\varrho(A) + \varepsilon}.$$

Como $\rho(A_\varepsilon) < 1$, aplicando el teorema VIII.2.19 obtenemos que $\lim_{m\to+\infty} A_\varepsilon^m = 0$, es decir,

$$0 = \lim_{m\to+\infty} \|\!|A_\varepsilon^m|\!\| = \lim_{m\to+\infty} \left\|\!\left| \frac{A^m}{(\varrho(A)+\varepsilon)^m} \right|\!\right\| = \lim_{m\to+\infty} \frac{\|\!|A^m|\!\|}{(\varrho(A)+\varepsilon)^m}.$$

De donde se sigue que existe $m_0 \in \mathbb{N}$ tal que $\|\!|A^m|\!\| < (\varrho(A)+\varepsilon)^m$, para todo $m \geq m_0$. Tomando ahora raíces m-ésimas se obtiene la desigualdad buscada. ∎

3. Número de condición de una matriz

Diremos que un problema está *mal condicionado* cuando pequeños cambios en los datos dan lugar a grandes cambios en las respuestas. Las técnicas que se emplean en el condicionamiento de un problema están fuertemente ligadas a la estructura del mismo. En general, a la hora de resolver un problema $y = P(x)$ se intenta definir un *número de condición*[2] $\kappa = \kappa(x) \geq 0$ de forma que

$$\left\| \frac{P(\bar{x}) - P(x)}{P(x)} \right\| \simeq \kappa(x) \left\| \frac{\bar{x} - x}{x} \right\|$$

Este número indicará, según sea cercano a 1 o esté alejado de éste, si el problema está bien o mal condicionado, respectivamente. Si el número de condición es menor que 1 o está próximo a 1, el error del dato no se amplificará mucho y el error del resultado será, a lo sumo, del mismo orden que el error en el dato; por el contrario, si este número de condición toma valores muy grandes, el error final será una gran amplificación del dato.

[2]Aquí la doble barra no significa necesariamente una norma, sino una "medida" de las magnitudes en cuestión.

Para casos concretos, podemos definir fácilmente el número de condición. Como por ejemplo ocurre con la resolución de sistemas lineales $A\mathbf{x} = \mathbf{b}$ con $A \in \mathcal{M}_n(\Bbbk)$, como veremos en breve.

Ejemplo VIII.3.1. **(R.S. Wilson)**
Consideremos el sistema lineal $A\mathbf{x} = \mathbf{b}$ donde $\mathbf{b}$ es el vector $\mathbf{b} = (32, 23, 33, 31)^{\mathrm{t}}$ y A es la matriz simétrica

$$A = \begin{pmatrix} 10 & 7 & 8 & 7 \\ 7 & 5 & 6 & 5 \\ 8 & 6 & 10 & 9 \\ 7 & 5 & 9 & 10 \end{pmatrix}$$

que tiene por matriz inversa a

$$A^{-1} = \begin{pmatrix} 25 & -41 & 10 & -6 \\ -41 & 68 & -17 & 10 \\ 10 & -17 & 5 & -3 \\ -6 & 10 & -3 & 2 \end{pmatrix}$$

y cuyo determinante es 1. La solución exacta de dicho sistema es $\mathbf{u} = (1, 1, 1, 1)^{\mathrm{t}}$. Si consideramos las perturbaciones de los datos A y $\mathbf{b}$

$$A + \Delta A = \begin{pmatrix} 10 & 7 & 8{,}1 & 7{,}2 \\ 7{,}08 & 5{,}04 & 6 & 5 \\ 8 & 5{,}98 & 8{,}89 & 9 \\ 6{,}99 & 4{,}99 & 9 & 9{,}98 \end{pmatrix} \quad \text{y} \quad \begin{pmatrix} 32{,}1 \\ 22{,}9 \\ 33{,}1 \\ 30{,}9 \end{pmatrix}$$

las soluciones exactas de los sistemas lineales $(A + \Delta A)\mathbf{x} = \mathbf{b}$ y $A\mathbf{x} = \mathbf{b} + \delta\mathbf{b}$ vienen dadas, respectivamente, por

$$\mathbf{u} + \Delta\mathbf{u} = \begin{pmatrix} -81 \\ 137 \\ -34 \\ 22 \end{pmatrix} \quad \text{y} \quad \mathbf{u} + \delta\mathbf{u} = \begin{pmatrix} 9{,}2 \\ -12{,}6 \\ 4{,}5 \\ -1{,}1 \end{pmatrix}.$$

Como se aprecia pequeños cambios en el dato A han producido un resultado muy alejado de la solución original $\mathbf{u}$. Análogamente, cuando se perturba ligeramente el dato $\mathbf{b}$ se obtiene un resultado $\mathbf{u} + \delta\mathbf{u}$ muy distante de $\mathbf{u}$.

En esta sección daremos la justificación de estas propiedades sorprendentes, así como la forma precisa de medir el tamaño de las perturbaciones y de los errores, mediante la introducción del *número de condición de una matriz.*

Sean $A \in \mathcal{M}_n(\Bbbk)$ una matriz invertible y $\mathbf{b} \in \Bbbk^n$ no nulo. Veamos cómo definir el condicionamiento de un sistema lineal

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

En el supuesto de que se tome como segundo miembro, en lugar del vector $\mathbf{b}$, una perturbación de éste $\mathbf{b} + \delta\mathbf{b}$, si denotamos $\mathbf{u}$ a la solución del sistema $A\mathbf{x} = \mathbf{b}$ y $\mathbf{u} + \delta\mathbf{u}$ a la solución del sistema perturbado, se verifica que

$$A(\mathbf{u} + \delta\mathbf{u}) = \mathbf{b} + \delta\mathbf{b} \Rightarrow A\,\delta\mathbf{u} = \delta\mathbf{b} \Rightarrow \delta\mathbf{u} = A^{-1}\,\delta\mathbf{b},$$

luego a partir de la norma matricial $|||\cdot|||$ subordinada a una norma vectorial $\|\cdot\|$, se tiene que

$$\|\delta\mathbf{u}\| \leq |||A^{-1}|||\,\|\delta\,\mathbf{b}\|;$$

como, por otra parte,

$$A\mathbf{u} = \mathbf{b} \Rightarrow \|\mathbf{b}\| \leq |||A|||\,\|\mathbf{u}\| \Rightarrow \frac{1}{\|\mathbf{u}\|} \leq \frac{|||A|||}{\|\mathbf{b}\|},$$

se tiene que

$$\frac{\|\delta\mathbf{u}\|}{\|\mathbf{u}\|} \leq |||A|||\,|||A^{-1}|||\,\frac{\|\delta\mathbf{b}\|}{\|\mathbf{b}\|}.$$

Parece claro, pues, que la cantidad $|||A|||\,|||A^{-1}|||$ servirá como número de condición para resolver un sistema lineal $A\mathbf{x} = \mathbf{b}$. De hecho, se tiene la siguiente definición:

Definición VIII.3.2. Sea $|||\cdot|||$ una norma matricial y $A \in \mathcal{M}_n(\Bbbk)$ una matriz invertible. El número

$$\mathrm{cond}(A) = |||A|||\,|||A^{-1}|||$$

se denomina **número de condición** (o **condicionamiento**) de la matriz A respecto de la norma $|||\cdot|||$.

En general, cuando escribamos $\mathrm{cond}(A)$ nos estaremos refiriendo al condicionamiento de una matriz respecto de una norma matricial $|||\cdot|||$. En el caso particular en que tomemos la norma $|||\cdot|||_p,\ 1 \leq p \leq \infty$, escribiremos

$$\mathrm{cond}_p(A) = |||A|||_p\,|||A^{-1}|||_p,\ 1 \leq p \leq \infty.$$

Teorema VIII.3.3. *Sean $|||\cdot|||$ la norma matricial subordinada a una norma vectorial $\|\cdot\|$ y $A \in \mathcal{M}_n(\Bbbk)$ una matriz invertible. Si $\mathbf{u}$ y $\mathbf{u} + \delta\mathbf{u}$ son las soluciones respectivas de los sistema*

$$A\mathbf{x} = \mathbf{b} \quad y \quad \mathbf{A}\mathbf{x} = \mathbf{b} + \delta\mathbf{b},$$

con $\mathbf{b} \neq 0$ y $\delta\mathbf{b} \in \Bbbk^n$, entonces se verifica que

$$\frac{\|\delta\mathbf{u}\|}{\|\mathbf{u}\|} \leq \mathrm{cond}(A)\,\frac{\|\delta\mathbf{b}\|}{\|\mathbf{b}\|}.$$

Además, $\mathrm{cond}(A)$ *es el número más pequeño que verifica la desigualdad anterior en el siguiente sentido: para cada matriz* A *invertible existen* $\mathbf{b}$ *y* $\delta\mathbf{b} \in \Bbbk^n \setminus \{\mathbf{0}\}$ *tales que*

$$\frac{\|\delta\mathbf{u}\|}{\|\mathbf{u}\|} = \mathrm{cond}(A)\,\frac{\|\delta\mathbf{b}\|}{\|\mathbf{b}\|},$$

donde $\mathbf{u}$ *y* $\mathbf{u}+\delta\mathbf{u}$ *son las soluciones de los sistemas* $A\mathbf{x}=\mathbf{b}$ *y* $A\mathbf{x}=\mathbf{b}+\delta\mathbf{b}$, *respectivamente.*

Demostración. La desigualdad propuesta en el enunciado ya se ha demostrado previamente. Veamos la optimalidad. Por la proposición VIII.2.5 existe $\mathbf{u} \in \Bbbk^n$ tal que

$$\|A\mathbf{u}\| = |\!|\!|A|\!|\!|\,\|\mathbf{u}\|.$$

A partir de este vector $\mathbf{u}$, definimos

$$\mathbf{b} = A\mathbf{u}.$$

Por otro lado, aplicando nuevamente la proposición VIII.2.5, existe $\delta\mathbf{b} \in \Bbbk^n$ tal que

$$\|A^{-1}\delta\mathbf{b}\| = |\!|\!|A^{-1}|\!|\!|\,\|\delta\mathbf{b}\|.$$

Así pues, considerando los sistemas lineales

$$A\mathbf{x}=\mathbf{b} \quad \text{y} \quad A\mathbf{x}=\mathbf{b}+\delta\mathbf{b},$$

tendremos, como antes, que

$$A\delta\mathbf{u} = \delta\mathbf{b}$$

y así

$$\delta\mathbf{u} = A^{-1}\delta\mathbf{b},$$

con lo que

$$\|\delta\mathbf{u}\| = \|A^{-1}\delta\mathbf{b}\| = |\!|\!|A^{-1}|\!|\!|\,\|\delta\mathbf{b}\| \quad \text{y} \quad \|\mathbf{b}\| = \|A\mathbf{u}\| = |\!|\!|A|\!|\!|\,\|\mathbf{u}\|.$$

Por tanto,

$$\frac{\|\delta\mathbf{u}\|}{\|\mathbf{u}\|} = |\!|\!|A|\!|\!|\,|\!|\!|A^{-1}|\!|\!|\,\frac{\|\delta\mathbf{b}\|}{\|\mathbf{b}\|} = \mathrm{cond}(A)\,\frac{\|\delta\mathbf{b}\|}{\|\mathbf{b}\|}.$$

■

Por tanto, según el resultado anterior, el número de condición es un medida de la sensibilidad del sistema a las perturbaciones en el término independiente. Cuando se consideran perturbaciones de la matriz A en lugar de perturbaciones del vector $\mathbf{b}$, el resultado que se obtiene no es tan nítido, pero el número $\mathrm{cond}(A)$ sigue siendo una buena herramienta para medir el condicionamiento del problema. En concreto, se tiene el siguiente resultado:

Teorema VIII.3.4. *Sean* $|||\cdot|||$ *la norma matricial subordinada a una norma vectorial* $\|\cdot\|$ *y* $A \in \mathcal{M}_n(\Bbbk)$ *una matriz invertible. Si* $\mathbf{u}$ *y* $\mathbf{u}+\Delta\mathbf{u}$ *son las soluciones respectivas de los sistemas lineales*

$$A\mathbf{x} = \mathbf{b} \quad y \quad (A+\Delta A)\mathbf{x} = \mathbf{b},$$

con $\mathbf{b} \neq \mathbf{0}$, *se verifica que*

$$\frac{\|\Delta\mathbf{u}\|}{\|\mathbf{u}+\Delta\mathbf{u}\|} \leq \operatorname{cond}(A)\frac{|||\Delta A|||}{|||A|||};$$

es más,

$$\frac{\|\Delta\mathbf{u}\|}{\|\mathbf{u}\|} \leq \operatorname{cond}(A)\frac{|||\Delta A|||}{|||A|||}\left(1+O(|||A|||)\right).$$

Además, $\operatorname{cond}(A)$ *el número más pequeño que verifica la desigualdad anterior en el siguiente sentido: para toda matriz* A *invertible existen* $\mathbf{b} \in \Bbbk^n \setminus \{\mathbf{0}\}$ *y* $\Delta A \in \mathcal{M}_n(\Bbbk)$ *tales que*

$$\frac{\|\Delta\mathbf{u}\|}{\|\mathbf{u}+\Delta\mathbf{u}\|} = \operatorname{cond}(A)\frac{|||\Delta A|||}{|||A|||},$$

donde $\mathbf{u}$ *y* $\Delta\mathbf{u}$ *son las soluciones de los sistemas* $A\mathbf{x} = \mathbf{b}$ *y* $(A+\Delta A)\mathbf{x} = \mathbf{b}$, *respectivamente.*

Demostración. La demostración de este resultado puede consultarse en [**Cia82**]. ∎

Otros resultados similares a los anteriores sobre el número de condición como medida de sensibilidad de un sistema de ecuaciones lineales a cambios en los datos se pueden encontrar en el apartado 3.1.2 de [**QSS07**].

A continuación recogemos algunas propiedades de demostración inmediata que verifica el número de condición de una matriz.

Proposición VIII.3.5. *Sea* $|||\cdot|||$ *una norma matricial (subordinada o no) y* $A \in \mathcal{M}_n(\Bbbk)$ *una matriz invertible. Se verifican las siguientes propiedades:*

(a) $\operatorname{cond}(A) \geq 1$.
(b) $\operatorname{cond}(A) = \operatorname{cond}(A^{-1})$.
(c) $\operatorname{cond}(\lambda A) = \operatorname{cond}(A)$, *para todo* $\lambda \in \Bbbk \setminus \{0\}$.

Demostración. Por el teorema VIII.2.16(a), $|||B||| \geq \varrho(B)$, para todo matriz $B \in \mathcal{M}_n(\Bbbk)$; en particular, $|||I_n||| \geq \varrho(I_n) = 1$, de modo que se verifica que

$$1 \leq |||I_n||| = |||A\,A^{-1}||| \leq |||A|||\,|||A^{-1}||| = \operatorname{cond}(A).$$

Por otra parte,

$$\operatorname{cond}(A) = |||A|||\,|||A^{-1}||| = |||A^{-1}|||\,|||A||| = \operatorname{cond}(A^{-1})$$

y, finalmente, para todo $\lambda \in \Bbbk$ no nulo se tiene que

$$\operatorname{cond}(\lambda A) = |||\lambda A||| \, |||(\lambda A)^{-1}||| = |\lambda| \, |\lambda^{-1}| \, |||A||| \, |||A^{-1}||| = \operatorname{cond}(A).$$

■

Además, si consideramos como norma matricial la subordinada a $|||\cdot|||_2$ se tiene que:

Proposición VIII.3.6. *Sea $A \in \mathcal{M}_n(\Bbbk)$ una matriz invertible. Se verifica que*

$$\operatorname{cond}_2(A) = \sqrt{\frac{\lambda_{máx}}{\lambda_{mín}}}$$

donde $\lambda_{máx}$ y $\lambda_{mín}$ son, respectivamente, el menor y el mayor de los autovalores de la matriz A^*A.

Demostración. En primer lugar hemos de tener en cuenta que A^*A es hermítica y definida positiva por ser A una matriz invertible (véase la proposición V.5.18), por lo que los autovalores de A^*A son reales y positivos. Por otra parte, aplicando el teorema VIII.2.7 se verifica que

$$|||A|||_2^2 = \varrho(A^*A) = \lambda_{\text{máx}}$$

y

$$|||A^{-1}|||_2^2 = \varrho((A^{-1})^*A^{-1}) = \varrho(A^{-1}(A^{-1})^*) = \varrho((A^*A)^{-1}) = \frac{1}{\lambda_{\text{mín}}}.$$

■

Nota VIII.3.7. Sea $A \in \mathcal{M}_n(\Bbbk)$ una matriz invertible. De la proposición VIII.2.9 se deduce que:

(a) Si A es normal y $\operatorname{sp}(A) = \{\lambda_1, \ldots, \lambda_n\}$, entonces

$$\operatorname{cond}_2(A) = |||A|||_2 \, |||A^{-1}|||_2 = \varrho(A)\varrho(A^{-1}) = \frac{\varrho(A)}{\mu(A)}$$

siendo $\mu(A) = \text{mín}_{1 \le i \le n} |\lambda_i|$.

(b) Si $A \in \mathcal{M}_n(\Bbbk)$ es una matriz invertible y normal se verifica que

$$\operatorname{cond}(A) = |||A||| \, |||A^{-1}||| \ge \varrho(A)\varrho(A^{-1}) = \operatorname{cond}_2(A)$$

para cualquier norma matricial subordinada $|||\cdot|||$ (véase el teorema VIII.2.16 y el apartado (a) anterior). Es decir, para matrices normales el número de condición cond_2 es el menor de todos.

(c) En el caso particular de que A sea unitaria entonces $\operatorname{cond}_2(A) = 1$.

(d) Como la norma $\|\!| \cdot |\!\|_2$ es invariante por transformaciones unitarias, se tiene que $\mathrm{cond}_2(A)$ es invariante por transformaciones unitarias, es decir,

$$\mathrm{cond}_2(A) = \mathrm{cond}_2(A\,Q) = \mathrm{cond}_2(Q\,A) = \mathrm{cond}_2(Q^*AQ),$$

si $Q^*Q = I_n$.

Hagamos unas consideraciones finales respecto al problema que nos ocupa en esta sección.

- Como hemos visto en la proposición VIII.3.6, siempre se verifica que el numero de condición de una matriz es un número mayor o igual que 1. Por tanto, el sistema lineal $A\mathbf{x} = \mathbf{b}$ estará tanto mejor condicionado cuando más próximo a 1 esté $\mathrm{cond}(A)$.
- En el caso de que A sea una matriz unitaria, el sistema $A\mathbf{x} = \mathbf{b}$ siempre está bien condicionado para $\|\!| \cdot |\!\|_2$, ya que $\mathrm{cond}_2(A) = 1$; además, las transformaciones unitarias conservan el número $\mathrm{cond}_2(A)$.
- Cuando se necesita resolver un sistema lineal $A\mathbf{x} = \mathbf{b}$ siendo A una matriz invertible con un número de condición elevado, se hace necesario utilizar un *precondicionador*. La idea básica es sencilla: tomar una matriz invertible M de forma que la matriz $A' = MA$ tenga un condicionamiento pequeño; después, bastará resolver el sistema $A'\mathbf{x} = \mathbf{b}'$ siendo $\mathbf{b}' = M\mathbf{b}$. Sin embargo, lo que no es sencillo es, precisamente, encontrar esta matriz M. Un posible elección, de fácil cálculo, es considerar $M = D^{-1}$ siendo $D = \mathrm{diag}(A)$.

  La idea aquí expuesta es la de un precondicionador por la izquierda. También se suelen tomar precondicionadores:
  - Por la derecha: $A' = AM,\ A'\mathbf{y} = \mathbf{b},\ \mathbf{x} = M\mathbf{y}$.
  - Por ambos lados: $M = C^2,\ A' = CAC, \mathbf{b}' = C\mathbf{b},\ \mathbf{x} = C\mathbf{y}$.
  - Simétricos: $M = CC^{\mathrm{t}},\ A' = CAC^{\mathrm{t}},\ \mathbf{b}' = C\mathbf{b}, A'\mathbf{y} = \mathbf{b},\ \mathbf{x} = C^{\mathrm{t}}\mathbf{y}$.

  lo que puede dar una idea de lo sofisticado de estas técnicas.

Analicemos ahora, con más detalle, el ejemplo de Wilson.

Ejemplo VIII.3.8. Consideremos

$$A = \begin{pmatrix} 10 & 7 & 8 & 7 \\ 7 & 5 & 6 & 5 \\ 8 & 6 & 10 & 9 \\ 7 & 5 & 9 & 10 \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} 32 \\ 23 \\ 33 \\ 31 \end{pmatrix} \quad \text{y} \quad \delta\mathbf{b} = \begin{pmatrix} 0{,}1 \\ -0{,}1 \\ 0{,}1 \\ -0{,}1 \end{pmatrix}.$$

La solución del sistema $A\mathbf{x} = \mathbf{b}$ es $\mathbf{u} = (1,1,1,1)^{\mathrm{t}}$, mientras que la solución del sistema $A\mathbf{x} = \mathbf{b} + \delta\mathbf{b}$ es

$$\mathbf{u} + \delta\mathbf{u} = (9{,}2, -12{,}6, 4{,}5, -1{,}1)^{\mathrm{t}}.$$

El polinomio característico de A viene dado por

$$\aleph_A(x) = \det(A - xI_4) = x^4 - 35x^3 + 146x^2 - 100x + 1$$

y tiene como raíces aproximadas los números

$$\lambda_1 \simeq 0{,}01015004839789,\ \lambda_2 \simeq 0{,}84310714985503,$$

$$\lambda_3 \simeq 3{,}85805745594495 \text{ y } \lambda_4 \simeq 30{,}28868534580212.$$

De esta forma, por ser A simétrica, el apartado (a) de la nota VIII.3.7 determina

$$\mathrm{cond}_2(A) = \frac{\lambda_4}{\lambda_1} \simeq 2984{,}092701675342.$$

Por tanto, no es de extrañar el mal comportamiento que, tras las pequeñas modificaciones en los datos, se observó anteriormente.

Ejercicios del tema VIII

Ejercicio 1. Probar que un espacio normado toda sucesión de Cauchy está acotada.

Ejercicio 2. Probar que

$$\|\mathbf{v}\|_\infty \leq \|\mathbf{v}\|_2 \leq \sqrt{n}\ \|\mathbf{v}\|_\infty$$
$$\frac{1}{n}\ \|\mathbf{v}\|_1 \leq \|\mathbf{v}\|_\infty \leq n\ \|\mathbf{v}\|_1$$
$$\|\mathbf{v}\|_2 \leq \|\mathbf{v}\|_1 \leq \sqrt{n}\ \|\mathbf{v}\|_2$$

para todo $\mathbf{v} \in \mathbb{C}^n$.

Ejercicio 3. Probar que para todo $p \geq 1$ se verifica que

$$\|\mathbf{v}\|_\infty \leq \|\mathbf{v}\|_p \leq \sqrt[p]{n}\|\mathbf{v}\|_\infty,$$

para cualquier $\mathbf{v} = (v_1, \ldots, v_n)^{\mathrm{t}} \in \mathbb{C}^n$. Concluir que

$$\|\mathbf{v}\|_\infty = \lim_{p\to\infty} \|\mathbf{v}\|_p,$$

para cualquier $\mathbf{v} \in \mathbb{C}^n$.

Ejercicio 4. Sea $A \in \mathcal{M}_n(\Bbbk)$. Probar que $\varrho(A^m) = \varrho(A)^m$, para todo $m \in \mathbb{N}$.

Ejercicio 5. Sea $A \in \mathcal{M}_n(\Bbbk)$ una matriz hermítica con $\mathrm{sp}(A) = \{\lambda_1, \ldots, \lambda_n\}$.

1. Probar que para todo $\lambda \in \mathbb{R}$ y $\mathbf{v} \in \Bbbk^n$ no nulo, se verifica que
$$\min_{1\leq j\leq n} |\lambda - \lambda_j| \leq \frac{\|A\mathbf{v} - \lambda\mathbf{v}\|_2}{\|\mathbf{v}\|_2}.$$
2. Estudiar cómo se puede aplicar el resultado anterior para obtener aproximaciones de los autovalores de la matriz A.

Ejercicio 6. Sean $A \in \mathcal{M}_n(\mathbb{C})$ una matriz invertible tal que $A = B^2$. Probar que:

1. $\mathrm{cond}_2(A) \leq \mathrm{cond}_2(B)^2$.
2. Si A es normal y $\mathrm{cond}_2(A) > 1$, entonces $\mathrm{cond}_2(B) < \mathrm{cond}_2(A)$.

Ejercicio 7. Sea $A \in \mathcal{M}_n(\Bbbk)$ una matriz invertible. Demostrar las siguientes desigualdades:

$$\frac{1}{n}\ \mathrm{cond}_2(A) \leq \mathrm{cond}_1(A) \leq n\ \mathrm{cond}_2(A).$$
$$\frac{1}{n}\ \mathrm{cond}_\infty(A) \leq \mathrm{cond}_2(A) \leq n\ \mathrm{cond}_\infty(A).$$
$$\frac{1}{n^2}\ \mathrm{cond}_1(A) \leq \mathrm{cond}_\infty(A) \leq n^2\ \mathrm{cond}_1(A).$$

TEMA IX

Métodos directos de resolución de sistemas lineales de ecuaciones

Hemos estudiado los sistema de ecuaciones lineales en varias ocasiones a lo largo de la asignatura; por ejemplo, en los temas II y VI dimos condiciones necesarias y suficientes para que un sistema tuviese una, infinita o ninguna solución, y tratamos algunos aspectos relacionados con su resolución y la forma de sus soluciones. En este tema nos vamos a ocupar de los métodos numéricos directos para la resolución de tales sistemas.

Cuando hablamos de métodos directos nos referimos a aquellos "procedimientos algorítmicos" que en un número finito de pasos alcanzan la solución exacta del sistema. Si bien el término exacto sólo tendrá sentido desde un punto de vista teórico, ya que el mal condicionamiento del sistema o la propagación de errores de redondeo sólo nos permitirán trabajar con buenas aproximaciones en el mejor de los casos.

Es fundamental tener en cuenta que este tipo de métodos adquiere su mayor interés cuando tratamos resolver sistemas con matrices de órdenes altos, donde el coste computacional de otros métodos (como, por ejemplo, la fórmula de Cramer) collevan un número de operaciones prohibitivo.

En este tema estudiaremos el método de eliminación gaussina (que ya apareció en el tema II cuando estudiamos las formas reducidas de una matriz) y las factorizaciones LU, de Cholesky y QR. La clave del uso de la eliminación gaussina y las tres factorizaciones citadas como métodos de resolución de sistema de ecuaciones lineales reside en una misma idea: reducir la resolución del sistema a dado a la resolución de uno varios sistemas de ecuaciones lineales en forma triangular. Estos métodos no son de validez general (a excepción de la resolución basada en la factorización QR) por lo que en cada caso daremos condiciones necesarias y/o suficientes para su aplicación. Por otra parte, si bien no nos ocuparemos del estudio de la complejidad de estos métodos, sí procuraremos incluir el coste computacional de cada uno de los métodos estudiados.

Con el ánimo de contextualizar los métodos y factorizaciones que aquí estudiaremos, comentamos que la factorizáción LU (y su variante PA = LU) consiste en descomponer la matriz del sistema como el producto de matriz triangular inferior L

por una triangular superior U, por lo que guarda una estrecha relación con el cálculo de las formas reducidas y escalonadas de una matriz. La factorización de Cholesky es, en cierto sentido, la análoga a la LU para matrices simétricas definidas positivas (está factorización ya apareció en el tema V). Ambas factorizaciones se apoyan en el método de eliminación gaussiana. La factorización QR consiste en descomponer la matriz del sistema como producto de matriz ortogonal por una triangular superior, el método usado para calcular tal descomposición es la versión numérica del método de ortonormalización de Gram-Schmidt, estudiado en el tema IV.

La bibliografía empleada para la elaboración de este tema ha sido [**Cia82**], [**IR99**], [**QSS07**] y algunas pinceladas de [**QS06**].

1. Eliminación Gaussiana y factorización LU

Comenzaremos estudiando algunos métodos para resolución de los sistemas de ecuaciones lineales de la forma $A\mathbf{x} = \mathbf{b}$ con $A \in \mathcal{M}_n(\mathbb{k})$ invertible y $\mathbf{b} \in \mathbb{k}^n$, es decir, de los sistemas compatibles de determinados (véase la definición II.5.1).

Resolución de sistemas triangulares.

Consideremos un sistema de tres ecuaciones lineales con tres incógnitas cuya matriz asociada es triangular inferior e invertible:

$$\begin{pmatrix} l_{11} & 0 & 0 \\ l_{21} & l_{22} & 0 \\ l_{31} & l_{32} & l_{33} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix},$$

de forma abreviada $L\mathbf{x} = \mathbf{b}$.

Como L es invertible por hipótesis, las entradas de su diagonal principal $l_{ii},\ i = 1, 2, 3$, son no nulas, de donde se sigue que podemos calcular secuencialmente los valores de las incógnitas $x_i,\ i = 1, 2, 3$, como sigue

$$\begin{aligned} x_1 &= b_1/l_{11}, \\ x_2 &= (b_2 - l_{21}x_1)/l_{22}, \\ x_3 &= (b_3 - l_{31}x_1 - l_{32}x_2)/l_{33}. \end{aligned}$$

Este algoritmo se puede extender a sistemas con n ecuaciones y n incógnitas se llama **sustitución hacia adelante**. En el caso de un sistema $L\mathbf{x} = \mathbf{b}$, donde L es una matriz triangular inferior e invertible de orden $n \geq 2$, el método tiene la siguiente

forma:

$$\begin{aligned} x_1 =& \frac{b_1}{l_{11}}, \\ x_i =& \frac{1}{l_{ii}} \left(b_i - \sum_{j=1}^{i-1} l_{ij} x_j \right), \ i = 2, \ldots, n. \end{aligned}$$

El número de multiplicaciones y divisiones para ejecutar este algoritmo es $n(n+1)/2$, mientras que el número de sumas y restas es $n(n-1)/2$. Por lo que la cuenta global de las operaciones del algoritmo de sustitución hacia atrás es del orden de n^2.

Unas conclusiones similares se pueden obtener para un sistema de ecuaciones lineales $U\mathbf{x} = \mathbf{b}$, donde U es una matriz triangular superior e invertible de orden $n \geq 2$. En este caso el algoritmo se llama **sustitución hacia atrás** y en sus versión general puede escribirse como:

$$\begin{aligned} x_n =& \frac{b_n}{u_{11}}, \\ x_i =& \frac{1}{u_{ii}} \left(b_i - \sum_{j=i+1}^{n} u_{ij} x_j \right), \ i = 2, \ldots, n. \end{aligned}$$

De nuevo el coste computacional es del orden de n^2 operaciones.

En la práctica 11 exploraremos la implementación de los algoritmos sustitución hacia atrás y hacia adelante. Por otra parte, en el apartado 3.2.2 de [**QSS07**] se pueden encontrar referencias sobre la propagación de errores de redondeo tanto para la resolución de sistemas triangulares mediante sustitución hacia adelante como hacía atrás.

Eliminación gaussiana y factorización LU.

El método de eliminación gaussiana consiste el reducir un sistema de ecuaciones lineales $A\mathbf{x} = \mathbf{b}$, con $A \in \mathcal{M}_n(\mathbb{k})$ invertible y $\mathbf{b} \in \mathbb{k}^n$ en otro equivalente (es decir, que tenga las mismas soluciones) de la forma $U\mathbf{x} = \hat{\mathbf{b}}$, donde $U \in \mathcal{M}_n(\mathbb{k})$ es una matriz triangular superior y $\hat{\mathbf{b}} \in \mathbb{k}^n$. Este último sistema se podrá resolver usando el algoritmo de sustitución hacia atrás, ya que U será invertible al serlo A. Como veremos a continuación el método de eliminación gaussiana no es más que una variante del método de Gauss-Jordan estudiando en el tema II.

Vamos a denotar $A^{(1)}\mathbf{x} = \mathbf{b}^{(1)}$ al sistema original, y supongamos que $a_{11}^{(1)} = a_{11}$ es distinto de cero. Introduciendo los multiplicadores

$$l_{i1} = a_{i1}^{(1)} / a_{i1}^{(1)}, \ i = 2, 3, \ldots, n,$$

donde $a_{ij}^{(1)} = a_{ij}$, es posible eliminar la incógnita x_1 es las filas distintas de las primera, sencillamente restándole a la fila i-esima la primera multiplicada por l_{i1} y haciendo los mismo en el término independiente. Si definimos

$$a_{ij}^{(2)} = a_{ij}^{(1)} - l_{i1}a_{ij}^{(1)},\ i,j = 2,\dots,n,$$
$$b_{i}^{(2)} = b_{i}^{(1)} - l_{i1}b_{1}^{(1)},\ i = 2,\dots,n,$$

donde $b_i^{(1)}$ denota los elementos de $\mathbf{b}^{(1)}$. De este modo obtenemos un sistema sistema

$$\begin{pmatrix} a_{11}^{(1)} & a_{12}^{(1)} & \dots & a_{1n}^{(1)} \\ 0 & a_{22}^{(2)} & \dots & a_{2n}^{(2)} \\ \vdots & \vdots & & \vdots \\ 0 & a_{n2}^{(2)} & \dots & a_{nn}^{(2)} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1^{(1)} \\ b_2^{(2)} \\ \vdots \\ b_n^{(2)} \end{pmatrix},$$

que denotaremos $A^{(2)}\mathbf{x} = \mathbf{b}^{(2)}$. Obsérvese que este sistema es equivalente al anterior, ya que solamente hemos realizado operaciones elementales por filas de tipo III en la matrices $A^{(1)}$ y $\mathbf{b}^{(1)}$.

De forma análoga, podemos transformar el sistema $A^{(2)}\mathbf{x} = \mathbf{b}^{(2)}$ en otro equivalente donde la incógnita x_2 haya sido eliminada de las filas $3,\dots,n$. En general, obtenemos una sucesión finita de sistema equivalentes

$$A^{(k)}\mathbf{x} = \mathbf{b}^{(k)},\ k = 1,\dots,n,$$

donde para $k \geq 2$ la matriz $A^{(k)}$ es de la forma

$$A^{(k)} = \begin{pmatrix} a_{11}^{(1)} & a_{12}^{(1)} & \dots & \dots & \dots & a_{1n}^{(1)} \\ 0 & a_{22}^{(2)} & \dots & \dots & \dots & a_{2n}^{(2)} \\ \vdots & & \ddots & & & \vdots \\ 0 & \dots & 0 & a_{kk}^{(k)} & \dots & a_{kn}^{(k)} \\ \vdots & & \vdots & \vdots & & \vdots \\ 0 & \dots & 0 & a_{nk}^{(k)} & \dots & a_{nn}^{(k)} \end{pmatrix}$$

suponiendo que $a_{ii}^{(i)} \neq 0$, para $i = 1,\dots,k-1$. Es claro que para $k = n$ se consigue un sistema triangular superior $A^{(n)}\mathbf{x} = \mathbf{b}^{(n)}$

$$\begin{pmatrix} a_{11}^{(1)} & a_{12}^{(1)} & \dots & \dots & a_{1n}^{(1)} \\ 0 & a_{22}^{(2)} & \dots & \dots & a_{2n}^{(2)} \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ 0 & & & & a_{nn}^{(n)} \end{pmatrix}$$

Siendo consistentes con la notación hemos introducido previamente, denotamos U la matriz triangular superior $A^{(n)}$. Las entradas $a_{kk}^{(k)}$ se llaman **pivotes** y deben ser no nulos para $k = 1, \ldots, n-1$.

Con el objeto de resaltar la fórmula que transforma el sistema k-ésimo en el $(k+1)$-ésimo, para $k = 1, \ldots, n-1$ suponiendo que $a_{kk}^{(k)} \neq 0$, definimos el multiplicador

$$l_{ik} = a_{ik}^{(k)} / a_{kk}^{(k)}, \ i = k+1, \ldots, n$$

y tomamos

$$a_{ij}^{(k+1)} = a_{ij}^{(k)} - l_{ik} a_{ij}^{(k)}, \ i, j = k+1, \ldots, n, \tag{IX.1.1}$$

$$b_{i}^{(k+1)} = b_{i}^{(k)} - l_{ik} b_{k}^{(k)}, \ i = k+1, \ldots, n, \tag{IX.1.2}$$

El método de eliminación gaussiana requiere $2(n-1)n(n+1)/3 + n(n-1)$ operaciones (sumas, restas, multiplicaciones y divisiones) a lo que tendremos que añadir las $n(n+1)/2$ necesarias para resolver el sistema $U\mathbf{x} = \mathbf{b}^{(n)}$. Por tanto, serán necesarias alrededor de $1/6\, n\, (4\, n^2 - 7 + 9\, n)$ operaciones para resolver el sistema $A\mathbf{x} = \mathbf{b}$ usando el método de eliminación gaussiana. Ignorando los términos de menor grado en la expresión anterior podemos concluir que el método de eliminación gaussiana tiene un coste de $2n^3/3$ operaciones. El lector interesado puede encontrar un estudio sobre la propagación de errores de redondeo para el método de eliminación gaussiana en el apartado 3.2.2 de [**QSS07**].

Como hemos ido remarcando, el método de eliminación gaussiana termina satisfactoriamente si, y sólo si, todos los pivotes $a_{kk}^{(k)}, \ k = 1, \ldots, n-1$ son distintos de cero. Desafortunadamente, que A tenga todas las entradas sus entradas en diagonal no nulas no es suficiente para garantizar que los pivotes sean no nulos durante el proceso de eliminación.

Ejemplo IX.1.1. La matriz

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 5 \\ 7 & 8 & 9 \end{pmatrix}$$

tiene todas las entradas de su diagonal no nulas, sin embargo se cumple que

$$A^{(2)} = \begin{pmatrix} 1 & 2 & 3 \\ 0 & \boxed{0} & -1 \\ 0 & -6 & -12 \end{pmatrix}.$$

Por lo que el método de eliminación gaussiana se ve interrumpido en el segundo paso, ya que $a_{22}^{(2)} = 0$.

Por consiguiente, se necesitan condiciones más restrictivas sobre A para garantizar la aplicabilidad del método. En breve demostraremos que una condición necesaria y suficiente para que todos los pivotes sean no nulos es que la matriz A tenga todos sus menores principales de orden $i = 1, \ldots, n-1$, distintos de cero (véase el teorema IX.1.4); nótese que la matriz de ejemplo anterior no tiene esta propiedad. Otros tipos de matrices en las que la eliminación gaussiana se puede aplicar con total seguridad de éxito son las siguientes:

- Las matrices diagonalmente dominantes por filas o por columnas[1].
- Las matrices simétricas definidas positivas.

Volveremos a estas cuestiones más adelante. Ahora nos vamos a ocupar de utilizar la eliminación gaussiana para calcular una factorización de la matriz A en producto de dos matrices, $A = LU$, con $U = A^{(n)}$. Como L y U sólo dependen de A y no del vector de términos independientes, la misma factorización puede ser utilizada para resolver los diferentes sistemas de ecuaciones lineales que se obtienen al variar $\mathbf{b}$. Esto supone una considerable reducción de número de operaciones necesarias para resolverlos, ya que el mayor esfuerzo computacional (entorno a $2n^3/3$ operaciones) se consume en el proceso de eliminación.

Según la igualdad IX.1.1, la matriz de paso a izquierda de $A^{(k)}$ a $A^{(k+1)}$ es

$$L_k = \begin{pmatrix} 1 & \ldots & 0 & 0 & \ldots & 0 \\ \vdots & \ddots & \vdots & \vdots & & \vdots \\ 0 & & 1 & 0 & & 0 \\ 0 & & -l_{k+1,\,k} & 1 & & 0 \\ \vdots & & \vdots & \vdots & & \vdots \\ 0 & \ldots & -l_{n,\,k} & 0 & \ldots & 1 \end{pmatrix}$$

con $l_{ik} = a_{ik}^{(k)}/a_{kk}^{(k)}$, para cada $k = 1, \ldots, n-1$.

Obsérvese que $L_k = I_n - \vec{\ell}_k \mathbf{e}_k^{\mathrm{t}}$ donde $\vec{\ell}_k = (0, \ldots, 0, l_{k+1,\,k}, \ldots, l_{n,\,k})^{\mathrm{t}} \in \Bbbk^n$ y $\mathbf{e}_k$ es el vector k-ésimo de la base usual de $\Bbbk^n$.

Lema IX.1.2. *Con la notación anterior, se cumple que:*

(a) *La matriz L_k es invertible y $L_k^{-1} = I_n + \vec{\ell}_k \mathbf{e}_k^{\mathrm{t}}$.*

[1] Una matriz $A = (a_{ij}) \in \mathcal{M}_n(\Bbbk)$ es **diagonalmente dominante por filas (por columnas, resp.)** si

$$|a_{ii}| > \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}|,$$

para todo $i = 1, \ldots, n$ (si $|a_{ii}| > \sum_{\substack{i=1 \\ i \neq j}}^{n} |a_{ij}|$, para todo $j = 1, \ldots, n$, resp.).

(b) $L_{n-1}L_{n-2}\cdots L_1 = (I_n + \sum_{i=1}^{n-1} \mathbf{e}_i^{\mathrm{t}})\ y$

$$(L_{n-1}L_{n-2}\cdots L_1)^{-1} = \begin{pmatrix} 1 & 0 & \dots & \dots & 0 \\ l_{21} & 1 & & & 0 \\ \vdots & l_{32} & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & 0 \\ l_{n1} & l_{n2} & \dots & l_{n,n-1} & 1 \end{pmatrix}.$$

Demostración. Para el apartado (a) basta tener en cuenta que la matrices $L_j,\ j = 1,\dots,n-1$, son producto de matrices elementales de tipo III, el apartado (b) se comprueba de forma directa por inducción sobre n; los detalles de ambas demostraciones se proponen como ejercicio al lector (ejercicio 1). ∎

Según lo anterior, si denotamos $L = (L_{n-1}L_{n-2}\cdots L_1)^{-1} = L_1^{-1}\cdots L_{n-2}^{-1}L_{n-1}^{-1}$, se sigue que

$$A = LU \tag{IX.1.3}$$

donde U es triangular superior y L es triangular inferior con unos en sus diagonal principal.

Nótese que una vez que hemos calculado las matrices L y U, para hallar la solución del sistema $A\mathbf{x} = \mathbf{b}$ sólo hay que resolver sucesivamente los dos sistemas triangulares siguientes

$$L\mathbf{y} = \mathbf{b}$$
$$U\mathbf{x} = \mathbf{y}.$$

En la práctica 11, veremos la implementación de un algoritmo para el cálculo de la factorización LU, así como diversos ejemplos de resolución de sistemas de ecuaciones lineales usando dicha factorización.

Definición IX.1.3. Dada una matriz $A \in \mathcal{M}_n(\Bbbk)$ se llama **factorización LU** de A, a $LU = A$ tal que L es triangular inferior con unos en su diagonal principal y U es triangular superior.

El siguiente resultado establece una relación entre los menores principales de una matriz cuadrada y su factorización LU. De hecho nos da una condición necesaria y suficiente para que exista una única factorización LU de una matriz cuadrada.

Teorema IX.1.4. *Sea $A \in \mathcal{M}_n(\Bbbk)$. La factorización LU de A existe y es única si, y sólo si, los menores principales de orden $i = 1,\dots,n-1$ de A son no nulos*[2].

[2]Un caso importante de matrices con esta propiedad son las simétricas (y hermíticas) definidas positivas (véase la proposición V.5.13).

Demostración. Sea

$$A_i = \begin{pmatrix} a_{11} & \dots & a_{1i} \\ \vdots & & \vdots \\ a_{i1} & \dots & a_{ii} \end{pmatrix} \in \mathcal{M}_i(\mathbb{R}),$$

es decir, A_i es la submatriz de A que se obtiene al eliminar las últimas $n-i$ filas y columnas.

En primer lugar supongamos que los menores principales , $|A_i|,\ i = 1, \dots, n-1$, de A son no nulos, y veamos por inducción sobre i que existe una única factorización LU de A. Es claro que el resultado es cierto para $i = 1$. Supongamos, pues, que A_{i-1} posee una única factorización LU, $A_{i-1} = L^{(i-1)}U^{(i-1)}$, y demostremos que A_i también tiene una única factorización LU. Para ello consideramos la siguiente partición de la matriz A_i,

$$A_i = \left(\begin{array}{c|c} A_{i-1} & \mathbf{c} \\ \hline \mathbf{d}^{\mathrm{t}} & a_{ii} \end{array}\right)$$

y busquemos una factorización de A_i de la forma

$$A_i = L^{(i)}U^{(i)} = \left(\begin{array}{c|c} L^{(i-1)} & 0 \\ \hline \vec{\ell}^{\,\mathrm{t}} & 1 \end{array}\right)\left(\begin{array}{c|c} U^{(i-1)} & \mathbf{u} \\ \hline \mathbf{0}^{\mathrm{t}} & u_{ii} \end{array}\right). \tag{IX.1.4}$$

Si calculamos el producto de estos dos factores e igualamos los bloques a los de A_i, concluimos que los vectores $\vec{\ell}$ y $\mathbf{u}$ son las soluciones de los sistemas $L^{(i-1)}\mathbf{x} = \mathbf{c}$ y $\mathbf{y}U^{(i-1)} = \mathbf{d}^{\mathrm{t}}$. Teniendo ahora en cuenta que que $0 \neq |A_{i-1}| = |L^{(i-1)}|\,|U^{(i-1)}|$, concluimos que la existencia y unicidad de $\mathbf{u}$ y de $\vec{\ell}$, por el teorema de Rouché-Fröbenius. Luego, existe una única factorización LU de A_i, con $u_{ii} = a_{ii} - \vec{\ell}\mathbf{u}$.

Recíprocamente, supongamos que existe una única factorización LU de A. Queremos demostrar que los menores principales de A son no nulos. Vamos a distinguir dos casos según A sea invertible o no.

Comencemos suponiendo que A es invertible. Según la igualdad (IX.1.4)

$$0 \neq |A_i| = |L^{(i)}|\,|U^{(i)}| = |U^{(i)}| = u_{11}u_{22}\cdots u_{ii},$$

de donde se sigue, tomando $i = n$ que $|A| = |A_n| = u_{11}u_{22}\cdots u_{nn} \neq 0$, y por consiguiente que $|A_i| \neq 0,\ i = 1, \dots, n-1$.

Sea ahora A no invertible y supongamos que, al menos, una de las entradas de la diagonal principal de U es no nula. Si u_{kk} es la entrada no nula de la diagonal de U de menor índice k. Por (IX.1.4), podemos garantizar que la factorización se puede calcular sin problemas hasta la etapa $k+1$. A partir de entonces, al ser la matriz $U^{(k)}$ no invertible, por el teorema de Rouché-Fröbenius se tiene que o bien no existe $\vec{\ell}$ o bien no es único, y lo mismo ocurre con la factorización. De modo que para que esto no ocurra (como es nuestro caso) las entradas de la diagonal principal u_{kk} de U

tienen que ser no nulas hasta el índice $k = n-1$ inclusive, y por consiguiente, de la igualdad $|A_i| = u_{11}u_{22}\cdots u_{ii}$, se sigue que $|A_i| \neq 0,\ i = 1,\ldots,n-1$. ∎

Nótese que en el caso en que la factorización LU sea única, tenemos que $|A| = |LU| = |L||U| = |U|$, es decir, el determinante de A es el producto de los pivotes:

$$|A| = u_{11}\cdots u_{nn} = \prod_{k=1}^{n} a_{kk}^{(k)}.$$

Terminamos esta sección mostrando algunos resultados sobre la factorización LU de ciertos tipos especiales de matrices.

Proposición IX.1.5. *Sea $A \in \mathcal{M}_n(\Bbbk)$ si A es diagonalmente semidominante por filas o por columnas*[3]*, entonces existe factorización LU. En particular, si A es diagonalmente dominante por columnas, entonces $|l_{ij}| \leq 1$, para todo $i, j = 1,\ldots,n$.*

Demostración. El lector interesado puede encontrar una demostración de este resultado en [Golub G.; Loan C. V. *Matrix Computations*. The John Hopkins Univ. Press, Baltimore and London. 1989] o en [Higham N. *Accuracy and Stability of Numerical Algorithms.* SIAM Publications, Philadelphia, PA. 1996]. ∎

Finalmente, consideremos el caso de una matriz tridiagonal

$$A = \begin{pmatrix} b_1 & c_1 & & & \\ a_2 & b_2 & c_2 & & \\ & \ddots & \ddots & \ddots & \\ & & \mu a_{n-1} & b_{n-1} & c_{n-1} \\ & & & a_n & b_n \end{pmatrix}.$$

En este caso, las matrices L y U de la factorización LU de A son bidiagonales de la forma

$$L = \begin{pmatrix} 1 & & & \\ \alpha_2 & 1 & & \\ & \ddots & \ddots & \\ & & \alpha_n & 1 \end{pmatrix} \quad \text{y} \quad U = \begin{pmatrix} \beta_1 & c_1 & & \\ & \beta_2 & \ddots & \\ & & \ddots & c_{n-1} \\ & & & \beta_n \end{pmatrix}.$$

[3]Una matriz $A = (a_{ij}) \in \mathcal{M}_n(\Bbbk)$ es **diagonalmente semidominante por filas (por columnas, resp.)** si

$$|a_{ii}| \geq \sum_{\substack{j=1 \\ j\neq i}}^{n} |a_{ij}|,$$

para todo $i = 1,\ldots,n$ (si $|a_{ii}| \geq \sum_{\substack{i=1 \\ i\neq j}}^{n} |a_{ij}|$, para todo $j = 1,\ldots,n$, resp.).

Los coeficientes α_i y β_i pueden ser fácilmente calculados a partir de las siguientes relaciones:

$$\beta_1 = b_1,\ \alpha_i = \frac{a_i}{\beta_{i-1}},\ \beta_i = b_i - \alpha_i c_{i-1},\ i = 2, \ldots, n.$$

Este algoritmo se puede aplicar a la resolución de sistema tridiagonales $A\mathbf{x} = \mathbf{f}$ resolviendo los correspondientes sistemas bidiagonales $L\mathbf{y} = \mathbf{f}$ y $U\mathbf{x} = \mathbf{y}$, para los que se cumplen las siguientes fórmulas:

$$y_1 = f_1,\ y_i = f_i - \alpha_i y_{i-1},\ i = 2, \ldots, n,$$

$$x_n = \frac{y_n}{\beta_n},\ x_i = (y_i - c_i x_{i+1})/\beta_i,\ i = n-1, \ldots, 1.$$

El algoritmo requiere $8n - 7$ operaciones; precisamente $3(n-1)$ para la factorización y $5n - 4$ para la resolución de los sistemas bidiagonales.

2. Factorización PA = LU. Técnicas de pivoteo

Como se ha apuntado anteriormente, el método de eliminación gaussiana (y por lo tanto la factorización LU) falla cuando uno encontramos un pivote nulo. En estos casos, se requiere lo que se conoce como *técnica de pivoteo* que consiste en intercambiar filas (o columnas[4]) para evitar los pivotes nulos.

Ejemplo IX.2.1. Consideremos de nuevo la matriz del ejemplo IX.1.1:

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 5 \\ 7 & 8 & 9 \end{pmatrix}$$

en el que el método de eliminación gaussiana fallaba en la segunda etapa al aparecer un pivote nulo. En este caso, sin más que intercambiar la fila segunda y la tercera de $A^{(2)}$ (es decir, haciendo una operación elemental por filas de tipo I) obtenemos la matriz triangular buscada

$$A^{(2')} = \begin{pmatrix} 1 & 2 & 3 \\ 0 & -6 & 12 \\ 0 & 0 & -1 \end{pmatrix} = U.$$

En esta sección consideramos el caso de los sistemas de ecuaciones lineales de la forma $A\mathbf{x} = \mathbf{b}$ con $A \in \mathcal{M}_n(\Bbbk)$ no necesariamente invertible y $\mathbf{b} \in \Bbbk^n$; por lo que se admite la posibilidad de que sistema tenga infinitas soluciones, es decir, que sea compatible indeterminado, o que no tenga ninguna solución, es decir, que sea incompatible (véase la definición II.5.1).

[4]Como hacíamos en el tema II para calcular la forma reducida por filas.

Teorema IX.2.2. *Sea $A \in \mathcal{M}_n(\Bbbk)$. Existen una matriz permutación P, una matriz L triangular inferior con unos en su diagonal principal y una matriz U triangular superior tales que*

$$PA = LU.$$

Demostración. Supongamos que en la etapa k-ésima del método de eliminación gaussiana nos encontramos con un pivote nulo, es decir,

$$L_{k-1}\cdots L_1 A = A^{(k)} = \begin{pmatrix} a_{11}^{(1)} & a_{12}^{(1)} & \dots & \dots & \dots & a_{1n}^{(1)} \\ 0 & a_{22}^{(2)} & \dots & \dots & \dots & a_{2n}^{(2)} \\ \vdots & & \ddots & & & \vdots \\ 0 & \dots & 0 & a_{kk}^{(k)} & \dots & a_{kn}^{(k)} \\ \vdots & & \vdots & \vdots & & \vdots \\ 0 & \dots & 0 & a_{nk}^{(k)} & \dots & a_{nn}^{(k)} \end{pmatrix}$$

con $a_{kk}^{(k)} = 0$. Si $a_{ik}^{(k)} = 0$, para todo $i = k, \dots, n$, tomamos $L_k = I_n$ y pasamos a la siguiente etapa; en otro caso, existe $l > k$ tal que $a_{lk} \neq 0$, entonces intercambiado las filas k-ésima y l-ésima de $A^{(k)}$ conseguimos una matriz $A^{(k')}$ equivalente a $A^{(k)}$ (y por lo tanto equivalente a A) con $a_{kk}^{(k')} \neq 0$. Obsérvese que

$$A^{(k')} = P_k A^{(k)},$$

donde $P_k = T_{il}$ es la matriz elemental de permutación de las filas i-ésima y l-ésima. Salvado este obstáculo podemos continuar con el método de eliminación gaussiana, del forma que podemos hallar la matriz L_k, a partir de $A^{(k')}$, tal que

$$A^{(k+1)} = L_k A^{(k')} = L_k P_k L_{k-1} \cdots L_1 A$$

es de la forma deseada.

Por consiguiente, podemos afirmar que existen $n-1$ matrices de permutación[5], P_k, $k = 1, \dots, n-1$, tales que

$$L_{n-1}P_{n-1}\cdots L_1 P_1 A = A^{(n)} = U.$$

Tomando ahora $M = L_{n-1}P_{n-1}\cdots L_1 P_1$ y $P = P_{n-1}\cdots P_1$, concluimos que $MA = U$, y por lo tanto que $MP^{-1}PA = U$. Teniendo ahora en cuenta que $L = (MP^{-1})^{-1} = PM^{-1}$ es triangular inferior con unos en su diagonal principal (ejercicio 5), concluimos que $PA = LU$. ∎

Según el teorema anterior, podemos establecer que una permutación adecuada de las filas la matriz A original hace factible el proceso de factorización completo. Desafortunadamente, no podemos conocer a priori qué filas debe permutarse, por lo que esta decisión ha de tomarse en cada etapa k en la que aparezca una entrada

[5]Recuérdese que la matriz identidad es una matriz de permutación.

diagonal $a_{kk}^{(k)}$ nula tal y como hemos hecho en la demostración del teorema. Puesto que una permutación de filas implica cambiar el elemento pivotal, esta técnica recibe el nombre de **pivoteo por filas**. La factorización generada de esta forma devuelve la matriz original original salvo una permutación de filas, concretamente obtenemos

$$PA = LU,$$

donde P es una matriz de permutación (es decir, un producto de matrices elementales de tipo I). Si en el curso del proceso las filas k y l de A se permutan, la misma permutación debe realizarse sobre las filas homólogas de P. En correspondencia con ello, ahora deberíamos resolver los siguientes sistemas triangulares

$$L\mathbf{y} = P\mathbf{b}$$

$$\mathbf{U}\mathbf{x} = \mathbf{y}.$$

En importante destacar que el sistema $U\mathbf{x} = \mathbf{y}$ podría no tener solución o poseer infinitas soluciones, ya que es posible que las entradas de la diagonal principal de U sean nulas.

Si bien hemos usado la técnica de pivoteo por filas para salvar la aparición de pivotes nulos. Existen otros casos en los que es conveniente aplicar esta técnica; por ejemplo, un pivote $a_{kk}^{(k)}$ es demasiado pequeño puede amplificar la propagación de errores de redondeo. Por tanto, para asegurar una mejor estabilidad, se suele elegir como elemento pivotal k-ésimo la mayor (en módulo) de las entradas $a_{ik}^{(k)},\ i = k, \ldots, n$ de la matriz $A^{(k)}$ ejecutando la correspondiente permutación de las filas de $A^{(k)}$. Alternativamente, el proceso de busqueda de un pivote óptimo se puede extender a todas las entradas $a_{ij}^{(k)},\ i, j = k, \ldots, n$, esta estrategia se conoce como técnica de **pivoteo total** y requiere permutaciones de columnas, por lo que el tipo de factorización obtenida en este caso seria de la forma

$$PAQ = LU.$$

3. Factorización de Cholesky

En el tema V vimos que cualquier matriz simétrica definida positiva $A \in \mathcal{M}_n(\mathbb{R})$ factoriza como sigue

$$A = QQ^{\mathrm{t}}$$

con Q triangular inferior (véase el corolario V.5.12). Veamos que tal descomposición, llamada **factorización de Cholesky**, existe y es única para cualquier matriz hermítica (simétrica si es de entradas reales) definida positiva (véase la definición V.5.16).

Teorema IX.3.1. *Sea $A \in \mathcal{M}_n(\Bbbk)$ una matriz hermítica definida positiva. Existe una única matriz triangular inferior H con diagonal real postiva tal que*

$$A = HH^*.$$

Demostración. Sea

$$A_i = \begin{pmatrix} a_{11} & \dots & a_{1i} \\ \vdots & & \vdots \\ a_{i1} & \dots & a_{ii} \end{pmatrix} \in \mathcal{M}_i(\mathbb{R}),$$

es decir, A_i es la submatriz de A que se obtiene al eliminar las últimas $n-i$ filas y columnas. Obsérvese que A_i es hermítica y definida positiva por serlo A.

Al igual que en la demostración del teorema IX.1.4 procederemos por inducción sobre i.

Para $i = 1$ el resultado es obviamente cierto. Supongamos, pues, que se cumple para $i-1$ y veamos que también es válido para i. Por la hipótesis de inducción, existe una matriz triangular inferior H_{i-1} tal que $A_{i-1} = H_{i-1}H_{i-1}^*$. Consideremos la siguiente partición de A_i

$$A_i = \left(\begin{array}{c|c} A_{i-1} & \mathbf{v} \\ \hline \mathbf{v}^* & \alpha \end{array}\right),$$

con $\alpha \in \mathbb{R}_+$ y $\mathbf{v} \in \mathbb{C}^{i-1}$, y busquemos una factorización de A_i de la forma

$$A_i = H_iH_i^* = \left(\begin{array}{c|c} H_{i-1} & \mathbf{0} \\ \hline \mathbf{h}^* & \beta \end{array}\right)\left(\begin{array}{c|c} H_{i-1}^* & \mathbf{h} \\ \hline \mathbf{0} & \beta \end{array}\right).$$

Forzando la igualdad con las entradas de A_i se obtienen las ecuaciones $H_{i-1}\mathbf{h} = \mathbf{v}$ y $\mathbf{h}^*\mathbf{h} + \beta^2 = \alpha$. El vector $\mathbf{h}^*$ está unívocamente determinado porque H_{i-1} es invertible. Por otra parte,

$$\mathbf{h}^*\mathbf{h} = \mathbf{v}^*(H_{i-1}^{-1})^*H_{i-1}\mathbf{v} = \mathbf{v}^*(H_{i-1}H_{i-1}^*)^{-1}\mathbf{v} = \mathbf{v}^*A_{i-1}^{-1}\mathbf{v}$$

y, según vimos al final del tema I

$$0 < |A_i| = \alpha(\alpha - \mathbf{v}^*A_{i-1}^{-1}\mathbf{v}).$$

Como $\alpha > 0$, ambos hechos implican que $\alpha - \mathbf{h}^*\mathbf{h} > 0$ y por lo tanto que existe un único número real positivo β tal que $\beta^2 = \alpha - \mathbf{h}^*\mathbf{h}$. ∎

Las entradas de la matriz triangular inferior H en la factorización de Cholesky de una matriz hermítica definida positiva $A = (a_{ij}) \in \mathcal{M}_n(\mathbb{R})$ se pueden calcular

mediante el siguiente algoritmo: ponemos $h_{11} = \sqrt{a_{11}}$ y para $i = 2, \ldots, n$,

$$h_{ij} = \frac{1}{h_{jj}} \left(a_{ij} - \sum_{k=1}^{j-1} h_{ik} \overline{h_{jk}} \right), \ j = 1, \ldots, i-1,$$
$$h_{ii} = \left(a_{ii} - \sum_{k=1}^{i-1} |h_{ik}|^2 \right)^{1/2}.$$

El algoritmo anterior requiere del orden de $(n^3/3)$ operaciones (la mitad de las requeridas por la factorización LU). Además, notemos que debido a la "simetría" sólo hace falta almacenar la parte inferior de A y así H puede ser almacenada en la misma área. Además, se trata un algoritmo bastante estable respecto a la propagación de errores de redondeo tal y como se ilustrará en la práctica 12.

4. Matrices de Householder. El método de Householder

Existen versiones para los números complejos de las definiciones y resultados que veremos a continuación. Nosotros nos centraremos en el caso real, pero el lector interesado puede consultar [Stoer, J.; Bulirsch, R. *Introduction to numerical analysis.* Third edition. Texts in Applied Mathematics, 12. Springer-Verlag, New York, 2002], para el caso complejo.

Definición IX.4.1. Llamaremos **matriz de Householder** a una matriz de la forma

$$\mathcal{H}(\mathbf{w}) = I_n - 2\,\frac{\mathbf{w}\,\mathbf{w}^{\mathrm{t}}}{\mathbf{w}^{\mathrm{t}}\mathbf{w}},$$

siendo $\mathbf{w}$ un vector no nulo de $\mathbb{R}^n$.

Obsérvese que $\mathcal{H}(\mathbf{w}) = \mathcal{H}(\lambda\mathbf{w})$, para todo $\lambda \in \mathbb{R}$ no nulo. Por otra parte, si $\mathbf{w}$ tiene módulo 1, entonces la correspondiente matriz de Householder es

$$\mathcal{H}(\mathbf{w}) = I_n - 2\,\mathbf{w}\,\mathbf{w}^{\mathrm{t}}.$$

De aquí, que muchos autores adopten esta última expresión como definición de matriz de Householder.

Por convenio, supondremos que la matriz identidad es una matriz de Householder (más concretamente, la matriz de Householder para el vector cero), con el fin de simplificar algunos de los enunciados posteriores.

Las matrices de Householder son *simétricas y ortogonales*[6], luego, en particular, conservan el producto escalar usual de $\mathbb{R}^n$ (ejercicio 6), por eso son muy estables en su aplicación numérica.

Desde un punto de vista geométrico la matriz de Householder $\mathcal{H}(\mathbf{w})$ es la matriz de una simetría (o reflexión) respecto del hiperplano perpendicular a $\mathbf{w}$; su interés en Análisis Numérico Matricial proviene del siguiente resultado que nos permite elegir una simetría que alinea a un vector $\mathbf{v} \in \mathbb{R}^n$ dado con el vector $\mathbf{e}_1$ de la base canónica de $\mathbb{R}^n$.

Teorema IX.4.2. *Sea* $\mathbf{v} = (v_1, v_2, \dots, v_n) \in \mathbb{R}^n$ *tal que* $\sum_{i=2}^n v_i^2 > 0$. *Existe una matriz de Householder* H *tal que las últimas* $n-1$ *componentes del vector* $H\mathbf{v}$ *son nulas. Más concretamente, si* $\mathbf{w} = \mathbf{v} \pm \|\mathbf{v}\|_2 \mathbf{e}_1$ *y* $H = \mathcal{H}(\mathbf{w})$, *entonces*

$$H\mathbf{v} = \mp \|\mathbf{v}\|_2 \mathbf{e}_1,$$

donde $\mathbf{e}_1$ *denota el primer vector de la base canónica de* $\mathbb{R}^n$.

Demostración. En primer lugar observamos que la hipótesis $\sum_{i=2}^n v_i^2 > 0$ garantiza que los vectores $\mathbf{v} \pm \|\mathbf{v}\|_2 \mathbf{e}_1$ no son nulos (condición necesaria para poder definir las correspondientes matrices de Householder). Veamos ahora que las matrices de Householder propuestas verifican el resultado deseado:

$$\begin{aligned} H\mathbf{v} = \mathcal{H}(\mathbf{v} \pm \|\mathbf{v}\|_2 \mathbf{e}_1)\mathbf{v} &= \mathbf{v} - 2\,\frac{(\mathbf{v} \pm \|\mathbf{v}\|_2 \mathbf{e}_1)(\mathbf{v}^{\mathrm{t}} \pm \|\mathbf{v}\|_2 \mathbf{e}_1^{\mathrm{t}})}{(\mathbf{v}^{\mathrm{t}} \pm \|\mathbf{v}\|_2 \mathbf{e}_1^{\mathrm{t}})(\mathbf{v} \pm \|\mathbf{v}\|_2 \mathbf{e}_1)}\,\mathbf{v} \\ &= \mathbf{v} - 2\,\frac{(\mathbf{v} \pm \|\mathbf{v}\|_2 \mathbf{e}_1)(\mathbf{v}^{\mathrm{t}} \pm \|\mathbf{v}\|_2 \mathbf{e}_1^{\mathrm{t}})\mathbf{v}}{(\mathbf{v}^{\mathrm{t}} \pm \|\mathbf{v}\|_2 \mathbf{e}_1^{\mathrm{t}})(\mathbf{v} \pm \|\mathbf{v}\|_2 \mathbf{e}_1)} \\ &= \mathbf{v} - 2\,\frac{\|\mathbf{v}\|_2(\|\mathbf{v}\|_2 \pm v_1)(\mathbf{v} \pm \|\mathbf{v}\|_2 \mathbf{e}_1)}{2\|\mathbf{v}\|_2(\|\mathbf{v}\|_2 \pm v_1)} \\ &= \mathbf{v} - (\mathbf{v} \pm \|\mathbf{v}\|_2 \mathbf{e}_1) \\ &= \mp \|\mathbf{v}\|_2 \mathbf{e}_1 \end{aligned}$$

■

El vector $\mathbf{w} = \mathbf{v} \pm \|\mathbf{v}\|_2 \mathbf{e}_1$ se dice que es un **vector de Householder** de $\mathbf{v}$.

[6] En efecto, sea $\mathbf{w}$ un vector de $\mathbb{R}^n$ de módulo 1. Entonces

$$\mathcal{H}(\mathbf{w})^{\mathrm{t}} = \left(I_n - 2\mathbf{w}\mathbf{w}^{\mathrm{t}}\right)^{\mathrm{t}} = I_n - 2(\mathbf{w}\mathbf{w}^{\mathrm{t}})^{\mathrm{t}} = I_n - 2\mathbf{w}\mathbf{w}^{\mathrm{t}} = \mathcal{H}(\mathbf{w}),$$

es decir, $H(\mathbf{w})$ es simétrica; por otra parte,

$$\begin{aligned} \mathcal{H}(\mathbf{w})\mathcal{H}(\mathbf{w})^{\mathrm{t}} = \mathcal{H}(\mathbf{w})^2 &= \left(I_n - 2\,\mathbf{w}\,\mathbf{w}^{\mathrm{t}}\right)^2 = I_n - 4\mathbf{w}\,\mathbf{w}^{\mathrm{t}} + 4(\mathbf{w}\,\mathbf{w}^{\mathrm{t}})^2 \\ &= I_n - 4\mathbf{w}\,\mathbf{w}^{\mathrm{t}} + 4(\mathbf{w}\,\mathbf{w}^{\mathrm{t}})(\mathbf{w}\,\mathbf{w}^{\mathrm{t}}) = I_n - 4\mathbf{w}\,\mathbf{w}^{\mathrm{t}} + 4\mathbf{w}\,(\mathbf{w}^{\mathrm{t}}\,\mathbf{w})\mathbf{w}^{\mathrm{t}} \\ &= I_n - 4\mathbf{w}\,\mathbf{w}^{\mathrm{t}} + 4\mathbf{w}\,\mathbf{w}^{\mathrm{t}} = I_n, \end{aligned}$$

esto es, $\mathcal{H}(\mathbf{w})$ es ortogonal.

Nota IX.4.3. Si $\sum_{i=2}^{n} v_i^2 = 0$; entonces

$$\begin{aligned} I_n\mathbf{v} &= \|\mathbf{v}\|_2\mathbf{e}_1 \quad \text{si} \quad v_1 \geq 0; \\ \mathcal{H}(\mathbf{v} - \|\mathbf{v}\|_2\mathbf{e}_1)\mathbf{v} &= \|\mathbf{v}\|_2\mathbf{e}_1 \quad \text{si} \quad v_1 < 0. \end{aligned}$$

De tal forma que podemos concluir que el teorema 1 es cierto *en todos los casos*, y además, que *la primera componente del vector* $H\mathbf{v}$ *siempre se puede tomar no negativa.*

En la práctica, procedemos de la siguiente forma: calculamos la norma de $\mathbf{v}$ para el producto escalar usual de $\mathbb{R}^n$, $\|\mathbf{v}\|_2$, después hallamos el vector $\mathbf{w} = \mathbf{v} \pm \|\mathbf{v}\|_2\mathbf{e}_1$, y luego el número

$$\beta := \frac{\mathbf{w}^{\mathrm{t}}\mathbf{w}}{2} = \|\mathbf{v}\|_2(\|\mathbf{v}\|_2 \pm v_1),$$

esto es, el módulo de $\mathbf{w}$ al cuadrado divido por dos.

Para la elección del signo (que precede a $\|\mathbf{v}\|_2\mathbf{e}_1$) nos guiamos por la presencia de la expresión $(\mathbf{w}^{\mathrm{t}}\mathbf{w})$ en el denominador de la matriz de Householder: para evitar divisiones por números demasiado "pequeños" (lo que puede tener consecuencias desastrosas en la propagación de errores de redondeo), elegimos $\mathbf{w} = \mathbf{v} + \|\mathbf{v}\|_2\mathbf{e}_1$, si $v_1 \geq 0$ y $\mathbf{w} = \mathbf{v} - \|\mathbf{v}\|_2\mathbf{e}_1$, si $v_1 < 0$.

Siguiendo con la notación anterior, sea $H = \mathcal{H}(\mathbf{w})$ con $\mathbf{w} \neq 0$ (en otro caso, tómese $H = I_n$). Si $\mathbf{a}$ es un vector de $\mathbb{R}^n$, el cálculo del vector $H\mathbf{a}$ se efectúa hallando primero el producto escalar $\alpha := \mathbf{w}^{\mathrm{t}}\mathbf{a}$, y a continuación el vector

$$\begin{aligned} H\mathbf{a} = \mathbf{a} - 2\,\frac{\mathbf{w}\mathbf{w}^{\mathrm{t}}}{\mathbf{w}^{\mathrm{t}}\,\mathbf{w}}\,\mathbf{a} = \mathbf{a} - \frac{(\mathbf{w}\mathbf{w}^{\mathrm{t}})\mathbf{a}}{\beta} = \mathbf{a} - \frac{\mathbf{w}(\mathbf{w}^{\mathrm{t}}\mathbf{a})}{\beta} = \mathbf{a} - \frac{\alpha\mathbf{w}}{\beta} \\ = \mathbf{a} - \frac{\alpha}{\beta}\,\mathbf{w}. \end{aligned}$$

Nótese que si $\alpha = 0$, entonces $\mathbf{a}$ pertenece al hiperplano perpendicular a $\mathbf{w}$, por lo que $H\mathbf{a} = \mathbf{a}$.

El método de Householder.

Sea $A \in \mathcal{M}_n(\mathbb{R})$. El método de Householder consiste en encontrar $n-1$ matrices de Householder, $H_1, \ldots, H_{n-1}$, tales que la matriz

$$H_{n-1} \cdots H_2 H_1 A$$

sea triangular superior.

Si denotamos $A_1 = A$, cada matriz $A_k = H_{k-1} \cdots H_2 H_1 A,\ k \geq 1$, es de la forma

$$A_k = (a_{ij})^{(k)} = \begin{pmatrix} \times & \times & \times & \times & \times & \times & \times & \times \\ & \times & \times & \times & \times & \times & \times & \times \\ & & \times & \times & \times & \times & \times & \times \\ & & & \times & \times & \times & \times & \times \\ & & & \times & \times & \times & \times & \times \\ & & & \times & \times & \times & \times & \times \\ & & & \times & \times & \times & \times & \times \\ & & & \times & \times & \times & \times & \times \end{pmatrix} \leftarrow \text{fila } k\text{-ésima}$$

$\mathbf{v}^{(k)}$

$\uparrow$

columna k-ésima

Nota IX.4.4. La distribución de los ceros en la matriz A_k es la misma que la que se obtiene en la etapa $(k-1)$-ésima del método de Gauss. Sin embargo, el paso de la matriz A_k a la matriz A_{k+1} es completamente diferente; por ejemplo, los elementos de la fila k-ésima se ven modificados, a diferencia de lo que ocurre con el método de Gauss.

Designemos por $\mathbf{v}^{(k)}$ al vector de $\mathbb{R}^{n-k+1}$ cuyas componentes son los elementos $a_{ik}^{(k)},\ k \leq i \leq n$, de la matriz $A_k = (a_{ij})^{(k)}$. Si $\sum_{i=k+1}^{n} (a_{ik}^{(k)})^2 > 0$, por el teorema 1, existe un vector $\tilde{\mathbf{w}}^{(k)} \in \mathbb{R}^{n-k+1}$ tal que el vector $\mathcal{H}(\tilde{\mathbf{w}}^{(k)})\mathbf{v}^{(k)} \in \mathbb{R}^{n-k+1}$ tiene todas sus componentes nulas excepto la primera.

Sea $\mathbf{w}^{(k)}$ el vector de $\mathbb{R}^n$ tal que sus primeras $(k-1)$ componentes son nulas y las $(n-k+1)$ restantes son las del vector $\tilde{\mathbf{w}}^{(k)}$. Bajo estas condiciones, las matriz

$$H_k = \begin{pmatrix} I_{k-1} & 0 \\ 0 & \mathcal{H}(\tilde{\mathbf{w}}^{(k)})\mathbf{v}^{(k)} \end{pmatrix}$$

es la matriz de Householder $\mathcal{H}(\mathbf{w}^{(k)})$ y se cumple que $A_{k+1} = H_k A_k$.

Naturalmente, si $\sum_{i=k+1}^{n} (a_{ik}^{(k)})^2 = 0$, es decir, si $a_{ik}^{(k)} = 0$, para todo $i = k+1, \ldots, n$, la matriz A_k ya tiene la forma deseada por lo que podemos tomar $A_{k+1} = A_k$ y $H_k = I_n$ (véase la nota 1 para más detalle).

Factorización QR.

La interpretación matricial del método de Householder nos conduce a un resultado tremendamente importante sobre factorización de matrices (cuadradas). Un primera versión del siguiente resultado ya apareció en el tema V como consecuencia del método de ortonormalización de Gram-Schmidt (véase el corolario V.3.11).

Teorema IX.4.5. *Sea $A \in \mathcal{M}_n(\mathbb{R})$. Existen una matriz ortogonal Q, producto de matrices de Householder, y una matriz triangular superior R tales que*

$$A = QR.$$

Además, los elementos de R se pueden elegir no negativos; en cuyo caso, si A es invertible, la ***factorización*** *QR es única.*

Demostración. En primer lugar, observamos que la existencia de las matrices de Householder $H_1, H_2, \ldots, H_{n-1}$ es independiente de que A sea invertible[7], por lo que toda matriz $A \in \mathcal{M}_n(\mathbb{R})$ se puede escribir de la forma

$$A = (H_{n-1} \cdots H_2 H_1)^{-1} A_n,$$

tal que la matriz $R := A_n$ sea triangular superior. La matriz $Q := (H_{n-1} \cdots H_2 H_1)^{-1} = H_1 H_2 \cdots H_{n-1}$ es ortogonal (recuérdese que las matrices de Householder cumplen que $H_k^{-1} = H_k^{\mathrm{t}} = H_k$). Luego, la existencia de una descomposición QR ya está demostrada.

El hecho de se puedan elegir los primeros $n-1$ elementos de la diagonal principal de $R = (r_{ij}) \in \mathcal{M}_n(\mathbb{R})$ no negativos es consecuencia del teorema 1 y de la nota 1. Si el elemento $r_{nn} = a_{nn}^{(n)}$ fuese negativo, basta tomar la siguiente matriz de Householder

$$H_n = \mathcal{H}(\mathbf{w}^{(n)}) \quad \text{con} \quad \mathbf{w}^{(n)} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ a_{nn}^{(n)} - |a_{nn}^{(n)}| \end{pmatrix}.$$

Si la matriz A es invertible, al menos, existe una factorización $A = QR$ tal que $r_{ii} > 0$, para todo $i = 1, \ldots, n$. Demostremos, pues, la unicidad de tal descomposición. De las igualdades

$$A = Q_1 R_1 = Q_2 R_2,$$

se deduce que

$$Q_2^{\mathrm{t}} Q_1 = R_2 R_1^{-1} =: B,$$

en particular B es una matriz triangular superior por ser producto de matrices triangular superiores. Por otra parte,

$$B^{\mathrm{t}} B = Q_1^{\mathrm{t}} Q_2 Q_2^{\mathrm{t}} Q_1 = I_n,$$

de donde se sigue que B ha de ser diagonal; ya que $B^{\mathrm{t}} = B^{-1}$ es triangular inferior, pero la inversa de una matriz triangular superior es triangular superior. Además, cómo

$$(B^{\mathrm{t}})_{ii} \cdot (B)_{ii} = 1, \ i = 1, \ldots, n,$$

[7] ¡De hecho tampoco depende de que A sea cuadrada!

y

$$(B^{\mathrm{t}})_{ii} = (B)_{ii} = \frac{(R_2)_{ii}}{(R_1)_{ii}} > 0,\ i = 1, \ldots, n,$$

concluimos que $(B)_{ii} = 1$, para todo $i = 1, \ldots, n$, y por consiguiente que $B = I_n$. Luego, $R_1 = R_2$ y $Q_1 = Q_2$. ∎

La factorización QR también se puede llevar a cabo en matrices no necesariamente cuadradas.

Corolario IX.4.6. *Sea $A \in \mathcal{M}_{m\times n}(\mathbb{R})$, con $m \geq n$. Existen una matriz ortogonal $Q \in \mathcal{M}_m(\mathbb{R})$ y una matriz $R \in \mathcal{M}_{m\times n}(\mathbb{R})$ con sus n primeras filas formando una matriz triangular superior y las $m - n$ últimas nulas tales que $A = QR$.*

Demostración. Si $A' = (A|\mathbf{0}_{m\times(m-n)}) \in \mathcal{M}_m(\mathbb{R})$ y $A' = QR'$ es su factorización QR, entonces $A = QR$ donde R es la matriz de orden $m \times n$ formada por las n primeras columnas de R'. ∎

El número de operaciones necesarias para llevar a cabo la factorización QR de una matriz de orden $m \times n,\ m \geq n$ es del orden de $2mn^2$. La implementación del algoritmo para hallar la factorización QR de una matriz cuadrada que se deduce de la demostración del teorema IX.4.5 se verá en la práctica 12.

Al igual que la factorización LU, la descomposición QR se utiliza para resolver sistemas de ecuaciones lineales $A\mathbf{x} = \mathbf{b}$.

- Calcula la factorización QR de A.
- Calcula $\mathbf{c} = Q^{\mathrm{t}}\mathbf{b}$.
- Resuelve el sistema triangular $R\mathbf{x} = \mathbf{c}$, por ejemplo, mediante sustitución hacia atrás.

Para terminar indicamos una interpretación muy importante de la factorización QR de una matriz invertible A. Si $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$ y $\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_n$ son los vectores columna de la matrices A y Q respectivamente, la relación $A = QR$ se escribe de la siguiente manera

$$\begin{cases} \mathbf{a}_1 &= r_{11}\mathbf{q}_1; \\ \mathbf{a}_2 &= r_{12}\mathbf{q}_1 + r_{22}\mathbf{q}_2; \\ &\vdots \\ \mathbf{a}_n &= r_{1n}\mathbf{q}_1 + r_{2n}\mathbf{q}_2 + \ldots + r_{nn}\mathbf{q}_n, \end{cases}$$

donde $R = (r_{ij}) \in \mathcal{M}_n(\mathbb{R})$. Ahora bien, como los vectores $\mathbf{q}_i$ forman un sistema ortogonal (pues son las columnas de una matriz ortogonal), las relaciones anteriores equivalen a un *proceso de ortonormalización de Gram-Schmidt.*

Ejercicios del tema IX

Ejercicio 1. Demostrar el lema IX.1.2.

Ejercicio 2. Sea $A = (a_{ij}) \in \mathcal{M}_n(\mathbb{R})$ tal que $a_{ij} = 1$ si $i = j$ o $j = n$, $a_{ij} = -1$ si $i > j$ y cero en otro caso. Probar que A admite factorización LU con $|l_{ij}| \leq 1$ y $u_{nn} = 2^{n-1}$.

Ejercicio 3. Sea

$$A_\epsilon = \begin{pmatrix} 1 & 1-\epsilon & 3 \\ 2 & 2 & 2 \\ 3 & 6 & 4 \end{pmatrix}.$$

Halalr para qué valores de ϵ no se satisfacen las hipótesis del teorema IX.1.4. ¿Para qué valores de ϵ esta matriz no es invertible? ¿Es posible calcular factorización LU en este caso?

Ejercicio 4. Verificar que el número de operaciones necesarias para calcular la factorización LU de una matriz cuadrada de orden n es aproximadamente $2n^3/3$.

Ejercicio 5. Sean $l_{ij} \in \Bbbk$, $1 \leq j < i \leq n$ y $L_k = I_n - \vec{\ell}_k \mathbf{e}_k^{\mathrm{t}}$ donde

$$\vec{\ell}_k = (0, \ldots, 0, l_{k+1,k}, \ldots, l_{n,k})^{\mathrm{t}} \in \Bbbk^n$$

y $\mathbf{e}_k$ es el vector k-ésimo de la base usual de $\Bbbk^n$, $k = 1, \ldots, n-1$. Probar que

1. Si $T_{ij} \in \mathcal{M}_n(\Bbbk)$ es la matriz elemental de tipo I que intercambia las filas i y j, entonces $T_{ij} L_k T_{ij} = L'_k$, donde $L'_k = I_n - \vec{\ell}'_k \mathbf{e}_k^{\mathrm{t}}$ siendo $\vec{\ell}'_k$ el vector $\vec{\ell}_k$ al que se le han intercambiado las coordenadas i y j.
2. Si $P \in \mathcal{M}_n(\Bbbk)$ es una matriz de permutación (es decir, producto de matrices elementales de tipo I), entonces $P L_k P^{-1} = L'_k$, donde $L'_k = I_n - \vec{\ell}'_k \mathbf{e}_k^{\mathrm{t}}$ siendo $\vec{\ell}'_k$ el vector $\vec{\ell}_k$ al que se le han intercambiado las coordenadas según la permutación definida por P.

3. Si $P_1, \ldots, P_{n-1} \in \mathcal{M}_n(\Bbbk)$ son matrices de permutación, $P = P_{n-1} \cdots P_1$ y $M = L_{n-1}P_{n-1} \cdots L_2P_2L_1P_1$, entonces

$$\begin{aligned} MP^{-1} &= L_{n-1}P_{n-1} \cdots L_2P_2L_1P_2^{-1}P_3^{-1} \cdots P_{n-1}^{-1} \\ &= L_{n-1}P_{n-1} \cdots L_2P_2L_1P_2^{-1}P_3^{-1} \cdots P_{n-1}^{-1} \\ &= L_{n-1}P_{n-1} \cdots P_3L_2L_1'P_3^{-1} \cdots P_{n-1}^{-1} \\ &= L_{n-1}P_{n-1} \cdots P_3L_2P_3^{-1}P_3L_1'P_3^{-1} \cdots P_{n-1}^{-1} \\ &= L_{n-1}P_{n-1} \cdots P_3L_2P_3^{-1}L_1''P_4^{-1} \cdots P_{n-1}^{-1} \\ &= L_{n-1}P_{n-1} \cdots P_3L_2P_3^{-1}P_4^{-1} \cdots P_{n-1}^{-1}L_1^{(n-2)} \\ &= \ldots \\ &= L_{n-1}L_{n-2}' \cdots L_2^{(n-3)}L_1^{(n-2)}. \end{aligned}$$

De donde se sigue que MP^{-1} y PM^{-1} son triangulares inferiores con unos en su diagonal principal.

Ejercicio 6. Sean $\mathbf{w} \in \mathbb{R}^n$ de módulo 1 y $\mathcal{H}(\mathbf{w}) = I_n - 2\mathbf{w}\mathbf{w}^{\mathrm{t}}$ la correspondiente matriz de Householder. Probar que dados $\mathbf{u}$ y $\mathbf{v} \in \mathbb{R}^n$, se cumple que

$$(\mathcal{H}(\mathbf{w})\mathbf{u})^{\mathrm{t}}(\mathcal{H}(\mathbf{w})\mathbf{v}) = \mathbf{u}^{\mathrm{t}}\mathbf{v}.$$

TEMA X

Métodos iterativos de resolución de sistemas lineales de ecuaciones

En este damos una breve introducción a los métodos iterativos para la resolución de sistemas lineales, mostrando aquellos métodos que tienen un comportamiento asintótico relativamente "ejemplar". Los métodos iterativos que consideraremos en este tema serán de la forma

$$\mathbf{u}^{(k+1)} = B\mathbf{u}^{(k)} + \mathbf{c}, \quad k \geq 1,$$

siendo el valor inicial $\mathbf{u}^{(0)}$ arbitrario, y tal que la matriz B y el vector $\mathbf{c}$ se construyen a partir de un sistema $A\mathbf{x} = \mathbf{b}$. Tal es el comienzo de la primera sección de este tema, donde exponemos la idea general sobre los métodos iterativos y estudiamos condiciones necesarias y suficientes para que la sucesión de vectores $(\mathbf{u}^{(k)})_{k\in\mathbb{N}}$ converja a la solución del sistema $A\mathbf{x} = \mathbf{b}$. Aquí son fundamentales el estudio espectral de la matriz de B y los resultados sobre convergencia de las potencias de una matriz estudiados en el tema VIII.

En la segunda sección mostramos un método muy general para construir métodos iterativos que consiste en descomponer la matriz A del sistema en la forma $A = M - N$ con M invertible, y tomar $B = M^{-1}N$. La matriz M se llama precondionador del método, y su elección será crucial para garantizar la convergencia. A continuación en la siguientes secciones mostramos distintos métodos iterativos derivados de distintas elecciones de M. En la tercera sección, se estudian los métodos de Jacobi, Gauss-Seidel y de relajación (método SOR), estos tres métodos parten de la idea común de descomponer la matriz A como la suma matriz diagonal D, una triangular inferior $-E$ y otra triangular superior $-F$, y a continuación considerar distintas combinaciones en esta descomposición para elección de M; así si, por ejemplo, tomamos $D = M$, se consigue el llamado método de Jacobi. En esta sección mostramos algunos resultados sobre la convergencia de estos métodos y exploramos algunos resultados que nos permiten su comparación, para familias de matrices espaciales (esencialmente, para las matrices hermíticas definidas positivas y las matrices tridiagonales). Al final de la sección consideramos el problema la condición de parada de un método iterativo para dar una buena aproximación de la solución del sistema.

En la última sección del tema, damos un pequeño paso más allá y estudiamos la generalización de los métodos anteriores. Tal generalización se conoce como método de Richardson, cuya aportación principal, en forma general, es la introducción un determinado parámetro que se irá actualizando en cada iteración. Casos particulares de este método, no contemplados en las sección anterior, son el método del gradiente y del gradiente conjugado. Nosotros solamente nos ocuparemos de estudiar el primero con detalle, mostrando resultados sobre su convergencia y precisión.

Para la elaboración de este tema hemos seguido el capítulo 4 de [**QSS07**] y el capítulo 5 de [**Cia82**]. También hemos usado [**QS06**], tangencialmente. En [**Cia82**] se da una introducción general a los métodos iterativos de Jacobi, Gauss-Seidel y de relajación. En [**QSS07**] se muestran éstos métodos, además de los de Richardson y otras variantes de éste (distintas del método del gradiente) de las que no nos ocuparemos en esta asignatura.

1. Sobre la convergencia de los métodos iterativos

Usaremos la siguiente notación en todo el tema $V = \mathbb{k}^n$, $A \in \mathcal{M}_n(\mathbb{k})$ invertible y $\mathbf{b} \in V$ no nulo.

Para entender en qué consisten los *métodos iterativos* para resolución de sistemas lineales, supongamos que, dado un sistema lineal $A\mathbf{x} = \mathbf{b}$, encontramos una matriz $B \in \mathcal{M}_n(\mathbb{k})$ y un vector $\mathbf{c} \in V$ tal que

- la matriz $I - B$ es invertible
- la única solución[1] del sistema lineal $\mathbf{x} = \mathbf{B}\mathbf{x} + \mathbf{c}$ es la solución de $A\mathbf{x} = \mathbf{b}$.

La forma del sistema $\mathbf{x} = B\mathbf{x} + \mathbf{c}$ sugiere abordar la resolución del sistema lineal $A\mathbf{x} = \mathbf{b}$ mediante un *método iterativo* asociado a la matriz B del siguiente modo: dado un vector inicial $\mathbf{u}^{(0)} \in V$ arbitrario, se construye la sucesión de vectores $(\mathbf{u}^{(k)})_{k\in\mathbb{N}}$ de V dada por

$$\mathbf{u}^{(k+1)} = B\mathbf{u}^{(k)} + \mathbf{c} \tag{X.1.1}$$

para $k \in \mathbb{N} \cup \{0\}$, con la esperanza de que converja a la solución del sistema lineal.

Definición X.1.1. El método iterativo dado por la expresión (X.1.1) es **convergente** si existe $\mathbf{u} \in V$ tal que

$$\lim_{m\to\infty} \mathbf{u}^{(k)} = \mathbf{u}$$

para cualquier vector inicial $\mathbf{u}^{(0)} \in V$. Nótese que, en tal caso, este vector $\mathbf{u}$ verifica $\mathbf{u} = \mathbf{B}\mathbf{u} + \mathbf{c}$ ó, equivalentemente, $A\mathbf{u} = \mathbf{b}$.

[1] Nótese que la condición $I - B$ invertible garantiza que la solución del sistema $\mathbf{x} = \mathbf{B}\mathbf{x} + \mathbf{c}$ existe y es única.

En otra palabras, un método iterativo consiste en construir una sucesión de vectores $(\mathbf{u}^{(k)})_{k\in\mathbb{N}}$ de V (mediante la expresión (X.1.1), por ejemplo) que converja a la solución exacta. Por esta razón B se llama **matriz de la iteración** asociada al sistema lineal $A\mathbf{x} = \mathbf{b}$.

Por otra parte, si para cada $k \in \mathbb{N} \cup \{0\}$ denotamos el vector de errores cometido en cada iteración por

$$\vec{\varepsilon}_k := \mathbf{u}^{(k)} - \mathbf{u}$$

se verifica que

$$\vec{\varepsilon}_k = \mathbf{u}^{(k)} - \mathbf{u} = (B\mathbf{u}^{(k-1)} + \mathbf{c}) - (B\mathbf{u} + \mathbf{c}) = B(\mathbf{u}^{(k-1)} - \mathbf{u}) = B\vec{\varepsilon}_{k-1}$$

y, por tanto,

$$\vec{\varepsilon}_k = B\vec{\varepsilon}_{k-1} = B^2\vec{\varepsilon}_{k-2} = \ldots = B^k\vec{\varepsilon}_0. \tag{X.1.2}$$

Además, si $\vec{\varepsilon}_0$ fuese de norma 1, entonces

$$\|\vec{\varepsilon}_k\| = \|B^k\vec{\varepsilon}_0\| \leq |\!|\!| B^k |\!|\!| \leq |\!|\!| B |\!|\!|^k,$$

para la norma matricial $|\!|\!| \cdot |\!|\!|$ subordinada a una norma vectorial $\|\cdot\|$ cualquiera.

Así pues, el error en la iteraciones depende, en esencia, de la matriz B. Obsérvese que el resultado siguiente, que da un criterio fundamental de convergencia de los métodos iterativos, sólo involucra la matriz de iteración B considerada.

Criterios de convergencia para métodos iterativos. *Sea $B \in \mathcal{M}_n(\Bbbk)$. Son equivalentes:*

a) *El método iterativo asociado a la matriz B es convergente.*
b) $\varrho(B) < 1$.
c) *Existe una norma matricial $|\!|\!| \cdot |\!|\!|$ (que se puede tomar subordinada) tal que $|\!|\!| B |\!|\!| < 1$*

Demostración. A partir del teorema VIII.2.19 y de la relación (X.1.2), se tienen las equivalencias:

$$\begin{aligned}
\text{El método es convergente} &\iff \lim_{m\to\infty} \vec{\varepsilon}_k = \mathbf{0}, \text{ para todo } \vec{\varepsilon}_0 \in V \\
&\iff \lim_{m\to\infty} B^k\vec{\varepsilon}_0 = \mathbf{0}, \text{ para todo } \vec{\varepsilon}_0 \in V \\
&\iff \varrho(B) < 1 \\
&\iff |\!|\!| B |\!|\!| < 1 \text{ para una norma matricial } |\!|\!| \cdot |\!|\!|.
\end{aligned}$$

■

Se plantea la cuestión de cómo elegir entre diversos métodos iterativos convergentes para la resolución de un mismo sistema lineal $A\mathbf{x} = \mathbf{b}$. En esta línea, se tiene la siguiente:

Proposición X.1.2. *Sean $\|\cdot\|$ una norma sobre V y $\mathbf{u} \in V$ tal que $\mathbf{u} = B\mathbf{u} + \mathbf{c}$. Para el método iterativo*

$$\begin{cases} \mathbf{u}^{(0)} \in V \text{ arbitrario} \\ \mathbf{u}^{(k+1)} = B\mathbf{u}^{(k)} + \mathbf{c}, \ k \in \mathbb{N} \cup \{0\}. \end{cases}$$

se verifica que

$$\lim_{k\to+\infty} \left(\sup_{\|\vec{\varepsilon}_0\|=1} \|\vec{\varepsilon}_k\|^{1/m} \right) = \rho(B)$$

donde $\vec{\varepsilon}_k$ está definido en (X.1.2).

Demostración. En el teorema VIII.2.20 vimos que $\lim_{k\to+\infty} |||B^k|||^{1/m} = \rho(B)$. Luego, basta tener en cuenta que por (X.1.2) se tiene que

$$|||B^k|||^{1/m} = \sup_{\|\vec{\varepsilon}_0\|=1} \|B^k\vec{\varepsilon}_0\| = \sup_{\|\vec{\varepsilon}_0\|=1} \|\vec{\varepsilon}_k\|.$$

■

Este último resultado afirma que $\sup_{\|\mathbf{u}^{(0)}-\mathbf{u}\|=1} \|\mathbf{u}^{(k)} - \mathbf{u}\|$ tiene el mismo comportamiento asintótico que $\varrho(B)^k$. Por tanto, en el caso de que el método iterativo converja, la convergencia de la sucesión $(\mathbf{u}^{(k)})_{k\in\mathbb{N}}$ será igual de rápida que la convergencia a cero de la sucesión de número reales $(\varrho(B)^k)_{k\in\mathbb{N}}$ y, por consiguiente, tanto más rápida cuanto menor sea el radio espectral de matriz B que define el método.

A la hora de resolver un sistema lineal mediante un método iterativo deberemos, en primer lugar, asegurar su convergencia (por ejemplo, encontrando alguna norma para la cual $|||B||| < 1$ o viendo que $\varrho(B) < 1$). Para luego, en caso de disponer de varios a nuestro alcance, elegir aquel cuyo radio espectral sea menor (véase el teorema VIII.2.16). En resumen, *para un método iterativo de la forma (X.1.1) cuya matriz de iteración satisface las dos condiciones del principio, se verifica que la convergencia para cualquier $\mathbf{u}^{(0)}$ si, y sólo si, $\varrho(B) < 1$*. Además, como consecuencia del teorema VIII.2.16, cuando más pequeño sea $\varrho(B)$, menor será el número de iteraciones necesario para reducir el error inicial.

2. Cómo construir métodos iterativos

La estrategia que se va a utilizar para construir métodos iterativos consistirá en descomponer la matriz A en la forma

$$\boxed{A = M - N}$$

donde M va a ser una matriz invertible tal que su matriz inversa sea fácil de calcular (en el sentido de que sea fácil de resolver el sistema asociado $MX = I_n$ como ocurre, por ejemplo, cuando M es una matriz diagonal, diagonal por bloques, triangular

o triangular por bloques, hermítica o simétrica definida positiva, ...). Con esta descomposición se verifica que:

$$A\mathbf{u} = \mathbf{b} \iff (M-N)\mathbf{u} = \mathbf{b} \iff M\mathbf{u} = N\mathbf{u} + \mathbf{b} \iff \mathbf{u} = B\mathbf{u} + \mathbf{c}$$

donde

$$\boxed{B = M^{-1}N = I_n - M^{-1}A} \quad \text{y} \quad \boxed{\mathbf{c} = M^{-1}\mathbf{b}}$$

De esta forma podemos considerar el método iterativo

$$\text{(X.2.3)} \qquad \begin{cases} \mathbf{u}^{(0)} \in V \text{ arbitrario} \\ \mathbf{u}^{(k+1)} = B\mathbf{u}^{(k)} + \mathbf{c}, \ k \in \mathbb{N} \cup \{0\}. \end{cases}$$

Como $N = M - A$, entonces $B = M^{-1}N = M^{-1}(M - A) = I - M^{-1}A$. Así,

$$I - B = M^{-1}A$$

es una matriz invertible, por lo que el sistema $(I - B)\mathbf{x} = \mathbf{c}$ tiene solución única. En la práctica, para calcular $\mathbf{u}^{(k+1)}$, se resolverá el sistema

$$M\mathbf{u}^{(k+1)} = N\mathbf{u}^{(k)} + \mathbf{b}$$

en vez de trabajar directamente con (X.2.3). Es por esto por lo que requerimos que M sea una matriz cuya matriz inversa sea fácil de calcular. La matriz M se suele llamar **precondicionador** de A.

Nota X.2.1. Como ya se ha comentado, todos los métodos iterativos que vamos a estudiar responden a una descomposición $M - N$ de la matriz A. Intuitivamente, cuanto más de A haya en M, tanto más se parecerá cada iteración al cálculo de la solución exacta (de hecho, en el caso límite $M = A$ la solución se obtiene en la primera iteración). No obstante, esto va en contra de la idea inicial de que el coste de cada iteración sea bajo. Un método iterativo será aquel que mantenga un equilibrio entre estas dos estrategias enfrentadas.

3. Métodos de Jacobi, Gauss-Seidel y relajación

En esta sección vamos a introducir tres de los métodos iterativos más usuales para la resolución de un sistema lineal $A\mathbf{x} = \mathbf{b}$. Todos ellos comparten una idea básica en su construcción: separar la matriz del sistema en suma de dos matrices.

A continuación describiremos una determinada descomposición de A que será la que usaremos como base en los diversos métodos iterativos que vamos a estudiar en esta sección.

Notación X.3.1. Dada una matriz $A = (a_{ij}) \in \mathcal{M}_n(\Bbbk)$ invertible con

$$\text{(X.3.4)} \qquad a_{ii} \neq 0$$

para $i = 1, 2, \ldots, n$, consideramos la siguiente descomposición de la matriz

$$A = \begin{pmatrix} & & -F \\ & D & \\ -E & & \end{pmatrix}$$

que podemos escribir en la forma

$$\boxed{A = D - E - F}$$

donde

$$D = \operatorname{diag}(a_{11}, a_{22}, \ldots, a_{nn}), \quad E = (e_{ij}) \in \mathcal{M}_n(\Bbbk), \quad \text{y} \quad F = (f_{ij}) \in \mathcal{M}_n(\Bbbk)$$

siendo

$$e_{ij} = \begin{cases} -a_{ij} & \text{si} \quad i > j \\ 0 & \text{si} \quad i \leq j \end{cases} \quad \text{y} \quad f_{ij} = \begin{cases} -a_{ij} & \text{si} \quad i < j \\ 0 & \text{si} \quad i \geq j \end{cases}$$

A esta descomposición de A la denominaremos **descomposición $D - E - F$ (por puntos) de la matriz A.**

Ejemplo X.3.2. Consideremos la matriz

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

donde $\epsilon \in \mathbb{R}$. Claramente, $A = D - E - F$ siendo

$$D = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 2 \end{pmatrix}, \quad E = \begin{pmatrix} 0 & 0 & 0 \\ -2 & 0 & 0 \\ -\epsilon & 0 & 0 \end{pmatrix} \quad \text{y} \quad F = \begin{pmatrix} 0 & 2 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}.$$

De forma análoga se podrían considerar descomposiciones $D - E - F$ de A por bloques; en este caso, las matrices D, E y F se eligen, respectivamente, diagonal, triangular inferior y triangular superior por bloques de ordenes apropiados para que sea $A = D - E - F$. Nosotros sólo nos ocuparemos de las descomposiciones por bloques de orden 1, es decir, descomposiciones por puntos. El lector interesado puede encontrar la versión por bloques de los métodos iterativos que estudiaremos a continuación en el apartado 5.3.4 de [**IR99**].

Método de Jacobi.

Consiste en tomar

$$\boxed{M = D} \quad \text{y} \quad \boxed{N = E + F}$$

Así pues,

$$A\mathbf{u} = \mathbf{b} \Longleftrightarrow D\mathbf{u} = (E+F)\mathbf{u} + \mathbf{b} \Longleftrightarrow \mathbf{u} = D^{-1}(E+F)\mathbf{u} + D^{-1}\mathbf{b}$$

que conduce al llamado **método iterativo de Jacobi** o método JOR (*Jacobi Over-Relaxation method*)

$$\begin{cases} \mathbf{u}^{(0)} \in V \quad \text{arbitrario} \\ \mathbf{u}^{(k+1)} = D^{-1}(E+F)\mathbf{u}^{(k)} + D^{-1}\mathbf{b}, \ k \in \mathbb{N} \cup \{0\} \end{cases}$$

o, equivalentemente,

$$\text{(X.3.5)} \qquad \begin{cases} \mathbf{u}^{(0)} \in V \quad \text{arbitrario} \\ D\mathbf{u}^{(k+1)} = (E+F)\mathbf{u}^{(k)} + \mathbf{b}, \ k \in \mathbb{N} \cup \{0\} \end{cases}$$

Nótese que la hipótesis (X.3.4) determina que la matriz $M = D$ es invertible. La matriz de este método es

$$\boxed{\mathcal{J} = D^{-1}(E+F) = I - D^{-1}A}$$

que se denomina **matriz de Jacobi**. La iteración definida en (X.3.5) puede escribirse, coordenada a coordenada, como

$$\begin{aligned} a_{ii}(\mathbf{u}^{(k+1)})_i &= b_i - a_{i1}\,(\mathbf{u}^{(k)})_1 - \ldots - a_{i,i-1}(\mathbf{u}^{(k)})_{i-1} - a_{i,i+1}(\mathbf{u}^{(k)})_{i+1} - \ldots - a_{in}(\mathbf{u}^{(k)})_n \\ &= b_i - \sum_{j=1}^{i-1} a_{ij}(\mathbf{u}^{(k)})_j - \sum_{j=i+1}^{n} a_{ij}(\mathbf{u}^{(k)})_j \end{aligned}$$

para $i = 1, 2, \ldots, n$, donde $(\mathbf{u}^{(k)})_j$ denota la coordenada j-ésima del vector $\mathbf{u}^{(k)}$.

Como se puede observar, las n componentes del vector $\mathbf{u}^{(k+1)}$ pueden calcularse de forma simultánea a partir de las n componentes del vector $\mathbf{u}^{(k)}$; de hecho, el método de Jacobi también se conoce como *método de las iteraciones simultáneas*.

Ejemplo X.3.3. Volviendo al ejemplo X.3.2, la matriz de Jacobi en este caso es

$$\mathcal{J} = D^{-1}(E+F) = \begin{pmatrix} 1/2 & 0 & 0 \\ 0 & 1/3 & 0 \\ 0 & 0 & 1/2 \end{pmatrix} \begin{pmatrix} 0 & 2 & 0 \\ -2 & 0 & 1 \\ -\epsilon & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ -2/3 & 0 & 1/3 \\ -\epsilon/2 & 0 & 0 \end{pmatrix}.$$

Así, por ejemplo, para $\epsilon = -1$ el radio espectral de $\mathcal{J}$ es 0,84865653915700, para $\epsilon = -3$, es 0,97263258335935 y, para $\epsilon = -5$, es 1,08264845639125. Luego, por los criterios de convergencia para métodos iterativos, se tiene, en este caso, que para los

dos primeros valores de ϵ el método de Jacobi es convergente y que para el último no lo es.

Método de Gauss-Seidel.

A la vista del cálculo de la componente $(\mathbf{u}^{(k+1)})_i$ en el método de Jacobi, parece claro que una estrategia adecuada para mejorar la convergencia de ese método sería emplear las componentes ya calculadas

$$\{(\mathbf{u}^{(k+1)})_1, (\mathbf{u}^{(k+1)})_2, \ldots, (\mathbf{u}^{(k+1)})_{i-1}\}$$

en vez de utilizar las "antiguas"

$$\{(\mathbf{u}^{(k)})_1, (\mathbf{u}^{(k)})_2, \ldots, (\mathbf{u}^{(k)})_{i-1}\}$$

Esta consideración nos sugiere la siguiente modificación en la descripción coordenada a coordenada de la k-ésima iteración del método de Jacobi:

$$a_{ii}(\mathbf{u}^{(k+1)})_i = b_i - \sum_{j=1}^{i-1} a_{ij}(\mathbf{u}^{(k+1)})_j - \sum_{j=i+1}^{n} a_{ij}(\mathbf{u}^{(k)})_j$$

para $i = 1, 2, \ldots, n$. Matricialmente, estas ecuaciones se escriben

$$D\mathbf{u}^{(k+1)} = E\mathbf{u}^{(k+1)} + F\mathbf{u}^{(k)} + \mathbf{b},$$

es decir,

$$(D - E)\mathbf{u}^{(k+1)} = F\mathbf{u}^{(k)} + \mathbf{b}.$$

Tenemos así definido un nuevo método iterativo tomando

$$\boxed{M = D - E} \quad \text{y} \quad \boxed{N = F}$$

De esta forma

$$A\mathbf{u} = \mathbf{b} \Longleftrightarrow (D - E)\mathbf{u} = F\mathbf{u} + \mathbf{b} \Longleftrightarrow \mathbf{u} = (D - E)^{-1}F\mathbf{u} + (D - E)^{-1}\mathbf{b}$$

que conduce al **método iterativo de Gauss-Seidel**

$$\begin{cases} \mathbf{u}^{(0)} \in V \quad \text{arbitrario} \\ \mathbf{u}^{(k+1)} = (D - E)^{-1}F\mathbf{u}^{(k)} + (D - E)^{-1}\mathbf{b}, \ k \in \mathbb{N} \cup \{0\} \end{cases}$$

o, en forma equivalente,

$$\begin{cases} \mathbf{u}^{(0)} \in V \quad \text{arbitrario} \\ (D - E)\mathbf{u}^{(k+1)} = F\mathbf{u}^{(k)} + \mathbf{b}, \ k \in \mathbb{N} \cup \{0\} \end{cases}$$

Nótese que, por (X.3.4), la matriz $M = D - E$ es invertible. La matriz de este método es

$$\boxed{\mathcal{L}_1 = (D - E)^{-1}F = I_n - (D - E)^{-1}A}$$

que se denomina **matriz de Gauss-Seidel**.

Contrariamente a lo que sucedía en el método de Jacobi, las n componentes del vector $\mathbf{u}^{(k+1)}$ debe obtenerse de manera sucesiva a partir de las componentes ya calculadas de $\mathbf{u}^{(k+1)}$ y las restantes del vector $\mathbf{u}^{(k)}$; por ello, a este método se le denomina método de las *iteraciones sucesivas*. Además, según lo dicho anteriormente, el método de Gauss-Seidel será, en principio, más "rápido" pues la matriz M contiene más elementos de A. Aunque no siempre ocurre así:

Ejemplo X.3.4. Retornando de nuevo al ejemplo X.3.2, la matriz de Gauss-Seidel en este caso es

$$\mathcal{L}_1 = (D-E)^{-1}F = \begin{pmatrix} 1/2 & 0 & 0 \\ -1/3 & 1/3 & 0 \\ -\epsilon/4 & 0 & 1/2 \end{pmatrix} \begin{pmatrix} 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & -2/3 & -1/3 \\ 0 & -\epsilon/2 & 0 \end{pmatrix}.$$

Así, por ejemplo, para $\epsilon = -1$ el radio espectral de $\mathcal{L}_1$ es 0,86037961002806; para $\epsilon = -3$, es 1,11506929330390 y para $\epsilon = -5$, es 1,30515864914088. Luego, por los criterios de convergencia para métodos iterativos, se tiene que para el primer valor de ϵ el método de Gauss-Seidel es convergente y que para los dos últimos no lo es. Luego, para $\epsilon = -3$, el método de Jacobi es mejor que el de Gauss-Seidel (véase el ejemplo X.3.3).

Veamos ahora un ejemplo en el que el método de Gauss-Seidel sí funciona mejor que el método de Jacobi, lo que pone manifiesto que, en general, la conveniencia de usar uno u otro está ligada al problema, es decir, no podemos asegurar que un método iterativo sea mejor que otro.

Ejemplo X.3.5. Consideremos la matriz

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

donde $\epsilon \in \mathbb{R}$. Podemos escribir, $A = D - E - F$ siendo

$$D = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 2 \end{pmatrix}, \quad E = \begin{pmatrix} 0 & 0 & 0 \\ -2 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \quad \text{y} \quad F = \begin{pmatrix} 0 & 2 & 0 \\ 0 & 0 & -\epsilon \\ 0 & 0 & 0 \end{pmatrix}.$$

Así,

$$\mathcal{J} = D^{-1}(E+F) = \begin{pmatrix} 1/2 & 0 & 0 \\ 0 & 1/3 & 0 \\ 0 & 0 & 1/2 \end{pmatrix} \begin{pmatrix} 0 & 2 & 0 \\ -2 & 0 & -\epsilon \\ -1 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ -2/3 & 0 & -\epsilon/3 \\ -1/2 & 0 & 0 \end{pmatrix}$$

y

$$\mathcal{L}_1 = (D-E)^{-1}F = \begin{pmatrix} 1/2 & 0 & 0 \\ -1/3 & 1/3 & 0 \\ -1/4 & 0 & 1/2 \end{pmatrix} \begin{pmatrix} 0 & 2 & 0 \\ 0 & 0 & -\epsilon \\ 0 & 0 & -\epsilon \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & -2/3 & -\epsilon/3 \\ 0 & -1/2 & 0 \end{pmatrix}.$$

Así, por ejemplo, para $\epsilon = -1$ los radios espectrales de $\mathcal{J}$ y de $\mathcal{L}_1$ son

$$0{,}84865653915700 \quad \text{y} \quad 0{,}40824829046386,$$

respectivamente, para $\epsilon = -4$, son

$$1{,}03018084965341 \quad \text{y} \quad 0{,}81649658092773,$$

respectivamente, y, para $\epsilon = -7$, son

$$1{,}17502381317383 \quad \text{y} \quad 1{,}08012344973464,$$

respectivamente. Luego, por los criterios de convergencia para métodos iterativos, se tiene que para el primer valor de ϵ ambos métodos son convergentes, para el segundo valor de ϵ el método de Jacobi es divergente, mientras que el de Gauss-Seidel es convergente, y para el tercer valor de ϵ ambos métodos son divergentes.

Método de relajación.

La idea que subyace en el método de relajación es tomar como valor siguiente, en cada paso del método iterativo, no el que resultaría de aplicar directamente el método, sino una media ponderada de éste y el valor anteriormente hallado, es decir,

$$\boxed{\text{Valor anterior: } \mathbf{u}^{(k)}} \Longrightarrow \boxed{\text{Método: } \mathbf{u}^{(k+1)}} \Longrightarrow \boxed{\text{Valor siguiente: } \alpha\mathbf{u}^{(k+1)} + (1-\alpha)\,\mathbf{u}^{(k)}}$$

para un factor de peso $\alpha \neq 0$. Así, por ejemplo, aplicando esta estrategia al método de Jacobi se obtiene

$$\mathbf{u}^{(k+1)} = \alpha(\mathbf{u}^{(k+1)})^{\mathcal{J}} + (1-\alpha)\,\mathbf{u}^{(k)}, \quad \alpha \neq 0$$

donde $(\mathbf{u}^{(k+1)})^{\mathcal{J}}$ es el valor obtenido al realizar una iteración en el método de Jacobi a partir de $\mathbf{u}^{(k)}$. En términos de coordenadas, tendríamos:

$$(\text{X.3.6}) \qquad (\mathbf{u}^{(k+1)})_i = \frac{\alpha}{a_{ii}}\left(b_i - \sum_{j=1}^{i-1} a_{ij}(\mathbf{u}^{(k)})_j - \sum_{j=i+1}^{n} a_{ij}(\mathbf{u}^{(k)})_j\right) + (1-\alpha)(\mathbf{u}^{(k)})_i$$

para $i = 1, 2, \ldots, n$, lo que matricialmente se escribe como

$$\begin{aligned} \mathbf{u}^{(k+1)} &= \alpha D^{-1}(\mathbf{b} + (E+F)\mathbf{u}^{(k)}) + (1-\alpha)\mathbf{u}^{(k)} \\ &= \alpha D^{-1}\left(\frac{1-\alpha}{\alpha}D + E + F\right)\mathbf{u}^{(k)} + \alpha D^{-1}\mathbf{b}. \end{aligned}$$

Este método, conocido como *método de relajación-Jacobi*, no se utiliza apenas debido a que no constituye una mejora sustancial del método de Jacobi. A la vista de las ecuaciones dadas en (X.3.6) es razonable pensar (siguiendo la idea del método de Gauss-Seidel) que los resultados obtenidos se mejorarían si usáramos cada coordenada de $\mathbf{u}^{(k+1)}$ desde el primer momento en que se haya calculado. Esto conduciría a las ecuaciones

$$(\mathbf{u}^{(k+1)})_i = \frac{\alpha}{a_{ii}}\left(b_i - \sum_{j=1}^{i-1} a_{ij}(\mathbf{u}^{(k+1)})_j - \sum_{j=i+1}^{n} a_{ij}(\mathbf{u}^{(k)})_j\right) + (1-\alpha)(\mathbf{u}^{(k)})_i$$

para $i = 1, 2, \ldots, n$, lo que, en términos matriciales, es

$$\mathbf{u}^{(k+1)} = \alpha D^{-1}(\mathbf{b} + E\mathbf{u}^{(k+1)} + F\mathbf{u}^{(k)}) + (1-\alpha)\mathbf{u}^{(k)}.$$

Agrupando se tiene que

$$(D - \alpha E)\mathbf{u}^{(k+1)} = ((1-\alpha)D + \alpha F)\mathbf{u}^{(k)} + \alpha b$$

o, equivalentemente,

$$\left(\frac{D}{\alpha} - E\right)\mathbf{u}^{(k+1)} = \left(\frac{1-\alpha}{\alpha}D + F\right)\mathbf{u}^{(k)} + b.$$

Veamos ahora que la solución obtenida mediante el uso iterado de esta fórmula coincide con la del sistema $A\mathbf{x} = \mathbf{b}$. La matriz de A puede ser escrita como $A = M - N$ siendo

$$\boxed{M = \tfrac{D}{\alpha} - E} \quad \text{y} \quad \boxed{N = \tfrac{1-\alpha}{\alpha}D + F}$$

Por tanto,

$$\begin{aligned} A\mathbf{u} = \mathbf{b} &\iff \left(\frac{D}{\alpha} - E\right)\mathbf{u} = \left(\frac{1-\alpha}{\alpha}D + F\right)\mathbf{u} + \mathbf{b} \\ &\iff \mathbf{u} = \left(\frac{D}{\alpha} - E\right)^{-1}\left(\frac{1-\alpha}{\alpha}D + F\right)\mathbf{u} + \left(\frac{D}{\alpha} - E\right)^{-1}\mathbf{b}, \end{aligned}$$

lo que conduce al **método iterativo de relajación**

$$\begin{cases} \mathbf{u}^{(0)} \in V \quad \text{arbitrario} \\ \mathbf{u}^{(k+1)} = \left(\frac{D}{\alpha} - E\right)^{-1}\left(\frac{1-\alpha}{\alpha}D + F\right)\mathbf{u}^{(k)} + \left(\frac{D}{\alpha} - E\right)^{-1}\mathbf{b}, \quad k \in \mathbb{N} \cup \{0\}. \end{cases}$$

o equivalentemente,

$$\begin{cases} \mathbf{u}^{(0)} \in V \quad \text{arbitrario} \\ \left(\frac{D}{\alpha} - E\right)\mathbf{u}^{(k+1)} = \left(\frac{1-\alpha}{\alpha}D + F\right)\mathbf{u}^{(k)} + \mathbf{b}, \quad k \in \mathbb{N} \cup \{0\} \end{cases}$$

La hipótesis (X.3.4) hace que la matriz $M = \frac{D}{\alpha} - E$ con $\alpha \neq 0$ sea invertible. La matriz de este método es

$$\boxed{\mathcal{L}_\alpha = \left(\tfrac{D}{\alpha} - E\right)^{-1}\left(\tfrac{1-\alpha}{\alpha}D + F\right) = (D - \alpha E)^{-1}((1-\alpha)D + \alpha F)}$$

denominada **matriz de relajación**. Algunos autores distinguen y denominan **sobrerrelajación** cuando $\alpha > 1$ y **subrelajación** si $\alpha < 1$. Nótese que para $\alpha = 1$ se tiene el método de Gauss-Seidel, lo que hace coherente la notación $\mathcal{L}_1$ para la matriz asociada al mismo.

En inglés el método de relajación se conoce como *Successive Over-Relaxation method*, de aquí que en muchas ocasiones se le denomine *método SOR*.

Nota X.3.6. El estudio del método de relajación consiste en determinar (si existen):

- un *intervalo* $I \subset \mathbb{R}$, que no contenga al origen, tal que
$$\alpha \in I \Longrightarrow \varrho(\mathcal{L}_\alpha) < 1;$$
- un *parámetro de relajación óptimo* $\alpha_0 \in I$ tal que
$$\varrho(\mathcal{L}_{\alpha_0}) = \inf\{\varrho(\mathcal{L}_\alpha) \mid \alpha \in I\}$$

Análisis de convergencia.

El estudio de la convergencia de los métodos iterativos puede ser demasiado prolijo puesto que no existen teoremas que aseguren la convergencia para una clase general de matrices. No obstante, pueden darse resultados parciales para determinados tipos de matrices; aquí presentamos un resultado de carácter general y sendas condiciones de convergencia para el método de relajación y el de Jacobi, recogiendo otros resultados en los ejercicios.

Lema X.3.7. *Sea $A \in \mathcal{M}_n(\Bbbk)$ una matriz hermítica definida positiva escrita como $A = M - N$ con $M \in \mathcal{M}_n(\Bbbk)$ invertible. Si la matriz $M^* + N$ es definida positiva, entonces*

$$\varrho(M^{-1}N) < 1.$$

Por consiguiente, en la situación anterior, el método iterativo definido por la matriz $B = M^{-1}N$ es convergente.

Demostración. En primer lugar, por ser A hermítica,

$$\begin{aligned}(M^* + N)^* &= M + N^* = (A + N) + N^* = (A^* + N^*) + N \\ &= (A + N)^* + N = M^* + N\end{aligned}$$

por lo que la matriz $M^* + N$ es hermítica. Por otra parte, sea $\lambda \in \mathrm{sp}(M^{-1}N)$ y $\mathbf{v} \in V \setminus \{0\}$ un autovector asociado al autovalor λ, es decir,

$$M^{-1}N\mathbf{v} = \lambda\mathbf{v}. \tag{X.3.7}$$

A partir de $\mathbf{v}$ construyamos el vector

$$\mathbf{w} = M^{-1}N\mathbf{v} \tag{X.3.8}$$

En primer lugar, nótese que $\mathbf{w} \neq \mathbf{v}$. En efecto, en caso contrario se obtendría, a partir de (X.3.8),

$$\mathbf{v} = M^{-1}N\mathbf{v} \Longrightarrow M\mathbf{v} = N\mathbf{v} \Longrightarrow A\mathbf{v} = (M-N)\mathbf{v} = \mathbf{0},$$

lo que contradice que A sea invertible al ser $\mathbf{v}$ no nulo. Por otra parte, como

$$M\mathbf{w} = N\mathbf{v},$$

se verifica que

$$\begin{aligned}(\mathbf{v}-\mathbf{w})^*(M^*+N)(\mathbf{v}-\mathbf{w}) &= (\mathbf{v}-\mathbf{w})^*M^*(\mathbf{v}-\mathbf{w}) + (\mathbf{v}-\mathbf{w})^*N(\mathbf{v}-\mathbf{w}) \\ &= (M\mathbf{v}-M\mathbf{w})^*(\mathbf{v}-\mathbf{w}) + (\mathbf{v}-\mathbf{w})^*(N\mathbf{v}-N\mathbf{w}) \\ &= (M\mathbf{v}-N\mathbf{v})^*(\mathbf{v}-\mathbf{w}) + (\mathbf{v}-\mathbf{w})^*(M\mathbf{w}-N\mathbf{w}) \\ &= \mathbf{v}^*A^*(\mathbf{v}-\mathbf{w}) + (\mathbf{v}-\mathbf{w})^*A\mathbf{w} \\ &= \mathbf{v}^*A\mathbf{v} - \mathbf{v}^*A\mathbf{w} + \mathbf{v}^*A\mathbf{w} - \mathbf{w}^*A\mathbf{w} \\ &= \mathbf{v}^*A\mathbf{v} - \mathbf{w}^*A\mathbf{w}\end{aligned}$$

por ser $A = M - N$ y $M^* + N$ matrices hermíticas. Por tanto,

$$\mathbf{v}^*A\mathbf{v} - \mathbf{w}^*A\mathbf{w} = (\mathbf{v}-\mathbf{w})^*(M^*+N)(\mathbf{v}-\mathbf{w}) > 0 \tag{X.3.9}$$

ya que $\mathbf{v} - \mathbf{w} \neq 0$ y $M^* + N$ es definida positiva. Ahora bien, a partir de (X.3.7), (X.3.8) y (X.3.9) se obtiene que

$$\begin{aligned}0 < \mathbf{v}^*A\mathbf{v} - \mathbf{w}^*A\mathbf{w} &= \mathbf{v}^*A\mathbf{v} - (\lambda\mathbf{v})^*A(\lambda\mathbf{v}) \\ &= \mathbf{v}^*A\mathbf{v} - (\bar{\lambda}\mathbf{v}^*)A(\lambda\mathbf{v}) \\ &= (1-|\lambda|^2)\,\mathbf{v}^*A\mathbf{v}.\end{aligned}$$

Como $\mathbf{v}^*A\mathbf{v} > 0$ por ser A definida positiva y $\mathbf{v} \neq 0$, entonces $1 - |\lambda|^2 > 0$, de donde se sigue que $|\lambda| < 1$, obteniéndose así el resultado buscado ∎

A continuación vamos a dar una condición necesaria y suficiente para la convergencia del método de relajación.

Teorema de Ostrowski-Reich. *Si $A \in \mathcal{M}_n(\mathbb{k})$ es una matriz hermítica definida positiva y $0 < \alpha < 2$, entonces el método de relajación es convergente. En particular, cuando A es hermítica y definida positiva el método de Gauss-Seidel es convergente.*

Demostración. La descomposición $A = M - N$ asociada al método de relajación es

$$A = \left(\frac{D}{\alpha} - E\right) - \left(\frac{1-\alpha}{\alpha}D + F\right),\ \alpha \neq 0.$$

Como la matriz A es hermítica se tiene que

$$D - E - F = A = A^* = D^* - E^* - F^*.$$

Identificando en la igualdad anterior los elementos diagonales y los que quedan en la parte triangular inferior y superior de A, se verifica que $D^* = D$ y $E^* = F$. Por tanto,

$$M^* + N = \frac{D}{\alpha} - E^* + \frac{1-\alpha}{\alpha}D + F = \frac{2-\alpha}{\alpha}D;$$

de modo que para valores del parámetro $0 < \alpha < 2$ se tiene que

$$\mathbf{v}^*(M^* + N)\mathbf{v} = \frac{2-\alpha}{\alpha}\mathbf{v}^*D\mathbf{v} > 0$$

pues D es definida positiva.[2] Aplicando el lema X.3.7 concluimos el resultado. ∎

Existen extensiones del teorema de Ostrowski-Reich a situaciones más generales; por ejemplo, el lector interesado puede encontrar en el artículo [J.M. Ortega y R.J. Plemmons *Extensions of the Ostrowski-Reich theorem for SOR iterations.* Linear Algebra Appl. **28** (1979), 177–191] generalizaciones del teorema de Ostrowski-Reich al caso en que A sea hermítica pero no necesariamente definida positiva, o al caso en que $A + A^*$ sea definida positiva pero A no sea hermítica.

Veamos ahora que la condición $0 < \alpha < 2$ es necesaria para la convergencia del método de relajación.

Teorema de Kahan. *El radio espectral de la matriz de la relajación siempre verifica*

$$\varrho(\mathcal{L}_\alpha) \geq |\alpha - 1|,\ \alpha \neq 0.$$

Consecuentemente, el método de relajación sólo puede ser convergente cuando $0 < \alpha < 2$.

Demostración. Por definición

$$\det(\mathcal{L}_\alpha) = \det\left(\left(\frac{D}{\alpha} - E\right)^{-1}\left(\frac{1-\alpha}{\alpha}D + F\right)\right) = \frac{\det\left(\frac{1-\alpha}{\alpha}D + F\right)}{\det\left(\frac{D}{\alpha} - E\right)}.$$

[2]En efecto, si $A = (a_{ij}) \in \mathcal{M}_n(\Bbbk)$ es hermítica, entonces $\mathbf{e}_i^* A \mathbf{e}_i = a_{ii} > 0$, para todo $i = 1, 2, \ldots, n$, siendo $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ la base usual de $\Bbbk^n$. Por otra parte, como $D = \operatorname{diag}(a_{11}, a_{22}, \ldots, a_{nn})$ se sigue que

$$\mathbf{v}^* D \mathbf{v} = \sum_{i=1}^{n} a_{ii}|v_i|^2 > 0.$$

Como

$$\det\left(\frac{1-\alpha}{\alpha}D+F\right)=\det\left(\frac{1-\alpha}{\alpha}D\right)\quad\text{y}\quad\det\left(\frac{D}{\alpha}-E\right)=\det\left(\frac{D}{\alpha}\right),$$

entonces

$$\det(\mathcal{L}_\alpha)=\frac{\det\left(\frac{1-\alpha}{\alpha}D\right)}{\det\left(\frac{D}{\alpha}\right)}=\frac{\frac{(1-\alpha)^n}{\alpha^n}\det(D)}{\frac{1}{\alpha^n}\det(D)}=(1-\alpha)^n. \tag{X.3.10}$$

Por otra parte, si $\mathrm{sp}(\mathcal{L}_\alpha)=\{\lambda_1,\lambda_2,\ldots,\lambda_n\}$ entonces $\det(\mathcal{L}_\alpha)=\lambda_1\cdot\lambda_2\cdots\lambda_n$. Así, usando (X.3.10) se obtiene que

$$\prod_{i=1}^{n}|\lambda_i|=|1-\alpha|^n,$$

lo que permite concluir que

$$\varrho(\mathcal{L}_\alpha)\geq\left(\prod_{i=1}^{n}|\lambda_i|\right)^{\frac{1}{n}}\geq|1-\alpha|.$$

■

En la aplicaciones concretas de los sistemas de ecuaciones lineales aparecen, con mucha frecuencia, matrices $A=(a_{ij})\in\mathcal{M}_n(\Bbbk)$ diagonalmente dominante por filas[3]. Estas matrices son invertibles y además, $a_{ii}\neq 0,\ i=1,2,\ldots,n$. Para este tipo de matrices se tiene el siguiente resultado de convergencia para el método de Jacobi.

Proposición X.3.8. *Si $A\in\mathcal{M}_n(\Bbbk)$ es una matriz diagonalmente dominante por filas, el método de Jacobi es convergente.*

Demostración. La matriz de iteración del método de Jacobi $\mathcal{J}=D^{-1}(E+F)$ verifica que

$$(\mathcal{J})_{ij}=\begin{cases}-a_{ij}/a_{ii} & \text{si}\quad i\neq j;\\ 0 & \text{si}\quad i=j.\end{cases}$$

Por tanto, a partir del teorema VIII.2.7, se tiene que

$$\|\mathcal{J}\|_\infty=\max_{1\leq i\leq n}\sum_{j=1}^{n}\left|(\mathcal{J})_{ij}\right|=\max_{1\leq i\leq n}\sum_{\substack{j=1\\ j\neq i}}^{n}\frac{|a_{ij}|}{|a_{ii}|}=\max_{1\leq i\leq n}\left(\frac{1}{|a_{ii}|}\sum_{\substack{j=1\\ j\neq i}}^{n}|a_{ij}|\right)<1.$$

[3]Recuérdese, que una matriz $A=(a_{ij})\in\mathcal{M}_n(\Bbbk)$ es diagonalmente dominante por filas si

$$|a_{ii}|>\sum_{\substack{j=1\\ j\neq i}}^{n}|a_{ij}|,$$

para todo $i=1,\ldots,n$.

De esta forma, aplicando los criterios de convergencia para métodos iterativos, se concluye el resultado. ∎

Comparación de los métodos iterativos.

Veamos a continuación que, en el caso en que la matriz A es tridiagonal, se pueden comparar de forma muy precisa los radios espectrales de las matrices de Jacobi, Gauss-Seidel y de relajación, tanto en el caso convergente como en el divergente. El caso $\alpha \neq 1$ es técnicamente más difícil que el caso $\alpha = 1$, por lo que solamente demostraremos el teorema de comparación de los radios espectrales de los métodos de Jacobi y Gauss-Seidel, y nos limitaremos a enunciar el resto de los teoremas.

Lema X.3.9. *Sean $\mu \in \Bbbk \setminus \{0\}$ y $A(\mu) \in \mathcal{M}_n(\Bbbk)$ una matriz tridiagonal de la forma*

$$A(\mu) = \begin{pmatrix} b_1 & \mu^{-1}c_1 & & & \\ \mu a_2 & b_2 & \mu^{-1}c_2 & & \\ & \ddots & \ddots & \ddots & \\ & & \mu a_{n-1} & b_{n-1} & \mu^{-1}c_{n-1} \\ & & & \mu a_n & b_n \end{pmatrix}.$$

Entonces,

$$\det(A(\mu)) = \det(A(1)),$$

para todo $\mu \in \Bbbk$ no nulo.

Demostración. Sea $Q(\mu) = \operatorname{diag}(\mu, \mu^2, \ldots, \mu^n) \in \mathcal{M}_n(\Bbbk)$. Se comprueba fácilmente que $A(\mu) = Q(\mu)A(1)Q(\mu)^{-1}$, de donde se sigue el resultado buscado. ∎

Comparación de los métodos de Jacobi y Gauss-Seidel. *Si A es tridiagonal, entonces los radios espectrales de las correspondientes matrices de Jacobi y de Gauss-Seidel está relacionados por*

$$\varrho(\mathcal{L}_1) = \varrho(\mathcal{J})^2,$$

de tal forma que los métodos convergen o divergen simultáneamente; además, en caso de convergencia, el método de Gauss-Seidel es más rápido que el método de Jacobi.

Demostración. Los autovalores de la matriz de Jacobi $\mathcal{J} = D^{-1}(E+F)$ son las raíces del polinomio

$$\aleph_{\mathcal{J}}(x) = \det(D^{-1}(E+F) - xI_n)$$

que coinciden con la raíces del polinomio

$$q_{\mathcal{J}}(x) = \det(xD - E - F) = \det(-D)\,\aleph_{\mathcal{J}}(x).$$

De la misma forma, los autovalores de la matriz de Gauss-Seidel $\mathcal{L}_1 = (D-E)^{-1}F$ son las raíces del polinomio

$$\aleph_{\mathcal{L}_1}(x) = \det((D-E)^{-1}F - xI_n),$$

que coinciden con las raíces del polinomio

$$q_{\mathcal{L}_1}(x) = \det(xD - xE - F) = \det(E-D)\,\aleph_{\mathcal{L}_1}(x).$$

Teniendo en cuenta la estructura tridiagonal de la matriz A, del lema X.3.9 se sigue que

$$q_{\mathcal{L}_1}(x^2) = \det(x^2D - x^2E - F) = \det(x^2D - xE - xF) = x^n\, q_{\mathcal{J}}(x),$$

para todo $x \in \Bbbk$, pues por continuidad esta expresión también es válida en $x = 0$. De esta relación funcional, se deducen las siguientes implicaciones

$$\lambda \in \mathrm{sp}(\mathcal{L}_1) \text{ no nulo} \Rightarrow \pm\sqrt{\lambda} \in \mathrm{sp}(\mathcal{J});$$

$$\{\lambda \in \mathrm{sp}(\mathcal{J}) \iff -\lambda \in \mathrm{sp}(\mathcal{J})\} \Rightarrow \lambda^2 \in \mathrm{sp}(\mathcal{L}_1).$$

De donde se sigue el resultado deseado. ∎

Comparación de los métodos de Jacobi y de relajación. *Sea A una matriz tridiagonal tal que todos los autovalores de la matriz de Jacobi correspondiente son reales. Entonces, el método de Jacobi, y el método de relajación para $0 < \alpha < 2$, convergen o divergen simultáneamente; además, en caso de convergencia, la función $\alpha \in (0,2) \mapsto \varrho(\mathcal{L}_\alpha)$ alcanza un mínimo absoluto en*

$$\alpha_0 = \frac{2}{1 + \sqrt{1 - \varrho(\mathcal{J})^2}}.$$

Demostración. Véase el teorema 5.3-5 de [**Cia82**]. ∎

Uniendo los resultados del Teorema de Kahan y la anterior comparación de métodos, obtenemos un resultado donde se pueden comparar los radios espectrales de las matrices $\mathcal{J}, \mathcal{L}_1, \mathcal{L}_{\alpha_0}$.

Corolario X.3.10. *Sea A una matriz hermítica, definida positiva y tridiagonal por bloques. Entonces los métodos de Jacobi, Gauss-Seidel y de relajación para $\alpha \in (0,2)$, son convergentes. Además, existe un único parámetro de relajación óptimo α_0 y se tiene que*

$$\varrho(\mathcal{L}_{\alpha_0}) = \inf_{0<\alpha<2}(\mathcal{L}_\alpha) = \alpha_0 - 1 < \varrho(\mathcal{L}_1) = \varrho(\mathcal{J})^2 < \varrho(\mathcal{J})$$

si $\varrho(\mathcal{J}) > 0$; si $\varrho(\mathcal{J}) = 0$, entonces $\alpha_0 = 1$ y $\varrho(\mathcal{L}_1) = \varrho(\mathcal{J}) = 0$.

Demostración. Véase el teorema 5.3-6 de [**Cia82**]. ∎

Test de parada de las iteraciones.

Como ya se ha dicho, cuando un método iterativo es convergente, la solución del sistema lineal $A\mathbf{x} = \mathbf{b}$ se obtiene como límite de la sucesión $(\mathbf{u}^{(k)})_{k\in\mathbb{N}}$ de iteraciones. Ante la imposibilidad de calcular todas las iteraciones, se plantea el problema de determinar $k \in \mathbb{N}$ tal que $\mathbf{u}^{(k)}$ sea una "buena" aproximación de $\mathbf{u}$. Es decir, si se desea que el error relativo sea inferior a una cantidad prefijada $\varepsilon > 0$, el valor de $k \in \mathbb{N}$ debe cumplir

$$\|\vec{\varepsilon}_k\| = \|\mathbf{u}^{(k)} - \mathbf{u}\| < \varepsilon\|\mathbf{u}\|$$

para alguna norma vectorial $\|\cdot\|$. Por supuesto, al ser el vector $\mathbf{u}$ desconocido, no se puede trabajar directamente con esas cantidades.

El test más sencillo que podemos emplear es detener el proceso cuando la diferencia entre dos iteraciones consecutivas sea, en términos relativos, menor que la **tolerancia** admisible ε, es decir,

$$\|\mathbf{u}^{(k+1)} - \mathbf{u}^{(k)}\| < \varepsilon\|\mathbf{u}^{(k+1)}\|. \tag{X.3.11}$$

Si embargo, este test tiene el inconveniente de que puede cumplirse la relación (X.3.11) sin que el vector $\mathbf{u}^{(k+1)}$ esté próximo a $\mathbf{u}$.

Una condición de parada de las iteraciones más adecuada viene dada a partir del vector residual.

Definición X.3.11. Con la notación anterior. Se llama **vector residual** k-ésimo de un método iterativo a

$$\mathbf{r}^{(k)} := \mathbf{b} - A\mathbf{u}^{(k)} = A(\mathbf{u} - \mathbf{u}^{(k)}),\ k \in \mathbb{N} \cup \{0\}.$$

En general, Si $\tilde{\mathbf{u}}$ es una aproximación de la solución de $A\mathbf{x} = \mathbf{b}$ se llama vector residual a $\mathbf{b} - A\tilde{\mathbf{u}}$.

Proposición X.3.12. *Si* $\tilde{\mathbf{u}}$ *es una aproximación de la solución* $\mathbf{u}$ *del sistema* $A\mathbf{x} = \mathbf{b}$, *entonces, para la norma subordinada* $|||\cdot|||$ *a una norma vectorial* $\|\cdot\|$ *cualquiera, se tiene que*

$$\|\mathbf{u} - \tilde{\mathbf{u}}\| \leq |||A^{-1}||| \cdot \|\mathbf{b} - A\tilde{\mathbf{u}}\|$$

y

$$\frac{\|\mathbf{u} - \tilde{\mathbf{u}}\|}{\|\mathbf{u}\|} \leq \mathrm{cond}(A)\,\frac{\|\mathbf{b} - A\tilde{\mathbf{u}}\|}{\|\mathbf{b}\|}.$$

Demostración. Es una consecuencia directa de la proposición VIII.2.5. ∎

A la vista del proposición anterior, es razonable pensar que si $\mathbf{u}^{(k)}$ está próximo a $\mathbf{u}$, entonces $A\mathbf{u}^{(k)}$ está próximo a $\mathbf{b}$. Por tanto, pararemos las iteraciones cuando

$$\frac{\|\mathbf{r}^{(k)}\|}{\|\mathbf{b}\|} = \frac{\|A\mathbf{u}^{(k)} - A\mathbf{u}\|}{\|A\mathbf{u}\|} < \varepsilon,$$

es decir, para valores de $k \in \mathbb{N}$ tales que

$$\|\mathbf{r}^{(k)}\| < \varepsilon \, \|\mathbf{b}\|.$$

Obviamente, debe procurarse que la comprobación de los test de parada no incremente en exceso el número de operaciones necesarias para realizar una iteración. Veamos cómo organizando los cálculos de forma adecuada esto puede conseguirse tanto en el método de Jacobi como en el de relajación:

a) En el *método de Jacobi* podemos reescribir la iteración cómo

$$D\mathbf{u}^{(k+1)} = \mathbf{b} + (E+F)\mathbf{u}^{(k)} = \mathbf{b} - A\mathbf{u}^{(k)} + D\mathbf{u}^{(k)} = \mathbf{r}^{(k)} + D\mathbf{u}^{(k)},$$

es decir,

$$D(\mathbf{u}^{(k+1)} - \mathbf{u}^{(k)}) = \mathbf{r}^{(k)}.$$

De esta forma calculando en primer lugar el vector $\mathbf{r}^{(k)}$ (mediante la fórmula $\mathbf{r}^{(k)} = \mathbf{b} - A\mathbf{u}^{(k)}$), resolviendo a continuación el sistema $D\mathbf{d}^{(k)} = \mathbf{r}^{(k)}$ y tomando

$$\mathbf{u}^{(k+1)} = \mathbf{u}^{(k)} + \mathbf{d}^{(k)}$$

obtenemos la información necesaria para los test de parada así como la iteración siguiente $\mathbf{u}^{(k+1)}$ sin ver incrementado sustancialmente el número de operaciones. En el caso particular del método de Jacobi, para cada $i \in \{1, 2, \ldots, n\}$, se calculan

$$\begin{array}{rcl} (\mathbf{r}^{(k)})_i & = & \mathbf{b}_i - \sum_{j=1}^{n} a_{ij}(\mathbf{u}^{(k)})_j \\ (\mathbf{d}^{(k)})_i & = & (\mathbf{r}^{(k)})_i / a_{ii} \\ (\mathbf{u}^{(k+1)})_i & = & (\mathbf{u}^{(k)})_i + (\mathbf{d}^{(k)})_i \end{array}$$

(b) En el *método de relajación* podemos reescribir la iteración como

$$\left(\frac{D}{\alpha} - E\right)\mathbf{u}^{(k+1)} = \left(\frac{1-\alpha}{\alpha}D + F\right)\mathbf{u}^{(k)} + \mathbf{b},$$

es decir,

$$\frac{D}{\alpha}\mathbf{u}^{(k+1)} = E\mathbf{u}^{(k+1)} - D\mathbf{u}^{(k)} + F\mathbf{u}^{(k)} + \frac{D}{\alpha}\mathbf{u}^{(k)} + \mathbf{b} = \tilde{\mathbf{r}}^{(k)} + \frac{D}{\alpha}\mathbf{u}^{(k)}$$

siendo

$$\tilde{\mathbf{r}}^{(k)} = \mathbf{b} - \left((D-F)\mathbf{u}^{(k)} - E\mathbf{u}^{(k+1)}\right)$$

y, de esta forma,

$$D(\mathbf{u}^{(k+1)} - \mathbf{u}^{(k)}) = \alpha\tilde{\mathbf{r}}^{(k)}.$$

En el caso particular del método de relajación se tiene que

$$(\tilde{\mathbf{r}}^{(k)})_i = \mathbf{b}_i - (A\mathbf{u}'^{(k)})_i$$

para $i = 1, 2, \ldots, n$, donde

$$\mathbf{u}'^{(k)} = \left((\mathbf{u}^{(k+1)})_1, (\mathbf{u}^{(k+1)})_2, \ldots, (\mathbf{u}^{(k+1)})_{i-1}, (\mathbf{u}^{(k)})_i, \ldots, (\mathbf{u}^{(k)})_n\right)^{\mathrm{t}}.$$

Es decir, para cada $i \in \{1, 2, \ldots, n\}$, se calculan

$$\begin{array}{rcl} (\tilde{\mathbf{r}}^{(k)})_i & = & \mathbf{b}_i - \sum_{j=1}^{i-1} a_{ij}(\mathbf{u}^{(k+1)})_j - \sum_{j=i}^{n} a_{ij}(\mathbf{u}^{(k)})_j \\ (\mathbf{d}^{(k)})_i & = & \alpha(\tilde{\mathbf{r}}^{(k)})_i / a_{ii} \\ (\mathbf{u}^{(k+1)})_i & = & (\mathbf{u}^{(k)})_i + (\mathbf{d}^{(k)})_i \end{array}$$

Para acabar, simplemente reseñar que las normas vectoriales que suelen emplearse con mayor frecuencia en este tipo de test son $\|\cdot\|_2$ y $\|\cdot\|_\infty$.

4. Métodos iterativos estacionarios y no estacionarios

Como en las secciones anteriores, denotemos por

$$B = I_n - M^{-1}A$$

la matriz de iteración asociada con el método iterativo (X.2.3). Procediendo como el caso del método de relajación, (X.2.3) puede ser generalizado introduciendo un parámetro α de relajación (o aceleración); de tal modo que consideremos descomposiciones de A de la forma

$$\boxed{A = \tfrac{1}{\alpha}M - N}.$$

De esta forma podemos considerar el método iterativo

$$\text{(X.4.12)} \qquad \begin{cases} \mathbf{u}^{(0)} \in V \text{ arbitrario} \\ \mathbf{u}^{(k+1)} = \mathbf{u}^{(k)} + \alpha M^{-1}\mathbf{r}^{(k)}, \ k \in \mathbb{N} \cup \{0\}, \end{cases}$$

donde $\mathbf{r}^{(k)}$ es el k-ésimo vector residual (véase la definición X.3.11). Este método se conoce como el **método de Richardson estacionario**.

De forma más general, permitiendo que α dependa del índice de iteración, se consigue el que se conoce como **método de Richardson no estacionario**

$$\text{(X.4.13)} \qquad \begin{cases} \mathbf{u}^{(0)} \in V \text{ arbitrario} \\ \mathbf{u}^{(k+1)} = \mathbf{u}^{(k)} + \alpha_k M^{-1}\mathbf{r}^{(k)}, \ k \in \mathbb{N} \cup \{0\}, \end{cases}$$

La matriz de iteración en la etapa k-ésima para este tipo de métodos es

$$B_{\alpha_k} = I_n - \alpha_k M^{-1}A,$$

con $\alpha_k = \alpha$ en el caso estacionario. Obsérvese que los métodos de Jacobi y Gauss-Seidel se pueden considerar métodos de Richardson estacionarios para $\alpha = 1$, $M = D$ y $M = D - E$, respectivamente.

Podemos escribir (X.4.13) (y, por lo tanto, (X.4.12) también) en una forma más apropiada para el cálculo. Sea

$$\mathbf{z}^{(k)} = M^{-1}\mathbf{r}^{(k)}$$

el llamado **vector residual precondicionado**. Entonces se tiene que $\mathbf{u}^{(k+1)} = \mathbf{u}^{(k)} + \alpha_k \mathbf{z}^{(k)}$ y $\mathbf{r}^{(k+1)} = \mathbf{b} - A\mathbf{u}^{(k+1)} = \mathbf{r}^{(k)} - \alpha_k A\mathbf{z}^{(k)}$. En resumen, un método de Richardson no estacionario requiere en su etapa $(k+1)$-ésima las siguientes operaciones

- resolver el sistema $M\mathbf{z}^{(k)} = \mathbf{r}^{(k)}$;
- calcular el parámetro de aceleracion;
- actualizar la solución $\mathbf{u}^{(k+1)} = \mathbf{u}^{(k)} + \alpha_k \mathbf{z}^{(k)}$;
- actualizar el vector residual $\mathbf{r}^{(k+1)} = \mathbf{r}^{(k)} - \alpha_k A\mathbf{z}^{(k)}$.

Por el momento sólo nos ocuparemos del método de Richardson estacionario, es decir, $\alpha_k = \alpha$, para todo k. En este caso se cumple el siguiente resultado de convergencia.

Teorema X.4.1. *Para cualquier matriz invertible M, el método de Richardson estacionario (X.4.12) es convergente si, y sólo si,*

$$\frac{2\mathrm{Re}(\lambda_i)}{\alpha|\lambda_i|^2} > 1,\ i = 1, \ldots, n,$$

donde $\mathrm{sp}(M^{-1}A) = \{\lambda_1, \ldots, \lambda_n\}$.

Demostración. Por el criterio de convergencia para método iterativos, tenemos que el método de Richardson estacionario es convergente si, y sólo si, el radio espectral de la matriz de iteración $B_\alpha = I_n - \alpha M^{-1}A$ es estrictamente menor que 1. Equivalentemente, cuando $|1 - \alpha\lambda_i| < 1,\ i = 1, \ldots, n$. De donde se sigue la desigualdad

$$(1 - \alpha\mathrm{Re}(\lambda_i))^2 + \alpha^2(\mathrm{Im}(\lambda_i))^2 < 1,$$

que implica de forma clara la desigualdad buscada. ∎

Obsérvese que si los signos de las partes de reales de los autovalores de $M^{-1}A$ no son constantes, el método de Richardson estacionario no convergerá.

Se pueden obtener resultados de convergencia más específicos bajo ciertas condiciones sobre el espectro de $M^{-1}A$.

Teorema X.4.2. *Si M es invertible y $M^{-1}A$ tiene todos sus autovalores reales positivos, $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n > 0$, entonces el método de Richardson estacionario (X.4.12) es convergente si, y sólo si $0 < \alpha < 2/\lambda_1$. Además, si $\alpha_{opt} = \frac{2}{\lambda_1 + \lambda_n}$ el radio espectral de $B_{\alpha_{opt}}$ es mínimo:*

$$\rho_{opt} = \min_{\alpha} B_\alpha = \frac{\lambda_1 - \lambda_n}{\lambda_1 + \lambda_n}. \tag{X.4.14}$$

Demostración. Los autovalores de B_α son $1-\alpha\lambda_i,\ i=1,\dots,n$, luego (X.4.12) es convergente si, y sólo si, $|1-\alpha\lambda_i|<1,\ i=1,\dots,n$, es decir, si $0<\alpha<2/\lambda_1$. Por otra parte, se comprueba que $\rho(B_\alpha)$ es mínimo cuando $1-\alpha\lambda_n=\alpha\lambda_1-1$ (véase la figura 4.2 en [**QSS07**] p. 138), es decir, para $\alpha=2/(\lambda_1+\lambda_n)$, lo que proporciona el valor deseado para α_{opt}. Sin más que sustituir, se obtiene el valor de ρ_{opt} buscado. ∎

Los resultados anteriores ponen de manifiesto que la elección del precondicionador es fundamental en el método de Richardson. El lector interesado en profundizar en este tema puede consultar el apartado 4.3.2 de [**QSS07**].

Corolario X.4.3. *Sea A una matriz simétrica definida positiva. El método de Richardson estacionario para $M=I_n$ es convergente y*

$$\left\|\vec{\varepsilon}^{(k+1)}\right\|_A \leq \rho(B_\alpha)\left\|\vec{\varepsilon}\right\|_A,\ k\geq 0.$$

Demostración. La convergencia es consecuencia del teorema X.4.1. Además, observamos que

$$\left\|\vec{\varepsilon}^{(k+1)}\right\|_A = \|B_\alpha\vec{\varepsilon}_k\|_A = \left\|A^{1/2}B_\alpha\vec{\varepsilon}_k\right\|_2 \leq \left\|A^{1/2}B_\alpha A^{-1/2}\right\|_2\ \left|A^{1/2}\vec{\varepsilon}_k\right|_2.$$

La matriz B_α es simétrica y definida positiva y semejante a $A^{1/2}B_\alpha A^{-1/2}$. Por lo tanto

$$\left\|A^{1/2}B_\alpha A^{-1/2}\right\| = \rho(B_\alpha).$$

Basta observa que $\|A^{1/2}\vec{\varepsilon}\|_2=\|\vec{\varepsilon}\|_A$, para obtener la desigualdad buscada. ∎

Un resultado similar se obtiene para cualquier M siempre que M, A y $M^{-1}A$ sean simétricas y definidas positivas (ejercicio 9).

El método del gradiente.

La expresión óptima del parámetro α dada en el teorema X.4.2 es de un uso muy limitado en casos prácticos, puesto que requiere el conocimiento del mayor y el menor autovalor de $M^{-1}A$. En el caso especial de las matrices simétricas definidas positivas, el parámetro de aceleración óptimo se puede calcular *dinámicamente* en cada etapa k como sigue.

En primer lugar observamos que, para las matrices simétricas definidas positivas, resolver el sistema $A\mathbf{x}=\mathbf{b}$ es equivalente a calcular el valor mínimo de la forma cuadrática (no homogénea)

$$\Phi(\mathbf{x})=\frac{1}{2}\mathbf{x}^tA\mathbf{x}-\mathbf{x}^t\mathbf{b}$$

que se denomina **energía del sistema** $A\mathbf{x}=\mathbf{b}$. En efecto, el gradiente de Φ es

$$\nabla\Phi(\mathbf{x})=\frac{1}{2}(A^t+A)\mathbf{x}-\mathbf{b}. \tag{X.4.15}$$

Como consecuencia, si $\nabla\Phi(\mathbf{u}) = \mathbf{0}$, entonces $\mathbf{u}$ es solución del sistema original. Por consiguiente, si $\mathbf{u}$ es solución, entonces

$$\Phi(\mathbf{u}) = \Phi(\mathbf{u} + (\mathbf{v} - \mathbf{u})) = \psi(\mathbf{u}) + \frac{1}{2}(\mathbf{v} - \mathbf{u})^{\mathrm{t}} A(\mathbf{v} - \mathbf{u}),$$

para todo $\mathbf{u} \in \mathbb{R}^n$, y por tanto $\Phi(\mathbf{v}) > \Phi(\mathbf{u})$, si $\mathbf{u} \neq \mathbf{v}$, es decir, $\mathbf{u}$ es el mínimo de la función Φ. Nótese que la relación anterior es equivalente a

$$\frac{1}{2}\|\mathbf{v} - \mathbf{u}\|_A^2 = \Phi(\mathbf{v}) - \Phi(\mathbf{u}), \tag{X.4.16}$$

donde $\|\cdot\|_A$ es la norma asociada al producto escalar cuya matriz respecto de la base usual de $\mathbb{R}^n$ es A.

El problema consiste, pues, en determinar el valor mínimo $\mathbf{u}$ de Φ a partir de un punto $\mathbf{u}^{(0)} \in \mathbb{R}^n$, es decir, seleccionar las direcciones apropiadas que nos permitan aproximarnos a la solución tanto como queramos. El valor óptimo de la dirección que une el punto de partida $\mathbf{u}^{(0)}$ con la solución $\mathbf{u}$ es obviamente desconocido a priori. Por consiguiente, debemos de dar un paso desde $\mathbf{u}^{(0)}$ en una dirección $\mathbf{d}^{(0)}$ que nos permita fijar un nuevo punto $\mathbf{u}^{(1)}$ desde el cual iterando este proceso alcancemos $\mathbf{u}$.

De este modo, en la etapa genérica k-ésima, $\mathbf{u}^{(k+1)}$ se calcula como

$$\mathbf{u}^{(k+1)} = \mathbf{u}^{(k)} + \alpha_k \mathbf{d}^{(k)},$$

donde α_k es el valor que fija la longitud de la dirección $\mathbf{d}^{(k)}$. La idea más natural es tomar la dirección descendiente de mayor pendiente $\nabla\Phi(\mathbf{u}^{(k)})$, lo que produce el llamada **método del gradiente**.

Por otra parte, según (X.4.15), $\nabla\Phi(\mathbf{u}^{(k)}) = A\mathbf{u}^{(k)} - \mathbf{b} = -\mathbf{r}^{(k)}$, por consiguiente, la dirección del gradiente de Φ coincide con la del vector residual y puede ser calculada usando $\mathbf{u}^{(k)}$. Esto demuestra que el método del gradiente se mueve en cada etapa k a lo largo de la dirección $\mathbf{d}^{(k)} = \mathbf{r}^{(k)}$.

Para calcular el parámetro α escribamos explícitamente $\Phi(\mathbf{u}^{(k+1)})$ como una función del parámetro α

$$\Phi(\mathbf{u}^{(k+1)}) = \frac{1}{2}(\mathbf{u}^{(k)} + \alpha\mathbf{r}^{(k)})^{\mathrm{t}} A(\mathbf{u}^{(k)} + \alpha\mathbf{r}^{(k)}) - (\mathbf{u}^{(k)} + \alpha\mathbf{r}^{(k)})^{\mathrm{t}}\mathbf{b}.$$

Derivando respecto de α e igualando a cero el resultado, obtenemos que el valor buscado de α es

$$\alpha_k = \frac{(\mathbf{r}^{(k)})^{\mathrm{t}}\mathbf{r}^{(k)}}{(\mathbf{r}^{(k)})^{\mathrm{t}} A\mathbf{r}^{(k)}} = \left(\frac{\|\mathbf{r}^{(k)}\|_2}{\|\mathbf{r}^{(k)}\|_A}\right)^2 \tag{X.4.17}$$

que depende exclusivamente del vector residual en la etapa k-ésima. Por esta razón, el método de Richardson no estacionario que emplea (X.4.17) para evaluar el parámetro de aceleración, se conoce como *método del gradiente con parámetro dinámico* o *método*

de gradiente, para distinguirlo del método de Richardson estacionario o *método del gradiente con parámetro constante*.

Resumiendo, el método del gradiente se puede describir como sigue: dado $\mathbf{u}^{(0)} \in \mathbb{R}^n$, para $k = 0, 1, \ldots$, hasta la convergencia, calcular

$$\mathbf{r}^{(k)} = \mathbf{b} - A\mathbf{u}^{(k)}$$

$$\alpha_k = \left(\frac{\|\mathbf{r}^{(k)}\|_2}{\|\mathbf{r}^{(k)}\|_A} \right)^2$$

$$\mathbf{u}^{(k+1)} = \mathbf{u}^{(k)} + \alpha_k \mathbf{r}^{(k)}.$$

Teorema X.4.4. *Sea A una matriz simétrica y definida positiva. El método del gradiente es convergente para cualquier elección del dato inicial $\mathbf{u}^{(0)}$ y*

$$\left\| \vec{\varepsilon}^{(k+1)} \right\|_A \leq \frac{\mathrm{cond}_2(A) - 1}{\mathrm{cond}_2(A) + 1} \left\| \vec{\varepsilon}_k \right\|_A, \ k = 0, 1 \ldots,$$

donde $\vec{\varepsilon}_k = \mathbf{u}^{(k)} - \mathbf{u}$ es el error cometido en cada iteración.

Demostración. Sean $\mathbf{u}^{(k)}$ las solución generada por el método del gradiente en la etapa k-ésima, y sea $\mathbf{u}_E^{(k+1)}$ igual al vector generado al aplicar el método de Richardson estacionario para $M = I_n$ con el parámetro óptimo a partir de $\mathbf{u}^{(k)}$, es decir, $\mathbf{u}^{(k)} + \alpha_{opt}\mathrm{r}^{(k)}$.

Por el corolario X.4.3 y por la igualdad (X.4.14), tenemos que[4]

$$\left\| \vec{\varepsilon}_E^{(k+1)} \right\| \leq \frac{\mathrm{cond}_2(A) - 1}{\mathrm{cond}_2(A) + 1} \left\| \vec{\varepsilon}_k \right\|$$

donde $\vec{\varepsilon}_E^{(k+1)} = \mathbf{u}_E^{(k+1)} - \mathbf{u}$. Además, por (X.4.16), tenemos que el vector $\mathbf{u}^{(k+1)}$, generado por el método del gradiente, es el que minimiza la norma $\|\cdot\|_A$ del error entre todos los vectores de la forma $\mathbf{u}^{(k)} + \gamma \mathbf{r}^{(k)}$, con $\gamma \in \mathbb{R}$. Por consiguiente, $\|\vec{\varepsilon}^{(k+1)}\|_A \leq \|\vec{\varepsilon}_E^{(k+1)}\|_A$ lo que completa la demostración. ∎

El método del gradiente consiste esencialmente en dos fases: elegir una dirección descendente $(-\mathbf{r}^{(k)})$ y seleccionar, mediante la elección del parámetro α_k, un mínimo local para Φ en esa dirección. La segunda fase es independiente de la primera, ya que, dada una dirección $\mathbf{p}^{(k)}$, podemos determinar un parámetro α_k que minimize la función $\Phi(\mathbf{u}^{(k)} + \alpha \mathbf{p}^{(k)})$.

En este sentido, una variante del método del gradiente es el **método del gradiente conjugado** que consiste esencialmente en elegir la sucesión de direcciones descendentes de la siguiente manera: $\mathbf{p}^{(0)} = \mathbf{r}^{(0)}$ y $\mathbf{p}^{(k+1)} = \mathrm{r}^{(k+1)} - \beta_k \mathbf{p}^{(k)}, \ k = 0, 1, \ldots,$

[4]Recuérdese que, cuando A es simétrica, $\mathrm{cond}_2(A) = \lambda_1/\lambda_n$, donde λ_1 y λ_n son los autovalores mayor y menor de A, respectivamente (véase la nota VIII.3.7).

de tal forma que las direcciones $\mathbf{p}^{(0)}, \ldots, \mathbf{p}^{(k+1)},\ k = 0, 1, \ldots$, sean mutuamente A-ortogonales[5] y a continuación determinar el parámetro α_k que minimize la función $\Phi(\mathbf{u}^{(k)} + \alpha \mathbf{p}^{(k)})$. La principal ventaja de este método es que ha de finalizar en n etapas (en aritmética exacta) ya que como sabemos el número máximo de vectores A-ortogonales en $\mathbb{R}^n$ es n. El lector interesado en los detalles del método del gradiente conjugado puede consultar el apartado 4.3.4 de [**QSS07**].

[5]Obsérvese que en el método de gradiente dos direcciones consecutivas $\mathbf{r}^{(k)}$ y $\mathbf{r}^{(k+1)}$ siempre son A-ortogonales.

Ejercicios del tema X

Ejercicio 1. A partir de un vector $\mathbf{u}^{(0)} \in V$ dado, se considera el método iterativo

$$\mathbf{u}^{(k+1)} = B\mathbf{u}^{(k)} + \mathbf{c}.$$

Estudiar el comportamiento de la sucesión $(\mathbf{u}^{(k)})$ cuando $\varrho(B) = 0$.

Ejercicio 2. Sea $A \in \mathcal{M}_n(\Bbbk)$ una matriz triangular superior. Estudiar la convergencia de los métodos de Jacobi, Gauss-Seidel y de relajación para A. Idem si A es triangular inferior.

Ejercicio 3. Demostrar que si $A = (a_{ij}) \in \mathcal{M}_n(\Bbbk)$ verifica

$$|a_{jj}| > \sum_{\substack{i=1 \\ i \neq j}} |a_{ij}|$$

para $j = 1, \ldots, n$, entonces el método de Jacobi para A es convergente.

Ejercicio 4. Analizar las propiedades de convergencia de los métodos de Jacobi y Gauss-Seidel para la resolución de un sistema lineal cuya matriz es

$$A_\epsilon = \begin{pmatrix} \epsilon & 0 & 1 \\ 0 & \epsilon & 0 \\ 1 & 0 & \epsilon \end{pmatrix},$$

con $\epsilon \in \mathbb{R}$.

Ejercicio 5. Proporcionar una condición suficiente sobre β tal que los métodos de Jacobi y Gauss-Seidel converjan cuando se apliquen a la resolución de un sistema cuya matriz es

$$A = \begin{pmatrix} -10 & 2 \\ \beta & 5 \end{pmatrix}.$$

Ejercicio 6. Sea $A \in \mathcal{M}_n(\Bbbk)$. Probar que

1. si $0 < \alpha < 1$ y $\lambda \in \Bbbk$ con $|\lambda| \geq 1$, entonces
$$\left| \frac{1 - \alpha - \lambda}{\lambda \alpha} \right| \geq 1;$$
2. si A es de diagonal estrictamente dominante, el método de relajación para A es convergente si $0 < \alpha \leq 1$.

Ejercicio 7. Sea $A \in \mathcal{M}_n(\Bbbk)$ una matriz hermítica e invertible descompuesta en la forma $A = M - N$ con M invertible.

1. Se considera la sucesión
$$\mathbf{u}^{(k+1)} = M^{-1}N\mathbf{u}^{(n)}$$
con $\mathbf{u}^{(0)} \in V \setminus \{0\}$ arbitrario. Probar que si la matriz $M^* + N$ es definida positiva, entonces la sucesión $((\mathbf{u}^{(k)})^* A\mathbf{u}^{(k)})$ es monótona creciente.
2. Demostrar que si $M^* + N$ es definida positiva y $\varrho(M^{-1}N) < 1$, entonces A es definida positiva.

Ejercicio 8. Construir matrices para las cuales el método de Jacobi asociado sea convergente y el método de Gauss-Seidel diverja y recíprocamente.

Ejercicio 9. Sean A, M y $M^{-1}A$ matrices simétricas definidas positivas. Probar que el método de Richardson estacionario es convergente y
$$\left\|\vec{\varepsilon}^{(k+1)}\right\|_A \leq \rho(B_\alpha)\left\|\vec{\varepsilon}\right\|_A,\ k \geq 0.$$

TEMA XI

Métodos iterativos para el cálculo de autovalores (y autovectores)

En este tema damos una breve semblanza de los métodos iterativos para el cálculo de los autovalores (y autovectores) de una matriz. Una observación previa digna a tener en cuenta es que debido a la imposibilidad de resolver por radicales de forma exacta una ecuación polinómica de grado mayor o igual que 5 (Teorema de Abel) es igualmente imposible calcular los autovalores de una matriz de orden mayor o igual que 5 mediante métodos directos, al tratarse esencialmente del mismo problema.

Para calcular aproximaciones numéricas de los autovalores de una matriz $A \in \mathcal{M}_n(\Bbbk)$, se suele construir una sucesión de matrices $U_k^{-1}AU_k$ convergente a una matriz cuyos autovalores sean conocidos, por ejemplo, a una matriz diagonal o a una triangular.

Esta idea es la base del *método de Jacobi* que estudiaremos en la primera sección del tema. Este método se emplea cuando buscamos aproximaciones de *todos los autovalores, y* (eventualmente) *todos los autovectores* de una matriz simétrica[1]. Las matrices U_k consideradas serán productos de matrices ortogonales elementales muy fáciles de construir. En este caso, demostraremos que

$$\lim_{k\to\infty} U_k^{-1}AU_k = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n),$$

donde los números reales λ_i son los autovalores de la matriz A. Además, cuando estos últimos son todos distintos, veremos que cada una de las columnas de las matrices U_k forma una sucesión convergente de vectores que converge a un autovector de la matriz A.

En general, para cualquier tipo de matrices, el método QR revela la misma idea. Utilizando en cada iteración la factorización QR de la matriz $U_k^{-1}AU_k$ obtenida, se obtiene un método iterativo general (y no sólo válido para las matrices simétricas). En la segunda sección sólamente consideraremos el caso de las matrices reales con todos sus autovalores reales, que viene a ser completamente análogo al caso de las matrices complejas. En todo caso, conviene advertir que esta condición es puramente

[1] Recuérdese que las matrices simétricas son diagonalizables con matriz de paso ortogonal (véase el teorema V.5.3).

técnica a afecto de simplificar nuestra breve introducción a los métodos iterativos para el cálculo de autovalores. Así mismo, al final de la sección mostramos cómo se pueden calcular los autovalores bajo ciertas condiciones.

En la última sección estudiamos el método de la potencia para el cálculo de autovalores y autovectores, aunque quizá sería más apropiado decir, para el cálculo de un autovalor y un autovector, ya que este método se caracteriza por su eficiencia a la hora de calcular el autovalor de mayor o menor módulo. Esto es a menudo suficiente si lo que nos interesa es conocer el radio espectral de una matriz dada. La sección finaliza con un pequeño análisis sobre la convergencia del método de la potencia y mostrando un método recurrente para el cálculo de pares autovalor/autovector a partir de pares previamente calculados.

Este tema se ha elaborado a partir del capítulo 4 de [**QSS07**] y usando también algunos aspectos del capítulo 6 de [**Cia82**]. Como se ha comentado, en este tema sólo hemos entreabierto la puerta al estudio de los métodos iterativos para el cálculo de autovalores y autovectores. El lector interesado en profundizar en este tema puede comenzar con el capítulo 6 de [**QS06**].

1. El método de Jacobi

Partiendo de una matriz simétrica $A_1 := A \in \mathcal{M}_n(\mathbb{R})$ el método de Jacobi consiste en construir una sucesión $(Q_k)_{k\in\mathbb{N}}$ de matrices ortogonales "elementales" (en un cierto sentido que se determinará en breve) tales que la sucesión de matrices (también simétricas)

$$A_{k+1} := Q_k^{\mathrm{t}} A_k Q_k = (Q_1 Q_2 \cdots Q_k)^{\mathrm{t}} A (Q_1 Q_2 \cdots Q_k),\ k \geq 1,$$

sea convergente a la matriz $D = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$, salvo permutación de los subíndices. Además, en ciertos casos, se puede concluir que la sucesión de matrices ortogonales

$$U_k := Q_1 Q_2 \cdots Q_k,\ k \geq 1, \tag{XI.1.1}$$

converge a una matriz ortogonal cuyas columnas forman una base ortonormal de autovectores de la matriz A.

El principio de cada transformación

$$A_k \longrightarrow A_{k+1} = Q_k^{\mathrm{t}} A_k Q_k,\ k \geq 1,$$

consiste en anular dos elementos extradiagonales de la matriz A_k en posición simétrica, $(A_k)_{pq}$ y $(A_k)_{qp}$, siguiendo un proceso bastante simple que vamos a describir y a estudiar a continuación. Por el momento no nos preocuparemos de la elección *efectiva* de la pareja (p, q).

Comenzamos con un lema técnico que es la clave del método de Jacobi.

Lema XI.1.1. *Sean p y q dos números enteros tales que $1 \leq p < q \leq n$, θ un número real y*

$$Q = I_n + R, \tag{XI.1.2}$$

donde $R \in \mathcal{M}_n(\mathbb{R})$ tiene como entrada (i,j)-ésima a

$$r_{ij} = \begin{cases} \cos(\theta) - 1 & si \quad i = j = p \ \acute{o} \ i = j = q \\ \operatorname{sen}(\theta) & si \quad i = p \quad y \quad j = q \\ -\operatorname{sen}(\theta) & si \quad i = q \quad y \quad j = p \\ 0 & en \ otro \ caso. \end{cases}$$

Si $A = (a_{ij}) \in \mathcal{M}_n(\mathbb{R})$ es una matriz simétrica y $B = Q^{\mathrm{t}}AQ = (b_{ij}) \in \mathcal{M}_n(\mathbb{R})$, entonces

(a) *B es simétrica y $\|B\|_F = \|A\|_F$, es decir,*

$$\sum_{i,j=1}^{n} b_{ij}^2 = \sum_{i,j=1}^{n} a_{ij}^2.$$

(b) *si $a_{pq} \neq 0$, existe un único valor de θ en $(-\pi/4, 0) \cup (0, \pi/4]$ tal que $b_{pq} = 0$; tal valor es la única solución de la ecuación*

$$\operatorname{cotan}(2x) = \frac{a_{pp} - a_{qq}}{2a_{pq}}$$

en $(-\pi/4, 0) \cup (0, \pi/4]$. Además, para este valor de θ se cumple que

$$\sum_{i=1}^{n} b_{ii}^2 = \sum_{i=1}^{n} a_{ii}^2 + 2a_{pq}^2.$$

Demostración. (a) Es claro que B es simétrica, pues

$$B^{\mathrm{t}} = (Q^{\mathrm{t}}AQ)^{\mathrm{t}} = Q^{\mathrm{t}}A^{\mathrm{t}}Q = Q^{\mathrm{t}}AQ = B.$$

Por otra parte, se comprueba fácilmente que la matriz Q es ortogonal; luego, en particular, es unitaria. Ahora, como la norma de Fröbenius es invariante por transformaciones unitarias (véase la proposición VIII.2.14), se sigue que

$$\sum_{i,j=1}^{n} b_{ij}^2 = \|B\|_F = \|Q^{\mathrm{t}}AQ\|_F = \|A\|_F = \sum_{i,j=1}^{n} a_{ij}^2.$$

(b) La transformación de los elementos de índices (p,p), (p,q), (q,p) y (q,q), se puede escribir de la siguiente manera

$$\begin{pmatrix} b_{pp} & b_{pq} \\ b_{qp} & b_{qq} \end{pmatrix} = \begin{pmatrix} \cos(\theta) & -\operatorname{sen}(\theta) \\ \operatorname{sen}(\theta) & \cos(\theta) \end{pmatrix} \begin{pmatrix} a_{pp} & a_{pq} \\ a_{qp} & a_{qq} \end{pmatrix} \begin{pmatrix} \cos(\theta) & \operatorname{sen}(\theta) \\ -\operatorname{sen}(\theta) & \cos(\theta) \end{pmatrix},$$

de tal forma que el mismo razonamiento que el apartado (a) nos permite asegurar que

$$b_{pp}^2 + b_{qq}^2 + 2b_{pq}^2 = a_{pp}^2 + a_{qq}^2 + 2a_{pq}^2,$$

para todo valor de θ.

Por otra parte, como $b_{pq} = b_{qp}$ es

$$a_{pq}\cos(2\theta) + \frac{a_{pp} - a_{qq}}{2}\operatorname{sen}(2\theta),$$

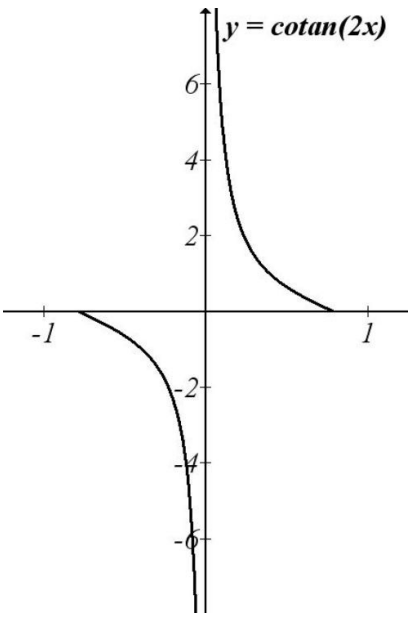
se sigue que si θ se pudiese elegir tal y como se indica en el enunciado, tendríamos que

$$b_{pq} = b_{qp} = 0$$

y por lo tanto que

$$b_{pp}^2 + b_{qq}^2 = a_{pp}^2 + a_{qq}^2 + 2a_{pq}^2.$$

Pero, tal valor de θ siempre existe y es único ya que la función $y = \operatorname{cotan}(2x)$ es continua y estrictamente decreciente en los intervalos $(-\pi/4, 0)$ y $(0, \pi/4]$, y su imagen es $(-\infty, 0)$ en el primer intervalo y $[0, +\infty)$ en el segundo.



Luego, la función

$$y = \operatorname{cotan}(2x) - \frac{a_{qq} - a_{pp}}{2a_{pq}}$$

corta al eje OX en un único punto.

Finalmente, como $a_{ii} = b_{ii}$ para todo $i \neq p$ e $i \neq q$, concluimos que

$$\sum_{i=1}^{n} b_{ii}^2 = \sum_{i=1}^{n} a_{ii}^2 + 2a_{pq}^2.$$

■

Nota XI.1.2.

i) La matriz Q es ortogonal para todo $\theta \in \mathbb{R}$.

ii) Solamente las filas y columnas p-ésima y q-ésima de la matriz A son modificadas por la transformación $A \to B = Q^{\mathrm{t}}AQ$. De forma más precisa, para

todo $\theta \in \mathbb{R}$ se tiene que $b_{ij} = b_{ji}$ es igual a

$$\begin{cases} a_{ij} & \text{si } i \neq p,q \text{ y } j \neq p,q \\ a_{pj}\cos(\theta) - a_{qj}\,\mathrm{sen}(\theta) & \text{si } i = p \text{ y } j \neq p,q \\ a_{pj}\,\mathrm{sen}(\theta) + a_{qj}\cos(\theta) & \text{si } i = q \text{ y } j \neq p,q \\ a_{pp}\cos^2(\theta) + a_{qq}\,\mathrm{sin}^2(\theta) - a_{pq}\mathrm{sin}(2\theta) & \text{si } i = j = p \\ a_{pp}\,\mathrm{sin}^2(\theta) + a_{qq}\cos^2(\theta) + a_{pq}\mathrm{sin}(2\theta) & \text{si } i = j = q \\ a_{pq}\cos(2\theta) + \frac{a_{pp}-a_{qq}}{2}\,\mathrm{sin}(2\theta) & \text{si } i = p \text{ y } j = q \end{cases},$$

para todo $\theta \in \mathbb{R}$.

iii) Gracias a las relaciones existentes entre las funciones trigonométricas, los elementos de la matriz B son, a pesar de las apariencias, determinados por *relaciones algebraicas* obtenidas a partir de los elementos de A; para ello, calculemos los siguientes números reales:

$$x_0 = \frac{a_{qq} - a_{pp}}{2a_{pq}}\ (= \mathrm{cotan}(2\theta)),$$

$$t_0 = \begin{cases} \left.\begin{array}{c}\text{la raíz de menor módulo} \\ \text{del polinomio } t^2 + 2x_0 t - 1\end{array}\right\} & \text{si } x_0 \neq 0 \\ 1 & \text{si } x_0 = 0 \end{cases}$$

es decir, $t_0 = \tan(\theta)$ con $|\theta| \leq \pi/4$, y finalmente,

$$c = \frac{1}{\sqrt{1+t_0^2}} (= \cos(\theta))$$

$$s = \frac{t_0}{\sqrt{1+t_0^2}} (= \mathrm{sen}(\theta)).$$

La fórmula dadas en ii) para los elementos de B se pueden escribir de la forma siguiente

$$b_{ij} = b_{ji} = \begin{cases} a_{ij} & \text{si } i \neq p,q \text{ y } j \neq p,q \\ a_{pj}\,c - a_{qj}\,s & \text{si } i = p \text{ y } j \neq p,q \\ a_{pj}\,s + a_{qj}\,c & \text{si } i = q \text{ y } j \neq p,q \\ a_{pp} - a_{pq}\,t_0 & \text{si } i = j = p \\ a_{qq} + a_{pq}\,t_0 & \text{si } i = j = q \\ 0 & \text{si } i = p \text{ y } j = q \end{cases},$$

cuando el valor de θ es la única solución de la ecuación

$$\mathrm{cotan}(2x) = \frac{a_{pp} - a_{qq}}{2a_{pq}}$$

en $(-\pi/4, 0) \cup (0, \pi/4]$.

Ahora ya estamos en disposición de describir la etapa k-ésima del método de Jacobi.

Proposición XI.1.3. *Dada la matriz* $A_k = (a_{ij}^{(k)}) \in \mathcal{M}_n(\mathbb{R})$ *y fijado un par* (p, q) *con* $p \neq q$ *tal que* $a_{pq}^{(k)} \neq 0$, *se puede construir una matriz ortogonal* $Q_k \in \mathcal{M}_n(\mathbb{R})$ *tal que*

$$A_{k+1} = Q_k^{\mathrm{t}} A_k Q_k$$

con $a_{pq}^{(k+1)} = a_{qp}^{(k+1)} = 0$. *En particular,* $\mathrm{sp}(A_{k+1}) = \mathrm{sp}(A_k)$.

Demostración. Por el lema XI.1.1 basta tomar Q_k de la forma (XI.1.2) con $\theta \in (-\pi/4, 0) \cup (0, \pi/4]$ verificando la ecuación

$$\mathrm{cotan}(2x) = \frac{a_{pp}^{(k)} - a_{qq}^{(k)}}{2a_{pq}^{(k)}}.$$

■

A continuación distinguiremos tres estrategias para la elección de la pareja (p, q).

Método de Jacobi clásico. La pareja (p, q) se elige de tal forma que

$$|a_{pq}^{(k)}| = \max_{i \neq j} |a_{ij}^{(k)}|.$$

Entiéndase que la elección pareja (p, q) va variando en cada una de las etapas, es decir, depende de k.

La principal desventaja del método de Jacobi clásico es el coste en tiempo que supone la búsqueda del elemento extradiagonal de mayor absoluto en la matriz A_k.

Método de Jacobi cíclico. En este caso vamos recorriendo todos los elementos extradiagonales mediante un *barrido cíclico*, sucesivamente aunque usando siempre el mismo; por ejemplo, elegimos las parejas (p, q) con el siguiente orden

$$(1,2), (1,3), \ldots, (1,n);\ (2,3), \ldots, (2,n); \ldots, (n-1,n).$$

Naturalmente, si en la etapa k-ésima el elemento $a_{pq}^{(k)}$ es cero, pasamos al siguiente (desde el punto de vista matricial esto equivale a tomar $Q_k = I_n$).

Método de Jacobi con umbral. Procedemos como en el método de Jacobi cíclico, pero *saltándonos* aquellas parejas (p, q) tales que $|a_{p,q}^{\bullet}| < \varepsilon$, para un cierto número real $\varepsilon > 0$ dado; pues parece inútil anular aquellos elementos extradiagonales cuyo valor absoluto sea muy pequeño, mientras existan otro elementos de orden elevado.

Nota XI.1.4. Independientemente de la estrategia (e incluso del método) elegida, es muy importante tener en cuenta que los elementos anulados en una etapa dada puede ser reemplazados por elementos no nulos en una etapa posterior. En otro caso, obtendríamos que la reducción a una matriz diagonal se podría realizar en un número finito de iteraciones, lo que no es posible en general.

Análisis de convergencia.

A continuación vamos a estudiar la convergencia del método de Jacobi, aunque nos restringiremos al caso más sencillo (es decir, al método clásico) y sin preocuparnos por la estimación de errores. En la página 114 de [**Cia82**] se pueden encontrar las referencias a algunos trabajos de P. Henrici y de H.P.M van Kempen realizados entre 1958 y 1968 sobre la convergencia de los métodos de Jacobi clásico y cíclico.

Sea $A \in \mathcal{M}_n(\mathbb{R})$ una matriz simétrica y $(A_k)_{k\in\mathbb{N}} \subset \mathcal{M}_n(\mathbb{R})$ la sucesión de matrices simétricas obtenidas mediante la aplicación del método de Jacobi clásico. Al igual que antes, denotaremos $a_{ij}^{(k)}$ a la entrada (i,j)-ésima de la matriz A_k. Para evitar situaciones triviales, a partir de ahora supondremos que $\max_{i\neq j} |a_{ij}^{(k)}| > 0$, para todo $k \geq 1$.

Como es habitual, designaremos por S_n al conjunto de todas las permutaciones del conjunto $\{1, 2, \ldots, n\}$, esto es el *grupo simétrico n-ésimo.*

Antes de demostrar el teorema de convergencia de los autovalores para el método de Jacobi clásico, necesitamos recordar el siguiente resultado sobre espacios normados que desempeñará un papel crucial en las demostraciones de los dos teoremas siguientes.

Lema XI.1.5. *Sea $(V, \|\cdot\|)$ un espacio normado de dimensión finita. Si $(\mathbf{v}_n)_{n\in\mathbb{N}} \subset V$ es una sucesión acotada tal que*

(a) *$(\mathbf{v}_n)_{n\in\mathbb{N}}$ posee un número finito de puntos de acumulación,*

(b) $\lim_{n\to\infty} \|\mathbf{v}_{n+1} - \mathbf{v}_n\| = 0$.

entonces la sucesión $(\mathbf{v}_n)_{n\in\mathbb{N}}$ es convergente (a un único punto de acumulación).

Demostración. La demostración se propone como ejercicio a lector. ∎

Teorema XI.1.6. *Con la notación anterior, la sucesión $(A_k)_{k\in\mathbb{N}}$ es convergente, y*

$$\lim_{k\to\infty} A_k = \operatorname{diag}(\lambda_{\sigma(1)}, \lambda_{\sigma(2)}, \ldots, \lambda_{\sigma(n)})$$

para alguna permutación $\sigma \in S_n$, siendo $\lambda_1, \lambda_2, \ldots, \lambda_n \in \mathbb{R}$ los autovalores de A.

Demostración. Dado un entero $k \geq 1$, escribiremos

$$A_k = (a_{ij}^{(k)}) = D_k + C_k$$

con $D_k := \operatorname{diag}(a_{11}^{(k)}, a_{22}^{(k)}, \ldots, a_{nn}^{(k)})$.

Demostremos en primer lugar que $\lim_{k\to\infty} C_k = 0$.

Los números

$$\varepsilon_k := \sum_{i\neq j} |a_{ij}^{(k)}|^2 = \|\!| C_k |\!\|_F^2, \quad k \geq 1,$$

verifican, por el lema XI.1.1(b), que

$$\varepsilon_{k+1} = \varepsilon_k + 2|a_{pq}^{(k)}|^2,$$

y, por la estrategia adoptada por el método de Jacobi clásico, que $\varepsilon_k \leq n(n-1)|a_{pq}^{(k)}|^2$, ya que hay $n(n-1)$ elementos extradiagonales. Combinando estas expresiones, se obtiene que

$$\varepsilon_{k+1} \leq \left(1 - \frac{2}{n(n-1)}\right)\varepsilon_k,$$

de donde se sigue que $\lim_{k\to\infty} \varepsilon_k = 0$.

Según lo anterior, como $A_k = D_k + C_k,\ k \geq 1$, se tiene que $\lim_{k\to\infty} A_k = \lim_{k\to\infty} D_k$. De modo que basta demostrar que la sucesión (D_k) es convergente a $\mathrm{diag}(\lambda_{\sigma(1)}, \lambda_{\sigma(2)}, \ldots, \lambda_{\sigma(n)})$ para alguna permutación $\sigma \in S_n$, y habremos terminado.

En primer lugar, observamos que la sucesión (D_k) es acotada. En efecto, por el lema XI.1.1, $\|A_k\|_F = \|A\|_F$; luego,

$$\|D_k\|_F \leq \|A_k\|_F = \|A\|_F,$$

para todo $k \geq 1$.

Veamos ahora que la sucesión $(D_k)_{k\in\mathbb{N}}$ tiene un número finito de puntos de acumulación, que han de ser de la forma $\mathrm{diag}(\lambda_{\sigma(1)}, \lambda_{\sigma(2)}, \ldots, \lambda_{\sigma(n)})$ para algún $\sigma \in S_n$.

Si $(D'_k)_{k\in\mathbb{N}}$ es una subsucesión de $(D_k)_{k\in\mathbb{N}}$ convergente a una matriz D, entonces se tiene que

$$\lim_{k\to\infty} A'_k = D \quad \text{con} \quad A'_k = D'_k + C'_k \quad \text{y} \quad \lim_{k\to\infty} C'_k = 0,$$

de modo que, considerando los coeficientes de los polinomios característicos, se tiene que

$$\aleph_D(x) = \det(D - xI_n) = \lim_{k\to\infty} \det(A'_k - xI_n) = \lim_{k\to\infty} \aleph_{A'_k}(x).$$

Pero, como

$$\det(A_k - xI_n) = \det(A - xI_n),$$

para todo k pues $\mathrm{sp}(A_k) = \mathrm{sp}(A)$, concluimos que las matrices A y $D = \lim_{k\to\infty} D'_k$ tienen los mismos autovalores con idénticas multiplicidades. Por consiguiente, como D es una matriz diagonal (por ser límite de una sucesión de matrices diagonales), existe una permutación $\sigma \in S_n$ tal que

$$D = \mathrm{diag}(\lambda_{\sigma(1)}, \lambda_{\sigma(2)}, \ldots, \lambda_{\sigma(n)}).$$

La siguiente etapa en nuestra demostración consiste en ver que $\lim_{k\to\infty}(D_{k+1} - D_k) = 0$. Para ello, observamos que

$$a_{ii}^{(k+1)} - a_{ii}^{(k)} = \begin{cases} 0 & \text{si } i \neq p, q \\ -\tan(\bar{\theta}_k)\, a_{pq}^{(k)} & \text{si } i = p \\ \tan(\bar{\theta}_k)\, a_{pq}^{(k)} & \text{si } i = q \end{cases}$$

Como

$$|\bar{\theta}_k| \leq \frac{\pi}{4} \quad \text{y} \quad |a_{pq}^{(k)}| \leq \|C_k\|_F$$

se concluye que $\lim_{k\to}(D_{k+1} - D_k) = 0$, al ser $\lim_{k\to\infty} B_k = 0$.

De todo lo anterior, por el lema XI.1.5, se sigue que la sucesión $(D_k)_{k\in\mathbb{N}}$ es convergente, y necesariamente $\lim_{k\to\infty} D_k = \operatorname{diag}(\lambda_{\sigma(1)}, \lambda_{\sigma(2)}, \ldots, \lambda_{\sigma(n)})$, para alguna permutación $\sigma \in S_n$. ∎

Terminamos esta sección mostrando un resultado sobre la convergencia del método de Jacobi para el cálculo de aproximaciones de los autovectores de una matriz simétrica con todos sus autovalores distintos. En primer lugar, recordemos que

$$A_{k+1} = Q_k^{\mathrm{t}} A_k Q_k = Q_k^{\mathrm{t}} Q_{k-1}^{\mathrm{t}} A_{k-1} Q_{k-1} Q_k = \ldots = U_k^{\mathrm{t}} A U_k,$$

donde $U_k = Q_1 Q_2 \cdots Q_k$.

Teorema XI.1.7. *Con la notación anterior, si todos los autovalores de la matriz A son distintos, entonces la sucesión $(U_k)_{k\in\mathbb{N}}$ de matrices ortogonales converge a una matriz cuyas columnas forman un sistema ortonormal de autovectores de A.*

Demostración. En primer lugar, como todas las matrices U_k son ortogonales (y, en particular, unitarias) se tiene que $\|U_k\|_2 = 1$. Luego, la sucesión $(U_k)_{k\in\mathbb{N}}$ es acotada.

Veamos que la sucesión (U_k) tiene un número finito de puntos de acumulación, que han de ser de la forma

$$\left(\mathbf{v}_{\sigma(1)} | \mathbf{v}_{\sigma(2)} | \ldots | \mathbf{v}_{\sigma(n)}\right) \in \mathcal{M}_n(\mathbb{R}), \quad \sigma \in S_n,$$

donde $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n \in \mathbb{R}^n$ son los vectores columna de la matriz ortogonal $Q \in \mathcal{M}_n(\mathbb{R})$ dada por la relación

$$Q^{\mathrm{t}} A Q = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n).$$

Sea $(U'_k)_{k\in\mathbb{N}}$ una subsucesión de $(U_k)_{k\in\mathbb{N}}$ convergente a una matriz (ortogonal) U'. Según el teorema anterior, existe una permutación $\sigma \in S_n$ tal que

$$\operatorname{diag}(\lambda_{\sigma(1)}, \lambda_{\sigma(2)}, \ldots, \lambda_{\sigma(n)}) = \lim_{k\to\infty} A'_k = \lim_{k\to\infty} ((U'_k)^{\mathrm{t}} A U'_k) = (U')^{\mathrm{t}} A U',$$

lo cual demuestra nuestro aserto. Obsérvese que la hipótesis referente a que los autovalores de A son todos distintos se utiliza como hecho esencial para concluir la existencia de un número finito de puntos de acumulación.

Finalmente demostremos que $\lim_{k\to\infty} U_{k+1} - U_k = 0$. Por construcción, θ_k verifica

$$\tan(2\theta_k) = \frac{2a_{pq}^{(k)}}{a_{qq}^{(k)} - a_{pp}^{(k)}}, \quad |\theta_k| \leq \frac{\pi}{4}.$$

Usando el teorema anterior y de nuevo el hecho de que todos los autovalores de A son distintos, concluimos la existencia de un entero l tal que

$$k \geq l \Rightarrow |a_{qq}^{(k)} - a_{pp}^{(k)}| \geq \frac{1}{2}\min|\lambda_i - \lambda_j| > 0$$

(como las parejas (p, q) varían en cada etapa m, no podemos afirmar que las sucesiones $(a_{pp}^{(k)})_{k\in\mathbb{N}}$ y $(a_{qq}^{(k)})_{k\in\mathbb{N}}$ sea convergentes). Sin embargo, como $\lim_{k\to\infty} a_{pq}^{(k)} = 0$, tenemos que

$$\lim_{k\to 0} \theta_k = 0, \quad \text{y por tanto que} \quad \lim_{k\to\infty} Q_k = I_n$$

(recuérdese que la expresión dada de la matriz Q_k depende de θ). Por consiguiente,

$$U_{k+1} - U_k = U_k(Q_{k+1} - I_n),$$

de donde si sigue que $\lim_{k\to\infty} U_{k+1} - U_k = 0$ al ser $(U_k)_{k\in\mathbb{N}}$ una sucesión acotada.

Ahora ya tenemos todos los ingredientes necesarios para aplicar el lema XI.1.5 y terminar la demostración. ∎

2. El método QR

En esta sección mostraremos el método QR para calcular aproximaciones de los autovalores y los autovectores de una matriz cuadrada con entradas reales que tenga todos sus autovalores en $\mathbb{R}$. El caso de las matrices con entradas complejas es esencialmente similar, sólo que habría que adaptar la factorización QR a este caso, tomando Q unitaria en vez de ortogonal. El lector interesado en conocer los detalles del caso complejo puede consultar la sección 6.3 de [**Cia82**].

Sea $A \in \mathcal{M}_n(\mathbb{R})$. Dada una matriz ortogonal $Q_0 \in \mathcal{M}_n(\mathbb{R})$ definimos $T_0 = Q_0^{\mathrm{t}} A Q_0$. Para cada $k = 1, 2, \ldots$, el **método QR** consiste en:

(XI.2.3)
$$\begin{aligned} &\text{determinar } Q_k \text{ y } R_k \text{ tales que} \\ &Q_k R_k = T_{k-1} \quad \text{(factorización QR);} \\ &\text{entonces, sea} \\ &T_k = R_k Q_k \end{aligned}$$

En cada etapa $k \geq 1$, la primera fase del método es la factorización QR de la matriz $T^{(k-1)}$ (véase el teorema IX.4.5). La segunda fase es simplemente el cálculo de un

producto de matrices. Obsérvese que

$$\begin{aligned} T_k = R_k Q_k = Q_k^{\mathrm{t}}(Q_k R_k) Q_k = Q_k^{\mathrm{t}} T_{k-1} Q_k = \dots \\ = (Q_0 Q_1 \cdots Q_k)^{\mathrm{t}} A (Q_0 Q_1 \cdots Q_k), \quad k \geq 0, \end{aligned}$$

es decir, T_k es congruente con A con matriz de paso ortogonal. Esto es particularmente interesante para garantizar la estabilidad del método, ya que el número de condición de T_k no será peor que el de A (véase la la nota VIII.3.7(d)).

Una implementación básica del método QR consiste en tomar Q_0 igual a la matriz identidad de orden n, de tal forma que $T_0 = A$. En cada etapa $k \geq 1$ la factorización QR de la matriz $T^{(k-1)}$ se puede calcular usando el algoritmo descrito en el teorema IX.4.5, cuyo coste computacional es del orden de $2n^3$ operaciones. En el capítulo 5 de [**QSS07**] se pueden encontrar otras implementaciones, así como variantes, del método QR. En el caso que nos ocupa, $Q_0 = I_n$, se tiene el siguiente resultado de convergencia:

Proposición XI.2.1. *Sea $A \in \mathcal{M}_n(\mathbb{R})$ invertible y tal que sus autovalores son reales y son diferentes en módulo $|\lambda_1| > |\lambda_2| > \ldots > |\lambda_n|$. Entonces*

$$\lim_{k\to\infty} T_k = \begin{pmatrix} \lambda_1 & t_{12} & \dots & t_{1n} \\ 0 & \lambda_2 & \dots & t_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{pmatrix}.$$

Además, si A es simétrica la sucesión $\{T_k\}_{k\in\mathbb{N}}$ tiende a una matriz diagonal.

Demostración. Para la demostración, véase el teorema 6.3-1 de [**Cia82**]. ∎

Las hipótesis de la proposición anterior puede verificarse *a priori* usando los círculos de Gerhsgorin (véase la sección 5.1 de [**QSS07**] o el apartado 6.3 de [**QS06**]). No obstante, si los autovalores, aún siendo distintos, no están *bien separados*, puede ocurrir que la convergencia sea demasiado lenta, ya que $|t_{i,i-1}^{(k)}|$ es del orden de $|\lambda_i/\lambda_{i-1}|^k$, $i = 2, \dots, n$, para k suficientemente alto (véase la Propiedad 5.9 de [**QSS07**]).

Supongamos ahora que tenemos una aproximación de la igualdad $Q^{\mathrm{t}}AQ = T$ siendo T triangular superior. Entonces, si $A\mathbf{x} = \lambda\mathbf{x}$, se tiene que $Q^{\mathrm{t}}AQQ^{\mathrm{t}}(\lambda\mathbf{x})$, es decir, tomando $\mathbf{y} = Q^{\mathrm{t}}\mathbf{x}$, se cumple que $T\mathbf{y} = \lambda\mathbf{y}$. Por tanto, $\mathbf{y}$ es un autovector de T, luego para calcular los autovalores de A podemos trabajar directamente con la matriz T.

Supongamos por simplicidad que $\lambda = t_{kk} \in \mathbb{C}$ es un autovalor simple de A. entonces la matriz triangular superior T se puede descomponer como

$$T = \begin{pmatrix} T_{11} & \mathbf{v} & T_{13} \\ 0 & \lambda & \mathbf{w}^{\mathrm{t}} \\ 0 & 0 & T_{33} \end{pmatrix},$$

donde $T_{11} \in \mathcal{M}_{k-1}(\mathbb{C})$ y $T_{33} \in \mathcal{M}_{n-k}(\mathbb{C})$ son matrices triangulares superiores, $\mathbf{v} \in \mathbb{C}^{k-1}, \mathbf{w} \in \mathbb{C}^{n-k}$ y $\lambda \notin \mathrm{sp}(T_{11}) \cup \mathrm{sp}(T_{33})$.

De esta forma tomando $\mathbf{y} = (\mathbf{y}_{k-1}^{\mathrm{t}}, y, \mathbf{y}_{n-k}^{\mathrm{t}})$, con $\mathbf{y}_{k-1}^{\mathrm{t}} \in \mathbb{C}^{k-1}$, $y \in \mathbb{C}$ e $\mathbf{y}_{n-k}^{\mathrm{t}} \in \mathbb{C}^{n-k}$, el sistema homogéneo $(T - \lambda I_n)\mathbf{y} = \mathbf{0}$ se puede escribir como

$$\left\{ \begin{array}{rcl} (T_{11} - \lambda I_{k-1})\mathbf{y}_{k-1} + \mathbf{v}y + T_{13}\mathbf{y}_{n-k} & = & \mathbf{0} \\ \mathbf{w}^{\mathrm{t}}\mathbf{y}_{n-k} & = & 0 \\ (T_{33} - \lambda I_{n-k})\mathbf{y}_{n-k} & = & \mathbf{0} \end{array} \right.$$

Como λ tiene multiplicidad 1, las matrices $T_{11} - \lambda I_{k-1}$ y $T_{33} - \lambda I_{n-k}$ son invertibles, por consiguiente $\mathbf{y}_{n-k} = \mathbf{0}$ y la primera ecuación se transforma en

$$(T_{11} - \lambda I_{k-1})\mathbf{y}_{k-1} = -\mathbf{v}y.$$

De donde se sigue, tomando $y = 1$ que una solución del sistema triangular anterior es

$$y = \begin{pmatrix} -(T_1 1 - \lambda I_{k-1})^{-1}\mathbf{v} \\ 1 \\ 0 \end{pmatrix}.$$

El autovector $\mathbf{x}$ buscado es, por tanto, $\mathbf{x} = Q\mathbf{y}$.

3. El método de la potencia

Sea $A \in \mathcal{M}_n(\mathbb{C})$ una matriz diagonalizable. Supongamos que los autovalores de A están ordenados como sigue

$$|\lambda_1| > |\lambda_2| \geq \ldots \geq |\lambda_n|. \tag{XI.3.4}$$

Nótese que, en particular, $|\lambda_1|$ es distinto de los otros módulos de los autovalores de A, es decir, que λ_1 es el autovalor dominante de A.

Sea $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ una base de $\mathbb{C}^n$ tal que $\mathbf{u}_j$ es un autovector (de norma usual 1, es decir $\|\mathbf{u}_j\|_2 = \sqrt{\mathbf{u}_j^* \mathbf{u}_j}$) asociado a λ_j, $j = 1, \ldots, n$ y denotemos por P a la matriz de orden n cuya columna j-ésima es $\mathbf{u}_j$. Obsérvese que para garantizar la existencia de una base de $\mathbb{C}^n$ de autovectores de A es fundamental que A sea diagonalizable (véase el teorema III.3.4).

Dado un vector inicial arbitrario $\mathbf{q}^{(0)} \in \mathbb{C}^n$ de norma usual 1, consideremos para $k = 1, 2, \ldots$, la siguiente iteración basada en el cálculo de potencias de matrices,

comúnmente llamado el **método de la potencia**:

$$\begin{aligned} \mathbf{z}(k) &= A\mathbf{q}^{(k-1)} \\ \mathbf{q}^{(k)} &= \mathbf{z}^{(k)}/\|\mathbf{z}^{(k)}\|_2 \\ \nu^{(k)} &= (\mathbf{q}^{(k)})^* A\mathbf{q}^{(k)}. \end{aligned} \tag{XI.3.5}$$

Análisis de convergencia.

Analicemos la convergencia de (XI.3.5). Por inducción sobre k podemos comprobar que

$$\mathbf{q}^{(k)} = \frac{A^k\mathbf{q}^{(0)}}{\|A^k\mathbf{q}^{(0)}\|_2}, \quad k \geq 1. \tag{XI.3.6}$$

Esta relación explica el papel jugado por la potencias de A en el método iterativo descrito.

Supongamos que

$$\mathbf{q}^{(0)} = \sum_{i=1}^{n} \alpha_i \mathbf{u}_i$$

con $\alpha_i \in \mathbb{C},\ i = 1, \ldots, n$. Como $A\mathbf{u}_i = \lambda_i \mathbf{u}_i,\ i = 1, \ldots, n$, tenemos que

$$A^k\mathbf{q}^{(0)} = \alpha_1\lambda_1^k \left(\mathbf{u}_1 + \sum_{i=2}^{n} \frac{\alpha_i}{\alpha_1} \left(\frac{\lambda_i}{\lambda_1} \right)^k \mathbf{u}_i \right), \quad k = 1, 2, \ldots \tag{XI.3.7}$$

Como $|\lambda_i/\lambda_1| < 1,\ i = 2, \ldots, n$, cuando k aumenta el vector $A^k\mathbf{q}^{(0)}$ (y por tanto $\mathbf{q}^{(k)}$, por XI.3.6) tiende a poseer una componente significativamente grande en la dirección de $\mathbf{u}_1$, mientras que las componentes en las otras direcciones $\mathbf{u}_j,\ j \neq 1$, disminuyen. Usando (XI.3.6) y (XI.3.7), obtenemos

$$\mathbf{q}^{(k)} = \frac{\alpha_1\lambda_1^k\big(\mathbf{u}_1 + \mathbf{v}^{(k)}\big)}{\|\alpha_1\lambda_1^k\big(\mathbf{u}_1 + \mathbf{v}^{(k)}\big)\|_2} = \mu_k \frac{\mathbf{u}_1 + \mathbf{v}^{(k)}}{\|\mathbf{u}_1 + \mathbf{v}^{(k)}\|_2},$$

donde μ_k es el signo de $\alpha_1\lambda_1^k$ y $\mathbf{v}^{(k)}$ denota un vector que se tiende a cero cuando k tiende hacia infinito.

Cuando k tiende hacia infinito, el vector $\mathbf{q}^{(k)}$ se alinea, pues, con la dirección del autovector $\mathbf{u}_1$, y se tiene la siguiente estimación del error en la etapa k-ésima.

Teorema XI.3.1. *Con la notación anterior, si* $\alpha_1 \neq 0$, *existe una constante* $C > 0$ *tal que*

$$\|\tilde{\mathbf{q}}^{(k)} - \mathbf{u}_1\|_2 \leq C \left| \frac{\lambda_2}{\lambda_1} \right|^k, \quad k \geq 1, \tag{XI.3.8}$$

donde

$$\tilde{\mathbf{q}}^{(k)} = \frac{\mathbf{q}^{(k)}\|A^k\mathbf{q}^{(0)}\|_2}{\alpha_1\lambda_1^k} = \mathbf{u}_1 + \sum_{i=2}^{n}\frac{\alpha_i}{\alpha_1}\left(\frac{\lambda_i}{\lambda_1}\right)^k\mathbf{u}_i, \quad k = 1, 2, \ldots,$$

Demostración. De (XI.3.7) se sigue que

$$\begin{aligned}\left\|\mathbf{u}_1 + \sum_{i=2}^{n}\frac{\alpha_i}{\alpha_1}\left(\frac{\lambda_i}{\lambda_1}\right)^k\mathbf{u}_i - \mathbf{u}_1\right\|_2 &= \left\|\sum_{i=2}^{n}\frac{\alpha_i}{\alpha_1}\left(\frac{\lambda_i}{\lambda_1}\right)^k\mathbf{u}_i\right\|_2 \\ &\leq \left(\sum_{i=2}^{n}\left(\frac{\alpha_i}{\alpha_1}\right)^2\left(\frac{\lambda_i}{\lambda_1}\right)^{2k}\right)^{1/2} \\ &\leq \left|\frac{\lambda_2}{\lambda_1}\right|^k\left(\sum_{i=2}^{n}\left(\frac{\alpha_i}{\alpha_1}\right)^2\right)^{1/2},\end{aligned}$$

que no es más que (XI.3.8) para $C = \left(\sum_{i=2}^{n}(\alpha_i/\alpha_1)^2\right)^{1/2}$. ∎

La estimación (XI.3.8) expresa la convergencia de $\tilde{\mathbf{q}}^{(k)}$ hacia $\mathbf{u}_1$. Por consiguiente, la sucesión de *cocientes de Rayleigh*

$$\frac{(\tilde{\mathbf{q}}^{(k)})^* A\tilde{\mathbf{q}}^{(k)}}{\|\tilde{\mathbf{q}}^{(k)}\|_2^2} = (\mathbf{q}^{(k)})^* A\mathbf{q}^{(k)} = \nu^{(k)}$$

convergerá a λ_1. Como consecuencia, $\lim_{k\to\infty}\nu^{(k)} = \lambda_1$, y la convergencia será más rápida cuanto menor sera el cociente $|\lambda_2|/|\lambda_1|$.

Ejemplo XI.3.2. Consideremos la familia de matrices

$$A_\alpha = \begin{pmatrix} \alpha & 2 & 3 & 13 \\ 5 & 11 & 10 & 8 \\ 9 & 7 & 6 & 12 \\ 4 & 14 & 15 & 1 \end{pmatrix}, \ \alpha \in \mathbb{R}.$$

Queremos aproximar el autovalor con mayor módulo por el método de la potencia. Cuando $\alpha = 30$, los autovalores de la matriz son $\lambda_1 = 39{,}396$, $\lambda_2 = 17{,}8208, \lambda_3 = -9{,}5022$ y $\lambda_4 = 0{,}2854$ aproximadamente. El método aproxima λ_1 en menos de 30 iteraciones con $\mathbf{q}^{(0)} = (1, 1, 1, 1)^{\mathrm{t}}$. Sin embargo, si $\alpha = -30$ necesitamos más de 700 iteraciones. El diferente comportamiento puede explicarse observando que en el último caso se tiene que $\lambda_1 = -30{,}634$ y $\lambda_2 = 29{,}7359$. Así, $|\lambda_2|/|\lambda_1| = 0{,}9704$, que está próximo a la unidad.

En la sección 5.3 de [**QSS07**] se puede encontrar un test de parada para las iteraciones del método de la potencia, así como una variante de este método denominado **método de la potencia inversa** que consiste en aplicar el método de la potencia

a la matriz $(A-\mu I_n)^{-1}$ donde μ se elige próximo a un autovalor de A. Este método tiene un coste computacional más elevado que el método de la potencia, pero tiene la ventaja de que podemos elegir μ de tal forma que converja a cualquier autovalor de A. La elección de μ para este propósito se puede realizar usando los círculos de Gerhsgorin (véase la sección 5.1 de [**QSS07**] o el apartado 6.3 de [**QS06**]), por ejemplo. Los aspectos sobre la implementación de los métodos de la potencia y de la potencia inversa se pueden consultar en el apartado 5.3.3 de [**QSS07**] o en el capítulo 6 de [**QS06**].

Deflación.

Supongamos que los autovalores de $A \in \mathcal{M}_n(\mathbb{R})$ está ordenados como en (XI.3.4) y supongamos que el par autovalor/autovector $(\lambda_1, \mathbf{u}_1)$ es conocido. La matriz A se puede transformar la siguiente matriz particionada en bloques

$$A_1 = HAH = \begin{pmatrix} \lambda_1 & \mathbf{b}^{\mathrm{t}} \\ 0 & A_2 \end{pmatrix},$$

donde $\mathbf{b} \in \mathbb{R}^{n-1}$, H es la matriz de Householder tal que $H\mathbf{u}_1 = \alpha\mathbf{u}_1$ para algún $\alpha \in \mathbb{R}$ y la matriz $A_2 \in \mathcal{M}_n(\mathbb{R})$ tiene los mismos autovalores que A excepto λ_1. La matriz H se puede calcular usando $\mathbf{w} = \mathbf{u}_1 \pm \|\mathbf{u}_1\|_2\mathbf{e}_1$ (véase la definición IX.4.1).

La **deflación** consiste en calcular el segundo autovalor λ_2 de A aplicando el método de la potencia a A_2 (supuesto que $|\lambda_2| \neq |\lambda_3|$). Una vez que conocemos λ_2, el autovector correspondiente $\mathbf{u}_2$ se puede calcular aplicando el método de la potencia inversa a la matriz A tomando μ próximo a λ_2 y así sucesivamente con el resto de pares autovalor/autovector (si fuese posible).

Ejercicios del tema XI

Ejercicio 1. Aplicar el método de Jacobi a las siguientes matrices:

$$\begin{pmatrix} 9 & 1 & -2 & 1 \\ 1 & 8 & -3 & -2 \\ -2 & -3 & 7 & -1 \\ 1 & -2 & -1 & 6 \end{pmatrix}, \quad \begin{pmatrix} 1 & -1 & 3 & 4 \\ -1 & 4 & 0 & -1 \\ 3 & 0 & 0 & -3 \\ 4 & -1 & -3 & 1 \end{pmatrix},$$

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \\ 3 & 4 & 1 & 2 \\ 4 & 3 & 2 & 1 \end{pmatrix}, \quad \begin{pmatrix} 9 & 1 & -2 & 4 \\ 1 & 8 & -3 & -2 \\ -2 & -3 & 7 & -1 \\ 4 & -2 & -1 & 6 \end{pmatrix},$$

y calcular (aproximaciones) de sus autovalores y autovectores.

Ejercicio 2. Aplicar el método QR a las matrices del ejercicio 1 y calcular (aproximaciones) de sus autovalores y autovectores.

Ejercicio 3. Este ejercicio muestra el método de Jacobi-Corbato, que, a partir del método de Jacobi clásico permite acelerar la busqueda de una pareja (p, q) verificando $\left|a_{pq}^{(m)}\right| = \text{máx}_{i\neq j}\left|a_{ij}^{(m)}\right|$.

1. Consideremos los vectores $\mathbf{a}_m$ y $\mathbf{b}_m$ de componentes

$$a_i^{(m)} = \max_{j>i}\left|a_{ij}^{(m)}\right| = \left|a_{ij_i^{(m)}}^{(m)}\right|, \ i = 1, \ldots, n,$$

$$b_i^{(m)} = j_i^{(m)}, \ i = 1, \ldots, n,$$

respectivamente. Explicar cómo se pueden construir los vectores $\mathbf{a}_{m+1}$ y $\mathbf{b}_{m+1}$ a partir de los vectores $\mathbf{a}_m$ y $\mathbf{b}_m$.

2. Deducir un proceso para determinar, a partir de $\mathbf{a}_m$ y $\mathbf{b}_m$, una pareja (p, q) tal que

$$\left|a_{pq}^{(m+1)}\right| = \max_{i\neq j}\left|a_{ij}^{(m+1)}\right|.$$

Ejercicio 4. Verificar que el método de la potencia no es capaz de calcular el autovalor de módulo máximo de la matriz siguiente, y explicar porqué:

$$A = \begin{pmatrix} 1/3 & 2/3 & 2 & 3 \\ 1 & 0 & -1 & 2 \\ 0 & 0 & -5/3 & -2/3 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

Ejercicio 5. Supongamos que se satisfacen todas condiciones necesarias para aplicar el método de la potencias excepto que $\alpha \neq 0$. Probar que en este caso la sucesión (XI.3.5) converge al par autovalor/autovector $(\lambda_2, \mathbf{u}_2)$. Entonces, estudiar experimentalmente el comportamiento del método calculando el par $(\lambda_1, \mathbf{u}_1)$ para la matriz

$$A = \begin{pmatrix} 1 & -1 & 2 \\ -2 & 0 & 5 \\ 6 & -3 & 6 \end{pmatrix}$$

TEMA XII

Espacios de Hilbert

El análisis funcional es una de las áreas centrales en la matemática moderna, y la teoría de los espacios de Hilbert es núcleo alrededor del cual el análisis funcional se ha desarrollado. Los espacios de Hilbert tienen una estructura geométrica bastante rica, ya que son espacios vectoriales dotados de un producto escalar que permite definir el concepto de ortogonalidad. De hecho el objetivo de este tema se centrará en la construcción de bases ortonormales (en un sentido a precisar que generalice el estudiado en los temas anteriores).

Uno de los ejemplos más importantes de espacio de Hilbert es el espacio L^2 de la funciones de cuadrado Lebesgue integrable que se estudiará en la asignatura de teorías de la medida y de la probabilidad, así como el espacio ℓ^2 la sucesiones de cuadrado sumable que será el que estudiaremos con cierto detalle en este tema. Otro ejemplo, también importante de espacio de Hilbert es el de espacio vectorial de dimensión finita dotado de un producto escalar. Estos espacios de Hilbert han ido apareciendo a lo largo de la asignatura desde el tema V.

En la primera sección del tema estudiamos los espacios vectorial dotados de un producto escalar (sin preocuparnos de la dimensión). Éstos son los llamados espacios prehilbertianos. En esta sección definimos este tipo de espacios y mostramos sus propiedades más importantes. Es destacable que, al igual que en el caso de dimensión finita, el producto escalar define una norma, por lo que podremos concluir que todo espacio prehilbertiano es un espacio normado, y por lo tanto métrico, es decir, podremos definir una noción de distancia entre sus vectores. Tras estudiar algunas propiedades interesantes de la norma y la métrica definidas en los espacios prehilbertiano, finalizamos el tema estudiando con detalle el ejemplo de los espacios ℓ^2 que, como se verá, será el ejemplo fundamental de espacio de Hilbert en esta asignatura.

En la segunda sección nos ocupamos de la ortogonalidad. En este caso aparentemente no hay una diferencia sustancial con lo estudiado sobre ortogonalidad en el caso de dimensión finita; sin embargo, a poco que lo pensemos se echa en falta la noción de base ortonormal. Téngase en cuenta que en todo espacio vectorial existen bases, y que dado una sucesión de vectores linealmente independiente demostramos que podemos calcular un sistema ortogonal que genere el mismo espacio que la sucesión; luego, ¿qué ingrediente nos falta? El ingrediente que nos falta es la numerabilidad:

todo espacio vectorial posee una base pero no necesariamente numerable. Así, todos nuestros esfuerzos hasta el final del tema consistirán en comprender qué condiciones hay que suponer en un espacio prehilbertiano para que exista una base ortonormal; lo que nos llevará primero a definir la noción de espacio de Hilbert y posteriormente la de espacio de Hilbert separable. El resultado final del tema será que esencialmente existen dos espacios de Hilbert separables sobre $\Bbbk = \mathbb{R}$ ó $\mathbb{C}$, a saber, $\Bbbk^n$ y ℓ^2.

Para la elaboración de este tema hemos utilizado el capítulo II de [**Ber77**] y algunas cuestiones puntuales del capítulo 3 de [**DP99**].

1. Espacios prehilbertianos

Definición XII.1.1. Un **espacio prehilbertiano** es un espacio vectorial V sobre $\Bbbk$ junto con una aplicación $V \times V \to \Bbbk; (\mathbf{u}, \mathbf{v}) \mapsto \mathbf{u} \cdot \mathbf{v}$, llamada **producto escalar**, tal que

(a) $\mathbf{u} \cdot \mathbf{v} = \overline{\mathbf{v} \cdot \mathbf{u}}$, para todo $\mathbf{u}$ y $\mathbf{v} \in V$;
(b) $(\mathbf{u} + \mathbf{v}) \cdot \mathbf{w} = \mathbf{u} \cdot \mathbf{w} + \mathbf{v} \cdot \mathbf{w}$, para todo $\mathbf{u}, \mathbf{v}$ y $\mathbf{w} \in V$;
(c) $(\lambda \mathbf{u}) \cdot \mathbf{v} = \lambda\, \mathbf{u} \cdot \mathbf{v}$, para todo $\lambda \in \Bbbk$ y $\mathbf{u}$ y $\mathbf{v} \in V$.
(d) $\mathbf{u} \cdot \mathbf{u} \geq 0$, para todo $\mathbf{u} \in V$, y $\mathbf{u} \cdot \mathbf{u} = 0$, si, y sólo si, $\mathbf{u} = \mathbf{0}$.

Ejemplos XII.1.2.

i) El espacio vectorial $\mathbb{R}^n$ es un espacio prehilbertiano con el producto escalar

$$\mathbf{u} \cdot \mathbf{v} = \mathbf{v}^{\mathrm{t}} \mathbf{u} = \sum_{i=1}^{n} u_i v_i,$$

donde $\mathbf{v} = (v_1, v_2, \ldots, v_n)^{\mathrm{t}}$ y $\mathbf{u} = (u_1, u_2, \ldots, u_n)^{\mathrm{t}} \in \mathbb{R}^n$. Nótese que este espacio prehilbertiano no es más que el espacio vectorial euclídeo $\mathbb{R}^n$ con el producto escalar usual que fue estudiado con detalle en el tema V.

ii) El espacio vectorial $\mathbb{C}^n$ es un espacio prehilbertiano con el producto escalar

$$\mathbf{u} \cdot \mathbf{v} = \mathbf{v}^* \mathbf{u} = \sum_{i=1}^{n} u_i \overline{v}_i,$$

donde $\mathbf{v} = (v_1, v_2, \ldots, v_n)^{\mathrm{t}}$ y $\mathbf{u} = (u_1, u_2, \ldots, u_n)^{\mathrm{t}} \in \mathbb{C}^n$, y $\mathbf{v}^*$ es el adjunto (es decir, el conjugado y traspuesto) de $\mathbf{v}$. Recuérdese que este espacio prehilbertiano ya apareció en el tema V cuando comentamos el caso de las matrices hermíticas.

iii) En el espacio vectorial complejo de las funciones $f : \{1, \ldots, n\} \subset \mathbb{R} \to \mathbb{C}$ el producto escalar

$$f \cdot g = \sum_{t=1}^{n} f(t) \overline{g(t)}$$

define una estructura de espacio prehilbertiano.

iv) El espacio vectorial V de las sucesiones de números reales casi nulas, esto es el conjunto de sucesiones de números reales $\mathbf{x} = (x_n)_{n\in\mathbb{N}}$ que son cero a partir de un cierto subíndice, con el producto escalar

$$\mathbf{x}\cdot\mathbf{y} = \sum_{n=1}^{\infty} x_n y_n$$

tiene una estructura de espacio prehilbertiano.

v) El espacio vectorial de dimensión infinita[1]

$$\ell^2 = \{(x_n)_{n\in\mathbb{N}} \mid x_n \in \mathbb{C} \text{ tales que } \sum_{n=1}^{\infty} |x_n|^2 < \infty\},$$

con el producto escalar

$$\mathbf{x}\cdot\mathbf{y} = \sum_{n=1}^{\infty} x_n \bar{y}_n$$

es un espacio prehilbertiano. Tal y como veremos en el siguiente tema este espacio es, en un cierto sentido, el ejemplo más importante de espacio prehilbertiano.

vi) El espacio vectorial de las funciones continuas en el intervalo $[a, b]$, donde $a < b$, con el producto escalar

$$f\cdot g = \int_a^b f(x)\overline{g(x)}\mathrm{d}x$$

tiene estructura de espacio prehilbertiano.

Los axiomas (b) y (c) para un espacio prehilbertiano se pueden expresar como sigue: el producto escalar $\mathbf{u}\cdot\mathbf{v}$ es *aditivo* y *homogéneo* en el primer factor. Las dos primeras propiedades recogidas en el siguiente resultado afirman que el producto escalar es aditivo y *homogéneo-conjugado* en el segundo factor.

Notación XII.1.3. En lo sucesivo, escribiremos $\mathcal{P}$ para denotar un espacio prehilbertiano genérico.

Obsérvese que, si V es un espacio prehilbertiano cualquiera y L es un subespacio vectorial de V, entonces L también es un espacio prehilbertiano.

Proposición XII.1.4. *Sea $\mathcal{P}$ un espacio prehilbertiano.*

(a) $\mathbf{u}\cdot(\mathbf{v}+\mathbf{w}) = \mathbf{u}\cdot\mathbf{v} + \mathbf{u}\cdot\mathbf{w}$, *para todo* $\mathbf{u},\mathbf{v}$ *y* $\mathbf{w}\in\mathcal{P}$.

(b) $\mathbf{u}\cdot(\lambda\mathbf{v}) = \bar{\lambda}\,\mathbf{u}\cdot\mathbf{v}$, *para todo* $\mathbf{u}$ *y* $\mathbf{v}\in\mathcal{P}$ *y* $\lambda\in\Bbbk$.

(c) $\mathbf{u}\cdot\mathbf{0} = \mathbf{0}\cdot\mathbf{u} = 0$, *para todo* $\mathbf{u}\in\mathcal{P}$.

[1]La demostración de que $\ell^2 = \{(x_n)_{n\in\mathbb{N}} \mid x_n \in \mathbb{C} \text{ tales que } \sum_{n=1}^{\infty} |x_n|^2 < \infty\}$, es un espacio vectorial no es trivial, por lo que la hemos añadido al final de esta sección.

(d) $(\mathbf{u}-\mathbf{v})\cdot\mathbf{w}=\mathbf{u}\cdot\mathbf{w}-\mathbf{v}\cdot\mathbf{w}$ *y* $\mathbf{u}\cdot(\mathbf{v}-\mathbf{w})=\mathbf{u}\cdot\mathbf{v}-\mathbf{u}\cdot\mathbf{w}$, *para todo* $\mathbf{u},\mathbf{v}$ *y* $\mathbf{w}\in\mathcal{P}$.

(e) *Si* $\mathbf{u}\cdot\mathbf{w}=\mathbf{v}\cdot\mathbf{w}$, *para todo* $\mathbf{w}\in\mathcal{P}$, *entonces* $\mathbf{u}=\mathbf{v}$.

Demostración. (a) Usando los axiomas (a) y (b) de la definición de espacio prehilbertiano,

$$\mathbf{u}\cdot(\mathbf{v}+\mathbf{w})=\overline{(\mathbf{v}+\mathbf{w})\cdot\mathbf{u}}=\overline{\mathbf{v}\cdot\mathbf{u}+\mathbf{w}\cdot\mathbf{u}}=\overline{\mathbf{v}\cdot\mathbf{u}}+\overline{\mathbf{w}\cdot\mathbf{u}}=\mathbf{u}\cdot\mathbf{v}+\mathbf{u}\cdot\mathbf{w}.$$

(b) Usando los axiomas (a) y (c) de la definición de espacio prehilbertiano,

$$\mathbf{u}\cdot(\lambda\mathbf{v})=\overline{(\lambda\mathbf{v})\cdot\mathbf{u}}=\overline{\lambda\mathbf{v}\cdot\mathbf{u}}=\bar{\lambda}\,\overline{\mathbf{v}\cdot\mathbf{u}}=\bar{\lambda}\,(\mathbf{u}\cdot\mathbf{v}).$$

(c) $\mathbf{u}\cdot\mathbf{0}=\mathbf{u}\cdot(\mathbf{0}+\mathbf{0})=\mathbf{u}\cdot\mathbf{0}+\mathbf{u}\cdot\mathbf{0}$, de donde se sigue que $\mathbf{u}\cdot\mathbf{0}=0$. Análogamente se demuestra que $\mathbf{0}\cdot\mathbf{u}=0$.

(d) $(\mathbf{u}-\mathbf{v})\cdot\mathbf{w}=(\mathbf{u}+(-\mathbf{v}))\cdot\mathbf{w}=\mathbf{u}\cdot\mathbf{w}+(-\mathbf{v})\cdot\mathbf{w}=\mathbf{u}\cdot\mathbf{w}+(-1)\mathbf{v}\cdot\mathbf{w}=\mathbf{u}\cdot\mathbf{w}-\mathbf{v}\cdot\mathbf{w}$. La otra igualdad se demuestra de forma análoga.

(e) Supongamos que $\mathbf{u}\cdot\mathbf{w}=\mathbf{v}\cdot\mathbf{w}$, para todo $\mathbf{w}\in\mathcal{P}$. Entonces $(\mathbf{u}-\mathbf{v})\cdot\mathbf{w}=\mathbf{u}\cdot\mathbf{w}-\mathbf{v}\cdot\mathbf{w}=0$, para todo $\mathbf{w}\in\mathcal{P}$; en particular, $(\mathbf{u}-\mathbf{v})\cdot(\mathbf{u}-\mathbf{v})=0$, de donde se sigue que $\mathbf{u}=\mathbf{v}$, por el axioma (d) de la definición de espacio prehilbertiano. ∎

Definición XII.1.5. En un espacio prehilbertiano $\mathcal{P}$ se define la **norma** de $\mathbf{v}\in\mathcal{P}$ como

$$\|\mathbf{v}\|:=(\mathbf{v}\cdot\mathbf{v})^{1/2}.$$

Veamos que la definición anterior se ajusta a la definición de norma estudiada anteriormente.

Proposición XII.1.6. *Sea* $\mathcal{P}$ *un espacio prehilbertiano.*

(a) $\|\mathbf{v}\|>0$, *cuando* $\mathbf{v}\neq 0$, *y* $\|\mathbf{v}\|=0$ *si, y sólo si,* $\mathbf{v}=\mathbf{0}$.

(b) $\|\mathbf{v}\|=|\lambda|\|\mathbf{v}\|$, *para todo* $\mathbf{v}\in\mathbb{k}$ *y* $\mathbf{v}\in\mathcal{P}$.

Demostración. El apartado (a) es inmediato a partir del axioma (d) de la definición de espacio prehilbertiano y de la relación $\mathbf{0}\cdot\mathbf{0}=0$. En cuanto al (b), basta observar que $\|\lambda\mathbf{v}\|^2=(\lambda\mathbf{v})\cdot(\lambda\mathbf{v})=\lambda\bar{\lambda}\,(\mathbf{v}\cdot\mathbf{v})=|\lambda|^2\|\mathbf{v}\|^2$. ∎

Veamos finalmente que nuestra definición de norma en un espacio prehilbertiano verifica la desigualdad triangular.

Desigualdad triangular. *Sea* $\mathcal{P}$ *un espacio prehilbertiano. Entonces*

$$\|\mathbf{u}+\mathbf{v}\|\leq\|\mathbf{u}\|+\|\mathbf{v}\|,$$

para todo $\mathbf{u}$ *y* $\mathbf{v}\in\mathcal{P}$.

Demostración. Si designamos por $\mathrm{Re}(\lambda)$ la parte real de un número complejo λ, es evidente que $|\mathrm{Re}(\lambda)| \leq |\lambda|$. Aplicando la desigualdad de Cauchy-Schwarz en los pasos adecuados,

$$\begin{aligned}\|\mathbf{u}+\mathbf{v}\|^2 &= \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 + \mathbf{u}\cdot\mathbf{v} + \mathbf{v}\cdot\mathbf{u} = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 + \mathbf{u}\cdot\mathbf{v} + \overline{\mathbf{u}\cdot\mathbf{v}} \\ &= \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 + 2\mathrm{Re}(\mathbf{u}\cdot\mathbf{v}) \leq \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 + 2|\mathbf{u}\cdot\mathbf{v}| \\ &\leq \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 + 2\|\mathbf{u}\|\|\mathbf{v}\| = (\|\mathbf{u}\| + \|\mathbf{v}\|)^2.\end{aligned}$$

■

De todo lo anterior se deduce que

Corolario XII.1.7. *Todo espacio prehilbertiano $\mathcal{P}$ tiene una estructura natural de espacio normado determinada por la norma $\|\mathbf{v}\| := (\mathbf{v}\cdot\mathbf{v})^{1/2}$.*

Recuérdese que todo espacio normado $(V, \|\cdot\|)$ tiene una estructura natural de espacio métrico determinada por la métrica $\mathrm{d}(\mathbf{u},\mathbf{v}) := \|\mathbf{u}-\mathbf{v}\|$. Luego, podemos concluir que *todo espacio prehilbertiano es un espacio métrico.*

En el tema VIII vimos algunos ejemplos de espacios normados, otro ejemplo de espacio normado es el siguiente:

Ejemplos XII.1.8.

i) Sea p un entero positivo. En el espacio vectorial, ℓ^p, de la sucesiones $x = (x_n)_{n\in\mathbb{N}}$ de número complejos tales que

$$\sum_{n=1}^{\infty} |x_n|^p < +\infty,$$

la aplicación $\|x\| = \left(\sum_{n=1}^{\infty} |x_n|^p\right)^{1/p}$ es una norma. La desigualdad triangular para esta norma es la desigualdad de Minkowski que veremos más adelante.

La norma que hemos definido en un espacio prehilbertiano verifica la siguiente propiedad:

Regla del paralelogramo. *Sea $\mathcal{P}$ un espacio prehilbertiano. Entonces*

$$\|\mathbf{u}+\mathbf{v}\|^2 + \|\mathbf{u}-\mathbf{v}\|^2 = 2(\|\mathbf{u}\|^2 + \|\mathbf{v}\|^2),$$

para todo $\mathbf{u}$ y $\mathbf{v} \in \mathcal{P}$.

Demostración. Se tiene que $\|\mathbf{u}+\mathbf{v}\|^2 = (\mathbf{u}+\mathbf{v})\cdot(\mathbf{u}+\mathbf{v}) = \mathbf{u}\cdot\mathbf{u}+\mathbf{u}\cdot\mathbf{v}+\mathbf{v}\cdot\mathbf{u}+\mathbf{v}\cdot\mathbf{v} = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 + (\mathbf{u}\cdot\mathbf{v}) + (\mathbf{v}\cdot\mathbf{u})$, y sustituyendo $\mathbf{v}$ por $-\mathbf{v}$, que $\|\mathbf{u}-\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - (\mathbf{u}\cdot\mathbf{v}) - (\mathbf{v}\cdot\mathbf{u})$. Por consiguiente $\|\mathbf{u}+\mathbf{v}\|^2 + \|\mathbf{u}-\mathbf{v}\|^2 = 2\|\mathbf{u}\|^2 + 2\|\mathbf{v}\|^2$. ■

Desigualdad de Cauchy-Schwarz. *Sea $\mathcal{P}$ un espacio prehilbertiano. Entonces,*

$$|\mathbf{u}\cdot\mathbf{v}| \leq \|\mathbf{u}\|\,\|\mathbf{v}\|,$$

para todo $\mathbf{u}$ *y* $\mathbf{v} \in \mathcal{P}$, y se da la igualdad cuando $\mathbf{u} = \alpha\mathbf{v}$, para $\alpha = (\mathbf{u}\cdot\mathbf{v})/(\mathbf{v}\cdot\mathbf{v})$.

Demostración. Sea $\lambda \in \Bbbk$ tal que $|\lambda| = 1$ y $\lambda(\mathbf{v}\cdot\mathbf{u}) = |\mathbf{v}\cdot\mathbf{u}|$.

Si $\mu \in \mathbb{R}$, entonces

$$(\mathbf{v}\cdot\mathbf{v})\mu^2 - 2|\mathbf{u}\cdot\mathbf{v}|\mu + (\mathbf{u}\cdot\mathbf{u}) = (\mu\lambda\mathbf{v} - \mathbf{u})\cdot(\mu\lambda\mathbf{v} - \mathbf{u}) \geq 0. \tag{XII.1.1}$$

Entendiendo $(\mathbf{v}\cdot\mathbf{v})\mu^2 - 2|\mathbf{u}\cdot\mathbf{v}|\mu + (\mathbf{u}\cdot\mathbf{u})$ como un polinomio de segundo grado en μ, de la desigualdad (XII.1.1) se sigue que su discriminante ha de ser negativo o cero, es decir,

$$4|\mathbf{u}\cdot\mathbf{v}|^2 - 4(\mathbf{u}\cdot\mathbf{u})(\mathbf{v}\cdot\mathbf{v}) \leq 0,$$

y concluimos que

$$|\mathbf{u}\cdot\mathbf{v}|^2 \leq (\mathbf{u}\cdot\mathbf{u})(\mathbf{v}\cdot\mathbf{v}).$$

La segunda parte de la demostración se deja como ejercicio al lector. ∎

Terminamos esta sección mostrando un resultado sobre convergencia en espacios prehilbertianos.

Proposición XII.1.9. *Sea $\mathcal{P}$ un espacio prehilbertiano.*

(a) *Si* $\mathbf{u}_n \to \mathbf{u}$ *y* $\mathbf{v}_n \to \mathbf{v}$, *entonces* $\mathbf{u}_n\cdot\mathbf{v}_n \to \mathbf{u}\cdot\mathbf{v}$.

(b) *Si* $(\mathbf{u}_n)_{n\in\mathbb{N}}$ *y* $(\mathbf{v}_n)_{n\in\mathbb{N}}$ *son sucesiones de Cauchy, entonces la sucesión* $\mathbf{u}_n\cdot\mathbf{v}_n$ *es una sucesión de Cauchy de escalares (y por lo tanto convergente).*

Demostración. (a) Para todo $n \geq 1$, se tiene que $\mathbf{u}_n\cdot\mathbf{v}_n - \mathbf{u}\cdot\mathbf{v} = (\mathbf{u}_n - \mathbf{u})\cdot(\mathbf{v}_n - \mathbf{v}) + \mathbf{u}\cdot(\mathbf{v}_n - \mathbf{v}) + (\mathbf{u}_n - \mathbf{u})\cdot\mathbf{v}$. Empleando la desigualdad triangular del módulo y la desigualdad de Cauchy-Schwarz, se tiene que $|\mathbf{u}_n\cdot\mathbf{v}_n - \mathbf{u}\cdot\mathbf{v}| \leq \|\mathbf{u}_n - \mathbf{u}\|\|\mathbf{v}_n - \mathbf{v}\| + \|\mathbf{u}\|\|\mathbf{v}_n - \mathbf{v}\| + \|\mathbf{u}_n - \mathbf{u}\|\|\mathbf{v}\|$; evidentemente el segundo miembro tiende a cero cuando n tiende hacia infinito.

(b) Análogamente, $|\mathbf{u}_n\cdot\mathbf{v}_n - \mathbf{u}_m\cdot\mathbf{v}_m| \leq \|\mathbf{u}_n - \mathbf{u}_m\|\|\mathbf{v}_n - \mathbf{v}_m\| + \|\mathbf{u}_m\|\|\mathbf{v}_n - \mathbf{v}_m\| + \|\mathbf{u}_n - \mathbf{u}_m\|\|\mathbf{v}_m\|$, para todo m y como $\|\mathbf{u}_m\|$ y $\|\mathbf{v}_m\|$ están acotados (pues toda sucesión de Cauchy en un espacio normado, y los prehilberianos lo son, está acotada), el segundo miembro tiende a cero cuando n y m tienden hacia infinito. ∎

Espacios ℓ^p.

El conjunto de todas las sucesiones (x_n) de escalares con la suma y multiplicación definidas como sigue

$$(x_1, x_2, \ldots) + (y_1, y_2, \ldots) = (x_1 + y_1, x_2 + y_2, \ldots)$$
$$\lambda(x_1, x_2, \ldots) = (\lambda x_1, \lambda x_2, \ldots)$$

es un espacio vectorial sobre $\Bbbk$. El conjunto de todas las sucesiones de escalares acotadas es un subespacio vectorial propio del espacio vectorial de la sucesiones de escalares. El conjunto de todas la sucesiones de escalares convergentes es un subespacio vectorial propio del espacio vectorial de la sucesiones de escalares acotadas.

La verificación de que los anteriores son realmente espacios vectoriales es muy fácil. En el siguiente caso la tarea es mucho más difícil.

Denotaremos por $\ell^p,\ p \geq 1$ al conjunto de todas las sucesiones (x_n) de números complejos tales que $\sum_{n=1}^{\infty} |x_n|^p < \infty$.

Vamos a demostrar que ℓ^p es un espacio vectorial. Como ℓ^p es un subconjunto de un subespacio vectorial, concretamente el espacio vectorial de todas las sucesiones de números complejos, basta demostrar que si (x_n) e $(y_n) \in \ell^p$ y $\lambda \in \mathbb{C}$, entonces $(x_n + y_n) \in \ell^p$ y $(\lambda x_n) \in \ell^p$. Para comprobar la segunda propiedad es suficiente observar que

$$\sum_{n=1}^{\infty} |\lambda x_n|^p = |\lambda|^p \sum_{n=1}^{\infty} |x_n|^p < \infty.$$

La condición $\sum_{n=1}^{\infty} |x_n+y_n|^p < \infty$ se sigue de la siguiente desigualdad de Minkowski

$$\left(\sum_{n=1}^{\infty} |x_n + y_n|^p\right)^{1/p} \leq \left(\sum_{n=1}^{\infty} |x_n|^p\right)^{1/p} + \left(\sum_{n=1}^{\infty} |y_n|^p\right)^{1/p}.$$

La demostración de la desigualdad de Minkowski se basa en la desigualdad de Hölder. Ambas desigualdades están demostradas a continuación.

Desigualdad de Hölder. Sean $p > 1, q > 1$ y $1/p + 1/q = 1$. Para cualquier par de sucesiones de números complejos (x_n) e (y_n) se tiene que

$$\sum_{n=1}^{\infty} |x_n y_n| \leq \left(\sum_{n=1}^{\infty} |x_n|^p\right)^{1/p} \left(\sum_{n=1}^{\infty} |y_n|^q\right)^{1/q}.$$

Demostración. En primer lugar observamos que

$$x^{1/p} \leq \frac{1}{p}x + \frac{1}{q}$$

para $0 \leq x \leq 1$. Sean a y b dos números reales no negativos tales que $a^p \leq b^q$. Entonces $0 \leq a^p/b^q \leq 1$ y por consiguiente tenemos que

$$a\, b^{-q/p} \leq \frac{1}{p}\frac{a^p}{b^q} + \frac{1}{q}.$$

Como $-q/p = 1 - q$, obtenemos que

$$a\, b^{1-q} \leq \frac{1}{p}\frac{a^p}{b^q} + \frac{1}{q}.$$

Multiplicando en ambos miembros por b^q obtenemos

$$a\, b \leq \frac{a^p}{p} + \frac{b^q}{q}. \tag{XII.1.2}$$

Hemos demostrado (XII.1.2) suponiendo que $a^p \leq b^q$. Un argumento similar sirve para demostrar (XII.1.2) cuando $b^q \leq a^p$. Por consiguiente la desigualdad puede ser usada para cualesquiera a y $b \geq 0$. Usando (XII.1.2) con

$$a = \frac{|x_j|}{\left(\sum_{k=1}^{n} |x_k|^p\right)^{1/p}} \qquad \text{y} \qquad b = \frac{|y_j|}{\left(\sum_{k=1}^{n} |y_k|^q\right)^{1/q}},$$

donde $n \in \mathbb{N}$ y $1 \leq j \leq n$, obtenemos que

$$\frac{|x_j|}{\left(\sum_{k=1}^{n} |x_k|^p\right)^{1/p}} \frac{|y_j|}{\left(\sum_{k=1}^{n} |y_k|^q\right)^{1/q}} \leq \frac{1}{p}\frac{|x_j|^p}{\sum_{k=1}^{n} |x_k|^p} + \frac{1}{q}\frac{|y_j|^q}{\sum_{k=1}^{n} |y_k|^q}.$$

Sumando estas desigualdades para $j = 1, \ldots, n$ obtenemos

$$\frac{\sum_{k=1}^{n} |x_j||y_j|}{\left(\sum_{k=1}^{n} |x_k|^p\right)^{1/p} \left(\sum_{k=1}^{n} |y_k|^q\right)^{1/q}} \leq \frac{1}{p} + \frac{1}{q} = 1;$$

tomando ahora $n \to \infty$ conseguimos la desigualdad de Hölder. ∎

Desigualdad de Minkowski. Sea $p \geq 1$. Para cualesquiera dos sucesiones (x_n) e (y_n) de números complejos se tiene que

$$\left(\sum_{n=1}^{\infty} |x_n + y_n|^p\right)^{1/p} \leq \left(\sum_{n=1}^{\infty} |x_n|^p\right)^{1/p} + \left(\sum_{n=1}^{\infty} |y_n|^p\right)^{1/p}.$$

Demostración. Para $p = 1$ basta con usar la desigualdad triangular para el valor absoluto. Si $p > 1$, entonces existe q tal que $1/p+1/q = 1$. Entonces, por la desigualdad de Hölder, tenemos que

$$\begin{aligned}\sum_{n=1}^{\infty} |x_n + y_n|^p &= \sum_{n=1}^{\infty} |x_n + y_n|\, |x_n + y_n|^{p-1} \\ &\leq \sum_{n=1}^{\infty} |x_n|\, |x_n + y_n|^{p-1} + \sum_{n=1}^{\infty} |y_n|\, |x_n + y_n|^{p-1} \\ &\leq \left(\sum_{n=1}^{\infty} |x_n|^p\right)^{1/p} \left(\sum_{n=1}^{\infty} |x_n + y_n|^{q(p-1)}\right)^{1/q} \\ &\quad + \left(\sum_{n=1}^{\infty} |y_n|^p\right)^{1/p} \left(\sum_{n=1}^{\infty} |x_n + y_n|^{q(p-1)}\right)^{1/q}.\end{aligned}$$

Como $q(p-1) = p$,

$$\sum_{n=1}^{\infty} |x_n + y_n|^p \leq \left(\left(\sum_{n=1}^{\infty} |x_n|^p\right)^{1/p} + \left(\sum_{n=1}^{\infty} |y_n|^p\right)^{1/p}\right) \left(\sum_{n=1}^{\infty} |x_n + y_n|^p\right)^{1-(1/p)}$$

de donde se sigue la desigualdad de Minkowski. ∎

2. Sistemas ortogonales. Sucesiones ortonormales

Definición XII.2.1. Sea $\mathcal{P}$ un espacio prehilbertiano. Se dice que dos vectores $\mathbf{u}$ y $\mathbf{v} \in \mathcal{P}$ son **ortogonales** cuando $\mathbf{u} \cdot \mathbf{v} = 0$.

La relación de ortogonalidad es simétrica, pero no es reflexiva. Además, todo vector es ortogonal a $\mathbf{0}$.

Proposición XII.2.2. *Sea $\mathcal{P}$ un espacio prehilbertiano. Si $\mathbf{v} \in \mathcal{P}$ es ortogonal a cada uno de los vectores $\mathbf{u}_1, \ldots, \mathbf{u}_n \in \mathcal{P}$, entonces es ortogonal a cualquier combinación lineal suya.*

Demostración. Si $\mathbf{u} = \sum_{i=1}^{n} \lambda_i \mathbf{u}_i,\ \lambda_i \in \Bbbk,\ i = 1, \ldots, n$, entonces se tiene que $\mathbf{v} \cdot \mathbf{u} = \sum_{i=1}^{n} \bar{\lambda}_i (\mathbf{v} \cdot \mathbf{u}_i) = 0$. ∎

Definición XII.2.3. Sea $\mathcal{P}$ un espacio prehilbertiano. Se dice que un subconjunto arbitrario S de $\mathcal{P} \setminus \{\mathbf{0}\}$ es un **sistema ortogonal** cuando $\mathbf{u} \cdot \mathbf{v} = 0$ para cualquier par de elementos distintos de S. Si, además, $\|\mathbf{v}\| = 1$, para todo $\mathbf{v} \in S$, entonces se dice que S es un **sistema ortonormal**.

Cualquier sistema ortogonal de vectores puede ser normalizado. En efecto, si S es un sistema ortogonal, entonces la familia

$$S_1 = \left\{ \frac{\mathbf{v}}{\|\mathbf{v}\|} \mid \mathbf{v} \in S \right\}$$

es un sistema ortonormal. Ambos sistemas son equivalentes en el sentido de que generan el mismo subespacio vectorial de $\mathcal{P}$.

Corolario XII.2.4. *En un espacio prehilbertiano todo sistema ortogonal es linealmente independiente.*

Demostración. Sean $\mathcal{P}$ un espacio prehilbertiano y $S \subseteq \mathbb{P}$ un sistema ortogonal. Supongamos que $\sum_{i=1}^{n} \lambda_i \mathbf{v}_i = \mathbf{0}$, para ciertos $\mathbf{v}_1, \ldots, \mathbf{v}_n \in S$ y $\lambda_1, \ldots, \lambda_n \in \Bbbk$. Entonces,

$$0 = \sum_{i=1}^{n} \mathbf{0} \cdot (\lambda_i \mathbf{v}_i) = \sum_{i=1}^{n} \left(\sum_{j=1}^{n} \lambda_j \mathbf{v}_j \right) \cdot (\lambda_i \mathbf{v}_i) = \sum_{i=1}^{n} |\lambda_i|^2 \|\mathbf{v}_i\|^2,$$

como $\|\mathbf{v}_i\| > 0$, para todo $i = 1, \ldots, n$, se sigue que $\lambda_i = 0$, para todo $i = 1, \ldots, n$. Luego, $\mathbf{v}_1, \ldots, \mathbf{v}_n$ son linealmente independientes. ∎

Ejemplos XII.2.5.

i) Sea $(\lambda_i)_{i \in \mathbb{N}}$ una sucesión cualquiera de escalares. En el espacio prehilbertiano de las sucesiones casi nulas, la sucesión $\mathbf{v}^{(1)} = (\lambda_1, 0, \ldots,)$, $\mathbf{v}^{(2)} = (0, \lambda_2, 0, \ldots,)$, $\mathbf{v}^{(3)} = (0, 0, \lambda_3, 0, \ldots,), \ldots$ forma un sistema ortogonal.

ii) En el espacio prehilbertiano de funciones continuas en el intervalo $[-\pi, \pi]$, la sucesión de funciones $(\mathbf{s}_n)_{n \in \mathbb{N}}$ de término general $\mathbf{s}_n(t) = \operatorname{sen}(nt)$ constituye un sistema ortogonal, es decir,

$$\int_{-\pi}^{\pi} \operatorname{sen}(mt) \operatorname{sen}(nt) \mathrm{d}t = 0 \quad \text{si} \quad m \neq n.$$

Análogamente, la sucesión $(\mathbf{c}_n)_{n \in \mathbb{N}}$ de término general $\mathbf{c}_n(t) = \cos(nt)$ forma un sistema ortogonal. Además, $\mathbf{s}_n \cdot \mathbf{c}_m = 0$, para todo m y n.

iii) En el espacio prehilbertiano de las funciones $f : \{1, \ldots, n\} \subset \mathbb{R} \to \mathbb{C}$, las n funciones no nulas del conjunto

$$\mathcal{S} = \left\{ \operatorname{sen}\left(\frac{2\pi k t}{n}\right), \ \cos\left(\frac{2\pi k t}{n}\right) \ \mid \ k = 0, 1, \ldots \left[\frac{n}{2}\right] \right\},$$

donde $[x]$ denota el mayor entero menor o igual que x, forman un sistema ortogonal.

Teorema XII.2.6. *Sea $\mathcal{P}$ un espacio prehilbertiano.*

(a) ***Teorema de Pitágoras***. *Si* $\mathbf{u}$ *y* $\mathbf{v} \in \mathcal{P}$ *son ortogonales, entonces*

$$\|\mathbf{u}+\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$$

(b) ***Teorema de Pitágoras generalizado.*** *Si* $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ *es un sistema ortogonal de vectores de* $\mathcal{P}$*, entonces*

$$\left\| \sum_{i=1}^{n} \mathbf{v}_i \right\|^2 = \sum_{i=1}^{n} \|\mathbf{v}_i\|^2.$$

Demostración. (a) Como $\mathbf{u}$ y $\mathbf{v}$ son ortogonales, $\mathbf{u}\cdot\mathbf{v} = 0 = \mathbf{v}\cdot\mathbf{u}$, de donde se sigue que

$$\begin{aligned}\|\mathbf{u}+\mathbf{v}\|^2 &= (\mathbf{u}+\mathbf{v})\cdot(\mathbf{u}+\mathbf{v}) = \mathbf{u}\cdot\mathbf{u}+\mathbf{u}\cdot\mathbf{v}+\mathbf{v}\cdot\mathbf{u}+\mathbf{v}\cdot\mathbf{v}\\ &= \mathbf{u}\cdot\mathbf{u}+\mathbf{v}\cdot\mathbf{v} = \|\mathbf{u}\|^2+\|\mathbf{v}\|^2.\end{aligned}$$

(b) Procedemos por inducción sobre n. Si $n = 2$, entonces $\|\mathbf{v}_1+\mathbf{v}_2\|^2 = \|\mathbf{v}_1\|^2 + \|\mathbf{v}_2\|^2$ por el teorema de Pitágoras. Supongamos que $n > 2$ y que el teorema es cierto para $n-1$ vectores, es decir,

$$\left\| \sum_{i=1}^{n-1} \mathbf{v}_i \right\|^2 = \sum_{i=1}^{n-1} \|\mathbf{v}_i\|^2.$$

Sea $\mathbf{u} = \sum_{i=1}^{n-1} \mathbf{v}_i$ y $\mathbf{v} = \mathbf{v}_n$. Como $\mathbf{u}$ y $\mathbf{v}$ son ortogonales, tenemos que

$$\begin{aligned}\left\| \sum_{i=1}^{n} \mathbf{v}_i \right\|^2 = \|\mathbf{u}+\mathbf{v}\|^2 = \|\mathbf{u}\|^2+\|\mathbf{v}\|^2 &= \left\| \sum_{i=1}^{n-1} \mathbf{v}_i \right\|^2 + \|\mathbf{v}_n\|^2\\ = \sum_{i=1}^{n-1} \|\mathbf{v}_i\|^2 + \|\mathbf{v}_n\|^2 &= \sum_{i=1}^{n} \|\mathbf{v}_i\|^2.\end{aligned}$$

■

Igualdad de Parseval (caso finito). *Si* $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ *es un sistema ortogonal de vectores de* $\mathbb{P}$ *y* $\mathbf{v} = \sum_{i=1}^{n} \lambda_i \mathbf{v}_i$*, entonces*

$$\|\mathbf{v}\|^2 = \sum_{i=1}^{n} |\lambda_i|^2 \|\mathbf{v}_i\|^2$$

y $\lambda_i = (\mathbf{v}\cdot\mathbf{v}_i)/\|\mathbf{v}_i\|^2$*, para cada* $k \in \{1, \ldots, n\}$.

Demostración. Es una consecuencia inmediata del teorema de Pitagoras generalizado, por lo que los detalles de su demostración se dejan como ejercicio al lector. ■

Estamos ya en disposición de enunciar y demostrar el resultado principal de esta sección.

Igualdad y desigualdad de Bessel. *Sean $\mathcal{P}$ un espacio prehilbertiano y $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ un sistema ortonormal de vectores de $\mathbb{P}$. Para todo $\mathbf{u} \in \mathcal{P}$ se cumple que*

$$\left\| \mathbf{u} - \sum_{i=1}^{n} (\mathbf{u} \cdot \mathbf{u}_i)\mathbf{u}_i \right\|^2 = \|\mathbf{u}\|^2 - \sum_{i=1}^{n} |\mathbf{u} \cdot \mathbf{u}_i|^2;$$

en particular,

$$\sum_{i=1}^{n} |\mathbf{u} \cdot \mathbf{u}_i|^2 \leq \|\mathbf{u}\|^2.$$

Demostración. Dados $\lambda_1, \ldots, \lambda_n \in \Bbbk$, se tiene que $\| \sum_{i=1}^n \lambda_i \mathbf{u}_i\|^2 = \sum_{i=1}^n \|\lambda_i \mathbf{u}_i\|^2 = \sum_{i=1}^n |\lambda_i|^2$, por la igualdad de Parseval. Por otra parte,

$$\begin{aligned}
\left\| \mathbf{u} - \sum_{i=1}^{n} \lambda_i \mathbf{u}_i \right\|^2 &= \|\mathbf{u}\|^2 - \Big(\sum_{i=1}^{n} \lambda_i \mathbf{u}_i\Big) \cdot \mathbf{u} - \mathbf{u} \cdot \Big(\sum_{i=1}^{n} \lambda_i \mathbf{u}_i\Big) + \sum_{i=1}^{n} |\lambda_i|^2 \\
&= \|\mathbf{u}\|^2 - \sum_{i=1}^{n} \lambda_i \overline{\mathbf{u} \cdot \mathbf{u}_i} - \sum_{i=1}^{n} \bar{\lambda}_i \mathbf{u} \cdot \mathbf{u}_i + \sum_{i=1}^{n} \lambda_i \bar{\lambda}_i \\
&= \|\mathbf{u}\|^2 - \sum_{i=1}^{n} |\mathbf{u} \cdot \mathbf{u}_i|^2 + \sum_{i=1}^{n} |\mathbf{u} \cdot \mathbf{u}_i - \lambda_i|^2
\end{aligned}$$

En particular, haciendo $\lambda_i = \mathbf{u} \cdot \mathbf{u}_i,\ i = 1, \ldots, n$, obtenemos la igualdad de Bessel; la desigualdad se deduce inmediatamente.

■

Obsérvese que la desigualdad de Bessel para $n = 1$ es esencialmente la desigualdad de Cauchy-Schwarz.

Nota XII.2.7. **Proyección ortogonal.** Según la demostración de la igualdad de Bessel, resulta claro que la elección $\lambda_i = \mathbf{u} \cdot \mathbf{u}_i,\ i = 1, \ldots, n$, hace *mínimo* al número $\|\mathbf{u} - \sum_{i=1}^n \lambda_i \mathbf{u}_i\|$, y por lo tanto proporciona la *mejor aproximación* a $\mathbf{u}$ mediante una combinación lineal de $\mathbf{u}_1, \ldots, \mathbf{u}_n$. Además, solamente un conjunto de coeficientes da la mejor aproximación. Obsérvese también que si $n > m$, entonces en dicha aproximación mediante $\mathbf{u}_1, \ldots, \mathbf{u}_n$, los m primeros coeficientes son precisamente los requeridos par la mejor aproximación mediante $\mathbf{u}_1, \ldots, \mathbf{u}_m$.

Por otra parte, si $\mathbf{v} = \sum_{i=1}^n (\mathbf{u} \cdot \mathbf{u}_i)\mathbf{u}_i$ y $\mathbf{w} = \mathbf{u} - \mathbf{v}$, es claro que $\mathbf{w} \cdot \mathbf{u}_i = 0$, para todo $i = 1, \ldots, n$, luego $\mathbf{w} \cdot \mathbf{v} = 0$. Por lo tanto, se tiene una descomposición $\mathbf{u} = \mathbf{v} + \mathbf{w}$, tal que $\mathbf{v}$ es combinación lineal de $\mathbf{u}_1, \ldots, \mathbf{u}_n$ y $\mathbf{w}$ es ortogonal a $\mathbf{u}_i,\ i = 1, \ldots, n$. Es fácil ver que esta descomposición es única. El vector $\mathbf{v}$ se llama **proyección ortogonal** de $\mathbf{u}$ en el subespacio L generado por $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$.

Obsérvese que, según lo dicho anteriormente, la proyección ortogonal $\mathbf{v}$ de $\mathbf{u}$ es el vector de L tal que $\mathrm{d}(\mathbf{u}, \mathbf{v})$ es mínima.

Sucesiones ortonormales.

Definición XII.2.8. Sea $\mathcal{P}$ un espacio prehilbertiano. Una sucesión de vectores $(\mathbf{v}_n)_{n\in\mathbb{N}}$ de $\mathcal{P}$ se llama **sucesión ortonormal** si $\{\mathbf{v}_n \mid n \in \mathbb{N}\}$ es un sistema ortonormal, es decir, si $\mathbf{v}_i \cdot \mathbf{v}_j = 0$, para $i \neq j$, y $\|\mathbf{v}_i\| = 1$, para todo $i \in \mathbb{N}$.

La condición de ortonormalidad de una sucesión de vectores se puede expresar en términos de la función delta de Kronecker:

$$\mathbf{v}_i \cdot \mathbf{v}_j = \delta_{ij} = \begin{cases} 0 & \text{si} \quad i \neq j, \\ 1 & \text{si} \quad i = j. \end{cases}$$

Ejemplos XII.2.9.

i) Si $(\mathbf{v}_n)_{n\in\mathbb{N}}$ es una sucesión de vectores no nulos ortogonales entre sí, la sucesión $(\mathbf{u}_n)_{n\in\mathbb{N}}$ tal que $\mathbf{u}_n = \mathbf{v}_n/\|\mathbf{v}_n\|,\ n \in \mathbb{N}$, es ortonormal.

ii) Con la notación del ejemplo XII.2.5.ii), se tiene que $\|\mathbf{c}_0\| = 2\pi$ y $\|\mathbf{s}_n\|^2 = \|\mathbf{c}_n\|^2 = \pi$, para $n \in \mathbb{N}$. Definimos

$$\begin{aligned} \mathbf{v}_0(t) &= \frac{1}{\sqrt{2\pi}}, \\ \mathbf{v}_{2n}(t) &= \frac{1}{\sqrt{\pi}}\cos(nt), \quad n \in \mathbb{N}, \\ \mathbf{v}_{2n+1}(t) &= \frac{1}{\sqrt{\pi}}\operatorname{sen}(nt), \quad n \in \mathbb{N}. \end{aligned}$$

Entonces, $(\mathbf{v}_m)_{m\geq 0}$ es una sucesión ortonormal.

iii) En el espacio prehilbertiano de las sucesiones casi nulas, sea $\mathbf{e}_1 = (1, 0, \dots,)$, $\mathbf{e}_2 = (0, 1, 0, \dots,), \mathbf{e}_3 = (0, 0, 1, 0, \dots,), \dots$ La sucesión $(\mathbf{e}_i)_{i\in\mathbb{N}}$ es ortonormal.

iv) Con la misma notación que en el apartado anterior, en ℓ^2 la sucesión $(\mathbf{e}_i)_{i\in\mathbb{N}}$ es ortonormal.

Es claro que, dado $\mathbf{x} = (\lambda_i)_{i\in\mathbb{N}} \in \ell^2$, se cumple que $\mathbf{x} \cdot \mathbf{e}_i = \lambda_i$, donde $\mathbf{e}_1 = (1, 0, \dots,), \mathbf{e}_2 = (0, 1, 0, \dots,), \mathbf{e}_3 = (0, 0, 1, 0, \dots,), \dots$ En particular, se cumple que $\sum_{i=1}^{\infty} |\mathbf{x} \cdot \mathbf{e}_i|^2 < \infty$; este resultado es válido para cualquier sucesión ortonormal en un espacio de Hilbert, como consecuencia de la *desigualdad de Bessel*.

Corolario XII.2.10. *Sea $\mathcal{P}$ un espacio prehilbertiano. Si $(\mathbf{u}_i)_{i\in\mathbb{N}}$ es una sucesión ortonormal, entonces para todo $\mathbf{u} \in \mathcal{P}$ se cumple que*

$$\sum_{i=1}^{\infty} |\mathbf{u} \cdot \mathbf{u}_i|^2 \leq \|\mathbf{u}\|^2.$$

En particular, la sucesión $(\mathbf{u} \cdot \mathbf{u}_i)_{i\in\mathbb{N}}$ converge a cero cuando i tiende hacia infinito.

Demostración. Para la primera afirmación, basta tener en cuenta que la desigualdad de Bessel se verifica para todo $n \in \mathbb{N}$. La segunda es consecuencia de la condición necesaria de convergencia de series de números reales. ∎

El corolario anterior asegura que la serie $\sum_{i=1}^{\infty} |\mathbf{u} \cdot \mathbf{u}_i|^2$ es convergente para todo $\mathbf{u} \in \mathcal{P}$. En otras palabras, la sucesión $(\mathbf{u} \cdot \mathbf{u}_i)_{i\in\mathbb{N}}$ es un elemento de ℓ^2. De modo que podemos decir que una sucesión ortonormal en $\mathcal{P}$ induce una aplicación de $\mathcal{P}$ a ℓ^2.

La expansión

$$\mathbf{u} \sim \sum_{i=1}^{\infty} (\mathbf{u} \cdot \mathbf{u}_i)\mathbf{u}_i \tag{XII.2.3}$$

se llama **serie de Fourier generalizada de u**. Los escalares $\lambda_i = \mathbf{u} \cdot \mathbf{u}_i$ son los **coeficientes generalizados de Fourier** de $\mathbf{u}$ respecto de la sucesión ortonormal $(\mathbf{u}_i)_{i\in\mathbb{N}}$. Como hemos reseñado, este conjunto de coeficientes proporciona la mejor aproximación de $\mathbf{u}$ en el espacio vectorial generado por $\{\mathbf{u}_i \mid i \in \mathbb{N}\}$. Sin embargo, en general, no sabemos cuándo la serie (XII.2.3) es convergente; volveremos a esta cuestión en el siguiente tema.

Terminamos esta sección mostrando un procedimiento sistemático (aunque infinito) para "ortonormalizar" cualquier sucesión linealmente independiente de vectores de un espacio prehilbertiano:

Proceso de ortonormalización de Gram-Schmidt (caso general). *Sea $\mathcal{P}$ un espacio prehilbertiano. Si $(\mathbf{v}_i)_{i\in\mathbb{N}}$ es una sucesión de vectores linealmente independientes de $\mathcal{P}$, existe una sucesión ortonormal $(\mathbf{u}_i)_{i\in\mathbb{N}}$ tal que $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ genera el mismo subespacio vectorial que $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$, para cada $n \in \mathbb{N}$.*

Demostración. Los vectores $\mathbf{u}_n$ se definen recursivamente. Sea $\mathbf{u}_1 = \mathbf{v}_1/\|\mathbf{v}_1\|$. Supongamos que ya hemos construido los vectores ortonormales $\mathbf{u}_1, \ldots, \mathbf{u}_{n-1}$, de forma que el espacio vectorial que genera $\{\mathbf{u}_1, \ldots, \mathbf{u}_j\}$, es el mismo que el generado por $\{\mathbf{v}_1, \ldots, \mathbf{v}_j\}$, para cada $j = 1, \ldots, n-1$. Sea $\mathbf{w} = \mathbf{v}_n - \sum_{i=1}^{n-1} (\mathbf{v}_n \cdot \mathbf{u}_i)\mathbf{u}_i$; entonces $\mathbf{w}$ es ortogonal a $\mathbf{u}_1, \ldots, \mathbf{u}_{n-1}$. Definamos $\mathbf{u}_n = \mathbf{w}/\|\mathbf{w}\|$; esto es válido, ya que $\mathbf{w} = \mathbf{0}$ implicaría que $\mathbf{v}_n$ es una combinación lineal de $\mathbf{u}_1, \ldots, \mathbf{u}_{n-1}$ y por tanto de $\mathbf{v}_1, \ldots, \mathbf{v}_{n-1}$, en contra de la independencia de la sucesión $(\mathbf{v}_i)_{i\in\mathbb{N}}$. El lector puede verificar fácilmente que toda combinación lineal de $\mathbf{u}_1, \ldots, \mathbf{u}_n$ es también una combinación lineal de $\mathbf{v}_1, \ldots, \mathbf{v}_n$, y viceversa. ∎

El proceso de Gram-Schmidt se puede aplicar a un conjunto finito de vectores $\mathbf{v}_1, \ldots, \mathbf{v}_n$ linealmente independientes; en este caso, se trata de un algoritmo que proporciona un sistema ortonormal de vectores $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ tal que el espacio vectorial generado por $\mathbf{u}_1, \ldots, \mathbf{u}_j$ es el mismo que el generado por $\mathbf{v}_1, \ldots, \mathbf{v}_j$. En particular:

Corolario XII.2.11. *Si $\mathcal{P}$ es un espacio prehilbertiano de dimensión finita, entonces $\mathcal{P}$ posee una base de vectores ortonormales.*

3. Espacios de Hilbert

Definición XII.3.1. Un espacio prehilbertiano completo se llama **espacio de Hilbert**.

El siguiente ejemplo muestra que no todos los espacios prehilbertianos son espacios de Hilbert, es decir, que existen espacios prehilbertianos que no son completos.

Ejemplo XII.3.2. Sabemos que el espacio vectorial V de las sucesiones de números reales casi nulas, con el producto escalar

$$\mathbf{u}\cdot\mathbf{v}=\sum_{i\geq 1}u_i\,v_i$$

tiene una estructura de espacio prehilbertiano. Veamos que V no es completo construyendo una sucesión de Cauchy que no tenga límite en V.

La sucesión propuesta es $(\mathbf{v}^{(n)})_{n\in\mathbb{N}}$ con

$$\begin{array}{rcl}\mathbf{v}^{(1)} & = & (1,0,0,\ldots)\\ \mathbf{v}^{(2)} & = & (1,1/2,0,\ldots)\\ \mathbf{v}^{(3)} & = & (1,1/2,1/3,0\ldots)\\ & \vdots & \\ \mathbf{v}^{(n)} & = & (1,1/2,1/3,\ldots,1/n,0)\\ & \vdots & \end{array}$$

Para todo $m>n\geq 1$,

$$\left\|\mathbf{v}^{(m)}-\mathbf{v}^{(n)}\right\|^2=\left\|(0,\ldots,0,\frac{1}{n+1},\ldots,\frac{1}{m},0,\ldots)\right\|^2=\sum_{k=n+1}^{m}\left(\frac{1}{k}\right)^2.$$

Dado que la serie $\sum_{k\geq 1}1/k^2$ es convergente, se cumple que $\mathrm{d}(\mathbf{v}^{(m)},\mathbf{v}^{(n)})=\|\mathbf{v}^{(m)}-\mathbf{v}^{(n)}\|$ tiende a cero cuando n tiene hacia infinito. Luego, $(\mathbf{v}^{(n)})_{n\in\mathbb{N}}$ es una sucesión de Cauchy de elementos de V.

Supongamos ahora que la sucesión es convergente en V, entonces existe un elemento de V, $\mathbf{v}=(\lambda_1,\lambda_2,\ldots,\lambda_N,0,\ldots)$, tal que $\lim_{n\to\infty}\mathbf{v}^{(n)}=\mathbf{v}$. Si $n\geq N$,

$$\left\|\mathbf{v}^{(n)}-\mathbf{v}\right\|^2=\sum_{k=1}^{n}\left|\frac{1}{k}-\lambda_k\right|^2,$$

haciendo tender n hacia infinito, obtenemos que $\sum_{k\geq 1}|1/k-\lambda_k|^2=0$, de donde se sigue que $\lambda_k=1\in\Bbbk$, para todo $k\geq 1$, en contradicción con que $\mathbf{v}$ esté en V.

Ejemplos de espacios de Hilbert son $\mathbb{R}^n$ y $\mathbb{C}^n$ con sus productos escalares usuales (véase el ejemplo XII.1.2.i)-ii)). Sin embargo, el ejemplo más importante es el siguiente.

Ejemplo XII.3.3. **El espacio de Hilbert** ℓ^2. Veamos que el espacio vectorial del conjunto de todas las sucesiones $\mathbf{x} = (x_n)_{n\in\mathbb{N}}$ de números complejos tales que $\sum_{n=1}^{\infty} |x_n|^2 < \infty$ con el producto escalar

$$\mathbf{x} \cdot \mathbf{y} = \sum_{n=1}^{\infty} x_n \bar{y}_n$$

es completo.

Supongamos que $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(n)}, \ldots$, es una sucesión de Cauchy en ℓ^2. Sea $\mathbf{x}^{(n)} = (x_i^{(n)})_{i\in\mathbb{N}}$. Para todo $i \in \mathbb{N}$, se tiene que

$$\left|x_i^{(m)} - x_i^{(n)}\right|^2 \leq \sum_{j=1}^{\infty} \left|x_j^{(m)} - x_j^{(n)}\right|^2 = \left\|\mathbf{x}^{(m)} - \mathbf{x}^{(n)}\right\|^2,$$

luego la sucesión $x_i^{(1)}, x_i^{(2)}, \ldots, x_i^{(n)}, \ldots$, de componentes i-ésimas es una sucesión de Cauchy. Como el conjunto de los números complejos es completo, existe $x_i \in \mathbb{C}$ tal que $\lim_{n\to\infty} x_i^{(n)} = x_i$. Vamos a demostrar que $\sum_{i=1}^{\infty} |x_i|^2 < \infty$, es decir, que la sucesión $\mathbf{x} = (x_i)_{i\in\mathbb{N}}$ está en ℓ^2 y que $(\mathbf{x}^{(n)})_{n\in\mathbb{N}}$ converge a $\mathbf{x}$.

Dado $\varepsilon > 0$, sea $N \in \mathbb{N}$ tal que $\|\mathbf{x}^{(m)} - \mathbf{x}^{(n)}\|^2 < \varepsilon$, para todo $m, n \geq N$. Fijemos un entero positivo r; entonces se tiene que

$$\sum_{i=1}^{r} \left|x_i^{(m)} - x_i^{(n)}\right|^2 \leq \left\|\mathbf{x}^{(m)} - \mathbf{x}^{(n)}\right\|^2 < \varepsilon,$$

supuesto que $m, n \geq N$; haciendo tender m hacia infinito,

$$\sum_{i=1}^{r} \left|x_i - x_i^{(n)}\right|^2 < \varepsilon$$

supuesto que $n \geq N$; como r es arbitrario,

$$\sum_{i=1}^{\infty} \left|x_i - x_i^{(n)}\right|^2 < \varepsilon, \quad \text{siempre que} \quad n \geq N. \tag{XII.3.4}$$

En particular, $\sum_{i\geq 1} \left|x_i - x_i^{(N)}\right|^2 < \varepsilon$, por lo tanto la sucesión $(x_i - x_i^{(N)})_{i\in\mathbb{N}}$ pertenece a ℓ^2; sumándole la sucesión $(x_i^{(N)})_{i\in\mathbb{N}}$ se obtiene $(x_i)_{i\in\mathbb{N}}$, por lo tanto, $\mathbf{x} = (x_i)_{i\in\mathbb{N}}$ pertenece a ℓ^2. Luego, de (XII.3.4) se sigue que $\|\mathbf{x} - \mathbf{x}^{(n)}\| < \varepsilon$, para todo $n \geq N$. Por lo tanto, $\mathbf{x}^{(n)}$ converge a $\mathbf{x}$.

Base ortonormal de un espacio de Hilbert.

En el espacio de Hilbert ℓ^2, consideramos la sucesión ortogonal $(\mathbf{e}_n)_{n\in\mathbb{N}}$ tal que $\mathbf{e}_1 = (1, 0, \ldots), \mathbf{e}_2 = (0, 1, 0, \ldots), \mathbf{e}_3 = (0, 0, 1, 0, \ldots), \ldots$. Si $\mathbf{x} = (\lambda_1, \lambda_2, \ldots, \lambda_n, 0, \ldots)$ es una sucesión que tiene a lo sumo un número finito de términos no nulos, es claro que $\mathbf{x} = \sum_{i=1}^{n} \lambda_i \mathbf{e}_i$; por tanto, se podría escribir

$$\mathbf{x} = \sum_{i=1}^{\infty} \lambda_i \mathbf{e}_i,$$

entendiendo que $\lambda_i = 0$ para todo $i > n$.

Consideremos ahora $\mathbf{x} = (\lambda_i)_{i\in\mathbb{N}} \in \ell^2$ ¿Qué sentido se le puede dar a la expresión $\mathbf{x} = \sum_{i=1}^{\infty} \lambda_i \mathbf{e}_i$? Parece natural definir $\sum_{i=1}^{\infty} \lambda_i \mathbf{e}_i$ como el límite de la sucesión de "sumas parciales" $\mathbf{x}_n = \sum_{i=1}^{n} \lambda_i \mathbf{e}_i$; este límite existe y su valor es $\mathbf{x}$, ya que

$$\|\mathbf{x} - \mathbf{x}_n\|^2 = \|(0, \ldots, 0, \lambda_{n+1}, \lambda_{n+2}, \ldots)\|^2 = \sum_{i=n+1}^{\infty} |\lambda_i|^2$$

tiende a cero cuando n tiende hacia infinito.

Veamos que esta situación es general para sucesiones ortonormales arbitrarias en espacios de Hilbert.

Notación XII.3.4. Si $(\mathbf{v}_i)_{i\in\mathbb{N}}$ una sucesión de vectores en un espacio prehilbertiano $\mathcal{P}$ tal que $\lim_{n\to\infty} \sum_{i=1}^{n} \mathbf{v}_i = \mathbf{v} \in \mathcal{P}$, escribiremos $\mathbf{v} = \sum_{i=1}^{\infty} \mathbf{v}_i$.

Lema XII.3.5. *Sean $\mathcal{P}$ un espacio prehilbertiano, $(\mathbf{u}_i)_{i\in\mathbb{N}}$ una sucesión ortonormal y $(\lambda_i)_{i\in\mathbb{N}}$ una sucesión de escalares tales que $\sum_{i=1}^{\infty} |\lambda_i|^2 < \infty$. La sucesión $(\mathbf{x}_n)_{n\in\mathbb{N}}$ de término general $\mathbf{x}_n = \sum_{i=1}^{n} \lambda_i \mathbf{u}_i$ es de Cauchy.*

Demostración. Basta tener en cuenta que, por la igualdad de Parseval (caso finito), se tiene que

$$\|\mathbf{x}_m - \mathbf{x}_n\|^2 = \left\| \sum_{i=n+1}^{m} \lambda_i \mathbf{u}_i \right\|^2 = \sum_{i=n+1}^{m} \|\lambda_i \mathbf{u}_i\|^2 = \sum_{i=n+1}^{m} |\lambda_i|^2,$$

para $m > n > 0$, que tiende a cero cuando n tiende hacia infinito. ∎

Teorema XII.3.6. *Sean $\mathcal{H}$ un espacio de Hilbert, $(\mathbf{u}_i)_{i\in\mathbb{N}}$ una sucesión ortonormal y $(\lambda_i)_{i\in\mathbb{N}}$ una sucesión de escalares. La serie $\sum_{i=1}^{\infty} \lambda_i \mathbf{u}_i$ es convergente si, y sólo si, la serie $\sum_{i=1}^{\infty} |\lambda_i|^2$ es convergente.*

Demostración. Si $\sum_{i=1}^{\infty} |\lambda_i|^2 < \infty$, entonces la sucesión $\mathbf{x}_n = \sum_{i=1}^{n} \lambda_i \mathbf{u}_i$ es una sucesión de Cauchy, por el lema XII.3.5. Esto implica la convergencia de la serie $\sum_{n=1}^{\infty} \lambda_i \mathbf{u}_i$ debido a la completitud de $\mathcal{H}$.

Recíprocamente, si la serie $\sum_{i=1}^{\infty} \lambda_i \mathbf{u}_i$ es convergente, entonces de la igualdad de Parseval (caso finito)

$$\left\| \sum_{i=n+1}^{m} \lambda_i \mathbf{u}_i \right\|^2 = \sum_{i=n+1}^{m} |\lambda_i|^2,$$

para $m > n > 0$, se sigue la convergencia de la serie $\sum_{i=1}^{\infty} |\lambda_i|^2$, pues los números $\mu_n = \sum_{i=1}^{n} |\lambda_i|^2$ forman una sucesión de Cauchy en $\mathbb{R}$. ∎

Proposición XII.3.7. *Sean $\mathcal{H}$ un espacio de Hilbert y $(\mathbf{u}_i)_{i\in\mathbb{N}}$ una sucesión ortonormal. Supongamos que $\mathbf{x} = \sum_{i=1}^{\infty} \lambda_i \mathbf{u}_i$ e $\mathbf{y} = \sum_{i=1}^{\infty} \mu_i \mathbf{u}_i$, en el sentido del teorema XII.3.6. Entonces,*

(a) $\mathbf{x} \cdot \mathbf{y} = \sum_{i=1}^{\infty} \lambda_i \bar{\mu}_i$, *siendo la serie absolutamente convergente.*
(b) $\mathbf{x} \cdot \mathbf{u}_i = \lambda_i$.
(c) $\|\mathbf{x}\|^2 = \sum_{i=1}^{\infty} |\lambda_i|^2 = \sum_{i=1}^{\infty} |\mathbf{x} \cdot \mathbf{u}_i|^2$.

Demostración. (a) Sean $\mathbf{x}_n = \sum_{i=1}^{n} \lambda_i \mathbf{u}_i$ e $\mathbf{y}_n = \sum_{i=1}^{n} \mu_i \mathbf{u}_i$. Por definición $\lim_{n\to\infty} \mathbf{x}_n = \mathbf{x}$ y $\lim_{n\to\infty} \mathbf{y}_n = \mathbf{y}$, de donde se sigue que $\mathbf{x}_n \cdot \mathbf{y}_n \to \mathbf{x} \cdot \mathbf{y}$, por el apartado (a) de la proposición XII.1.9. Dado que $\mathbf{x}_n \cdot \mathbf{y}_n = \sum_{i,j}^{n} \lambda_i \bar{\mu}_j (\mathbf{u}_i \cdot \mathbf{u}_j) = \sum_{i=1}^{n} \lambda_i \bar{\mu}_j$, se tiene que $\mathbf{x} \cdot \mathbf{y} = \sum_{i=1}^{\infty} \lambda_i \bar{\mu}_j$. Además, sustituyendo $(\lambda_i)_{i\in\mathbb{N}}$ y $(\mu_i)_{i\in\mathbb{N}}$ por $(|\lambda_i|)_{i\in\mathbb{N}}$ y $(|\mu_i|)_{i\in\mathbb{N}}$, respectivamente, resulta claro que la convergencia es absoluta.

(b) Es un caso particular del apartado (a), con $\mu_i = 1$ y $\mu_j = 0$, para todo $i \neq j$.

(c) Basta tomar $\mathbf{x} = \mathbf{y}$ en el apartado (a). ∎

De los resultados anteriores y del corolario XII.2.10 se sigue que en un espacio de Hilbert $\mathcal{H}$ la serie $\sum_{i=1}^{\infty} (\mathbf{x} \cdot \mathbf{u}_i)\mathbf{u}_i$ es convergente para todo $\mathbf{x} \in \mathcal{H}$, siendo $(\mathbf{u}_i)_{i\in\mathbb{N}}$ una sucesión ortonormal. Sin embargo, puede ocurrir que converja a un vector distinto de $\mathbf{x}$.

Supongamos, pues, que $(\mathbf{u}_n)_{n\in\mathbb{N}}$ es una sucesión ortogonal en un espacio de Hilbert $\mathcal{H}$. Dado $\mathbf{x} \in \mathcal{H}$, por el corolario XII.2.10, los escalares $\lambda_i = \mathbf{x} \cdot \mathbf{u}_i$, $i \in \mathbb{N}$, verifican que

$$\sum_{i=1}^{\infty} |\lambda_i|^2 \leq \|\mathbf{x}\|^2 < \infty.$$

Luego, de acuerdo con el teorema XII.3.6, se puede considerar el vector $\mathbf{y} = \sum_{i=1}^{\infty} \lambda_i \mathbf{u}_i$, y, por la proposición XII.3.7, $\mathbf{y} \cdot \mathbf{u}_i = \lambda_i = \mathbf{x} \cdot \mathbf{u}_i$, para todo $i \in \mathbb{N}$.

¿Cuándo se puede concluir que $\mathbf{x} = \mathbf{y}$? Desde luego se tiene que $(\mathbf{y} - \mathbf{x}) \cdot \mathbf{u}_i = \mathbf{y} \cdot \mathbf{u}_i - \mathbf{x} \cdot \mathbf{u}_i = 0$, para todo $i \in \mathbb{N}$; por lo tanto, se podría concluir que $\mathbf{x} = \mathbf{y}$ si los vectores de la sucesión $(\mathbf{u}_i)_{i\in\mathbb{N}}$ tuviesen la siguiente propiedad: *el único vector de $\mathcal{H}$ que es ortogonal a $\mathbf{u}_i$, para todo $i \in \mathbb{N}$, es el cero.*

Definición XII.3.8. Sea $\mathcal{H}$ un espacio de Hilbert. Se dice que un subconjunto arbitrario S de $\mathcal{H}$ es un **conjunto total** cuando el único vector $\mathbf{z} \in \mathcal{H}$ tal que $\mathbf{z} \cdot \mathbf{v} = 0$, para todo $\mathbf{v} \in S$, es $\mathbf{z} = \mathbf{0}$. En particular, una sucesión de vectores $(\mathbf{v}_i)_{i\in\mathbb{N}} \subset \mathcal{H}$ se llama **sucesión total** cuando

$$\mathbf{z} \cdot \mathbf{v}_i = 0, \text{ para todo } i \in \mathbb{N} \Longrightarrow \mathbf{z} = \mathbf{0}.$$

Aquí el nombre de total hace referencia a la siguiente propiedad: *un sistema ortogonal de un espacio de Hilbert $\mathcal{H}$ es total si, y sólo si, no está contenido en ningún otro sistema ortogonal de $\mathcal{H}$*, cuya comprobación proponemos como ejercicio al lector (ejercicio 12).

Ejemplos XII.3.9.

i) En un espacio prehilbertiano $\mathcal{P}$ cualquiera, el propio $\mathcal{P}$ es un conjunto total de vectores, pues si $\mathbf{z} \cdot \mathbf{x} = 0$, para todo $\mathbf{x} \in \mathcal{P}$, en particular $\mathbf{z} \cdot \mathbf{z} = 0$, luego $\mathbf{z} = \mathbf{0}$.

ii) Cualquier sistema de generadores $\mathcal{S}$ de un espacio prehilbertiano $\mathcal{P}$ es un conjunto total. En efecto, si $\mathbf{z} \in \mathcal{P}$ es ortogonal a todo vector de $\mathcal{S}$, será ortogonal a cualquier combinación lineal de vectores de $\mathcal{S}$; en particular, $\mathbf{z} \cdot \mathbf{z} = 0$, luego $\mathbf{z} = \mathbf{0}$.

iii) En el espacio de Hilbert ℓ^2, la sucesión de vectores

$$\mathbf{e}_1 = (1, 0, \ldots), \mathbf{e}_2 = (0, 1, 0, \ldots), \mathbf{e}_3 = (0, 0, 1, 0, \ldots), \ \ldots$$

es total. También lo es la sucesión

$$\mathbf{v}_1 = (1, 0, \ldots), \mathbf{v}_2 = (1, 1, 0, \ldots), \mathbf{v}_3 = (1, 1, 1, 0, \ldots), \ \ldots$$

Proposición XII.3.10. *Sea $\mathcal{H}$ un espacio de Hilbert. Entonces, una sucesión ortonormal $(\mathbf{u}_i)_{i\in\mathbb{N}}$ de vectores de $\mathcal{H}$ es total si, y sólo si,*

$$\mathbf{x} = \sum_{i=1}^{\infty} (\mathbf{x} \cdot \mathbf{u}_i)\mathbf{u}_i,$$

para todo $\mathbf{x} \in \mathcal{H}$.

Demostración. Si cada $\mathbf{x} \in \mathcal{H}$ admite la representación

$$\mathbf{x} = \sum_{i=1}^{\infty} (\mathbf{x} \cdot \mathbf{u}_i)\mathbf{u}_i,$$

entonces es claro que $\mathbf{x} \cdot \mathbf{u}_i = 0$, para todo $i \in \mathbb{N}$, implica que $\mathbf{x} = \mathbf{0}$.

Recíprocamente, sea $\mathbf{x} \in \mathcal{H}$ y supongamos que la sucesión ortonormal $(\mathbf{u}_i)_{i\in\mathbb{N}}$ es total. Sea

$$\mathbf{y} = \sum_{i=1}^{\infty} (\mathbf{x} \cdot \mathbf{u}_i)\mathbf{u}_i.$$

Esta suma existe en $\mathcal{H}$ por el corolario XII.2.10 y el teorema XII.3.6. Como, para todo $j \in \mathbb{N}$, se tiene que

$$\begin{aligned}(\mathbf{x}-\mathbf{y})\cdot\mathbf{u}_i &= \mathbf{x}\cdot\mathbf{u}_j - \left(\sum_{i=1}^{\infty}(\mathbf{x}\cdot\mathbf{u}_i)\mathbf{u}_i\right)\cdot\mathbf{u}_j = \mathbf{x}\cdot\mathbf{u}_j - \left(\sum_{i=1}^{\infty}(\mathbf{x}\cdot\mathbf{u}_i)\cdot(\mathbf{u}_i\cdot\mathbf{u}_j)\right)\\ &= \mathbf{x}\cdot\mathbf{u}_j - \mathbf{x}\cdot\mathbf{u}_j = 0,\end{aligned}$$

entonces, al ser $(\mathbf{u}_i)_{i\in\mathbb{N}}$ total se sigue que $\mathbf{x}-\mathbf{y}=\mathbf{0}$ y, por consiguiente, que $\mathbf{x} = \sum_{i=1}^{\infty}(\mathbf{x}\cdot\mathbf{u}_i)\mathbf{u}_i$. ∎

Igualdad de Parseval (caso general). *Una sucesión ortonormal $(\mathbf{u}_i)_{i\in\mathbb{N}}$ en un espacio de Hilbert $\mathcal{H}$ es total si, y sólo si,*

$$\|\mathbf{x}\|^2 = \sum_{i=1}^{\infty}|\mathbf{x}\cdot\mathbf{u}_i|^2, \tag{XII.3.5}$$

para todo $\mathbf{x}\in\mathcal{H}$.

Demostración. La implicación directa es consecuencia inmediata de las proposiciones XII.3.10 y XII.3.7(c). Recíprocamente, si se cumple (XII.3.5), el término de la derecha de la igualdad de Bessel,

$$\left\|\mathbf{u}-\sum_{i=1}^{n}(\mathbf{u}\cdot\mathbf{u}_i)\mathbf{u}_i\right\|^2 = \|\mathbf{u}\|^2 - \sum_{i=1}^{n}|\mathbf{u}\cdot\mathbf{u}_i|^2,$$

converge a cero cuando n tiende hacia infinito, y por lo tanto

$$\lim_{n\to\infty}\left\|\mathbf{u}-\sum_{i=1}^{n}(\mathbf{u}\cdot\mathbf{u}_i)\mathbf{u}_i\right\|^2 = 0;$$

de donde se sigue que la sucesión $(\mathbf{u}_i)_{i\in\mathbb{N}}$ es total, por la proposición XII.3.10. ∎

Definición XII.3.11. Se dice que una sucesión ortonormal $(\mathbf{u}_i)_{i\in\mathbb{N}}$ en un espacio de Hilbert $\mathcal{H}$ es una **base ortonormal** si todo $\mathbf{x}\in\mathcal{H}$ admite una única representación

$$\mathbf{x} = \sum_{i=1}^{\infty}\lambda_i\mathbf{u}_i,$$

con $\lambda_i \in \Bbbk$ para todo $i\in\mathbb{N}$.

Nota XII.3.12. Sea $\mathcal{H}$ un espacio de Hilbert que contiene un conjunto finito de vectores $\{\mathbf{u}_1,\ldots,\mathbf{u}_n\}$ que es ortonormal y total. Si $\mathbf{v}\in\mathcal{H}$ es un vector arbitrario, entonces $\mathbf{v}-\sum_{i=1}^{n}(\mathbf{v}\cdot\mathbf{u}_i)\,\mathbf{u}_i$ es ortogonal a $\mathbf{u}_i$, $i=1,\ldots,n$, y por lo tanto es nulo. Así, $\mathbf{v}=\sum_{i=1}^{n}(\mathbf{v}\cdot\mathbf{u}_i)\,\mathbf{u}_i$, de donde se sigue que $\{\mathbf{u}_1,\ldots,\mathbf{u}_n\}$ es una base de $\mathcal{H}$ y, por lo tanto, que $\mathcal{H}$ es de dimensión finita. Por consiguiente, *en un espacio de Hilbert de dimensión finita una base ortonormal es una base formada por vectores ortonormales.*

Ejemplo XII.3.13. La sucesión $(\mathbf{e}_n)_{n\in\mathbb{N}}$ descrita en el ejemplo XII.3.9 es una base ortonormal del espacio de Hilbert ℓ^2 que denominaremos **base usual** (o **base canónica**) de ℓ^2.

Proposición XII.3.14. *Una sucesión ortonormal en un espacio de Hilbert es base ortonormal si, y sólo si, es total.*

Demostración. Supongamos que $(\mathbf{u}_i)_{i\in\mathbb{N}}$ es una base ortonormal en un espacio de Hilbert $\mathcal{H}$. Sea $\mathbf{z}\in\mathcal{H}$, tal que $\mathbf{z}\cdot\mathbf{u}_i=0$, para todo $i\in\mathbb{N}$. Por ser $(\mathbf{u}_i)_{i\in\mathbb{N}}$ una base ortonormal, existen unos únicos $\lambda_j\in\Bbbk,\ j=1,2,\ldots$, tales que $\mathbf{z}=\sum_{j=1}^{\infty}\lambda_j\mathbf{u}_j$. Teniendo en cuenta que

$$0=\mathbf{z}\cdot\mathbf{u}_i=\left(\sum_{j=1}^{\infty}\lambda_j\mathbf{u}_j\right)\cdot\mathbf{u}_i=\lambda_i,$$

para todo $i\in\mathbb{N}$, concluimos que $\mathbf{z}=\mathbf{0}$.

Veamos ahora que una sucesión ortonormal total $(\mathbf{u}_i)_{i\in\mathbb{N}}$ en un espacio de Hilbert $\mathcal{H}$ es una base ortonormal de $\mathcal{H}$. En efecto, según la proposición XII.3.10 se tiene que

$$\mathbf{x}=\sum_{i=1}^{\infty}(\mathbf{x}\cdot\mathbf{u}_i)\mathbf{u}_i,$$

para todo $\mathbf{x}\in\mathcal{H}$; luego, basta comprobar la unicidad de tal representación. Si

$$\mathbf{x}=\sum_{i=1}^{\infty}\lambda_i\mathbf{u}_i,$$

para ciertos $\lambda_i\in\Bbbk$, entonces

$$\begin{aligned}0=\|\mathbf{x}-\mathbf{x}\|^2&=\left\|\sum_{i=1}^{\infty}(\mathbf{x}\cdot\mathbf{u}_i)\mathbf{u}_i-\sum_{i=1}^{\infty}\lambda_i\mathbf{u}_i\right\|^2=\left\|\sum_{i=1}^{\infty}\big((\mathbf{x}\cdot\mathbf{u}_i)-\lambda_i\mathbf{u}_i\big)\right\|^2\\&=\sum_{i=1}^{\infty}|(\mathbf{x}\cdot\mathbf{u}_i)-\lambda_i|^2,\end{aligned}$$

por la proposición XII.3.7. De donde se sigue que $(\mathbf{x}\cdot\mathbf{u}_i)=\lambda_i$, para todo $i\in\mathbb{N}$. ∎

Espacios de Hilbert separables.

No todos los espacios de Hilbert tienen bases ortonormales; a continuación vamos a dar una condición necesaria y suficiente para que un espacio de Hilbert tenga una base ortonormal. Pero antes necesitamos introducir algunos conceptos generales sobre espacios métricos.

Definición XII.3.15. Sea (X, d) un espacio métrico. Se dice que una sucesión $(x_n)_{n\in\mathbb{N}}$ de elementos de X es **densa**, si para cada $x \in X$ existe una subsucesión de $(x_n)_{n\in\mathbb{N}}$ que converge a x.

Definición XII.3.16. Se dice que un espacio métrico (X, d) es **separable**[2] si contiene alguna sucesión densa.

Dado que todo espacio de Hilbert es, en particular, un espacio métrico, diremos que un **espacio de Hilbert** es **separable** si es separable como espacio métrico.

Lema XII.3.17. *Toda sucesión densa en un espacio de Hilbert es total.*

Demostración. Sean $(\mathbf{v}_i)_{i\in\mathbb{N}}$ una sucesión densa en un espacio de Hilbert $\mathcal{H}$ y $\mathbf{z} \in \mathcal{H}$ tal que $\mathbf{z}\cdot\mathbf{v}_i = 0$, para todo $i \in \mathbb{N}$. Por hipótesis, existe una subsucesión $(\mathbf{v}'_i)_{i\in\mathbb{N}}$ de $(\mathbf{v}_i)_{i\in\mathbb{N}}$ convergente a $\mathbf{z}$. Luego, por la proposición XII.1.9, $\lim_{i\to\infty}(\mathbf{v}'_i\cdot\mathbf{z}) = \mathbf{z}\cdot\mathbf{z} = \|\mathbf{z}\|^2$; pero $\mathbf{v}'_i\cdot\mathbf{z} = 0$, para todo $n \in \mathbb{N}$. Por consiguiente, $\|\mathbf{z}\|^2 = 0$, es decir, $\mathbf{z} = 0$; y concluimos que la sucesión $(\mathbf{v}_i)_{i\in\mathbb{N}}$ es total. ∎

Teorema XII.3.18. *Sea $\mathcal{H}$ un espacio de Hilbert. Las siguientes condiciones son equivalentes:*

(a) *$\mathcal{H}$ es separable.*
(b) *$\mathcal{H}$ tiene una base ortonormal $(\mathbf{u}_i)_{i\in\mathbb{N}}$.*

Demostración. $\boxed{\text{(a)} \Rightarrow \text{(b)}}$ Si $\mathcal{H}$ es separable, entonces contiene alguna sucesión densa; luego, por el lema XII.3.17, contiene una sucesión total. Sea $(\mathbf{v}_i)_{i\in\mathbb{N}}$ una sucesión total de elementos de $\mathcal{H}$. De cualquier conjunto de vectores podemos extraer un subconjunto linealmente independiente que genera el mismo espacio vectorial, sea $\mathcal{S}$ un subconjunto linealmente independiente de $\{\mathbf{v}_i \mid i \in \mathbb{N}\}$ que genera al mismo espacio vectorial que $\{\mathbf{v}_i \mid i \in \mathbb{N}\}$. Es claro que $\mathcal{S}$ es total; en efecto, si $\mathbf{z}$ es ortogonal a todos los elementos de $\mathcal{S}$ también lo será a cualquier combinación lineal suya, y por lo tanto a todo $\mathbf{v}_i$, $i \in \mathbb{N}$, de donde se sigue que $\mathbf{z} = \mathbf{0}$. Si $\mathcal{S}$ no es un conjunto finito, podemos considerarlo como una subsucesión de $\{\mathbf{v}_i \mid i \in \mathbb{N}\}$. En cualquier caso, por el proceso de ortonormalización de Gram-Schmidt, existe un sistema ortonormal $\mathcal{B}$ que genera el mismo espacio vectorial que $\mathcal{S}$. Este sistema ortonormal es total, por el razonamiento anterior; en consecuencia, $\mathcal{B}$ es una base ortonormal de $\mathcal{H}$ (veánse la proposición XII.3.14 y la nota XII.3.12)

$\boxed{\text{(b)} \Rightarrow \text{(a)}}$ Sea $(\mathbf{u}_i)_{i\in\mathbb{N}}$ una sucesión ortonormal total en $\mathcal{H}$. Basta tener en cuenta que los elementos del subconjunto $\mathcal{S} = \{\alpha_1\mathbf{u}_1 + \ldots + \alpha_i\mathbf{u}_i \mid \alpha_i \in \mathbb{Q},\ i \in \mathbb{N}\}$ forman una sucesión densa en $\mathcal{H}$ si $\Bbbk = \mathbb{R}$, y el subconjunto $\mathcal{S} = \{(\alpha_1 + \mathrm{i}\beta_1)\mathbf{u}_1 + \ldots +$

[2]Recuérdese que un espacio topológico es denso si posee un subconjunto denso y numerable.

$(\alpha_i + \mathrm{i}\beta_i)\mathbf{u}_i \mid \alpha_i, \beta_i \in \mathbb{Q},\ i \in \mathbb{N}\}$ forman un sucesión densa en $\mathcal{H}$ si $\Bbbk = \mathbb{C}$; ya que $\lim_{n\to\infty} \sum_{i=1}^{n} (\mathbf{x} \cdot \mathbf{u}_i)\mathbf{u}_i = \mathbf{x}$, para todo $\mathbf{x} \in \mathcal{H}$. ∎

Ejemplo XII.3.19. Sea $\mathcal{H}$ el espacio vectorial de todas las funciones $f : \mathbb{R} \to \mathbb{C}$ que se anulan en todo $\mathbb{R}$ excepto en una cantidad numerable de puntos y tales que

$$\sum_{f(x)\neq 0} |f(x)|^2 < \infty.$$

$\mathcal{H}$ tiene estructura de espacio de Hilbert con el producto escalar

$$f \cdot g = \sum_{f(x)g(x)\neq 0} f(x)\overline{g(x)}.$$

Sin embargo, este espacio de Hilbert no es separable ya que para cualquier sucesión de funciones $(f_n)_{n\in\mathbb{N}}$ de $\mathcal{H}$ existen funciones no nulas $f \in \mathcal{H}$ tales que $f \cdot f_n = 0$, para todo $n \in \mathbb{N}$.

Nota XII.3.20. Se puede demostrar que todo espacio de Hilbert (separable o no) contiene un subconjunto ortonormal total $\mathcal{B}$; tal conjunto se llama *base ortonormal* del espacio. Sin embargo; puede ser imposible enumerar los elementos de $\mathcal{B}$ en forma de sucesión. Es más, en virtud del teorema XII.3.18, sólo podremos encontrar subconjuntos ortonormales totales numerables en los espacios de Hilbert separables.

Espacios de Hilbert isomorfos. El espacio de Hilbert clásico.

Definición XII.3.21. Se dice que un espacio de Hilbert $\mathcal{H}_1$ es **isomorfo** a un espacio de Hilbert $\mathcal{H}_2$ si existe una aplicación lineal biyectiva[3] $T : \mathcal{H}_1 \to \mathcal{H}_2$ tal que

$$T(\mathbf{x}) \cdot T(\mathbf{y}) = \mathbf{x} \cdot \mathbf{y},$$

para todo $\mathbf{x}$ e $\mathbf{y} \in \mathcal{H}_1$. La aplicación T se dice que es un **isomorfismo de espacios de Hilbert**.

Se comprueba fácilmente que el isomorfismo de espacios de Hilbert es una relación de equivalencia.

Teorema XII.3.22. *Sea $\mathcal{H}$ un espacio de Hilbert separable.*

(a) *Si $\mathcal{H}$ es de dimensión infinita, entonces es isomorfo a ℓ^2.*
(b) *Si $\mathcal{H}$ tiene dimensión $n > 0$, entonces es isomorfo a $\Bbbk^n$.*

[3]Esto es, un isomorfismo de espacios vectoriales.

Demostración. (a) Sea $(u_n)_{n\in\mathbb{N}}$ una sucesión ortonormal total en $\mathcal{H}$. Sea $\mathbf{x} \in \mathcal{H}$. Definimos $T(\mathbf{x}) = (\lambda_n)_{n\in\mathbb{N}}$, donde $\lambda_n = \mathbf{x}\cdot\mathbf{u}_n,\ i = 1, 2, \ldots$. Por el teorema XII.3.6, T es una aplicación biyectiva de $\mathcal{H}$ a ℓ^2. Se comprueba fácilmente que es lineal. Además, para $\lambda_n = \mathbf{x}\cdot\mathbf{u}_n$, y $\mu_n = \mathbf{y}\cdot\mathbf{u}_n$, con $\mathbf{x}$ e $\mathbf{y} \in \mathcal{H}$ y $n \in \mathbb{N}$, se tiene que

$$T(\mathbf{x})\cdot T(\mathbf{y}) = (\lambda_n)_{n\in\mathbb{N}}\cdot(\mu_n)_{n\in\mathbb{N}} = \sum_{n=1}^{\infty}\lambda_n\bar{\mu_n} = \sum_{n=1}^{\infty}(\mathbf{x}\cdot\mathbf{u}_n)\overline{(\mathbf{y}\cdot\mathbf{u}_n)}$$

$$= \sum_{n=1}^{\infty}(\mathbf{x}\cdot\big((\mathbf{y}\cdot\mathbf{u}_n)\mathbf{u}_n\big) = \mathbf{x}\cdot\left(\sum_{n=1}^{\infty}(\mathbf{y}\cdot\mathbf{u}_n)\mathbf{u}_n\right) = \mathbf{x}\cdot\mathbf{y}.$$

Así, concluimos que T es un isomorfismo de $\mathcal{H}$ a ℓ^2.

(b) La demostración de este apartado se deja como ejercicio al lector. ∎

Como cualquier espacio de Hilbert separable de dimensión infinita sobre los complejos es isomorfo al espacio ℓ^2 complejo, se sigue que cualesquiera dos espacios de Hilbert de este tipo son isomorfos. Lo mismo ocurre para los espacios de Hilbert reales; cualquier espacio de Hilbert separable de dimensión infinita es isomorfo al espacio ℓ^2 sobre $\mathbb{R}$. De modo que, en cierto sentido, *existe un único espacio de Hilbert separable de dimensión infinita real y un único espacio de Hilbert separable de dimensión infinita complejo,* que se llaman **espacios de Hilbert clásicos real** y **complejo**, respectivamente.

Ejercicios del tema XII

Ejercicio 1. Comprobar que los espacios prehilbertianos del ejemplo XII.1.2 son efectivamente espacios prehilbertianos.

Ejercicio 2. Sea $V = \mathcal{M}_n(\mathbb{C})$. Probar que la aplicación $V \times V \longrightarrow \mathbb{C};\ (A, B) \longmapsto \operatorname{tr}(B^*A)$, donde B^* es la matriz adjunta (esto es, la traspuesta conjugada) de B es un producto escalar.

Ejercicio 3. Probar que, en cualquier espacio prehilbertiano, se cumple que

$$\|\mathbf{w}-\mathbf{u}\|^2 + \|\mathbf{w}-\mathbf{v}\|^2 = \frac{1}{2}\|\mathbf{u}-\mathbf{v}\|^2 + 2\left\|\mathbf{w} - \frac{\mathbf{u}+\mathbf{v}}{2}\right\|^2,$$

para todo $\mathbf{u}, \mathbf{v}$ y $\mathbf{w}$. Esta igualdad se conoce como *identidad de Apolonio*.

Ejercicio 4. Sea $(V, \|\cdot\|)$ un espacio normado. Probar que $\|\cdot\|$ proviene de un producto escalar si, y sólo si, cumple la regla del paralelogramo. En este caso, probar que

1. si V está definido sobre los reales,
$$\mathbf{u}\cdot\mathbf{v} = \frac{1}{4}\Big(\|\mathbf{u}+\mathbf{v}\|^2 - \|\mathbf{u}-\mathbf{v}\|^2\Big).$$
2. si V está definido sobre los complejos
$$\mathbf{u}\cdot\mathbf{v} = \frac{1}{4}\Big(\|\mathbf{u}+\mathbf{v}\|^2 - \|\mathbf{u}-\mathbf{v}\|^2 + \mathrm{i}\,\|\mathbf{u}+\mathrm{i}\,\mathbf{v}\|^2 - \mathrm{i}\,\|\mathbf{u}-\mathrm{i}\,\mathbf{v}\|^2\Big).$$

La igualdades anteriores se conocen como *identidades de polarización*.

Ejercicio 5. Probar que, en cualquier espacio prehilbertiano, $\|\mathbf{u}-\mathbf{v}\|+\|\mathbf{v}-\mathbf{w}\| = \|\mathbf{u}-\mathbf{w}\|$ si, y sólo si, $\mathbf{v} = \alpha\mathbf{u} + (1-\alpha)\mathbf{w}$, para algún $\alpha \in [0,1]$.

Ejercicio 6. Sean $\mathcal{P}$ un espacio prehilbertiano y $(x_n)_{n\in\mathbb{N}}$ e $(y_n)_{n\in\mathbb{N}}$ dos sucesiones de elementos de $\mathcal{P}$. Probar que, si $\lim_{n\to\infty} x_n = 0$ e $(y_n)_{n\in\mathbb{N}}$ es acotada, entonces $\lim_{n\to\infty}(x_n \cdot y_n) = 0$.

Ejercicio 7. En el espacio prehilbertiano de las sucesiones eventualmente nulas, ortonormalizar la sucesión de vectores $\mathbf{v}_1 = (1, 0, \dots,)$, $\mathbf{v}_2 = (1, 1, 0, \dots,)$, $\mathbf{v}_3 = (1, 1, 1, 0, \dots,)$, $\dots$

Ejercicio 8. Sea $\mathcal{P}$ el espacio prehilbertiano de las funciones continuas en el intervalo $[-1, 1]$. Probar que

1. La sucesión $(x_n)_{n\in\mathbb{N}}$ de término general
$$x_n(t) = \begin{cases} 0 & \text{si} \quad -1 \le t \le 0; \\ nt & \text{si} \quad 0 < t < 1/n; \\ 1 & \text{si} \quad 1/n \le t \le 1. \end{cases}$$

es de Cauchy.
2. La sucesión anterior no es convergente en $\mathcal{P}$.
3. $\mathcal{P}$ no es un espacio de Hilbert.

Ejercicio 9. Sea $\mathcal{H} = \mathcal{C}^1([a,b])$, esto es, el espacio vectorial de las funciones reales diferenciables de derivada continua en $[a,b]$.

1. Para f y $g \in \mathcal{H}$ se define
$$f \cdot g = \int_a^b f'(x)g'(x)\mathrm{d}x$$
¿Es $\cdot$ un producto escalar en $\mathcal{H}$?
2. Sea $\mathcal{H}' = \{f \in \mathcal{H} \mid f(a) = 0\}$. ¿Es $\cdot$ un producto escalar en $\mathcal{H}'$? ¿Es $\mathcal{H}'$ un espacio de Hilbert?

Ejercicio 10. Probar que para cualquier x en un espacio de Hilbert se cumple que
$$\|x\| = \sup_{\|y\|=1}|x \cdot y|.$$

Ejercicio 11. Sean $\mathcal{H}_1, \ldots, \mathcal{H}_n$ espacios prehilbertianos y $\mathcal{H} = \mathcal{H}_1 \times \ldots \times \mathcal{H}_n$. Probar que

1. Si $\mathbf{x} = (x_1, \ldots, x_n)$ e $\mathbf{y} = (y_1, \ldots, y_n) \in \mathcal{H}$, entonces
$$\mathbf{x} \cdot \mathbf{y} = x_1 \cdot y_1 + \ldots + x_n \cdot y_n,$$
define un producto escalar en $\mathcal{H}$.
2. Si $\mathcal{H}_1, \ldots, \mathcal{H}_n$ son espacios de Hilbert, entonces $\mathcal{H}$ tiene una estructura natural de espacio de Hilbert donde la norma de $\mathbf{x} = (x_1, \ldots, x_n) \in \mathcal{H}$ es
$$\|\mathbf{x}\| = \sqrt{\|x_1\|^2 + \ldots + \|x_n\|^2}.$$

Ejercicio 12. Probar que un sistema ortogonal de un espacio de Hilbert $\mathcal{H}$ es completo si, y sólo si, no está contenido en ningún otro sistema ortogonal de $\mathcal{H}$.

PRÁCTICA 1

Vectores y MATLAB

ESTA y todas las demás prácticas están pensadas para ser trabajadas delante de un ordenador con MATLAB instalado, y no para ser leídas como una novela. En vez de eso, cada vez que se presente un comando de MATLAB, se debe introducir el comando, pulsar la tecla "Enter" para ejecutarlo y ver el resultado. Más aún, se desea que se verifique el resultado. Asegúrese de que se comprende perfectamente lo que se obtiene antes de continuar con la lectura.

Aunque MATLAB es un entorno que trabaja con matrices, en esta práctica se aprenderá cómo introducir vectores por filas o por columnas y a manejar algunas operaciones con vectores.

Prerrequisitos: ninguno.

1. Vectores fila

La introducción de vectores fila en MATLAB es muy fácil. Introdúzcase el siguiente comando en la pantalla de MATLAB [1]

```
>> v=[1 2 3]
```

Hay una serie de ideas a destacar en este comando. Para introducir un vector, se escribe una apertura de corchete, los elementos del vector separados por espacios y un cierre de corchete. Se pueden usar también comas para delimitar las componentes del vector

```
>> v=[1,2,3]
```

El signo = es el operador de asignación de MATLAB. Se usa este operador para asignar valores a variables. Para comprobar que el vector fila [1,2,3] ha sido asignado a la variable v introdúzcase el siguiente comando en el indicador de MATLAB.

[1] El símbolo >> es el indicador de MATLAB. Se debe introducir lo que aparece tras el indicador. Entonces se pulsa la tecla "Enter" para ejecutar el comando.

```
>> v
```

1.1. Rangos.

Algunas veces es necesario introducir un vector con componentes a intervalos regulares. Esto se realiza fácilmente con MATLAB con la estructura `inicio:incremento:fin`. Si no se proporciona un incremento, MATLAB asume que es 1.

```
>> x1=0:10
```

Se puede seleccionar el propio incremento.

```
>> x2=0:2:10
```

Se puede ir incluso hacia atrás.

```
>> x3=10:-2:1
```

O se le puede echar imaginación.

```
>> x4=0:pi/2:2*pi
```

Hay veces, sobre todo cuando hay que pintar funciones, que se precisan un gran número de componentes en un vector.

```
>> x=0:.1:10
```

1.2. Elimina la salida.

Se puede suprimir la salida de un comando de MATLAB añadiendo un punto y coma.

```
>> x=0:.1:10;
```

Es muy útil cuando la salida es muy grande y no se desea verla.

1.3. Espacio de trabajo de MATLAB.

Es posible obtener una lista de las variables en el espacio de trabajo en cualquier momento mediante el comando

```
>> who
```

Se puede obtener incluso más información acerca de las variables con

```
>> whos
```

Se eliminar la asignación hecha a una variable con

```
>> clear x
>> who
```

Obsérvese que también se da el tamaño de cada variable. Es posible mantener una ventana con la lista de variables usadas y su tamaño. Para ello, en la barra superior selecciónese el menú `Desktop` y actívese la opción `Workspace`.

Se puede obtener el tamaño de un vector `v` con el comando

```
>> size(v)
```

La información que devuelve indica que el vector `v` tiene 1 fila y 3 columnas. Aunque se puede entender al vector `v` como una matriz con 1 fila y 3 columnas, también se puede entender como un vector fila de longitud 3. Por ejemplo, pruébese el siguiente comando:

```
>> length(v)
```

2. Vectores columna

Es también fácil escribir vectores columna en MATLAB. Introdúzcase el siguiente comando en el indicador.

```
>> w=[4;5;6]
```

Observe que los símbolos de punto y coma delimitan las filas de un vector columna. Pruébense los siguientes comandos.

```
>> w
>> who
>> whos
>> size(w)
```

El resultado indica que el vector `w` tiene 3 filas y 1 columna. Aunque se puede ver al vector `w` como una matriz de 3 filas y 1 columna, también es posible pensar en él como un vector columna de longitud 3. Pruébese el siguiente comando.

```
>> length(w)
```

2.1. Transposición.

El operador en MATLAB para transponer es el apóstrofe simple '. Se puede cambiar así un vector fila a un vector columna.

```
>> y=(1:10)'
```

O un vector columna a un vector fila.

```
>> y=y'
```

2.2. Indexado de vectores.

Una vez que se ha definido un vector, es posible acceder fácilmente a cada una de sus componentes con los comandos de MATLAB. Por ejemplo, introdúzcase el siguiente vector.

```
>> x=[10,13,19,23,27,31,39,43,51]
```

Ahora pruébense los siguientes comandos.

```
>> x(2)
>> x(7)
```

Se puede cambiar fácilmente el contenido de una componente.

```
>> x(6)=100
```

Se puede también acceder a un rango de elementos

```
>> x([1,3,5])
>> x(1:3)
>> x(1:2:length(x))
```

3. Operaciones con vectores

Un gran número de operaciones en las que intervienen vectores y escalares se pueden ejecutar con MATLAB.

3.1. Operaciones entre vector y escalar.

Las operaciones entre escalares y vectores son directas. Desde el punto de vista teórico, no se puede sumar un escalar a un vector. Sin embargo, MATLAB sí lo permite. Por ejemplo, si y es un vector, el comando y+2 añadirá 2 a cada componente del vector. Estúdiense las salidas de los siguientes comandos.

```
>> y=1:5
>> y+2
>> y-2
>> 2*y
>> y/2
```

Por supuesto, estas operaciones son igualmente válidas para vectores columna.

```
>> w=(1:3:20)'
>> w+3
>> w-11
>> .1*w
>> w/10
```

3.2. Operaciones entre vectores.

En primer lugar, considérense los siguientes vectores.

```
>> a=1:3
>> b=4:6
```

La adición y sustracción de vectores es natural y fácil. Introdúzcanse los siguientes comandos.[2]

```
>> a,b,a+b
>> a,b,a-b
```

De nuevo, estas operaciones son válidas para vectores columna.

```
>> a=(1:3)',b=(4:6)'
>> a+b,a-b
```

Sin embargo, se pueden obtener resultados no esperados si no se recuerda que MATLAB es un entorno que trabaja con matrices.

```
>> a,b,a*b
```

El último comando devuelve un error porque $*$ es el símbolo de MATLAB para la multiplicación de matrices, y en este caso hay un problema de compatibilidad entre los ordenes de las "matrices" a y b. También pueden ocurrir errores si se intenta añadir vectores de diferente tamaño.

```
>> a=1:3,b=4:7,a+b
```

3.3. Operaciones con componentes.

Para multiplicar los vectores a y b componente a componente, ejecútese el siguiente comando de MATLAB.

[2]Como no aparece punto y coma que suprima la salida, el comando a,b,a+b mostrará primero el vector a, luego el vector b y por último el a+b

```
>> a=(1:3)',b=(4:6)'
>> a,b,a.*b
```

El símbolo .∗ es el operador de MATLAB para la multiplicación elemento a elemento. La salida se calcula multiplicando las primeras componentes de los vectores a y b, a continuación las segundas componentes, etc. El operador de MATLAB para la división componente a componente es ./

```
>> a,b,a./b
```

Para elevar cada componente de un vector a una potencia, úsese .^

```
>> a,a.^2
```

3.4. Expresiones más complicadas.

Con un poco de práctica se aprenderá como evaluar expresiones más complejas. Supongamos, por ejemplo, para evaluar la expresión $x^2 - 2x - 3$ para valores de x entre 1 y 10, con incremento de 1 escríbase

```
>> x=1:10
>> y=x.^2-2*x-3
```

Supóngase ahora que se quiere evaluar la expresión $\text{sen}(x)/x$ para valores de x entre -1 y 1 con incrementos de 0,1 unidades.[3]

```
>> x=-1:.1:1
>> y=sin(x)./x
```

Los operadores por componentes también funcionan con vectores columna.

```
>> xdata=(1:10)'
>> xdata.^2
```

[3]Escribiendo `help elfun` se obtiene una lista de las funciones elementales de MATLAB.

Ejercicios de la práctica 1

Ejercicio 1. Escribe el comando `MATLAB` que genera cada uno de los siguientes vectores.

1. $\begin{pmatrix} 1 \\ 2 \\ -3 \end{pmatrix}$.
2. $(1, 2, -1, 3)$.
3. Un vector columna que contenga los números impares entre 1 y 1000.
4. Un vector fila que contenga los números pares entre 2 y 1000.

Ejercicio 2. Si `x=0:2:20`, escribe el comando de `MATLAB` que eleva al cuadrado cada componente de `x`.

Ejercicio 3. Si `x=[0,1,4,9,16,25]`, escribe el comando `MATLAB` que calcula la raíz cuadrada de cada componente de `x`.

Ejercicio 4. Si `x=0:.1:1`, escribe el comando de `MATLAB` que eleva cada componente de `x` a 2/3.

Ejercicio 5. Si `x=0:pi/2:2*pi`, escribe el comando `MATLAB` que calcula el coseno de cada componente de `x`.

Ejercicio 6. Si `x=-1:.1:1`, escribe el comando `MATLAB` que calcula el arcoseno de cada componente de `x`.

Ejercicio 7. Si `x=linspace(0,2*pi,1000)`, ¿cuál es la entrada 50 de `x`? ¿Cuál es la longitud de `x`?

Ejercicio 8. Si `k=0:100`, ¿cuál es la entrada número 12 de `y=0.5.^k`?

PRÁCTICA 2

Matrices y MATLAB

En esta práctica se aprenderá a introducir y editar matrices en MATLAB. Se experimentará con algunas funciones de construcción de matrices incorporadas en MATLAB. Se aprenderá a construir matrices a partir de vectores y bloques de matrices.

Prerrequisitos: ninguno.

1. Entrada de matrices

La entrada de matrices en MATLAB es fácil. Escríbase lo siguiente en el indicador de MATLAB.

```
>> A=[1,2,3;4,5,6;7,8,9]
```

Obsérvese cómo los símbolos de punto y coma indican el final de la fila, mientras que las comas se usan para separar las entradas en la fila. Se pueden usar también espacios para delimitar las entradas de cada fila.

```
>> A=[1 2 3;4 5 6;7 8 9]
```

1.1. Matrices especiales.

MATLAB tiene una serie de rutinas incorporadas para crear matrices.[1] Es posible crear una matriz de ceros de cualquier tamaño.

```
>> A=zeros(5)
>> B=zeros(3,5)
```

Es fácil crear una matriz de ceros con el mismo tamaño que una dada.

[1]Para obtener una lista de todas las matrices elementales de MATLAB, escríbase `help elmat` en el indicador de MATLAB; para obtener información detallada sobre una en concreto escríbase `help` seguido del tipo de matriz, por ejemplo, `help magic`.

```
>> C=magic(5)
>> D=zeros(size(C))
```

Se pueden crear matrices de unos de manera análoga.

```
>> A=ones(6)
>> B=ones(2,10)
>> C=hilb(5)
>> D=ones(size(C))
```

Cuando se realizan simulaciones en MATLAB es útil construir matrices de números aleatorios. Se puede crear una matriz de números aleatorios con distribución uniforme, cada uno entre 0 y 1, con los siguientes comandos.

```
>> A=rand(6)
>> B=rand(5,3)
```

La multiplicación por escalares es exactamente igual que para vectores.

```
>> C=10*rand(5)
```

MATLAB proporciona unas rutinas para el redondeo de números.

```
>> D=floor(C)
>> D=ceil(C)
>> D=round(C)
>> D=fix(C)
```

La matriz identidad tiene unos en su diagonal principal y ceros en el resto.

```
>> I=eye(5)
```

Se pueden generar otros tipos de matrices diagonales con el comando diag.

```
>> E=diag([1,2,3,4,5])
>> F=diag([1,2,3,4,5],-1)
```

```
>> G=diag(1:5,1)
```

1.2. Trasposición.

El operador de trasposición, que es ’ (comilla simple), tiene el mismo efecto que sobre vectores. Se intercambian filas y columnas.

```
>> J=[1 2 3;4 5 6;7 8 9]
>> J’
```

1.3. Elimina la salida.

Recuérdese que finalizando un comando de MATLAB con punto y coma se elimina la salida. Es útil cuando el resultado es grande y se desea ocultarlo.

```
>> K=rand(100);
```

1.4. Espacio de trabajo de MATLAB.

Examínese el espacio de trabajo con el comando `whos`, o activando la opción “Workspace” del menú “View” de la barra superior.

```
>> whos
```

Obsérvese que aparece el tamaño de cada una de las variables. Por supuesto, se puede obtener el tamaño de la matriz `I` con

```
>> size(I)
```

2. Indexado de matrices

La siguiente notación es la que se usa para representar una matriz con 3 filas y 3 columnas.

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

o en forma reducida $A = (a_{ij}) \in \mathcal{M}_3(\Bbbk)$, donde $\Bbbk$ es cuerpo (por ejemplo, $\Bbbk = \mathbb{R}$ o $\Bbbk = \mathbb{C}$. El símbolo a_{ij} se refiere a la entrada situada en la fila i y columna j. MATLAB usa una notación similar para representar los elementos de una matriz.

```
%pascal no funciona en Octave

>> A=pascal(5)
>> A(1,2)
>> A(3,4)
```

En general, `A(i,j)` se refiere al elemento de la fila i, columna j de la matriz A. También es fácil cambiar el valor de una entrada.

```
>> A(3,3)=11111
```

2.1. Algo más sobre indexado.

Cuando se indexa una matriz, los subíndices pueden ser vectores. Esta es una herramienta de gran alcance que permite extraer fácilmente una submatriz de una matriz.

```
>> A=magic(6)
>> A([1,2],[3,4,5])
```

La notación `A([1,2],[3,4,5])` referencia a la submatriz formada por los elementos que aparecen en las filas 1 y 2 y en las columnas 3, 4 y 5 de la matriz A.

El comando

```
>> A([1,3,5],[1,2,3,4,5,6])
```

produce una submatriz con las filas 1, 3 y 5 de la matriz A. Si se recuerda que la notación `1:6` representa al vector `[1,2,3,4,5,6]` y que la notación `1:2:6` representa al vector `[1,3,5]`, de este modo se tiene que `A([1:2:6],[1:6])` es equivalente a `A([1,3,5],[1,2,3,4,5,6])`.

```
>> A([1:2:6],[1:6])
```

Si se usa el símbolo dos puntos en lugar de subíndices, se indica todo el rango. Así,

```
>> A(:,1)
```

produce la primera columna de la matriz A, y

```
>> A(3,:)
```

genera la tercera fila de la matriz A. En cierto sentido, la notación `A(3,:)` se puede leer como "Tercera fila, todas las columnas." El comando

```
>> A(1:3,:)
```

produce una submatriz compuesta de las tres primeras filas de la matriz A. El comando

```
>> A(:,1:2:6)
```

produce una submatriz compuesta de las columnas 1, 3 y 5 de la matriz A.

3. Construcción de matrices

Con MATLAB se pueden crear matrices más complejas a partir de otras matrices y vectores.

3.1. Construcción de matrices con vectores.

Créense tres vectores fila con los comandos

```
>> v1=1:3
>> v2=4:6
>> v3=7:9
```

El comando

```
>> M=[v1;v2;v3]
```

construye una matriz con los vectores `v1, v2` y `v3`, cada uno formando una fila de la matriz M. El comando

```
>> N=[v1,v2,v3]
```

produce un resultado completamente diferente, pero con sentido.

Cámbiense los vectores `v1,v2,v3` en vectores columna con el operador de trasposición.

```
>> v1=v1'
>> v2=v2'
>> v3=v3'
```

El comando

```
>> P=[v1,v2,v3]
```

construye una matriz con los vectores `v1,v2,v3` como columnas de la matriz P. Se puede obtener el mismo resultado con la transpuesta de la matriz M.

```
>> P=M'
```

Téngase en cuenta que **las dimensiones deben coincidir:** cuando se construyen matrices, hay que asegurarse que cada fila y columna tengan el mismo número de elementos. Por ejemplo, la siguiente secuencia de comandos producirá un error.

```
>> w1=1:3;w2=4:6;w3=7:10;
>> Q=[w1;w2;w3]
```

3.2. Construcción de matrices con otras matrices.

Es una cuestión simple aumentar una matriz con un vector fila o columna. Por ejemplo,

```
>> A=[1,2,3,4;5,6,7,8;9,10,11,12]
>> b=[1,1,1]'
>> M=[A,b]
```

es válido, pero

```
>> M=[A;b]
```

no lo es; aunque sí lo es

```
>> c=[1,1,1,1]
>> M=[A;c]
```

Se pueden concatenar dos o más matrices. Así,

```
>> A=magic(3),B=ones(3,4)
>> M=[A,B]
```

es válido, pero

```
>> N=[A;B]
```

no lo es; aunque sí lo es

```
>> C=[1,2,3;4,5,6]
>> P=[A;C]
```

3.3. La imaginación es el límite.

Las capacidades de construir matrices de MATLAB son muy flexibles. Considérese el siguiente ejemplo.

```
>> A=zeros(3),B=ones(3),C=2*ones(3),D=3*ones(3)
>> M=[A,B;C,D]
```

Se puede construir una matriz de Vandermonde de la siguiente manera

```
>> x=[1,2,3,4,5]'
>> N=[ones(size(x)),x,x.^2,x.^3,x.^4]
```

O también matrices por bloques

```
>> B=zeros(8)
>> B(1:3,1:3)=[1,2,3;4,5,6;7,8,9]
>> B(4:8,4:8)=magic(5)
```

Ejercicios de la práctica 1

Ejercicio 1. Escribe el comando `MATLAB` que genera cada uno de los siguientes vectores.

1. $\begin{pmatrix} 1 \\ 2 \\ -3 \end{pmatrix}$.
2. $(1, 2, -1, 3)$.
3. Un vector columna que contenga los números impares entre 1 y 1000.
4. Un vector fila que contenga los números pares entre 2 y 1000.

Ejercicio 2. Si `x=0:2:20`, escribe el comando de `MATLAB` que eleva al cuadrado cada componente de `x`.

Ejercicio 3. Si `x=[0,1,4,9,16,25]`, escribe el comando `MATLAB` que calcula la raíz cuadrada de cada componente de `x`.

Ejercicio 4. Si `x=0:.1:1`, escribe el comando de `MATLAB` que eleva cada componente de `x` a 2/3.

Ejercicio 5. Si `x=0:pi/2:2*pi`, escribe el comando `MATLAB` que calcula el coseno de cada componente de `x`.

Ejercicio 6. Si `x=-1:.1:1`, escribe el comando `MATLAB` que calcula el arcoseno de cada componente de `x`.

Ejercicio 7. Si `x=linspace(0,2*pi,1000)`, ¿cuál es la entrada 50 de `x`? ¿Cuál es la longitud de `x`?

Ejercicio 8. Si `k=0:100`, ¿cuál es la entrada número 12 de `y=0.5.^k`?

PRÁCTICA 3

Formas escalonadas de una matriz

En esta práctica aprenderemos a manejar el comando `rref` de `MATLAB`, que calcula la forma escalonada por filas de una matriz; también se verán algunas de sus aplicaciones.

Prerrequisitos: cierta familiaridad con cálculos a mano de la forma escalonada por filas de una matriz.

1. Resolución de sistemas con `MATLAB`

Hasta ahora, hemos invertido cierto tiempo para resolver sistemas de ecuaciones lineales a mano, con lo que advertimos que es un proceso largo y con tendencia a que se produzcan errores. En cuanto la matriz de coeficientes es de un tamaño superior a 5×5, lo más probable es que nos equivoquemos en el resultado. Vamos a ver cómo puede `MATLAB` ayudarnos en el proceso.

En primer lugar, recordemos algunas definiciones. El primer elemento no nulo en cada fila de una matriz se denomina **pivote**. Una matriz se dice que está en forma *escalonada* por filas si

- Las filas de ceros aparecen en la parte inferior de la matriz.
- Cada pivote es 1.
- Cada pivote aparece en una columna estrictamente a la derecha del pivote de la fila anterior.

Se dice que una matriz está en forma *escalonada* por filas si satisface además otra propiedad

- Cada pivote es el único elemento no nulo en su columna.

Se sabe que *toda matriz es equivalente a una matriz en forma escalonada por filas*, es decir, que mediante transformaciones elementales (por filas) toda matriz se puede convertir en una matriz escalonada por filas. De hecho la forma escalonada por filas de una matriz se diferencia de la forma reducida por filas en que en esta última se permiten las permutaciones de columnas.

Por otra parte, es de sobra conocido que cuando se resuelve un sistema de ecuaciones de la forma

$$\left.\begin{array}{r} a_{11}x_1,+a_{12}x_2,+\ldots,+a_{1n}x_n,=,b_1 \\ a_{21}x_1,+a_{22}x_2,+\ldots,+a_{2n}x_n,=,b_2 \\ ,,,,\vdots, \\ a_{m1}x_1,+a_{m2}x_2,+\ldots,+a_{mn}x_n,=,b_m \end{array}\right\} \tag{3.1.1}$$

puede ocurrir que

- el sistema tenga una única solución, o
- el sistema no tenga solución, o
- el sistema tenga infinitas soluciones.

Veamos un ejemplo de cada caso.

1.1. Solución única.

Consideremos el sistema

$$\left.\begin{array}{r} x_1,+x_2,+x_3,=,6 \\ x_1,,-2x_3,=,4 \\ ,x_2,+x_3,=,2 \end{array}\right\} \tag{3.1.2}$$

La matriz ampliada de este sistema es

$$\left(\begin{array}{rrr} 1,1,1,6 & & \\ 1,0,-2,4 & & \\ & 0,1 & 1,2 \end{array}\right), \tag{3.1.3}$$

que podemos introducirla en el espacio de trabajo de `MATLAB` con

```
>> A=[1,1,1,6;1,0,-2,4;0,1,1,2]
```

El comando `rref` de `MATLAB` calcula la forma escalonada por filas de la matriz A.

```
>> R=rref(A)
```

El commando `rrefmovie` de `MATLAB` nos muestra paso a paso cómo ha obtenido la forma escalonada por filas.

```
>> rrefmovie(A)
```

Hemos obtenido que la forma escalonada por filas de la matriz ampliada (3.1.3) es

$$\left(\begin{array}{cc} 1,0,0,4 & \\ 0,1,0,2 & \\ 0,0 & 1,0 \end{array}\right). \tag{3.1.4}$$

Esta matriz representa al sistema

$$\left.\begin{array}{c} x_1,,,=,4 \\ ,x_2,,=,2 \\ ,,x_3,=,0 \end{array}\right\} \tag{3.1.5}$$

que es equivalente al sistema (3.1.2). Por tanto, el sistema (3.1.2) tiene solución única $(4,2,0)$.

Es interesante considerar la geometría de este ejemplo. Cada una de las ecuaciones del sistema (3.1.2) representa un plano en el espacio de 3 dimensiones. Como se puede ver en la Figura (1), las tres ecuaciones del sistema (3.1.2) producen tres planos. Observemos además que la intersección de los tres planos en la Figura (1) es un único punto, lo que coincide con nuestro resultado.

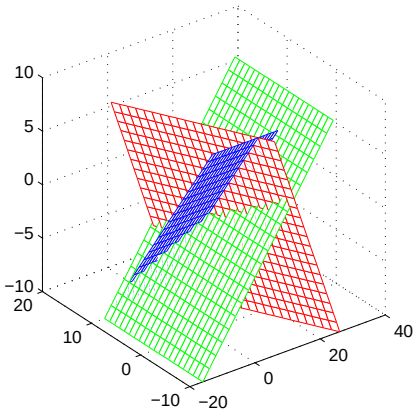


FIGURA 1. Un sistema con solución única. Los tres planos se cortan en un punto.

1.2. Sin soluciones.

Consideremos ahora el sistema

$$\left.\begin{array}{c} x_1,+x_2,+x_3,=,-6 \\ x_1,,-2x_3,=,4 \\ 2x_1,+x_2,-x_3,=,18 \end{array}\right\} \tag{3.1.6}$$

La matriz ampliada del sistema es

$$\begin{pmatrix} 1,1,1,-6 \\ 1,0,-2,4 \\ 2,1,-1,18 \end{pmatrix}, \tag{3.1.7}$$

que podemos introducirla en MATLAB con el comando

```
>> A=[1,1,1,-6;1,0,-2,4;2,1,-1,18]
```

Usamos el comando `rref` para calcular la forma escalonada por filas.

```
>> R=rref(A)
```

Por tanto, la forma escalonada por filas de la matriz (3.1.7) es

$$\begin{pmatrix} 1,0,-2,0 & \\ 0,1,3,0 & \\ 0,0 & 0,1 \end{pmatrix} \tag{3.1.8}$$

Observemos la última fila de la matriz 3.1.8. Representa la ecuación

$$0x_1 + 0x_2 + 0x_3 = 1 \tag{3.1.9}$$

Es claro que la ecuación 3.1.9 no tiene solución. Por tanto, el sistema 3.1.6 tampoco. Decimos que el sistema 3.1.6 es *incompatible*.

De nuevo, la representación geométrica aporta luz a lo anterior. Como podemos ver en la figura 2, cada plano corta a otro en una recta, pero esa recta es paralela al otro plano. Por tanto, no hay puntos comunes a los tres planos, que coincide con nuestro resultado algebraico.

1.3. Infinitas soluciones.

Como ejemplo final, consideremos el sistema

$$\left.\begin{array}{c} x_1,+x_2,+x_3,=,6 \\ x_1,,-2x_3,=,4 \\ 2x_1,+x_2,-x_3,=,10 \end{array}\right\} \tag{3.1.10}$$

La matriz ampliada del sistema es

$$\begin{pmatrix} 1,1,1,6 & \\ 1,0,-2,4 & \\ 2,1 & -1,10 \end{pmatrix} \tag{3.1.11}$$

y en MATLAB queda

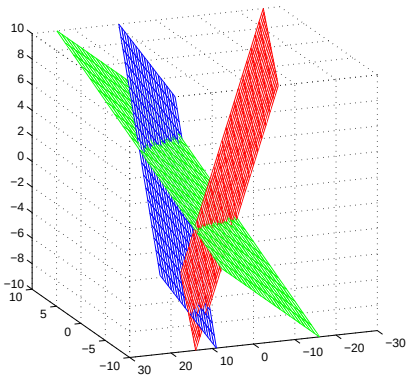


FIGURA 2. Dos planos se cortan en una recta, paralela al otro. No hay puntos comunes en la intersección.

```
>> A=[1,1,1,6;1,0,-2,4;2,1,-1,10]
```

Usamos el comando `rref`

```
>> R=rref(A)
```

y la forma escalonada por filas de la matriz 3.1.11 es

$$\begin{pmatrix} 1,0,-2,4 & \\ 0,1,3,2 & \\ 0,0 & 0,0 \end{pmatrix}. \tag{3.1.12}$$

Observemos que tenemos una fila de ceros en la parte inferior de la matriz. Además, tenemos solamente dos pivotes. Es muy importante, en este momento, identificar las variables pivotes y las variables libres. Observemos que las columnas 1 y 2 tienen pivotes. Por tanto, x_1 y x_2 son variables pivote. La columna 3 no tiene pivote. Así, la variable x_3 es libre.

Como la última fila de la matriz representa la ecuación

$$0x_1 + 0x_2 + 0x_3 = 0, \tag{3.1.13}$$

que se verifica para cualesquiera valores de x_1, x_2 y x_3, únicamente necesitamos encontrar los valores de x_1, x_2 y x_3 que satisfacen las ecuaciones representadas por las dos primeras filas de la matriz 3.1.12

$$\left.\begin{array}{l} x_1, -2x_3, = 4 \\ x_2, +3x_3, = 2 \end{array}\right\} \tag{3.1.14}$$

Ahora el método es simple y directo. Resolvemos cada ecuación para su variable pivote en función de la variable libre. Así nos queda

$$\left.\begin{array}{l} x_1, =, 4 + 2x_3 \\ x_2, =, 2 - 3x_3. \end{array}\right\} \tag{3.1.15}$$

Es habitual colocar parámetros para representar la variable libre. Por ejemplo, si hacemos $x_3 = \lambda$, el sistema 3.1.10 tiene infinitas soluciones, descritas por

$$x_1 = 4 + 2\lambda, \quad x_2 = 2 - 3\lambda, \quad x_3 = \lambda \tag{3.1.16}$$

donde λ es cualquier número real. Por cada valor que demos a λ obtenemos una solución. Por ejemplo, para $\lambda = 0$ obtenemos la solución $(4, 2, 0)$. Para $\lambda = 1$ nos queda $(6, -1, 1)$.

De nuevo, la visualización geométrica nos aclara lo anterior. Como podemos ver en la figura 3, los tres planos se cortan a lo largo de una recta. Por tanto, hay un número infinito de soluciones, que coincide con nuestra conclusión anterior.

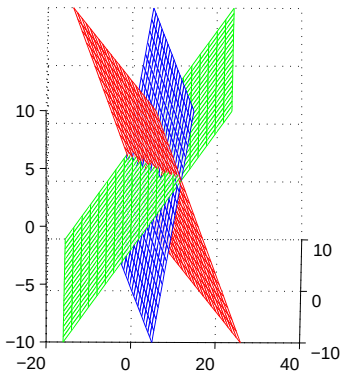


FIGURA 3. Los tres planos se cortan en una recta, que contiene un número infinito de puntos.

2. Más difícil todavía

El pánico suele crecer cuando el número de ecuaciones e incógnitas se incrementa. Por supuesto, este aumento hace las cosas un poco más difíciles, pero si seguimos una sencillas reglas estas dificultades desaparecen.

- Identifica las variables pivot. Esto se consigue observando las columnas que son pivote.
- Identifica las variables libres. Esto se obtiene observando las columnas que no tienen pivote.
- Resuelve cada ecuación colocando cada variable pivot en función de la libres.
- Cambia las variables libres por parámetros.

Por ejemplo, consideremos el siguiente sistema

$$\begin{array}{c} -4x_1, -2x_2, , +2x_4, -4x_5, +4x_6, =, 2 \\ 4x_1, +x_2, , -3x_4, +4x_5, -4x_6, =, -3 \\ x_1, -2x_2, , -3x_4, +x_5, -x_6, =, -3 \\ , -2x_2, , -2x_4, , , =, -2 \end{array} \tag{3.2.17}$$

A simple vista, el problema puede echar para atrás por su tamaño. Si seguimos las reglas anteriores, no tendremos problema para encontrar la solución. En primer lugar, consideremos la matriz ampliada,

$$\begin{pmatrix} -4, -2, 0, 2, -4, 4, 2 \\ 4, 1, 0, -3, 4, -4, -3 \\ 1, -2, 0, -3, 1, -1, -3 \\ 0, -2, 0, -2, 0, 0, -2 \end{pmatrix} \tag{3.2.18}$$

y la introducimos en MATLAB.

```
>> A=[-4,-2,0,2,-4,4,2;4,1,0,-3,4,-4,-3; ...
>> 1,-2,0,-3,1,-1,-3;0,-2,0,-2,0,0,-2]
```

Calculamos la forma escalonada por filas con `rref`.

```
>> R=rref(A)
```

Las columnas uno y dos tienen pivotes. Por tanto, x_1 y x_2 son variables pivote. Las restantes incógnitas, x_3, x_4, x_5 y x_6 son variables libres.

Las últimas filas de ceros se pueden ignorar, porque estas ecuaciones las verifican todos los valores. Así, solamente debemos resolver el sistema

$$\left.\begin{array}{c} x_1, -x_4, +x_5, -x_6, =, -1 \\ , , x_2, +x_4, =, 1 \end{array}\right\} \tag{3.2.19}$$

Resolvemos cada ecuación para su variable pivote.

$$\left.\begin{array}{c} x_1, =, -1, +x_4, -x_5, +x_6 \\ x_2, =, 1, -x_4 \end{array}\right\} \tag{3.2.20}$$

Pongamos las variables libres como parámetros. Por ejemplo, $x_3 = \alpha, x_4 = \beta, x_5 = \gamma, x_6 = \delta$ y nos queda

$$\begin{array}{c} x_1, =, -1 + \beta - \gamma + \delta, \\ x_2, =, 1 - \beta, \\ x_3, =, \alpha, \\ x_4, =, \beta, \\ x_5, =, \gamma, \\ x_6, =, \delta, \end{array} \tag{3.2.21}$$

donde $\alpha, \beta, \gamma, \delta$ son números reales arbitrarios. Entonces el sistema 3.2.17 tiene infinitas soluciones, y las podemos obtener dando valores a los parámetros de 3.2.21.

Como podemos ver, cuando el número de incógnitas y ecuaciones crece, el problema se vuelve más difícil. No obstante, también observamos que con estas simples reglas, el tamaño no debe ser un problema.

3. Matriz inversa y forma escalonada por filas

Sea $A = (a_{ij}) \in \mathcal{M}_n(\Bbbk)$ una matriz invertible. Por ejemplo,

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

```
>> A = [1, -1, 0; 2, 0, -3; 0, 2, 1]
```

La orden `inv` de `MATLAB` calcula la matriz inversa de A.

```
>> B = inv(A)
>> A*B
```

Veamos otra forma de calcular la inversa de A usando forma escalonada por filas. Para ello basta tener en cuenta que, por definición, la matriz inversa de A es la única matriz $X = (x_{ij}) \in \mathcal{M}_n(\Bbbk)$ tal que

$$AX = I_n;$$

por lo que la columna j-ésima de X es la (única) solución del sistema

$$A(x_{1j}, \ldots, x_{nj})^{\mathrm{t}} = (0, \ldots, 0, \overset{j)}{1}, 0, \ldots, 0)^{\mathrm{t}}.$$

Por consiguiente, si partimos de la matriz $(A|I_n) \in \mathcal{M}_{n\times 2n}(\Bbbk)$ y calculamos su forma escalonada por filas llegaremos a la matriz $(I_n|A^{-1})$.

```
>> I = eye(3)
>> AI = [A,I]
>> rAI = rref(AI)
>> P = rAI(1:3,4:6)
>> A*P
```

De hecho, los programas de ordenador usan este método (o variantes del mismo) para calcular la matriz inversa, y no la fórmula por todos conocida que tiene un coste de tiempo prohibitivo.

4. Cálculo de matrices de paso

Sea $A = (a_{ij}) \in \mathcal{M}_{m\times n}(\Bbbk)$ una matriz invertible. Por ejemplo,

$$A = \begin{pmatrix} 0, 0, 1, 1 \\ -2, -1, 2, -1 \\ 2, 1, 4, 2 \\ 4, 2, 3, 0 \end{pmatrix}$$

```
>> format rat
>> A = [ 0, 0, 1, 1; -2, -1, 2, -1; 2, 1, 4, 2; 4, 2, 3, 0]
```

Veamos como podemos usar el commando `rref` para calcular, no sólo la forma escalonada R de A, sino además una matrices invertibles $P \in \mathcal{M}_n(\Bbbk)$ y $Q \in \mathcal{M}_m(\Bbbk)$ tales que

$$Q^{-1}AP = R.$$

La clave de nuestra construcción consistirá en tener en cuenta que la forma escalonada de A es la forma escalonada por columnas de la forma escalonada por filas de A.

Pero, ¿cómo se calcula la forma escalonada por columnas con MATLAB? La respuesta es bien sencilla, basta calcular la traspuesta de la forma escalonada por filas de la traspuesta de A.

```
>> C = rref(A')'
```

Ya sabemos calcular la forma escalonada por columnas; sin embargo, seguimos sin conocer cómo se calculan las matrices de paso. Para calcular una matriz invertible $Q \in \mathcal{M}_m(\Bbbk)$ tal que $F = Q^{-1}A$ es la forma escalonada por las filas de A, es suficiente observar que la forma escalonada por filas de $(A|I_m)$ es $(F|Q^{-1})$ (que es lo que sucedía antes cuando calculábamos la inversa).

```
>> F = rref(A)
>> AI = [A,eye(4)]
>> FAI = rref(AI)
>> Q1 = FAI(:,5:8)
>> Q = inv(Q1)
```

La explicación es bien sencilla, como el comando `rref` no permuta columnas, las sucesivas operaciones elementales por filas que se hacen en A para obtener su forma escalonada por filas quedan recogidas en la matriz identidad de la derecha. De forma más precisa

$$Q^{-1}(A|I_m) = (Q^{-1}A|Q^{-1}I_m) = (Q^{-1}A|Q^{-1}) = (F|Q^{-1}).$$

Ahora, para calcular matriz invertible $P \in \mathcal{M}_n(\Bbbk)$, tal que AP es la forma escalonada por columnas C de A, repetimos el proceso anterior con la traspuesta de A; y trasponemos el resultado obtenido.

```
>> B = A'
>> BI = [B,eye(4)]
>> FBI = rref(BI)
>> P1 = FBI(:,5:8)
>> P = P1'
```

Una vez que sabemos calcular matrices de paso para la forma escalonada por filas y para la forma escalonada por columnas de A, veamos cómo se calculan unas matrices de paso $P \in \mathcal{M}_n(\Bbbk)$ y $Q \in \mathcal{M}_m(\Bbbk)$ tales que $Q^{-1}AP$ es la forma escalonada de A. Para ello, basta calcular la forma escalonada por columnas de la forma escalonada por filas de A y unas matrices de paso.

En nuestro caso, ya teníamos calculada la forma escalonada por filas F de A y la matriz de paso Q, luego sólo nos queda calcular la forma escalonada por columnas de F y una matriz de paso.

```
>> E = F'
>> EI = [E,eye(4)]
>> FEI = rref(EI)
>> P1 = FEI(:,5:8)
>> P = P1'
```

Obsérvese que MATLAB ha escrito * en vez de algunas entradas de las matrices que hemos ido obteniendo, esto ocurre cuando usamos el formato racional y el tamaño de la entrada tiene demasiada longitud; por ejemplo, cuando se trata de una fracción con un denominador muy grande, como es nuestro caso. En nuestro ejemplo, estos asteriscos deben ser tratados como ceros; aunque en realidad lo que ponen de manifiesto es la propagación de errores de redondeo en nuestras operaciones.

Ejercicios de la práctica 3

Ejercicio 1. Consideremos la siguiente matriz

$$A = \begin{pmatrix} -4 & -2 & -4 & 0 \\ -2 & -10 & -22 & 4 \\ -5 & 2 & 5 & -2 \\ -24 & 6 & 16 & -8 \end{pmatrix}.$$

Si R es la forma escalonada por filas de A, calcular, usando `MATLAB`, las matrices Q y P tales que $Q^{-1}AP = R$.

Calcular la forma escalonada por columnas de A, la forma reducida de A y las matrices de paso cada caso.

Ejercicio 2. El comando `null` de `MATLAB`, calcula una base del núcleo de A, $\ker(A)$. Usando este comando, calcula la solución general del sistema $A\mathbf{x} = \mathbf{b}$, con

$$A = \begin{pmatrix} 1 & 1 & -1 & 0 & 2 \\ 2 & 1 & 1 & 1 & 1 \end{pmatrix} \quad \text{y} \quad \mathbf{b} = \begin{pmatrix} 3 \\ 1 \end{pmatrix}.$$

Ejercicio 3. Dadas la siguientes matrices

$$A = \begin{pmatrix} 1 & 2 & -1 & 3 \\ 2 & 4 & -2 & 6 \\ 3 & 6 & -3 & 9 \\ 1 & 3 & 1 & 2 \end{pmatrix} \quad \text{y} \quad B = \begin{pmatrix} 8 & 2 & 0 & 9 \\ 16 & 4 & 0 & 18 \\ 24 & 6 & 0 & 27 \\ 9 & -3 & 4 & 14 \end{pmatrix},$$

estudiar si existe una matrix invertible $P \in \mathcal{M}_4(\mathbb{R})$ tal que $AP = B$.

Dar una condición necesaria y suficiente para que fijadas dos matrices A y $B \in \mathcal{M}_{m\times n}(\mathbb{R})$ exista una matriz invertible $P \in \mathcal{M}_n(\mathbb{R})$ tal que $AP = B$.

Ejercicio 4. Considerar el sistema de ecuaciones $AXB = C$, donde X es una matriz de orden 3 de incógnitas y

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

Hallar, si es posible, la solución general de este sistema.

Ejercicio 5.

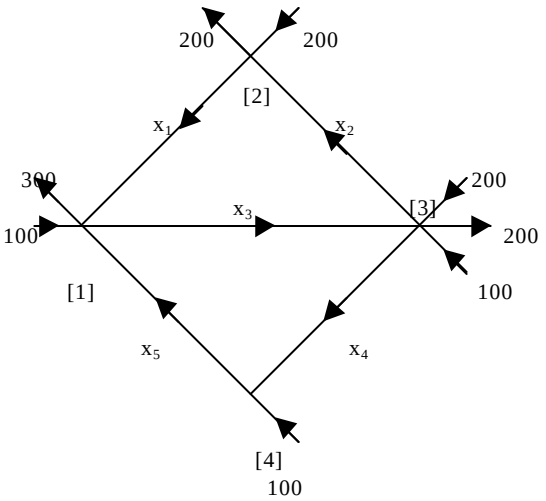
1. Hallar las inversas de las siguientes matrices utilizando el método de Gauss-Jordan con ayuda del comando `rref` de `MATLAB`.

$$A = \begin{pmatrix} 4 & -6 & -9 \\ -2 & -1 & 1 \\ -1 & 1 & 2 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & -5 & -11 \\ -1 & -2 & -18 \\ 1 & -1 & 6 \end{pmatrix}$$

$$C = \begin{pmatrix} 0 & -1 & -5 & 1 \\ -1 & -1 & 5 & -5 \\ 1 & 1 & -4 & 4 \\ -1 & -3 & -5 & -1 \end{pmatrix}$$

2. Usar la función `inv` de `MATLAB` para comprobar dichos resultados.

Ejercicio 6. Flujos de Tráfico. Considerar el siguiente diagrama de una malla de calles de un sentido con vehículos que entran y salen de las intersecciones. La intersección k se denota $[k]$. Las flechas a lo largo de las calles indican la dirección del flujo de tráfico. Sea x_i el número de vehículos por hora que circulan por la calle i. Suponiendo que el tráfico que entra a una intersección también sale, establecer un sistema de ecuaciones que describa el diagrama del flujo de tráfico. Por ejemplo, en la intersección [1] $x_1 + x_5 + 100 =$ tráfico que entra $=$ tráfico que sale $= x_3 + 300$, lo que da $x_1 - x_3 + x_5 = 200$. Estudiar la compatibilidad de dicho sistema y resolverlo usando la función `rref` de `MATLAB`.



Ejercicio 7. Considerar el sistema de ecuaciones lineales

$$\left.\begin{array}{rcr} x - 2y + 3z & = & 1 \\ 4x + y - 2z & = & -1 \\ 2x - y + 4z & = & 2 \end{array}\right\}$$

1. Definir la matriz A del sistema y la matriz b de términos independientes, a las que llamaremos `A` y `b`, respectivamente, y la matriz ampliada del sistema, a que llamaremos `Ab`.
2. Estudiar la compatibilidad del sistema usando la función `rref`.
3. Escribir `A\b` en la línea de comandos de `MATLAB`, y explicar el resultado.

Considerar ahora el sistema ecuaciones lineales, y repetir los apartados anteriores.

$$\left.\begin{array}{rcr} x - 2y & = & 1 \\ 4x + y & = & -1 \\ 5x - y & = & 1 \end{array}\right\}$$

Ejercicio 8. El sistema de ecuaciones

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

tiene como matriz de coeficientes y vector de términos independientes a

$$A = \begin{pmatrix} 1 & 2 & -3 \\ 2 & 0 & -3 \\ 0 & 1 & 1 \end{pmatrix}, \mathbf{b} = \begin{pmatrix} 4 \\ -2 \\ 0 \end{pmatrix}$$

respectivamente. Construye la matriz ampliada `M=[A,b]` y obtenga su forma reducida por filas con el comando `rref(M)`.

Ejercicio 9.

Cada una de las siguientes matrices representa la matriz ampliada de un sistema lineal. Realiza las siguientes tareas para cada caso.

- Define una matriz cuadrada de orden 9 con coeficientes enteros entre -4 y 4. Realiza las siguientes tareas para cada caso.
  - Introduce la matriz en `MATLAB` y con el comando `rref` calcula la forma escalonada por filas. Cópiala en un papel.
  - Identifica variables pivote y variables libres.
  - Resuelve cada ecuación para su variable pivote.
  - Asigna parámetros a las variables libres.

(a)

$$\begin{pmatrix} 3 & -1 & 0 & -1 & -3 & -1 & -2 & -3 \\ -2 & 0 & 0 & 0 & 2 & 0 & 2 & 2 \\ 3 & 0 & 0 & -1 & -1 & -2 & -1 & -1 \\ 0 & 0 & 0 & 1 & -2 & 2 & -2 & -2 \\ 3 & 1 & 0 & 0 & -1 & -1 & -2 & -1 \\ 1 & -4 & 0 & -2 & -5 & 0 & -1 & -5 \end{pmatrix}.$$

(b)

$$\begin{pmatrix} -2 & -2 & 2 & -1 & 1 & -2 & -1 & -1 & 0 \\ -1 & -2 & 2 & 1 & 3 & -1 & -2 & -1 & 0 \\ 0 & 0 & 1 & 0 & 3 & -2 & -1 & -1 & 0 \\ 1 & 0 & 0 & 2 & 2 & 1 & -1 & 0 & 0 \\ -2 & 1 & 0 & -1 & -2 & -1 & 0 & -1 & -2 \\ 0 & 1 & -2 & -1 & -4 & 1 & 2 & 1 & 0 \\ 0 & 1 & 2 & 1 & 2 & 1 & -2 & -1 & -2 \\ -2 & -1 & 0 & 1 & 0 & -1 & -1 & -1 & -1 \end{pmatrix}.$$

Ejercicio 10. Juan tiene 4 euros en monedas de $1, 2, 5$ y 10 céntimos de euro. Tiene igual número de monedas de 2 céntimos y de 5 céntimos, y en total tiene 100 monedas. ¿De cuántas formas es esto posible?

Ejercicio 11.

- Define una matriz cuadrada de orden 9 con coeficientes enteros entre -4 y 4.
- Con el comando `rref` calcula la forma escalonada por filas.
- Identifica variables pivote y variables libres.
- Resuelve cada ecuación para su variable pivote.
- Asigna parámetros a las variables libres.

Ejercicio 12. Usar el método de Gauss para resolver simultáneamente los sistemas

$$\begin{array}{rcl|c|c} 4x-8y+5z & = & 1 & 0 & 0 \\ 4x-7y+4z & = & 0 & 1 & 0 \\ 3x-4y+2z & = & 0 & 0 & 1 \end{array}$$

Ejercicio 13. Supongamos que 100 insectos se distribuyen en una cámara que consta de 4 habitaciones con pasajes entre ellos tal como aparece en la figura (4). Al final de un minuto, los insectos se han redistribuido. Supongamos que un minuto no es bastante tiempo para que un insecto visite más de una habitación y al final de un minuto el 40 % de los insectos de cada habitación permanece en ella. Los insectos que la abandonan se distribuyen uniformemente entre las demás habitaciones que son accesibles desde la que ocupan inicialmente. Por ejemplo, desde la habitación 3, la mitad de los que se mueven van a 2 y la otra mitad a 4.

1. Si al final de un minuto hay $12, 25, 26$ y 37 insectos en las habitaciones $1, 2, 3$ y 4, respectivamente, determinar la distribución inicial.
2. Si la distribución inicial es $20, 20, 20$ y 40 ¿Cuál es la distribución al final de un minuto?

Ejercicio 14. En la figura (5) aparece una placa de acero. La temperatura en cada punto de la placa es constante (no cambia con el tiempo). La temperatura en cada punto del retículo en el borde de la placa aparece en la figura. Sea t_i la temperatura en grados en cada punto del retículo en el interior de la placa. Supongamos que la temperatura en cada punto interior del retículo es la media de las temperaturas de sus cuatro puntos vecinos. Calcula la temperatura t_i en cada punto interior del retículo.

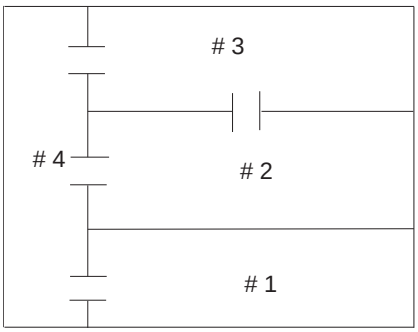


FIGURA 4. Distribución de las cámaras y los pasajes.

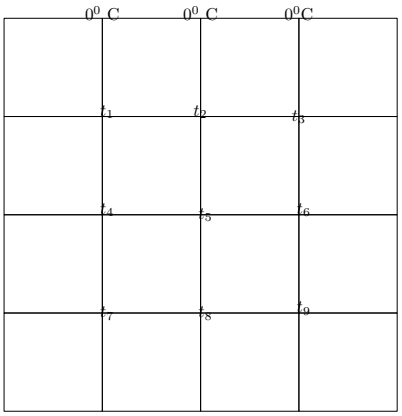


FIGURA 5. Distribución de temperatura en una placa de metal.

Ejercicio 15. Consideremos la siguiente matriz

$$A = \begin{pmatrix} -4 & -2 & -4 & 0 \\ -2 & -10 & -22 & 4 \\ -5 & 2 & 5 & -2 \\ -24 & 6 & 16 & -8 \end{pmatrix}.$$

Si R es la forma escalonada por filas de A, calcular, usando `MATLAB`, las matrices Q y P tales que $Q^{-1}AP = R$.

Calcular la forma escalonada por columnas de A, la forma reducida de A y las matrices de paso cada caso.

PRÁCTICA 4

Comportamiento asintótico de sistemas dinámicos

La forma cerrada de la solución de un sistema de ecuaciones en diferencias se puede usar para determinar el comportamiento a largo plazo o asintótico de un sistema dinámico. El concepto de autovalor dominante aparece entonces.

Pre-requisitos: conocimiento de autovalores y autovectores. Forma canónica de Jordan. Ecuaciones en diferencias homogéneas finitas con coeficientes constantes (caso diagonalizable).

1. Comportamiento de la sucesión λ^n

Para una comprensión de lo que viene después, necesitamos estudiar en primer lugar el comportamiento asintótico de la sucesión $(\lambda^n)_{n\in\mathbb{N}}$, con $\lambda \in \mathbb{C}$. Hay que distinguir varios casos.

1.1. Cuando λ es un número real.

Vamos a realizar varios experimentos cuando λ es un número real. Por ejemplo, estudiemos el límite de la sucesión $(0{,}5^n)_{n\in\mathbb{N}}$ cuando $n \to \infty$. El siguiente código en `MATLAB` genera los 15 primeros términos de la sucesión.

```
>> n=(1:15)';
>> (0.5).^n
```

Este resultado nos indica que $\lim_{n\to\infty}(0{,}5)^n = 0$. De forma análoga, se puede estimar el límite de la sucesión $((-0{,}75)^n)_{n\in\mathbb{N}}$.

```
>> n=(1:30)';
>> (-0.75).^n
```

Observemos que la sucesión definida por $((-0{,}75)^n)_{n\in\mathbb{N}}$ oscila entre valores positivos y negativos. Vemos también que converge a cero, aunque la velocidad es menor que la sucesión definida por $(0{,}5^n)_{n\in\mathbb{N}}$.

Conjetura. *Si λ es un número real con* $\text{abs}(\lambda) < 1$, *entonces* $\lim_{n\to\infty} \lambda^n = 0$.

Experimento. En MATLAB, verificar que las siguientes sucesiones converge a cero cuando $n \to \infty$.

- $(0{,}25)^n$.
- $(-0{,}8)^n$.
- $(0{,}99)^n$.

Conjetura. *Si λ es un número real tal que* $\text{abs}(\lambda) > 1$, *entonces los términos de la sucesión $\{\lambda^n\}$ se hacen tan grandes como queramos en valor absoluto.*

Experimento. En MATLAB, verificar que las siguientes sucesiones producen términos de valor absoluto tan grande como queramos cuando $n \to \infty$.

- $2{,}3^n$.
- $(-1{,}4)^n$.
- $(1{,}05)^n$.

1.2. Cuando λ es un número complejo.

Si $\lambda = a + b\text{i}$ entonces su norma es $|\lambda| = \sqrt{a^2 + b^2}$. Por ejemplo, si $\lambda = 0{,}3 + 0{,}4\text{i}$ entonces la norma de λ es $|\lambda| = \sqrt{0{,}3^2 + 0{,}4^2} \approx 0{,}5$. Observemos que en este caso $|\lambda| < 1$. Con MATLAB podemos calcular fácilmente la norma de un número complejo con los comandos `norm` o `abs`

```
>> norm(0.3+0.4i)
```

Y las siguientes instrucciones en MATLAB generan los 15 primeros términos de la sucesión definida por $((0{,}3 + 0{,}4\text{i})^n)_{n\in\mathbb{N}}$.

```
>> n=(1:15)';
>> (0.3+0.4i).^n
```

La siguiente figura (obtenida con el comando `plot((0.3+0.4i).^n)` de MATLAB) se observa que los términos de la sucesión convergen a $0 + 0\text{i}$.

Conjetura. *Si $|\lambda| < 1$ entonces la sucesión $(\lambda^n)_{n\in\mathbb{N}}$ converge a* 0.

Experimento. Usar MATLAB para probar que el término general de las siguientes sucesiones tiene norma menor que 1, y que convergen a cero.

- $\{(0{,}25 + 0{,}45i)^n\}$.
- $\{(-0{,}5 - 0{,}2i)^n\}$.

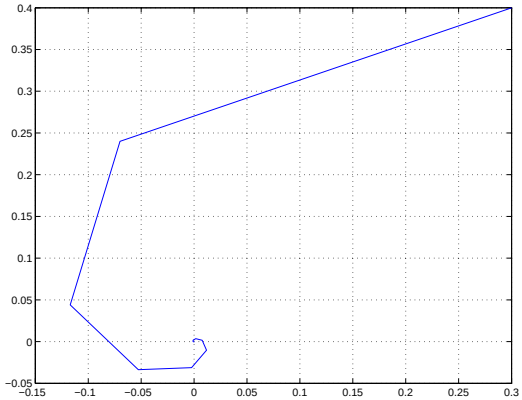


FIGURA 1. Convergencia a 0 de la sucesión $((0{,}3+0{,}4\mathrm{i})^n)_{n\in\mathbb{N}}$.

Conjetura. *Si* $|\lambda|>1$ *entonces la sucesión* $(\lambda^n)_{n\in\mathbb{N}}$ *toma valores de norma tan grandes como se quiera.*

Por ejemplo, si $\lambda=0{,}8+1{,}2\mathrm{i}$ entonces $|\lambda|=\sqrt{0{,}8^2+1{,}2^2}\approx 1{,}4422$, que es mayor que uno.

```
>> norm(0.8+1.2i)
```

Con las siguientes instrucciones generamos los primeros términos de la sucesión.

```
>> n=(1:15)';
>> S=(0.8+1.2i).^n
```

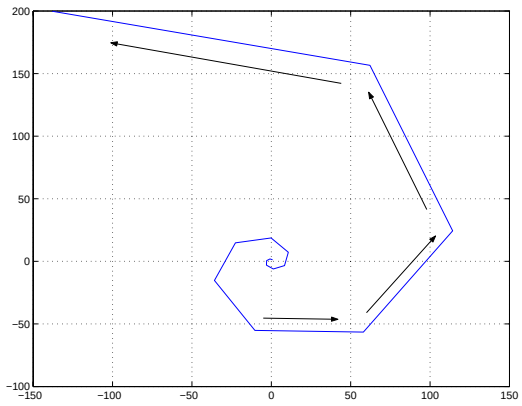


FIGURA 2. Comportamiento de la sucesión $((0{,}8+1{,}2)^n)_{n\in\mathbb{N}}$.

Podemos ver las normas de cada término de la sucesión.

```
>> abs(S)
```

Es claro que las normas de los términos de la sucesión van creciendo en tamaño.
Experimento. Usar MATLAB para probar que el término general de las siguientes sucesiones tiene norma mayor que 1, y la sucesión $(\lambda^n)_{n\in\mathbb{N}}$ alcanza valores de norma cada vez mayor.

- $((1{,}25 + 0{,}8\mathrm{i})^n)_{n\in\mathbb{N}}$.
- $((-1{,}4 - 0{,}8\mathrm{i})^n)_{n\in\mathbb{N}}$.

2. Sistemas de ecuaciones en diferencias: comportamiento asintótico

Consideremos el sistema de ecuación en diferencias con condición inicial definida por

$$\left\{ \begin{array}{rcl} x_{n1} & = & 1{,}0\, x_{n-1\,1} + 0{,}2\, x_{n-1\,2} \\ x_{n\,2} & = & 0{,}2\, x_{n-1\,1} + 1{,}0\, x_{n-1\,2} \end{array} \right. \tag{4.2.1}$$

con $x_{01} = 0$ y $x_{02} = 1$. En notación matricial

$$\mathbf{x}_n = \begin{pmatrix} 1{,}0 & 0{,}2 \\ 0{,}2 & 1{,}0 \end{pmatrix} \mathbf{x}_{n-1}, \quad \mathbf{x}_0 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}. \tag{4.2.2}$$

siendo $\mathbf{x}_n = (x_{n1}, x_{n2})^{\mathrm{t}},\ n \geq 0$.

Los autovalores y autovectores asociados de la matriz

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

son

$$\lambda_1 = 1{,}2 \quad \text{y} \quad \mathbf{v}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix},$$
$$\lambda_2 = 0{,}8 \quad \text{y} \quad \mathbf{v}_2 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

En efecto,

```
>> A = [1.0, 0.2; 0.2, 1.0]
>> lambda = eig(A)
>> v1 = null(lambda(1)*eye(2)-A,'r')
>> v2 = null(lambda(2)*eye(2)-A,'r')
```

Por consiguiente, si $P = (\mathbf{v}_1, \mathbf{v}_2) \in \mathcal{M}_2(\mathbb{R})$, entonces $P^{-1}AP = D = \text{diag}(\lambda_1, \lambda_2)$.

Como en la práctica sobre ecuaciones en diferencias, si la condición inicial se puede escribir como combinación lineal de los autovectores, es decir, $\mathbf{x}_0 = c_1\mathbf{v}_1 + c_2\mathbf{v}_2$, entonces la forma cerrada de la solución ecuación (4.2.1) es

$$\mathbf{x}_n = c_1\lambda_1^n\mathbf{v}_1 + c_2\lambda_2^n\mathbf{v}_2. \tag{4.2.3}$$

En efecto, si $\mathbf{c} = P^{-1}\mathbf{x}_0$, entonces

$$\mathbf{x}_n = A\mathbf{x}_{n-1} = \ldots = A^n\mathbf{x}_0 = P\begin{pmatrix} \lambda_1^n & 0 \\ 0 & \lambda_2^n \end{pmatrix} P^{-1}\mathbf{x}_0 = P\begin{pmatrix} \lambda_1^n & 0 \\ 0 & \lambda_2^n \end{pmatrix}\mathbf{c}.$$

Nota.- Como, en nuestro caso, $|\lambda_1| > |\lambda_2|$, decimos que λ_1 es el **autovalor dominante de** A.

Ahora dividimos ambos lados de la ecuación (4.2.3) por λ_1^n. Nos queda entonces

$$\frac{1}{\lambda_1^n}\mathbf{x}_n = c_1\mathbf{v}_1 + c_2\left(\frac{\lambda_2}{\lambda_1}\right)^n\mathbf{v}_2 \tag{4.2.4}$$

Tomemos límite cuando $n \to \infty$ en la expresión anterior.

$$\begin{aligned} \lim_{n\to\infty}\frac{1}{\lambda_1^n}\mathbf{x}_n &= \lim_{n\to\infty}\left(c_1\mathbf{v}_1 + c_2\left(\frac{\lambda_2}{\lambda_1}\right)^n\mathbf{v}_2\right) \\ &= c_1\mathbf{v}_1 + c_2\lim_{n\to\infty}\left(\frac{\lambda_2}{\lambda_1}\right)^n\mathbf{v}_2 \end{aligned} \tag{4.2.5}$$

Pero como $|\lambda_1| > |\lambda_2|$ sabemos que $|\lambda_2/\lambda_1| < 1$ y en consecuencia

$$\lim n \to \infty\left(\frac{\lambda_2}{\lambda_1}\right)^n = 0 \qquad \text{y} \qquad \lim_{n\to\infty}\frac{1}{\lambda_1^n}\mathbf{x}_n = c_1\mathbf{v}_1.$$

Entonces, para valores grandes de n se tiene que

$$\begin{aligned} \frac{1}{\lambda_1^n}\mathbf{x}_n &\approx c_1\mathbf{v}_1 \\ \mathbf{x}_n &\approx c_1\lambda_1^n\mathbf{v}_1. \end{aligned} \tag{4.2.6}$$

Como c_1 y λ_1^n son escalares, la ecuación (4.2.6) indica que el vector $\mathbf{x}_n$ es, aproximadamente, un múltiplo de $\mathbf{v}_1$. Así, cuando iteramos la ecuación (4.2.1), el vector $\mathbf{x}_n$ se va colocando de forma paralela al autovalor $\mathbf{v}_1$.

2.1. Dibujo de trayectorias.

Vamos a usar el m-fichero `tray.m` (cuyo código se incluye al final de esta sección), que nos ayudará a dibujar soluciones de la ecuación (4.2.1). Ejecutamos el programa tecleando `tray` en la pantalla de comandos de `MATLAB`. Introducimos entonces la matriz de la ecuación 4.2.1 cuando nos la pidan.

```
>> tray
```

El programa responde creando una figura con ejes. Coloca el puntero del ratón, aproximadamente, en el punto $(1, 0)$, que va a ser la condición inicial $\mathbf{x}_0 = (1, 0)^{\mathrm{t}}$, y haga 'click' con el botón derecho. Se dibuja la trayectoria solución, primero hacia adelante en el tiempo desde la condición inicial $\mathbf{x}_0$ y luego hacia atrás en el tiempo. Observa que esta solución, tal como aparece en la figura, se acerca de forma paralela al autovalor $\mathbf{v}_1$. Crea ahora más trayectorias de la ecuación (4.2.1) pulsando condiciones iniciales $\mathbf{x}_0$ con el ratón. Note que las trayectorias futuras se acercan a un vector paralelo a $\mathbf{v}_1$.

Fichero `tray.m`

```
function tray(action)
global AxHndl FigNum AA
if nargin<1
  action='initialize';
end
if strcmp(action,'initialize')
  home
  AA= input('Introduzca una matriz 2x2 en la forma [a,b;c,d] --> ');
  FigNum=figure(gcf);
  clf
  set(FigNum,...
    'units','normalized',...
    'position',[.1 .1 .8 .8],...
    'Name','Sistemas Din´amicos',...
    'NumberTitle','off',...
    'WindowButtonDownFcn','tray(''gotraj'')');
  AxHndl=axes(...
    'xlim',[-10 10],...
    'ylim',[-10,10],...
    'xtick',-10:10,...
```

```
  'ytick',-10:10,...
  'units','normalized',...
  'position',[.1 .1 .7 .8]);
xax=line([-10 10],[0 0],'color','black');
yax=line([0 0],[-10 10],'color','black');
grid
axhndl2=axes(...
  'units','normalized',...
  'position',[.85,.7,.1,.2],...
  'visible','off',...
  'xlim',[-1 1],...
  'ylim',[0 1]);
y=[0 .1 .2 .4 .8];
x=zeros(size(y));
line(x,y,...
  'linestyle','-',...
  'marker','o',...
  'color','b');
%line(x,y,...
  %'linestyle','-',...
  %'color','b');
textfwd=uicontrol(...
  'style','text',...
  'units','normalized',...
  'position',[.85 .6 .1 .05],...
  'string','futuro',...
  'ForegroundColor','b');
axhndl3=axes(...
  'units','normalized',...
  'position',[.85,.3,.1,.2],...
  'visible','off',...
  'xlim',[-1 1],...
  'ylim',[0 1]);
y=[0 .1 .2 .4 .8];
x=zeros(size(y));
line(x,y,...
  'linestyle','-',...
  'marker','x',...
  'color','r');
```

```
  %line(x,y,...
    %'linestyle','-',...
    %'color','r');
  textbwd=uicontrol(...
    'style','text',...
    'units','normalized',...
    'position',[.85 .2 .1 .05],...
    'string','pasado',...
    'ForegroundColor','r');
  qbut=uicontrol(...
    'style','pushbutton',...
    'string','Salida',...
    'units','normalized',...
    'position',[.85 .05 .1 .05],...
    'callback','tray(''quit'')');
  figure(FigNum);
  axes(AxHndl)
elseif strcmp(action,'gotraj')
  N=20;
  points=zeros(2,N);
  figure(FigNum);
  axes(AxHndl);
  p=get(gca,'CurrentPoint');
  x=p(1,1);y=p(1,2);
  points(:,1)=[x,y]';
  for k=2:N
    points(:,k)=AA*points(:,k-1);
  end
  fwdpt=line(points(1,:),points(2,:),...
    'linestyle','o',...
    'color','b',...
    'erasemode','background',...
    'clipping','on');
  fwdseg=line(points(1,:),points(2,:),...
    'linestyle','-',...
    'color','b',...
    'erasemode','background',...
    'clipping','on');
  for k=2:N
```

```
    points(:,k)=inv(AA)*points(:,k-1);
  end
  bwdpt=line(points(1,:),points(2,:),...
    'linestyle','x',...
    'color','r',...
    'erasemode','background',...
    'clipping','on');
  bwdseg=line(points(1,:),points(2,:),...
    'linestyle','-',...
    'color','r',...
    'erasemode','background',...
    'clipping','on');
elseif strcmp(action,'quit')
  close(FigNum)
end
```

Ejercicios de la práctica 4

Ejercicio 1. Para cada una de las siguientes ecuaciones en diferencia (sistemas dinámicos) realizar las siguientes tareas:

- Usar el comando `eig` para calcular los autovalores y autovectores de la matriz asociada.
- Escribir en forma cerrada
$$\mathbf{x}_{n+2} = c_1\lambda_1^n\mathbf{v}_1 + c_2\lambda_2^n\mathbf{v}_2$$
la solución de la ecuación.
- Dividir ambos lados de la solución $\mathbf{x}_{n+2} = c_1\lambda_1^n\mathbf{v}_1 + c_2\lambda_2^n\mathbf{v}_2$ por la n-ésima potencia del autovalor dominante y tome el límite cuando $n \to \infty$. Usar el resultado para aproximar $\mathbf{x}_n$ para valores grandes de n y prediga el comportamiento de la solución.
- Ejecutar el m-fichero `tray.m` y verificar que las trayectorias de la solución se comportan como se indicó en el apartado anterior.

$$\mathbf{x}_n = \begin{pmatrix} 0{,}6 & 0{,}2 \\ 0{,}0 & 0{,}8 \end{pmatrix}\mathbf{x}_{n-1}, \quad \mathbf{x}_0 = \begin{pmatrix} 5 \\ 3 \end{pmatrix}.$$

$$\mathbf{x}_n = \begin{pmatrix} 1{,}42 & 0{,}16 \\ 0{,}16 & 1{,}18 \end{pmatrix}\mathbf{x}_{n-1}, \quad \mathbf{x}_0 = \begin{pmatrix} 1 \\ 4 \end{pmatrix}.$$

PRÁCTICA 5

Ecuaciones en diferencias

En esta práctica ilustraremos con algunos sencillos ejemplos como se puede calcular la forma cerrada de la solución una ecuación lineal en diferencias con coeficientes constantes con condición inicial.

Pre-requisitos: conocimiento de autovalores y autovectores. Forma canónica de Jordan.

1. Ecuaciones en diferencias de primer orden

Consideremos la siguiente expresión:

$$\begin{cases} a_{n+1} &= (6/5)a_n,\ n \geq 1 \\ a_1 &= 2 \end{cases} \tag{5.1.1}$$

Esto es una ecuación en diferencias de primer orden con condición inicial. Este tipo de expresiones son las que aparecen cuando se definen relaciones por recurrencia. La ecuación y su condición inicial dada por la ecuación (5.1.1) sirven para calcular fácilmente los términos de la sucesión:

$$\begin{aligned} a_2 &= (6/5)a_1 = (6/5)\cdot 2, \\ a_3 &= (6/5)a_2 = (6/5)^2\cdot 2, \\ a_4 &= (6/5)a_3 = (6/5)^3\cdot 2, \\ &\vdots \end{aligned} \tag{5.1.2}$$

Tal como aparece en la ecuación (5.1.2), el término $(n+1)$-ésimo de la sucesión definida en la ecuación (5.1.1) viene dado por $a_{n+1} = (6/5)^n \cdot 2$. La expresión $a_{n+1} = (6/5)^n \cdot 2$ se llama **solución forma cerrada** de la ecuación (5.1.1). Dar la solución en forma cerrada es útil para calcular directamente cualquier término de la sucesión generada por la ecuación (5.1.1). Por ejemplo, el término undécimo es:

$$a_{11} = (6/5)^{10} \cdot 2 \approx 12{,}3835.$$

En efecto,

```
>> a11=(6/5)^10*2
```

Ahora vamos a usar MATLAB para producir los primeros once términos de la sucesión generada por la ecuación en diferencias de 5.1.1. En primer lugar, declaramos un vector con ceros que usaremos para almacenar los once términos de la sucesión. En la ecuación (5.1.1), vemos que el primer valor de la sucesión es $a_1 = 2$. Colocamos este valor en la primera componente del vector **a**.

```
>> a=zeros(11,1);
>> a(1)=2
```

Según la ecuación (5.1.1), el $(n+1)$-ésimo término se obtiene multiplicando el n-ésimo por $6/5$. Esto se puede hacer en MATLAB con un bucle `for`.

```
>> for n=1:10,a(n+1)=(6/5)*a(n);end
>> a
```

2. Ecuaciones en diferencias de orden $p \geq 2$

Las soluciones de las ecuaciones en diferencias de orden $p \geq 2$ también admiten una expresión cerrada. En este caso, la clave consiste en escribir la ecuación en diferencias en forma matricial. La forma cerrada de la solución dependerá de si la correspondiente matriz asociada es diagonalizable o no.

2.1. Caso diagonalizable.

Consideremos la ecuación en diferencias de segundo orden

$$x_{n+2} = 3x_{n+1} - 2x_n,\ n \geq 1, \tag{5.2.3}$$

con las condiciones iniciales $x_1 = 1$ y $x_2 = 0$. Sabemos que esta ecuación en diferencias se puede escribir

$$\begin{pmatrix} x_{n+2} \\ x_{n+1} \end{pmatrix} = \begin{pmatrix} 3 & -2 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x_{n+1} \\ x_n \end{pmatrix}.$$

De tal forma que si denotamos

$$\mathbf{x}_n = \begin{pmatrix} x_{n+2} \\ x_{n+1} \end{pmatrix},\ n \geq 1, \quad \text{y} \quad A = \begin{pmatrix} 3 & -2 \\ 1 & 0 \end{pmatrix},$$

tenemos que nuestra ecuación en diferencias se ha transformado en el siguiente sistema de ecuaciones en diferencias

$$\mathbf{x}_n = A\mathbf{x}_{n-1},\ n \geq 1 \tag{5.2.4}$$

con la condición inicial $\mathbf{x}_0 = (0,1)^{\mathrm{t}}$. Por consiguiente el término general de la solución de nuestra ecuación en diferencias será la primera coordenada de $\mathbf{x}_n$.

La ecuación en diferencia (5.2.4) se puede usar para producir una sucesión de vectores en forma similar a como hicimos con la ecuación (5.1.1).

$$\mathbf{x}_1 = \begin{pmatrix} 3 & -2 \\ 1 & 0 \end{pmatrix} \mathbf{x}_0 = \begin{pmatrix} 3 & -2 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} -2 \\ 0 \end{pmatrix},$$

$$\mathbf{x}_2 = \begin{pmatrix} 3 & -2 \\ 1 & 0 \end{pmatrix} \mathbf{x}_1 = \begin{pmatrix} 3 & -2 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} -2 \\ 0 \end{pmatrix} = \begin{pmatrix} -6 \\ -2 \end{pmatrix},$$

$$\mathbf{x}_3 = \begin{pmatrix} 3 & -2 \\ 1 & 0 \end{pmatrix} \mathbf{x}_2 = \begin{pmatrix} 3 & -2 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} -6 \\ -2 \end{pmatrix} = \begin{pmatrix} -14 \\ -6 \end{pmatrix},$$

y así sucesivamente.

Con `MATLAB` es muy sencillo generar términos de la sucesión determinada por la ecuación (5.2.4). En primer lugar, definimos la matriz A y el vector inicial $\mathbf{x}_0$.

```
>> A=[3,-2;1,0]
>> x0=[0;1]
```

Vamos a generar una sucesión con once términos. Esta vez, cada término de la sucesión es un vector 2×1. Por tanto, reservamos espacio en una matriz X para esos once vectores, y cada uno de ellos se almacenará en una columna. La condición inicial $\mathbf{x}_0$ irá en la primera columna de X.

```
>> X=zeros(2,11);
>> X(:,1)=x0;
```

Recordemos que la notación `X(:,1)` hace referencia a "todas las filas, primera columna"de la matriz X. De forma similar al ejemplo anterior, el k-ésimo término de la sucesión se calcula multiplicando el $(k-1)$-ésimo término por la matriz A. Usamos un bucle `for`.

```
>> for n=2:11,X(:,n)=A*X(:,n-1);end
>> X
```

Es claro del cálculo anterior que

$$\mathbf{x}_{10} = \begin{pmatrix} -2046 \\ -1022 \end{pmatrix}. \tag{5.2.5}$$

A continuación vamos a calcular la forma cerrada de la solución de la ecuación en diferencias con condición inicial $\mathbf{u}_0$:

$$\begin{cases} \mathbf{x}_n &= A\mathbf{x}_{n-1},\ n \geq 1, \\ \mathbf{x}_0 &= \mathbf{u}_0 \end{cases}$$

cuando la matriz A es diagonalizable.

Por ejemplo si la matriz $A \in \mathcal{M}_2(\mathbb{R})$ y tiene dos autovalores λ_1, λ_2 distintos. Supongamos que $\mathbf{v}_1$ y $\mathbf{v}_2$ son autovectores de A asociados a λ_1 y λ_2 respectivamente. Como A es diagonalizable, la condición inicial $\mathbf{u}_0$ se puede escribir como combinación lineal de $\mathbf{v}_1$ y $\mathbf{v}_2$.

$$\mathbf{u}_0 = c_1\mathbf{v}_1 + c_2\mathbf{v}_2.$$

Podemos calcular $\mathbf{x}_1$ como sigue:

$$\begin{aligned} \mathbf{x}_1 &= A\mathbf{x}_0 = A\mathbf{u}_0 \\ &= A(c_1\mathbf{v}_1 + c_2\mathbf{v}_2) \\ &= c_1A\mathbf{v}_1 + c_2A\mathbf{v}_2 \\ &= c_1\lambda_1\mathbf{v}_1 + c_2\lambda_2\mathbf{v}_2 \end{aligned}$$

Para $\mathbf{x}_2$ podemos hacer algo análogo.

$$\begin{aligned} \mathbf{x}_2 &= A\mathbf{x}_1 \\ &= A(c_1\lambda_1\mathbf{v}_1 + c_2\lambda_2\mathbf{v}_2) \\ &= c_1\lambda_1A\mathbf{v}_1 + c_2\lambda_2A\mathbf{v}_2 \\ &= c_1\lambda_1^2\mathbf{v}_1 + c_2\lambda_2^2\mathbf{v}_2 \end{aligned}$$

Así, si continuamos de esta forma es claro que una forma cerrada de la ecuación (5.2.4) está dada por

$$\begin{cases} \mathbf{x}_n &= c_1\lambda_1^n\mathbf{v}_1 + c_2\lambda_2^n\mathbf{v}_2,\ n \geq 1 \\ \mathbf{x}_0 &= c_1\mathbf{v}_1 + c_2\mathbf{v}_2 \end{cases} \tag{5.2.6}$$

Y por lo tanto, el término general la solución de la ecuación en diferencias (5.2.3) es

$$x_{n+2} = c_1\lambda_1^n v_{11} + c_2\lambda_2^n v_{21},$$

donde v_{11} y v_{21} son las primeras coordenadas de los vectores $\mathbf{v}_1$ y $\mathbf{v}_2$, respectivamente.

Usando los datos de nuestro ejemplo, vamos a usar la ecuación (5.2.6) para encontrar la forma cerrada de la ecuación (5.2.4). Recordemos que la forma matricial de nuestra ecuación en diferencias es

$$\begin{cases} \mathbf{x}_n &= \begin{pmatrix} 3 & -2 \\ 1 & 0 \end{pmatrix}\mathbf{x}_{n-1},\ n \geq 1, \\ \mathbf{x}_0 &= (0,1)^{\mathrm{t}} \end{cases}$$

Para calcular su forma cerrada, realizamos el siguiente procedimiento:

1. Calcular los autovalores y autovectores de la matriz A y comprobar si A es diagonalizable.
2. Expresar la condición inicial $\mathbf{x}_0$ como combinación lineal de los autovectores.
3. Usar la ecuación (5.2.6) para escribir la forma cerrada y verificar los resultados.

El polinomio característico de la matriz A es $\aleph_A(x) = x^2 - 3x + 2$. Los autovalores, raíces del polinomio $\aleph_A(x)$, son $\lambda_1 = 2$ y $\lambda_2 = 1$. El siguiente comando calcula el polinomio característico de A.

```
>> p=poly(A)
```

Observemos que los coeficientes están escritos en orden decreciente de grado. Así, `[1 -3 2]` representa al polinomio $p(x) = x^2 - 3x + 2$. El siguiente comando calcula las raíces del polinomio característico, que son los autovalores de la matriz A.

```
>> roots(p)
```

Otra posibilidad es utilizar el comando `eig`

```
>> lambda = eig(A)
```

Obsérvese que A es diagonalizable, pues tiene tantos autovalores distintos como su orden. Luego, podemos continuar sin problemas.

El subespacio de autovectores asociado a cada autovalor λ es el núcleo de $\lambda I_2 - A$. Aunque es fácil hacerlo a mano, vamos a usar el comando `null` de `MATLAB` para obtener los autovectores asociados a cada autovalor. Teclea `help null` para ver una descripción del comando.

```
>> v1=null(lambda(1)*eye(2)-A,'r')
>> v2=null(lambda(2)*eye(2)-A,'r')
```

Por tanto, el autovector asociado a $\lambda_1 = 2$ es $\mathbf{v}_1 = (2, 1)^t$ y el autovector asociado a $\lambda_2 = 1$ es $\mathbf{v}_2 = (1, 1)^t$.

La opción `'r'` hace que `MATLAB` calcule el autovalor de una forma similar a como se haría a mano. Si no se usa la opción `'r'`, `MATLAB` calcula una base ortonormal del núcleo.

El comando

```
>> [P,D] = eig(A)
```

devuelve de forma directa la matriz diagonal $D = \mathrm{diag}(\lambda_1, \lambda_2)$ y la matriz de paso P tal que $D = P^{-1}AP$. En efecto,

```
>> inv(P)*A*P
```

Nuestra segunda tarea es escribir $\mathbf{x}_0$ como combinación lineal de $\mathbf{v}_1$ y $\mathbf{v}_2$. Así, queremos calcular c_1 y $c_2 \in \mathbb{R}$ tales que

$$\mathbf{x}_0 = c_1\mathbf{v}_1 + c_2\mathbf{v}_2,$$

que en nuestro caso es

$$\begin{pmatrix} 0 \\ 1 \end{pmatrix} = c_1 \begin{pmatrix} 2 \\ 1 \end{pmatrix} + c_2 \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

Esta ecuación entre vectores se puede escribir en forma matricial como

$$\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

$$P\mathbf{c} = \mathbf{x}_0.$$

Observemos que lo que estamos haciendo es un cambio de base. Conocemos las coordenadas respecto de la base $\mathcal{B} = \{\mathbf{e}_1, \mathbf{e}_2\}$ y queremos obtener las coordenadas respecto de la nueva base $\mathcal{B}' = \{\mathbf{v}_1, \mathbf{v}_2\}$. En este caso, P es la matriz del cambio de base de $\mathcal{B}'$ a $\mathcal{B}$. La solución del sistema es $\mathbf{c} = P^{-1}\mathbf{x}_0$. Vamos a ver cómo se puede calcular con MATLAB.

En primer lugar, definimos la matriz de paso $P \in \mathcal{M}_2(\mathbb{R})$

```
>> P=[v1,v2]
```

También se puede usar la matriz P calculada mediante el comando `[P,D] = eig(A)` aunque los resultados intermedios serán distintos, no así el resultado final debe ser el mismo.

Escribamos la condición inicial y calculemos $\mathbf{c}$

```
>> x0=[0;1];
>> c=inv(P)*x0
```

Por tanto,

$$\mathbf{c} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} -1 \\ 2 \end{pmatrix}.$$

Por último, sustituimos los valores de c_1, c_2, autovalores y autovectores en la ecuación (5.2.6), y obtenemos que

$$\mathbf{x}_n = (-1)(2)^n \begin{pmatrix} 2 \\ 1 \end{pmatrix} + (2)(1)^n \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

Tras simplificar,

$$\mathbf{x}_n = -2^n \begin{pmatrix} 2 \\ 1 \end{pmatrix} + \begin{pmatrix} 2 \\ 2 \end{pmatrix} = \begin{pmatrix} 2 - 2^{n+1} \\ 2 - 2^n \end{pmatrix}. \tag{5.2.7}$$

Podemos verificar que es correcto. Por ejemplo, para calcular $\mathbf{x}_{10}$ sustituimos $n = 10$ en 5.2.7 y nos queda

$$\mathbf{x}_{10} = -2^{10} \begin{pmatrix} 2 \\ 1 \end{pmatrix} + \begin{pmatrix} 2 \\ 2 \end{pmatrix} = \begin{pmatrix} -2048 \\ -1024 \end{pmatrix} + \begin{pmatrix} 2 \\ 2 \end{pmatrix} = \begin{pmatrix} -2046 \\ -1022 \end{pmatrix}.$$

En efecto,

```
>> x10 = -2^10*[2;1]+2*[1;1]
```

Observemos que coincide con el resultado obtenido en (5.2.5). Podemos usar también MATLAB para generar términos de la sucesión a partir de la ecuación (5.2.7).

```
>> Y=zeros(2,11);
>> for n=1:11,Y(:,n)=-2^(n-1)*[2;1]+2*[1;1];end
>> Y
```

Notemos que esta salida coincide con la que encontramos anteriormente al usar la ecuación (5.2.4). En efecto,

```
>> X == Y
```

De todo lo anterior se deduce que el término general de la solución de la ecuación en diferencias (5.2.3) es

$$x_{n+2} = 2 - 2^{n+1},\ n \geq 1,\ x_2 = 0, x_1 = 1.$$

Teniendo en cuenta que una sucesión de números reales es, en particular, una función $\mathbb{N} \to \mathbb{R}; n \mapsto x_n$, podemos definir una sucesión en MATLAB como una función. Para ello abrimos el editor de MATLAB y escribimos

```
function y = x(n)
  y = 2-2^(n-1);
```

y lo guardamos con el nombre `x.m`

Si escribimos ahora

```
>> x(12)
```

en MATLAB obtendremos el valor que toma la sucesión $(x_n)_{n\in\mathbb{N}}$ que es solución de la ecuación en diferencias (5.2.3) en $n = 12$.

2.2. Caso A no diagonalizable.

En apartado anterior vimos con un ejemplo cómo se podía obtener una forma cerrada de la solución de una ecuación en diferencias cuando su matriz asociada era diagonalizable. Exploremos ahora con otro ejemplo qué ocurre en el caso no diagonalizable. Para ello, consideremos el siguiente caso

$$x_{n+2} = 4x_{n+1} - 4x_n,\ n \geq 1,\ x_2 = 1, x_1 = -1, \tag{5.2.8}$$

cuya expresión matricial con la notación habitual es

$$\begin{cases} \mathbf{x}_n &= \begin{pmatrix} 4 & -4 \\ 1 & 0 \end{pmatrix} \mathbf{x}_{n-1},\ n \geq 1, \\ \mathbf{x}_0 &= (1,-1)^{\mathrm{t}} \end{cases} \tag{5.2.9}$$

Calculemos la forma canónica de Jordan de A tal y como se explicó en las clases de teoría. En principio podríamos tratar de calcularla con la orden `eig`

```
>> A = [4, -4; 1 ,0]
>> [P,J] = eig(A)
```

Hasta el momento no parece haber ningún problema; a menos que tratemos de comprobar la igualdad $J = P^{-1}AP$

```
>> inv(P)*A*P
```

ya que la matriz P que nos ha devuelto `MATLAB` no es invertible. Esto ocurre en general con `MATLAB` cuando usamos el comando `eig` con matrices no diagonalizables, y la matriz A de nuestro ejemplo no lo es, ya que su polinomio característico es $\aleph_A(x) = x^2 - 4x + 4 = (x-2)^2$ pero la dimensión del subespacio propio asociado a $\lambda = 2$ es uno; veámoslo:

```
>> poly(A)
```

Luego, A tiene un autovalor $\lambda = 2$ de multiplicidad 2.

```
>> lambda = eig(A)
```

Sin embargo la dimensión del subespacio propio $\ker(\lambda I_2 - A)$ es uno.

```
>> 2-rank(lambda(1)*eye(2) - A)
```

Por consiguiente, para calcular la forma canónica de Jordan de A necesitamos considerar los subespacios invariantes generalizados asociados a λ

$$L_0 = \{0\} \subseteq L_1 = \ker(\lambda I_2 - A) \subseteq L_2 = \ker((\lambda I_2 - A)^2) \subseteq \dots$$

En este caso, basta L_2 pues su dimensión ya coincide con la multiplicidad de λ.

```
>> 2-rank((lambda(1)*eye(2) - A)^2)
```

Dado que $n_2 = \dim(L_2) = 2, n_1 = \dim(L_1) = 1$ y $n_0 = \dim(L_0) = 0$, tenemos que $p_2 = n_2 - n_1 = 1$ y $p_1 = n_1 - n_0 = 1$. Luego, hay sabemos que hay $p_2 = 1$ bloques de Jordan de orden 2 y $p_1 - p_2 = 0$ bloques de Jordan de orden 1, es decir, la forma canónica de Jordan de A es

$$J = \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}$$

```
>> J = [2, 1; 0, 2]
```

Calculemos ahora una matriz $P \in \mathcal{M}_2(\mathbb{R})$ tal que $P^{-1}AP = J$; para ello elegimos p_2 vectores de L_2 que sean linealmente independientes módulo L_1, en nuestro caso, basta tomar un vector de L_2 que no esté en L_1, por ejemplo, $\mathbf{v}_{21} = \mathbf{e}_1 = (1,0)^{\mathrm{t}}$, y calculamos $\mathbf{v}_{11} = -(\lambda I_2 - A)\mathbf{v}_{21}$. Así, la matriz buscada no es más que $P = (\mathbf{v}_{11}|\mathbf{v}_{21})$

```
>> v21 = [1;0]
>> v11 = -(lambda(1)*eye(2)-A)*v21
>> P = [v11,v21]
>> inv(P)*A*P
>> J
```

Pasemos entonces a resolver la ecuación (5.2.9). Observemos que

$$\begin{aligned} \mathbf{x}_n &= A\mathbf{x}_{n-1} \\ &= A^2\mathbf{x}_{n-2} \\ &\vdots \\ &= A^n\mathbf{x}_0. \end{aligned}$$

El problema, por tanto, se reduce a encontrar una expresión de A^n. Aquí viene en nuestra ayuda la forma canónica de Jordan. Se tiene que

$$A^n = (P \cdot J \cdot P^{-1})^k = PJP^{-1}PJP^{-1}\cdots PJP^{-1} = PJ^nP^{-1}.$$

La cuestión ahora es si podemos encontrar fácilmente la expresión de J^n. Veamos el comportamiento:

```
>> J^2, J^3, J^4
```

Tal y como vimos en las clases de teoría, tenemos que

$$J^n = \begin{pmatrix} 2^n & n2^{n-1} \\ 0 & 2^n \end{pmatrix}.$$

Entonces la solución de la ecuación (5.2.9) es

$$\begin{aligned} \mathbf{x}_n = A^n\mathbf{x}_0 = PJ^nP^{-1}\mathbf{x}_0 &= \begin{pmatrix} 2^{n+1} & (n+1)2^n \\ 2^n & n2^{n-1} \end{pmatrix} P^{-1}\mathbf{x}_0 \\ = \begin{pmatrix} 2^{n+1} & (n+1)2^n \\ 2^n & n2^{n-1} \end{pmatrix}\begin{pmatrix} -1 \\ 3 \end{pmatrix} &= \begin{pmatrix} -2^{n+1} + 3(n+1)2^n \\ -2^n + 3n2^{n-1} \end{pmatrix} \end{aligned}$$

y el término general de la solución de la ecuación en diferencias (5.2.8) es, por lo tanto, $x_{n+2} = -2^{n+1} + 3(n+1)2^n = n2^{n+1} + (n+1)2^n,\ n \geq 1$.

Al igual que antes podemos definir la sucesión como una función de MATLAB

```
function y=x(n)
  y = (n-1)*2^(n-2) + (n-2)*2^(n-1)
```

que debemos de guardar con el nombre `x.m` para luego poder invocarla en la ventana de `MATLAB`

Ejercicios de la práctica 5

Ejercicio 1. Dar la forma cerrada de la solución y definir la correspondiente sucesión como función de MATLAB para cada una de siguiente ecuaciones en diferencias con la condición inicial dada.

1. $x_{n+3} = 5x_{n+2} - 8x_{n+1} + 4x_n, \quad x_2 = 3,\ x_1 = 2,\ x_0 = 1$
2. $x_{n+3} = 3x_{n+2} - 3x_{n+1} + x_n, \quad x_2 = 3,\ x_1 = 2,\ x_0 = 1$
3. $x_{n+3} = 2x_{n+2} + x_{n+1} - 2x_n, \quad x_2 = 3,\ x_1 = 2,\ x_0 = 1.$

PRÁCTICA 6

Matrices de Leslie

El modelo matricial de Leslie es una herramienta usada para determinar el crecimiento de una población así como la distribución por edad a lo largo del tiempo.

Esta práctica está centrada en el uso de la matriz de Leslie para determinar el crecimiento de una población y los porcentajes de distribución por edad a lo largo del tiempo. Esta descripción fue hecha por P.H. Leslie en 1945 (Biometrika, vol. 33, (1945), pp. 183-212). Se ha usado para estudiar la dinámica de poblaciones de una amplia variedad de organismos, como truchas, conejos, escarabajos, piojos, orcas, humanos o pinos.

Pre-requisitos: multiplicación de matrices e indexado en `MATLAB`, *autovalores y autovectores. Matrices no negativas irreducibles.*

1. Planteamiento y discusión del modelo

El modelo de Leslie para el estudio de una población una cierta especie de salmón parte de las siguiente hipótesis:

- Solamente se consideran las hembras en la población de salmones.
- La máxima edad alcanzada por un individuo son tres años.
- Los salmones se agrupan en tres tramos de un año cada uno.
- La probabilidad de sobrevivir un salmón de un año para otro depende de su edad.
- La tasa de supervivencia, s_i, en cada grupo es conocida.
- La fecundidad (tasa de reproducción), f_i, en cada grupo es conocida.
- La distribución de edad inicial es conocida.

Con este punto de partida es posible construir un modelo determinista con matrices. Como la edad máxima de un salmón es tres años, la población entera puede dividirse en tres clases de un año cada una. La clase 1 contiene los salmones en su primer año de vida, la clase 2 a los salmones entre 1 y 2 años, y la clase 3 a los salmones de más de dos años.

Supongamos que conocemos el número de hembras en cada una de las tres clases en un momento inicial. Llamemos $p_1(0)$ al número de hembras en la primera clase,

$p_2(0)$ al número de hembras en la segunda clase y $p_3(0)$ al número de hembras en la tercera clase. Con estos tres números formamos el vector

$$\mathbf{p}(0) = \begin{pmatrix} p_1(0) \\ p_2(0) \\ p_3(0) \end{pmatrix}.$$

Llamamos a $\mathbf{p}(0)$ el vector inicial de distribución por edad, o vector de distribución de edad en el instante inicial o instante 0.

A medida que el tiempo pasa, el número de hembras en cada una de las tres clases cambia por la acción de tres procesos biológicos: nacimiento, muerte y envejecimiento. Mediante la descripción de estos procesos de forma cuantitativa podremos estimar el vector de distribución por edad en el futuro.

Observaremos la población en intervalos discretos de un año, definidos por $0, 1, 2, \ldots$ Los procesos de nacimiento y muerte entre dos observaciones sucesivas se pueden describir a través de los parámetros *tasa media de reproducción* y *tasa de supervivencia*.

Sea f_1 el número medio de hembras nacidas de una hembra en la primera clase, f_2 el número medio de hembras nacidas de una hembra en la segunda clase, y f_3 el número medio de hembras nacidas de una hembra en la tercera clase. Cada f_i es la tasa media de reproducción de una hembra en la clase i-ésima.

Sea s_1 la fracción de hembras en la primera clase que sobreviven el año para pasar a la segunda clase. Sea s_2 la fracción de hembras en la segunda clase que sobreviven el año para pasar a la tercera clase. No hay s_3. Tras cumplir 3 años, el salmón muere tras desovar, y ninguno sobrevive para llegar a una cuarta clase. En general,

f_i es la tasa media de reproducción de una hembra en la clase i.

s_i es la tasa de supervivencia de hembras en la clase i.

Por su definición $f_i \geq 0$, porque la descendencia no puede ser negativa. En el caso de esta población de salmones, $f_1 = 0, f_2 = 0$, porque el salmón solamente produce huevos en su último año de vida. Por ello, únicamente f_3 tiene un valor positivo. Tenemos también que $0 < s_i \leq 1$ para $i = 1, 2$, porque suponemos que alguno de los salmones debe sobrevivir para llegar a la siguiente clase. Esto es cierto excepto para la última clase, donde el salmón muere.

Definimos el vector de distribución por edad en el instante j por

$$\mathbf{p}(j) = \begin{pmatrix} p_1(j) \\ p_2(j) \\ p_3(j) \end{pmatrix},$$

donde $p_i(j)$ es el número de salmones hembra en la clase i en el instante j.

En el instante j, el número de salmones en la primera clase, $p_1(j)$, es igual a los salmones nacidos entre los instantes $j-1$ y j. El número de descendientes producidos por cada clase se puede calcular multiplicando la tasa media de reproducción de la

clase por el número de hembras en la clase de edad. La suma de todos estos valores proporciona el total de descendientes. Así, escribimos

$$p_1(j) = f_1 p_1(j-1) + f_2 p_2(j-1) + f_3 p_3(j-1),$$

que indica que el número de hembras en la clase 1 es igual al número de hijas nacidas de hembras en la clase 1 entre los instantes $j-1$ y j más el número de hijas nacidas de hembras en la clase 2 entre $j-1$ y j, más el número de hijas nacidas de hembras en la clase 3 entre $j-1$ y j. En este ejemplo, como los salmones solamente producen huevos en su último año de vida, tenemos que $f_1 = f_2 = 0$, y nos queda la ecuación

$$p_1(j) = 0 \cdot p_1(j-1) + 0 \cdot p_2(j-1) + f_3 p_3(j-1).$$

El número de hembras en la segunda clase de edad en el instante j se obtiene a partir de las hembras de la primera clase en el instante $j-1$ que sobreviven al instante j. En forma de ecuación, nos queda

$$p_2(j) = s_1\, p_1(j-1).$$

El número de hembras en la tercera clase de edad en el instante j procede del número de hembras de la segunda clase de edad en el instante $j-1$ que sobreviven al instante j. Como antes, esto nos lleva a

$$p_3(j) = s_2\, p_2(j-1).$$

Por tanto, llegamos a la siguiente expresión:

$$\begin{array}{rcl} p_1(j) & = & f_1\, p_1(j-1) + f_2\, p_2(j-1) + f_3\, p_3(j-1) \\ p_2(j) & = & s_1\, p_1(j-1) \\ p_3(j) & = & s_2\, p_2(j-1) \end{array}$$

que en términos matriciales se puede expresar como

$$\begin{pmatrix} p_1(j) \\ p_2(j) \\ p_3(j) \end{pmatrix} = \begin{pmatrix} f_1 & f_2 & f_3 \\ s_1 & 0 & 0 \\ 0 & s_2 & 0 \end{pmatrix} \begin{pmatrix} p_1(j-1) \\ p_2(j-1) \\ p_3(j-1) \end{pmatrix}.$$

En notación vectorial nos queda

$$\mathbf{p}(j) = A\, \mathbf{p}(j-1),$$

donde

$$\mathbf{p}(j) = \begin{pmatrix} p_1(j) \\ p_2(j) \\ p_3(j) \end{pmatrix}$$

es la distribución por edad en el instante j y

$$A = \begin{pmatrix} f_1 & f_2 & f_3 \\ s_1 & 0 & 0 \\ 0 & s_2 & 0 \end{pmatrix}$$

se denomina matriz de Leslie.

Como en nuestro ejemplo $f_1 = f_2 = 0$, la matriz de Leslie para la población de salmones es

$$A = \begin{pmatrix} 0 & 0 & f_3 \\ s_1 & 0 & 0 \\ 0 & s_3 & 0 \end{pmatrix}.$$

Podemos generar ahora una sucesión de ecuaciones matriciales para calcular el vector de distribución por edad en cualquier instante j.

$$\begin{aligned} \mathbf{p}(1) &= A\,\mathbf{p}(0) \\ \mathbf{p}(2) &= A\,\mathbf{p}(1) = A(A\,\mathbf{p}(0)) = A^2\mathbf{p}(0) \\ \mathbf{p}(3) &= A\,\mathbf{p}(2) = A(A^2\mathbf{p}(0)) = A^3\mathbf{p}(0) \\ &\vdots \\ \mathbf{p}(j) &= A\,\mathbf{p}(j-1) = A(A^{j-1}\mathbf{p}(0)) = A^j\mathbf{p}(0) \end{aligned}$$

Por tanto, si conocemos el vector de distribución por edad inicial

$$\mathbf{p}(0) = \begin{pmatrix} p_1(0) \\ p_2(0) \\ p_3(0) \end{pmatrix}$$

y la matriz de Leslie podemos determinar el vector de distribución por edad de la población de hembras en cualquier instante posterior con la multiplicación de una potencia apropiada de la matriz de Leslie por el vector de distribución por edad inicial $\mathbf{p}(0)$.

2. Un ejemplo concreto con MATLAB

Supongamos que hay 1 000 hembras en cada una de las tres clases. Entonces

$$\mathbf{p}(0) = \begin{pmatrix} p_1(0) \\ p_2(0) \\ p_3(0) \end{pmatrix} = \begin{pmatrix} 1000 \\ 1000 \\ 1000 \end{pmatrix}.$$

Supongamos que la tasa de supervivencia del salmón en la primera clase es de $0,5\,\%$, la tasa de supervivencia del salmón en la segunda clase es $10\,\%$, y que cada hembra de

la tercera clase produce 2000 hembras en su puesta. Entonces $s_2 = 0,005$, $s_3 = 0,10$ y $f_3 = 2000$. La matriz de Leslie es entonces

$$A = \begin{pmatrix} 0 & 0 & 2000 \\ 0,005 & 0 & 0 \\ 0 & 0,10 & 0 \end{pmatrix}.$$

Para calcular el vector de distribución por edad después de un año, usamos la ecuación $\mathbf{p}(1) = L\mathbf{p}(0)$. Vamos a emplear MATLAB para dicho cálculo. Primero, introducimos el vector de distribución de edad inicial y la matriz de Leslie.

```
>> p0=[1000;1000;1000];
>> A=[0,0,2000;0.005, 0,0;0,0.1,0]
```

Notemos que MATLAB usa notación científica. El valor `1.0e+003` que precede a la matriz indica que debemos multiplicar cada entrada de la matriz por 1×10^3, es decir, hay que mover la coma decimal tres lugares a la derecha. Vamos a probar un nuevo formato para la salida (con `help format` se obtiene una lista completa de todas las posibilidades).

```
>> format short g
>> A=[0,0,2000;0.005, 0,0;0,0.1,0]
```

El comando `format short g` indica a MATLAB que use el mejor entre formato fijo o en coma flotante, según cada entrada de la matriz. Ahora calculamos $\mathbf{p}(1)$ como sigue.

```
>> p1=A*p0
```

El vector de distribución de edad $\mathbf{p}(1)$ muestra que tras el primer año hay 2 000 000 de salmones en la primera clase, 5 en la segunda clase y 100 en la tercera clase. Procedemos ahora a calcular $\mathbf{p}(2)$, el vector de distribución por edad después de 2 años.

```
>> p2=A*p1
```

El mismo resultado lo tendríamos con

```
>> p2=A^2*p0
```

El vector de distribución por edad $\mathbf{p}(2)$ indica que después de 2 años hay 200 000 salmones en la primera clase de edad, 10 000 en la segunda clase de edad y $0,5$ en la tercera clase. En la realidad, es imposible tener medio salmón. Sin embargo, apartemos de momento esta cuestión y calculemos la población tras 3 años.

```
>> p3=A*p2
```

Observemos que la población de salmones ha vuelto a su configuración original, con 1 000 peces en cada categoría. Usa MATLAB para realizar 4 iteraciones más $\mathbf{p}(4), \mathbf{p}(5)$, $\mathbf{p}(6)$ y $\mathbf{p}(7)$. ¿Qué pauta sigue?

2.1. El gráfico de un vector de distribución por edad.

Una de las mejores formas de examinar tendencias en el crecimiento de una población es dibujar el gráfico del vector de distribución por edad a lo largo del tiempo. También es deseable hacer un seguimiento de la población por más de tres o cuatro años.

La iteración de la ecuación $\mathbf{p}(j) = A\,\mathbf{p}(j-1)$ como lo hemos hecho antes es ineficiente. Si conocemos de antemano el número de veces que queremos realizar la iteración debemos usar un bucle `for` de MATLAB para realizarla.

La iteración de $\mathbf{p}(j) = A\,\mathbf{p}(j-1)$ un total de 24 veces producirá 24 generaciones del vector de distribución por edad. En MATLAB es recomendable reservar espacio en memoria para almacenar los resultados. Creamos entonces una matriz de ceros de orden 3×24. Las 3 filas se deben a que cada vector tiene tres componentes y las 24 columnas por las generaciones que deseamos calcular.

```
>> P=zeros(3,24);
```

Ahora colocamos el vector de distribución por edad inicial en la primera columna de la matriz P.

```
>> P(:,1)=p0;
```

Recordemos que la notación `P(:,1)` indica "todas las filas, primera columna". Por tanto, el comando `P(:,1)=p0;` pone las condiciones iniciales, contenidas en `p0`, en la primera columna de la matriz P.

Calculamos el contenido de las columnas 2 a 24 de la matriz P por iteración de la ecuación $\mathbf{p}(j) = A\,\mathbf{p}(j-1)$, con j variando de 2 a 24.

```
>> for j=2:24, P(:,j)=A*P(:,j-1); end
```

cuando el número de iteraciones se conoce de antemano, el bucle `for` de `MATLAB` es la solución más adecuada. Recordemos que `2:24` produce un vector fila, que comienza en 2 y con incremento de 1 llega a 24. Entonces el comando `for j=2:24` inicia un bucle que empieza con un valor de j igual a 2. En el siguiente paso del bucle j tendrá un valor de 3. La iteración continúa y el último paso por el bucle j tendrá un valor de 24. El comando `end` indica el final de las sentencias a ejecutar dentro del bucle.

El comando `P(:,j)=A*P(:,j-1)` merece una explicación. Recordemos que `P(:,j)` se lee como "matriz P, todas las filas, j-ésima columna". De igual forma, el comando `P(:,j-1)` se lee como "matriz P, todas las filas, $(j-1)$-ésima columna". Por tanto, el comando `P(:,j)=A*P(:,j-1)` calcula el producto de la matriz de Leslie A y la columna $(j-1)$-ésima de la matriz P, y almacena el resultado en la columna j-ésima de la matriz P. Hemos finalizado el comando con ";", pero puede resultar instructivo ejecutarlo sin él.

Una vez que la iteración está completa, podemos mostrar el contenido de la matriz P.

```
>> P
```

Teclea `help plot` en el indicador de `MATLAB` y lee la ayuda. Prestemos atención a la línea que indica `PLOT(Y) plots the columns of Y versus their index`. Sin embargo, la primera fila de la matriz P contiene el número de salmones hembra en la primera clase de edad, la segunda fila contiene la segunda clase de edad, y la tercera fila contiene el número de salmones hembra en la tercera y última clase de edad. Queremos pintar las filas de P a lo largo de su índice, pero `plot(P)` dibuja las columnas de P a lo largo de su índice.

Para ver la diferencia hagamos el siguiente experimento. Introducimos

```
>> Y=[1,2,3,4,5;2,3,4,5,1];
>> plot(Y,'*-'),figure,plot(Y','*-')
```

y veamos las matrices de la siguiente forma

$$Y \to \begin{matrix} 1 \\ 2 \end{matrix} \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 3 & 4 & 5 & 1 \end{pmatrix}, Y' \to \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{matrix} \begin{pmatrix} 1 & 2 \\ 2 & 3 \\ 3 & 4 \\ 4 & 5 \\ 5 & 1 \end{pmatrix}$$

Observamos que en la primera figura de la izquierda están representados los pares de puntos

$$\{(1,1),(2,2)\},\{(1,2),(2,3)\},\{(1,3),(2,4)\},\{(1,4),(2,5)\} \text{ y } \{(1,5),(2,1)\}.$$

En la figura de la derecha encontramos a los conjuntos de 5 puntos

$$\{(1,1),(2,2)),(3,3),(4,4),(5,5)\} \text{ y } \{(1,2),(2,3),(3,4),(4,5),(5,1)\}.$$

Por tanto, la solución para pintar lo que queremos de la matriz P es considerar su transpuesta.

```
>> plot(P')
```

Si hemos dicho que el comando `plot(P')` dibuja cada una de las columnas de la matriz P', ¿dónde están los otros dos gráficos en la figura? Si miramos con cuidado, observaremos que cerca del eje x hay algo por encima. Notemos que el valor superior de del eje y es 2×10^6. Cuando hay un rango tan amplio en los datos, como en este caso, que aparecen desde $1/2$ hasta $2\,000\,000$, podemos obtener una mejor visión dibujando el logaritmo de la población de salmones a lo largo del tiempo.

```
>> semilogy(P')
```

A menudo es útil añadir una leyenda al gráfico.

```
>> legend('Alevines','Pre-adultos','Adultos')
```

Se ve claramente a partir de la última figura que cada división por edad de la población de salmones oscila con periodo 3.

Podemos mejorar un poco el gráfico cambiando el tipo de línea. Ejecutemos los siguientes comandos.

```
>> h=semilogy(P')
>> set(h(1),'LineStyle','--')
>> set(h(2),'LineStyle',':')
```

```
>> legend('Alevines','Pre-adultos','Adultos')
>> grid off
```

Nota.- A partir de a versión 6 de MATLAB es posible cambiar el estilo de línea de forma interactiva, editando el gráfico y pulsando el botón derecho del ratón sobre la línea. Un menú desplegable nos muestra estilos de línea, color y otras propiedades.

3. Otro ejemplo con MATLAB

Consideremos ahora otra población también divida en tres clases de edad. Supongamos que cada hembra de la segunda y tercera clases producen una descendencia femenina de 4 y 3 miembros, respectivamente, en cada iteración. Supongamos además que el 50 % de las hembras de la primera clase sobreviven a la segunda clase, y que el 25 % de las hembras de la segunda clase llegan vivas a la tercera clase. La matriz de Leslie de esta población es

$$A = \begin{pmatrix} 0 & 4 & 3 \\ 0{,}5 & 0 & 0 \\ 0 & 0{,}25 & 0 \end{pmatrix}.$$

Supongamos que el vector inicial de población es

$$\mathbf{p}(0) = \begin{pmatrix} 10 \\ 10 \\ 10 \end{pmatrix}.$$

```
>> A=[0,4,3;0.5,0,0;0,0.25,0];
>> p0=[10;10;10];
```

Vamos a seguir los cambios en la población sobre un periodo de 10 años. Empezamos en el año cero y acabamos en el año 11. Hay tres clases que calcular en cada iteración. Empezamos creando una matriz que contendrá los datos de la población. La matriz tendrá tres filas, y cada fila contendrá los datos de una clase de edad. La matriz tendrá 11 columnas, y la primera de ellas tendrá el vector inicial de distribución por edad. Las diez restantes columnas almacenarán los vectores de distribución por edad en cada paso de la iteración (desde el año 1 hasta el año 10).

```
>> P=zeros(3,11);
```

Ponemos el vector inicial en la primera columna de la matriz P.

```
>> P(:,1)=p0;
```

Ahora usaremos la ecuación

$$\mathbf{p}(j) = A\,\mathbf{p}(j-1) \tag{6.3.1}$$

para calcular el vector de distribución por edad en los siguientes 10 años. Estos diez vectores se pondrán en las columnas 2 a la 11 de la matriz P. En el paso j-ésimo, calculamos el vector de distribución por edad número j multiplicando el correspondiente $j-1$ por la matriz A. Esto admite el siguiente bucle `for`.

```
>> for j=2:11, P(:,j)=A*P(:,j-1);end
```

Podemos ver el resultado introduciendo la variable que contiene los datos.

```
>> P
```

Recordemos que el prefijo `1.0e+003` significa que cada número en la salida debe multiplicarse por 10^3. Para el resto de la actividad, usaremos otro formato.

```
>> format short g
>> P
```

La distribución de población en cada año aparece como un vector columna de la matriz P. La gráfica de la evolución de la población a lo largo del tiempo se puede obtener como sigue.

```
>> j=0:10;
>> plot(j,P')
>> xlabel('Tiempo')
>> ylabel('Población')
```

El gráfico se aclara si añadimos una leyenda a cada color.

```
>> legend('Primera clase de edad','Segunda clase de edad', ...
'Tercera clase de edad')
```

Observemos que el número de hembras en cada grupo de edad en la figura se incrementa con el tiempo, con cierto comportamiento oscilatorio. Podemos dibujar el logaritmo de la población a lo largo del tiempo, tal como aparece en una figura obtenida con la siguiente secuencia de comandos.

```
>> j=(0:10);
>> semilogy(j,P')
>> xlabel('Tiempo')
>> ylabel('Log Poblaci´on')
>> legend('Primera clase de edad','Segunda clase de edad', ...
'Tercera clase de edad')
```

Nota.- Sabemos que las matrices de Leslie son irreducibles, por lo que posee un autovalor real positivo que es mayor que cualquiera de sus otros autovalores. Además, este autovalor tiene multiplicidad uno y tiene un autovector positivo asociado.

Vamos a usar `MATLAB` para calcular los autovalores y autovectores de A.

```
>> [V,D]=eig(A)
```

Denotemos $\lambda_1 = 1{,}5, \lambda_2 = -1{,}309$ y $\lambda_3 = -0{,}19098$, y $\mathbf{v}_j$ la columna j-ésima de V, $j = 1, 2, 3$.

En este caso, vemos que $\rho := \lambda_1 = 1{,}5$ es el autovalor dominante, y un autovector asociado positivo a ρ es

$$\mathbf{v} = -\mathbf{v}_1 = \begin{pmatrix} 0{,}947370 \\ 0{,}315790 \\ 0{,}052632 \end{pmatrix},$$

que es la primera columna de la matriz V cambiada de signo.

Por lo que hemos visto en clase de teoría, el límite de las proporciones de cada clase de edad sobre la población total es igual a $\mathbf{v}/\sum_{i=1}^{n} v_i$. En este caso podemos calcular

```
>> v=-V(:,1)
>> v/sum(v)
```

Por tanto, la primera clase de edad compondrá el 72 % de la población, la segunda clase el 24 % y la tercera clase el 4 % de la población total.

Vamos a comprobar con MATLAB que, en efecto, el comportamiento a largo plazo de la población sigue este esquema.

Desarrollando la expresión (6.3.1) obtenemos que

$$\mathbf{p}(j) = A\,\mathbf{p}(j-1) = A^j\mathbf{p}(0) = VDV^{-1}\mathbf{p}(0) = c_1\rho^j\mathbf{v}_1 + c_2\lambda_2^j\mathbf{v}_2 + c_3\lambda_3^j\mathbf{v}_3 \tag{6.3.2}$$

En nuestro caso queda

$$\begin{aligned}\mathbf{p}(j) =& c_1(1{,}5)^j\begin{pmatrix} -0{,}94737 \\ -0{,}31579 \\ -0{,}052632 \end{pmatrix} + c_2(-1{,}309)^j\begin{pmatrix} 0{,}93201 \\ -0{,}356 \\ -0{,}067989 \end{pmatrix} \\ &+ c_3(-0{,}19098)^j\begin{pmatrix} 0{,}22588 \\ -0{,}59137 \\ 0{,}77412 \end{pmatrix}\end{aligned}$$

```
>> p100=A^100*p0
>> p100/sum(p100)
```

Los comandos anteriores han calculado el porcentaje de población de cada clase de edad tras 100 años. Vemos que coincide con lo que habíamos deducido a partir de $\mathbf{v}$.

Vamos a dibujar la evolución de los porcentajes de cada clase de edad en los primeros 100 años. Primero almacenamos los vectores de distribución por edad.

```
>> P=zeros(3,101);
>> P(:,1)=p0;
>> for j=2:101,P(:,j)=A*P(:,j-1);end
```

Ahora podemos obtener los porcentajes de cada clase de edad sobre la población total dividiendo cada columna por su suma.

```
>> G=zeros(3,101);
>> for j=1:101, G(:,j)=P(:,j)/sum(P(:,j));end
```

La gráfica de estas poblaciones "normalizadas.es interesante.

```
>> j=0:100;
>> plot(j,G')
>> xlabel('Tiempo')
>> ylabel('Porcentajes')
>> legend('Primera clase de edad','Segunda clase de edad',...
```

```
'Tercera clase de edad')
```

Después de un número suficiente de años, el porcentaje de organismos en cada clase se aproxima a 74 %, 24 % y 4 %.

El autovalor dominante $\rho = 1{,}5$ nos dice cómo cambia el vector de población de un año para otro. Veamos los siguientes comandos.

```
>> p99=A^99*p0
>> p100./p99
```

El comando `p100./p99` divide cada componente del vector `p100` por la correspondiente del vector `p99`. En este caso vemos que el número de hembras en cada clase de edad después de 100 años es 1,5 veces el número de hembras en cada clase tras 99 años. En general, tras un periodo largo de tiempo, $\mathbf{p}(j) = 1{,}5\mathbf{p}(j-1)$. Esta fórmula se puede deducir de la ecuación 6.3.2 como sigue. Por la ecuación 6.3.2 para j suficientemente grande tenemos que

$$\mathbf{p}(j) \approx c_1 \rho^j \mathbf{v}_1.$$

De forma análoga tenemos que

$$\mathbf{p}(j-1) \approx c_1 \rho^{j-1} \mathbf{v}_1,$$

o de forma equivalente

$$\mathbf{v}_1 \approx \frac{1}{c_1 \rho^{j-1}} \mathbf{p}(j-1).$$

Entonces

$$\mathbf{p}(j) \approx c_1 \rho^j \frac{1}{c_1 \rho^{j-1}} \mathbf{p}(j-1) = \rho\, \mathbf{p}(j-1).$$

4. Resumen

El modelo de Leslie está definido por la ecuación $\mathbf{p}(j) = L^j \mathbf{p}(0)$, donde $\mathbf{p}(0)$ es el vector inicial de distribución de la población, y $\mathbf{p}(j)$ el el vector de distribución de población en el instante j. Si A es diagonalizable, entonces $A = VDV^{-1}$, donde D es una matriz diagonal formada por los autovalores de A. Las columnas de V son los autovectores correspondientes. En este caso, el modelo de Leslie se puede escribir como

$$\mathbf{p}(j) = c_1 \lambda_1^j \mathbf{v}_1 + c_2 \lambda_2^j \mathbf{v}_2 + \ldots + c_n \lambda_n^j \mathbf{v}_n,$$

donde $\lambda_i, \mathbf{v}_i$ son autovalor y autovector asociados. Si $\rho = \lambda_1$ es autovalor estrictamente dominante de A, entonces para valores grandes de j se tiene que

$$\mathbf{p}(j) \approx c_1 \rho^j \mathbf{v}_1,$$

y la proporción de hembras en cada clase de edad tiende a una constante. Estas proporciones límites se pueden determinar a partir de las componentes de $\mathbf{v}_1$. Por último, el autovalor dominante ρ determina la tasa de cambio de un año para otro. Como

$$\mathbf{p}(j) \approx \rho \mathbf{p}(j-1)$$

para valores grandes de j, el vector de población en el instante j es un múltiplo del vector de población en ele instante $j-1$. Si $\lambda_1 > 1$ entonces la población tendrá un crecimiento indefinido. Si $\lambda_1 < 1$, entonces la población se extinguirá.

Ejercicios de la práctica 6

Ejercicio 1. Supongamos que una especie de salmón vive cuatro años. Además, supongamos que la tasa de supervivencia en sus primero, segundo y tercer años son, respectivamente, $0,5\,\%, 7\,\%$ y $15\,\%$. Sabemos también que cada hembra en la cuarta clase de edad produce 5 000 huevos de hembra. Las otras clases de edad no tienen descendencia.

1. Calcular la matriz de Leslie de la población.
2. Si se introducen en el sistema 1 000 salmones hembra en cada clase de edad, calcular el vector de distribución de edad inicial.
3. Usar un bucle `for` para iterar la ecuación de Leslie 25 veces. Usar los gráficos de `MATLAB` para dibujar el logaritmo de cada clase de edad a lo largo del tiempo. ¿Cuál es el destino de esta población de salmones?
4. Calcular la población de salmones en la iteración número 50, sin calcular las 49 iteraciones anteriores.

Ejercicio 2. En la misma situación anterior, pero con tasas de supervivencia iguales a $2\,\%, 15\,\%$ y $25\,\%$, respectivamente. Cada hembra de la cuarta clase produce 5 000 huevos hembra. Responder a las mismas cuestiones del ejercicio anterior.

Ejercicio 3. En la misma situación anterior, pero con tasas de supervivencia iguales a $1\,\%, 10\,\%$ y $2\,\%$, respectivamente. Cada hembra de la cuarta clase produce 5 000 huevos hembra. Responder a las mismas cuestiones del ejercicio anterior.

Ejercicio 4. Las hembras de cierta especie animal viven tres años. Supongamos que la tasa de supervivencia de hembras en sus primero y segundo años es del $60\,\%$ y $25\,\%$, respectivamente. Cada hembra del segundo grupo de edad tiene 4 hijas al año de media, y cada hembra del tercer grupo tiene una media de 3 hijas por año.

1. Calcular la matriz de Leslie de esta población.
2. Supongamos que al inicio hay 10 hembras en cada clase de edad. Usar `MATLAB` para calcular el vector de distribución por edad para los primeros 100 años, y dibujar los vectores de distribución por edad con los comandos `plot` y `semilogy`.
3. Usar `MATLAB` para calcular los autovalores y autovectores de la matriz de Leslie. ¿Qué le ocurre a esta población a lo largo del tiempo?
4. Tras 100 años, ¿cuál es el porcentaje de hembras en cada clase?
5. A largo plazo, ¿cuál es el factor de aumento o disminución?

Ejercicio 5. Igual que el ejercicio anterior, con tasas de supervivencia iguales a $20\,\%$ y $25\,\%$ y resto de datos iguales.

Ejercicio 6. Supongamos que una población de salmones vive tres años. Cada salmón adulto produce 800 huevos hembras. La probabilidad de que un salmón sobreviva el primer año y pase al segundo año es del 5 %, y la probabilidad de que un salmón sobreviva el segundo año y llegue al tercero es 2,5 %.

1. Calcule la matriz de Leslie de esta población.
2. Supongamos que al inicio hay 10 hembras en cada clase de edad. Use `MATLAB` para calcular el vector de distribución por edad para los primeros 100 años.
3. Use `MATLAB` para calcular los autovalores y autovectores de la matriz de Leslie. ¿Hay un autovalor dominante?
4. Describir el comportamiento de la población a lo largo del tiempo.

Ejercicio 7. Supongamos que la población de un país se divide en clases de 6 años de duración. Los valores de las tasas de reproducción f_i y supervivencia s_i para cada clase se muestran en la siguiente tabla:

i	f_i	s_i
1	0	0.99670
2	0.00102	0.99837
3	0.08515	0.99780
4	0.30574	0.99672
5	0.40002	0.99607
6	0.28061	0.99472
7	0.15260	0.99240
8	0.06420	0.98867
9	0.01483	0.98274
10	0.00089	0

Supongamos que hay 10 hembras en cada una de las 10 clases al principio. Resolver las mismas preguntas que en el ejercicio 4.

PRÁCTICA 7

Cadenas de Markov

En líneas generales, un proceso estocástico consiste en una serie de sucesos que cambian con el tiempo de una forma secuencial y con ciertas probabilidades. Los sucesos no suelen ser independientes, y lo que ocurra en el instante t depende de lo ocurrido en los instantes $t-1, t-2, \ldots$ Cuando la probabilidad asociada a un suceso depende solamente de su estado anterior, el proceso se denomina cadena de Markov.

En esta actividad analizamos diversos procesos que pueden ser modelizados por una cadena de Markov, y estudiaremos la situación límite.

Pre-requisitos: Autovalores y autovectores. Matrices no negativas

1. Un ejemplo con MATLAB

Supongamos que los procesos migratorios entre dos zonas geográficas, que llamaremos Norte y Sur, son como siguen. Cada año, el 50 % de la población del Norte emigra al Sur, mientras que el 25 % de la población del Sur emigra al Norte. Este proceso se puede representar como aparece en la figura 1.

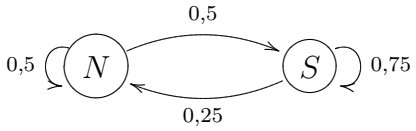


Figura 1. Procesos migratorios.

Queremos estudiar la evolución de la población a largo plazo. Sea n_t la proporción de la población total que vive en el Norte al final del año t, y s_t la correspondiente para la que vive en el Sur. El modelo de migración establece que las proporciones de población en cada región al final del año $t+1$ son

$$\left\{ \begin{array}{rcl} n_{t+1} & = & n_t(,5) + s_t(,25) \\ s_{t+1} & = & n_t(,5) + s_t(,75) \end{array} \right. \tag{7.1.1}$$

Si escribimos

$$\mathbf{p}_t = \begin{pmatrix} n_t \\ s_t \end{pmatrix}$$

para indicar el vector de población en el instante m, entonces la ecuación (7.1.1) se puede escribir como

$$\mathbf{p}_{t+1} = P\mathbf{p}_t \tag{7.1.2}$$

donde

$$P = \begin{pmatrix} ,5 & ,25 \\ ,5 & ,75 \end{pmatrix},$$

es la matriz de transición, porque contiene las probabilidades de transición de un estado a otro en el sistema. Supongamos que el vector de población inicial es $\mathbf{p}_0 = \begin{pmatrix} 0{,}9 \\ 0{,}1 \end{pmatrix}$. Calculemos la evolución en los próximos 10 años.

```
>> P=[0.5,0.25;0.5,0.75]
>> p0=[9/10;1/10];
>> X=zeros(2,10);X(:,1)=p0;
>> for t=2:10,X(:,t)=P*X(:,t-1);end
>> plot(X')
>> legend('Pobl. en el Norte','Pobl. en el Sur')
```

Observamos que el sistema se vuelve estable. El vector de estado converge a un vector fijo. En este caso decimos que el proceso ha alcanzado el equilibrio. El vector fijo recibe el nombre de vector de estado estacionario. En este caso tenemos lo siguiente.

```
>> X(:,8:10)
```

Podemos calcular la expresión exacta del vector estacionario a partir de la forma canónica de Jordan. Sea $\mathbf{p}_0 = \begin{pmatrix} n_0 \\ s_0 \end{pmatrix}$ un vector de población inicial. Los autovalores de la matriz P son $\lambda_1 = 1/4$ y $\lambda_2 = 1$. Los autovectores asociados respectivos son

$$\mathbf{v}_1 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}.$$

```
> format rat
> lambda = eig(P)
```

Observamos que la matriz T es diagonalizable. Calculemos ahora la forma canónica de Jordan J y matriz de paso P.

```
> J = diag(lambda);
```

```
> v1 = null(lambda(1)*eye(2)-P, 'r');
> v2 = null(lambda(2)*eye(2)-P, 'r');
> Q = [v1,v2]
```

Entonces la forma canónica de Jordan es

$$J = \begin{pmatrix} 1/4 & 0 \\ 0 & 1 \end{pmatrix}$$

y la matriz de paso es

$$Q = \begin{pmatrix} -1 & 1 \\ 1 & 2 \end{pmatrix}.$$

Se tiene que $P = QJQ^{-1}$, y es claro que

$$\lim_{t\to\infty} J^t = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

De aquí se deduce que

$$\lim_{t\to\infty} P^t = \lim_{t\to\infty} QJ^tQ^{-1} = Q\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}Q^{-1} = \begin{pmatrix} 1/3 & 1/3 \\ 2/3 & 2/3 \end{pmatrix}.$$

```
> Jinf = [0,0;0,1]
> Pinf = Q*Jinf*inv(Q)
> format
```

Entonces si escribimos $\mathbf{p}^\infty = \lim_{t\to\infty} \mathbf{p}_m$ obtenemos que

$$\begin{aligned} \mathbf{p}_\infty &= \lim \mathbf{x}_m \\ &= \lim_{t\to\infty} P^t\mathbf{p}_0 \\ &= \begin{pmatrix} 1/3 & 1/3 \\ 2/3 & 2/3 \end{pmatrix}\begin{pmatrix} n_0 \\ s_0 \end{pmatrix} \\ &= \begin{pmatrix} 1/3n_0 + 1/3s_0 \\ 2/3n_0 + 2/3s_0 \end{pmatrix} \\ &= \begin{pmatrix} 1/3 \\ 2/3 \end{pmatrix}, \end{aligned}$$

porque recordemos que $n_0 + s_0 = 1$.

Existen procesos de este tipo que no tienen ese equilibrio. Por ejemplo, consideremos un dispositivo electrónico que puede estar en tres estados 1, 2 y 3, y supongamos

que el dispositivo cambia a unos ciclos regulares de reloj. Si se encuentra en los estados 1 o 3 cambia a 2 en el siguiente ciclo. Si se encuentra en 2 cambiará a 1 o a 3 en el siguiente ciclo con igual probabilidad. La matriz de transición es

$$P = \begin{pmatrix} 0 & 0{,}5 & 0 \\ 1 & 0 & 1 \\ 0 & 0{,}5 & 0 \end{pmatrix}.$$

Si partimos de $\mathbf{p}_0 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$, el comportamiento del sistema es periódico.

$$\mathbf{p}_1 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \mathbf{p}_2 = \begin{pmatrix} 0{,}5 \\ 0 \\ 0{,}5 \end{pmatrix}, \mathbf{p}_3 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \ldots$$

En efecto,

```
>> format short g
>> P = [0, 0.5, 0; 1, 0, 1; 0, 0.5, 0]
>> p0 = [1;0;0]
>> X=zeros(3,10);X(:,1)=p0;
>> for t=2:10,X(:,t)=P*X(:,t-1);end
>> plot(X')
>> legend('Primer estado','Segundo estado','Tercer estado')
```

Sin embargo, si pedimos que la matriz de transición satisfaga una propiedad razonable (por ejemplo que sea primitiva), obtenemos unos procesos que sí alcanzan el equilibrio.

2. Otros ejemplos con MATLAB

2.1. Procesos de movilidad social.

Consideremos el problema de la movilidad social que involucra la transición entre distintas clases sociales a través de las generaciones sucesivas de una familia. Supongamos que cada individuo es clasificado socialmente según su ocupación como clase alta, media o baja, que etiquetamos como estados $1, 2$ y 3, respectivamente. Supongamos que la matriz de transición que relaciona la clase de un hijo con la de su padre

es

$$P = \begin{pmatrix} 0{,}45 & 0{,}05 & 0{,}05 \\ 0{,}45 & 0{,}70 & 0{,}50 \\ 0{,}10 & 0{,}25 & 0{,}45 \end{pmatrix},$$

de tal forma que, por ejemplo, la probabilidad de que un hijo sea clase alta, media o baja cuando su padre es de clase baja viene dada por la última columna de P. Como P es primitiva (pues es positiva), podemos aplicar los resultados discutidos anteriormente. Un simple análisis de los autovalores y autovectores de P revela que el autovector positivo $\mathbf{p}$ tal que $p_1 + p_2 + p_3 = 1$ es

$$\mathbf{p} = \begin{pmatrix} 0{,}0833 \\ 0{,}6198 \\ 0{,}2969 \end{pmatrix}.$$

En efecto,

```
>> P = [0.45, 0.05, 0.05; 0.45, 0.70, 0.50; 0.10, 0.25, 0.45]
>> p = null(eye(3) - P, 'r')
>> p = p/sum(p)
```

Por consiguiente, si este proceso verifica las condiciones de una cadena de Markov homogénea y finita, después de una cantidad considerable de generaciones, la población masculina consistiría en un 8.3 % de clase alta, un 62 % de clase media y un 29.7 % de clase baja.

Veamos experimentalmente que el resultado es el mismo para cualquier dato inicial.

```
>> p0 = rand(3,1)
>> p0 = p0/sum(p0)
>> p100 = P*p0
```

2.2. Sistemas de seguridad.

Consideremos un sistema que tiene dos controles independientes, A y B, que previene que el sistema sea destruido. El sistema se activa en momentos discretos $t_1, t_2, t_3, \ldots$, y el sistema se considera bajo control si alguno de los controles A o B funciona en el momento de la activación. El sistema se destruye si A y B fallan simultáneamente. Por ejemplo, un automóvil tiene dos sistemas de frenado independientes, el freno de pedal y el freno de mano. El automóvil está bajo control si al

menos uno de los sistemas de frenado está operativo cuando intentamos parar, pero choca si ambos sistemas fallan simultáneamente.

Si uno de los controles falla en un punto de activación pero el otro control funciona, entonces el control defectuoso es reemplazado antes de la siguiente activación. Si un control funciona en el momento t entonces se considera fiable en un 90 % para la activación $t+1$. Sin embargo, si un control falla en el instante t, entonces su recambio no probado se considera fiable en un 60 % para $t+1$.

La pregunta que nos planteamos es: ¿Puede el sistema funcionar indefinidamente sin ser destruido? Si no, ¿cuánto tiempo se espera que el sistema funcione antes de la destrucción?

Este problema se puede modelizar con una cadena de Markov con cuatro estados, definidos por los controles que funcionen en un momento de activación. Podemos poner entonces que el espacio de estados es el conjunto de pares (a, b) tales que

$$a = \begin{cases} 1 & \text{si } A \text{ funciona,} \\ 0 & \text{si } A \text{ falla,} \end{cases} \quad \text{y } b = \begin{cases} 1 & \text{si } B \text{ funciona,} \\ 0 & \text{si } B \text{ falla.} \end{cases}$$

El estado $(0, 0)$ es absorbente, es decir, si se llega a él no se puede salir.

Por simplicidad, escribiremos $1, 2, 3$ y 4 en vez de $(1, 1)$, $(1, 0)\,(0, 1)$ y $(0, 0)$, respectivamente. De este modo la matriz de transición es

$$P = \begin{pmatrix} 0{,}81 & 0{,}54 & 0{,}54 & 0 \\ 0{,}09 & 0{,}36 & 0{,}06 & 0 \\ 0{,}09 & 0{,}06 & 0{,}36 & 0 \\ 0{,}01 & 0{,}04 & 0{,}04 & 1 \end{pmatrix}$$

En este caso, P no es primitiva. Sin embargo, los autovalores de la matriz P son 0,9827, 0,2473, 0,3 y 1.

```
>> P = [0.81, 0.54, 0.54, 0; ...
        0.09, 0.36, 0.06, 0; ...
        0.09, 0.06, 0.36, 0; ...
        0.01, 0.04, 0.04, 1]
>>eig(P);
```

Entonces, existe el límite $\lim_{t\to\infty} P^t$, y es igual a

$$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix}.$$

Esto significa que el estado absorbente se alcanza siempre, partamos de donde partamos. Así que tenemos respondida a la primera pregunta: el sistema se destruirá, a la larga, con probabilidad 1. La segunda cuestión que planteábamos es en cuántos procesos de activación llegaremos al desastre. Se puede probar que si escribimos

$$P = \begin{pmatrix} P_{11} & \mathbf{0} \\ \mathbf{p}_{12} & 1 \end{pmatrix},$$

donde P_{11} es la submatriz de P formada por las tres primeras filas y columnas, entonces el número medio de pasos antes de caer en el estado absorbente, si partimos del estado i-ésimo, es igual a $(\mathbf{u}^{\mathrm{t}}(I_3 - P_{11})^{-1})_i$, donde $\mathbf{u}$ es el vector con todas sus componentes iguales a 1 (esto es, la suma de las entradas de la columna i-ésima). En efecto, la submatriz P_{11} da la probabilidad de ir desde cualquier estado no absorbente a otro estado no absorbente en un paso exactamente, P_{11}^2 da las probabilidades de ir desde cualquier estado no absorbente hasta otro estado no absorbente en dos pasos exactamente. P_{11}^3 da información similar para tres pasos, ... Por lo tanto, P_{11}^n da esta misma información para n pasos. Para hallar el número esperado de pasos antes que el proceso sea absorbido, consiste en calcular el número esperado de veces que el proceso puede estar en cada estado no absorbente y sumarlos. Esto totalizaría el número de pasos antes de que el proceso fuera absorbido y por consiguiente el número esperado de pasos hacia la absorción. Como

$$I_3 + P_{11} + P_{11}^2 + P_{11}^3 + \ldots = (I_3 - P_{11})^{-1}$$

se sigue que $(I_3 - P_{11})^{-1}$ representa el número esperado de períodos que el sistema estará en cada estado no absorbente antes de la absorción, por lo tanto la suma de cada fila de $(I_3 - P_{11})^{-1}$ representa el promedio de períodos que transcurren antes de ir a un estado absorbente. En nuestro caso,

$$(I_3 - P_{11})^{-1} = \begin{pmatrix} 44{,}615 & 41{,}538 & 41{,}538 \\ 6{,}9231 & 8{,}022 & 6{,}5934 \\ 6{,}9231 & 6{,}5934 & 8{,}022 \end{pmatrix},$$

y

$$\mathbf{u}^t(I_3 - P_{11})^{-1} = \begin{pmatrix} 58{,}462 & 56{,}154 & 56{,}154 \end{pmatrix}.$$

```
>> P11 = P(1:3,1:3)
>> X = inv(eye(3)-P11)
>> u = ones(3,1)
>> u'*X
```

Interpretemos los resultados. El tiempo medio para fallo si partimos con los dos controles probados es algo más de 58 pasos, mientras que el tiempo medio para

fallo si partimos con uno de los controles no probado está alrededor de los 56 pasos. La diferencia no parece significativa, pero vamos a considerar qué ocurre usamos solamente un control en el sistema. En este caso, solamente hay dos estados en la cadena de Markov: 1 (control que funciona) y 2 (control que no funciona). La matriz de transición queda

$$P = \begin{pmatrix} 0{,}9 & 0 \\ 0{,}1 & 1 \end{pmatrix}$$

por lo que el tiempo medio de fallo es únicamente de $\mathbf{u}^{\mathrm{t}}(I - P_{11})^{-1} = 10$ pasos ¿Qué ocurrirá si usamos tres controles independientes?

Ejercicios de la práctica 7

Ejercicio 1. Determinar cuáles de las siguientes matrices son matrices de transición.

$$\text{(a)} \begin{pmatrix} 0{,}3 & 0{,}7 \\ 0{,}4 & 0{,}6 \end{pmatrix}, \quad \text{(b)} \begin{pmatrix} 0{,}2 & 0{,}3 & 0{,}1 \\ 0{,}8 & 0{,}5 & 0{,}7 \\ 0{,}0 & 0{,}2 & 0{,}2 \end{pmatrix}$$

Ejercicio 2. En un experimento, se coloca todos los días una rata en una jaula con dos puertas A y B. La rata puede pasar por la puerta A, y recibe una descarga eléctrica, o por la puerta B, y obtiene cierto alimento. Se registra la puerta por la que pasa la rata. Al inicio del experimento, la rata tiene la misma probabilidad de pasar por la puerta A que por la puerta B. Después de pasar por la puerta A y recibir una descarga, la probabilidad de seguir pasando por la misma puerta al día siguiente es 0,3. Después de pasar por la puerta B y recibir alimento, la probabilidad de pasar por la misma puerta al día siguiente es 0,6.

1. Escribir la matriz de transición para el proceso de Markov.
2. ¿Cuál es la probabilidad de que la rata continúe pasando por la puerta A el tercer día después del inicio del experimento?
3. ¿Cuál es el vector de estado estacionario?

Ejercicio 3. Un país está dividido en tres regiones demográficas. Se calcula que cada año un 5 % de residentes de la región 1 se mudan a la región 2, y un 5 % se desplazan a la región 3. De los residentes de la región 2, el 15 % van a la región 1 y el 10 % a la región 3. Y de los residentes de la región 3, el 10 % se mueven a la región 1 y el 5 % a la región 2. ¿Qué porcentaje de población reside en cada una de las tres regiones tras un largo periodo de tiempo?

Ejercicio 4. Usar las mismas premisas del ejemplo del sistema de seguridad, pero con tres controles A, B y C. Determinar el tiempo medio de fallo si partimos de tres controles probados, con dos probados y uno sin probar, y con uno probado y dos sin probar.

PRÁCTICA 8

Proyección ortogonal. Mínimos cuadrados

En esta práctica ilustraremos con algunos ejemplos los conceptos de proyección ortogonal sobre un vector y sobre un subespacio vectorial. Además, usaremos la proyección ortogonal y la inversa de Moore-Penrose para calcular la solución aproximada mínimo cuadrática de diversos sistemas de ecuaciones lineales.

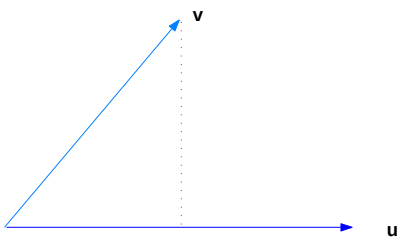
Pre-requisitos: Sistemas de ecuaciones lineales. Proyección ortogonal. Inversa de Moore-Penrose.

1. Proyección ortogonal

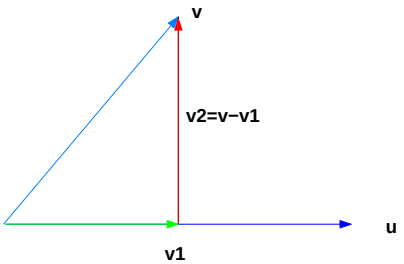
Comencemos recordando algunos cuestionas relacionadas con la proyección ortogonal en $\mathbb{R}^n$ con el producto escalar usual.

1.1. Proyección de un vector sobre una recta.

Queremos proyectar un vector $\mathbf{v}$ sobre otro vector $\mathbf{u}$. En términos geométricos, esto significa que queremos calcular el vector sobre $\mathbf{u}$ que es más "próximo" al vector $\mathbf{v}$. El concepto de ortogonalidad entra entonces en juego. En la figura de la derecha "proyectamos" el vector $\mathbf{v}$ sobre el $\mathbf{u}$.



En este caso se trata de una proyección ortogonal. El resultado es el vector $\mathbf{v}_1$ que aparece en la siguiente figura.



Observemos que esta elección de $\mathbf{v}_1$ hace que el vector $\mathbf{v}_2 = \mathbf{v} - \mathbf{v}_1$ tan pequeño de norma como sea posible.

Como proyectamos sobre el vector $\mathbf{u}$, el vector $\mathbf{v}_1$ debe ser de la forma $\mathbf{v}_1 = \alpha\mathbf{u}$, un múltiplo escalar de $\mathbf{u}$. Nuestro objetivo es calcular α. Como el vector $\mathbf{v}_2$ es ortogonal a $\mathbf{u}$. Entonces, exigimos que

$$0 = \mathbf{u} \cdot \mathbf{v}_2 = \mathbf{u} \cdot (\mathbf{v} - \mathbf{v}_1) = \mathbf{u} \cdot (\mathbf{v} - \alpha\mathbf{u}) = \mathbf{u} \cdot \mathbf{v} - \alpha\mathbf{u} \cdot \mathbf{u},$$

luego

$$\alpha = \frac{\mathbf{u} \cdot \mathbf{v}}{\mathbf{u} \cdot \mathbf{u}} = \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\|^2}.$$

Con este valor de α, el vector de proyección $\mathbf{v}_1 = \alpha\mathbf{u}$ se obtiene fácilmente. Pues *si* $\mathbf{u}$ *y* $\mathbf{v}$ *son vectores en* $\mathbb{R}^n$ *entonces el vector proyección del vector* $\mathbf{v}$ *sobre el vector* $\mathbf{u}$ *es*

$$\mathbf{v}_1 = \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\|^2}\mathbf{u}. \tag{8.1.1}$$

Ejemplo 8.1.1. Supongamos que queremos proyectar el vector $\mathbf{v} = (2, -1, 0)$ sobre el vector $\mathbf{u} = (1, 1, 1)$. Entonces, con la fórmula (8.1.1) y `MATLAB` nos queda

```
>> u=[1;1;1];v=[2;-1;0];
>> v1=dot(u,v)/dot(u,u) * u
```

La aplicación que realiza la proyección de un vector sobre otro es lineal. Buscamos la matriz P que aplica el vector $\mathbf{v}$ sobre el vector $\mathbf{v}_1$ calculado. Lo podemos hacer a partir de la expresión (8.1.1). Recordemos que $\mathbf{u} \cdot \mathbf{v}$ lo podemos escribir en forma matricial como $\mathbf{u}^{\mathrm{t}}\mathbf{v}$. Entonces tenemos

$$\mathbf{v}_1 = \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\|^2}\mathbf{u} = \frac{\mathbf{u} \cdot \mathbf{v}}{\mathbf{u} \cdot \mathbf{u}}\mathbf{u} = \mathbf{u}\frac{\mathbf{u}^{\mathrm{t}}\mathbf{v}}{\mathbf{u}^{\mathrm{t}}\mathbf{u}} = \frac{\mathbf{u}\mathbf{u}^{\mathrm{t}}}{\mathbf{u}^{\mathrm{t}}\mathbf{u}}\mathbf{v} = P\mathbf{v}$$

Por tanto, la matriz de proyección P es igual a

$$P = \frac{\mathbf{u}\mathbf{u}^{\mathrm{t}}}{\mathbf{u}^{\mathrm{t}}\mathbf{u}}.$$

Ejemplo 8.1.2. Vamos a realizar un pequeño experimento cuando usamos la matriz de proyección para proyectar un número aleatorio de puntos en el plano sobre el vector $\mathbf{u} = (1, 1)^{\mathrm{t}}$.

En primer lugar definimos una matriz X de orden 2×100 con entradas aleatorias en el intervalo $[-1, 1]$

```
>> X=2*rand(2,100)-1;
```

Podemos dibujar estos puntos en el plano; la primera fila de la matriz X contiene las coordenadas x de los puntos aleatorios, y la segunda fila las coordenadas y. Una vez que hayas obtenido el dibujo no cierres la ventana.

```
>> x = X(1,:);
>> y = X(2,:);
>> plot(x,y,'b.')
```

Vamos a proyectar sobre el vector $\mathbf{u} = (1, 1)^{\mathrm{t}}$. En la figura anterior dibujamos la recta de dirección $\mathbf{u}$.

```
>> hold on
>> plot([1,-1],[1,-1],'y')
```

Ahora calculamos la matriz P de proyección sobre $\mathbf{u}$. Ahora, con la fórmula para calcular la matriz P proyectaremos sobre el vector $\mathbf{u} = (1, 1)^{\mathrm{t}}$.

```
>> u = [1;1]
>> P=(u*u')/dot(u,u)
```

Por último, vamos a aplicar a cada punto definido por la matriz X la matriz P, y dibujaremos el resultado. Si calculamos PX, la primera columna de PX contiene el resultado de aplicar la proyección sobre (x_1, y_1), la segunda columna el proyectado de (x_2, y_2), y así con todas. Realizamos la operación.

```
>> PX=P*X;
```

Tal como hemos explicado, las columnas de PX contienen la proyección de cada punto de la matriz X sobre el vector $\mathbf{u}$. Las coordenadas x de estos proyectados están en la primera fila de PX, y las coordenadas y en la segunda fila de PX.

```
>> Px=PX(1,:);
>> Py=PX(2,:);
```

Ahora podemos dibujar los puntos originales en azul y sus proyectados en rojo en la misma figura.

```
>> plot(Px,Py,'r.')
>> hold off
```

1.2. Proyección de un vector sobre un subespacio.

En este apartado vamos a recordar como se proyecta un vector $\mathbf{v} \in \mathbb{R}^m$ sobre un subespacio vectorial L de $\mathbb{R}^m$.

Sean L es subespacio vectorial generado por los vectores $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n \in \mathbb{R}^m$,y sea

$$A = \left(\begin{array}{c|c|c|c} \mathbf{u}_1 & \mathbf{u}_2 & \ldots & \mathbf{u}_n \end{array} \right) \in \mathcal{M}_{m\times n}(\mathbb{R}).$$

Nótese que $L = \mathrm{im}(A)$.

Si $\mathbf{v} \in L$, entonces no hay más nada que hacer; la proyección de $\mathbf{v}$ sobre L es el propio $\mathbf{v}$. Por ello, supongamos que $\mathbf{v}$ no es combinación lineal de los vectores $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$, y calculemos su proyección sobre $L = \mathrm{im}(A)$.

Para tener una idea gráfica de la situación, pensemos por un momento que L es un plano de $\mathbb{R}^3$. El vector $\mathbf{v}$ no está en ese plano. Esto lo representamos en la figura 1. En la figura 1 proyectamos el vector $\mathbf{v}$ sobre el vector $\mathbf{v}_1$, que sí está en el plano generado por los vectores $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$. Observemos, de nuevo, que el vector $\mathbf{v}_2 = \mathbf{v} - \mathbf{v}_1$ es ortogonal a L. En términos geométricos, lo que queremos es que el

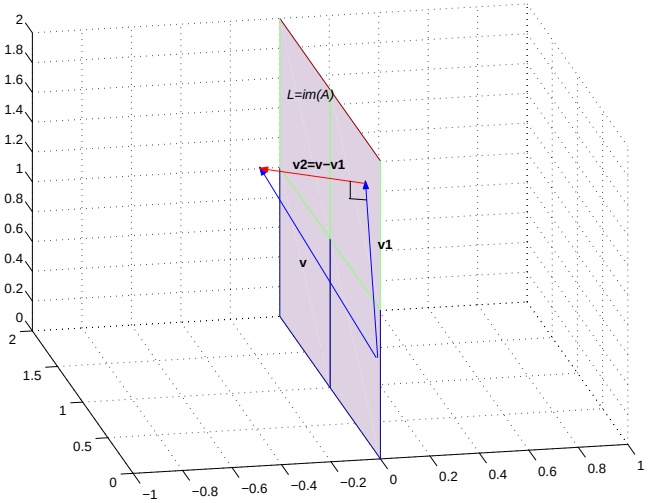


FIGURA 1. Proyección de $\mathbf{v}$ sobre $L = \mathrm{im}(A)$.

vector $\mathbf{v}_2$ sea ortogonal a cada vector de L. Esto se cumplirá si $\mathbf{v}_2$ es ortogonal a cada uno de los vectores $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$. Por tanto, las ecuaciones que nos quedan son

$$\mathbf{u}_1 \cdot \mathbf{v}_2 = \mathbf{u}_2 \cdot \mathbf{v}_2 = \ldots = \mathbf{u}_n \cdot \mathbf{v}_2 = 0$$

En notación matricial esto es

$$\mathbf{u}_1^{\mathrm{t}}\mathbf{v}_2 = \mathbf{u}_2^{\mathrm{t}}\mathbf{v}_2 = \ldots = \mathbf{u}_n^{\mathrm{t}}\mathbf{v}_2 = 0$$

Como $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$ son las columnas de la matriz A, entonces $\mathbf{u}_1^{\mathrm{t}}, \mathbf{u}_2^{\mathrm{t}}, \ldots, \mathbf{u}_n^{\mathrm{t}}$ son las filas de la matriz A^{t}, por lo que podemos expresar lo anterior como

$$\begin{pmatrix} \mathbf{u}_1^{\mathrm{t}} \\ \mathbf{u}_2^{\mathrm{t}} \\ \vdots \\ \mathbf{u}_n^{\mathrm{t}} \end{pmatrix} \mathbf{v}_2 = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Es claro que esto es lo mismo que

$$A^{\mathrm{t}}\mathbf{v}_2 = 0.$$

En la figura 1 vemos que el vector $\mathbf{v}_1$ tiene que estar en el $\mathrm{im}(A)$. Así, $\mathbf{v}_1$ se puede escribir como combinación lineal de los vectores $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$.

$$\begin{aligned} \mathbf{v}_1 &= w_1\mathbf{u}_1 + w_2\mathbf{u}_2 + \ldots + w_n\mathbf{u}_n \\ &= \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \ldots & \mathbf{u}_n \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix} \\ &= A\mathbf{w}. \end{aligned}$$

Entonces $\mathbf{v}_2 = \mathbf{v} - A\mathbf{w}$ y podemos escribir

$$A^{\mathrm{t}}(\mathbf{v} - A\mathbf{w}) = \mathbf{0}.$$

Si desarrollamos esta expresión, obtenemos

$$\begin{aligned} A^{\mathrm{t}}\mathbf{v} - A^{\mathrm{t}}A\mathbf{w} &= \mathbf{0} \\ A^{\mathrm{t}}A\mathbf{w} &= A^{\mathrm{t}}\mathbf{v}. \end{aligned}$$

Sea $(A^{\mathrm{t}}A)^+$ la inversa de Moore-Penrose de $A^{\mathrm{t}}A$. Usando las propiedades de la inversa generalizada (concretamente, que $(A^{\mathrm{t}}A)^+A^{\mathrm{t}} = A^+$ y que $A\,A^+A = A$) concluimos que

$$\mathbf{v}_1 = A\mathbf{w} = AA^+A\mathbf{w} = A(A^{\mathrm{t}}A)^+A^{\mathrm{t}})A\mathbf{w} = A(A^{\mathrm{t}}A)^+A^{\mathrm{t}}A\mathbf{w} = A(A^{\mathrm{t}}A)^+A^{\mathrm{t}}\mathbf{v}.$$

Esta expresión tiene la fórmula $\mathbf{v}_1 = P\mathbf{v}$, donde

$$\begin{aligned} P &= A(A^{\mathrm{t}}A)^{+}A^{\mathrm{t}} = AA^{+}(A^{+})^{\mathrm{t}}A^{\mathrm{t}} \\ &= AA^{+}(AA^{+})^{\mathrm{t}} = AA^{+}AA^{+} = AA^{+} \end{aligned}$$

es la matriz de proyección.

Vamos a hacer un ejemplo similar al del apartado anterior, pero ahora en tres dimensiones.

Ejemplo 8.1.3. En primer lugar, generamos un conjunto de puntos en el espacio.

```
>> X=3*rand(3,100)-1;
```

Extraemos las coordenadas x, y y z.

```
>> x=X(1,:);
>> y=X(2,:);
>> z=X(3,:);
```

Dibujamos estos puntos en el espacio, y no cerramos la figura

```
>> plot3(x,y,z,'b.')
>> box on
>> grid on
>> hold on
```

Vamos a proyectar los puntos definidos por X sobre el subespacio vectorial de $\mathbb{R}^3$ generado por la columnas de

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

Introducimos en primer lugar la matriz A.

```
>> u1=[1;1;0];u2=[0;1;1];u3=[1;0;-1];
>> A=[u1,u2,u3];
```

Ahora calculamos la matriz de proyección. El comando `pinv` de `MATLAB` calcula la inversa de Moore-Penrose.

```
>> P=A*pinv(A)
```

Ahora, si multiplicamos la matriz X por la matriz P proyectaremos cada punto sobre el espacio de columnas de A.

```
>> PX=P*X;
```

Tomamos las componentes de cada punto.

```
>> Px=PX(1,:);
>> Py=PX(2,:);
>> Pz=PX(3,:);
```

Ya podemos dibujar los puntos originales y sus proyecciones.

```
>> plot3(Px,Py,Pz,'r.')
```

La pregunta es si realmente hemos conseguido lo que buscábamos. Es difícil de decir a partir de la figura obtenida. Sin embargo, podemos hacer dos cosas para convencernos de que la proyección se ha efectuado sobre el subespacio vectorial generado por los vectores $\mathbf{u}_1, \mathbf{u}_2$ y $\mathbf{u}_3$. Primero dibujemos los vectores $\mathbf{u}_1 = (1,1,0)^{\mathrm{t}}, \mathbf{u}_2 = (0,1,1)^{\mathrm{t}}$ y $\mathbf{u}_3 = (1,0,-1)$ sobre la figura con los siguientes comandos.

```
>> line([0,1],[0,1],[0,0],'linewidth',2,'color','k')
>> line([0,0],[0,1],[0,1],'linewidth',2,'color','k')
>> line([0,1],[0,0],[0,-1],'linewidth',2,'color','k')
>> hold off
```

El comando `line` permite añadir más gráficos sobre el dibujo. Los vectores $\mathbf{u}_1, \mathbf{u}_2$ y $\mathbf{u}_3$ aparecen en la nueva figura sobre el plano $\mathrm{im}(A)$.

Si ahora pulsamos el icono de rotación en la pantalla de la figura, podemos experimentar con diferentes puntos de vista. En la figura obtenida, usamos el ratón para colocar la figura con acimut 29 y elevación -40. Esto se puede hacer sin el ratón mediante el comando `view([29,-40])`. Vemos que los vectores $\mathbf{u}_1, \mathbf{u}_2$ y $\mathbf{u}_3$ se ocultan por la nube de puntos proyectados sobre el plano.

2. Soluciones aproximadas mínimo cuadráticas de sistemas de ecuaciones lineales

En algunas situaciones en las nos encontramos con sistema de ecuaciones $A\mathbf{x} = \mathbf{b}$, puede ser conveniente hallar un vector $\hat{\mathbf{x}}$ que este "cerca de ser solución del sistema"; entendiendo por esto que $A\hat{\mathbf{x}} - \mathbf{b}$ sea próximo a cero. Una de las formas más comunes de medir la proximidad de $A\hat{\mathbf{x}} - \mathbf{b}$ a cero es mediante el cálculo de la suma de los cuadrados de las componentes de $A\hat{\mathbf{x}} - \mathbf{b}$. Cualquier vector que minimice esta suma de cuadrados se llama **solución aproximada mínimo cuadrática**.

Ejemplo 8.2.1. Supongamos que queremos calcular la solución del sistema de ecuaciones lineales

$$\begin{aligned} m \cdot 0 + c &= 6 \\ m \cdot 1 + c &= 0 \\ m \cdot 2 + c &= 0 \end{aligned}$$

Este sistema está sobre-determinado: hay más ecuaciones que incógnitas. Es más, es incompatible.

```
>> M=[0,1,6;1,1,0;2,1,0]
>> R=rref(M)
```

La última fila de R representa la ecuación $0 \cdot m + 0 \cdot c = 1$, que no tiene solución.

Como es habitual el sistema se puede escribir en la forma

$$\begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} m \\ c \end{pmatrix} = \begin{pmatrix} 6 \\ 0 \\ 0 \end{pmatrix},$$

o bien $A\mathbf{x} = \mathbf{b}$, donde

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 2 & 1 \end{pmatrix}, \mathbf{x} = \begin{pmatrix} m \\ c \end{pmatrix}, \mathbf{v} = \begin{pmatrix} 6 \\ 0 \\ 0 \end{pmatrix}.$$

Como el sistema no tiene solución, $\mathbf{b}$ no puede escribirse como combinación lineal de las columnas de A; en otras palabras, $\mathbf{b} \notin \mathrm{im}(A)$.

Teniendo en cuenta que una recta en el plano es de la forma $y = mx + c$, podemos reenunciar nuestro problema en términos geométricos como el de calcular una recta que se ajuste lo mejor posible, en sentido mínimo cuadrático, a los datos de la siguiente

tabla:

x	0	1	2
y	6	0	0

Si dibujamos los datos de la tabla como puntos en un plano

```
>> plot([0,1,2],[6,0,0],'s')
>> axis([-1,7,-1,7])
>> grid on
>> hold on
```

se ve claramente que los puntos no están alineados, por lo que no es posible dibujar una recta a través de ellos como ya sabíamos. De modo que tendremos que contentarnos con hallar una solución aproximada.

Vamos a calcular la solución aproximada mínimo cuadrática de nuestro sistema. Para ello, en primer lugar, calculamos la proyección ortogonal de $\mathbf{b}$ el espacio vectorial que generan las columnas de A tal y como hicimos en la sección anterior.

```
>> A = [0,1;1,1;2,1]
>> b = [6;0;0]
>> P = A*pinv(A)
>> bb = P*b
```

Así, obtenemos un vector $\mathbf{b}'$ que sí está en $\mathrm{im}(A)$, de hecho, es el vector de $\mathrm{im}(A)$ tal que $\mathrm{d}(\mathbf{b}', \mathbf{b}) = \|\mathbf{b}'-\mathbf{b}\|$ es mínima. De este modo, garantizamos que el sistema $A\hat{\mathbf{x}} = \mathbf{b}'$ tiene solución y que la suma al cuadrado de las componentes de $\mathbf{b}' - \mathbf{b} = A\hat{\mathbf{x}} - \mathbf{b}$, esto es, su norma al cuadrado es mínima.

```
>> Abb = [A,bb]
>> rref(Abb)
>> xgorro = A\bb
```

Nota.- Aunque sabemos que el sistema $A\mathbf{x} = \mathbf{b}$ es incompatible, observemos la salida de la siguiente sentencia.

```
>> A\b
```

Es la solución $\hat{\mathbf{x}}$ que habíamos obtenido. Esto ocurre porque el comando \ calcula la solución mínimo cuadrática del sistema $A\mathbf{x} = \mathbf{b}$. Teclea `help mldivide` para una descripción más completa.

En términos geométricos la solución aproximada mínimo cuadrática, $\hat{\mathbf{x}}$, obtenida nos da la ecuación de la recta que andábamos buscando. Como $m = -3$ y $b = 5$, la ecuación de la recta que mejor se ajusta es $y = -3x + 5$.

```
>> x=linspace(-1,2)
>> plot(x,-3*x+5,'r')
>> hold off
```

Es interesante examinar el error cometido al aproximar los datos con la recta de mejor ajuste. Los puntos originales eran $(0, 6), (1, 0)$ y $(2, 0)$, y sus proyecciones ortogonales sobre la recta son $(0, 5), (1, 2)$ y $(2, -1)$, respectivamente. Así, tenemos que n $x = 0$, el valor del dato es $y = 6$, y el punto sobre la recta correspondientes es $\hat{y} = 5$; entonces, el error cometido es $y - \hat{y} = 6 - 5 = 1$. Análogamente, en $x = 1$ tenemos que $y - \hat{y} = 0 - 2 = -2$, y en $x = 2$ obtenemos $y - \hat{y} = 0 - (-1) = 1$. Realmente estos errores se pueden calcular directamente con el vector $\mathbf{e} = \mathbf{b} - \mathbf{b}'$.

```
>> e=b-bb
```

Por tanto, el error total cometido es

```
>> norm(e)^2
```

2.1. Otros ejemplos con MATLAB.

Ejemplo 8.2.2. Supongamos que en un experimento físico, colgamos unas masas de un muelle, y medimos la distancia que el muelle elonga desde su punto de equilibrio para cada masa. Los datos los tenemos en la siguiente tabla.

m	10	20	30	40	50	60
d	1,4	2,8	3,6	5,0	6,4	7,2

Vamos a usar MATLAB para calcular la curva más simple que mejor ajusta a los datos de la tabla anterior.

En primer lugar, introducimos los datos en MATLAB y los dibujamos.

```
>> clear all
>> close all
>> m=(10:10:60)';
>> d=[1.4, 2.8, 3.6, 5.0, 6.4, 7.2]';
>> plot(m,d,'*')
>> hold on
```

Usamos el operador de transposición para formar vectores columna. Se ve en la figura que existe una tendencia lineal. En concreto,

```
>> corrcoef(m,d)
```

indica que el coeficiente de correlación es 0,9969.

Vamos a ajustar los datos con una recta de la forma $d = a + b\,m$. Primero, sustituimos cada punto en la ecuación:

$$\begin{aligned} 1{,}4 &= a + b \cdot 10 \\ 2{,}8 &= a + b \cdot 20 \\ 3{,}6 &= a + b \cdot 30 \\ 5{,}0 &= a + b \cdot 40 \\ 6{,}4 &= a + b \cdot 50 \\ 7{,}2 &= a + b \cdot 60 \end{aligned}$$

y escribimos el sistema matricialmente.

$$\begin{pmatrix} 1 & 10 \\ 1 & 20 \\ 1 & 30 \\ 1 & 40 \\ 1 & 50 \\ 1 & 60 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 1{,}4 \\ 2{,}8 \\ 3{,}6 \\ 5{,}0 \\ 6{,}4 \\ 7{,}2 \end{pmatrix},$$

$$A\mathbf{x} = \mathbf{d}$$

El vector $\mathbf{d}$ ya lo tenemos definido en MATLAB. La segunda matriz de A contiene los datos de masa almacenados en el vector $\mathbf{m}$.

```
>> A=[ones(size(m)),m]
```

Luego ya sólo nos queda calcular la solución aproximada mínimo cuadrática del sistema $A\mathbf{x} = \mathbf{b}$ tal y como hicimos en el ejemplo anterior

```
>> P = A*pinv(A)
>> xgorro= A\P*d
```

Nota.- Notemos de nuevo que

```
>> A\d
```

nos da la solución correcta.

Entonces $a = 0{,}2800$ y $b = 0{,}1177$. Con estos valores vamos a dibujar la recta de mejor ajuste en nuestra figura.

```
>> ygorro=xgorro(1)+xgorro(2)*m;
>> plot(m,ygorro,'r')
>> hold off
```

Ejemplo 8.2.3. En otro experimento, un cohete de juguete es lanzado al aire. La altura del cohete a instantes determinados aparece en la tabla siguiente.

t	5	10	15	20	25	30
s	722	1073	1178	1117	781	102

Debemos examinar los datos y decidir un modelo apropiado para su ajuste por mínimos cuadrados.

Empecemos introduciendo los datos en vectores columna **t** y **s**.

```
>> clear all
>> close all
>> t=(5:5:30)';
>> s=[722, 1073, 1178, 1117, 781, 102]';
```

Podemos dibujar nuestros datos como sigue:

```
>> plot(t,s,'bs','MarkerFaceColor','b')
>> hold on
```

Aparentemente los datos forman una parábola. Intentemos entonces ajustar los datos a una ecuación de la forma $s = a + bt + ct^2$. Sustituimos los datos de la tabla en la ecuación $s = a + bt + ct^2$.

$$\begin{aligned} 722 &= a + b \cdot 5 + c \cdot (5)^2 \\ 1073 &= a + b \cdot 10 + c \cdot (10)^2 \\ 1178 &= a + b \cdot 15 + c \cdot (15)^2 \\ 1117 &= a + b \cdot 20 + c \cdot (20)^2 \\ 781 &= a + b \cdot 25 + c \cdot (25)^2 \\ 102 &= a + b \cdot 30 + c \cdot (30)^2 \end{aligned}$$

La expresión matricial del sistema es de la forma

$$\begin{pmatrix} 1 & 5 & 5^2 \\ 1 & 10 & 10^2 \\ 1 & 15 & 15^2 \\ 1 & 20 & 20^2 \\ 1 & 25 & 25^2 \\ 1 & 30 & 30^2 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 722 \\ 1073 \\ 1178 \\ 1117 \\ 781 \\ 102 \end{pmatrix},$$

$$A\mathbf{x} = \mathbf{s}.$$

Podemos introducir en MATLAB los valores de A de una forma sencilla.

```
>> A=[ones(size(t)),t,t.^2]
```

Vamos entonces a calcular la solución aproximada mínimo cuadrática del sistema $A\mathbf{x} = \mathbf{s}$.

```
>> xgorro = A\s
```

Entonces $a = 80{,}2000, b = 149{,}7814$ y $c = -4{,}9386$. Con estos coeficientes vamos a pintar la parábola de mejor ajuste. Además, queremos hacer dos estimaciones. Por un lado, vamos a averiguar la altura inicial del cohete, y por otro queremos saber en qué momento volvió a tierra. Por ello, extendemos el intervalo de t para que nos aparezcan esos datos.

```
>> tt=linspace(0,35);
>> sgorro=xgorro(1)+xgorro(2)*tt+xgorro(3)*tt.^2;
>> plot(tt,sgorro)
```

```
>> grid
>> hold off
```

El vector de errores es igual a $\mathbf{e} = \mathbf{s} - A\hat{\mathbf{x}}$, y podemos calcular su norma.

```
>> p=A*xgorro;
>> e=s-p
>> norm(e)
```

Finalmente, podemos preguntarnos por qué no realizamos, por ejemplo, un ajuste con una cúbica. La ecuación buscada es $s = a + bt + ct^2 + dt^3$ y, en ese caso, el sistema queda de la forma

$$\begin{pmatrix} 1 & 5 & 5^2 & 5^3 \\ 1 & 10 & 10^2 & 10^3 \\ 1 & 15 & 15^2 & 15^3 \\ 1 & 20 & 20^2 & 20^3 \\ 1 & 25 & 25^2 & 25^3 \\ 1 & 30 & 30^2 & 30^3 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} = \begin{pmatrix} 722 \\ 1073 \\ 1178 \\ 1117 \\ 781 \\ 102 \end{pmatrix},$$

$$B\mathbf{x} = \mathbf{s}.$$

Veamos qué resulta siguiendo los pasos anteriores.

```
>> B=[ones(size(t)),t,t.^2,t.^3]
>> xgorroB=B\s
```

Observamos que el coeficiente d es de orden de 10^{-2}, lo que nos dice que la aportación de término en t^3 es pequeña. Si calculamos el error cometido, debe salir más pequeño que en el ajuste por una parábola.

```
>> eB=s-B*xgorroB;
>> norm(eB)
```

Por ello, no se trata de encontrar el modelo que dé el menor error, sino el que sea más sencillo y nos permita construir un modelo teórico.

Ejercicios de la práctica 8

Ejercicio 1. Calcular la matriz P que proyecta todos los puntos del plano sobre el subespacio generado por el vector $\mathbf{u} = (1, 2)^{\mathrm{t}}$.

Ejercicio 2. Calcular la matriz P que proyecta todos los puntos de $\mathbb{R}^3$ sobre el subespacio generado por

(a) $\mathbf{u} = (1, 1, 1)^{\mathrm{t}}$.
(b) $\mathbf{u}_1 = (1, 0, 0)^{\mathrm{t}}$ y $\mathbf{u}_2 = (1, 1, 1)^{\mathrm{t}}$.

Ejercicio 3. Calcule la recta de mejor ajuste a los datos de la siguiente tabla:

x	5	10	15	20	25	30
y	28	39	48	65	72	82

Ejercicio 4. Calcule la parábola de mejor ajuste a los datos de siguiente tabla:

x	2	6	10	14	18	22
y	286	589	749	781	563	282

Ejercicio 5. Si cada ecuación en un sistema es lineal, entonces hemos visto que el Álgebra Lineal nos permite encontrar el ajuste por mínimos cuadrados. En principio, si intentamos calcular un ajuste de una ecuación exponencial $y = a\mathrm{e}^{bx}$ a los datos de la tabla siguiente parece que no seremos capaces.

x	1	2	3	4	5	6
y	128	149	214	269	336	434

Sin embargo, si tomamos logaritmos en ambos lados la ecuación queda lineal.

$$\begin{aligned} y &= a\mathrm{e}^{bx} \\ \log(y) &= \log(a) + bx \end{aligned}$$

1. Prepare un gráfico que muestre la relación lineal entre $\log(y)$ y x.
2. Calcule la recta de ajuste de los datos transformados del apartado anterior.
3. Usando el apartado anterior, calcule la ecuación exponencial $y = a\mathrm{e}^{bx}$ que mejor ajusta a los datos originales.

Ejercicio 6. Calcule una función de la forma $y = ax^b$ que ajuste los datos de la siguiente tabla:

x	1	2	3	4	5	6
y	117	385	920	1608	2518	3611

PRÁCTICA 9

Calculando inversas generalizadas

En esta práctica veremos algunos métodos computacionales para las inversas generalizadas.

Pre-requisitos: Inversas generalizadas. Forma reducida

1. La fórmula de Greville

T.N.E. Greville[1] obtuvo en 1960 la siguiente expresión de la inversa de Moore-Penrose de una matriz $A \in \mathcal{M}_{m\times n}(\mathbb{R})$ particionada en la forma $(B|\mathbf{b})$, donde B es una matriz de orden $m \times (n-1)$ y $\mathbf{b}$ es un vector columna no nulo con m filas:

$$A^+ = \begin{pmatrix} B^+ - \mathbf{d}\,\mathbf{c}^+ \\ \hline \mathbf{c}^+ \end{pmatrix},$$

donde $\mathbf{d} = B^+\mathbf{b}$ y

$$\mathbf{c} = \begin{cases} \mathbf{b} - B\mathbf{d} & \text{si } \mathbf{b} \neq B\mathbf{d} \\ \dfrac{1 + \|\mathbf{d}\|_2^2}{\|(B^+)^{\mathrm{t}}\mathbf{d}\|_2^2}(B^+)^{\mathrm{t}}\mathbf{d} & \text{en otro caso} \end{cases}$$

El lector interesado puede encontrar una demostración de la fórmula de Greville en [Udwadia, F.E.; Kabala, R.E. *An alternative proof of the Greville formula.* J. Optim. Theory Appl. **94** (1997), no. 1, 23–28.].

Comprobemos con un ejemplo que la fórmula funciona correctamente. Consideremos la matriz

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

e introduzcámosla en `MATLAB`.

```
>> A = [1,1,2,3;1,-1,0,1;1,1,2,3]
```

[1] Greville, T.N.E. *Some applications of the pseudoinverse of a matrix.* SIAM Rev. **2**, 1960, 15–22.

A continuación dividimos nuestra matriz A en dos bloques B y : el primero formado por las tres primeras columnas de A y el segundo por la última columna de A.

```
>> B = A(1:3,1:3)
>> b = A(1:3,4)
```

Ahora calculamos los vectores **d** y **c**. Recuérdese que el comando `pinv` de `MATLAB` calcula la inversa de Moore-Penrose.

```
>> d = pinv(B)*b
>> c = b - B*d
>> %% Observamos que b = B*d, por tanto
>> c = (1+norm(d)^2)/(norm(pinv(B)'*d)^2)*pinv(B)'*d
```

Y finalmente

```
>> cc = pinv(c)
>> AA = [pinv(B) - d*cc; cc]
```

Obsérvese que, `AA` coincide esencialmente con `pinv(A)`.

```
>> pinv(A)
```

Así vista, la fórmula de Geville no parece que dé un método para calcular la inversa de Moore-Penrose de A, ya que necesitamos conocer la inversa de Moore-Penrose de una submatriz de A. La clave está en usar la fórmula de Greville recursivamente como explicamos a continuación.

Consideremos la matriz $A \in \mathcal{M}_{m\times n}(\mathbb{R})$, denotemos $\mathbf{a}_j$ a la columna j-ésima de A y definamos $A_j = (\mathbf{a}_1|\ldots|\mathbf{a}_j)$, de tal forma que $A_j \in \mathcal{M}_{m\times j}(\mathbb{R})$ es la submatriz de A formada por sus j primeras columnas. La fórmula de Greville nos dice que si conocemos la inversa de Moore-Penrose de A_{j-1} podemos calcular la inversa de Moore-Penrose de A_j. Por consiguiente, la inversa de Moore-Penrose de A se puede calcular hallando sucesivamente las inversas generalizadas de $A_1^+ = \mathbf{a}_1^+,\ A_2^+, A_3^+, \ldots, A_n^+ = A$.

Teniendo además en cuenta que la inversa de Moore-Penrose de $\mathbf{a}_1^+$ no es más que

$$\mathbf{a}_1^+ = \mathbf{a}_1^{\mathrm{t}}/(\mathbf{a}_1^{\mathrm{t}}\mathbf{a}_1);$$

podemos afirmar que tenemos un algoritmo para cálculo del inversa de Moore-Penrose de A, mediante el uso recursivo de la fórmula de Greville.

Pongamos en práctica nuestro algoritmo con la matriz A del ejemplo anterior. Si no hemos borrado el valor de la variable `A` no tendremos que volver a introducirla, esto lo podemos saber viendo nuestro `Workspace`, con el comando `who` o simplemente escribiendo

```
>> A
```

Si la variable `A` no está definida, obtendremos el mensaje `??? Undefined function or variable 'A'` y tendremos que volver a introducirla.

Consideremos ahora la primera columna de `A`, llamémosla `A1` y calculemos su inversa de Moore-Penrose a la que llamaremos `AA1`.

```
>> A1 = A(1:3,1)
>> AA1 = a1'/(a1'*a1)
```

Calculemos a continuación la inversa de Moore-Penrose de $A_2 = (\mathbf{a}_1|\mathbf{a}_2) = (A_1|\mathbf{a}_2)$ usando la fórmula de Greville.

```
>> a2 = A(1:3,2)
>> A2 = [A1,a2]
>> d2 = AA1*a2
>> c2 = a2 - A1*d2
```

Como $\mathbf{a}_2 \neq A_1\mathbf{d}_2$, se tiene que

```
>> cc2 = c2'/(c2'*c2)
>> AA2 = [AA1-d2*cc2;cc2]
```

De modo que la inversa de Moore-Penrose de A_2 es

$$A_2^+ = \begin{pmatrix} 1/4 & 1/2 & 1/4 \\ 1/4 & -1/2 & 1/4 \end{pmatrix}.$$

La inversa de Moore-Penrose de $A_3 = (A_2|\mathbf{a}_3)$ se puede calcular ahora usando A_2^+

```
>> a3 = A(1:3,3)
>> A3 = [A2,a3]
>> d3 = AA2*a3
>> c3 = a3 - A2*d3
```

Como, en este caso, $\mathbf{a}_3 = A_2\mathbf{d}_3$ tenemos que definir $\mathbf{c}_3$ correctamente (siguiendo la fórmula de Greville)

```
>> c3 = (1+norm(d3)^2)/(norm(AA2'*d3)^2)*AA2'*d3
```

y por lo tanto

```
>> cc3 = c3'/(c3'*c3)
>> AA3 = [AA2-d3*cc3;cc3]
```

Luego la inversa generalizada de A_3 es

$$A_2^+ = \begin{pmatrix} 1/12 & 1/2 & 1/12 \\ 1/12 & -1/2 & 1/12 \\ 1/6 & 0 & 1/6 \end{pmatrix}.$$

Finalmente, para obtener la inversa de Moore-Penrose de $A = A_4 = (A_3|\mathbf{a}_4)$, calculamos

```
>> a4 = A(1:3,4)
>> A4 = [A3,a4]
>> d4 = AA3*a4
>> c4 = a4 - A3*d4
```

Al igual que antes, tenemos que definir correctamente el valor de $\mathbf{c}_4$, pues $\mathbf{a}_4 = A_3\mathbf{d}_4$.

```
>> c4 = (1+norm(d4)^2)/(norm(AA3'*d4)^2)*AA3'*d4
```

y para terminar

```
>> cc4 = c4'/(c4'*c4)
>> AA4 = [AA3-d4*cc4;cc4]
```

Por lo que podemos concluir que la inversa de Moore-Penrose de A es

$$A^+ = \begin{pmatrix} 0 & 1/3 & 0 \\ 1/12 & -1/2 & 1/12 \\ 1/12 & -1/6 & 1/12 \\ 1/12 & 1/6 & 1/12 \end{pmatrix}$$

Nota 9.1.1. Como se indicó en la introducción, este método basado en la fórmula de Greville no se suele utilizar para calcular la inversa de Moore-Penrose, la principal razón es la propagación de errores de redondeo. Lo general es utilizar la descomposición en valores singulares (véase demostración del teorema VI.2.2). Así es como funciona realmente el comando `pinv` de `MATLAB`, usando a su vez el comando `svd` para calcula la descomposición en valores singulares. Básicamente el comando `svd` funciona como describimos en el siguiente ejemplo

En primer lugar definimos una matriz aleatoriamente con entradas entre -10 y 10 de orden también aleatorio $m \times n,\ 1 \leq m, n \leq 11$.

```
>> m = round(10*rand+1);
>> n = round(10*rand+1);
>> A = 20*rand(m,n)-10;
```

A continuación calculamos su descomposición en valores singulares $A = PDQ^{\mathrm{t}}$

```
>> [Pt,D,Q] = svd(A);
```

y finalmente la inversa de Moore-Penrose de A usando la fórmula $A^+ = QD'P^{\mathrm{t}}$, donde D' se obtiene al sustituir su submatriz formada por la r primeras filas y columnas por la inversa de la submatriz de las r primeras filas y columnas de D, siendo r el rango de A.

```
>> DD = zeros(n,m);
>> r = rank(A)
>> DD(1:r,1:r) = inv(D(1:r,1:r))
>> AA = Q*DD*Pt'
```

Podemos comprobar que el resultado obtenido es esencialmente el mismo que el que se obtiene con el comando `pinv` de `MATLAB`

```
>> pinv(A)
```

2. Cálculo de inversas generalizadas

Un método común para calcular inversas generalizadas, esto es, {1}-inversas, de un matriz dada se basa en el cálculo de la forma reducida.

2.1. Inversas generalizadas de matrices cuadradas.

Sabemos que dada una matriz $A \in \mathcal{M}_n(\mathbb{R})$ de rango r, existen P y $Q \in \mathcal{M}_n(\mathbb{R})$ invertibles tales que

$$P^{-1}AQ = R = \left(\begin{array}{c|c} I_r & 0 \\ \hline 0 & 0 \end{array}\right).$$

Es claro que la matriz R es idempotente, esto es, $R^2 = R$. Por consiguiente,

$$P^{-1}A(QP^{-1})AQ = (P^{-1}AQ)(P^{-1}AQ) = R^2 = R = P^{-1}AQ.$$

Entonces, multiplicando a izquierda por P y a la derecha por Q^{-1} en la igualdad anterior, obtenemos que

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

Es decir, QP^{-1} es una inversa generalizada de A.

Veamos con un ejemplo que el método propuesto funciona.

Sea

$$A = \begin{pmatrix} 2 & 2 & 4 \\ 4 & -2 & 2 \\ 2 & -4 & -2 \end{pmatrix}.$$

Usando `MATLAB` podemos calcular la forma reducida R de A y matrices de paso P y Q invertibles tales $P^{-1}AQ = R$ (véase la práctica 3).

```
>> A = [2,2,4;4,-2,2;2,-4,-2]
>> F = rref(A)
>> AI = [A,eye(3)]
>> FAI = rref(AI)
>> invP = FAI(:,4:6) %Inversa de P
>> E = F'
>> EI = [E,eye(3)]
>> FEI = rref(EI)
>> Q1 = FEI(:,4:6)
>> Q = Q1'
>> R = invP*A*Q
```

Así obtenemos que

$$P^{-1} = \begin{pmatrix} 0 & 1/3 & -1/6 \\ 0 & 1/6 & -1/3 \\ 1 & -1 & 1 \end{pmatrix} \quad \text{y} \quad Q = \begin{pmatrix} 0 & 0 & 1 \\ -1 & 1 & 1 \\ 1 & 0 & -1 \end{pmatrix}$$

son matrices invertibles tales que

$$P^{-1}AQ = R = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Por consiguiente, una inversa generalizada de A es

$$A^- = QP^{-1} = \begin{pmatrix} 1 & -1 & 1 \\ 1 & -7/6 & 5/6 \\ -1 & 4/3 & -7/6 \end{pmatrix}.$$

En efecto,

```
>> B = Q*invP
>> A*B*A
```

2.2. Inversas generalizadas, caso general.

Sea $A \in \mathcal{M}_{m\times n}(\mathbb{R})$, donde $m < n$. Definimos la matriz A_* como sigue

$$A_* = \left(\frac{A}{0_{(n-m)\times n}}\right)$$

donde $0_{(n-m)\times n}$ es una matriz de ceros de orden $(n-m)\times n$. Es claro que si

$$P^{-1}A_*Q = R$$

es la forma reducida de A y $P^{-1} = (P_1|P_2)$ es una partición por bloques de P^{-1} compatible con la partición de A^*, entonces $Q\,P_1$ es una inversa generalizada de A.

En efecto,

$$\begin{aligned} A_*QP^{-1}A_* &= \left(\frac{A}{0_{(n-m)\times n}}\right) Q(P_1|P_2) \left(\frac{A}{0_{(n-m)\times n}}\right) \\ &= \left(\frac{A}{0_{(n-m)\times n}}\right) (QP_1|QP_2) \left(\frac{A}{0_{(n-m)\times n}}\right) \\ &= \left(\begin{array}{c|c} AQP_1 & AQP_2 \\ \hline 0_{(n-m)\times m} & 0_{(n-m)\times(n-m)} \end{array}\right) \left(\frac{A}{0_{(n-m)\times n}}\right) \\ &= \left(\frac{AQP_1A}{0_{(n-m)\times n}}\right) \end{aligned}$$

Igualando esta identidad a A_*, obtenemos que $AQP_1A = A$.

Una expresión análoga se obtiene cuando $m > n$, ampliando A a la derecha con ceros hasta hacerla cuadrada. Veamos este caso en un ejemplo.

Supongamos que queremos calcular una inversa generalizada de la matriz

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

Consecuentemente consideramos la matriz ampliada

$$A_* = (A|\mathbf{0}) = \begin{pmatrix} 1 & 1 & 2 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 2 & 0 \\ 2 & 0 & 2 & 0 \end{pmatrix}.$$

Procediendo como antes obtenemos matrices invertibles P^{-1} y Q tales que $P^{-1}A_*Q$ es la forma reducida de A_*.

$$P^{-1} = \begin{pmatrix} 0 & 0 & 0 & 1/2 \\ 0 & 0 & 1 & -1/2 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1/2 \end{pmatrix} \quad \text{y} \quad Q = \begin{pmatrix} 0 & 0 & 1 & 0 \\ -1 & 1 & 1 & 0 \\ 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Particionando la matriz Q como sigue

$$Q = \left(\frac{Q_1}{Q_2}\right) = \left(\begin{array}{cccc} 0 & 0 & 1 & 0 \\ -1 & 1 & 1 & 0 \\ 1 & 0 & -1 & 0 \\ \hline 0 & 0 & 0 & 1 \end{array}\right)$$

Encontramos que una inversa generalizada de A es Q_1P^{-1}.

```
>> A = [1, 1, 2; 1, 0, 1; 1, 1, 2; 2, 0, 2]
>> invP = [0, 0, 0, 1/2; 0, 0, 1, -1/2; 1, 0, -1, 0; 0, 1, 0, -1/2]
>> Q = [0, 0, 1, 0; -1, 1, 1, 0; 1, 0, -1, 0]
>> Q1 = Q(1:3,1:4)
>> B = Q1*invP
>> A*B*A
```

Obsérvese que es lo mismo considerar las primeras $m-n$ filas de Q y realizar el producto con P^{-1} que tomar las primeras $m-n$ del filas de producto QP^{-1}.

```
>> C = Q*invP
>> D = C(1:3,1:4)
```

3. Cálculo de inversas mínimo cuadráticas

Según lo estudiado en clase de teoría, se puede calcular una inversa mínimo cuadrática de una matriz $A \in \mathcal{M}_{m\times n}(\mathbb{R})$ calculando primero una inversa generalizada de $A^{\mathrm{t}}A$ y usando la igualdad $A^{\square} = (A^{\mathrm{t}}A)^{-}A^{\mathrm{t}}$ (véase la proposición VI.3.8). Ilustremos con un ejemplo este procedimiento.

Consideremos la matriz del ejemplo anterior

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

```
>> A = [1, 1, 2; 1, 0, 1; 1, 1, 2; 2, 0, 2]
```

Definamos $A^{\mathrm{t}}A$, llamémosla B y calculemos una de sus inversas generalizadas.

```
>> B = A'*A
>> F = rref(B)
>> BI = [B,eye(3)]
>> FBI = rref(BI)
>> invP = FBI(:,4:6) %Inversa de P
>> E = F'
>> EI = [E,eye(3)]
>> FEI = rref(EI)
```

```
>> Q1 = FEI(:,4:6)
>> Q = Q1'
>> BB = Q*invP
```

Usando ahora la expresión $A^{\square} = (A^{\mathrm{t}}A)^{-}A^{\mathrm{t}}$

```
>> AA = BB*A'
```

se obtiene que una inversa mínimo cuadrática de A es

$$A^{\square} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1/2 & -2/5 & 1/2 & -4/5 \\ 0 & 1/5 & 0 & 2/5 \end{pmatrix}$$

Ejercicios de la práctica 9

Ejercicio 1. Usar el método recursivo basado en la fórmula de Greville para calcular una inversa de Moore-Penrose de la matriz

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

Ejercicio 2. Hallar una inversa generalizada de la matriz A del ejercicio anterior calculando su forma reducida.

Ejercicio 3. Hallar una inversa generalizada de la matriz

$$A = \begin{pmatrix} 1 & -1 & -2 & 1 \\ -2 & 4 & 3 & -2 \\ 1 & 1 & -3 & 1 \end{pmatrix}$$

calculando su forma reducida.

Ejercicio 4. Hallar una inversa mínimo cuadrática de la matriz A del ejercicio anterior distinta de la inversa de Moore-Penrose.

PRÁCTICA 10

Número de condición de una matriz y MATLAB

En esta práctica se mostrará la interpretación gráfica que tiene la resolución de un sistema de ecuaciones en relación con el número de condición de la matriz del sistema. Se expondrán también las funciones que incorpora MATLAB para calcular la norma de un vector y el número de condición de una matriz.

Pre-requisitos: resolución de sistemas de ecuaciones lineales, normas matriciales.

1. Número de condición de una matriz y MATLAB

Consideremos el sistema de ecuaciones $A\mathbf{x} = \mathbf{b}$ donde

$$A = \begin{pmatrix} 0{,}835 & 0{,}667 \\ 0{,}333 & 0{,}266 \end{pmatrix}, \mathbf{b} = \begin{pmatrix} 0{,}168 \\ 0{,}067 \end{pmatrix} \tag{10.1.1}$$

La solución del sistema se puede calcular como sigue[1]

```
>> A=[0.835,0.667;0.333,0.266]
>> b=[0.168;0.067]
>> sol1=A\b
```

Desde el punto de vista geométrico, se trata de dos rectas en $\mathbb{R}^2$, y la solución es el punto de corte. Para obtener la representación gráfica en MATLAB deben ser pasadas a paramétricas. Si r_1 es la recta representada por la primera ecuación y r_2 la representada por la segunda, se tiene que

$$r_1 : \begin{cases} x = 1 + 0{,}667\,t \\ y = -1 - 0{,}835\,t \end{cases} \quad \text{y que} \quad r_2 : \begin{cases} x = 1 + 0{,}266\,t \\ y = -1 - 0{,}333\,t \end{cases}$$

Teclea en MATLAB la siguiente secuencia de comandos para representar ambas rectas.

```
>> close all
>> t=linspace(-10,10);
>> x1=1+0.667*t;
```

[1] Téngase que la matriz A es invertible, por lo que el sistema de ecuaciones tiene solución única. En otro caso, si A no fuese invertible (o incluso si no fuese cuadrada) la orden `A\b` adquiere otro significado.

```
>> y1=-1-0.835*t;
>> x2=1+0.266*t;
>> y2=-1-0.333*t;
>> plot(x1,y1,'--r',x2,y2,':g')
>> axis([-2,2,-2,2])
>> grid
>> line(1,-1,'Marker','.','MarkerSize',16,'color','r')
>> text(1,-1,'(1,-1)','HorizontalAlignment','Left')
```

En la figura que produce MATLAB se ve que el punto de corte es el $(1,-1)$ y que las rectas son casi paralelas (y por lo tanto casi idénticas pues coinciden en un punto).

A continuación realizaremos una ligera modificación en los coeficientes de A. Consideremos ahora el sistema $(A+\Delta A)\mathbf{x}=\mathbf{b}$, donde

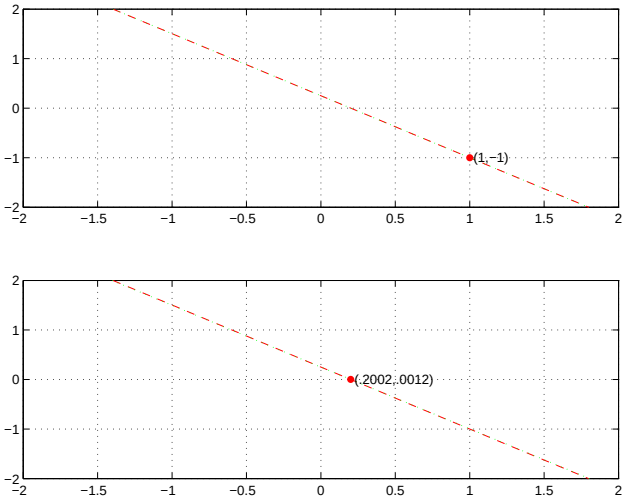
$$A+\Delta A=\begin{pmatrix} 0{,}835 & 0{,}667 \\ 0{,}333 & 0{,}267 \end{pmatrix}, \mathbf{b}=\begin{pmatrix} 0{,}168 \\ 0{,}067 \end{pmatrix}. \tag{10.1.2}$$

Observa que únicamente hemos alterado la entrada $(2,2)$.

Al igual que antes la solución del sistema se calcula como sigue:

```
>> A2=[0.835,0.667;0.333,0.267]
>> b=[0.168;0.067]
>> sol2=A2\b
```

Las representaciones gráfica de los sistemas (10.1.1) y (10.1.2) se pueden ver en la siguiente figura:



Este hecho induce a pensar que sistemas en los que las rectas sean casi paralelas (es decir, aquellos en los que el determinante de la matriz del sistema esté próximo a cero) tendrán números de condición muy grandes. En tal caso, el determinante de la matriz de coeficientes es pequeño, lo que hace que el menor autovalor de A^*A sea pequeño. Recuérdese que

$$\text{cond}_2(A) = \sqrt{\frac{\lambda_n(A^*A)}{\lambda_1(A^*A)}}, \tag{10.1.3}$$

siendo $\lambda_n(A^*A)$ y $\lambda_1(A^*A)$ el mayor y el menor autovalor de A^*A, respectivamente, lo que nos da entonces valores grandes de $\text{cond}_2(A)$.

1.1. Las funciones `cond` y `norm` de `MATLAB`.

Calculemos el número de condición de la matriz de coeficientes para la norma $\|\!|\cdot|\!\|_2$.

```
>> cond(A,2)
```

Observa que es un valor muy grande, tal como se esperaba. Para las restantes normas se obtienen resultados parecidos.

```
>> cond(A,1)
>> cond(A,inf)
```

En este tipo de cálculo, lo que interesa es el orden de magnitud, y no tanto el valor exacto. En nuestro caso, donde se ha efectuado una modificación en la matriz A, se tiene la siguiente acotación:

$$\frac{\|\Delta\mathbf{u}\|}{\|\mathbf{u}+\Delta\mathbf{u}\|} \leq \text{cond}(A)\frac{\|\!|\Delta A|\!\|}{\|\!|A|\!\|}.$$

donde $\mathbf{u}$ y $\mathbf{u}+\Delta\mathbf{u}$ son las soluciones de los sistemas $A\mathbf{x}=\mathbf{b}$ y $(A+\Delta A)\mathbf{x}=\mathbf{b}$, respectivamente.

Con `MATLAB` se puede comprobar que, en nuestro caso, la acotación es muy poco ajustada (como es de esperar).

```
>> sol1=A\b
>> sol2=A2\b
>> miembro_de_la_izquierda = norm(sol2-sol1,2)/norm(sol2,2)
>> miembro_de_la_derecha = cond(A,2)*norm(A2-A,2)/norm(A,2)
```

Nota.- Escribe `help norm` y `help cond` para saber más sobre las ordenes `norm` y `cond` de `MATLAB`.

2. Número de condición y transformaciones elementales.

Veamos cómo afectan las transformaciones elementales al número de condición de una matriz. Para ello consideraremos una nueva matriz, por ejemplo[2],

$$B = \begin{pmatrix} 0{,}4494 & 0{,}1426 \\ 0{,}7122 & 0{,}5643 \end{pmatrix}.$$

```
>> B = [0.4494, 0.1426; 0.7122, 0.5643]
>> cond(B)
```

Consideremos una matriz unitaria, por ejemplo,

$$U = \begin{pmatrix} \cos(\pi/5) & \text{sen}(\pi/5) \\ -\text{sen}(\pi/5) & \cos(pi/5) \end{pmatrix}$$

```
>> U = [cos(pi/5),sin(pi/5);-sin(pi/5),cos(pi/5)]
```

que, como podemos comprobar, es efectivamente unitaria[3]

```
>> U.'*U
```

Entonces, sabemos que se dan las siguiente igualdades

$$\text{cond}_2(B) = \text{cond}_2(B\,U) = \text{cond}_2(U\,B) = \text{cond}_2(U^*BU), \tag{10.2.4}$$

lo que significa que el número de condición respecto de la norma $\|\!|\cdot|\!\|_2$ es invariante por transformaciones unitarias.

Tratemos de comprobar la igualdad $\text{cond}_2(B) = \text{cond}_2(U^*BU)$ con `MATLAB` usando el símbolo lógico[4] `==`

```
>> k1 = cond(B)
>> k2 = cond(U.'*B*U)
>> k1 == k2
```

[2]Si lo deseas puedes elegir otra matriz, por ejemplo una matriz aleatoria con la orden `rand(2)`.

[3]Recuérdese que una matriz $U \in \mathcal{M}_n(\mathbb{C})$ es unitaria si $U^*U = I_n$.

[4]En `MATLAB` existe un tipo de dato llamado lógico que son también matrices de números pero que deben manipularse de distinta manera y tienen otras utilidades. La forma más sencilla de construirlo estos datos lógicos es aplicando la función `logical`.

Evidentemente algo no ha ido bien pues la respuesta de MATLAB ha sido negativa. La razón es la propagación de los errores de redondeo:

```
>> format long
>> k1
>> k2
>> format
```

Veamos ahora que el número de condición respecto de la norma $\|\cdot\|_\infty$ no es estable por transformaciones unitarias.

```
>> c1 = cond(B,inf)
>> c2 = cond(U*B,inf)
```

En este caso no hay dudas de que ambos números de condición son distintos.

Consideremos ahora $P \in \mathcal{M}_n(\mathbb{C})$ no unitaria, por ejemplo,

$$P = \begin{pmatrix} 1 & 0 \\ 455477 & -1142114 \end{pmatrix}$$

y calculemos el número de condición $\mathrm{cond}_2(P\,A)$ y $\mathrm{cond}_2(P^{-1}\,A)$ para

$$A = \begin{pmatrix} 0{,}8350 & 0{,}6670 \\ 0{,}3330 & 0{,}2660 \end{pmatrix}.$$

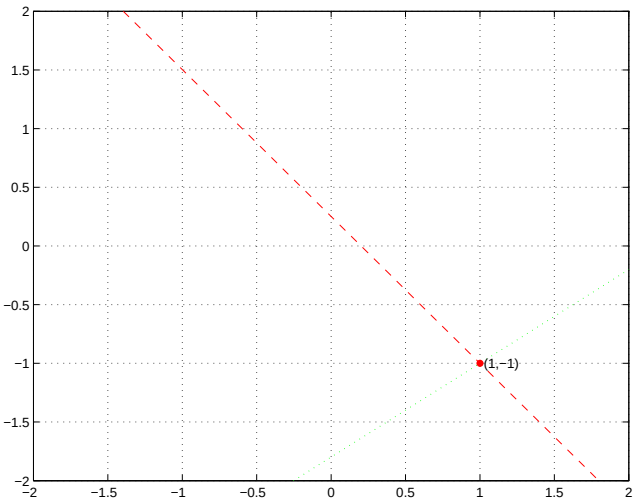
```
>> clear all
>> P = [1,0;455477,-1142114]
>> A = [0.835,0.667;0.333,0.266]
>> k1 = cond(P*A)
>> k2 = cond(inv(P)*A)
```

La comparación entre dos escalares produce un resultado de tipo lógico que vale 1 si es cierta y un 0 cuando es falsa. Las operaciones de relación en MATLAB son las siguientes

==	igualdad
~=	desigualdad
<	menor que
>	mayor que
<=	menor o igual que
>=	mayor o igual que

Observamos que PA tiene el mejor número de condición posible, mientras que $P^{-1}A$ tiene un número de condición mucho más grande que el que tenía A.

La primera opción, PA, representa la mejor situación que se nos puede dar, porque recordemos que el número de condición de una matriz siempre es mayor o igual que 1. Desde el punto de vista geométrico significa que las rectas determinadas por $PA\mathbf{x} = \mathbf{b}$ con $\mathbf{b}^{\mathrm{t}} = (0{,}168, 0{,}067)$ se cortan de forma casi perpendicular. La representación gráfica la tenemos en la siguiente figura:



3. Sistemas mal condicionados.

Consideremos ahora los sistemas lineales $H_n\mathbf{x}_n = \mathbf{b}_n$, donde $H_n \in \mathcal{M}_n(\mathbb{R})$ es la llamada *matriz de Hilbert* de orden n cuya entrada (i, j)-ésima es

$$h_{ij} = 1/(i + j - 1), \qquad i, j = 1, \ldots, n,$$

mientras que $\mathbf{b}_n \in \mathbb{R}^n$ se elige de tal forma que la solución exacta sea $\mathbf{x}_n = (1, 1, \ldots, 1)^{\mathrm{t}}$, es decir, $\mathbf{b}$ es el vector cuya coordenada i-ésima es

$$(\mathbf{b}_n)_i = \sum_{j=1}^{n} \frac{1}{i + j - 1}, \qquad i = 1, 2, \ldots, n.$$

La matriz H_n es claramente simétrica y se puede probar que es definida positiva (por consiguiente, su autovalores son reales y positivos).

Vamos a dibujar una gráfica (en escala semilogarítmica) para visualizar el comportamiento de los errores relativos

$$\varepsilon_n = \|\mathbf{x}_n - \hat{\mathbf{x}}_n\|/\|\mathbf{x}_n\|$$

cuando aumenta n, siendo $\hat{\mathbf{x}}_n$ la solución del sistema $H_n\mathbf{x} = \mathbf{b}_n$ que nos ha proporcionado MATLAB, usando el comando \

Usa el editor de MATLAB para introducir los siguientes comandos, y ejecutarlos todos juntos posteriormente. Guarda estos comandos en un fichero llamado mal_cond.m en tu disco (asegúrate de que el *Current directory* es A:\)

```
>> warning('off')
>> E_n = [];
>> for n = 1:100
>>   clear b xx;
>>   x = ones(n,1);
>>   for i = 1:n
>>     b(i) =  sum(1./(i+(1:n)-1));
>>   end
>>   xx = hilb(n)\b';
>>   E_n = [E_n, norm(x-xx)/norm(x)];
>> end
>> semilogy(1:100,E_n)
>> warning('on')
```

Sobre la base de la observación anterior podríamos especular diciendo que cuando el sistema lineal $A\mathbf{x} = \mathbf{b}$ se resuelve numéricamente, en realidad uno está buscando la solución *exacta* $\hat{\mathbf{x}}$ de un sistema perturbado

$$(A + \Delta A)\hat{\mathbf{x}} = \mathbf{b} + \delta\mathbf{b},$$

donde ΔA y $\delta\mathbf{b}$ son, respectivamente, una matriz y un vector que dependen del método numérico específico que se esté utilizando. Luego, según lo que hemos estudiado en clase, el número de condición de la matriz A explicaría el resultado experimental anterior (retomaremos esta cuestión en el ejercicio 5).

Ejercicios de la práctica 10

Ejercicio 1. Utiliza el comando `eig` de `MATLAB` y la fórmula 10.1.3 para calcular $\mathrm{cond}_2(A)$ siendo A la matriz de Wilson

$$A = \begin{pmatrix} 10 & 7 & 8 & 7 \\ 7 & 5 & 6 & 5 \\ 8 & 6 & 10 & 9 \\ 7 & 5 & 9 & 10 \end{pmatrix}.$$

Calcular también, usando el comando `cond` de `MATLAB`, los condicionamientos de dicha matriz respecto de las normas $\|\cdot\|_1$, $\|\cdot\|_\infty$ y $\|\cdot\|_F$; comprobar que los tres son mayores que $\mathrm{cond}_2(A)$.

Resolver los sistemas

$$A\mathbf{x} = \mathbf{b} \qquad \text{y} \qquad A\mathbf{x} = (\mathbf{b} + \delta\mathbf{b}),$$

para $\mathbf{b} = (32, 23, 33, 31)^{\mathrm{t}}$ y $\delta\mathbf{b} = (0{,}1, -0{,}1, 0{,}1, -0{,}1)^{\mathrm{t}}$. Explicar los resultados obtenidos.

Ejercicio 2. Consideremos el sistema

$$\begin{cases} 3x + 4y = 7 \\ 3x + 5y = 8 \end{cases}$$

1. Calcula su número de condición respecto a la norma 1.
2. Construye, si es posible, sistemas equivalentes que tengan un número de condición mayor y menor que el dado.

Ejercicio 3. Tomemos el sistema $A\mathbf{x} = \mathbf{b}$, donde

$$A = \begin{pmatrix} 1000 & 999 \\ 999 & 998 \end{pmatrix}, \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \mathbf{b} = \begin{pmatrix} 1999 \\ 1997 \end{pmatrix}.$$

Calcula $\|A\|_\infty, \|A^{-1}\|_\infty$ y el número de condición $\mathrm{cond}_\infty(A)$. ¿Se puede decir que el sistema está bien condicionado?

Ejercicio 4. Considerar el siguiente sistema de ecuaciones lineales

$$\begin{pmatrix} 1001 & 1000 \\ 1000 & 1001 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 2001 \\ 2001 \end{pmatrix}.$$

Comprobar que una pequeña variación $\delta\mathbf{b} = (1, 0)^{\mathrm{t}}$ en término independiente produce grandes cambios en la solución. Explicar por qué.

Ejercicio 5. Dibujar una gráfica donde se muestre el comportamiento del número de condición de la matriz de Hilbert de orden n para $n = 1, \ldots, 100$, primero en escala 1:1 y luego en escala semilogarítmica (distingue esta última de las anteriores dibujándola en rojo, por ejemplo)

Usando la orden `hold on`, comparar la gráfica en escala semilogarítmica obtenida con la del comportamiento de los errores relativos estudiada anteriormente (si guardaste aquella ordenes el disco solo tienes que escribir `mal_cond`). Explicar el resultado obtenido.

PRÁCTICA 11

Factorización LU

1. Introducción

En esta práctica aprenderemos a resolver sistemas lineales de ecuaciones con la descomposición LU de una matriz, junto a las sustituciones hacia adelante y hacia atrás. Además, veremos algunas funciones de MATLAB sobre ficheros y estructuras de control en programación.

Pre-requisitos: conocimiento de vectores y matrices en MATLAB. Familiaridad con la eliminación gaussiana, matrices elementales y factorización LU.

2. M-ficheros de ejecución y de funciones en MATLAB

Los *M-ficheros de ejecución (scripts)* son simplemente una forma conveniente de poner unos comandos de MATLAB que queremos ejecutar en secuencia. Por ejemplo, abrimos el editor de MATLAB e introducimos el siguiente código[1]

```
format rat % formato racional
A=[ 1, 2,-3, 4; ...
    2, 0,-1, 4; ...
    3, 1, 0, 6; ...
   -4, 4, 8, 0]
b=[ 1; 1; 1;-1]
M=[A,b]
R=rref(M)
x=R(:,5)
format % vuelta al formato original
```

Grabemos el fichero como `ejemplo.m`. Ahora, en el indicador de MATLAB, escribimos

```
>> ejemplo
```

y cada línea de `ejemplo.m` se ejecutará en el orden que aparece en el fichero.

Los *M-ficheros de funciones (funciones)* son similares a los scripts. El código introducido se ejecuta en secuencia. Sin embargo, mientras que los scripts permiten al

[1] Los puntos suspensivos son comandos de continuación en MATLAB. Todo lo que sigue al símbolo % es tratado como comentario. MATLAB no lo procesa.

usuario introducir datos, las funciones pueden devolver una respuesta a las funciones que las llamen (en algunos casos, la propia pantalla de MATLAB).

Supongamos, por ejemplo, que quisiéramos codificar la función definida por $f(x) = x^2$. Abrimos el editor de MATLAB e introducimos las siguientes líneas de código.

```
function y=f(x)
  y=x^2;
```

Lo grabamos como f.m. Ahora la probamos en el indicador de MATLAB con los siguientes comandos.

```
>> t=8;
>> z=f(t)
```

Observemos que en la llamada a la función no es necesario que se use el mismo nombre para la variable independiente. En la función es x y nosotros hemos usado t. Tampoco tienen que coincidir los nombres de las variables dependientes. Por ejemplo, es válido lo siguiente.

```
>> t=8;
>> t_cuadrado=f(t);
>> t_cuadrado
```

Evidentemente, la función no tiene por qué llamarse f, podemos darle el nombre que queramos. Abrimos el editor de MATLAB e introducimos las siguientes líneas de código.

```
function y=cuadrado(x)
  y=x^2;
```

No obstante, MATLAB requiere que grabemos el fichero con el mismo nombre que le demos a la función. Esto es, en el caso anterior, el fichero debe llamarse cuadrado.m. Esta función tendrá un comportamiento igual que la primera función. Solamente han cambiado los nombres.

```
>> t=8;
>> t_cuadrado = cuadrado(t);
>> t_cuadrado
```

Las funciones pueden tener más de una entrada y pueden tener una salida múltiple. Por ejemplo, consideremos la función definida por $g(x, y) = x^2 + y^2$. La codificamos como sigue en el editor de MATLAB.

```
function z = g(x,y)
  z=x^2+y^2;
```

Grabamos este fichero como g.m y ejecutamos los siguientes comandos en el indicador de MATLAB.

```
>> u=3;v=4;
>> z=g(u,v);
>> z
```

Aunque pronto encontraremos funciones con respuesta múltiple, veamos un ejemplo. Consideremos la función $h(x, y) = [x^2 + y^2, x^2 - y^2]$. La codificamos como sigue en el editor de MATLAB.

```
function [h1,h2] = h(x,y)
  h1=x^2+y^2;
  h2=x^2-y^2;
```

Grabamos este fichero como h.m y ejecutamos los siguientes comandos en el indicador de MATLAB.

```
>> u=5;v=2;
>> [a,b]=h(u,v);
>> [a,b]
```

Tradicionalmente MATLAB obliga a crear un M-fichero por cada función. El nombre de la función debe coincidir con el de la función. No obstante, a partir de la versión 5.0 se han introducido las *subfunciones*, que son funciones adicionales definidas en un mismo M-fichero con nombre diferentes del nombre del fichero (y del nombre de la función principal) y que sólo pueden ser llamadas por funciones contenidas en ese fichero, resultando "invisibles" par otra funciones externas.

Por ejemplo, si escribimos en el editor de MATLAB

```
function y = fun(x)
  y = x+subfun(x);

function y = subfun(x)
  y = x^2;
```

grabamos este fichero como fun.m y ejecutamos los siguientes comandos en el indicador de MATLAB

```
>> w=2;
>> fun(2)
>> subfun(2)
```

observamos que MATLAB "reconoce" la función fun, pero no así la función subfun; aunque esta última sea necesaria para el buen funcionamiento de la primera.

3. Métodos específicos para la resolución de sistemas triangulares.

3.1. Sustitución hacia atrás.

Consideremos el sistema de ecuaciones[2]

$$\left.\begin{array}{rrrcr} 2x_1 & +x_2 & -x_3 & = & 4 \\ & -2x_2 & +x_3 & = & -3 \\ & & 4x_3 & = & 8 \end{array}\right\} \tag{11.3.1}$$

En forma matricial, se puede representar como

$$\begin{pmatrix} 2 & 1 & -2 \\ 0 & -2 & 1 \\ 0 & 0 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 4 \\ -3 \\ 8 \end{pmatrix}.$$

Hemos escrito el sistema (11.3.1) como

$$U\mathbf{x} = \mathbf{c},$$

donde

$$U = \begin{pmatrix} 2 & 1 & -2 \\ 0 & -2 & 1 \\ 0 & 0 & 4 \end{pmatrix}, \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, \mathbf{c} = \begin{pmatrix} 4 \\ -3 \\ 8 \end{pmatrix}.$$

Observemos que la matriz U es cuadrada de orden 3 y triangular superior, porque cada coeficiente por debajo de la diagonal principal es nulo. Además, U es invertible. Estos sistemas se resuelven fácilmente con una técnica que se denomina *sustitución hacia atrás*. En primer lugar, resolvemos la última ecuación del sistema (11.3.1) para calcular el valor de x_3, y nos da $x_3 = 2$.

Este valor lo sustituimos en la segunda ecuación del sistema (11.3.1).

$$-2x_2 + x_3 = -3 \Rightarrow x_2 = (-3 - x_3)/(-2) = (-3 - 2)/(-2) = 5/2.$$

Por último, sustituimos $x_3 = 2$ y $x_2 = 5/2$ en la primera ecuación del sistema (11.3.1), y calculamos el valor de x_1.

$$2x_1 + x_2 - 2x_3 = 4 \Rightarrow x_1 = (4 - x_2 + 2x_3)/2 = (4 - 5/2 + 2 \cdot 2)/2 = 11/4.$$

En general, si U es triangular superior e invertible[3], entonces para cualquier $\mathbf{c}$ el sistema $U\mathbf{x} = \mathbf{c}$ tiene solución única. Ésta se encuentra fácilmente con la sustitución hacia atrás.

$$\left.\begin{array}{rrrcr} u_{11}x_1 + & u_{12}x_2 & + \ldots + & u_{1n}x_n & = c_1 \\ & u_{22}x_2 & + \ldots + & u_{2n}x_n & = c_2 \\ & & & \vdots & \\ & & & u_{nn}x_n & = c_n \end{array}\right\}$$

[2]Que podemos pensar que el sistema equivalente a uno dado, después de haber calculado la forma reducida por filas de la matriz ampliada del sistema original.

[3]Recuérdese que si una matriz es triangular, entonces su determinante coincide con el producto de los elementos de su diagonal principal. Por lo que la condición necesaria y suficiente para que sea invertible es que los elementos de su diagonal principal sean distintos de cero.

En primer lugar, resolvemos x_n de la última ecuación.

$$x_n = c_n/u_{nn}. \tag{11.3.2}$$

Con este dato y la penúltima ecuación encontramos x_{n-1}.

$$x_{n-1} = (c_{n-1} - u_{n-1,n}x_n)/u_{n-1,n-1}.$$

Si continuamos de esta manera, podemos resolver todo el sistema. Por ejemplo, la i-ésima ecuación

$$u_{ii}x_i + u_{i,i+1}x_{i+1} + \ldots + u_{in}x_n = c_i$$

nos lleva a

$$x_i = (c_i - u_{i,i+1}x_{i+1} - \ldots u_{in}x_n)/u_{ii}$$

y en notación sumatoria

$$x_i = (c_i - \sum_{j=i+1}^{n} u_{ij}x_j)/u_{ii}. \tag{11.3.3}$$

esta última ecuación es la que permite automatizar el proceso.

Vamos a sistematizar el proceso de sustitución hacia atrás definición una función de MATLAB. Para ello, abrimos el editor de MATLAB y comenzamos dando un nombre a la función y a sus entradas y salidas. Pasamos como dato de entrada la matriz de coeficientes U, que debe ser cuadrada de orden n, y el vector $\mathbf{c}$ de términos independientes. La función tiene que devolver la solución del sistema $U\mathbf{x} = \mathbf{c}$ en la variable $\mathbf{x}$.

```
function x=sust_atras(U,c)
```

Ahora, almacenamos el tamaño de la matriz U en las variables m (número de filas) y n (número de columnas).

```
[m,n]=size(U);
```

Si la matriz no es cuadrada, tenemos un problema. Puede ocurrir que el sistema tenga más de una solución. Verificamos tal condición, y si la matriz U no es cuadrada paramos la ejecución y damos un mensaje de aviso.

```
if m~=n
  disp('La matriz U no es cuadrada.')
  return;
end
```

Ahora reservamos un espacio que contendrá la solución del sistema.

```
x=zeros(n,1);
```

Usamos la ecuación (11.3.2) para calcular x_n y almacenar la solución

```
x(n)=c(n)/U(n,n);
```

Con la sustitución hacia atrás, podemos calcular los valores de x_i para $i = n - 1, \ldots, 2, 1$. Esta es una tarea iterada, que podemos programar con un bucle `for`. El bucle interno calcula la suma de la ecuación (11.3.3). Por último, la ecuación (11.3.3) se usa para obtener x_i.

```
for k=n-1:-1:1
  sum=0;
  for j=k+1:n
    sum=sum+U(k,j)*x(j);
  end
  x(k)=(c(k)-sum)/U(k,k);
end
```

El texto completo que debe aparecer escrito en el editor de `MATLAB` es el siguiente:

```
function x = sust_atras(U,c)
  [m,n]=size(U);
  if m~=n
    disp('La matriz U no es cuadrada.')
    return;
  end
  x=zeros(n,1);
  x(n)=c(n)/U(n,n);
  for k=n-1:-1:1
    sum=0;
    for j=k+1:n
      sum=sum+U(k,j)*x(j);
    end
    x(k)=(c(k)-sum)/U(k,k);
  end
```

Guardamos el fichero como `sust_atras.m` y lo probamos con la matriz del sistema (11.3.1). En primer lugar, introducimos U y $\mathbf{c}$.

```
>> U=[2,1,-2;0,-2,1;0,0,4]
>> c=[4;-3;8]
```

Observemos que $\mathbf{c}$ se define como un vector columna. Finalmente, obtenemos la solución con los siguientes comandos.

```
>> format rat
>> x=sust_atras(U,c)
>> format
```

y vemos que coincide con la solución del sistema (11.3.1) que habíamos calculado a mano.

3.2. Sustitución hacia adelante.

Consideremos ahora el sistema de ecuaciones

$$\left.\begin{array}{rcr} c_1 & = & 4 \\ 2c_1 + c_2 & = & 5 \\ -3c_1 + 2c_2 + c_3 & = & -10 \end{array}\right\} \tag{11.3.4}$$

En forma matricial, este sistema se puede escribir como

$$\begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -3 & 2 & 1 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} = \begin{pmatrix} 4 \\ 5 \\ -10 \end{pmatrix}. \tag{11.3.5}$$

y el sistema (11.3.4) toma la forma

$$L\mathbf{c} = \mathbf{b},$$

donde

$$L = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -3 & 2 & 1 \end{pmatrix}, \mathbf{c} = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix}, \mathbf{b} = \begin{pmatrix} 4 \\ 5 \\ -10 \end{pmatrix}.$$

Observemos que L es una matriz cuadrada 3×3, triangular inferior. Además, los elementos de la diagonal principal son iguales a 1, lo que simplifica el cálculo de la solución que se puede obtener mediante el *método de sustitución hacia adelante*. Empezamos por resolver la ecuación para c_1.

$$c_1 = 4.$$

Sustituímos c_1 en la segunda ecuación del sistema (11.3.4) y obtenemos c_2.

$$\begin{array}{rcl} 2c_1 + c_2 & = & 5 \\ c_2 & = & 5 - 2c_1 \\ c_2 & = & 5 - 2 \cdot 4 \\ c_2 & = & -3 \end{array}$$

Sustituímos ahora c_1 y c_2 en la tercera ecuación del sistema (11.3.4) y calculamos c_3.

$$\begin{array}{rcl} -3c_1 + 2c_2 + c_3 & = & -10 \\ c_3 & = & -10 + 3c_1 - 2c_2 \\ c_3 & = & -10 + 3 \cdot 4 - 2 \cdot (-3) \\ c_3 & = & 8 \end{array}$$

En general, si L es una matriz cuadrada, triangular inferior y con unos en la diagonal, entonces para cualquier $\mathbf{b}$ el sistema $L\mathbf{c} = \mathbf{b}$ tiene solución única. El sistema se resuelve fácilmente con sustitución hacia adelante.

$$\left.\begin{array}{rcl} c_1 & = & b_1 \\ l_{21}c_1 + c_2 & = & b_2 \\ & \vdots & \\ l_{n1}c_1 + l_{n2}c_2 + \ldots + c_n & = & b_n \end{array}\right\}$$

Resolvemos la primera ecuación para c_1.

$$c_1 = b_1.$$

Con este resultado calculamos c_2 en la segunda ecuación.

$$\begin{array}{rcl} l_{21}c_1 + c_2 & = & b_2; \\ c_2 & = & b_2 - l_{21}c_1. \end{array}$$

Continuando de esta forma calculamos el resto de incógnitas. La i-ésima ecuación

$$l_{i1}c_1 + l_{i2}c_2 + \ldots + l_{i,i-1}c_{i-1} + c_i = b_i;$$

nos da

$$c_i = b_i - l_{i1}c_1 - l_{i2}c_2 - \ldots - l_{i,i-1}c_{i-1}.$$

En notación de sumatorio es

$$c_i = b_i - \sum_{j=1}^{i-1} l_{ij}c_j.$$

Definimos la función `sust_adelante` sin explicación. Animamos al lector a usar la explicación de `sust_atras` para comprender el algoritmo antes de pasar a probar la rutina.

```
function c=sust_adelante(L,b)
  [m,n]=size(L);
  if m~=n
    disp('La matriz L no es cuadrada.')
    return;
  end
  c=zeros(n,1);
  c(1)=b(1);
  for k=2:n
    sum=0;
    for j=1:k-1
      sum=sum+L(k,j)*c(j);
    end
```

```
    c(k)=b(k)-sum;
  end
```

Grabamos el fichero como `sust_adelante.m` y comprobamos su funcionamiento con el sistema (11.3.4).

```
>> L=[1,0,0;2,1,0;-3,2,1]
>> b=[4;5;-10]
>> c=sust_adelante(L,b)
```

Como era de esperar, la solución coincide con la que habíamos obtenido previamente.

4. Factorización LU

Nota 11.4.1. La descripción y justificación teórica de la descomposición LU se puede encontrar en los apuntes de la asignatura Álgebra y Geometría.

Consideremos el sistema de ecuaciones

$$\left.\begin{array}{rrrcr} 2x_1 & +x_2 & -2x_3 & = & 4 \\ 4x_1 & & -3x_3 & = & 5 \\ -6x_1 & -7x_2 & +12x_3 & = & -10 \end{array}\right\}$$

En forma matricial, el sistema tiene la forma

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

donde

$$A = \begin{pmatrix} 2 & 1 & -2 \\ 4 & 0 & -3 \\ -6 & -7 & 12 \end{pmatrix}, \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, \mathbf{b} = \begin{pmatrix} 4 \\ 5 \\ -10 \end{pmatrix}$$

Vamos a usar operaciones elementales por filas para llevar la matriz A a una matriz triangular superior, y guardamos los multiplicadores de cada posición en una matriz triangular inferior L según vamos haciendo los cálculos.

Calculamos el primer multiplicador con

$$l_{21} := a_{21}/a_{11} = 4/2 = 2.$$

Restamos a la segunda fila la primera multiplicada por l_{21}.

$$E_1 A = \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 2 & 1 & -2 \\ 4 & 0 & -3 \\ -6 & -7 & 12 \end{pmatrix} = \begin{pmatrix} 2 & 1 & -2 \\ 0 & -2 & 1 \\ -6 & -7 & 12 \end{pmatrix}$$

Calculamos el segundo multiplicador con

$$l_{31} := a_{31}/a_{11} = -6/2 = -3.$$

Ahora le restamos a la tercera fila la primera multiplicada por l_{31}.

$$E_2(E_1A) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 3 & 0 & 1 \end{pmatrix} \begin{pmatrix} 2 & 1 & -2 \\ 0 & -2 & 1 \\ -6 & -7 & 12 \end{pmatrix} = \begin{pmatrix} 2 & 1 & -2 \\ 0 & -2 & 1 \\ 0 & -4 & 6 \end{pmatrix}$$

Calculamos ahora el siguiente multiplicador con

$$l_{32} = a_{32}^{(2)}/a_{22}^{(2)} = -4/(-2) = 2,$$

donde $a_{32}^{(2)}, a_{22}^{(2)}$ son las entradas correspondientes de $A^{(2)} = E_2E_1A$. Restamos a la tercera fila la segunda multiplicada por l_{32}.

$$E_3(E_2E_1A) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -2 & 1 \end{pmatrix} \begin{pmatrix} 2 & 1 & -2 \\ 0 & -2 & 1 \\ 0 & -4 & 6 \end{pmatrix} = \begin{pmatrix} 2 & 1 & -2 \\ 0 & -2 & 1 \\ 0 & 0 & 4 \end{pmatrix} = U.$$

Entonces

$$U = \begin{pmatrix} 2 & 1 & -2 \\ 0 & -2 & 1 \\ 0 & 0 & 4 \end{pmatrix}.$$

Construimos la matriz L a partir de la matriz identidad colocando los multiplicadores l_{ij} en sus posiciones correspondientes.

$$L = \begin{pmatrix} 1 & 0 & 0 \\ l_{21} & 1 & 0 \\ l_{31} & l_{32} & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -3 & 2 & 1 \end{pmatrix}.$$

Observemos que estas matrices U y L son las matrices triangulares que hemos usado en los sistemas de la sección 3. Entonces $A = LU$ y el sistema

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

se transforma en

$$(LU)\mathbf{x} = \mathbf{b}$$
$$L(U\mathbf{x}) = \mathbf{b}$$

Podemos escribirlo como dos sistemas

$$L\mathbf{c} = \mathbf{b} \text{ y } U\mathbf{x} = \mathbf{c}.$$

Estos sistemas fueron resueltos en la sección 3. Por tanto, la solución del sistema $A\mathbf{x} = \mathbf{b}$ es

$$\mathbf{x} = \begin{pmatrix} 11/4 \\ 5/2 \\ 2 \end{pmatrix}.$$

Vamos ahora a escribir una rutina para calcular la descomposición LU. La entrada es una matriz cuadrada A, y la salida son matrices triangulares L (inferior con unos en la diagonal) y U (superior) tales que $A = LU$.

```
function [L,U]=mi_lu(A)
```

Al igual que antes, si A no es cuadrada, devolvemos un mensaje de error y paramos.

```
[m,n]=size(A);
if m ~= n
  disp('A no es cuadrada.')
  return
end
```

La asignación inicial de la matriz L es la identidad.

```
L=eye(n);
```

Modificaremos las entradas de la matriz A, usando como pivotes los elementos de la diagonal para eliminar las entradas que están por debajo de ellos. Como no hay coeficientes por debajo de la fila n, el bucle de eliminación llega hasta $n-1$.

```
for k=1:n-1
```

En el paso k-ésimo, la matriz A tendrá la siguiente forma, donde hemos usado $a^{\bullet}_{ij}$ para nombrar a sus entradas, dado que los pasos previos de eliminación han alterado los valores originales a_{ij}.

$$\begin{pmatrix} a_{11} & \dots & a_{1k} & a_{1,k+1} & \dots & a_{1n} \\ \vdots & \ddots & & \dots & & \\ 0 & \dots & a^{\bullet}_{kk} & a^{\bullet}_{k,k+1} & \dots & a^{\bullet}_{kn} \\ 0 & \dots & a^{\bullet}_{k+1,k} & a^{\bullet}_{k+1,k+1} & \dots & a^{\bullet}_{k+1,n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & a^{\bullet}_{nk} & a^{\bullet}_{n,k+1} & \dots & a^{\bullet}_{nn} \end{pmatrix}.$$

Ahora notemos que las filas por debajo de $a^{\bullet}_{kk}$ van desde $k+1$ hasta n.

```
for i=k+1:n
```

A continuación, determinamos el multiplicador. Es importante observar que las entradas en A son las actuales, todas calculadas en los $k-1$ pasos previos de eliminación.

```
L(i,k)=A(i,k)/A(k,k);
```

Con este multiplicador eliminamos $a^{\bullet}_{i,k}$. Estamos en la columna k, y la eliminación afectará a las entradas a la **derecha** de esta columna, que corresponden a un índice inicial de $k+1$.

```
for j=k:n
  A(i,j)=A(i,j)-L(i,k)*A(k,j);
end
```

Cerramos los dos bucles anteriores

```
end
end
```

La matriz A se ha transformado en triangular superior. Basta asignar este valor a U

```
U=A;
```

Además, L está también completa, y no tenemos que hacer nada con ella.

Si no lo hemos hecho ya, abrimos el editor de MATLAB e introducimos el código completo.

```
function [L,U]=mi_lu(A)
  [m,n]=size(A);
  if m ~= n
    disp('A no es cuadrada.')
    return
  end
  L=eye(n);
  for k=1:n-1
    for i=k+1:n
    L(i,k)=A(i,k)/A(k,k);
      for j=k:n
        A(i,j)=A(i,j)-L(i,k)*A(k,j);
      end
    end
  end
  U=A;
```

Finalmente, grabamos el fichero como mi_lu.m y comprobamos su funcionamiento usando la matriz

$$A = \begin{pmatrix} 2 & 1 & -2 \\ 4 & 0 & -3 \\ -6 & -7 & 12 \end{pmatrix},$$

de la que sabemos que

$$A = LU = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -3 & 2 & 1 \end{pmatrix} \begin{pmatrix} 2 & 1 & -2 \\ 0 & -2 & 1 \\ 0 & 0 & 4 \end{pmatrix}.$$

Introducimos la matriz A.

```
>> A=[2,1,-2;4,0,-3;-6,-7,12]
```

Usamos nuestra función mi_lu para calcular la descomposición LU.

```
>> [L,U]=mi_lu(A)
```

y verificamos que el resultado obtenido concuerda con los anteriores.

5. MATLAB y la factorización LU

MATLAB tiene una rutina muy eficiente para calcular la factorización LU de una matriz. Si no se necesitan cambios de filas, el comando `[L,U]=lu(A)` calcula una matriz L triangular inferior y una matriz U triangular superior tales que $A = LU$.

```
>> format rat
>> A=[2,1,-1;0,1,-1;1,0,1]
>> [L,U]=lu(A)
```

Si hay que realizar cambios de filas para calcular la descomposición LU, entonces el comando `[L,U,P]=lu(A)` devuelve además una matriz de permutación P tal que

$$PA = LU.$$

```
>> A=[1,1,0;2,-2,1;0,1,5]
>> [L,U,P]=lu(A)
```

MATLAB usa *pivoteo por filas*[4] para el cálculo de la factorización LU. Observemos el siguiente ejemplo.

```
>> A=[1,2,-3,4;4,8,12,-8;2,3,2,1;-3,-1,1,-4]
>> lu(A)
>> [L,U]=lu(A)
>> [L,U,P]=lu(A)
>> format
```

5.1. Rendimiento.

Podemos pensar qué es más costoso, en términos de CPU, calcular las tres matrices de la factorización LU u obtener la forma reducida por filas de una matriz. Vamos a hacer algunos experimentos.

```
>> A=round(10*rand(50)-5);
>> tic;rref(A);toc
```

Para comparar, veamos el tiempo que tarda en realizar una descomposición LU de la matriz A.

[4]Se toma como pivote el elemento de mayor módulo entre los $n - j$ últimos elementos de la columna j-ésima; es decir, se elige $a_{ij}^{\bullet}$, $j \leq i \leq n$, de forma que

$$|a_{ij}^{\bullet}| = \max_{j \leq l \leq n} |a_{lj}^{\bullet}|.$$

```
>>tic;[L,U,P]=lu(A);toc
```

Como se ve, el comando `lu` es muy eficiente, y por ello `MATLAB` lo usa en muchas de sus rutinas.

5.2. Matrices con muchos ceros.

Consideremos una matriz A con un número elevado de ceros en sus entradas. Las llamaremos matrices dispersas (*sparse*). Uno de los problemas de la factorización LU es que si A es una matriz dispersa, las matrices L y U no lo serán en general. Veamos un ejemplo.

```
>> close all
>> B=bucky;
>> [L,U,P]=lu(B);
>> spy(B); % figura 1
>> figure
>> spy(L); % figura 2
>> figure
>> spy(U); % figura 3
```

Con el comando `gallery` de `MATLAB` podemos tener acceso a una colección de matrices que poseen diferentes estructuras y propiedades. Escribe

```
>> help gallery
```

para mayor información.

Ejercicios de la práctica 11

Ejercicio 1. Usa la descomposición LU para resolver los siguientes sistemas:

$$\left.\begin{array}{rrrcr} -2x_1 & & -3x_3 & = & 6 \\ x_1 & +2x_2 & +x_3 & = & 4 \\ -3x_1 & +x_2 & -5x_3 & = & 15 \end{array}\right\}, \qquad \left.\begin{array}{rrrcr} -2x_1 & -3x_2 & -4x_3 & = & 12 \\ -3x_1 & & +x_3 & = & 9 \\ 3x_1 & -x_2 & -x_3 & = & -3 \end{array}\right\}$$

Ejercicio 2. Construye una matriz A de orden 3×3 singular sin entradas nulas tal que la rutina `mi_lu` falle. ¿Qué mensaje de error da `MATLAB`? Explica por qué la rutina falla sobre la matriz.

Ejercicio 3. Construye una matriz A de orden 3×3 invertible sin entradas nulas tal que la rutina `mi_lu` falle. ¿Qué mensaje de error da `MATLAB`? Explica por qué la rutina falla sobre la matriz.

Ejercicio 4. Si una matriz A necesita intercambio de filas en la eliminación, entonces la rutina `mi_lu` falla.

- Observa qué ocurre al calcular una descomposición LU de la siguiente matriz:

$$A = \begin{pmatrix} -1 & -1 & -1 & 1 \\ 1 & 1 & 0 & -1 \\ 2 & -1 & -1 & 0 \\ 5 & -3 & -3 & 2 \end{pmatrix}.$$

- Calcula L y U tales que $PA = LU$.
- Considera el vector

$$\mathbf{b} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}.$$

  Explica cómo se pueden usar las funciones de `MATLAB lu`, `sust_adelante` y `sust_atras` para calcular la solución del sistema $A\mathbf{x} = \mathbf{b}$. Úsalas para calcularla.

Ejercicio 5. Consideremos los sistemas de ecuaciones lineales $A\mathbf{x} = \mathbf{b}$ con

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

y $\mathbf{b}$ tal que la solución correspondiente sea $\mathbf{u} = (1, 1, 1)^{\mathrm{t}}$, siendo ϵ un número real positivo. Calcular la factorización LU de A para distintos valores de ϵ y observar que $l_{32} \to \infty$ cuando $\epsilon \to 0$. A pesar de ello, verificar que las solución calculada posee una buena precisión.

PRÁCTICA 12

Otras factorizaciones de matrices

1. Introducción

En esta práctica estudiaremos las factorizaciones de Cholesky y QR. Además, veremos cómo se puede añadir comentarios de ayuda en nuestros M-ficheros que se puedan visualizar con el comando `help` de `MATLAB`.

Pre-requisitos: Factorización de Cholesky. Matrices de Householder. Factorización QR

2. Factorización de Cholesky

Consideremos la matriz simétrica

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 8 & 4 \\ 3 & 4 & 14 \end{pmatrix}.$$

Usando `MATLAB` podemos comprobar que es definida positiva, por ejemplo, calculando sus autovalores y observando que son estrictamente positivos

```
>> eig(A)
```

Por tanto, tenemos garantía que A admite una factorización de Cholesky, es decir, podemos afirmar que existe Q triangular inferior tal con entradas positivas en su diagonal principal tal que

$$A = QQ^{\mathrm{t}}.$$

Según vimos en clase de teoría, las entradas de Q se pueden calcular mediante el siguiente algoritmo: ponemos $q_{11} = \sqrt{a_{11}}$ y para $i = 2, \ldots, n$,

$$q_{ij} = \frac{1}{q_{jj}} \left(a_{ij} - \sum_{k=1}^{j-1} q_{ik} q_{jk} \right), \ j = 1, \ldots, i-1,$$

$$q_{ii} = \left(a_{ii} - \sum_{k=1}^{i-1} q_{ik}^2 \right)^{1/2}.$$

Veamos paso a paso como funciona este algoritmo con nuestro ejemplo. Vamos a usar MATLAB para hacer los cálculos, aunque dado el tamaño de la matriz bien se podrían hacer a mano.

Para comenzar definimos una matriz nula Q del mismo orden que A.

```
>> Q = zeros(size(A))
```

Según nuestro algoritmo

```
>> Q(1,1) = sqrt(A(1,1))
>> Q(2,1) = A(2,1)/Q(1,1)
>> Q(2,2) = sqrt(A(2,2)-Q(2,1)^2)
>> Q(3,1) = A(3,1)/Q(1,1)
>> Q(3,2) = (A(3,2)-Q(3,1)*Q(2,1))/Q(2,2)
>> Q(3,3) = sqrt(A(3,3)-Q(3,1)^2-Q(3,2)^2)
```

Ahora podemos comprobar que en efecto $A = QQ^{\mathrm{t}}$.

```
>> A == Q*Q'
```

Este proceso se puede automatizar en MATLAB definiendo una función adecuaamente. La siguiente función de MATLAB calcula la factorización de Cholesky de una matriz hermítica definida positiva.

```
function H = mi_chol(A)
%MI_CHOL:
% entrada: A    - matriz hermítica definida positiva.
% salida:  H    - matriz triangular inferior tal que A = H*H'
%
% Si la matriz A no es hermítica o definida positiva la función
% operará incorrectamente pudiéndose producir errores de división
% por cero.

  [n,n] = size(A);
  H = zeros(n);
  H(1,1) = sqrt(A(1,1));
  for i = 2:n
    for j = 1:i-1
```

```
      H(i,j) = (A(i,j)-H(i,1:j-1)*H(j,1:j-1)')/H(j,j);
    end
    H(i,i) = sqrt(A(i,i)-H(i,1:i-1)*H(i,1:i-1)');
  end
```

Usemos nuestro ejemplo para comprobar que nuestra función está bien definida:

```
>> H = mi_chol(A)
>> Q == H
```

En nuestra función `mi_chol` hemos añadido un comentario de ayuda. Observa qué ocurre cuando escribes

```
>> help mi_chol
```

en el indicador de `MATLAB`.

En esta ayuda advertimos que la función `mi_chol` no verifica si la matriz es hermítica o definida positiva, por lo que la salida de nuestra función puede no ser fiable a menos que tengamos garantía de que la matriz usada tenga estas propiedades. Evidentemente podríamos añadir un "test de hipótesis" en nuestra función, por ejemplo, escribiendo

```
>> if A == A'
```

antes de la línea 10 de `mi_chol` y

```
 else
   error('La matriz no es hermítica');
 end
```

al final de la función `mi_chol`, para verificar que la matriz es hermítica; o

```
 if A(i,i) < H(i,1:i-1)*H(i,1:i-1)' ...
 error('La matriz no es definida positiva'); end
```

antes de la línea 17 de `mi_chol`, para comprobar que la matriz es definida positiva.

No obstante, lo habitual es no incluir demasiados "tests de hipótesis" en favor de una mayor velocidad de cálculo. En todo caso, he aquí nuestra función `mi_chol` modificada, a la que llamamos `mi_chol2`

```
function H = mi_chol2(A)
%MI_CHOL2:
% entrada: A    - matriz hermítica definida positiva.
% salida:  H    - matriz triangular inferior tal que A = H*H'
%
% Si la matriz A no es hermítica o definida positiva la función
% devolverá un mensaje de error.

  if A == A'
    [n,n] = size(A);
    H = zeros(n);
    H(1,1) = sqrt(A(1,1));
    for i = 2:n
      for j = 1:i-1
        H(i,j) = (A(i,j)-H(i,1:j-1)*H(j,1:j-1)')/H(j,j);
      end
      if A(i,i) < H(i,1:i-1)*H(i,1:i-1)' ...
      error('La matriz no es definida positiva'); end
      H(i,i) = sqrt(A(i,i)-H(i,1:i-1)*H(i,1:i-1)');
    end
  else
    error('La matriz no es hermítica');
  end
```

MATLAB posee un comando propio para calcular la factorización de Cholesky de una matriz hermítica y definida positiva. Si leemos la ayuda de este comando

```
>> help chol
```

observamos que calcula una matriz triangular superior tal que $A = Q^{\mathrm{t}}Q$; además, esta función sí comprueba que la matriz introducida sea definida positiva, aunque no que sea hermítica.

```
>> A = [1,2;-1,5]
>> chol(A)
```

Nótese que la salida del comando `chol` de `MATLAB` es la traspuesta conjugada de la salida de nuestra función `mi_chol`.

2.1. Rendimiento.

El algoritmo para la factorización de Cholesky es muy estable respecto a la propagación de errores de redondeo, incluso para matrices mal condicionadas.

```
>> A = hilb(15);
>> Q = mi_chol(A);
>> spy(A-Q*Q')
```

Por otra parte, el algoritmo de factorizción de Cholesky es mucho más eficiente que el de factorización LU, aunque usemos la estrategia de pivoteo parcial.

```
>> A = rand(50); %%%%% Definimos una matriz aleatoria de orden 50
>> B = A*A'; %%%%% Definimos una matriz simétrica, que será
>>           %%%%% será definida positiva si A es invertible
>> Q = mi_chol(B);
>> spy(B-Q*Q')
>> [L,U] = lu(B);
>> spy(B-L*U)
>> [L,U,P] = lu(B);
>> spy(P*B-L*U)
```

3. Matrices de Householder

Comencemos definiendo la matriz de Householder asociada a un vector $\mathbf{v} \in \mathbb{R}^n$.

En primer lugar consideramos la siguiente función de `MATLAB` calcula un vector de Householder $\mathbf{w}$ y el factor β de un vector no nulo $\mathbf{v} \in \mathbb{R}^n$.

```
function [w,beta] = vector_householder(v)
%VECTOR_HOUSEHOLDER:
% entrada: v    - un vector no nulo.
% salida:  w    - un vector de Householder de v.
%          beta - el módulo de w al cuadrado divido por dos.

  n = length(v);
  nv = norm(v);
```

```
  w = v;
  c = nv^2 - v(1)^2;
  if c == 0
    w(1) = -min(2*v(1),0);
    beta = w(1)^2/2;
  else
    if v(1) >= 0
      w(1) = v(1) + nv;
    else
      w(1) = v(1) - nv;
    end
    beta = nv*abs(w(1));
  end
```

La siguiente función de MATLAB calcula la imagen de un vector $\mathbf{a} \in \mathbb{R}^n$ por la matriz de Householder un vector $\mathbf{v}$ dado.

```
function [Ha] = im_house(v,a)
%IM_HOUSE
% entrada: v    - un vector no nulo.
%          a    - un vector arbitrario.
% salida:  Ha   - imagen de a por la matriz de Householder H tal
%                 que Hv es un múltiplo del vector (1, 0, ..., 0);
%                 esto es, la imagen de a por la simetría respecto
%                 del hiperplano ortogonal a un vector de Householder
%                 de v.

  [w,beta] = vector_householder(v);
  alpha = w'*a;
  if beta == 0
    Ha = a;
  else
    Ha = a - alpha/beta*w;
  end
```

Comprobemos el buen funcionamiento de las dos funciones anteriores calculando un vector de Householder de $\mathbf{v} = (1, 2, 3)^{\mathrm{t}}$.

```
>> v = [1; 2; 3];
>> [w,beta] = vector_householder(v)
```

y la imagen de $\mathbf{a} = (0, 3, -2)^{\mathrm{t}}$ por la transformación de Householder de matriz $H = \mathcal{H}(\mathbf{w})$ para el vector $\mathbf{w}$ obtenido anteriormente:

```
>> a = [0; 3; 2]
>> Ha = im_house(v,a)
```

Obsérvese que $H\mathbf{a}$ coincide con $\mathbf{a}$ ¿por qué?

4. Factorización QR

La siguiente función de MATLAB calcula las matrices Q y R del teorema IX.4.5 tales que $A = QR$, de una matriz A dada.

```
function [Q,R] = mi_qr(A)
% ESCRIBE TÚ LA AYUDA

  [m,n] = size(A);
  Q = eye(m);
  R = A;
  for i = 1:n-1
    H = eye(m);
    v = R(i:m,i);
    [w,beta] = vector_householder(v);
    if beta == 0
      H = H;
    else
      H(i:m,i:m) = eye(m-i+1) - w*w'/beta;
    end
    R = H*R;
    Q = Q*H;
  end
```

Comprobemos nuestro algoritmo con la matriz

$$A = \begin{pmatrix} 4 & -1 & -1 & 0 \\ -1 & 4 & 0 & -1 \\ -1 & 0 & 4 & -1 \\ 0 & -1 & -1 & 4 \end{pmatrix}$$

```
>> A = [4, -1, -1, 0; -1, 4, 0, -1; -1, 0, 4, -1; 0, -1, -1, 4]
>> [Q,R] = mi_qr(A)
```

MATLAB tiene una rutina muy eficiente para calcular la factorización QR de una matriz.

```
>> help qr
```

4.1. Rendimiento.

Un hecho destacable del método de Householder es que el condicionamiento para la norma matricial euclídea de la matriz de partida no se ve modificado;

$$\mathrm{cond}_2(A) = \mathrm{cond}_2(A_k),\ k \geq 1;$$

ya que el $\mathrm{cond}_2(-)$ es invariante por transformaciones ortogonales (unitarias).

```
>> A = round(10*rand(50)-5);
>> [Q,R] = mi_qr(A);
>> cond(A)
>> cond(R)
```

Esto es una ventaja, del método de Householder respecto del método de eliminación gaussiana, desde el punto de vista de la "estabilidad numérica" compensada, sin embargo, por un mayor número (prácticamente el doble) de operaciones elementales con la consecuente propagación de errores de redondeo.

```
>> A = round(10*rand(50)-5);
>> tic;rref(A);toc
>> tic;lu(A);toc
>> tic;[Q,R] = mi_qr(A);toc
>> tic;[Q,R] = qr(A);toc
```

El método de Householder permite calcular de forma muy simple el determinante de la matriz A. En efecto, el determinante de una matriz de Householder es ± 1, de modo que

$$\det(A) = (-1)^r a_{11}^{(1)} a_{22}^{(2)} \cdots a_{nn}^{(n)},$$

siendo r el número de matrices de Householder utilizadas distintas de las unidad.

```
>> A = round(10*rand(10)-5);
>> [Q,R] = mi_qr(A);
>> det(A)
```

```
>> prod(diag(R))
```

Terminamos esta práctica mostrando que la propagación de errores de redondeo es similar si usamos la factorización LU o la factorización QR para resolver un sistema de ecuaciones lineales mal condicionado.

Consideremos los sistemas lineales $A_n\mathbf{x}_n = \mathbf{b}_n$ donde $A_n \in \mathcal{M}_n(\mathbb{R})$ es la matriz de Hilbert de orden n mientras que $\mathbf{b}_n$ se elige de tal forma que la solución exacta del sistema sea $\mathbf{u}_n = (1, 1, \ldots, 1)^{\mathrm{t}}$. La matriz A_n es claramente simétrica y se puede comprobar que es definida positiva.

Para $n = 1, \ldots, 100$, utilizamos las funciones `lu` y `qr` para factorizar la matriz A_n. Entonces, resolvemos los sistemas lineales asociados (mediante las sustitución hacia adelante y hacia atrás) y denotamos por $\mathbf{u} + \delta\mathbf{u}$ la solución calculada. En la figura que resulta recogemos (en escala semilogarítmica) los errores relativos

$$E_n = \|\delta\mathbf{u}_n\|_2/\|\mathbf{u}_n|_2$$

en cada caso.

```
>> warning('off')
>> close all
>> E1_n = [];
>> E2_n = [];
>> for n = 1:100
>>   clear b xx;
>>   x = ones(n,1);
>>   for i = 1:n
>>     b(i) =  sum(1./(i+(1:n)-1));
>>   end
>>   A_n =  hilb(n);
>>   [L,U,P] = lu(A_n);
>>   y = sust_adelante(L,P*b');
>>   xx = sust_atras(U,y);
>>   E1_n = [E1_n, norm(x-xx)/norm(x)];
>>   [Q,R] = qr(A_n);
>>   xx = sust_atras(R,Q'*b');
>>   E2_n = [E2_n, norm(x-xx)/norm(x)];
>> end
>> semilogy(1:100,E1_n,'r')
>> hold on
```

```
>> semilogy(1:100,E2_n)
>> legend('Error relativo con LU','Error relativo con QR')
>> warning('on')
```

Ejercicios de la práctica 12

Ejercicio 1. Calcular, si es posible, la factorización de Cholesky de la siguiente matriz

$$A = \begin{pmatrix} 2 & 1 & 0 & 0 \\ 1 & 4 & 1 & 0 \\ 0 & 1 & 4 & 1 \\ 0 & 0 & 1 & 2 \end{pmatrix}.$$

Comparar la factorización obtenida con su factorización LU.

Ejercicio 2. Sea

$$A = \begin{pmatrix} -5 & 2 & 2 & -1 & 4 \\ 2 & -1 & -2 & 4 & 3 \\ -1 & -2 & 0 & 4 & 1 \\ 3 & -3 & -3 & 1 & 3 \\ 0 & -3 & 2 & 0 & 2 \end{pmatrix}.$$

Calcular las matrices de Householder H_1, H_2, H_3 y H_4 tales que

$$H_4 H_3 H_2 H_1 A$$

es triangular superior.

Ejercicio 3. Usa las descomposiciones LU y QR para resolver el siguiente sistema:

$$\left.\begin{array}{rrrcr} x_1 & +1/2\, x_2 & +1/3\, x_3 & = & 6 \\ 1/2\, x_1 & +1/3\, x_2 & +1/4\, x_3 & = & 4 \\ 1/3\, x_1 & +1/4\, x_2 & +1/5\, x_3 & = & 15 \end{array}\right\}$$

Interpreta los resultados obtenidos.

Ejercicio 4. Modifica debidamente la función `mi_qr` para determinar cuantas matrices de Householder distintas de la identidad se han usado. Usando esta modificación, define una función de `MATLAB` que calcule el determinante de una matriz cuadrada.

Ejercicio 5. Define una matriz aleatoria de orden 3×5 con entradas enteras entre -10 y 10. ¿Se puede calcular una descomposición QR de esta matriz? Compruébalo con `MATLAB` y explica el resultado.

Ejercicio 6. Estudia el comportamiento de la descomposición QR para matrices dispersas (es decir, aquellas que tiene un número elevado de entradas nulas).

APÉNDICE A

Conceptos topológicos fundamentales

1. Espacios Métricos

Definición A.1.1. Sea X un conjunto no vacío. Una aplicación $\mathrm{d} : X \times X \to \mathbb{R}$ es una **distancia** (o **aplicación distancia**) sobre X, si para todo x, y y $z \in X$ verifica los axiomas siguientes:

(a) (Definida positiva) $\mathrm{d}(x, y) \geq 0$; además, $\mathrm{d}(x, y) = 0$ si, y sólo si, $x = y$.
(b) (Simetría) $\mathrm{d}(x, y) = \mathrm{d}(y, x)$.
(c) (Desigualdad triangular) $\mathrm{d}(x, z) \leq \mathrm{d}(x, y) + \mathrm{d}(y, z)$.

El número real $\mathrm{d}(x, y)$ recibe el nombre de **distancia** de x a y.

Nótese que (a) establece que la distancia de un elemento de X a otro elemento de X nunca es negativa, y es cero únicamente cuando ambos elementos son iguales, en particular, la distancia de un elemento a sí mismo es cero, y recíprocamente. El axioma (b) establece que la distancia de un elemento de $x \in X$ a un elemento $y \in X$ es la misma que la distancia de y a x, por esta razón $\mathrm{d}(x, y)$ se lee distancia entre x e y.

El axioma (c) se conoce desigualdad triangular porque si x, y y z son tres puntos de plano $\mathbb{R}^2$, entonces (c) establece que la longitud $\mathrm{d}(x, z)$ de uno de los lados del triángulo de vértices x, y y z es menor o igual que la suma $\mathrm{d}(x, y) + \mathrm{d}(y, z)$ de las longitudes de los otros dos lados del triángulo.

Veamos, a continuación, algunos ejemplos de distancias. Que estos ejemplos verifican realmente los axiomas requeridos se propone como ejercicio al lector.

Ejemplos A.1.2.

i) **Distancia discreta**. Sean X un conjunto no vacío y $\mathrm{d} : X \times X \to \mathbb{R}$ tal que

$$\mathrm{d}(x, y) = \begin{cases} 0 & \text{si } x = y; \\ 1 & \text{si } x \neq y. \end{cases}$$

ii) La aplicación $\mathrm{d}(x, y) = |x - y|$, donde x e y son números reales, es un distancia llamada **distancia usual de la recta real** $\mathbb{R}$. Además, la aplicación d definida por

$$\mathrm{d}(\mathbf{u}, \mathbf{v}) = \sqrt{(u_1 - v_1)^2 + (u_2 - v_2)^2}$$

donde $\mathbf{u} = (u_1, u_2)$ y $\mathbf{v} = (v_1, v_2)$ están en $\mathbb{R}^2$, es una distancia llamada distancia usual de $\mathbb{R}^2$. En general, la aplicación $\mathrm{d} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ definida por

$$\mathrm{d}(\mathbf{u}, \mathbf{v}) = \left(\sum_{i=1}^{n} |u_i - v_i|^2 \right)^{1/2},$$

donde $\mathbf{u} = (u_1, u_2, \ \ldots, \ u_n)$ y $\mathbf{v} = (v_1, v_2, \ \ldots, \ v_n)$, es una distancia llamada **distancia usual** de $\mathbb{R}^n$.

iii) En $\mathbb{R}^n$ se pueden definir otras distancias distintas de la usual; por ejemplo, las aplicaciones d definidas como sigue son distancias sobre $\mathbb{R}^n$

$$\mathrm{d}(\mathbf{u}, \mathbf{v}) = \sum_{i=1}^{n} |u_i - v_i|$$

$$\mathrm{d}(\mathbf{u}, \mathbf{v}) = \left(\sum_{i=1}^{n} |u_i - v_i|^p \right)^{1/p}, \quad p \geq 1.$$

$$\mathrm{d}(\mathbf{u}, \mathbf{v}) = \text{máx} \left\{ |u_i - v_i|, \ i = 1, \ \ldots, \ n \right\}.$$

iv) En $\mathcal{C}[0,1] = \{f : [0,1] \to \mathbb{R} \text{ continuas}\}$, se puede definir una distancia de la manera siguiente:

$$\mathrm{d}(f, g) = \int_0^1 \Big| f(x) - g(x) \Big| \, \mathrm{d}x.$$

Asimismo, se pueden definir las dos distancias siguientes

$$\mathrm{d}(f, g) = \left(\int_0^1 \Big| f(x) - g(x) \Big|^p \, \mathrm{d}x \right)^{1/p}, \ p \geq 1$$

y

$$\mathrm{d}(f, g) = \max_{x \in [0,1]} \Big| f(x) - g(x) \Big|$$

Definición A.1.3. Un **espacio métrico** es un par (X, d) formado por un conjunto no vacío X y una distancia sobre X.

Nótese que un mismo conjunto podemos definir varias distancias; por lo que, en un espacio métrico, tan importante es el conjunto como la distancia definida.

Nota A.1.4. Obsérvese que si (X, d) es un espacio métrico e Y es un subconjunto de X, entonces la restricción de d a $Y \times Y$ define una estructura natural de espacio métrico en Y.

Proposición A.1.5. *Sea* (X, d) *un espacio métrico. Entonces*

$$\Big| \mathrm{d}(x, z) - \mathrm{d}(y, z) \Big| \leq \mathrm{d}(x, y).$$

Demostración. Por la desigualdad triangular, $\mathrm{d}(x,z) \leq \mathrm{d}(x,y) + \mathrm{d}(y,z)$; por tanto, $\mathrm{d}(x,z) - \mathrm{d}(y,z) \leq \mathrm{d}(x,y)$. Intercambiando el papel de x e y, obtenemos que $\mathrm{d}(y,z) - \mathrm{d}(x,z) \leq \mathrm{d}(y,x)$, esto es, $-\mathrm{d}(x,y) \leq \mathrm{d}(x,z) - \mathrm{d}(y,z)$. En resumen,

$$-\mathrm{d}(x,y) \leq \mathrm{d}(x,z) - \mathrm{d}(y,z) \leq \mathrm{d}(x,y),$$

de donde se sigue la desigualdad buscada. ∎

Topología métrica.

Definición A.1.6. Sean (X, d) un espacio métrico, $x \in X$ y ε un número real positivo. Llamaremos **bola abierta de centro x y radio ε** al conjunto

$$B(x,\varepsilon) := \{y \in X \mid \mathrm{d}(x,y) < \varepsilon\}.$$

Llamaremos **bola cerrada de centro x y radio ε** al conjunto

$$B[x,\varepsilon] := \{y \in X \mid \mathrm{d}(x,y) \leq \varepsilon\}.$$

Ejemplos A.1.7. Veamos los ejemplos bolas abiertas en $\mathbb{R}^2$ para las distancias más comunes.

i) Si $\mathrm{d}(\mathbf{v},\mathbf{u}) = \sqrt{|v_1-u_1|^2 + |v_2-u_2|^2}$, con $\mathbf{v} = (v_1,v_2)$ y $\mathbf{u} = (u_1,u_2) \in \mathbb{R}^2$, entonces

$$\begin{aligned} B(\mathbf{v},\varepsilon) &= \{\mathbf{u} \in \mathbb{R}^2 \mid \mathrm{d}(\mathbf{v},\mathbf{u}) < \varepsilon\} \\ &= \{\mathbf{u} \in \mathbb{R}^2 \mid \sqrt{|v_1-u_1|^2 + |v_2-u_2|^2} < \varepsilon\} \\ &= \{\mathbf{u} \in \mathbb{R}^2 \mid |v_1-u_1|^2 + |v_2-u_2|^2 < \varepsilon^2\}. \end{aligned}$$

Esto es, el círculo (sin borde) de centro $\mathbf{u}$ y radio ε.

ii) Si $\mathrm{d}(\mathbf{v},\mathbf{u}) = \mathrm{máx}\{|v_1-u_1|, |v_2-u_2|\}$, con $\mathbf{v} = (v_1,v_2)$ y $\mathbf{u} = (u_1,u_2) \in \mathbb{R}^2$, entonces

$$\begin{aligned} B(\mathbf{0},1) &= \{\mathbf{u} \in \mathbb{R}^2 \mid \mathrm{d}(\mathbf{0},\mathbf{u}) < 1\} \\ &= \{\mathbf{u} \in \mathbb{R}^2 \mid \mathrm{máx}\{|u_1|, |u_2|\} < 1\} \\ &= \{\mathbf{u} \in \mathbb{R}^2 \mid u_1, u_2 \in (-1,1)\}. \end{aligned}$$

Esto es, el cuadrado (sin borde) de vértices $(1,1), (-1,1), (-1,-1)$ y $(1,-1)$.

iii) Si $\mathrm{d}(\mathbf{v},\mathbf{u}) = |v_1-u_1| + |v_2-u_2|$, con $\mathbf{v} = (v_1,v_2)$ y $\mathbf{u} = (u_1,u_2) \in \mathbb{R}^2$,

$$\begin{aligned} B(\mathbf{0},1) &= \{\mathbf{u} \in \mathbb{R}^2 \mid \mathrm{d}(\mathbf{0},\mathbf{u}) < 1\} \\ &= \{\mathbf{u} \in \mathbb{R}^2 \mid |u_1| + |u_2| < 1\}. \end{aligned}$$

Esto es, el cuadrado (sin borde) de vértices $(1,0), (0,1), (-1,0)$ y $(0,-1)$.

Definición A.1.8. Sea (X, d) un espacio métrico. Un subconjunto A de X es un **entorno** de un elemento $x \in X$ si existe una bola abierta centrada en x contenida en A, es decir, si existe $\varepsilon > 0$ tal que $B(x, \varepsilon) \subseteq A$.

Obsérvese que toda bola abierta contiene bolas cerradas del mismo centro y radio menor, y que toda bola cerrada contiene bolas abiertas del mismo centro y radio menor.

Definición A.1.9. Sea (X, d) un espacio métrico. Un subconjunto U de X se dice **abierto** cuando para cada $x \in U$ existe $\varepsilon > 0$ (que depende de x) tal que

$$B(x, \varepsilon) \subseteq U.$$

Luego, si U es un abierto de un espacio métrico (X, d), para cada punto de U se puede encontrar una bola abierta centrada en él contenida en U, dicho de otro modo, *U es entorno de todos sus puntos.*

Ejemplos A.1.10.

i) Las bolas abiertas de un espacio métrico son subconjuntos abiertos.
ii) En $\mathbb{R}$ con la distancia usual, los intervalos abiertos son subconjuntos abiertos.
iii) En cualquier conjunto X con la distancia discreta, cualquier punto $x \in X$ es un abierto, ya que $B(x, 1/2) = \{x\}$.

Propiedades de los subconjuntos abiertos de un espacio métrico. Sea (X, d) un espacio métrico.

(a) El conjunto vacío, $\varnothing$, y el total, X, son abiertos.
(b) La unión arbitraria de abiertos es un abierto, es decir, si $\{U_i\}_{i \in I}$ es una familia arbitraria de abiertos, entonces $\cup_{i \in I} U_i$ es abierto.
(c) La intersección finita de abiertos es un abierto, es decir, si $\{U_1, \ldots, U_n\}$ es una familia finita de abiertos, entonces $\cap_{i=1}^{n} U_i$ es abierto.

Demostración. La demostración de estas propiedades se deja como ejercicio al lector. ■

Definición A.1.11. Sea X un conjunto no vacío. Un clase $\mathcal{T}$ de subconjuntos de X es una **topología** en X si $\mathcal{T}$ verifica los axiomas siguientes.

(a) $\varnothing$ y X pertenecen a $\mathcal{T}$.
(b) La unión arbitraria de conjuntos de $\mathcal{T}$ pertenece a $\mathcal{T}$.
(c) La intersección de un número finito de conjuntos de $\mathcal{T}$ pertenece a $\mathcal{T}$.

Los elementos de $\mathcal{T}$ se llaman **conjuntos abiertos de la topología** $\mathcal{T}$ y el par $(X, \mathcal{T})$ se llama **espacio topológico**.

De la propiedades de los subconjuntos abiertos de un espacio métrico, se deduce que todo espacio métrico (X, d) tiene una estructura natural de espacio topológico, aquella que define la topología $\mathcal{T}$ formada por los abiertos de (X, d) que llamaremos **topología métrica**.

Definición A.1.12. Un espacio topológico $(X, \mathcal{T})$ es un **espacio de Hausdorff** si dados dos puntos cualesquiera x e $y \in X$ distintos, existen conjuntos abiertos U y $V \in \mathcal{T}$ tales que

$$x \in U,\ y \in V \quad \text{y} \quad U \cap V = \varnothing.$$

Proposición A.1.13. *Todo espacio métrico es de Hausdorff.*

Demostración. Sean (X, d) un espacio métrico y x e $y \in X$ dos puntos distintos; luego, de acuerdo con el axioma (a) de la definición de espacio métrico, $\mathrm{d}(x, y) = \varepsilon > 0$. Consideremos las bolas abiertas $U = B(x, \varepsilon/3)$ y $V = B(y, \varepsilon/3)$ y veamos que son disjuntas. En efecto, si $z \in U \cap V$, entonces $\mathrm{d}(x, z) < \varepsilon/3$ y $\mathrm{d}(z, y) < \varepsilon/3$, de donde se sigue que

$$\mathrm{d}(x, y) \leq \mathrm{d}(x, z) + \mathrm{d}(z, y) < \varepsilon/3 + \varepsilon/3 = 2\,\varepsilon/3,$$

lo que supone una contradicción. Por tanto, U y V son abiertos disjuntos tales que $x \in U$ e $y \in V$. ∎

Nota A.1.14. Aunque todos los espacios topológicos que consideraremos en esta asignatura serán espacios métricos, conviene advertir al lector que no todos los espacios topológicos son métricos. Por ejemplo, sean $X = \{0, 1\}$ y $\mathcal{T} = \{\varnothing, \{0\}, X\}$, el par $(X, \mathcal{T})$ es un espacio topológico, llamado **espacio topológico de Sierpinski**,en el que no se puede definir ninguna distancia.

Definición A.1.15. Sea (X, d) un espacio métrico. Un subconjunto $F \subseteq X$ se dice **cerrado** cuando su complementario, $X \setminus F$ es abierto.

Ejemplos A.1.16.

i) Las bolas cerradas de un espacio métrico son subconjuntos cerrados.
ii) En $\mathbb{R}$ con la distancia usual, los intervalos cerrados son subconjuntos cerrados.
iii) En cualquier conjunto X con la distancia discreta, cualquier punto $x \in X$ es un cerrado, ya que $B[x, 1/2] = \{x\}$.

Propiedades de los subconjuntos cerrados de un espacio métrico. Sea (X, d) un espacio métrico.

(a) El conjunto vacío, $\varnothing$, y el total, X, son cerrados.
(b) La unión finita de cerrados es un cerrado.
(c) La intersección arbitraria de cerrados es un cerrado.

Demostración. La demostración de estas propiedades se deja como ejercicio al lector. ■

Definición A.1.17. Sean (X, d) un espacio métrico y A un subconjunto de X.

- Un **elemento** $x \in A$ es **interior** de A cuando existe una bola de centro x y radio $\varepsilon > 0$ contenida en A, equivalentemente, si A es un entorno de x.
- El **interior** de A es el conjunto formado por todos sus puntos interiores

$$\mathrm{int}(A) := \{x \in X \mid B(x, \varepsilon) \subseteq A, \text{ para algún } \varepsilon > 0\}.$$

- Un **elemento** $x \in A$ es **adherente** a A cuando toda bola de centro x corta a A.
- la **clausura** de A es el conjunto de sus puntos adherentes,

$$\bar{A} := \{x \in X \mid B(x, \varepsilon) \cap A \neq \varnothing, \text{ para todo } \varepsilon > 0\}.$$

- Un **elemento** x está en la **frontera** de A cuando toda bola de centro x corta A y a su complementario $X \setminus A$.
- La **frontera** de A es el conjunto de sus puntos frontera

$$\mathrm{Fr}(A) := \{x \in X \mid B(x, \varepsilon) \cap A \neq \varnothing \text{ y } B(x, \varepsilon) \cap (X \setminus A) \neq \varnothing, \text{ para todo } \varepsilon > 0\}.$$

- Un elemento x es un **punto de acumulación** de A cuando toda bola de centro x corta a $A \setminus \{x\}$. El conjunto de puntos de acumulación de A se denota por A'.

Proposición A.1.18. *Sean* (X, d) *un espacio métrico y* A *un subconjunto de* X. *Se verifica que:*

(a) $\mathrm{int}(A) \subseteq A \subseteq \bar{A}$.
(b) *Si* $A \subseteq B$, *entonces* $\mathrm{int}(A) \subseteq \mathrm{int}(B)$ *y* $\bar{A} \subseteq \bar{B}$.
(c) A *es abierto si, y sólo si,* $A = \mathrm{int}(A)$.
(d) A *es cerrado si, y sólo si,* $A = \bar{A}$.
(e) $\mathrm{int}(A)$ *es el mayor abierto contenido en* A.
(f) $\bar{A}$ *es el menor cerrado que contiene a* A.
(g) $X \setminus \bar{A} = \mathrm{int}(X \setminus A)$.
(h) $\mathrm{Fr}(A) = \bar{A} \setminus \mathrm{int}(A)$.

Demostración. La demostración de esta proposición se deja como ejercicio al lector. ■

2. Sucesiones y continuidad

Sea X un conjunto. Usaremos la notación $(x_n)_{n\in\mathbb{N}}$ o $(x_1, x_2, \ldots)$ (o simplemente (x_n) cuando no exista posibilidad de confusión) para denotar la **sucesión** de elementos de X cuyo n-ésimo término es x_n y $\{x_n \mid n \in \mathbb{N}\}$ para denotar el conjunto de todos los elementos de la sucesión. Nótese que $\{x_n \mid n \in \mathbb{N}\}$ puede ser finito aunque la sucesión $(x_n)_{n\in\mathbb{N}}$ sea infinita.

Dada una sucesión $(x_n)_{n\in\mathbb{N}}$ y M un subconjunto infinito de $\mathbb{N}$, diremos que la sucesión $(x_m)_{m\in M}$ es una **subsucesión** de la primera.

Definición A.2.1. Sea (X, d) un espacio métrico. Un elemento $x \in X$ es un **valor de adherencia de una sucesión** (x_n) de elementos de X, si en cada bola de centro x hay infinitos términos de la sucesión.

Definición A.2.2. Diremos que una sucesión $(x_n)_{n\in\mathbb{N}}$ de elementos de un espacio métrico (X, d) **converge** a $x \in X$, y lo denotaremos $\lim_{n\to\infty} x_n = x$, si

$$\text{para cada } \varepsilon > 0 \text{ existe } N \in \mathbb{N} \text{ tal que } x_n \in B(x, \varepsilon) \text{ para todo } n \geq N,$$

es decir, cuando para cada bola de centro x existe un subíndice a partir de cual los términos de la sucesión "quedan dentro" de la bola.

En general, el concepto de convergencia depende de la distancia que determina la estructura métrica.

Nótese que el límite de una sucesión es un valor de adherencia. Aunque no al contrario, una sucesión puede tener valor de adherencia y no ser convergente; considérese, por ejemplo, la sucesión de números reales $x_n = (-1)^n,\ n \in \mathbb{N}$.

Proposición A.2.3. *Sean (X, d) un espacio métrico. El límite de $(x_n)_{n\in\mathbb{N}}$ una sucesión de elementos de X, si existe, es único*

Demostración. Supongamos que existen x e $y \in X$ distintos, tales que $\lim_{n\to\infty} x_n = x$ y $\lim_{n\to\infty} x_n = y$. Como, por la proposición A.1.13, (X, d) es un espacio Haussdorff, existen dos abiertos disjuntos U y V tales que $x \in U$ e $y \in V$. Por consiguiente, existen dos bolas abiertas disjuntas $B(x, \varepsilon)$ y $B(y, \varepsilon')$; lo que es del todo imposible ya que para N suficientemente grande $x_n \in B(x, \varepsilon)$ y $x_n \in B(y, \varepsilon')$, para todo $n \geq N$, por ser x e y límites de la sucesión $(x_n)_{n\in\mathbb{N}}$. ∎

Veamos ahora que los conjuntos cerrados de un espacio métrico se pueden caracterizar usando sucesiones.

Proposición A.2.4. *Sean (X, d) un espacio métrico y A un subconjunto de X.*

(a) *$x \in \bar{A}$ si, y sólo si, existe una sucesión de elementos de A que converge a x.*

(b) *A es cerrado si, y sólo si, cualquier sucesión convergente de elementos de A converge a un elemento de A.*

Demostración. (a) Si $x \in \bar{A}$, entonces x es un punto adherente a A, es decir, cualquier bola de centro de x corta A. Por consiguiente, para cada $n \in \mathbb{N}$, la intersección $B(x, 1/n) \cap A$ no es vacía. Por lo que podemos tomar un elemento $x_n \in B(x, 1/n) \cap A$, para cada $n \in \mathbb{N}$, y construir de este modo una sucesión, $(x_n)_{n\in\mathbb{N}}$ de elementos de A convergente a x. El recíproco se sigue de las definiciones de convergencia y de punto adherente.

(b) Si $(x_n)_{n\in\mathbb{N}} \subseteq A$ es una sucesión convergente a $x \in X$, entonces toda bola de centro x contiene (infinitos) términos de la sucesión, en particular, corta a A. Luego, $x \in \bar{A}$ y por ser A cerrado concluimos que $x \in A$. Recíprocamente, si $x \in \bar{A}$, por el apartado anterior, existe una sucesión en A que converge a x; luego, por hipótesis, $x \in A$ y concluimos que A es cerrado. ∎

Proposición A.2.5. *Sean (X, d) un espacio métrico, $(x_n)_{n\in\mathbb{N}}$ e $(y_n)_{n\in\mathbb{N}}$ dos sucesiones de elementos de X y x e $y \in X$. Entonces*

$$\left.\begin{array}{l} \lim_{n\to\infty} x_n = x \\ \lim_{n\to\infty} y_n = y \end{array}\right\} \Longrightarrow \lim_{n\to\infty} \mathrm{d}(x_n, y_n) = \mathrm{d}(x, y).$$

Demostración. Usando la proposición A.1.5 y la desigualdad triangular del valor absoluto,

$$\begin{aligned} |\mathrm{d}(x, y) - \mathrm{d}(x_n, y_n)| &\leq |\mathrm{d}(x, y) - \mathrm{d}(x_n, y)| + |\mathrm{d}(x_n, y) - \mathrm{d}(x_n, y_n)| \\ &\leq \mathrm{d}(x, x_n) + \mathrm{d}(y, y_n) \end{aligned}$$

que tiende a cero cuando n tiende hacia infinito. ∎

Definición A.2.6. Una **aplicación** $f : (X, \mathrm{d}) \to (Y, \mathrm{d}')$ entre dos espacios métricos se dice que es **continua en un elemento** $x \in X$, cuando

$$\text{para cada } \varepsilon > 0, \text{ existe } \delta > 0 \text{ tal que } \mathrm{d}(x, y) < \delta \text{ implica que } \mathrm{d}'(f(x), f(y)) < \varepsilon,$$

equivalentemente, si para cada $\varepsilon > 0$ existe $\delta > 0$ tal que $y \in B(x, \delta)$ implica $f(y) \in B(f(x), \varepsilon)$, es decir, $f(B(x, \delta)) \subseteq B(f(x), \varepsilon)$.

Nótese que δ depende tanto de ε como de x.

El concepto de continuidad de una aplicación en un punto es local. Se trata, intuitivamente, de que la aplicación conserve la noción de proximidad en torno a x.

Definición A.2.7. Una **aplicación** $f : (X, \mathrm{d}) \to (Y, \mathrm{d}')$ entre dos espacios métricos se dice que es **continua**, cuando es continua en cada elemento de X.

Proposición A.2.8. *Una aplicación $f : (X, \mathrm{d}) \to (Y, \mathrm{d}')$ entre dos espacios métricos es continua si, y sólo si, la imagen inversa de un abierto es un abierto.*

Demostración. Sea $U \subseteq Y$ un abierto, se trata de demostrar que $f^{-1}(U)$ es un abierto de X, es decir, que $f^{-1}(U)$ es entorno de cada uno de sus puntos. Sea $x \in f^{-1}(U)$, entonces $f(x) \in U$. Luego, existe $\varepsilon > 0$ tal que $B((f(x), \varepsilon) \subseteq U$. Ahora, por ser f continua, existe $\delta > 0$ tal que $f(B(x, \delta)) \subseteq B(f(x), \varepsilon) \subseteq U$. De donde se sigue que $B(x, \delta) \subseteq f^{-1}(U)$.

Recíprocamente, veamos que f es continua en $x \in X$. Para cada $\varepsilon > 0$, $B(f(x), \varepsilon)$ es un abierto de Y. Luego, $f^{-1}(B(f(x), \varepsilon))$ es un abierto de X que contiene a x. Por consiguiente, existe $\delta > 0$ tal que $B(x, \delta) \subseteq f^{-1}(B(f(x), \varepsilon))$, y concluimos que $f(B(x, \delta)) \subseteq B(f(x), \varepsilon)$. ∎

Otras caracterizaciones del concepto de concepto de continuidad son las siguientes:

- Una aplicación $f : (X, \mathrm{d}) \to (Y, \mathrm{d}')$ entre dos espacios métricos es continua si, y sólo si, la imagen inversa de un cerrado es un cerrado.
- Una aplicación $f : (X, \mathrm{d}) \to (Y, \mathrm{d}')$ entre dos espacios métricos es continua si, y sólo si, para todo subconjunto A de X se cumple que $f(\bar{A}) \subseteq \overline{f(A)}$.

Teorema A.2.9. *La composición de aplicaciones continuas es continua.*

Demostración. Sean $f : (X, \mathrm{d}) \to (Y, \mathrm{d}')$ y $g : (Y, \mathrm{d}') \to (Z, \mathrm{d}'')$ dos aplicaciones continuas entre espacios métricos. Si $U \subseteq Z$ es un abierto, entonces $g^{-1}(U)$ es un abierto en Y y $f^{-1}(g^{-1}(U))$ es un abierto en X. De donde se sigue que $(g \circ f)^{-1}(U) = f^{-1}(g^{-1}(U))$ es un abierto. ∎

Proposición A.2.10. *Una aplicación continua entre espacios métricos transforma sucesiones convergentes en sucesiones convergentes.*

Demostración. Sean $f : (X, \mathrm{d}) \to (Y, \mathrm{d}')$ una aplicación continua entre espacios métricos y sea $(x_n)_{n\in\mathbb{N}}$ una sucesión convergente de elementos de X, por ejemplo, $\lim_{n\to\infty} x_n = x \in X$.

Dado $\varepsilon > 0$, existe $\delta > 0$ tal que

$$\mathrm{d}(x, y) < \delta \Rightarrow \mathrm{d}'(f(x), f(y)) < \epsilon,$$

por ser f continua. Por otra parte, al ser $(x_n)_{n\in\mathbb{N}}$ convergente, existe $N \in \mathbb{N}$ tal que, para todo $n \geq N$, $\mathrm{d}(x, x_n) < \delta$; de donde se sigue que

$$\mathrm{d}'(f(x), f(x_n)) < \varepsilon,$$

y se concluye que la sucesión $(f(x_n))_{n\in\mathbb{N}}$ es convergente a $f(x)$. ∎

Definición A.2.11. Una aplicación $f : (X, \mathrm{d}) \to (Y, \mathrm{d}')$ entre dos espacios métricos

(a) es **abierta** si lleva abiertos en abiertos, es decir, si para todo abierto $U \subseteq X,\ f(U)$ es un abierto.
(b) es **cerrada** si lleva cerrados en cerrados, es decir, si para todo cerrado $F \subseteq X,\ f(F)$ es un cerrado.
(c) es un **homeomorfismo** si es biyectiva y tanto f como f^{-1} son continuas.

3. Sucesiones de Cauchy. Completitud

Definición A.3.1. Una **sucesión** $(x_n)_{n\in\mathbb{N}}$ en un espacio métrico (X, d) se dice que es **de Cauchy** si

$$\text{para cada } \varepsilon > 0 \text{ existe } n_0 \in \mathbb{N} \text{ tal que } n, m > n_0 \text{ implica que } \mathrm{d}(x_n, x_m) < \varepsilon,$$

es decir, si se pueden encontrar dos términos de la sucesión tan próximos como se quiera.

Nótese que toda sucesión convergente es de Cauchy[1], pero el recíproco no es cierto. Por ejemplo la sucesión de término general $x_n = (1 + 1/n)^n$ en el espacio métrico $\mathbb{Q}$ con la distancia usual (es decir, el valor absoluto) es de Cauchy, aunque no es convergente pues su "límite" sería el número e que no es racional.

Ejemplo A.3.2. Sea $(\mathbf{v}_m)_{m\in\mathbb{N}}$ una sucesión de Cauchy en $\mathbb{R}^n$ con la distancia usual; por ejemplo,

$$\mathbf{v}_1 = (v_1^{(1)}, v_2^{(1)},\ \ldots,\ v_n^{(1)}),\ \ldots,\ \mathbf{v}_m = (v_1^{(m)}, v_2^{(m)},\ \ldots,\ v_n^{(m)}),\ \ldots$$

Las proyecciones de los vectores $\mathbf{v}_m,\ m \in \mathbb{N}$, en cada uno de los n subespacios coordenados, es decir,

$$(v_1^{(m)})_{m\in\mathbb{N}},\ \ldots,\ (v_n^{(m)})_{m\in\mathbb{N}} \tag{A.3.1}$$

son sucesiones de Cauchy en $\mathbb{R}$. En efecto, para cada $\varepsilon > 0$, puesto que $(\mathbf{v}_m)_{m\in\mathbb{N}}$ es de Cauchy, existe $m_0 \in \mathbb{N}$ tal que si i y j son mayores que m_0, entonces

$$\mathrm{d}(\mathbf{v}_i, \mathbf{v}_j)^2 = |v_i^{(1)} - v_j^{(1)}|^2 + \ldots + |v_i^{(m)} - v_j^{(m)}|^2 < \varepsilon^2.$$

Luego, en particular, si i y j son mayores que m_0, entonces

$$|v_i^{(1)} - v_j^{(1)}|^2 < \varepsilon^2,\ \ldots,\ |v_i^{(m)} - v_j^{(m)}|^2 < \varepsilon^2.$$

[1]Sea $(x_n)_{n\in\mathbb{N}}$ una sucesión convergente en un espacio métrico (X, d); por ejemplo, $\lim_{n\to\infty} x_n = x \in X$. Entonces, $(x_n)_{n\in\mathbb{N}}$ es necesariamente una sucesión de Cauchy porque, para todo $\varepsilon > 0$, existe $n_0 \in \mathbb{N}$ tal que $n \geq n_0$ implica que $\mathrm{d}(x_n, x) < 1/2\,\varepsilon$. Luego, por la desigualdad triangular, dados n y m mayores que n_0, se cumple que

$$\mathrm{d}(x_n, x_m) \leq \mathrm{d}(x_n, x) + \mathrm{d}(x_m, x) < 1/2\varepsilon + 1/2\varepsilon = \varepsilon.$$

En otras palabras, $(x_n)_{n\in\mathbb{N}}$ es una sucesión de Cauchy.

En otras palabras, cada una de las m sucesiones dadas en (A.3.1) es una sucesión de Cauchy.

Lema A.3.3. *Sea* (X, d) *un espacio métrico. Toda sucesión de Cauchy de elementos de* X *con un valor de adherencia es convergente.*

Demostración. Sea $(x_n)_{n\in\mathbb{N}} \subset X$ una sucesión de Cauchy y $x \in X$ un valor de adherencia de la sucesión. Veamos que $\lim_{n\to\infty} x_n = x$. Sea $\varepsilon > 0$. Por ser $(x_n)_{n\in\mathbb{N}}$ un sucesión de Cauchy, existe $n_0 \in \mathbb{N}$ tal que

$$\mathrm{d}(x_n, x_m) < \varepsilon/2,$$

para todo $n, m \geq n_0$. Por otra parte, al ser x un valor de adherencia de la sucesión, existe $N \geq n_0$ tal que $x_N \in B(x, \varepsilon/2)$. De ambos hechos se sigue que, para todo $n \geq N$,

$$\mathrm{d}(x_n, x) \leq \mathrm{d}(x_n, x_N) + \mathrm{d}(x_N, x) < \varepsilon/2 + \varepsilon/2 = \varepsilon.$$

Luego, la sucesión es convergente al valor de adherencia. ∎

Definición A.3.4. Un espacio métrico (X, d) es **completo** si toda sucesión de Cauchy $(x_n)_{n\in\mathbb{N}}$ de elementos de X converge a un elemento de X.

Ejemplos A.3.5.

i) Veamos que $\mathbb{R}$ con la distancia usual, es decir, con el valor absoluto de la diferencia, es un espacio métrico completo.

Veamos en primer lugar que *toda sucesión de Cauchy de números reales es acotada*[2]. Sea $N \in \mathbb{N}$ tal que $|x_n - x_m| < 1 \quad \text{si} \quad n, m \geq N$. En particular, $|x_n - x_N| < 1 \quad \text{si} \quad n \geq N$. Por tanto, $|x_n| = |x_n - x_N| + |x_N| < 1 + |x_N|$. Por lo tanto, si K es el máximo de $|x_1|, \ \ldots, \ |x_{N-1}|$ y $1 + |x_N|$, concluimos que $|x_n| < K$, para todo $n \in \mathbb{N}$, es decir, $x_n \in (-K, K)$ para todo $n \in \mathbb{N}$.

A continuación demostraremos que *toda sucesión de Cauchy de números reales posee una subsucesión convergente.* Sea $(x_n)_{n\in\mathbb{N}}$ una sucesión de Cauchy de números reales. Como es acotada, existe $K > 0$ tal que $x_n \in (-K, K)$ para todo $n \in \mathbb{N}$. Ahora, podemos dividir $(-K, K)$ en dos mitades, y en una de ellas, que denotamos (a_1, b_1), encontraremos infinitos términos de nuestra sucesión. Elegimos un término de la sucesión $x_{i_1} \in (a_1, b_1)$. Dividimos ahora (a_1, b_1) en dos mitades, nuevamente habrá infinitos elementos de nuestra sucesión en una de las mitades, que denotamos (a_2, b_2); y elegimos un término de nuestra sucesión $x_{i_2} \in (a_2, b_2)$ con $i_1 \leq i_2$. Continuando de esta manera, obtenemos dos sucesiones $(a_n)_{n\in\mathbb{N}}$ y $(b_n)_{n\in\mathbb{N}}$, y una subsucesión $(x_{i_n})_{n\in\mathbb{N}}$ de $(x_n)_{n\in\mathbb{N}}$. Está tres sucesiones tienen las siguientes características:

[2]De hecho, esta propiedad es cierta para cualquier *espacio normado* como veremos más adelante.

(a) La sucesión $(a_n)_{n\in\mathbb{N}}$ es monótona creciente y acotada, luego es convergente (compruébese). Sea $a = \lim_{n\to\infty} a_n$.
(b) La sucesión $(b_n)_{n\in\mathbb{N}}$ es monótona decreciente y acotada, luego es convergente (compruébese). Sea $b = \lim_{n\to\infty} b_n$.
(c) La subsucesión $(x_{i_n})_{n\in\mathbb{N}}$ está comprendida entre las anteriores, es decir, $a_n < x_{i_n} < b_n$, para cada $n \geq 1$ (compruébese).

Veamos ahora, que a y b son iguales. Es claro que la longitud del intervalo (a_n, b_n) es $|a_n - b_n| = K/2^{n-1}$, que converge a 0 cuando n tiende hacia infinito. Por consiguiente, usando la desigualdad triangular del valor absoluto, obtenemos que

$$|a - b| \leq |a - a_n| + |a_n - b| \leq |a - a_n| + |a_n - b_n| + |b_n - b|.$$

De donde se sigue que $a = b$. Además, como $a_n < x_{i_n} < b_n$, para cada $n \geq 1$, concluimos que la subsucesión $(x_{i_n})_{n\in\mathbb{N}}$ es convergente; es más, $\lim_{n\to\infty} x_{i_n} = a = b$.

Hemos demostrado que toda sucesión de Cauchy de número reales posee una subsucesión convergente, es decir, *toda sucesión de Cauchy de números reales tiene un valor de adherencia*. Luego, por el lema A.3.3, concluimos que *toda sucesión de Cauchy de números reales es convergente*, y por lo tanto que $\mathbb{R}$ es un espacio métrico completo.

ii) El espacio vectorial $\mathbb{R}^n$ con la distancia usual es completo. En efecto, sea $(\mathbf{v}_m)_{m\in\mathbb{N}}$ una sucesión de Cauchy en $\mathbb{R}^n$, donde

$$\mathbf{v}_1 = (v_1^{(1)}, v_2^{(1)}, \ \ldots, \ v_n^{(1)}), \ \ldots, \ \mathbf{v}_m = (v_1^{(m)}, v_2^{(m)}, \ \ldots, \ v_n^{(m)}), \ldots$$

Entonces (véase el ejemplo A.3.2) las proyecciones de $(\mathbf{v}_m)_{m\in\mathbb{N}}$ en los m subespacio coordenados son sucesiones de Cauchy y, puesto que $\mathbb{R}$ es completo, convergen:

$$\lim_{m\to\infty} v_1^{(m)} = v_1, \ \ldots, \ \lim_{m\to\infty} v_n^{(m)} = v_n.$$

Así, pues $(\mathbf{v}_m)_{m\in\mathbb{N}}$ converge a $\mathbf{v} = (v_1, \ \ldots, \ v_n) \in \mathbb{R}^n$, ya que $\mathrm{d}(\mathbf{v}_m, v)^2 = |v_1^{(m)} - v_1|^2 + \ldots + |v_n^{(m)} - v_n|^2$.

iii) Tanto $\mathbb{C}$ como $\mathbb{C}^n$, con sus distancias usuales respectivas, son completos; basta tener en cuenta que $\mathbb{C}$ la distancia definida por el módulo de la diferencia es, topológicamente hablando, exactamente igual que $\mathbb{R}^2$ con la distancia usual.

Proposición A.3.6. *Sea* (X, d) *un espacio métrico. Si* $(x_n)_{n\in\mathbb{N}}$ *e* $(y_n)_{n\in\mathbb{N}}$ *son sucesiones de Cauchy, entonces* $\mathrm{d}(x_n, y_n)$ *es una sucesión convergente de números reales.*

Demostración. Usando la proposición A.1.5 y la desigualdad triangular del valor absoluto,

$$\begin{aligned}|\mathrm{d}(x_m, y_m) - \mathrm{d}(x_n, y_n)| &\leq |\mathrm{d}(x_m, y_m) - \mathrm{d}(x_n, y_m)| + |\mathrm{d}(y_m, x_n) - \mathrm{d}(y_n, x_n)| \\ &\leq \mathrm{d}(x_m, x_n) + \mathrm{d}(y_m, y_n)\end{aligned}$$

que tiende a cero cuando n y m tienden hacia infinito. Como los números reales con la distancia usual constituyen un espacio métrico completo, la sucesión de Cauchy $\mathrm{d}(x_n, y_n)$ es convergente. ∎

Veamos ahora que todo subconjunto completo de un espacio métrico es cerrado.

Proposición A.3.7. *Sea* (X, d) *un espacio métrico. Todo subconjunto completo de* X *es cerrado.*

Demostración. Toda sucesión convergente de elementos de Y es, en particular, de Cauchy. Luego, su límite pertenece a Y y, por la proposición A.2.4(b), podemos afirmar que Y es cerrado. ∎

Proposición A.3.8. *Sea* (X, d) *un espacio métrico completo. Un subconjunto de* X *es completo si, y sólo si, es cerrado.*

Demostración. Si Y es completo, entonces, por la proposición anterior, es cerrado. Recíprocamente, como toda sucesión de Cauchy de elementos de Y es convergente en X (pues, en particular, es una sucesión de elementos de X y X es completo) e Y es cerrado, por la proposición A.2.4(b), tiene su límite en Y. ∎

4. Conjuntos compactos

Definición A.4.1. Sea (X, d) un espacio métrico. Se dice que un subconjunto M de X es **acotado** si existen $x \in M$ y $\varepsilon > 0$ tales que

$$M \subseteq B(x, \varepsilon).$$

Obsérvese que las bolas abiertas y cerradas son conjuntos acotados.

Definición A.4.2. Sea (X, d) un espacio métrico. Se dice que un subconjunto M de X es **totalmente acotado** (o **precompacto**) cuando de cualquier sucesión de elementos de M se puede extraer una subsucesión de Cauchy.

También pueden describirse los conjuntos totalmente acotados de la siguiente manera:

Proposición A.4.3. *Sea* (X, d) *un espacio métrico. Un subconjunto* $M \subseteq X$ *es totalmente acotado si, y sólo si, para cada* $\varepsilon > 0$ *existe un número finito de elementos* $x_1, \ldots, x_n \in M$ *(que dependen de* ε*) tales que,*

$$M \subseteq \bigcup_{i=1}^{n} B(x_i, \varepsilon).$$

Demostración. $\boxed{\Rightarrow}$ Demostremos el contrarrecíproco. Supongamos que existe $\varepsilon > 0$ tal que para cualquier conjunto finito $x_1, \ldots, x_n \in M$ existe $x_{n+1} \in M$ con $\mathrm{d}(x_i, x_{n+1}) \geq \varepsilon$, $i \in \{1, \ldots, n\}$. Es decir, existe una sucesión $(x_n)_{n\in\mathbb{N}}$ tal que $\mathrm{d}(x_i, x_j) \geq \varepsilon$, para todo $j > i$. Es claro, que de esta sucesión no se puede extraer ninguna subsucesión de Cauchy por lo que M no es totalmente acotado.

$\boxed{\Leftarrow}$ Sean $(y_n)_{n\in\mathbb{N}}$ una sucesión de elementos de M y $\varepsilon > 0$. Por hipótesis, existen $x_1^{(j)}, \ldots, x_{n_j}^{(j)}$ tales que

$$M \subseteq \bigcup_{i=1}^{n_j} B(x_i^{(j)}, \varepsilon/2^j),$$

para cada $j \in \mathbb{N}$. Si reordenamos las bolas de tal forma que

$$U_k := \bigcap_{j=1}^{k} B(x_j^{(1)}, \varepsilon/2^j)$$

contenga infinitos términos de la sucesión, para cada $k \geq 1$, y elegimos $y_{n_1} \in U_1$, $y_{n_2} \in U_2$, con $n_2 > n_1, \ldots, y_{n_k} \in U_mathbbmssk$, con $n_k > n_{k-1}$, y así sucesivamente; obtenemos una subsucesión, $(y_{n_k})_{k\in\mathbb{N}}$, de $(y_n)_{n\in\mathbb{N}}$ que es de Cauchy. ∎

Corolario A.4.4. *Sea* (X, d) *un espacio métrico. Todo subconjunto de* X *totalmente acotado es acotado.*

Demostración. Si $M \subseteq X$ es totalmente acotado, para cada $\varepsilon' > 0$ existen $y_1, \ldots, y_n \in M$ (que dependen de ε') tales que, $M \subseteq \bigcup_{i=1}^{n} B(y_i, \varepsilon')$. Sean $\varepsilon'' = \sum_{i=1}^{n-1} \mathrm{d}(y_i, y_{i+1})$ y $x \in M$, sin perdida de generalidad, podemos suponer que $x \in B(y_1, \varepsilon')$. Si $y \in M$, existe $m \in \{1, \ldots, n\}$ tal que $\mathrm{d}(y_m, y) < \varepsilon$. Por consiguiente,

$$\mathrm{d}(x, y) = \mathrm{d}(x, y_1) + \ldots + \mathrm{d}(y_m, y) < 2\varepsilon' + \varepsilon''.$$

Luego, $y \in B(x, \varepsilon)$, con $\varepsilon = 2\varepsilon' + \varepsilon''$, y concluimos que M es acotado. ∎

El recíproco de la proposición anterior no es cierto en general. Por ejemplo, la recta real $\mathbb{R}$ con la distancia d definida por $\mathrm{d}(x, y) = \text{ínf}\{1, |x - y|\}$ es acotada pero no es totalmente acotada.

Definición A.4.5. Sea (X, d) un espacio métrico. Se dice que un subconjunto K de X es **compacto** cuando cualquier sucesión de elementos de K se puede extraer una subsucesión convergente a un elemento de K.

En particular, todo conjunto compacto es totalmente acotado.

Propiedad fundamental de los espacios métricos. *Sea* (X, d) *un espacio métrico. Un subconjunto de* X *es compacto si, y sólo si, es completo y totalmente acotado.*

Demostración. Sea $K \subset X$ compacto. Por hipótesis, de cualquier sucesión $(x_n)_{n\in\mathbb{N}}$ de elementos de K se puede extraer una subsucesión convergente a un elemento de K. Luego, en particular se puede extraer un subsucesión de Cauchy y K es totalmente acotado. Por otra parte, toda sucesión de Cauchy en K admite una subsucesión convergente a un elemento $x \in K$, luego x será un valor de adherencia de la sucesión de Cauchy y, por el lema A.3.3, el límite de la sucesión de Cauchy. Luego, K es completo.

Recíprocamente, si $K \subseteq X$ es totalmente acotado, de toda sucesión $(x_n)_{n\in\mathbb{N}}$ se puede extraer una subsucesión de Cauchy, que, por ser K completo, es convergente a un elemento de K. Luego, K es compacto. ∎

Nótese que de la Propiedad fundamental de los espacios métricos, se sigue que, *en un espacio métrico todo compacto es cerrado y acotado.*

Corolario A.4.6. *Sea* (X, d) *un espacio métrico. Si un subconjunto de* X *es compacto, entonces es cerrado.*

Demostración. Si $K \subseteq X$ es compacto, entonces, por la Propiedad fundamental de los espacios métricos, es completo y totalmente acotado. Luego, por la proposición A.3.7, es cerrado. ∎

Corolario A.4.7. *Sea* (X, d) *un espacio métrico compacto. Un subconjunto de* X *es compacto si, y sólo si, es cerrado.*

Demostración. Si $K \subseteq X$ es compacto, entonces, por el corolario anterior, es cerrado. Recíprocamente, si $K \subseteq X$ es cerrado, entonces es completo, por la proposición A.3.8, y es totalmente acotado por serlo X. Luego, por Propiedad fundamental de los espacios métricos, concluimos que K es compacto. ∎

Ejemplos A.4.8.

i) La recta real $\mathbb{R}$ con la distancia usual, no es compacta porque no es acotada.
ii) La bola cerrada de centro el origen y radio unidad de la recta real $\mathbb{R}$ con la distancia usual es compacta, pues es completa (al ser un cerrado de un espacio métrico completo), y es totalmente acotada.
iii) En la real $\mathbb{R}$ con la distancia usual ser totalmente acotado equivale a ser acotado, luego en este caso se tiene que un subconjunto es compacto si, y sólo si, es cerrado y acotado.

iv) En $\mathbb{R}^n$ con la distancia usual, se puede comprobar que los conjuntos cerrados y acotados son compactos. Luego, en $\mathbb{R}^n$ también se cumple que un subconjunto es compacto si, y sólo si, es cerrado y acotado.

Teorema A.4.9. *Sea* (X, d) *un espacio métrico compacto. Si* $f : X \to \mathbb{R}$ *es continua, entonces*

(a) f *es acotada, es decir, existe* $M > 0$ *tal que* $|f(x)| < M$, *para todo* $x \in X$.
(b) f *alcanza un máximo y un mínimo.*
(c) f *es cerrada.*

Demostración. La demostración de este teorema se deja como ejercicio al lector. ∎

Nota A.4.10. El lector interesado en profundizar en este tema puede consultar [**Lip70**] donde además encontrará multitud de ejercicios y ejemplos que puede ayudar a una mejor compresión de este apéndice.

APÉNDICE B

Estructuras algebraicas

A continuación repasemos brevemente los conceptos de grupo, cuerpo y anillo, centrándonos en algunos ejemplos conocidos. Un estudio más detallado de estas estructuras puede encontrarse en [**Nav96**].

1. Grupos y subgrupos

La suma en $\mathbb{Z}$ se puede entender como una aplicación

$$\begin{array}{rccl} \circ & : & \mathbb{Z}\times\mathbb{Z} & \longrightarrow & \mathbb{Z} \\ & & (m,n) & \longmapsto & \circ(m,n) := m+n \end{array}$$

que verifica las siguientes propiedades:

- Propiedad asociativa: si m, n y $p \in \mathbb{Z}$, entonces $(m+n)+p = m+(n+p)$.
- Propiedad de elemento neutro: existe $e \in \mathbb{Z}$ tal que $n+e = e+n = n$, para todo $n \in \mathbb{Z}$. Tómese, $e = 0 \in \mathbb{Z}$.
- Propiedad de elemento simétrico: existe $n' \in \mathbb{Z}$ tal que $n+n' = n'+n = e$, para cada $n \in \mathbb{Z}$. Tómese $n' = -n$, para cada $n \in \mathbb{Z}$.
- Propiedad conmutativa: $m+n = n+m$, para todo m y $n \in \mathbb{Z}$.

Este conocido ejemplo sirve como introducción a la noción de grupo.

Definición B.1.1. Un **grupo** es un par $(G, \circ)$ donde G es un conjunto no vacío y $\circ : G\times G \longrightarrow G; (a,b) \mapsto a\circ b$ es una aplicación que verifica las siguientes propiedades:

(G1) **Propiedad asociativa:** si a, b y $c \in G$, entonces $(a\circ b)\circ c = a\circ(b\circ c)$.

(G2) **Propiedad de elemento neutro:** existe $e \in G$ tal que $a\circ e = e\circ a = a$, para todo $a \in G$.

(G3) **Propiedad de elemento simétrico:** para cada $a \in G$ existe $a' \in G$ tal que $a\circ a' = a'\circ a = e$.

Además, si se cumple

(G4) **Propiedad conmutativa:** $a\circ b = b\circ a$, para todo a y $b \in G$.

se dice que el par $(G, \circ)$ es un **grupo conmutativo** ó **grupo abeliano**.

Ejemplo B.1.2. El par $(\mathbb{Z}, +)$ es un grupo conmutativo. El par $(\mathrm{Gl}_n(\mathbb{Q}), \cdot)$ con $n > 1$ es un grupo no conmutativo.

Nota B.1.3. Dado que la gran mayoría de los grupos con los que vamos a trabajar serán grupos conmutativos, a partir de ahora omitiremos el apelativo conmutativo y nos referimos a ellos como grupos sin más, especificando lo contrario cuando sea necesario.

Habitualmente, si $(G, \circ)$ es grupo, a la aplicación $\circ$ se le llama **operación interna** del grupo ó **ley de composición interna**. Es fundamental el hecho de que, dados dos elementos a y $b \in G$, la imagen por $\circ$ del par (a, b), es decir, $a \circ b$, es también un elemento de G.

Veamos ahora que los elementos de G cuya existencia aseguran los axiomas (G2) y (G3) de la definición B.1.1 son únicos.

Proposición B.1.4. *Si $(G, \circ)$ es un grupo, entonces se verifican las siguientes propiedades:*

(a) *Existe un único elemento $e \in G$, tal que $a \circ e = e \circ a = a$, para todo $a \in G$.*
(b) *Existe un único elemento $a' \in G$, tal que $a \circ a' = a' \circ a = e$, para cada $a \in G$.*

Demostración. (a) Si existen dos elementos neutros e y $e' \in G$, entonces $e \circ e' = e' \circ e = e$ y $e' \circ e = e \circ e' = e'$. De donde se sigue $e = e'$.

(b) Sea $a \in G$, si existen dos elementos simétricos a' y $a'' \in G$, entonces

$$\left.\begin{array}{rcl} a \circ a' = e & \Longrightarrow & a'' \circ (a \circ a') = a'' \circ e = a'' \\ a'' \circ a = e & \Longrightarrow & (a'' \circ a) \circ a' = e \circ a' = a' \end{array}\right\} \overset{\text{Asociativa}}{\Longrightarrow} a'' = a'.$$

■

Definición B.1.5. Sea $(G, \circ)$ un grupo. Al único elemento e de G tal que $a \circ e = e \circ a = a$, para todo $a \in G$, lo llamaremos **elemento neutro** de G. Si $a \in G$, al único elemento a' de G tal que $a \circ a' = a' \circ a = e$, para cada $a \in G$, lo llamaremos **elemento simétrico** de a.

Aunque a la operación interna del grupo la hayamos llamado $\circ$, es frecuente utilizar las notaciones habituales de la adición$(+)$ y de la multiplicación$(\cdot)$. En notación aditiva, el elemento neutro se llama **cero** y se expresa por 0, y el elemento simétrico de un elemento $a \in G$ se llama **opuesto** y se representa por $-a$. En notación multiplicativa, el elemento neutro se llama **unidad** y se representa por 1, y el elemento simétrico de un elemento $a \in G$ se llama **inverso** y se representa por a^{-1}.

Ejemplo B.1.6. Además de los ejemplos que han servido como introducción al concepto de grupo, se citan a continuación otros, de los que se deja al lector las comprobaciones correspondientes:

1. $(\mathbb{Q}, +), (\mathbb{R}, +)$ y $(\mathbb{C}, +)$ son grupos. Aquí la operación interna $+$ es la suma usual.

2. $(\mathbb{Q}\setminus\{0\},\cdot),(\mathbb{R}\setminus\{0\},\cdot)$ y $(\mathbb{C}\setminus\{0\},\cdot)$ son grupos. Es decir, el conjunto de los racionales (reales, complejos, respectivamente) no nulos junto con la multiplicación usual de números racionales (reales, complejos, respectivamente) tiene estructura de grupo ¿Por qué ha sido necesario prescindir del cero?
3. Sea $n\in\mathbb{N}$ fijo. $(\mathbb{Q}^n,+)$ es un grupo, donde

$$\mathbb{Q}^n:=\{(a_1,a_2,\ldots,a_n)\ \mid\ a_i\in\mathbb{Q},i=1,\ldots,n\}$$

y $+$ es la suma elemento a elemento, es decir, $(a_1,a_2,\ldots,a_n)+(b_1,b_2,\ldots,b_n)=(a_1+b_1,a_2+b_2,\ldots,a_n+b_n)$. Asimismo $(\mathbb{R}^n,+)$ y $(\mathbb{C}^n,+)$ son grupos, con la suma definida de igual forma que antes.

Nota B.1.7. En lo sucesivo, y mientras no se diga lo contrario, usaremos la notación aditiva para grupos. Así escribiremos $(G,+)$ en vez de $(G,\circ)$, entendiendo que $+$ denota la operación interna que dota G de estructura de grupo.

Ejercicio B.1.8. Dado un grupo $(G,+)$ cualquiera, no necesariamente conmutativo. Probar que las siguientes afirmaciones son equivalentes:

(a) G es conmutativo;
(b) $n(a+b)=na+nb$, para todo a y $b\in G$ y $n\in\mathbb{Z}$.
(c) $-(a+b)=-a-b$, para todo a y $b\in G$.

Generalmente, los subconjuntos de un grupo no heredan la estructura de grupo. Llamaremos subgrupo a los subconjuntos que sí la conserven.

Definición B.1.9. Sean $(G,+)$ un grupo (no necesariamente conmutativo) y H un subconjunto no vacío de G. Diremos que H es un **subgrupo** (no necesariamente conmutativo) de $(G,+)$ si $(H,+)$ es grupo, donde $+:H\times H\longrightarrow H$ es la restricción de $+:G\times G\longrightarrow G$ a $H\times H\subseteq G\times G$.

Obsérvese que, dado un grupo $(G,+)$, se tiene que tanto G como $\{0\}$ son subgrupos de $(G,+)$. Un subgrupo se dice **propio** si es distinto de G.

Según la definición anterior, para comprobar si $H\subseteq G$ es un subgrupo de $(G,+)$ tenemos que asegurarnos de que H es un subconjunto no vacío, que la restricción de $+:H\times H\longrightarrow H$ está bien definida, es decir, que es una aplicación, y que el par $(H,+)$ verifica los axiomas de grupo, (G1-G3) de la definición B.1.1. Sin embargo, en breve veremos que esto no va a ser necesario.

Ejemplo B.1.10. Consideramos el grupo $(\mathbb{Z},+)$ con $+$ la suma usual de números enteros.

1. El subconjunto de $\mathbb{Z}$ formado por todos los números enteros pares, es decir, $\{2z\mid z\in\mathbb{Z}\}\subset\mathbb{Z}$, que denotamos por $2\mathbb{Z}$ es un subgrupo de $(\mathbb{Z},+)$ (compruébese).

2. El subconjunto de todos los números enteros impares, es decir, $H := \{2z+1 \mid z \in \mathbb{Z}\} \subset \mathbb{Z}$, no es subgrupo de $\mathbb{Z}$. Basta observar que la correspondencia $+ : H \times H \longrightarrow H$ es el conjunto vacío y por tanto que no es aplicación.
3. El subconjunto de todos los números naturales, $\mathbb{N}$, no es subgrupo de $(\mathbb{Z}, +)$. En efecto, aunque la aplicación $+ : \mathbb{N} \times \mathbb{N} \longrightarrow \mathbb{N}$ está bien definida, no se verifica la propiedad de elemento simétrico ((G3) de la definición B.1.1).

El siguiente resultado proporciona una definición equivalente de subgrupo, que resulta mucho más manejable.

Proposición B.1.11. *Sean $(G, +)$ un grupo y H un subconjunto no vacío de G. Son equivalentes:*

(a) *H es un subgrupo de G.*
(b) *Si a y $b \in H$, entonces $a - b = a + (-b) \in H$.*

Demostración. $\boxed{\text{(a)} \Rightarrow \text{(b)}}$ Sean a y b elementos de H. Por ser H subgrupo de $(G, +)$ se tiene que $(H, +)$ es grupo. Luego por el axioma (G3) de la definición B.1.1, tenemos que $-b \in H$, de donde se sigue que $a + (-b) = a - b \in H$.

$\boxed{\text{(b)} \Rightarrow \text{(a)}}$ La propiedad asociativa, al verificarse en $(G, +)$, se verifica en cualquier subconjunto H de G. Por otro lado, si $a \in H$ (existe alguno pues $H \neq \emptyset$) tomando $b = a$, se tiene $a - a \in H$. O sea, $0 \in H$, y por lo tanto $0 - a \in H$, luego $-a \in H$. De manera que, si a y $b \in H$, en particular, $-b \in H$, y por tanto $a - (-b) = a + b \in H$, lo que completa la demostración. ∎

Operaciones con subgrupos.

A lo largo de este apartado consideramos fijado un grupo $(G, +)$.

Es claro que la intersección y unión de subconjuntos de G es de nuevo un subconjunto de G. Parece por tanto natural, que nos preguntemos si ocurre algo similar con la intersección y unión de subgrupos de $(G, +)$.

En este apartado veremos que la intersección de subgrupos de $(G, +)$ es un subgrupo de $(G, +)$, y que esto no ocurrirá en general para la unión de subgrupos. Haciéndose necesario introducir una nueva operación que llamaremos suma de subgrupos, y que jugará un papel análogo a la unión de subconjuntos.

Proposición B.1.12. *Si H_1 y H_2 son dos subgrupos de $(G, +)$, entonces el conjunto intersección de H_1 y H_2, es decir, $H_1 \cap H_2$, es un subgrupo de $(G, +)$.*

Demostración. En primer lugar, tenemos que asegurarnos de que $H_1 \cap H_2 \neq \emptyset$. Dado que el elemento neutro 0 pertenece a cualquier subgrupo de $(G, +)$, podemos afirmar que $H_1 \cap H_2 \neq \emptyset$. Ahora, por la proposición B.1.11, basta comprobar que si a y b

está en $H_1 \cap H_2$, entonces $a-b \in H_1 \cap H_2$, lo que es elemental y se deja como ejercicio. ■

El resultado anterior se puede generalizar a una familia arbitraria de subgrupos de $(G,+)$.

Corolario B.1.13. *Si $\{H_i\}_{i\in I}$ es una familia de subgrupos de $(G,+)$, entonces $\cap_{i\in I} H_i$ es un subgrupo de $(G,+)$.*

Ejercicio B.1.14. Sean H_1 y H_2 dos subgrupos de $(G,+)$. Probar que $H_1 \cap H_2$ es el mayor de todos los subgrupo de $(G,+)$ que están contenidos en H_1 y H_2 simultáneamente. Generalizar el resultado para una intersección arbitraria de subgrupos.

Por consiguiente, podemos afirmar que la intersección es el ínfimo de una familia de subgrupos dada.

Como ya hemos comentando, la unión de subgrupos no es subgrupo en general, tal y como puede deducirse del siguiente ejemplo.

Ejemplo B.1.15. Consideramos el grupo $(\mathbb{Z},+)$ donde $+$ es la suma usual de números enteros, y los subconjuntos $2\mathbb{Z} = \{2n \mid n \in \mathbb{Z}\}$ y $3\mathbb{Z} = \{2n \mid n \in \mathbb{Z}\}$ de $\mathbb{Z}$. Tanto $2\mathbb{Z}$ como $3\mathbb{Z}$ son subgrupos de $(\mathbb{Z},+)$ (compruébese). En cambio $2\mathbb{Z} \cup 3\mathbb{Z}$ no lo es, ya que 2 y $3 \in 2\mathbb{Z} \cup 3\mathbb{Z}$ pero $2-3=-1 \notin 2\mathbb{Z} \cup 3\mathbb{Z}$ pues -1 ni es par ni múltiplo de 3.

Única y exclusivamente se puede asegurar que la unión de dos subgrupos H_1 y H_2 de $(G,+)$ es un subgrupo de $(G,+)$ si y sólo si ó $H_1 \subseteq H_2$ ó $H_2 \subseteq H_1$, es decir, si y sólo si ó $H_1 \cup H_2 = H_2$ ó $H_1 \cup H_2 = H_1$.

Por consiguiente, a diferencia de lo que ocurría con los conjuntos la unión no podrá desempeñar el rol de supremo de una familia de subgrupos dada. Esta deficiencia se suple con la suma de subgrupos, que pasamos a definir a continuación.

Nota B.1.16. Advertimos al lector que en los siguientes resultados se hará uso de la propiedad conmutativa, y que por tanto no serán ciertos para grupos no conmutativos en general.

Comencemos definiendo la suma de dos subgrupos de $(G,+)$ y comprobando que efectivamente es subgrupo de $(G,+)$.

Definición B.1.17. Sean H_1 y H_2 dos subgrupos de $(G,+)$. Definimos la suma de H_1 y H_2 como el subconjunto

$$H_1 + H_2 := \{h_1 + h_2 \mid h_1 \in H_1 \text{ y } h_2 \in H_2\} \subseteq G.$$

Proposición B.1.18. *Sean H_1 y H_2 dos subgrupos de $(G,+)$. El conjunto suma de H_1 con H_2, es decir, $H_1 + H_2$, es subgrupo de G.*

Demostración. Obviamente $H_1+H_2 \neq \emptyset$, pues $0 = 0+0 \in H_1+H_2$. Por la proposición B.1.11, basta probar que si a y $b \in H_1 + H_2$, entonces $a - b \in H_1 + H_2$. Si a y $b \in H_1 + H_2$, entonces $a = a_1 + a_2$ y $b = b_1 + b_2$ con a_1 y $b_1 \in H_1$ y a_2 y $b_2 \in H_2$. De manera que tenemos la siguiente cadena de igualdades

$$a - b = (a_1 + a_2) - (b_1 + b_2) = a_1 + a_2 - b_2 - b_1 \overset{\text{Conmutativa}}{=} (a_1 - b_1) + (a_2 - b_2).$$

Como H_1 y H_2 son subgrupos de $(G, +)$ podemos asegurar que $(a_1 - b_1) \in H_1$ y $(a_2 - b_2) \in H_2$. De donde se sigue que $a - b = (a_1 - b_1) + (a_2 - b_2) \in H_1 + H_2$. ∎

La definición de suma de dos subgrupos se puede generalizar sin mayor complicación a una suma de una familia finita de subgrupos de $(G, +)$. Obteniéndose el siguiente resultado, cuya demostración se deja como ejercicio.

Corolario B.1.19. *Sean* $\{H_1, \ldots, H_n\}$ *una familia finita de subgrupos de* $(G, +)$. *El conjunto suma* $H_1 + \ldots + H_n$ *es un subgrupo de* $(G, +)$.

Nota B.1.20. Se puede definir la suma de una familia arbitraria de subgrupos de $(G, +)$, pero no de forma totalmente análoga. Y puesto que a lo más, trabajaremos con sumas finitas de subgrupos, es preferible sacrificar la generalidad por una mayor concreción.

Ejercicio B.1.21. Sean H_1 y H_2 dos subgrupos de $(G, +)$. Probar que $H_1 + H_2$ es el menor de todos los subgrupos de $(G, +)$ que contiene tanto a H_1 como a H_2, es decir, que contiene al conjunto $H_1 \cup H_2$. Generalizar el resultado para cualquier suma finita de subgrupos.

2. Cuerpos

En los apartados 1. y 2. del ejemplo B.1.6 vimos que las operaciones usuales de suma y producto de números racionales dotan a los conjuntos $\mathbb{Q}$ y $\mathbb{Q}\backslash\{0\}$ de estructura de grupo (conmutativo), respectivamente. Además, no es difícil comprobar que ambas operaciones verifican la siguiente propiedad:

$$\forall a, b \text{ y } c \in \mathbb{Q},\ a \cdot (b + c) = a \cdot b + a \cdot c \qquad (*).$$

Y análogamente ocurre en $\mathbb{R}$ y $\mathbb{R} \setminus \{0\}$ y en $\mathbb{C}$ y $\mathbb{C} \setminus \{0\}$.

Esta doble estructura de grupo (conmutativo) junto con la propiedad (*) recibe el nombre de cuerpo (conmutativo).

Definición B.2.1. Un **cuerpo** es una terna $(mathbbmssk, +, \cdot)$, donde $\Bbbk$ es un conjunto no vacío, y $+ : \Bbbk \times \Bbbk \longrightarrow \Bbbk,\ (a, b) \mapsto a + b$, y $\cdot : \Bbbk \times \Bbbk \longrightarrow \Bbbk, (a, b) \mapsto a \cdot b$, dos aplicaciones, verificando:

(a) $(mathbbmssk, +)$ es un grupo conmutativo, es decir:

- $(a + b) + c = a + (b + c)$, para todo a, b y $c \in \Bbbk$.

- Existe $e \in \Bbbk$ tal que $a + e = e + a = a$, para todo $a \in \Bbbk$. $(e = 0)$.
- Para cada $a \in \Bbbk$, existe $a' \in \Bbbk$ tal que $a + a' = a' + a = e$ $(a' = -a)$.
- $a + b = b + a$, para todo a y $b \in \Bbbk$.

(b) $(\Bbbk \setminus \{0\}, \cdot)$ es un grupo conmutativo, esto es:

- $(a \cdot b) \cdot c = a \cdot (b \cdot c)$, para todo a, b y $c \in \Bbbk \setminus \{0\}$.
- Existe $u \in \Bbbk \setminus \{0\}$ tal que $a \cdot u = u \cdot a = a$, para todo $a \in \Bbbk \setminus \{0\}$. $(u = 1)$.
- Para cada $a \in \Bbbk \setminus \{0\}$, existe $\tilde{a} \in \Bbbk \setminus \{0\}$ tal que $a \cdot \tilde{a} = \tilde{a} \cdot a = u$ $(\tilde{a} = a^{-1})$.
- $a \cdot b = b \cdot a$, para todo a y $b \in \Bbbk \setminus \{0\}$.

(c) **Propiedad distributiva:** $a \cdot (b + c) = a \cdot b + a \cdot c$, para todo a, b y $c \in \Bbbk$.

Conviene destacar que el conjunto cuyo único elemento es el cero no es un cuerpo, en otro caso, tendríamos que $(\{0\} \setminus \{0\} = \emptyset, \cdot)$ sería un grupo, lo que es del todo imposible. Luego, podemos afirmar que todo cuerpo tiene al menos dos elementos, el 0 y el 1.

Nota B.2.2. En lo sucesivo, dado un cuerpo conmutativo $(mathbbmssk, +, \cdot)$, nos referiremos a él como el cuerpo $\Bbbk$ a secas, sobrentendiendo las operaciones internas de suma$(+)$ y producto$(\cdot)$, y asumiendo que $\Bbbk$ es conmutativo salvo que se diga lo contrario.

Nota B.2.3. Obsérvese que $a \cdot 0 = 0 \cdot a = 0$, para todo $a \in \Bbbk$. En efecto, para cualesquiera a y $b \in \Bbbk \setminus \{0\}$ se tiene que

$$a \cdot b = a \cdot (b + 0) \overset{\text{Distributiva}}{=} a \cdot b + a \cdot 0,$$

de donde se sigue, por la unicidad del elemento neutro, que $a \cdot 0 = 0$ y, por la conmutatividad del producto, que $0 \cdot a = 0$.

De aquí que en mucho textos se sobrentienda esta propiedad y en el punto 2. de la definición B.2.1 se escriba:

2. La aplicación $\cdot : \Bbbk \times \Bbbk \longrightarrow \Bbbk$ cumple:
   - $(a \cdot b) \cdot c = a \cdot (b \cdot c)$, para todo a, b y $c \in \Bbbk$.
   - Existe $u \in \Bbbk$ tal que $a \cdot u = u \cdot a = a$, para todo $a \in \Bbbk$. $(u = 1)$.
   - Para cada $a \in \Bbbk \setminus \{0\}$, existe $\tilde{a} \in \Bbbk$ tal que $a \cdot \tilde{a} = \tilde{a} \cdot a = u$ $(\tilde{a} = a^{-1})$.
   - $a \cdot b = b \cdot a$, para todo a y $b \in \Bbbk$.

Lo que evidentemente implica que si $\Bbbk \neq \{0\}$, entonces $(\Bbbk \setminus \{0\}, \cdot)$ es grupo.

Ejemplo B.2.4. Como se comentado anteriormente, ejemplos de cuerpo son $\mathbb{Q}, \mathbb{R}$ y $\mathbb{C}$ con la suma y el producto habituales en cada uno de ellos. Sin embargo, $(\mathbb{Z}, +, \cdot)$ no es cuerpo, puesto que $(\mathbb{Z} \setminus \{0\}, \cdot)$ no es un grupo.

Nota B.2.5. La propiedad distributiva, junto con la unicidad de los elementos neutro y unidad (véase la proposición B.1.4), asegura que las dos aplicaciones que dotan a un conjunto de estructura de cuerpo han de ser necesariamente distintas.

3. Anillos

Finalmente recordamos que se entiende por anillo (conmutativo, con elemento unidad) y $\Bbbk$-álgebra.

Definición B.3.1. Un **anillo** es una terna $(A, +, \circ)$, donde A es un conjunto no vacío y

(a) $+ : A \times A \longrightarrow A,\ (a, b) \mapsto a + b$, es una aplicación, llamada **suma**, tal que $(A, +)$ es un grupo conmutativo, es decir:
- $(a + b) + c = a + (b + c)$, para todo a, b y $c \in \Bbbk$.
- Existe $e \in \Bbbk$ tal que $a + e = e + a = a$, para todo $a \in \Bbbk$. $(e = 0)$.
- Existe $a' \in \Bbbk$ tal que $a + a' = a' + a = e$, para cada $a \in \Bbbk$. $(a' = -a)$.
- $a + b = b + a$, para todo a y $b \in \Bbbk$.

(b) $\circ : A \times A \longrightarrow A, (a, b) \mapsto a \circ b$, es otra aplicación, llamada **producto**, verificando las propiedades asociativa y distributiva respecto a $+$, es decir:
- $(a \circ b) \circ c = a \circ (b \circ c)$, para todo a, b y $c \in \Bbbk$.
- $a \circ (b + c) = a \circ b + a \circ c$, para todo a, b y $c \in \Bbbk$.

Si la aplicación producto verifica la propiedad de elemento unidad, es decir,

- $u \in \Bbbk \setminus \{0\}$ tal que $a \circ u = u \circ a = a$, para todo $a \in \Bbbk$. $(u = 1)$.

se dice que $(A, +, \circ)$ es un **anillo con unidad**. Por otra parte, si la aplicación producto verifica la propiedad conmutativa, es decir,

- $a \circ b = a \circ (b \circ c)$, para todo a y $b \in \Bbbk \setminus \{0\}$.

se dice que $(A, +, \circ)$ es un **anillo conmutativo**.

Ejemplo B.3.2. Todo cuerpo (conmutativo) es, en particular, un anillo (conmutativo) con elemento unidad. Un ejemplo de un anillo conmutativo con elemento unidad que no es un cuerpo es $\mathbb{Z}$ con las operaciones usuales de suma y producto (compruébese).

El conjunto $\Bbbk[x]$ de polinomios en la indeterminada x con coeficientes en un cuerpo $\Bbbk$ es un anillo conmutativo con unidad para la suma y el producto habitual de polinomios (compruébese).

Definición B.3.3. Sean A y A' dos anillos. Diremos que una aplicación $f : A \longrightarrow A'$ es un **morfismo de anillos** si verifica que

(a) $f(a +_A b) = f(a) +_{A'} f(b)$, para todo a y $b \in A$;
(b) $f(a \circ_A b) = f(a) \circ_{A'} f(b)$, para todo a y $b \in A$,

y si además, A y A' son anillos con unidad, que

(c) $f(1_A) = 1_{A'}$.

Nota B.3.4. Este apéndice cubre con creces los conceptos y resultados elementales sobre estructuras algebraicas que usaremos en este manual. No obstante, el lector interesado en profundizar en este tema puede consultar [**Nav96**] donde además encontrará multitud de ejercicios y ejemplos que puede ayudar a una mejor compresión de este apéndice.

APÉNDICE C

Espacios vectoriales

1. Definiciones y propiedades. Ejemplos

DE ahora en adelante, mientras no se indique lo contrario, $\Bbbk$ denotará a un cuerpo.

Definición C.1.1. Un **espacio vectorial** sobre $\Bbbk$, también llamado $\Bbbk$-espacio vectorial, es un conjunto no vacío V junto con:

(a) Una **operación interna** $+ : V \times V \longrightarrow V$ que dota a V de estructura de grupo conmutativo, es decir, que cumple:
- $(\mathbf{u}+\mathbf{v})+\mathbf{w} = \mathbf{u}+(\mathbf{v}+\mathbf{w})$, para todo $\mathbf{u}, \mathbf{v}$ y $\mathbf{w} \in V$.
- Existe $\mathbf{e} \in V$ tal que $\mathbf{u}+\mathbf{e} = \mathbf{e}+\mathbf{u} = \mathbf{u}$, para todo $\mathbf{u} \in V$. ($\mathbf{e} = \mathbf{0}$).
- Existe $\mathbf{u}' \in V$ tal que $\mathbf{u}+\mathbf{u}' = \mathbf{u}'+\mathbf{u} = \mathbf{e}$, para cada $\mathbf{u} \in V$. ($\mathbf{u}' = -\mathbf{u}$).
- $\mathbf{u}+\mathbf{v} = \mathbf{v}+\mathbf{u}$, para todo $\mathbf{u}$ y $\mathbf{v} \in V$.

(b) Una aplicación u **operación externa**

$$\begin{array}{rccl} * & : & \Bbbk \times V & \longrightarrow & V \\ & & (\lambda, \mathbf{u}) & \longmapsto & *(\lambda, \mathbf{u}) := \lambda * \mathbf{u} \end{array}\ ,$$

que verifica:
- $\lambda * (\mathbf{u}+\mathbf{v}) = \lambda * \mathbf{u} + \lambda * \mathbf{v}$, para todo $\mathbf{u}$ y $\mathbf{v} \in V$ y $\lambda \in \Bbbk$.
- $(\lambda+\mu) * \mathbf{u} = \lambda * \mathbf{u} + \mu * \mathbf{u}$, para todo $\mathbf{u} \in V$ y λ y $\mu \in \Bbbk$.
- $(\lambda \cdot \mu) * \mathbf{u} = \lambda * (\mu * \mathbf{u})$, para todo $\mathbf{u} \in V$ y λ y $\mu \in \Bbbk$.
- $1 * \mathbf{u} = \mathbf{u}$, para todo $\mathbf{u} \in V$, donde 1 es el elemento unidad de $\Bbbk$.

Nota C.1.2. Sea $(V, +, *)$ un $\Bbbk$-espacio vectorial. Llamaremos **vectores** a los elementos de V y **escalares** a los elementos del cuerpo $\Bbbk$. La aplicación $* : \Bbbk \times V \longrightarrow V$ se llama **producto por escalares**. De aquí que en lo que sigue, abusemos de la notación multiplicativa y, si $\lambda \in \Bbbk$ y $\mathbf{u} \in V$, escribamos $\lambda \cdot \mathbf{u}$ ó $\lambda\mathbf{u}$ en vez de $\lambda * \mathbf{u}$.

Según de la definición anterior, un $\Bbbk$-espacio vectorial es una terna $(V, +, *)$ que verifica una serie de propiedades. Sin embargo, por simplicidad en la escritura, a partir de ahora diremos que V es $\Bbbk$-espacio vectorial, entendiendo por ello que V está dotado de una operación "+" con la que es grupo abeliano y de un producto por escalares del cuerpo $\Bbbk$.

Asimismo conviene destacar que estamos denotando por $\mathbf{0}$ al elemento neutro del espacio vectorial, el **vector cero**, y por 0 al elemento neutro del cuerpo $\Bbbk$, el **escalar**

cero. En cualquier caso, el propio contexto, delimitará claramente cuando se usa uno u otro.

Ejemplo C.1.3. Mostramos a continuación una serie de ejemplos de espacios vectoriales, de los que se deja al lector las comprobaciones correspondientes.

1. El cuerpo $\Bbbk$, con las operaciones suma y productos propias, es un espacio vectorial sobre sí mismo.
2. El conjunto cuyo único elemento es el cero $\{\mathbf{0}\}$ es un $\Bbbk$-espacio vectorial, que llamaremos **espacio vectorial trivial**.
3. Las matrices de m filas y n columnas con coeficientes en $\Bbbk$, $\mathcal{M}_{m\times n}(\Bbbk)$, junto con la operación suma de matrices y el producto de escalares habitual, es decir, $A+B=(a_{ij})+(b_{ij})=(a_{ij}+b_{ij})$ y $\lambda A=\lambda(a_{ij})=(\lambda a_{ij})$ con $A=(a_{ij})$ y $B=(a_{ij})\in\mathcal{M}_{m\times n}(\Bbbk)$ y $\lambda\in\Bbbk$, es un $\Bbbk$-espacio vectorial.
4. El conjunto de los polinomios en la variable x y coeficientes en $\Bbbk$, $\Bbbk[x]$ con las operaciones usuales, es un espacio vectorial sobre $\Bbbk$.
5. El conjunto de los polinomios en la variable x de grado menor o igual que $n\in\mathbb{N}$ y con coeficientes en $\Bbbk$, $\Bbbk[x]_{\leq n}$ con las operaciones usuales, es un espacio vectorial sobre $\Bbbk$.

De la definición C.1.1 se siguen de forma inmediata las siguientes propiedades.

Proposición C.1.4. *Sea V un $\Bbbk$-espacio vectorial. Para todo $\mathbf{u}$ y $\mathbf{v}\in V$ y λ y $\mu\in\Bbbk$, se verifica que:*

(a) $\lambda\cdot\mathbf{0}=\mathbf{0}$.
(b) $0\cdot\mathbf{u}=\mathbf{0}$.
(c) $\lambda\cdot(\mathbf{u}-\mathbf{v})=\lambda\cdot\mathbf{u}-\lambda\cdot\mathbf{v}$.
(d) $(\lambda-\mu)\cdot\mathbf{u}=\lambda\cdot\mathbf{u}-\mu\cdot\mathbf{u}$.
(e) $(-\lambda)\cdot\mathbf{u}=-(\lambda\cdot\mathbf{u})$.
(f) $\lambda\cdot(-\mathbf{u})=-(\lambda\cdot\mathbf{u})$.

Demostración. (a) Si $\mathbf{u}\in V$ es un vector arbitrario, entonces $\lambda\cdot\mathbf{u}=\lambda\cdot(\mathbf{u}+\mathbf{0})=\lambda\cdot\mathbf{u}+\lambda\cdot\mathbf{0}\Longrightarrow\lambda\cdot\mathbf{0}=\mathbf{0}$.

(b) Si $\lambda\in\Bbbk$ es un escalar cualquiera, entonces $\lambda\cdot\mathbf{u}=(\lambda+0)\cdot\mathbf{u}=\lambda\cdot\mathbf{u}+0\cdot\mathbf{u}\Longrightarrow 0\cdot\mathbf{u}=\mathbf{0}$.

(c) $\lambda\cdot(\mathbf{u}-\mathbf{v})+\lambda\mathbf{v}=\lambda\cdot((\mathbf{u}-\mathbf{v})+\mathbf{v})=\lambda\cdot(\mathbf{u}+(-\mathbf{v}+\mathbf{v}))=\lambda\cdot\mathbf{u}\Longrightarrow\lambda\cdot(\mathbf{u}-\mathbf{v})=\lambda\cdot\mathbf{u}-\lambda\cdot\mathbf{v}$.

(d) $(\lambda-\mu)\cdot\mathbf{u}+\mu\mathbf{u}=((\lambda-\mu)+\mu)\cdot\mathbf{u}=(\lambda+(-\mu+\mu))\cdot\mathbf{u}=\lambda\cdot\mathbf{u}\Longrightarrow(\lambda-\mu)\cdot\mathbf{u}=\lambda\cdot\mathbf{u}-\mu\cdot\mathbf{u}$.

(e) $(-\lambda)\cdot\mathbf{u}+\lambda\cdot\mathbf{u}=(-\lambda+\lambda)\cdot\mathbf{u}=0\cdot\mathbf{u}=\mathbf{0}\Longrightarrow(-\lambda)\cdot\mathbf{u}=-\lambda\cdot\mathbf{u}$.

(f) $\lambda\cdot(-\mathbf{u})+\lambda\cdot\mathbf{u}=\lambda\cdot(-\mathbf{u}+\mathbf{u})=\lambda\cdot\mathbf{0}=\mathbf{0}\Longrightarrow\lambda\cdot(-\mathbf{u})=-\lambda\cdot\mathbf{u}$. ∎

Ejemplo C.1.5. En un ejemplo anterior vimos que todo cuerpo $\Bbbk$ con sus propias operaciones de suma y producto es un $\Bbbk$-espacio vectorial. Por ejemplo $\mathbb{R}$ con la suma y producto usual de números reales es un $\mathbb{R}$-espacio vectorial.

Sin embargo, los siguientes productos por escalares $(*)$ de $\mathbb{R}$ no dotan a $\mathbb{R}$ con la suma usual de estructura de $\mathbb{R}$-espacio vectorial. Lo que pone de manifiesto que, en la definición de espacio vectorial, la operación externa tan importante como la estructura de grupo.

1. Si $\lambda * \mathbf{u} = \lambda^2 \mathbf{u}$, para todo $\mathbf{u} \in \mathbb{R}$ y $\lambda \in \mathbb{R}$, entonces $(\mathbb{R}, +, *)$ no es un espacio vectorial sobre $\mathbb{R}$, pues $(\lambda + \mu) * \mathbf{u} \neq \lambda * \mathbf{u} + \mu * \mathbf{u}$.
2. Si $\lambda * \mathbf{u} = \mathbf{0}$, para todo $\mathbf{u} \in \mathbb{R}$ y $\lambda \in \mathbb{R}$, entonces $(\mathbb{R}, +, *)$ no es un espacio vectorial sobre $\mathbb{R}$, pues $1 * \mathbf{u} \neq \mathbf{u}$.

Para finalizar esta sección veamos con detalle el ejemplo de los espacios vectoriales numéricos.

Ejemplo C.1.6. Sea $n \in \mathbb{N}$ fijo. Si consideramos

$$\Bbbk^n = \{\mathbf{u} = (u_1, \ldots, u_n) \mid u_i \in \Bbbk,\ i = 1, \ldots, n\}$$

con las operaciones suma y producto por escalares definidas como sigue:

$$\begin{array}{rcll} \mathbf{u} + \mathbf{v} & = & (u_1, \ldots, u_n) + (v_1, \ldots, v_n) & := & (u_1 + v_1, \ldots, u_n + v_n); \\ \lambda \cdot \mathbf{u} & = & \lambda(u_1, \ldots, u_n) & := & (\lambda u_1, \ldots, \lambda u_n), \end{array}$$

para todo $\mathbf{u}$ y $\mathbf{v} \in \Bbbk^n$ y $\lambda \in \Bbbk$, entonces $(\Bbbk^n, +, \cdot)$ es un $\Bbbk$-espacio vectorial. En efecto, $(\Bbbk^n, +)$ es un grupo (compruébese), veamos que se verifican el resto de axiomas de espacio vectorial. Si $\mathbf{u}$ y $\mathbf{v} \in \Bbbk^n$ y λ y $\mu \in \Bbbk$, entonces

- $\lambda \cdot (\mathbf{u} + \mathbf{v}) = \lambda \cdot ((u_1, \ldots, u_n) + (v_1, \ldots, v_n)) = \lambda \cdot (u_1 + v_1, \ldots, u_n + v_n) = (\lambda(u_1 + v_1), \ldots, \lambda(u_n + v_n)) = (\lambda u_1 + \lambda v_1, \ldots, \lambda u_n + \lambda v_n) = (\lambda u_1, \ldots, \lambda u_n) + (\lambda v_1, \ldots, \lambda v_n) = \lambda(u_1, \ldots, u_n) + \lambda(v_1, \ldots, v_n) = \lambda \cdot \mathbf{u} + \lambda \cdot \mathbf{v}$.
- $(\lambda + \mu) \cdot \mathbf{u} = (\lambda + \mu) \cdot (u_1, \ldots, u_n) = ((\lambda + \mu)u_1, \ldots, (\lambda + \mu)u_n) = (\lambda u_1 + \mu u_1, \ldots, \lambda u_n + \mu u_n) = (\lambda u_1, \ldots, \lambda u_n) + (\mu u_1, \ldots, \mu u_n) = \lambda(u_1, \ldots, u_n) + \mu(u_1, \ldots, u_n) = \lambda \cdot \mathbf{u} + \mu \cdot \mathbf{u}$.
- $(\lambda \cdot \mu) \cdot \mathbf{u} = (\lambda \cdot \mu) \cdot (u_1, \ldots, u_n) = ((\lambda \cdot \mu)u_1, \ldots, (\lambda \cdot \mu)u_n) = (\lambda \cdot (\mu \cdot u_1), \ldots, \lambda \cdot (\mu \cdot u_n)) = \lambda \cdot (\mu u_1, \ldots, \mu u_n) = \lambda \cdot (\mu(u_1, \ldots, u_n) = \lambda \cdot (\mu \cdot \mathbf{u})$.
- $1 \cdot \mathbf{u} = 1 \cdot (u_1, \ldots, u_n) = (1 \cdot u_1, \ldots, 1 \cdot u_n) = (u_1, \ldots, u_n) = \mathbf{u}$.

El espacio vectorial $(\Bbbk^n, +, \cdot)$ se llama **$\Bbbk$-espacio vectorial numérico** de *dimensión* n.

2. Subespacios vectoriales

Definición C.2.1. Sea V un espacio vectorial sobre $\Bbbk$. Diremos que un subconjunto no vacío L de V es un **subespacio vectorial** de V sobre $\Bbbk$, si L, con las operaciones interna y externa de V, es un espacio vectorial.

Por consiguiente, si L es un subespacio vectorial de $(V,+,\cdot)$, entonces $(L,+)$ es un grupo (conmutativo) y la restricción del producto por escalares definido en V dota a F de estructura de espacio vectorial sobre $\Bbbk$; en resumen, todo subespacio vectorial es de modo natural un espacio vectorial.

Veamos, en primer lugar, los ejemplos más sencillos de subespacios vectoriales.

Ejemplo C.2.2.

1. Todo espacio vectorial es un subespacio vectorial de él mismo; dicho subespacio se denomina subespacio vectorial **total** ó **impropio**. Un subespacio vectorial se dice **propio** si es distinto del total.
2. Todo espacio vectorial tiene un subespacio vectorial denominado **trivial**, aquel cuyo único elemento es el vector cero.

Como ocurre con la definición de subgrupo, existen definiciones equivalentes de subespacio vectorial que facilitan las comprobaciones a efectos prácticos.

Proposición C.2.3. *Sean V un $\Bbbk$-espacio vectorial y L un subconjunto no vacío de V. Las siguientes condiciones son equivalentes:*

(a) *L es subespacio vectorial de V.*
(b) *$(L,+)$ es un subgrupo de V cerrado para el producto por escalares, es decir, $\lambda\mathbf{u}\in L$, para todo $\lambda\in\Bbbk$ y $\mathbf{u}\in L$.*
(c) *L es un conjunto cerrado para combinaciones lineales, esto es, $\lambda\mathbf{u}+\mu\mathbf{v}\in L$, para todo $\lambda,\mu\in\Bbbk$ y $\mathbf{u},\mathbf{v}\in L$.*

Demostración. $\boxed{\text{(a)}\Longrightarrow\text{(b)}}$ Como L es subespacio vectorial, en particular $(L,+)$ es un subgrupo de V y además la restricción del producto por escalares a $\Bbbk\times L$ valora en L.

$\boxed{\text{(b)}\Longrightarrow\text{(c)}}$ Sean $\lambda,\mu\in\Bbbk$ y $\mathbf{u},\mathbf{v}\in L$. Al ser L cerrado para el producto por escalares tenemos que $\lambda\mathbf{u}$ y $\mu\mathbf{v}$ están en L. De donde se sigue $\lambda\mathbf{u}+\mu\mathbf{v}\in L$, pues $(L,+)$ es subgrupo.

$\boxed{\text{(c)}\Longrightarrow\text{(a)}}$ Tenemos que

$$\lambda\mathbf{u}+\mu\mathbf{v}\in L,\ \text{para todo } \lambda,\mu\in\Bbbk \text{ y } \mathbf{u},\mathbf{v}\in L. \tag{C.2.2}$$

Tomando $\lambda=1$ y $\mu=-1$ en (C.2.2), se prueba que $\mathbf{u}-\mathbf{v}\in L$, para todo $\mathbf{u},\mathbf{v}\in L$. Luego, por la proposición B.1.11, se sigue que $(L,+)$ es subgrupo de V. Tomando ahora $\mu=0$ en (C.2.2) se obtiene que $\lambda\mathbf{u}\in L$, para todo $\lambda\in\Bbbk$ y $\mathbf{u}\in L$, es decir, que

la restricción de $\cdot$ a $\Bbbk \times L$ valora en L. De todo esto se deduce que las operaciones interna y externa de V dotan a L de estructura de espacio vectorial sobre $\Bbbk$, si más que comprobar que la aplicación $\cdot : \Bbbk \times L \longrightarrow L$ verifica lo requerido en la definición C.1.1(b), lo que se deja como ejercicio. ∎

Ejercicio C.2.4. Probar que los únicos subespacios vectoriales de un cuerpo, considerado como espacio vectorial sobre sí mismo, son el trivial y el total.

Ejemplo C.2.5. A continuación mostramos un par de ejemplos no elementales de subespacios vectoriales.

1. El conjunto $L_I = \{(a_1, \ldots, a_n) \in \Bbbk^n \mid a_i = 0, \text{ si } i \in I\}$, es un subespacio vectorial de $\Bbbk^n$, para todo $I \subseteq \{1, \ldots, n\}$.
2. Sea $A = (a_{ij}) \in \mathcal{M}_n(\Bbbk)$. Se llama **traza** de A, $\mathrm{tr}(A)$, a la suma de los elementos de la diagonal principal de A, es decir, $\mathrm{tr}(A) := \sum_{i=1}^{n} a_{ii}$. El subconjunto de $\mathcal{M}_n(\Bbbk)$ formado por la matrices de traza 0 es un subespacio vectorial de $\mathcal{M}_n(\Bbbk)$ con la suma de matrices y producto por escalares habituales.

3. Bases de un espacio vectorial. Dimensión

En esta sección definiremos un primer "invariante" intrínseco asociado a un espacio vectorial, la dimensión. Para ello será necesario introducir una serie de conceptos que nos conducirán a la noción de base de un espacio vectorial, y de aquí a la de dimensión.

Definición C.3.1. Sea V un espacio vectorial sobre $\Bbbk$. Se dice que $\mathbf{u} \in V$ es **combinación lineal** de un conjunto de vectores $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ de V, si existen $\lambda_1, \lambda_2, \ldots, \lambda_r \in \Bbbk$ tales que

$$\mathbf{u} = \lambda_1 \mathbf{v}_1 + \lambda_2 \mathbf{v}_2 + \ldots + \lambda_r \mathbf{v}_r.$$

Ejemplo C.3.2.

1. El vector $\mathbf{0} \in V$ es combinación lineal de cualquier conjunto de vectores de V.
2. El vector $\mathbf{v} := 3x^2 + 2x - 2 \in V = \Bbbk[x]$ es combinación lineal del conjunto de vectores $\{\mathbf{v}_1 := x^2, \mathbf{v}_1 := x - 1\} \subset V$.

Obsérvese que dado un conjunto finito de vectores $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ de un $\Bbbk$-espacio vectorial V, cualquier combinación lineal suya es un vector de V. Este hecho dota de sentido a la siguiente:

Notación C.3.3. Sea $S \subseteq V$ un subconjunto no vacío (no necesariamente finito). Denotaremos por $\langle S \rangle$ al conjunto de combinaciones lineales de los subconjuntos finitos de S, es decir,

$$\langle S \rangle := \{\lambda_1 \mathbf{v}_1 + \lambda_2 \mathbf{v}_2 + \ldots + \lambda_r \mathbf{v}_r \mid \lambda_i \in \Bbbk \text{y } \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\} \subseteq S\}.$$

Proposición C.3.4. *Sean V un espacio vectorial sobre $\Bbbk$ y $S \subseteq V$ un subconjunto no vacío (no necesariamente finito). El conjunto de combinaciones lineales de los subconjuntos finitos de S, $\langle S \rangle$ es el menor subespacio vectorial de V que contiene a S.*

Demostración. Por la proposición C.2.3(c), basta probar que $\langle S \rangle$ es cerrado para combinaciones lineales, es decir, $\lambda \mathbf{u} + \mu \mathbf{v} \in \langle S \rangle$, parta todo $\lambda, \mu \in \Bbbk$ y $\mathbf{u}, \mathbf{v} \in \langle S \rangle$. Sean $\mathbf{u}$ y $\mathbf{v} \in \langle S \rangle$. Como $\mathbf{u} \in \langle S \rangle$, existirá un subconjunto finito de S, $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r\}$ tal que

$$\mathbf{u} = \lambda_1 \mathbf{u}_1 + \lambda_2 \mathbf{u}_2 + \ldots + \lambda_r \mathbf{u}_r,$$

para ciertos $\lambda_i \in \Bbbk,\ i = 1, \ldots, r$, y análogamente, con

$$\mathbf{v} = \mu_1 \mathbf{v}_1 + \mu_2 \mathbf{v}_2 + \ldots + \mu_s \mathbf{v}_s,$$

para algún subconjunto finito $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_s\}$ de S y $\mu_i \in \Bbbk,\ i = 1, \ldots, s$. Por ser V un espacio vectorial se sigue que

$$\lambda \mathbf{u} + \mu \mathbf{v} = (\lambda\lambda_1)\mathbf{u}_1 + (\lambda\lambda_2)\mathbf{u}_2 + \ldots + (\lambda\lambda_r)\mathbf{u}_r + (\mu\mu_1)\mathbf{v}_1 + (\mu\mu_2)\mathbf{v}_2 + \ldots + (\mu\mu_s)\mathbf{v}_s,$$

y por consiguiente que existe un subconjunto finito de S, $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r\} \cup \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_s\}$, tal que $\lambda \mathbf{u} + \mu \mathbf{v}$ es combinación lineal suya. Esto prueba que $\lambda \mathbf{u} + \mu \mathbf{v} \in \langle S \rangle$ y por tanto que es subespacio vectorial.

Queda ver que $\langle S \rangle$ es el menor subespacio vectorial que contiene a S. Pero esto es elemental, ya que si F es un subespacio vectorial que contiene a S, entonces contiene a cualquier combinación lineal (finita) de elementos de S. De donde se sigue que $\langle S \rangle \subseteq F$. ∎

Definición C.3.5. Sean V un espacio vectorial sobre $\Bbbk$ y $F \subseteq V$ un subespacio vectorial. Si S es un subconjunto tal que $F = \langle S \rangle$, diremos que F **está generado por** S, que S es un **sistema de generadores** de F ó que S **genera** a F, indistintamente.

Nota C.3.6. Todo subespacio vectorial F de V posee un sistema de generadores, ya que, por ejemplo, $F = \langle F \rangle$.

Ejemplo C.3.7. Veamos alguno ejemplos que ilustran el concepto de sistema de generadores.

1. Sea V un $\Bbbk$-espacio vectorial cualquiera y F el subespacio vectorial trivial, es decir, $F = \langle \mathbf{0} \rangle$. El conjunto cuyo único elemento es el vector cero, $S = \{0\}$, es un sistema de generadores de F.
2. Sea $V = \mathbb{R}^3$ y consideramos el subespacio vectorial $F = \{(0, a_2, a_3) \mid a_2, a_3 \in \mathbb{R}\}$. Los conjuntos $S_1 = \{\mathbf{v}_1 = (0, 1, 0), \mathbf{v}_2 = (0, 0, 1)\}$, $S_2 = \{\mathbf{v}_1 = (0, 2, 0), \mathbf{v}_2 = (0, 1/2, -1)\}$ y $S_3 = \{\mathbf{v}_1 = (0, 1, 1), \mathbf{v}_2 = (0, 1, -1), \mathbf{v}_3 = (0, 2, 3)\}$ son (cada uno ellos) sistemas de generadores de F. El conjunto $S_4 = \{\mathbf{v}_1 = (0, 1, 1)\}$

no genera a F. Obsérvese que la primera parte de este ejemplo señala que un mismo subespacio vectorial puede tener distintos sistemas de generadores.

3. Sea $V = \Bbbk[x]$. Los conjuntos de vectores $S_1 = \{1, x, x^2, \ldots, x^n, \ldots\}$ y $S_2 = \{1, (x-7), (x-7)^2, \ldots, (x-7)^n, \ldots\}$ son (cada uno) sistemas de generadores del subespacio vectorial impropio, es decir, del mismo V.
4. Sea $V = \Bbbk[x]$ y consideramos el subespacio vectorial $F = \Bbbk[x]_{\leq n}$. Los conjuntos de vectores $S_1 = \{1, x, x^2, \ldots, x^n\}$ y $S_2 = \{1, x, x^2, \ldots, x^n, x-1, x-2, \ldots, x-m, \ldots, \}$ son (cada uno) sistemas de generadores de F.
5. Sea $\Bbbk = \mathbb{Q}$ y consideramos $\mathbb{R}$ con estructura de $\mathbb{Q}$-espacio vectorial. El conjunto de vectores $S = \{1\} \cup (\mathbb{R} \setminus \mathbb{Q})$ es un sistema de generadores de $\mathbb{R}$ como $\mathbb{Q}$-espacio vectorial.

Antes vimos (véase la nota C.3.6) que todo subespacio vectorial tiene, al menos, un sistema de generadores. Además, en el caso de los espacios vectoriales se puede hablar de sistemas de "minimales" generadores. La siguiente definición precisará que entenderemos por "minimal".

Definición C.3.8. Sean V un espacio vectorial sobre $\Bbbk$ y $S \subseteq V$ un subconjunto no vacío (no necesariamente finito). Se dice que S es un conjunto **linealmente independiente** ó **libre** si toda combinación lineal (finita) de vectores de S nula tiene sus escalares nulos. Es decir,

$$\lambda_1 \mathbf{v}_1 + \lambda_2 \mathbf{v}_2 + \ldots + \lambda_r \mathbf{v}_r = \mathbf{0} \Longrightarrow \lambda_1 = \lambda_2 = \ldots = \lambda_r = 0,$$

para todo subconjunto finito $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\} \subseteq S$.

En otro caso se dice que S es un conjunto **linealmente dependiente**.

Ejemplo C.3.9. Sea V un $\Bbbk$-espacio vectorial. Un subconjunto S de V formado por único vector $\mathbf{v} \in V$, esto es $S = \{\mathbf{v}\}$, es linealmente independiente si, y sólo si, $\mathbf{v} \neq \mathbf{0}$.

Proposición C.3.10. *Sean V un espacio vectorial sobre $\Bbbk$ y $S \subseteq V$ un subconjunto no vacío (no necesariamente finito). Se cumple que:*

(a) *S es linealmente independiente si, y sólo si, $\mathbf{v} \notin \langle S \setminus \mathbf{v} \rangle$, para todo $\mathbf{v} \in S$.*
(b) *S es linealmente dependiente si, y sólo si, existe $\mathbf{v} \in S$ tal que $\mathbf{v} \in \langle S \setminus \mathbf{v} \rangle$.*

Demostración. Teniendo en cuenta que la equivalencia del apartado (a) es la negación de la del apartado (b), y viceversa, es suficiente demostrar una de las dos.

(b) Si S es linealmente dependiente, entonces existe $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\} \subseteq S$ tal que $\lambda_1 \mathbf{v}_1 + \lambda_2 \mathbf{v}_2 + \ldots + \lambda_r \mathbf{v}_r = \mathbf{0}$ con $\lambda_j \neq 0$ para algún $j \in \{1, 2, \ldots, r\}$ Sin perdida de generalidad, podemos suponer $\lambda_1 \neq 0$, en otro caso reordenaríamos los subíndices del conjunto $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$. Por consiguiente, como $\lambda_1 \mathbf{v}_1 + \lambda_2 \mathbf{v}_2 + \ldots + \lambda_r \mathbf{v}_r = \mathbf{0}$ implica

$\lambda_1\mathbf{v}_1 = -\lambda_2\mathbf{v}_2 - \ldots - \lambda_r\mathbf{v}_r$, y dado que $\lambda_1 \neq 0$, se sigue que $\mathbf{v}_1 = -\frac{\lambda_2}{\lambda_1}\mathbf{v}_2 - \ldots - \frac{\lambda_r}{\lambda_1}\mathbf{v}_r$ es un elemento de $\langle S \setminus \mathbf{v}_1\rangle$.

Recíprocamente, si $\mathbf{v} \in \langle S \setminus \mathbf{v}_1\rangle$, entonces existe un subconjunto finito $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ de S, tal que $\mathbf{v} = \lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 + \ldots + \lambda_r\mathbf{v}_r$, para ciertos $\lambda_i \in mathbbmssk$, y $\mathbf{v}_i \neq \mathbf{v}$, para todo $i = 1, 2, \ldots, r$. Ahora bien, como $\mathbf{v} = \lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 + \ldots + \lambda_r\mathbf{v}_r$ implica $\mathbf{0} = -\mathbf{v} + \lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 + \ldots + \lambda_r\mathbf{v}_r$ y $\mathbf{v}_i \neq \mathbf{v}$, para todo $i = 1, 2, \ldots, r$. Luego tenemos una combinación lineal nula de vectores de S con al menos un escalar no nulo (el -1 que acompaña a $\mathbf{v}$). Por consiguiente S es un conjunto linealmente dependiente. ∎

Corolario C.3.11. *Sean V un espacio vectorial sobre $\Bbbk$ y $S \subseteq V$ un subconjunto no vacío (no necesariamente finito). Si $\mathbf{0} \in S$, entonces S es linealmente dependiente.*

Demostración. Por la proposición C.3.10, la prueba es inmediata pues $\mathbf{0} \in \langle S \setminus \{\mathbf{0}\}\rangle$, ya que el vector cero pertenece a todo subespacio vectorial. ∎

Ejemplo C.3.12. Volviendo al ejemplo C.3.7 tenemos que:

1. El conjunto $S = \{0\}$ no es linealmente independiente.
2. Los conjuntos S_1, S_2 y S_4 son linealmente independientes, pero S_3 no lo es, es decir, S_3 es linealmente dependiente.
3. Los conjuntos S_1 y S_2 son conjuntos (con infinitos elementos) linealmente independientes. Pero, si tomamos $S_1 \cup S_2$ obtenemos un conjunto linealmente dependiente.
4. El conjunto de vectores S_1 es linealmente independiente y el conjunto de vectores S_2 es linealmente dependiente.
5. El conjunto de vectores S es no linealmente independiente, ya que por ejemplo $\{\pi, 2\pi\} \subset S$ no es linealmente independiente.

Si nos fijamos en el ejemplo anterior, observamos que la independencia lineal define una relación de orden (relación $\leq$ que verifica las propiedades reflexiva y transitiva y tal que $x \leq y$ e $y \geq x$ simultáneamente implica $x = y$) en el conjunto de todos los sistemas de generadores de un subespacio vectorial dado; $S \leq S' \iff \langle S\rangle = \langle S'\rangle$ y $S \subseteq S'$. Lo que nos permite definir un concepto de minimalidad entre sistemas de generadores: un sistema de generadores de un subespacio vectorial L es "minimal" si no existe ningún otro sistema de generadores de L contenido dentro de él. Un sistema de generadores "minimal" es lo que llamaremos una base.

Definición C.3.13. Sean V un $\Bbbk$-espacio vectorial, L un subespacio vectorial y $\mathcal{B}$ un conjunto de vectores de V. Diremos que $\mathcal{B}$ es una **base** L si genera a L y es linealmente independiente.

Nota C.3.14. Obsérvese que si S es un conjunto linealmente independiente, entonces es base de $\langle S\rangle$.

Ejemplo C.3.15. Por la definición de base de un subespacio vectorial y a la vista de los ejemplos C.3.7 y C.3.12, tenemos que:

1. El subespacio vectorial trivial no tiene base.
2. Los conjuntos S_1 y S_2 son bases de L. Luego un subespacio vectorial puede tener mas de una base.
3. Los conjuntos S_1 y S_2 son bases de $\Bbbk[x]$. Por lo tanto, hay bases con infinitos vectores.
4. El conjunto S_1 es una base de $\Bbbk[x]_{\leq n}$, es decir que un espacio vectorial con bases de infinitos vectores, contiene subespacios vectoriales cuyas bases tienen un número finito de vectores.
5. S no es una base de $\mathbb{R}$ como $\mathbb{Q}$-espacio vectorial.

Ejercicio C.3.16. Probar que una base del espacio vectorial de matrices de orden 2×3 con coeficientes en $\Bbbk$ es

$$\left\{\begin{pmatrix}1&0&0\\0&0&0\end{pmatrix},\begin{pmatrix}0&1&0\\0&0&0\end{pmatrix},\begin{pmatrix}0&0&1\\0&0&0\end{pmatrix},\begin{pmatrix}0&0&0\\1&0&0\end{pmatrix},\begin{pmatrix}0&0&0\\0&1&0\end{pmatrix},\begin{pmatrix}0&0&0\\0&0&1\end{pmatrix}\right\}.$$

La relevancia de las bases en espacios vectoriales no sólo radica en el hecho de que corresponda con la idea de sistema "minimal" de generadores minimal. El siguiente resultado muestra una importante propiedad de las bases que será fundamental en el transcurso de este curso.

Proposición C.3.17. *Sean V un $\Bbbk$-espacio vectorial, L un subespacio vectorial y $\mathcal{B}$ un conjunto de vectores de L. $\mathcal{B}$ es base de L si, y sólo si, todo vector de L se expresa de manera única como combinación lineal de elementos de $\mathcal{B}$.*

Demostración. Si $\mathcal{B}$ es base de L, en particular es sistema de generadores de L. Luego, dado $\mathbf{v}\in L$, existe $\{\mathbf{v}_1,\ldots,\mathbf{v}_r\}\subseteq\mathcal{B}$ tal que $\mathbf{v}=\lambda_1\mathbf{v}_1+\ldots\lambda_r\mathbf{v}_r$, para ciertos $\lambda_i\in\Bbbk,\ i=1,\ldots,r$. Sea $\{\mathbf{u}_1,\ldots,\mathbf{u}_s\}\subseteq\mathcal{B}$, otro conjunto de vectores tal que $\mathbf{v}=\mu_1\mathbf{u}_1+\ldots\mu_s\mathbf{u}_1+\ldots\mu_s\mathbf{u}_s$ para ciertos $\mu_i\in\Bbbk,\ i=1,\ldots,s$. Si un vector $\mathbf{v}_j$ no aparece en la segunda expresión, añadimos a ésta el sumando $0\mathbf{v}_j$; análogamente, si un $\mathbf{u}_j$ no aparece en la primera expresión, añadimos a ésta el sumando $0\mathbf{u}_j$. Consiguiendo de este modo dos combinaciones lineales de los mismos vectores, es decir,

$$\mathbf{v}=\lambda_1\mathbf{v}_1+\ldots+\lambda_r\mathbf{v}_r+\lambda_{r+1}\mathbf{v}_{r+1}+\ldots+\lambda_m\mathbf{v}_m=\mu_1\mathbf{u}_1+\ldots+\mu_s\mathbf{u}_s+\mu_{s+1}\mathbf{u}_{s+1}+\ldots+\mu_m\mathbf{u}_m$$

con $\{\mathbf{v}_1,\ldots,\mathbf{v}_m\}=\{\mathbf{u}_1,\ldots,\mathbf{u}_m\}$. Así, reordenando los subíndices de la segunda expresión si fuese necesario, obtenemos que $\mathbf{v}=\lambda_1\mathbf{v}_1+\ldots+\lambda_m\mathbf{v}_m$, que $\mathbf{v}=\mu_1\mathbf{v}_1+\ldots+\mu\mathbf{v}_m$ y, restando ambas expresiones, que

$$\mathbf{0}=(\lambda_1-\mu_1)\mathbf{v}_1+\ldots+(\lambda_m-\mu_m)\mathbf{v}_m.$$

El conjuto de vectores $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ está contenido en la base $\mathcal{B}$ que, en particular, es un conjunto linealmente independiente. Por la definición C.3.8, se sigue $\lambda_1 - \mu_1 = \ldots = \lambda_m - \mu_m = 0$, es decir, $\lambda_1 = \mu_1, \ldots, \lambda_m = \mu_m = 0$.

Recíprocamente, si todo vector de L se expresa como combinación lineal de elementos de $\mathcal{B}$, entonces, por la proposición C.3.4, tenemos que $\mathcal{B}$ genera a L. Por otro lado, si $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ es un subconjunto de vectores de $\mathcal{B}$ tal que $\mathbf{0} = \lambda_1\mathbf{v}_1 + \ldots + \lambda_r\mathbf{v}_r$, dado que también $\mathbf{0} = 0\mathbf{v}_1 + \ldots + 0\mathbf{v}_r$ y que la expresión debe ser única, se sigue $\lambda_1 = \ldots = \lambda_r = 0$. Luego $\mathcal{B}$ es linealmente independiente. ∎

Sabemos, por la proposición C.3.17, que todo vector $\mathbf{v} \in V$ se expresa de forma única como combinación lineal de los vectores de $\mathcal{B}$; es decir, existen unos únicos $\lambda_1, \ldots, \lambda_n \in \Bbbk$ tales que $\mathbf{v} = \lambda_1\mathbf{v}_1 + \ldots + \lambda_n\mathbf{v}_n$, llamados **coordenadas de $\mathbf{v} \in V$ respecto de $\mathcal{B}$**.

En lo que sigue centraremos nuestra atención en aquellos espacios vectoriales que está generados por un número finito de vectores. Probaremos que las bases de éstos son siempre finitas y que el número de vectores en cualquiera de sus bases es una constante. A esta constante la llamaremos dimensión del espacio vectorial.

Definición C.3.18. Sea V un espacio vectorial sobre $\Bbbk$. Diremos que V es de **dimensión finita** si posee sistemas de generadores finitos. En caso contrario diremos que es de **dimensión infinita**.

Proposición C.3.19. *Sea V un $\Bbbk$-espacio vectorial no trivial de dimensión finita. Si S es un sistema de generadores finito de V, entonces existe un subconjunto de S que es base de V. Es decir, todo espacio vectorial de dimensión finita tiene una base finita.*

Demostración. En primer lugar, conviene resaltar que existe $\mathbf{v} \in S$ no nulo, ya que $V = \langle S\rangle$ y $V \neq \{0\}$. Luego, al menos, hay un subconjunto S que es linealmente independiente.

Como V es de dimensión finita podemos asegurar que existe un conjunto finito $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ que genera a V. Si S es linealmente independiente, entonces $\mathcal{B} = S$ es una base de V. En caso contrario, hay al menos uno que es combinación lineal de los otros. Sin perdida de generalidad podemos suponer que éste es $\mathbf{v}_1$. Entonces $V = \langle S\rangle = \langle \mathbf{v}_2, \ldots, \mathbf{v}_r\rangle$. Si este nuevo conjunto es linealmente independiente, es una base de V. En otro caso, podemos volver a suprimir uno de ellos, obteniendo otro sistema de generadores de V. Repitiendo el proceso tantas veces como sea necesario, eliminando aquellos generadores que sean combinación lineal del resto. Llegaremos de esta manera a conseguir un conjunto linealmente independiente que genera a V, es decir, una base de V. ∎

Teorema C.3.20. ***de Steinitz.*** *Si V es un $\Bbbk$-espacio vectorial no trivial de dimensión finita, entonces cualesquiera dos bases finitas de V tienen el mismo número de vectores.*

Demostración. Sean $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ y $\mathcal{B}' = \{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ dos bases de V, y suponemos $n \leq m$.

Sustituiremos uno por uno n vectores de la base $\mathcal{B}'$ por los n vectores de la base $\mathcal{B}$.

Por ser $\mathcal{B}'$ un sistema de generadores de V tenemos que $\mathbf{v}_1 = \lambda_1\mathbf{u}_1 + \ldots + \lambda_m\mathbf{u}_m$, para ciertos $\lambda_i \in \Bbbk$. Como $\mathbf{v}_1 \neq \mathbf{0}$, al menos uno de los λ_j es distinto de cero. Sin perdida de generalidad podemos suponer $\lambda_1 \neq 0$. Entonces

$$\mathbf{u}_1 = \lambda_1^{-1}\mathbf{v}_1 + (\lambda_1^{-1}\lambda_2)\mathbf{u}_2 + \ldots + (\lambda_1^{-1}\lambda_m)\mathbf{u}_m.$$

Esta expresión asegura que $\langle \mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\rangle = \langle \mathbf{v}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\rangle$ y por consiguiente que $\{\mathbf{v}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\}$ genera a V. Además, $\{\mathbf{v}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\}$ es linealmente independiente. En efecto, si $\mathbf{0} = \mu_1\mathbf{v}_1 + \mu_2\mathbf{u}_2 + \ldots + \mu_m\mathbf{u}_m = \mu_1\left(\sum_{i=1}^m \lambda_i\mathbf{u}_i\right) + \mu_2\mathbf{u}_2 + \ldots + \mu_m\mathbf{u}_m = \mu_1\lambda_1\mathbf{u}_1 + (\mu_1\lambda_2 + \mu_2)\mathbf{u}_2 + \ldots + (\mu_1\lambda_m + \mu_m)\mathbf{u}_m$, entonces $\mu_1\lambda_1 = 0$ y $\mu_1\lambda_i + \mu_i = 0,\ i = 2, \ldots, m$, pues $\mathcal{B}'$ es linealmente independiente. Pero $\lambda_1 \neq 0$. Por tanto $\mu_1 = 0$ y $\mu_i = 0,\ i = 2, \ldots, m$. Así pues,$\{\mathbf{v}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\}$ es una base de V.

Tenemos que $\{\mathbf{v}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\}$ es una nueva base de V. Procedamos igual que antes y expresemos $\mathbf{v}_2$ como combinación lineal de esta base: $\mathbf{v}_2 = \lambda_1\mathbf{v}_1 + \lambda_2\mathbf{u}_2 + \ldots + \lambda_m\mathbf{u}_m$, para ciertos $\lambda_i \in \Bbbk$. A la vista de lo anterior, sólo tenemos que probar que $\mathbf{v}_2$ se puede sustituir por alguno de los $\mathbf{u}_j,\ j = 2, \ldots, m$. Para ello, y a la vista de lo anterior, basta asegurar que algún $\lambda_j,\ j = 2, \ldots, m$, es distinto de cero. Pero si fuese $\lambda_2 = \ldots = \lambda_m = 0$, entonces $\mathbf{v}_2 = \lambda_1\mathbf{v}_1$, es decir, $\mathbf{v}_2$ sería combinación lineal de $\{\mathbf{v}_1, \mathbf{v}_3, \ldots, \mathbf{v}_n\}$ y esto no es posible por ser $\mathcal{B}$ base.

Siguiendo el proceso descrito arriba sustituimos n vectores de la base $\mathcal{B}'$ por los vectores de $\mathcal{B}$, y reordenando los subíndices de los vectores de $\mathcal{B}'$ podemos suponer que hemos cambiado los n primeros vectores de $\mathcal{B}'$. Así obtenemos que $\mathcal{B}'' = \{\mathbf{v}_1, \ldots, \mathbf{v}_n, \mathbf{u}_{n+1}, \ldots \mathbf{u}_m\}$ es una base de V. Pero $\{\mathbf{u}_{n+1}, \ldots, \mathbf{u}_m\} \subseteq V = \langle\mathcal{B}\rangle = \langle \mathbf{v}_1, \ldots, \mathbf{v}_n\rangle$. Luego, necesariamente, $m = n$ y $\mathcal{B}'' = \mathcal{B}$. ∎

Corolario C.3.21. *Si V es un $\Bbbk$-espacio vectorial no trivial de dimensión finita, entonces cualesquiera dos bases de V tienen el mismo número de vectores. Es decir, en un espacio vectorial de dimensión finita distinto del trivial todas las bases son finitas y tienen el mismo número de vectores.*

Demostración. Basta repetir la demostración del teorema de Steinitz (teorema C.3.20) con $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ y $\mathcal{B}' = \{\mathbf{u}_j \mid j \in J\}$, con J es un conjunto arbitrario de índices. ∎

Este último corolario permite definir sin ambigüedad el concepto de dimensión.

Definición C.3.22. Llamaremos **dimensión** de un $\Bbbk$-espacio vectorial no trivial de dimensión finita V, y la denotaremos por $\dim_{\Bbbk} V$ (ó simplemente $\dim V$), al número de elementos de las bases de V.

Por convenio, se define la dimensión del espacio vectorial trivial como cero, es decir, $\dim\langle \mathbf{0}\rangle = 0$.

Ejemplo C.3.23.

1. La dimensión de $\Bbbk$ como $\Bbbk$-espacio vectorial es 1. Por ejemplo, $\mathbb{R}$ tiene dimensión 1 como $\mathbb{R}$-espacio vectorial. Sin embargo, tiene dimensión infinita como $\mathbb{Q}$-espacio vectorial.
2. $\Bbbk^n$ es un $\Bbbk$-espacio vectorial de dimensión n.
3. $\Bbbk[x]$ es un $\Bbbk$-espacio vectorial de dimensión infinita.
4. $\Bbbk[x]_{\leq n}$ es un $\Bbbk$-espacio vectorial dimensión $n+1$.
5. $\mathcal{M}_{m\times n}(\Bbbk)$ es un espacio vectorial de dimensión $m \cdot n$ sobre $\Bbbk$.

Por la proposición C.3.19 podemos afirmar que la dimensión de un espacio vectorial V coincide con el menor número de vectores que generan a V. Veamos que la dimensión de V también se puede entender como el mayor número de vectores linealmente independientes en V.

Proposición C.3.24. *Sean V un espacio vectorial sobre $\Bbbk$ y $\mathbf{v} \in V$. Si $S' \subseteq V$ es subconjunto linealmente independiente (no necesariamente finito) tal que $\mathbf{v} \notin \langle S'\rangle$, entonces $S = S' \cup \{\mathbf{v}\}$ también es linealmente independiente.*

Demostración. Consideramos $\lambda\mathbf{v} + \lambda_1\mathbf{v}_1 + \ldots + \lambda_r\mathbf{v}_r = \mathbf{0}$, donde $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\} \subset S'$, $\lambda \in \Bbbk$ y $\lambda_i \in \Bbbk$, $i = 1, \ldots, r$. Si $\lambda \neq 0$, entonces existe $\lambda^{-1} \in \Bbbk$ y por tanto $\mathbf{v} = -(\lambda^{-1}\lambda_1)\mathbf{v}_1 - \ldots - (\lambda^{-1}\lambda_r)\mathbf{v}_r \in \langle S'\rangle$, en contra de la hipótesis $\mathbf{v} \notin \langle S'\rangle$. Por tanto $\lambda = 0$. Pero entonces tenemos $\lambda_1\mathbf{v}_1 + \ldots + \lambda_r\mathbf{v}_r = \mathbf{0}$, con $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\} \subset S'$ y $\lambda_i \in \Bbbk$, $i = 1, \ldots, r$. Por ser S' linealmente se sigue que $\lambda_1 = \ldots = \lambda_r = 0$. ∎

Corolario C.3.25. *Si V es un $\Bbbk$-espacio vectorial no trivial de dimensión finita. Todo conjunto linealmente independiente de vectores de V o es una base de V o se puede ampliar a una base del espacio vectorial.*

Demostración. La prueba es inmediata a partir del teorema de Steinitz (teorema C.3.20 y la proposición C.3.24. ∎

Ejercicio C.3.26. Sea V un $\Bbbk$-espacio vectorial de dimensión n. Probar las siguientes afirmaciones:

1. Todo subconjunto linealmente independiente de n vectores es una base de V.

2. Todo conjunto de más de n vectores es linealmente dependiente.
3. Todo sistema de generadores de V tiene al menos n vectores.
4. Todo sistema de generadores con n elementos es una base de V.

Para terminar esta sección veamos que ocurre con la dimensión de los subespacios de un espacio vectorial de dimensión finita.

Proposición C.3.27. *Sea V un $\Bbbk$-espacio vectorial de dimensión finita. Si L es un subespacio vectorial de V, entonces L tiene dimensión finita y* $\dim L \leq \dim V$.

Demostración. Si $\mathcal{B}$ una base de L, en particular es un subconjunto de vectores de V linealmente independiente. Luego, por el corolario C.3.25 es ampliable a una base de V. De donde se sigue que $\mathcal{B}$ tiene, a lo sumo, tanto elementos como $\dim V$. ∎

Definición C.3.28. Sean V un $\Bbbk$-espacio vectorial de dimensión finita y L un subespacio vectorial de V. Se llama **Rango** de L, y se denota por $\text{rango}(L)$, a su dimensión como $\Bbbk$-espacio vectorial, es decir, $\text{rango}(L) = \dim L$.

Corolario C.3.29. *Sean V un $\Bbbk$-espacio vectorial de dimensión finita y L un subespacio vectorial de V. Toda base de L es ampliable a una base de V.*

Demostración. Sigue del corolario C.3.25. ∎

Corolario C.3.30. *Sean V un $\Bbbk$-espacio vectorial de dimensión finita y L un subespacio vectorial de V.* $\dim L = \dim V$, *si, y sólo si,* $L = V$.

Demostración. Si $L = V$ entonces, es claro, que $\dim L = \dim V$. Recíprocamente, si $\dim L = \dim V$, entonces toda base $\mathcal{B}$ de L es base de V. En otro caso, sería ampliable y por tanto $\dim L < \dim V$. Luego $L = \langle \mathcal{B} \rangle = V$. ∎

Anexo. Bases en un espacio vectorial de dimensión infinita.

Aunque en esta sección hemos centrado nuestra atención en los espacios vectoriales de dimensión finita con el objeto de definir su dimensión, se puede probar la existencia de bases para cualquier espacio vectorial independientemente de su dimensión. Es decir, todo espacio vectorial distinto del trivial tiene base.

Añadimos en este apartado la demostración de tal resultado, advirtiendo al lector que la clave de la prueba se base en el **Lema de Zorn**[1]

Teorema C.3.31. *Todo $\Bbbk$-espacio vectorial V distinto del trivial tiene base.*

[1] M.F.Atiyah, I.G.Macdonald, *Introducción al álgebra conmutativa* p.4. "Sea S un conjunto no vacío parcialmente ordenado (es decir se ha dado una relación $x \leq y$ en S que es reflexiva y transitiva y tal que $x \leq y$ e $y \geq x$ simultáneamente implica $x = y$). Un subconjunto T de S es una *cadena* si o $x \leq y$ o $y \leq x$ para cada par de elementos x, y en T. El Lema de Zorn se puede establecer como

Demostración. Sea Σ el conjunto de todos los subconjuntos linealmente independientes de V. Se ordena Σ por inclusión. Σ es no vacío, pues $\{\mathbf{v}\} \in \Sigma$, para todo $\mathbf{v} \in V$ no nulo. Para aplicar el lema de Zorn, se ha de probar que toda cadena en Σ tiene cota superior; sea $\{S_i\}_{i\in I}$ una cadena de subconjuntos de V linealmente independientes de forma que para cada par de índices j, k se tiene ó $S_j \subseteq S_k$ o $S_k \subseteq S_j$. Sea $S = \cup_{i\in I} S_i$. Entonces S subconjunto de V que es linealmente independiente (compruébese). Por tanto $S \in \Sigma$ y es una cota superior de la cadena. Por virtud del lema de Zorn Σ tiene elemento maximal, este elemento maximal es necesariamente una base de V. ∎

4. Intersección y suma de subespacios vectoriales

Similarmente a lo que ocurría con los subgrupos, tenemos que la **intersección de subespacios vectoriales** es siempre un subespacio vectorial:

Proposición C.4.1. *Sean V un $\Bbbk$-espacio vectorial. Si L_1 y L_2 son dos subespacios vectoriales de V, entonces $L_1 \cap L_2$ es un subespacio vectorial de V.*

Demostración. Por la proposición B.1.12, tenemos que $(L_1 \cap L_2, +)$ es un subgrupo de $(V, +)$. De modo que, por la proposición C.2.3(b), queda ver que el grupo $L_1 \cap L_2$ es cerrado para el producto por escalares. Sean $\mathbf{u} \in L_1 \cap L_2$ y $\lambda \in \Bbbk$. Como $\mathbf{u} \in L_1$ y $\mathbf{u} \in L_2$, y ambos son subespacios vectoriales, se sigue que $\lambda\mathbf{u} \in L_1$ y $\lambda\mathbf{u} \in L_2$. Luego $\lambda\mathbf{u} \in L_1 \cap L_2$. ∎

Ejercicio C.4.2. Generalizar el resultado anterior a cualquier intersección finita de subespacios vectoriales.

Como, en general, la unión de subgrupos no es un subgrupo, la unión de subespacios vectoriales no va a ser subespacio vectorial (véase la proposición C.2.3(b)). De modo que, para evitar trabajar con conjuntos que no son subespacios vectoriales, consideramos, en lugar de la unión, el subespacio vectorial generado por la unión. Veremos que este subespacio vectorial coincide con la noción de suma de subgrupos.

Proposición C.4.3. *Sean V un $\Bbbk$-espacio vectorial. Si L_1 y L_2 son dos subespacios vectoriales de V, entonces*

$$\langle L_1 \cup L_2 \rangle = \{\mathbf{u} + \mathbf{v} \ \mid \mathbf{u} \in L_1, \ \mathbf{v} \in L_2\}.$$

sigue: si cada cadena T de S tiene una cota superior en S (es decir, si existe un $x \in S$ tal que $t \leq x$ para todo $t \in T$), entonces S tiene, por lo menos, un elemento maximal.

Para una demostración de la equivalencia del Lema de Zorn con el axioma de elección, con el principio de buena de ordenación, etc. ver, por ejemplo, Paul R. Halmos. *Naive Set Theory.* Undergraduate Texts in Mathematics. Springer-Verlag 1974.

Demostración. Sabemos que la suma como subgrupos de L_1 y L_2, es decir $L_1 + L_2 = \{\mathbf{u} + \mathbf{v} \mid \mathbf{u} \in L_1,\ \mathbf{v} \in L_2\}$, es el menor subgrupo que contiene a L_1 y a L_2. Veamos además que es cerrada para el producto por escalares. En efecto, si $\lambda \in \Bbbk$ y $\mathbf{u} \in L_1 + L_2$, entonces existen $\mathbf{u}_1 \in L_1$ y $\mathbf{u}_2 \in L_2$ tales que $\mathbf{u} = \mathbf{u}_1 + \mathbf{u}_2$, por tanto $\lambda\mathbf{u} = \lambda(\mathbf{u}_1 + \mathbf{u}_2) = \lambda\mathbf{u}_1 + \lambda\mathbf{u}_2 \in L_1 + L_2$, pues L_1 y L_2 son subespacios vectoriales. Así, por la proposición C.2.3(b), tenemos que $L_1 + L_2$ es subespacio vectorial. De hecho tiene que ser el menor subespacio vectorial de V que contiene a L_1 y a L_2, luego, por definición, $\langle L_1 \cup L_2\rangle = L_1 + L_2$. ∎

Definición C.4.4. Sean V un $\Bbbk$-espacio vectorial. Si L_1 y L_2 son dos subespacios vectoriales de V, llamaremos la **suma** de L_1 y L_2, y la denotaremos por $L_1 + L_2$, a $\langle L_1 \cup L_2\rangle$.

En general, si $\{L_1, \ldots, L_r\}$ es una familia finita de subespacios vectoriales de V, se define la suma de $L_1, \ldots, L_r$, y se denota por $L_1 + \ldots + L_r$, como $\langle L_1 \cup \ldots \cup L_r\rangle$.

Ejercicio C.4.5. Sean V un $\Bbbk$-espacio vectorial y L_1 y L_2 dos subespacios vectoriales de V. Probar que, si $\mathcal{B}_1$ y $\mathcal{B}_2$ son bases de L_1 y L_2, respectivamente, entonces $\mathcal{B}_1 \cup \mathcal{B}_2$ genera a $L_1 + L_2$, pero, en general, no es base de $L_1 + L_2$.

Veamos a continuación el resultado principal de esta sección, conocido como **fórmula para la dimensión de la suma** ó **fórmula de Grassmann**.

Teorema C.4.6. ***(de Grassmann)***. *Sea V un $\Bbbk$-espacio vectorial. Si L_1 y L_2 son dos subespacios de V de dimensión finita, entonces $L_1 \cap L_2$ y $L_1 + L_2$ son de dimensión finita y*

$$\dim(L_1 + L_2) = \dim L_1 + \dim L_2 - \dim(L_1 \cap L_2).$$

Demostración. En primer lugar, como $L_1 \cap L_2$ es un subespacio vectorial de L_1 (y de L_2) y L_1 es de dimensión finita, podemos asegurar, por la proposición C.3.27, que $L_1 \cap L_2$ también tiene dimensión menor o igual que $\dim L_1$ (y que $\dim L_2$), y por lo tanto que es de dimensión finita. Sean $m = \dim(L_1 \cap L_2)$, $r = \dim L_1$ y $s = \dim L_2$, con $m \leq r$ y $m \leq s$. Dada una base $\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ de $L_1 \cap L_2$, por el corolario C.3.29, podemos ampliarla a una base de L_1 y a una base de L_2 : $\mathcal{B}_1 = \{\mathbf{u}_1, \ldots, \mathbf{u}_m, \mathbf{v}_1, \ldots, \mathbf{v}_{r-m}\}$ base de L_1 y $\mathcal{B}_2 = \{\mathbf{u}_1, \ldots, \mathbf{u}_m, \mathbf{w}_1, \ldots, \mathbf{w}_{s-m}\}$ base de L_2. Si probamos que

$$\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2 = \{\mathbf{u}_1, \ldots, \mathbf{u}_m, \mathbf{v}_1, \ldots, \mathbf{v}_{r-m}, \mathbf{w}_1, \ldots, \mathbf{w}_{s-m}\}$$

es base de $L_1 + L_2$, habremos terminado ya que tendríamos que $\dim(L_1 + L_2) = m + (r - m) + (s - m) = r + s - m \leq \infty$. Veamos que efectivamente $\mathcal{B}$ es base de $L_1 + L_2$.

Por el ejercicio C.4.5, tenemos que $L_1 + L_2 = \langle \mathcal{B}_1 \cup \mathcal{B}_2 \rangle = \langle \mathcal{B} \rangle$. Luego sólo nos queda probar que, $\mathcal{B}$ es linealmente independiente. Sea pues

$$\lambda_1 \mathbf{u}_1 + \ldots + \lambda_m \mathbf{u}_m + \mu_1 \mathbf{v}_1 + \ldots + \mu_{r-m} \mathbf{v}_{r-m} + \nu_1 \mathbf{w}_1 + \ldots + \nu_{s-m} \mathbf{w}_{s-m} = \mathbf{0}. \quad \text{(C.4.3)}$$

Entonces

$$\nu_1 \mathbf{w}_1 + \ldots + \nu_{s-m} \mathbf{w}_{s-m} = -\left(\lambda_1 \mathbf{u}_1 + \ldots + \lambda_m \mathbf{u}_m + \mu_1 \mathbf{v}_1 + \ldots + \mu_{r-m} \mathbf{v}_{r-m}\right).$$

Como el segundo miembro de la igualdad es un vector de L_1, entonces el primer miembro es un vector de L_1 que está en L_2, pues es combinación lineal de vectores de $\mathcal{B}_2$. Luego $\nu_1 \mathbf{w}_1 + \ldots + \nu_{s-m} \mathbf{w}_{s-m} \in L_1 \cap L_2$ y por tanto existen $\alpha_1, \ldots, \alpha_m \in \Bbbk$ tales que $\nu_1 \mathbf{w}_1 + \ldots + \nu_{s-m} \mathbf{w}_{s-m} = \alpha_1 \mathbf{u}_1 + \ldots \alpha_m \mathbf{u}_m$, y por ser $\mathcal{B}_2$ base de L_2, resulta $\nu_i = 0,\ i = 1, \ldots, s-m$ y $\alpha_j = 0,\ j = 1, \ldots, m$. Entonces, volviendo a (C.4.3), tenemos que

$$\lambda_1 \mathbf{u}_1 + \ldots + \lambda_m \mathbf{u}_m + \mu_1 \mathbf{v}_1 + \ldots + \mu_{r-m} \mathbf{v}_{r-m} = \mathbf{0},$$

que es una combinación lineal nula de vectores de $\mathcal{B}_1$. Por tanto, $\lambda_1 = \ldots = \lambda_m = \mu_1 = \ldots = \mu_{m-r} = 0$. En resumen, hemos probado que los coeficientes de la combinación lineal (C.4.3) son nulos. Luego $\mathcal{B}$ es linealmente independiente. ∎

Ejercicio C.4.7. Sean V un $\Bbbk$-espacio vectorial y $\mathcal{B}_1$ y $\mathcal{B}_2$ bases de dos subespacios vectoriales L_1 y L_2, respectivamente. Probar que $\mathcal{B}_1 \cup \mathcal{B}_2$ es base de $L_1 + L_2$ si, y sólo si, $\mathcal{B}_1 \cap \mathcal{B}_2$ es base de $L_1 \cap L_2$.

5. Suma directa de subespacios vectoriales. Subespacios suplementarios

Un caso especial de suma de subespacios vectoriales L_1 y L_2 de un $\Bbbk$-espacio vectorial V es aquel en que $L_1 \cap L_2 = \{\mathbf{0}\}$, pues, en esta situación, el teorema C.4.6 nos dice que la dimensión de $L_1 + L_2$ es igual a la suma de las dimensiones de L_1 y L_2.

Definición C.5.1. Sean V un $\Bbbk$-espacio vectorial y L_1 y L_2 dos subespacios vectoriales. Se dice que $L_1 + L_2$ están en **suma directa** (ó que la suma $L_1 + L_2$ es directa), y se denota $L_1 \oplus L_2$, cuando $L_1 \cap L_2 = \{\mathbf{0}\}$

La proposición que sigue caracteriza las sumas directas.

Proposición C.5.2. *Sean V un $\Bbbk$-espacio vectorial y L_1 y L_2 dos subespacios vectoriales. La suma $L_1 + L_2$ es directa si, y sólo si, la expresión de un vector de $L_1 + L_2$ como suma de un vector de L_1 y otro de L_2 es única.*

Demostración. $\boxed{\Rightarrow}$ Si tenemos dos expresiones $\mathbf{u}_1 + \mathbf{u}_2 = \mathbf{v}_1 + \mathbf{v}_2$ con $\mathbf{u}_1, \mathbf{v}_1 \in L_1$ y $\mathbf{u}_2, \mathbf{v}_2 \in L_2$, entonces $\mathbf{u}_1 - \mathbf{v}_1 = \mathbf{u}_2 - \mathbf{v}_2 \in L_1 \cap L_2 = \{\mathbf{0}\}$, de donde se sigue que $\mathbf{u}_1 - \mathbf{v}_1 = \mathbf{u}_2 - \mathbf{v}_2 = \{\mathbf{0}\}$ y, por tanto, que $\mathbf{u}_1 = \mathbf{v}_1$ y $\mathbf{u}_2 - \mathbf{v}_2$.

$\boxed{\Leftarrow}$ Si $\mathbf{v} \in L_1 \cap L_2$, resulta que $\mathbf{v}+\mathbf{0} = \mathbf{0}+\mathbf{v}$ son dos expresiones de un mismo vector de $L_1 + L_2$. Las dos expresiones deben coincidir. Por tanto, $\mathbf{v} = \mathbf{0}$. ∎

Nota C.5.3. Es conveniente destacar que la suma directa de subespacios vectoriales, pese a su nombre, no es una operación sino una propiedad de la suma de subespacios vectoriales.

La generalización de la suma directa presenta más dificultades. La forma correcta de hacerlo es usando la proposición C.5.2. Así pues, diremos que la suma $L_1+\ldots+L_m$ es directa y escribiremos $L_1 \oplus \ldots \oplus L_m$ si la expresión de todo vector de $L_1+\ldots+L_m$ como suma de vectores de $L_1, \ldots, L_m$ es única.

Proposición C.5.4. *Sean V un $\Bbbk$-espacio vectorial y $\{L_1, \ldots, L_m\}$ una familia de subespacios vectoriales de V. Los subespacios $L_1, \ldots, L_m$ está en suma directa si, y sólo si, se satisfacen las siguientes $m-1$ igualdades: $(L_1 + \ldots + L_i) \cap L_{i+1} = \{\mathbf{0}\}$, para cada $i = 1, \ldots, m-1$.*

Demostración. $\boxed{\Rightarrow}$ Sea $i \in \{1, \ldots, m-1\}$ fijo. Si $\mathbf{v} \in (L_1 + \ldots + L_i) \cap L_{i+1}$, entonces $\mathbf{v} = \mathbf{v}_1 + \ldots + \mathbf{v}_i = \mathbf{v}_{i+1}$ para ciertos vectores $\mathbf{v}_j \in L_j,\ j = 1, \ldots, i+1$. Luego

$$\mathbf{0} = \mathbf{v}_1 + \ldots + \mathbf{v}_i + (-\mathbf{v}_{i+1}) + \mathbf{0} + \ldots + \mathbf{0} \in L_1 + \ldots L_i + L_{i+1} + L_{i+2} + \ldots + L_m.$$

De donde se sigue, aplicando la hipótesis, que $\mathbf{v}_1 = \ldots = \mathbf{v}_i = \mathbf{v}_{i+1} = \mathbf{0} = \ldots = \mathbf{0}$, en particular $\mathbf{v} = \mathbf{0}$.

$\boxed{\Leftarrow}$ Sean $\mathbf{v}_j \in L_j,\ j = 1, \ldots, m$ tales que $\mathbf{v}_1 + \ldots + \mathbf{v}_m = \mathbf{0}$. Despejando $\mathbf{v}_m$ obtenemos que $\mathbf{v}_m = -(\mathbf{v}_1 + \ldots + \mathbf{v}_{m-1}) \in (L_1 + \ldots + L_{m-1}) \cap L_m = \{\mathbf{0}\}$ y por lo tanto que $\mathbf{v}_m = \mathbf{0}$ y $\mathbf{v}_1 + \ldots + \mathbf{v}_{m-1} = \mathbf{0}$. Despejando ahora $\mathbf{v}_{m-1}$ en está última igualdad obtenemos que $\mathbf{v}_{m-1} = -(\mathbf{v}_1+\ldots+\mathbf{v}_{m-2}) \in (L_1+\ldots+L_{m-2})\cap L_{m-1} = \{\mathbf{0}\}$, luego $\mathbf{v}_{m-1} = \mathbf{0}$ y $\mathbf{v}_1 + \ldots + \mathbf{v}_{m-2} = \mathbf{0}$. Repitiendo este razonamiento las veces que sea necesario se concluye que $\mathbf{v}_1 = \ldots = \mathbf{v}_m = \mathbf{0}$. ∎

Ejercicio C.5.5. Sean V un $\Bbbk$-espacio vectorial y $\{L_1, \ldots, L_m\}$ una familia de subespacios vectoriales de V. Probar que $L_1 \cap L_{i+1} + \ldots + L_i \cap L_{i+1} \subseteq (L_1 + \ldots + L_i) \cap L_{i+1}$, para cada $i = 1, \ldots, n-1$. Concluir que $(L_1 + \ldots + L_i) \cap L_{i+1} = \{\mathbf{0}\}$, para cada $i = 1, \ldots, n-1$, implica $L_i \cap L_j = \{\mathbf{0}\}$, para todo $i \neq j$.

Sin embargo la implicación contraria no es cierta en general. Por ejemplo, si $V = \mathbb{R}^2$ y $L_1 = \langle(1,0)\rangle,\ L_2 = \langle(0,1)\rangle$ y $L_3 = \langle(1,1)\rangle$, entonces $L_1 \cap L_3 = L_2 \cap L_3 = L_2 \cap L_3 = \{\mathbf{0}\}$, mientras que $(L_1 + L_2) \cap L_3 = L_3 \neq \{\mathbf{0}\}$.

Definición C.5.6. Sean V un $\Bbbk$-espacio vectorial y L_1 y L_2 dos subespacios vectoriales de V. Diremos que L_1 y L_2 son **suplementarios** si están en suma directa

y su suma es V. Es decir, según la definición de dos subespacios que están en suma directa, tenemos que L_1 y L_2 son suplementarios si

$$L_1 \cap L_2 = \{\mathbf{0}\} \quad \text{y} \quad L_1 + L_2 = V.$$

Proposición C.5.7. *Sea V un $\Bbbk$-espacio vectorial de dimensión finita. Si L es un subespacio vectorial de V, entonces existe otro subespacio vectorial L' de V tal que $L \oplus L' = V$, es decir, tal que L y L' son suplementarios.*

Demostración. Supongamos $\dim V = n$. Sea $\mathcal{B} = \{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ una base de L. Si Completamos $\mathbf{B}$ a una base de V; $\{\mathbf{u}_1, \ldots, \mathbf{u}_m, \mathbf{u}_{m+1}, \ldots, \mathbf{u}_n\}$, entonces el subespacio $L' = \langle \mathbf{u}_{m+1}, \ldots, \mathbf{u}_n \rangle$ cumple lo deseado (compruébese). ∎

Ejercicio C.5.8. Sean V un $\Bbbk$-espacio vectorial y L_1 y L_2 dos subespacios vectoriales de V. Probar que las siguiente afirmaciones son equivalentes:

(a) L_1 y L_2 son suplementarios.
(b) Para todo $\mathbf{v} \in V$ existe un único $\mathbf{v}_1 \in L_1$ tal que $\mathbf{v} - \mathbf{v}_1 \in L_2$. Al vector $\mathbf{v}_1$ se le llama **proyección de v sobre L_1 paralelamente a L_2**.

Anexo. Subespacios suplementarios en un espacio vectorial de dimensión infinita.

Teorema C.5.9. *Todo subespacio vectorial de un $\Bbbk$-espacio vectorial posee un subespacio suplementario.*

Demostración. Sea L' un subespacio vectorial de un $\Bbbk$-espacio vectorial V y consideramos el conjunto

$$\mathcal{L} = \{L \text{ subespacio vectorial de } V \mid L \cap L' = \{\mathbf{0}\}\};$$

dicho conjunto no es vacío y está ordenado por la inclusión. Si $\{L_i\}_{i\in I}$ es una cadena de $\mathcal{L}$, entonces $\cup_{i\in I} L_i$ es un elemento de $\mathcal{L}$ que es una cota superior para el conjunto $\{L_i\}_{i\in I}$ de $\mathcal{L}$. Por lo tanto, aplicando el Lema de Zorn, obtenemos que en $\mathcal{L}$ hay elementos maximales, es decir, existe un subespacio vectorial L de V que es elemento de $\mathcal{L}$ tal que ningún elemento de $\mathcal{L}$ contiene estrictamente a L. Veamos que L y L' son suplementarios, para lo cual basta probar que $V = L + L'$. Supongamos que no se satisface la igualdad, es decir, que existe un vector no nulo $\mathbf{v} \in V$ tal que $\mathbf{v} \notin L + L'$; entonces el subespacio vectorial $L' + \langle \mathbf{v} \rangle$ de V sería un elemento de $\mathcal{L}$ que contiene estrictamente a L, lo que claramente supone una contradicción. ∎

6. Suma directa de espacios vectoriales

Sean U y V dos espacios vectoriales sobre un cuerpo $\Bbbk$. Llamaremos **suma directa** de U y V al conjunto $U \times V$ con las operaciones

$$\begin{aligned}(\mathbf{u},\mathbf{v})+(\mathbf{u}',\mathbf{v}') &:= (\mathbf{u}+\mathbf{u}',\mathbf{v}+\mathbf{v}');\\ \lambda(\mathbf{u},\mathbf{v}) &:= (\lambda\mathbf{u},\lambda\mathbf{v}),\end{aligned}$$

donde $\mathbf{u},\mathbf{u}' \in U$, $\mathbf{v},\mathbf{v}' \in V$ y $\lambda \in \Bbbk$. Con estas dos operaciones $U \times V$ es un espacio vectorial, que designaremos por $U \times V$.

La suma directa una familia finita de $\Bbbk$-espacios vectoriales se define forma completamente análoga.

Ejemplo C.6.1. Un ejemplo ya conocido de suma directa de espacios vectoriales es el de los espacios vectoriales numéricos, $\Bbbk^n = \Bbbk \times \ldots \times \Bbbk$. En general, la suma de directa de un mismo $\Bbbk$-espacio vectorial V n veces, $V \times \ldots \times V$, se denota por V^n.

Proposición C.6.2. *Si U y V son dos $\Bbbk$-espacios vectoriales de dimensión finita, entonces $U \times V$ es de dimensión finita y* $\dim(U \times V) = \dim U + \dim V$.

Demostración. Sean $\mathcal{B}_U = \{\mathbf{u}_1,\ldots,\mathbf{u}_n\}$ una base de U y $\mathcal{B}_V = \{\mathbf{v}_1,\ldots,\mathbf{v}_m\}$ una base de V. Entonces $\mathcal{B} = \{(\mathbf{u}_1,\mathbf{0}_V),\ldots,(\mathbf{u}_n,\mathbf{0}_V),(\mathbf{0}_U,\mathbf{v}_1),\ldots,(\mathbf{0}_U,\mathbf{v}_m)\}$ es una base de $U \times V$. En efecto: estos vectores generan $U \times V$, ya que si $(\mathbf{u},\mathbf{v}) \in U \times V$ tenemos

$$\begin{aligned}(\mathbf{u},\mathbf{v}) &= (\mathbf{u},\mathbf{0}_V)+(\mathbf{0}_U,\mathbf{v}) = (\textstyle\sum_{i=1}^n \lambda_i\mathbf{u}_i,\mathbf{0}_V)+(\mathbf{0}_U,\sum_{j=1}^m \mu_j\mathbf{v}_j)\\ &= \textstyle\sum_{i=1}^n \lambda_i(\mathbf{u}_i,\mathbf{0}_V)+\sum_{j=1}^m \mu_j(\mathbf{0}_U,\mathbf{v}_j),\end{aligned}$$

y son linealmente independientes, ya que si

$$\sum_{i=1}^n \lambda_i(\mathbf{u}_i,\mathbf{0}_V)+\sum_{j=1}^m \mu_j(\mathbf{0}_U,\mathbf{v}_j) = (\mathbf{0}_U,\mathbf{0}_V)$$

entonces

$$(\sum_{i=1}^n \lambda_i\mathbf{u}_i,\sum_{j=1}^m \mu_j\mathbf{v}_j) = (\mathbf{0}_U,\mathbf{0}_V),$$

lo que implica $\sum_{i=1}^n \lambda_i\mathbf{u}_i = \mathbf{0}_U$ y $\sum_{j=1}^m \mu_j\mathbf{v}_j = \mathbf{0}_V$. De donde se sigue que $\lambda_1 = \ldots = \lambda_n = \mu_1 = \ldots = \mu_m = 0$, por ser $\mathcal{B}_U$ y $\mathcal{B}_V$ bases. ∎

Corolario C.6.3. *Si $\{V_1,\ldots,V_n\}$ es una familia de $\Bbbk$-espacios vectoriales de dimensión finita, entonces $V_1 \times \ldots \times V_n$ es de dimensión finita y* $\dim(V_1 \times \ldots \times V_n) = \dim V_1 + \ldots + \dim V_n$.

En algunos textos se usa el símbolo $\oplus$ en vez de $\times$ para expresar lo que hemos definido como suma directa de espacios vectoriales. Hemos optado por esta notación para evitar confusiones.

Nota C.6.4. En los capítulos 1 y 2 de [**BCR07**] se pueden encontrar diversos ejercicios y ejemplos que con seguridad ayudarán a la mejor compresión de este tema, sobre todo al lector poco familiarizado con los conceptos y resultados.

Bibliografía

[Bas83] A. Basilevsky, *Applied matrix algebra in the statistical sciences*, North-Holland, New York, 1983.

[BCR07] V.J. Bolós, J. Cayetano, and B. Requejo, *Álgebra lineal y geometría*, Manuales de Unex, vol. 50, Universidad de Extremadura, 2007.

[Ber77] S.K. Berberian, *Introducción al espacio de hilbert*, Editorial Teide, 1977.

[BS98] R. Barbolla and P. Sanz, *Álgebra lineal y teoría de matrices*, Prentice Hall, Madrid, 1998.

[Cia82] P.G. Ciarlet, *Introduction à l'analyse numérique matriciele et à l'optimisation*, Masson, Paris, 1982.

[CnR05] J. Arvesú Carballo, F. Marcellán Espa nol, and J. Sánchez Ruiz, *Problemas resueltos de álgebra lineal*, Thomson Editores Spain, Madrid, 2005.

[DP99] L. Debnath and P.Mikusiński, *Introduction to hilbert spaces with applications*, Academic Press, Inc., San Diego, CA, 1999.

[dR87] D. Peña Sánchez de Rivera, *Estadística. modelos y métodos*, Alianza Universidad Textos, vol. 110, Alianza Editorial, Madrid, 1987.

[FVV03] C. Fernández-Pérez, F.J. Vázquez-Hernández, and J.M. Vegas Montaner, *Ecuaciones diferencias y en diferencias*, Thomson Editores Spain, Madrid, 2003.

[Her85] D. Hernández, *Álgebra lineal*, Manuales de la Universidad de Salamanca, Universidad de Salamanca, 1985.

[IR99] J.A. Infante del Río and J.M. Rey Cabezas, *Metodos numericos: teoria, problemas y practicas con matlab*, Ed. Pirámide, S.A., Madrid, 1999.

[Lip70] S. Lipschutz, *Topología general*, Serie de Compendios Schaum, McGraw-Hill, México, 1970.

[Mey00] C. Meyer, *Matrix analysis and applied linear algebra*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2000.

[MN07] J.R. Magnus and H. Neudecker, *Matrix Differential Calculus with applications in statistics and econometrics*, second (revised) ed., Wiley Series in Probability and Statistics, John Wiley & Sons, Chichester, 2007.

[MS06] L. Merino and E. Santos, *Álgebra lineal con métodos elementales*, Thomson Editores Spain, Madrid, 2006.

[Nav96] J.A. Navarro González, *Álgebra conmutativa básica*, Manuales de Unex, vol. 19, Universidad de Extremadura, 1996.

[QS06] A. Quarteroni, , and F. Saleri, *Cálculo científico con MATLAB y Octave*, Springer-Verlag, Milano, 2006.

[QSS07] A. Quarteroni, R. Sacco, and F. Saleri, *Numerical mathematics*, second ed., Texts in Applied Mathematics, vol. 37, Springer-Verlag, Berlin, 2007.

[RM71] C.R. Rao and S.K. Mitra, *Generalized inverse of matrices and its applications*, John Wiley & Sons, New York-London-Sydney, 1971.

[Sch05] J.R. Schott, *Matrix analysis for statistics*, second ed., Wiley Series in Probability and Statistics, John Wiley & Sons, Hoboken, NJ, 2005.

[Sea82] S.R. Searle, *Matrix algebra useful for statistics*, Wiley Series in Probability and Statistics, John Wiley & Sons, Chichester, 1982.

[Sen81] E. Seneta, *Non-negative matrices and markov chains*, Springer Series in Statistics, Springer Verlag, Berlin, 1981.

[Spi88] M. Spivak, *Cálculo en variedades*, Editorial reverté, Barcelona, 1988.

[SV95] M.J. Soto and J.L. Vicente, *Algebra lineal con matlab y maple*, Prentice Hall International, Hertfordshire, Reino Unido, 1995.

Índice alfabético

ISBN 978-84-691-6429-7

UNIVERSIDAD DE EXTREMADURA

UNIÓN EUROPEA
Fondo Social Europeo

JUNTA DE EXTREMADURA

58