E. T. S. de Ingenieros Industriales
Universidad Politécnica de Madrid

Grado en Ingeniería en Tecnologías Industriales

Grado en Ingeniería Química

Apuntes de Álgebra

(Curso 2014/15)

Departamento de Matemática Aplicada a la Ingeniería Industrial

Índice general

Capítulo 1

Los espacios vectoriales $\mathbb{R}^n$ y $\mathbb{C}^n$

1.1 Definición. Combinaciones lineales. Clausura lineal. Dependencia e independencia lineal. Subespacios vectoriales.

1.2 Bases. Dimensión. Intersección y suma de subespacios. Suma directa. Subespacios suplementarios. La relación de Grassmann.

1.3 Ejercicios. Cuestiones.

1.1. Definición, propiedades y ejemplos

En el presente capítulo nos proponemos estudiar los espacios vectoriales $\mathbb{R}^n$ y $\mathbb{C}^n$, que son generalizaciones del plano y del espacio euclídeo ordinarios así como de las magnitudes vectoriales de la física. Se trata de conjuntos cuyos elementos, denominados *vectores*, pueden sumarse entre sí y asimismo multiplicarse por números, que denominaremos *escalares*, obteniéndose como resultado en ambos casos un nuevo vector.

Nota: En lo que sigue, la letra $\mathbb{K}$ denotará indistintamente el cuerpo $\mathbb{R}$ de los números reales o el cuerpo $\mathbb{C}$ de los números complejos. Utilizamos el mismo símbolo "+" para denotar la suma en el espacio $\mathbb{K}^n$ (*suma de vectores*) y la suma en el cuerpo $\mathbb{K}$. El símbolo "·" denota la ley externa en $\mathbb{K}^n$ (*producto de escalar por vector*)[1].

Definición 1 *Si n es cualquier entero positivo, definimos el espacio $\mathbb{K}^n$ como el conjunto de las n-uplas ordenadas de elementos de $\mathbb{K}$, es decir:*

$$\mathbb{K}^n := \{(x_1, x_2, \ldots, x_n) : x_j \in \mathbb{K}, j = 1, 2, \ldots, n\}. \tag{1.1}$$

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

[1] En la práctica, tanto para producto de escalares como de escalar por vector, se yuxtaponen los términos omitiendo el punto "·".

Definición 2 (suma de vectores) *En $\mathbb{K}^n$ definimos la operación suma mediante*

$$(x_1, x_2, \ldots, x_n) + (y_1, y_2, \ldots, y_n) := (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n). \tag{1.2}$$

Propiedades de la suma:

La suma de vectores de $\mathbb{K}^n$ es una ley de composición interna en $\mathbb{K}^n$ (es decir, a cada par de vectores de $\mathbb{K}^n$ le asocia un nuevo vector de $\mathbb{K}^n$) que verifica las siguientes propiedades:

1. $\forall \mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{K}^n$, $(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$ (propiedad asociativa).

2. $\exists \mathbf{0} \in \mathbb{K}^n$ tal que, $\forall \mathbf{x} \in \mathbb{K}^n$, $\mathbf{x} + \mathbf{0} = \mathbf{0} + \mathbf{x} = \mathbf{x}$ (existencia de elemento neutro).

3. $\forall \mathbf{x} \in \mathbb{K}^n$, $\exists -\mathbf{x} \in \mathbb{K}^n$ tal que $\mathbf{x} + (-\mathbf{x}) = -\mathbf{x} + \mathbf{x} = \mathbf{0}$ (existencia de elemento simétrico).

4. $\forall \mathbf{x}, \mathbf{y} \in \mathbb{K}^n$, $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$ (propiedad conmutativa).

Las cuatro propiedades enunciadas, que se deducen inmediatamente de las de la suma de números reales o complejos, dotan a la dupla $(\mathbb{K}^n, +)$ de la estructura de **grupo abeliano**.

Definición 3 (vector nulo) *Llamamos vector nulo de $\mathbb{K}^n$, y lo denotamos por $\mathbf{0}$, al vector $(0, 0, \ldots, 0)$, es decir, el elemento neutro de la suma de vectores.*

Definición 4 (vector opuesto) *Dado un vector $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in \mathbb{K}^n$, su opuesto es el vector $-\mathbf{x} := (-x_1, -x_2, \ldots, -x_n)$, es decir, su elemento simétrico respecto a la suma de vectores.*

Definición 5 (producto de escalar por vector) *Si $\alpha \in \mathbb{K}$ y $(x_1, x_2, \ldots, x_n) \in \mathbb{K}^n$, definimos su producto mediante*

$$\alpha \cdot (x_1, x_2, \ldots, x_n) = (\alpha x_1, \alpha x_2, \ldots, \alpha x_n). \tag{1.3}$$

*Esta operación también se denomina **ley externa** de $\mathbb{K}^n$.*

Propiedades de la ley externa:

La ley externa, en conjunción con la suma de vectores, verifica las siguiente propiedades:

5. $\forall \mathbf{x} \in \mathbb{K}^n,\ \forall \lambda, \mu \in \mathbb{K},\ (\lambda + \mu) \cdot \mathbf{x} = \lambda \cdot \mathbf{x} + \mu \cdot \mathbf{x}$.

6. $\forall \mathbf{x}, \mathbf{y} \in \mathbb{K}^n,\ \forall \lambda \in \mathbb{K},\ \lambda \cdot (\mathbf{x} + \mathbf{y}) = \lambda \cdot \mathbf{x} + \lambda \cdot \mathbf{y}$.

7. $\forall \mathbf{x} \in \mathbb{K}^n,\ \forall \lambda, \mu \in \mathbb{K},\ (\lambda\mu) \cdot \mathbf{x} = \lambda \cdot (\mu \cdot \mathbf{x})$.

8. $\forall \mathbf{x} \in \mathbb{K}^n,\ 1 \cdot \mathbf{x} = \mathbf{x}$.

Todas estas propiedades se deducen fácilmente de las de la suma y el producto de números reales o complejos. Las propiedades 1 a 8 dotan a la cuaterna $(\mathbb{K}^n, \mathbb{K}, +, \cdot)$ de la estructura de **espacio vectorial**. También se dice que $\mathbb{K}^n$ es un espacio vectorial sobre $\mathbb{K}$.

Propiedades adicionales:
De las 8 propiedades enunciadas —que constituyen los axiomas de espacio vectorial—, o bien de las propiedades de suma y producto de números reales y complejos, se deducen con facilidad las siguientes (nótese la distinción entre el escalar $0 \in \mathbb{K}$ y el vector $\mathbf{0} \in \mathbb{K}^n$):

- $\forall \mathbf{x} \in \mathbb{K}^n,\ 0 \cdot \mathbf{x} = \mathbf{0}$.
- Sean $\lambda \in \mathbb{K}$ y $\mathbf{x} \in \mathbb{K}^n$; $\lambda \cdot \mathbf{x} = \mathbf{0} \Leftrightarrow \lambda = 0 \vee \mathbf{x} = \mathbf{0}$.
- $\forall \lambda \in \mathbb{K},\ \forall \mathbf{x} \in \mathbb{K}^n,\ (-\lambda) \cdot \mathbf{x} = \lambda \cdot (-\mathbf{x}) = -(\lambda \mathbf{x})$.

1.1.1. Combinaciones lineales

Definición 6 (combinación lineal) *Una combinación lineal de los vectores* $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k$ *de* $\mathbb{K}^n$ *es cualquier vector de la forma*

$$\mathbf{u} = \lambda_1 \mathbf{u}_1 + \lambda_2 \mathbf{u}_2 + \ldots + \lambda_k \mathbf{u}_k, \tag{1.4}$$

siendo $\lambda_1, \lambda_2, \ldots, \lambda_k \in \mathbb{K}$.

Ejemplo: En $\mathbb{R}^3$, el vector $\left(\frac{1}{3}, 3, -\frac{4}{3}\right)$ es combinación lineal de los vectores $(1, 0, -1)$ y $(1, 3, -2)$, ya que

$$\left(\frac{1}{3}, 3, -\frac{4}{3}\right) = -\frac{2}{3}(1, 0, -1) + (1, 3, -2).$$

Observación: El vector nulo es combinación lineal de cualesquiera vectores. Todo vector es combinación lineal de sí mismo.

Definición 7 (clausura lineal) *Dado un conjunto* $M \subset \mathbb{K}^n$*, se define la clausura lineal de* M*, y se denota por* $L[M]$*, como el conjunto de todas las combinaciones lineales de vectores de* M*.*

Nota: Para $M = \varnothing$ se adopta el convenio $L[\varnothing] = \{\mathbf{0}\}$.

Ejemplo: En $\mathbb{R}^3$, sea $M = \{(1, 0, -1), (1, 3, -2)\}$. $L[M]$ es el conjunto de vectores de la forma $\alpha(1, 0, -1) + \beta(1, 3, -2)$ con α, β escalares reales, es decir:

$$L[M] = \{(\alpha + \beta, 3\beta, -\alpha - 2\beta) : \alpha, \beta \in \mathbb{R}\}.$$

El vector $\left(\frac{1}{3}, 3, -\frac{4}{3}\right)$ del ejemplo anterior pertenece a $L[M]$, pues se obtiene para $\alpha = -2/3$ y $\beta = 1$.

1.1.2. Dependencia e independencia lineal

Definición 8 (independencia lineal) *Los vectores $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k$ de $\mathbb{K}^n$ son linealmente independientes si la única combinación lineal de ellos igual al vector nulo es la que tiene todos los coeficientes nulos, es decir,*

$$\lambda_j \in \mathbb{K}, \quad \lambda_1\mathbf{u}_1 + \lambda_2\mathbf{u}_2 + \ldots + \lambda_k\mathbf{u}_k = \mathbf{0} \Rightarrow \lambda_1 = \lambda_2 = \ldots = \lambda_k = 0. \tag{1.5}$$

En caso contrario, se dice que los vectores anteriores son ***linealmente dependientes****.*

Observación: $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k$ son **linealmente dependientes** si y sólo si

$$\exists \lambda_1, \lambda_2, \ldots, \lambda_k \in \mathbb{K}, \text{ con algún } \lambda_i \neq 0, \text{ tales que } \sum_{j=1}^{k} \lambda_j \mathbf{u}_j = \mathbf{0}. \tag{1.6}$$

Lo anterior equivale a que se pueda despejar alguno de los vectores como combinación lineal de los restantes, es decir,

$$\mathbf{u}_i = \sum_{j \neq i} \alpha_j \mathbf{u}_j, \quad \alpha_j \in \mathbb{K}. \tag{1.7}$$

(Basta considerar un $\mathbf{u}_i$ cuyo $\lambda_i \neq 0$ en la combinación lineal que da el vector nulo).

Definición 9 (familia libre y ligada) *Sea $S = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\} \subset \mathbb{K}^n$; se dice que S es una familia libre si $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k$ son linealmente independientes; en caso contrario, se dice que S es una familia ligada.*

Ejemplos:

- En $\mathbb{R}^3$, $\{(1,1,1),(1,1,0),(1,0,0)\}$ es una familia libre, ya que

  $$\alpha(1,1,1)+\beta(1,1,0)+\gamma(1,0,0) = (0,0,0) \Leftrightarrow \alpha+\beta+\gamma = \alpha+\beta = \alpha = 0 \Leftrightarrow \alpha = \beta = \gamma = 0.$$

  Los tres vectores son linealmente independientes.

- En el mismo espacio, $\left\{(1,0,-1),(1,3,-2),\left(\frac{1}{3},3,-\frac{4}{3}\right)\right\}$ es ligada, ya que, como se vio anteriormente, $\left(\frac{1}{3},3,-\frac{4}{3}\right) = -\frac{2}{3}(1,0,-1) + (1,3,-2)$, lo que equivale a que

  $$-\frac{2}{3}(1,0,-1) + (1,3,-2) - \left(\frac{1}{3},3,-\frac{4}{3}\right) = (0,0,0).$$

  Los tres vectores son linealmente dependientes.

Nota: Por razones que se verán más adelante en este mismo tema, adoptamos el convenio de considerar el conjunto vacío, $\varnothing$, como una familia libre.

Observación: Recapitulando:

- Cualquier subconjunto de una familia libre es libre.
- Cualquier familia que contenga el vector nulo es una familia ligada.
- Cualquier familia que contenga dos vectores proporcionales es ligada.
- La unión de una familia ligada con cualquier otra siempre da una familia ligada.
- Una familia S es ligada si y sólo si algún vector de S es combinación lineal de los restantes vectores de S.
- S es libre si y sólo si ningún vector de S es combinación lineal de los restantes vectores de S.

Ejemplo: En $\mathbb{K}^n$, consideramos los n vectores $\mathbf{e}_1 = (1, 0, \ldots, 0), \mathbf{e}_2 = (0, 1, 0, \ldots, 0), \ldots, \mathbf{e}_n = (0, \ldots, 0, 1)$ (para cada $j = 1, 2, \ldots, n$, $\mathbf{e}_j$ es el vector que tiene nulas todas sus componentes excepto la j-ésima que vale 1). Estos vectores se conocen como **vectores canónicos** de $\mathbb{K}^n$.

La familia formada por los vectores canónicos es libre. En efecto, si $\lambda_j \in \mathbb{K}$,

$$\lambda_1 \mathbf{e}_1 + \lambda_2 \mathbf{e}_2 + \ldots + \lambda_n \mathbf{e}_n = \mathbf{0} \Leftrightarrow (\lambda_1, \lambda_2, \ldots, \lambda_n) = (0, 0, \ldots, 0) \Leftrightarrow \lambda_j = 0 \ \forall j = 1, 2, \ldots, n.$$

1.1.3. Subespacios vectoriales

El concepto de subespacio vectorial se refiere a todo subconjunto H de $\mathbb{K}^n$ que, por sí mismo, mantenga la *estructura* de espacio vectorial de aquél, como por ejemplo sucede en $\mathbb{R}^3$ con los planos coordenados. Esta idea se traduce en que, si se efectúan las operaciones propias de $\mathbb{K}^n$ con los vectores de H, se tienen que obtener en todos los casos vectores pertenecientes a H.

Definición 10 (subespacio vectorial) *Un subespacio vectorial de $\mathbb{K}^n$ es cualquier subconjunto no vacío de dicho espacio que sea cerrado para la suma y para el producto por escalares, es decir:*

$H \subset \mathbb{K}^n$; se dice que H es un subespacio vectorial de $\mathbb{K}^n$ si:

1. $\mathbf{0} \in H$.
2. $\forall \mathbf{u}, \mathbf{v} \in H, \ \mathbf{u} + \mathbf{v} \in H$.

3. $\forall \lambda \in \mathbb{K}, \forall \mathbf{u} \in H,\ \lambda \mathbf{u} \in H$.

Nótese que los apartados 2. y 3. pueden escribirse conjuntamente en la forma:

$$\forall \mathbf{u}, \mathbf{v} \in H,\ \forall \lambda, \mu \in \mathbb{K},\ \lambda \mathbf{u} + \mu \mathbf{v} \in H. \tag{1.8}$$

Así pues, un subespacio vectorial es un conjunto no vacío **cerrado para las combinaciones lineales**.

Observación: El propio espacio $\mathbb{K}^n$ es subespacio vectorial de sí mismo; además, el conjunto $\{\mathbf{0}\}$ formado únicamente por el vector nulo, constituye un subespacio que llamaremos **subespacio nulo**. Estos dos subespacios son los llamados *subespacios triviales* de $\mathbb{K}^n$.

Ejemplos:

1. En $\mathbb{R}^4$ consideramos el subconjunto $H = \{\mathbf{x} \in \mathbb{R}^4 : x_1 = 2x_3\}$.

   H, que obviamente contiene el $(0,0,0,0)$, es un subespacio vectorial; en efecto, sean $\mathbf{x}, \mathbf{y} \in H$, es decir, que satisfacen que $x_1 = 2x_3$ e $y_1 = 2y_3$. Para cualesquiera $\lambda, \mu \in \mathbb{R}$, consideramos el vector $\lambda\mathbf{x}+\mu\mathbf{y}$, cuya primera componente vale $\lambda x_1 + \mu y_1 = \lambda \cdot 2x_3 + \mu \cdot 2y_3 = 2(\lambda x_3 + \mu y_3)$; como $\lambda x_3 + \mu y_3$ es precisamente el valor de la tercera componente de $\lambda\mathbf{x} + \mu\mathbf{y}$, concluimos que dicho vector pertenece a H y, por tanto, éste es subespacio.

2. En $\mathbb{R}^3$, el plano de ecuación $x + y - 2z = 5$ no es subespacio vectorial, ya que el vector nulo (origen de coordenadas) no pertenece a dicho plano.

3. En $\mathbb{R}^3$, el paraboloide P de ecuación $z = x^2 + y^2$, pese a contener al $(0,0,0)$, no es subespacio vectorial. En efecto, si tomamos los puntos $(1,1,2)$ y $(-1,-1,2)$, ambos pertenecen a P, pero su suma $(0,0,4)$ no satisface la ecuación.

Observación: La clausura lineal de cualquier conjunto M es un subespacio vectorial, ya que las combinaciones lineales de vectores de $L[M]$ pueden reducirse a su vez a combinaciones lineales de vectores de M, y pertenecen por tanto a $L[M]$, el cual se denomina también subespacio generado por M. Cualquier subespacio que contenga a M necesariamente contiene a $L[M]$.

Observación: Como consecuencia de la observación anterior, en $\mathbb{R}^2$ los únicos subespacios no triviales son las rectas que pasan por el origen, y en $\mathbb{R}^3$ las rectas que pasan por el origen y los planos que contienen a dicho punto.

1.2. Bases

1.2.1. Base y dimensión

Definición 11 (sistema generador) *Dados H subespacio vectorial de $\mathbb{K}^n$ y $\mathcal{M} \subset H$, se dice que $\mathcal{M}$ es un sistema generador (o sistema de generadores) de H si $H = L[\mathcal{M}]$, es decir, si todo vector de H puede ponerse como combinación lineal de los de $\mathcal{M}$.*

Ejemplo: Los vectores canónicos de $\mathbb{K}^n$ constituyen un sistema generador de dicho espacio, ya que $\forall \mathbf{x} = (x_1, x_2, \ldots, x_n) \in \mathbb{K}^n$,

$$\mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \ldots + x_n\mathbf{e}_n.$$

Definición 12 (base) *Dado H subespacio vectorial de $\mathbb{K}^n$, una base de H es un* ***sistema generador, libre y ordenado****.*

Definición 13 (base canónica) *En $\mathbb{K}^n$, la familia formada por los vectores canónicos $(\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n)$, que ya se ha visto que constituye un sistema libre y generador de $\mathbb{K}^n$, es por tanto una base que denominaremos base canónica de $\mathbb{K}^n$.*

Ejemplo: Encontrar una base del subespacio de $\mathbb{R}^4$ determinado por la ecuación $x_1 - x_2 + 3x_3 + x_4 = 0$.

Solución: puesto que tenemos una única ecuación y cuatro incógnitas, podemos dar valores arbitrarios a tres de ellas y escribir la cuarta en función de esas tres, por ejemplo: $x_2 = \alpha, x_3 = \beta, x_4 = \gamma, x_1 = \alpha - 3\beta - \gamma$, es decir:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} \alpha - 3\beta - \gamma \\ \alpha \\ \beta \\ \gamma \end{pmatrix} = \alpha \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} + \beta \begin{pmatrix} -3 \\ 0 \\ 1 \\ 0 \end{pmatrix} + \gamma \begin{pmatrix} -1 \\ 0 \\ 0 \\ 1 \end{pmatrix}.$$

Los tres vectores anteriores engendran de manera evidente el subespacio y son linealmente independientes (véanse las tres últimas componentes), constituyendo por lo tanto una base del mismo.

Teorema 1 (existencia de bases) *Todo subespacio vectorial admite una base.*

Teorema 2 (equicardinalidad de bases) *Todas las bases de un mismo subespacio vectorial tienen el mismo número de vectores.*

Definición 14 (dimensión) *Llamamos dimensión de un subespacio vectorial H, y la denotamos por $\dim H$, al número de elementos de cualquiera de sus bases.*

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

Nota: Adoptamos el convenio de considerar al conjunto vacío, $\varnothing$, como la base del subespacio nulo $\{\mathbf{0}\}$. Así pues, $\dim(\{\mathbf{0}\}) = 0$. Con esta convención, la relación de Grassmann que se verá más adelante así como otros resultados de este curso siguen siendo válidos cuando alguno de los subespacios involucrados es el nulo.

Recuérdese en todo caso que $\{\mathbf{0}\}$ así como cualquier familia que lo contenga siempre es una familia ligada.

Proposición 1 *De todo sistema generador de un subespacio vectorial H puede extraerse una base.*

Apunte constructivo: Nos restringimos al caso en que el sistema generador es finito. Sea $\mathcal{G}$ un generador finito de H. Si es libre el resultado está probado. Si es ligado, existe un vector $\mathbf{v} \in \mathcal{G}$ que depende linealmente de los restantes vectores de $\mathcal{G}$; por tanto, al suprimir ese vector, el conjunto resultante $\mathcal{G} \setminus \{\mathbf{v}\}$ sigue engendrando el mismo subespacio H. Se repite el argumento y, dado que $\mathcal{G}$ es finito, en un número finito de pasos habremos extraído de $\mathcal{G}$ un generador libre.

Observación: La construcción anterior muestra que una base es un sistema generador "minimal", es decir, con el menor número posible de vectores. Así pues, **la dimensión de un subespacio equivale al mínimo número de vectores para un sistema generador** de dicho subespacio.

Igualmente, puede demostrarse que en un subespacio de dimensión m no puede existir un sistema libre con un número de vectores superior a m. Así, una base se caracteriza por ser un sistema libre "maximal", es decir, con el máximo número posible de vectores. **La dimensión es el mayor número posible de vectores linealmente independientes**.

Algunas ideas afines a lo anterior se recapitulan en la siguiente proposición:

Proposición 2 *Sea H un subespacio de $\mathbb{K}^n$ cuya dimensión es m:*

- *Si $\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ es una familia libre de H, entonces es generadora (y por tanto base) de H.*

- *Si $\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ es un sistema generador de H, entonces es libre (y por tanto base de H).*

- *Si $m = n$ entonces $H = \mathbb{K}^n$.*

Observación: Si un subespacio H tiene dimensión k, el vector genérico de dicho espacio puede escribirse como

$$\mathbf{x} = \alpha_1 \mathbf{u}_1 + \ldots + \alpha_k \mathbf{u}_k, \ \alpha_j \in \mathbb{K}, \tag{1.9}$$

en donde $(\mathbf{u}_1, \ldots, \mathbf{u}_k)$ es una base de H. La expresión anterior coincide con la de la solución general de un sistema de $n-k$ ecuaciones lineales homogéneas, que denominaremos **ecuaciones implícitas** de H, que caracterizan a los vectores de dicho subespacio. Para encontrar dichas ecuaciones basta con eliminar los parámetros de la ecuación vectorial anterior o, equivalentemente, forzar a que el vector genérico $\mathbf{x} = (x_1, \ldots, x_n)$ sea combinación lineal de los de la base de H.

Para un subespacio H de $\mathbb{K}^n$, el número de ecuaciones implícitas **independientes** que lo determinan es $n - \dim H$. Esto quedará demostrado en la proposición 14 del tema 2.

Ejemplo: En $\mathbb{R}^3$, encontrar la ecuación implícita que determina el subespacio $H = L[(1,0,1),(1,3,2)]$.

Solución: El vector genérico (x,y,z) de H se expresa mediante las ecuaciones paramétricas

$$\begin{aligned} x &= \alpha + \beta \\ y &= 3\beta \\ z &= \alpha + 2\beta \end{aligned}$$

siendo α, β parámetros reales. Procedemos a eliminar dichos parámetros (formalmente, esto equivale a forzar, en términos de x, y, z, a que el sistema lineal cuyas incógnitas son α, β sea compatible): $\beta = z - x \Rightarrow y = 3(z-x)$, de donde se concluye que la ecuación implícita de H es

$$3x + y - 3z = 0.$$

Otra forma: La pertenencia a H del vector (x,y,z) equivale a que la última columna de la matriz

$$\begin{pmatrix} 1 & 1 & x \\ 0 & 3 & y \\ 1 & 2 & z \end{pmatrix}$$

sea combinación de las dos primeras; si conocemos el concepto de rango, sabemos que lo anterior sucede si y solo si dicha matriz tiene rango 2 (ya que $(1,0,1)$ y $(1,3,2)$ son linealmente independientes). Triangulamos la matriz mediante operaciones de reducción gaussiana e imponemos que el rango sea 2:

$$\begin{pmatrix} 1 & 1 & x \\ 0 & 3 & y \\ 1 & 2 & z \end{pmatrix} \sim \begin{pmatrix} 1 & 1 & x \\ 0 & 3 & y \\ 0 & 1 & z-x \end{pmatrix} \sim \begin{pmatrix} 1 & 1 & x \\ 0 & 3 & y \\ 0 & 0 & 3(z-x)-y \end{pmatrix}$$

cuyo rango será 2 si y sólo si $3x + y - 3z = 0$.

Obsérvese que, puesto que la matriz es cuadrada, también podríamos haber procedido forzando a que su determinante se anulase.

Teorema 3 (compleción de la base) *Sea H un subespacio vectorial de dimensión m, y sea $\mathcal{G} = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ una familia libre de H, con $k < m$. Entonces existen $\mathbf{u}_{k+1}, \ldots, \mathbf{u}_m \in H$, tales que $(\mathbf{u}_1, \ldots, \mathbf{u}_k, \mathbf{u}_{k+1}, \ldots, \mathbf{u}_m)$ es una base de H.*

Demostración constructiva:

Puesto que todo sistema generador de H tiene al menos m elementos, el conjunto $\mathcal{G}$ no puede ser generador, luego existe algún vector $\mathbf{u}_{k+1}$ que no es combinación lineal de los de $\mathcal{G}$, es decir, $u_{k+1} \in H \setminus L[\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k]$.

El conjunto obtenido $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k, \mathbf{u}_{k+1}\}$ es un sistema libre de H, ya que $\mathbf{u}_{k+1}$ no depende linealmente de los k primeros vectores que, a su vez, son linealmente independientes entre sí.

Si $k+1 = m$, ya tenemos una base de H, pues el conjunto $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k, \mathbf{u}_{k+1}\}$ es una familia libre de m vectores; en caso contrario, volvemos a aplicar el mismo razonamiento a dicho conjunto (sea $\mathbf{u}_{k+2} \in H \setminus L[\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k, \mathbf{u}_{k+1}]$, etc.); en $m-k$ pasos, habremos obtenido una familia libre de m vectores, es decir, una base de H. $\square$

Ejemplo: Encontrar una base de $\mathbb{R}^4$ cuyos dos primeros vectores sean $(1, 2, 3, 4)$, $(1, 0, -1, 0)$.

Solución: En lugar de proceder paso a paso como en la construcción teórica anterior, vamos a intentar encontrar los dos vectores directamente. Probamos con aquéllos que nos pueden facilitar los cálculos, es decir, los canónicos, por ejemplo $(0, 0, 1, 0), (0, 0, 0, 1)$ y simplemente comprobamos si estos vectores completan la base. En efecto, la familia obtenida es libre ya que, con una simple operación de reducción gaussiana por filas,

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \sim \begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & -2 & -4 & -4 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

de donde concluimos que las cuatro filas de la matriz son linealmente independientes. Así pues, una base que satisface las condiciones pedidas es

$$((1, 2, 3, 4), (1, 0, -1, 0), (0, 0, 1, 0), (0, 0, 0, 1)).$$

Si los dos vectores escogidos nos hubiesen proporcionado una matriz de rango inferior a 4, escogeríamos otros distintos y repetiríamos la comprobación.

1.2.2. Coordenadas respecto a una base

Definición 15 (coordenadas respecto a una base) *Sea $\mathcal{B} = (\mathbf{u}_1, \ldots, \mathbf{u}_n)$ una base de $\mathbb{K}^n$. Si $\mathbf{v} \in \mathbb{K}^n$, $\mathbf{v} = x_1\mathbf{u}_1 + \ldots + x_n\mathbf{u}_n$ con $x_j \in \mathbb{K}$, los escalares $x_1, \ldots, x_n$ se denominan coordenadas del vector $\mathbf{v}$ respecto a la base $\mathcal{B}$.*

Nota: De igual modo pueden definirse las coordenadas respecto a una base de un subespacio de $\mathbb{K}^n$ de cualquier vector perteneciente a dicho subespacio.

Proposición 3 *Las coordenadas de un vector respecto a una base son únicas.*

Demostración: Supongamos que un vector $\mathbf{u}$ puede expresarse como $\mathbf{u} = x_1\mathbf{u}_1 + \ldots + x_n\mathbf{u}_n = x_1'\mathbf{u}_1 + \ldots + x_n'\mathbf{u}_n$. Entonces,

$$\mathbf{0} = x_1\mathbf{u}_1 + \ldots + x_n\mathbf{u}_n - (x_1'\mathbf{u}_1 + \ldots + x_n'\mathbf{u}_n) = (x_1 - x_1')\mathbf{u}_1 + \ldots + (x_n - x_n')\mathbf{u}_n, \quad (1.10)$$

lo que implica, por la independencia lineal de los $\mathbf{u}_j$, que todos los escalares $x_j - x_j'$ valen 0, es decir, que $x_j' = x_j \ \forall j = 1, 2, \ldots, n$. □

Definición 16 *El vector* $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{K}^n$ *se denomina* ***vector de coordenadas*** *de* $\mathbf{v}$ *respecto a* $\mathcal{B}$*, y se suele denotar por* $[\mathbf{v}]^{\mathcal{B}}$.

Observación: Las coordenadas de un vector respecto a la base canónica de $\mathbb{K}^n$ coinciden con las componentes del vector, es decir, el vector de coordenadas coincide con el propio vector.

Ejemplo: Calcular las coordenadas del vector $\mathbf{u} = (1, 2, 3)$ respecto a la base $\mathcal{B} = ((1, 0, 1), (1, 1, -1), (1, 2, 0))$ de $\mathbb{R}^3$.

Solución: Buscamos x, y, z tales que $x(1, 0, 1) + y(1, 1, -1) + z(1, 2, 0) = (1, 2, 3)$, es decir, tenemos que resolver el sistema de tres ecuaciones lineales cuya matriz ampliada es

$$\left(\begin{array}{ccc|c} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 2 \\ 1 & -1 & 0 & 3 \end{array}\right).$$

Realizamos operaciones de reducción gaussiana por filas (es decir, ecuaciones), obteniendo:

$$\left(\begin{array}{ccc|c} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 2 \\ 0 & -2 & -1 & 2 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 2 \\ 0 & 0 & 3 & 6 \end{array}\right),$$

que, despejando las incógnitas de abajo a arriba, nos da la solución $z = 2, y = -2, x = 1$. El vector de coordenadas es, por tanto,

$$[\mathbf{u}]^{\mathcal{B}} = (1, -2, 2).$$

1.2.3. Intersección y suma de subespacios

Recuérdese que la intersección de dos conjuntos es el conjunto formado por aquellos elementos que pertenecen simultáneamente a ambos, es decir:

$$A, B \text{ conjuntos}; x \in A \cap B \Leftrightarrow x \in A \wedge x \in B. \tag{1.11}$$

Proposición 4 *La intersección de dos subespacios vectoriales de $\mathbb{K}^n$ es subespacio vectorial de dicho espacio.*

Demostración: Sean los subespacios M y N de $\mathbb{K}^n$. $M \cap N$ contiene obviamente al vector nulo. Sean los vectores $\mathbf{u}, \mathbf{v} \in M \cap N$; así pues, $\mathbf{u}, \mathbf{v} \in M$ y $\mathbf{u}, \mathbf{v} \in N$; por ser estos conjuntos subespacios, para cualesquiera $\lambda, \mu \in \mathbb{K}$, $\lambda\mathbf{u} + \mu\mathbf{v} \in M$ y $\lambda\mathbf{u} + \mu\mathbf{v} \in N$, es decir, $\lambda\mathbf{u} + \mu\mathbf{v} \in M \cap N$. □

Observación: La **unión** de subespacios vectoriales **no** es, en general, subespacio vectorial. Considérense, por ejemplo, los dos ejes coordenados de $\mathbb{R}^2$, que son subespacios. Tanto el vector $\mathbf{e}_1$ como el $\mathbf{e}_2$ pertenecen a la unión de dichos subespacios, pero su vector suma $(1, 1)$ no pertenece a dicha unión.

De hecho, dados dos subespacios L, M de un mismo espacio vectorial E, se tiene que

$$L \cup M \text{ es subespacio } \Leftrightarrow (L \subset M \vee M \subset L). \tag{1.12}$$

Definición 17 (suma de subespacios) *Dados dos subespacios L y M de $\mathbb{K}^n$, definimos su suma como*

$$L + M := \{\mathbf{u} + \mathbf{v} : \mathbf{u} \in L \ , \ \mathbf{v} \in M\}. \tag{1.13}$$

En general, la definición anterior puede extenderse recursivamente a un número finito de subespacios de $\mathbb{K}^n$.

Ejemplo: En $\mathbb{R}^3$, se consideran los dos ejes coordenados $OX = \{(x, 0, 0) : x \in \mathbb{R}\}$ y $OY = \{(0, y, 0) : y \in \mathbb{R}\}$. La suma de dichos subespacios es el plano XOY, es decir,

$$OX + OY = XOY = \{(x, y, 0) : x, y \in \mathbb{R}\}.$$

Obsérvese que como resultado de la suma hemos obtenido un nuevo subespacio vectorial de $\mathbb{R}^3$.

Proposición 5 *La suma de subespacios vectoriales de $\mathbb{K}^n$ es un subespacio vectorial de $\mathbb{K}^n$.*

Demostración: $\mathbf{0} \in M + N$ ya que pertenece a ambos subespacios y $\mathbf{0} = \mathbf{0} + \mathbf{0}$. Sean $\mathbf{x}, \mathbf{y} \in M + N$, es decir, $\mathbf{x} = \mathbf{u}_1 + \mathbf{v}_1$ e $\mathbf{y} = \mathbf{u}_2 + \mathbf{v}_2$, con $\mathbf{u}_j \in M$ y $\mathbf{v}_j \in N$. Entonces, $\forall \lambda, \mu \in \mathbb{K}$,

$$\lambda\mathbf{x} + \mu\mathbf{y} = \lambda(\mathbf{u}_1 + \mathbf{v}_1) + \mu(\mathbf{u}_2 + \mathbf{v}_2) = \lambda\mathbf{u}_1 + \lambda\mathbf{v}_1 + \mu\mathbf{u}_2 + \mu\mathbf{v}_2 = \tag{1.14}$$

$$(\lambda\mathbf{u}_1 + \mu\mathbf{u}_2) + (\lambda\mathbf{v}_1 + \mu\mathbf{v}_2) \in M + N, \tag{1.15}$$

ya que $\lambda\mathbf{u}_1 + \mu\mathbf{u}_2$ pertenece a M por ser combinación lineal de vectores de M, e igualmente le sucede al segundo sumando respecto a N. □

Definición 18 (suma directa) *Dados dos subespacios vectoriales de $\mathbb{K}^n$, diremos que su suma es directa, y escribiremos $L \oplus M$, si $L \cap M = \{\mathbf{0}\}$.*

Nota: Para más de dos subespacios, se dice que su suma es directa si lo es la de cada subespacio con la suma de todos los demás.

Proposición 6 (Caracterización de suma directa) *Sean L, M subespacios vectoriales de $\mathbb{K}^n$; la suma $L+M$ es directa si y sólo si $\forall \mathbf{u} \in L+M$, existen* ***únicos*** *$\mathbf{v} \in L$, $\mathbf{w} \in M$, tales que $\mathbf{u} = \mathbf{v} + \mathbf{w}$.*

Demostración: ***(no exigible)***

$\Rightarrow$) Supongamos que se tiene $L \oplus M$, es decir, $L \cap M = \{\mathbf{0}\}$, y sea un vector $\mathbf{u} = \mathbf{v}_1 + \mathbf{w}_1 = \mathbf{v}_2 + \mathbf{w}_2$ con $\mathbf{v}_j \in L$ y $\mathbf{w}_j \in M$. Entonces $\mathbf{v}_1 - \mathbf{v}_2 = \mathbf{w}_2 - \mathbf{w}_1$ es un vector de L según la expresión del primer miembro y de M como consecuencia de la del segundo, es decir, que pertenece a $L \cap M$ y es por tanto el vector nulo. Por consiguiente, $\mathbf{v}_1 = \mathbf{v}_2$ y $\mathbf{w}_1 = \mathbf{w}_2$ y la descomposición de $\mathbf{u}$ es única.

$\Leftarrow$) Supongamos que todo vector de $L + M$ se descompone de manera única. Dado cualquier vector $\mathbf{u} \in L \cap M$ podemos escribirlo como $\mathbf{u} = \mathbf{u} + \mathbf{0}$ pero también como $\mathbf{u} = \mathbf{0} + \mathbf{u}$, siendo en ambos casos el primer sumando perteneciente a L y el segundo a M. La unicidad de la descomposición obliga a que sea $\mathbf{u} = \mathbf{0}$ y, consecuentemente, $L \cap M = \{\mathbf{0}\}$. □

Definición 19 (subespacios suplementarios) *Dos subespacios L y M de $\mathbb{K}^n$ se dicen suplementarios si $L \oplus M = \mathbb{K}^n$.*

Observación: Dados dos subespacios suplementarios L, M, todo vector de $\mathbb{K}^n$ se descompone de forma única como suma de un vector de L y otro de M.

Ejemplo: En $\mathbb{R}^2$ cualesquiera dos rectas distintas que pasen por el origen son suplementarias. En $\mathbb{R}^3$, cualquier plano que contenga al origen es suplementario a cualquier recta que lo corte únicamente en dicho punto.

1.2.4. La relación de Grassmann

Observación: Dados dos subespacios M_1 y M_2, de sistemas generadores respectivos $\mathcal{M}_1$ y $\mathcal{M}_2$, se tiene que $\mathcal{M}_1 \cup \mathcal{M}_2$ es un sistema generador de $M_1 + M_2$; es decir, que **la unión de los sistemas generadores es un sistema generador de la suma**; en particular, la unión de las bases es un sistema generador (pero no necesariamente base) de la suma.

Proposición 7 (Relación de Grassmann) *Dados dos subespacios vectoriales L, M de $\mathbb{K}^n$, se tiene que*

$$\dim(L + M) = \dim L + \dim M - \dim(L \cap M). \tag{1.16}$$

Demostración: ***(no exigible)***

Llamamos $l = \dim L,\ m = \dim M,\ k = \dim(L \cap M)$; queremos demostrar que $\dim(L + M) = l + m - k$.

Sea $(\mathbf{u}_1, \ldots, \mathbf{u}_k)$ una base de $L \cap M$. Puesto que $L \cap M \subset L$, aplicando el teorema de compleción de la base en este subespacio, sabemos que existen $\mathbf{v}_1, \ldots, \mathbf{v}_{l-k} \in L$ tales que $(\mathbf{u}_1, \ldots, \mathbf{u}_k, \mathbf{v}_1, \ldots, \mathbf{v}_{l-k})$ es una base de L; pero también, como $L \cap M \subset M$, podemos aplicar igualmente el teorema de compleción de la base en el espacio M: existen $\mathbf{w}_1, \ldots, \mathbf{w}_{m-k} \in M$ tales que $(\mathbf{u}_1, \ldots, \mathbf{u}_k, \mathbf{w}_1, \ldots, \mathbf{w}_{m-k})$ es una base de M.

A continuación construimos el sistema

$$(\mathbf{u}_1, \ldots, \mathbf{u}_k, \mathbf{v}_1, \ldots, \mathbf{v}_{l-k}, \mathbf{w}_1, \ldots, \mathbf{w}_{m-k}) \tag{1.17}$$

y probaremos que es una base de $L + M$:

- Es sistema generador de $L + M$, pues es la unión de sendas bases de L y de M.
- Es sistema libre: supongamos que, para $\lambda_j, \mu_j, \gamma_j \in \mathbb{K}$,

  $$\sum_{j=1}^{k} \lambda_j \mathbf{u}_j + \sum_{j=1}^{l-k} \mu_j \mathbf{v}_j + \sum_{j=1}^{m-k} \gamma_j \mathbf{w}_j = \mathbf{0}, \tag{1.18}$$

  es decir,

  $$\sum_{j=1}^{k} \lambda_j \mathbf{u}_j + \sum_{j=1}^{l-k} \mu_j \mathbf{v}_j = -\sum_{j=1}^{m-k} \gamma_j \mathbf{w}_j. \tag{1.19}$$

  El primer miembro de la anterior igualdad es combinación lineal de los elementos de la base de L; el segundo miembro lo es de algunos elementos de la base de M; por lo tanto, pertenece a $L \cap M$, es decir:

  $$\sum_{j=1}^{m-k} \gamma_j \mathbf{w}_j \in L \cap M. \tag{1.20}$$

Consecuentemente, el sumatorio anterior es combinación lineal de los vectores de la base $(\mathbf{u}_1, \ldots, \mathbf{u}_k)$ de $L \cap M$, esto es,

$$\sum_{j=1}^{m-k} \gamma_j \mathbf{w}_j = \sum_{j=1}^{k} \alpha_j \mathbf{u}_j \tag{1.21}$$

para algunos escalares $\alpha_j \in \mathbb{K}$; pero como los vectores $\mathbf{u}_j, \mathbf{w}_j$ son linealmente independientes por constituir todos juntos una base de M, tenemos que todos los coeficientes en la anterior igualdad son nulos; en particular,

$$\gamma_j = 0, \ j = 1, 2, \ldots, m-k. \tag{1.22}$$

Pero esto, sustituyendo en (1.18), nos lleva a que

$$\sum_{j=1}^{k} \lambda_j \mathbf{u}_j + \sum_{j=1}^{l-k} \mu_j \mathbf{v}_j = \mathbf{0}, \tag{1.23}$$

que, por la independencia lineal de los $\mathbf{u}_j, \mathbf{v}_j$ (pues todos juntos constituyen una base de L), nos permite concluir que todos los coeficientes λ_j, μ_j son nulos. Con lo que concluimos que todos los coeficientes $\lambda_j, \mu_j, \gamma_j$ de la combinación lineal (1.18) de partida son nulos.

Recapitulando: hemos construido una base de $L + M$ que tiene $k + (l-k) + (m-k) = l + m - k$ elementos, luego $\dim(L+M) = l + m - k$, como queríamos demostrar. □

Ejemplo: En $\mathbb{R}^4$, se consideran los subespacios M definido mediante las ecuaciones

$$M : \left\{ \begin{array}{rcl} x_1 + x_2 - x_3 & = & 0 \\ x_2 - x_4 & = & 0 \end{array} \right.$$

y $N = L[(1,2,1,0),(1,1,0,-1)]$. Encuéntrense sendas bases de los subespacios $M \cap N$ y $M + N$, así como unas ecuaciones implícitas de los mismos.

Solución: Calcularemos primero una base de $M \cap N$. El vector genérico de N es

$$\alpha(1,2,1,0) + \beta(1,1,0,-1) = (\alpha+\beta, 2\alpha+\beta, \alpha, -\beta)$$

al cual, para que pertenezca simultáneamente a M, le imponemos las ecuaciones de este subespacio:

$$\alpha + \beta + 2\alpha + \beta - \alpha = 2\alpha + 2\beta = 0 \Leftrightarrow \beta = -\alpha;$$

$$2\alpha + \beta + \beta = 2\alpha + 2\beta = 0 \Leftrightarrow \beta = -\alpha.$$

Así pues, los vectores de $M \cap N$ dependen de un único parámetro (es decir, $\dim(M \cap N) = 1$), siendo estos vectores de la forma $\alpha[(1,2,1,0) - (1,1,0,-1)] = \alpha(0,1,1,1)$ y, por tanto, una base de $N \cap M$ es $((0,1,1,1))$.

Ahora, por la relación de Grassmann, sabemos que $\dim(M+N) = 2+2-1 = 3$; para construir una base de $M+N$, partimos de la base de la intersección y la completamos con sendos vectores de $M \setminus N = M \setminus (M \cap N)$ y de $N \setminus M = N \setminus (M \cap N)$. Como ningún vector no proporcional a $(0,1,1,1)$ puede pertenecer a $N \cap M$, tomamos por ejemplo $(1,0,1,0) \in M \setminus N$ (pues satisface las ecuaciones de M) y $(1,1,0,-1) \in N \setminus M$. Hemos construido, pues, una base de $M+N$,

$$((0,1,1,1),(1,0,1,0),(1,1,0,-1))$$

en la que el primer vector constituye una base de $M \cap N$, el primero junto al segundo forman una base de M, y el primero más el tercero son base de N.

En cuanto a las ecuaciones implícitas, dado que $\dim(M \cap N) = 1$, sabemos que este subespacio estará caracterizado por $4-1=3$ ecuaciones independientes. Los vectores proporcionales a $(0,1,1,1)$ se caracterizan por ser $x_1 = 0, x_2 = x_3 = x_4$, así que

$$\left\{ \begin{array}{rcl} x_1 & = & 0 \\ x_2 - x_3 & = & 0 \\ x_3 - x_4 & = & 0 \end{array} \right.$$

son unas ecuaciones implícitas de $M \cap N$ (resolución sistemática: se elimina el parámetro α de las ecuaciones paramétricas $(x_1,x_2,x_3,x_4) = (0,\alpha,\alpha,\alpha)$).

Respecto a $M+N$, como su dimensión vale 3, estará caracterizado por $4-3=1$ ecuación. Para calcularla, obligamos a que el vector (x_1,x_2,x_3,x_4) sea combinación lineal de los de la base del subespacio, por ejemplo escribiendo

$$\left(\begin{array}{ccc|c} 1 & 1 & 0 & x_1 \\ 1 & 0 & 1 & x_2 \\ 0 & 1 & 1 & x_3 \\ -1 & 0 & 1 & x_4 \end{array} \right)$$

(se ha alterado el orden de los vectores de la base por comodidad en los cálculos) y obligando a que el sistema lineal cuya matriz ampliada es la antescrita sea compatible. Realizando operaciones de reducción gaussiana por filas,

$$\left(\begin{array}{ccc|c} 1 & 1 & 0 & x_1 \\ 0 & -1 & 1 & x_2 - x_1 \\ 0 & 1 & 1 & x_3 \\ 0 & 1 & 1 & x_4 + x_1 \end{array} \right) \sim \left(\begin{array}{ccc|c} 1 & 1 & 0 & x_1 \\ 0 & -1 & 1 & x_2 - x_1 \\ 0 & 1 & 1 & x_3 \\ 0 & 0 & 0 & x_4 + x_1 - x_3 \end{array} \right),$$

de modo que el sistema será compatible siempre que $x_1 - x_3 + x_4 = 0$ y ésta es la ecuación implícita de $M+N$.

Nótese que, puesto que la matriz es cuadrada, se obtiene el mismo resultado imponiendo que el determinante sea igual a 0.

Observación: Para cualesquiera subespacios L, M de intersección nula, $\dim(L \oplus M) = \dim L + \dim M$.

En particular, si L, M son dos subespacios suplementarios en $\mathbb{K}^n$, entonces $n = \dim L + \dim M$.

Ejemplo: Encuéntrese una base de un suplementario en $\mathbb{R}^4$ del subespacio

$$M = L[(1,2,-1,-2),(1,1,1,0)].$$

Solución: Sabemos que la dimensión del suplementario ha de ser $4-2=2$. Si completamos la base de M hasta una de $\mathbb{R}^4$, los vectores añadidos constituirán una base de un subespacio N tal que $M+N=\mathbb{R}^4$, pero también $\dim(M \cap N) = \dim M + \dim N - \dim(M+N) = 4-4 = 0$, luego la suma anterior será directa o, lo que es lo mismo, N será un suplementario de M.

Así pues, nos limitaremos a completar la base dada hasta una de $\mathbb{R}^4$. Probamos, por ejemplo, con los dos últimos vectores canónicos:

$$\begin{pmatrix} 1 & 2 & -1 & -2 \\ 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \sim \begin{pmatrix} 1 & 2 & -1 & -2 \\ 0 & -1 & 2 & 2 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

La forma escalonada de la matriz obtenida nos garantiza que la familia es libre, luego un suplementario de M en $\mathbb{R}^4$ es

$$N = L[(0,0,1,0),(0,0,0,1)].$$

1.3. Ejercicios

1.1.– Se consideran los vectores $(1,-1,2)\,,(2,0,1)\,,(a,1,0)$ de $\mathbb{R}^3$.

1. Determinar los valores del parámetro a para que sean linealmente independientes.
2. Para $a=3$, estudiar si el vector $(1,1,-1)$ es combinación lineal de ellos.

1.2.– Determinar los valores de los parámetros λ y μ que hacen que los tres vectores $(1,2,-1,3)\,,(1,-1,\lambda,3)$ y $(3,2,5,\mu)$ sean linealmente independientes.

1.3.– Comprobar la dependencia o independencia lineal de los vectores de $\mathbb{C}^3$

$$(1+i,1,0),(1+i,1-3i,1-i),(-1,1-i,1).$$

1.4.– Determinar justificadamente cuáles de los siguientes subconjuntos son subespacios vectoriales. Para aquellos que lo sean, encontrar una base.

1. $P=\{(x_1,x_2,x_3,x_4)\in\mathbb{R}^4 : x_1+x_2=x_3+x_4 \wedge x_1-x_2=x_3-x_4\}$.
2. $H=\{(x,y,z)\in\mathbb{R}^3 : (x+y-z)(x-y+z)=0\}$.
3. $M=\{(x,y,z)\in\mathbb{R}^3 : x+2y-z=1\}$.
4. $N=\{(x,y,z)\in\mathbb{R}^3 : (x,y,z)\times(1,2,-1)=(0,0,0)\}$.
5. $S=L[(1,-2,1),(1,-1,0)]\cup L[(1,-3,2)]$.
6. $J=L[(1,-2,1),(1,-1,0)]\cup L[(2,-3,0)]$.

1.5.– En $\mathbb{R}^4$ se considera el subespacio L generado por los vectores $(1,1,1,1)\,,(1,2,3,4)\,,$ $(5,3,1,-1)\,,(0,1,0,-1)$. Se pide:

1. Comprobar que esos cuatro vectores son linealmente dependientes.
2. Hallar una base de L.
3. Construir una base de $\mathbb{R}^4$ extendiendo la base anterior con vectores cuyas tres primeras componentes sean iguales a 2.

1.6.– Calcular las coordenadas del vector (a,b,c) de $\mathbb{R}^3$ respecto a la base $((1,1,0),(1,0,1),$ $(0,1,1))$.

1.7.– En $\mathbb{R}^3$ se considera una base $\mathcal{B}=(\mathbf{u}_1,\mathbf{u}_2,\mathbf{u}_3)$, y el conjunto

$$\mathcal{V}=(\mathbf{u}_1+\mathbf{u}_3,\mathbf{u}_2-2\mathbf{u}_1,\mathbf{u}_3-\mathbf{u}_2)\,.$$

Comprobar que $\mathcal{V}$ también es base de $\mathbb{R}^3$ y determinar las coordenadas del vector $\mathbf{u}_1+\mathbf{u}_2+\mathbf{u}_3$ respecto a dicha base.

1.8.– En $\mathbb{R}^5$ se considera el subespacio M que tiene por base $((1,0,1,0,1),(0,1,-1,1,0),(1,1,0,-1,-1))$. Se pide:

1. Hallar unas ecuaciones implícitas de M.
2. Determinar los valores de a,b para que el vector $(1,1,a,b,1)$ pertenezca a M.
3. Para esos valores, hallar las coordenadas de dicho vector respecto a la base del enunciado.

1.9.– Se considera el subespacio de $\mathbb{R}^4$ $H = L[(1,2,0,1),(1,0,1,-1),(1,-2,2,-3)]$. Se pide:

1. Hallar unas ecuaciones implícitas de H.
2. Hallar una base y unas ecuaciones implícitas de un suplementario de H cuyos vectores tengan sus dos últimas componentes iguales pero no todas nulas.

1.10.– Sea V el subespacio de $\mathbb{R}^4$ definido por las ecuaciones implícitas

$$x_1 + x_2 - x_3 = 0,\ \ x_1 + x_4 = 0,\ \ x_2 - x_3 - x_4 = 0.$$

Escribir las ecuaciones de un subespacio suplementario de V. ¿Es único este suplementario?

1.11.– Hallar una base de cada uno de los subespacios $U+V$ y $U \cap V$, siendo:

1. $U = L[(1,0,2,2),(1,1,0,1)]$, $V = L[(1,2,-2,0),(0,1,1,2)]$.
2. $U = L[(1,2,1,0),(0,-1,0,1)]$, $V = L[(0,0,1,1),(1,1,2,2)]$.

1.12.– Sean F y G dos subespacios de $\mathbb{R}^n$ cuyas dimensiones son p y q. Determinar todos los posibles valores de las dimensiones de los subespacios $F+G$ y $F \cap G$.

1.3.1. Cuestiones

1.13.– Se consideran los subespacios M, N de $\mathbb{R}^4$ definidos por:

$$M = L[(0,1,1,0),(-1,1,3,1),(1,1,-1,-1)]\ ;\ N = \begin{cases} x_1 - x_2 + x_3 - x_4 = 0 \\ x_1 + x_3 + x_4 = 0 \\ 2x_1 - x_2 + 2x_3 = 0 \end{cases}$$

Determínese la veracidad o falsedad de las siguientes afirmaciones:

1. $\dim M = 3$
2. $\dim N = 2$

3. $M + N = \mathbb{R}^4$
4. $(0, 1, 1, 0) \in M \cap N$
5. $(1, 2, 0, -1) \in M$

1.14.– Sean $\mathbf{u}, \mathbf{v}, \mathbf{w}$ tres vectores linealmente independientes de $\mathbb{R}^n$. Si α, β, γ denotan números reales, determínese la veracidad o falsedad de las siguientes afirmaciones:

1. Si $\alpha\mathbf{u} + \beta\mathbf{v} = \mathbf{0}$, entonces $\alpha = \beta = 0$.
2. No existen $\alpha, \beta \in \mathbb{R}$ tales que $\mathbf{w} = \alpha\mathbf{u} + \beta\mathbf{v}$.
3. Se cumple que $\alpha\mathbf{u} + \beta\mathbf{v} + \gamma\mathbf{w} = \mathbf{0} \Leftrightarrow \alpha = \beta = \gamma = 0$.
4. $(\alpha + \beta + \gamma)(\mathbf{u} + \mathbf{v} + \mathbf{w}) = \mathbf{0} \Rightarrow \alpha = \beta = \gamma = 0$.
5. $\mathbf{u} + \mathbf{v} + \mathbf{w} \neq \mathbf{0}$.

1.15.– Se considera la familia $\mathcal{M} = \{(1, 1, a), (1, a, a), (a, a, a)\}$, donde a es un número real. Determínese la veracidad o falsedad de las siguientes afirmaciones:

1. $\mathcal{M}$ no puede ser base de $\mathbb{R}^3$.
2. $\forall a \neq 0,\ \mathcal{M}$ es base de $\mathbb{R}^3$.
3. Si $a = \dfrac{\pi}{3}$, $\mathcal{M}$ es una familia libre.
4. $\forall a \in \mathbb{R},\ \mathcal{M}$ es un sistema generador de $\mathbb{R}^3$.
5. Si $a = 0$, $\mathcal{M}$ es una familia libre.

1.16.– En el espacio $\mathbb{R}^4$, se considera el subespacio H definido mediante la ecuación $x_1+2x_2+3x_3+4x_4 = 0$. Determínese la veracidad o falsedad de las siguientes afirmaciones:

1. $\dim H = 3$.
2. $L[(1, 1, 1, 1)]$ es un subespacio suplementario de H en $\mathbb{R}^4$.
3. $L[(1, 0, 0, 0), (1, 1, 1, 1)] + H = \mathbb{R}^4$.
4. Si M es un subespacio suplementario de H en $\mathbb{R}^4$, $\dim M = 1$.
5. $L[(1, 0, 0, 0), (1, 1, 1, 1)] \cap H = \{\mathbf{0}\}$.

1.17.– Se considera la base $\mathcal{B} = ((1, 1, 0), (1, 0, 1), (0, 1, 1))$ de $\mathbb{R}^3$. Determínese la veracidad o falsedad de las siguientes afirmaciones:

1. El vector de coordenadas de $(2, 0, 0)$ respecto a $\mathcal{B}$ es $(1, 1, -1)$.

2. El vector de coordenadas de $(1, 1, 0)$ respecto a $\mathcal{B}$ es $(1, 1, 0)$.
3. El vector de coordenadas de $(0, 0, 1)$ respecto a $\mathcal{B}$ es $\frac{1}{2}(1, 1, -1)$.
4. $(0, 1, 0)$ es combinación lineal de los dos primeros vectores de $\mathcal{B}$.
5. Todo vector de $\mathbb{R}^3$ es combinación lineal de los vectores de $\mathcal{B}$.

Capítulo 2

Matrices y sistemas lineales

2.1. Matrices

Definición 20 *Sean m, n enteros positivos. Una* ***matriz*** *de tamaño $m \times n$ es una tabla rectangular formada por mn escalares de $\mathbb{K}$ que se disponen como se indica:*

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}. \tag{2.1}$$

Cada a_{ij} se llama elemento, coeficiente o entrada de la matriz. Se dice que está situado en la ***fila*** *i y en la* ***columna*** *j de la matriz.*

La matriz está formada por m filas que son vectores de $\mathbb{K}^n$ y por n columnas que lo son de $\mathbb{K}^m$. Solemos utilizar letras mayúsculas para indicar una matriz $\mathbf{A}, \mathbf{B}, \mathbf{M}$, etc. y la correspondiente minúscula para sus entradas: a_{ij}, b_{ij}, m_{ij}. A veces expresaremos $a_{ij} = (\mathbf{A})_{ij}$. En general se intentará que la notación se explique a sí misma. El vector $F_i(\mathbf{A}) =$

$(a_{i1}, a_{i2}, \ldots, a_{in})$ constituye la fila i-ésima de la matriz y el vector

$$C_j(\mathbf{A}) = \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{pmatrix} \tag{2.2}$$

es la columna j-ésima de la matriz.

El conjunto de las matrices con m filas y n columnas y coeficientes en $\mathbb{K}$ se denota $\mathbb{K}^{m\times n}$. Dicho conjunto está dotado de las operaciones suma y producto por un escalar:

$$\begin{array}{rcl} \mathbf{A}, \mathbf{B} & \in & \mathbb{K}^{m\times n} \;\; (\mathbf{A}+\mathbf{B})_{ij} := a_{ij} + b_{ij}, \\ \lambda & \in & \mathbb{K},\, \mathbf{A} \in \mathbb{K}^{m\times n} \;\; (\lambda\mathbf{A})_{ij} := \lambda a_{ij}. \end{array}$$

A los elementos a_{ii} de igual índice de fila que de columna se les llama elementos diagonales de la matriz y constituyen la *diagonal principal* de la matriz.

En particular, si $m = n$ se dice que la matriz es **cuadrada** de orden n. La matriz cuadrada cuyos elementos diagonales son iguales a 1 y el resto nulos se llama **matriz identidad** (o **matriz unidad**) y se denota con la letra $\mathbf{I}$:

$$\mathbf{I} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}. \tag{2.3}$$

Observación: La matriz $\mathbf{I}$ de orden n corresponde a la aplicación identidad de $\mathbb{K}^n$, es decir, la aplicación que deja invariantes todos los vectores de $\mathbb{K}^n$:

$$\forall \mathbf{x} \in \mathbb{K}^n,\; \mathcal{I}(\mathbf{x}) = \mathbf{I}\mathbf{x} = \mathbf{x}.$$

Definición 21 (matrices triangulares)

- *Se dice que* $\mathbf{A}$ *es* ***triangular superior*** *si sus elementos por debajo de la diagonal principal son nulos, es decir, si* $a_{ij} = 0$ *para todo* $i > j$.
- *Se dice que* $\mathbf{A}$ *es* ***triangular inferior*** *si sus elementos por encima de la diagonal principal son nulos, es decir, si* $a_{ij} = 0$ *para todo* $i < j$.
- *Se dice que* $\mathbf{A}$ *es* ***diagonal*** *si es triangular superior e inferior simultáneamente.*

Ejemplo: La matriz identidad $\mathbf{I}$ es un ejemplo de matriz diagonal: obviamente, es tanto triangular superior como triangular inferior.

Definición 22 *Para cada matriz $\mathbf{A} \in \mathbb{K}^{m\times n}$, se define:*

- *su **traspuesta**, que se denota por $\mathbf{A}^t$ y es la matriz de $\mathbb{K}^{n\times m}$ que se obtiene a partir de $\mathbf{A}$ cambiando filas por columnas, esto es:*

$$(\mathbf{A}^t)_{ij} = a_{ji}\ ,\ i = 1, 2, \ldots, m,\ j = 1, 2, \ldots, n. \tag{2.4}$$

- *su **conjugada**, que se denota por $\overline{\mathbf{A}}$ y es la matriz de $\mathbb{K}^{m\times n}$ cuyos elementos son los conjugados de los de $\mathbf{A}$, $(\overline{\mathbf{A}})_{ij} = \overline{a}_{ij}$.*
- *su **traspuesta conjugada**, que se denota por $\mathbf{A}^h$ y es la traspuesta de la conjugada de $\mathbf{A}$ (o bien la conjugada de la traspuesta de $\mathbf{A}$), $(\mathbf{A}^h)_{ij} = \overline{a}_{ji}$.*

Definición 23 (matrices simétricas y antisimétricas, hermíticas y antihermíticas)

- *Una matriz cuadrada $\mathbf{A}$ se dice que es **simétrica si** $\mathbf{A}^t = \mathbf{A}$, es decir, si $a_{ji} = a_{ij}$ para todo i, j.*
- *Una matriz cuadrada $\mathbf{A}$ se dice que es **antisimétrica si** $\mathbf{A}^t = -\mathbf{A}$, es decir, si $a_{ji} = -a_{ij}$ para todo i, j.*
- *Una matriz cuadrada $\mathbf{A}$ se dice que es **hermítica si** $\mathbf{A}^h = \mathbf{A}$, es decir, si $a_{ji} = \overline{a_{ij}}$ para todo i, j.*
- *Una matriz cuadrada $\mathbf{A}$ se dice que es **antihermítica si** $\mathbf{A}^h = -\mathbf{A}$, es decir, si $a_{ji} = -\overline{a_{ij}}$ para todo i, j.*

Obsérvese que los elementos diagonales de una matriz real antisimétrica deben ser nulos.

2.1.1. Aplicaciones lineales

En numerosas ramas de la ciencia y la técnica son fundamentales las aplicaciones entre espacios vectoriales que **conservan las combinaciones lineales**, es decir, aquellas para las que el transformado de toda combinación lineal de vectores es la combinación lineal, con los mismos coeficientes, de los transformados de los vectores originales.

Definición 24 *Sea una aplicación $f : \mathbb{K}^n \to \mathbb{K}^m$; se dice que f es lineal si verifica las dos condiciones siguientes:*

$$\forall \mathbf{u}, \mathbf{v} \in \mathbb{K}^n,\ f(\mathbf{u} + \mathbf{v}) = f(\mathbf{u}) + f(\mathbf{v}); \tag{2.5}$$

$$\forall \mathbf{u} \in \mathbb{K}^n, \forall \lambda \in \mathbb{K},\ f(\lambda \mathbf{u}) = \lambda f(\mathbf{u}). \tag{2.6}$$

Equivalentemente, f es lineal si

$$\forall \mathbf{u}, \mathbf{v} \in \mathbb{K}^n, \forall \lambda, \mu \in \mathbb{K},\ f(\lambda \mathbf{u} + \mu \mathbf{v}) = \lambda f(\mathbf{u}) + \mu f(\mathbf{v}). \tag{2.7}$$

La condición (2.7) que caracteriza dichas aplicaciones se llama propiedad de **linealidad**.

$\mathbb{K}^n$ se denomina **espacio inicial** de la aplicación lineal f, mientras que $\mathbb{K}^m$ es el **espacio final**.

Observación:

$$f(\mathbf{0}_{\mathbb{K}^n}) = f(0 \cdot \mathbf{0}_{\mathbb{K}^n}) = 0 \cdot f(\mathbf{0}_{\mathbb{K}^n}) = \mathbf{0}_{\mathbb{K}^m}. \tag{2.8}$$

Observación: Aplicando reiteradamente la definición se tiene que, cualquiera que sea el número k de sumandos, si $\lambda_j \in \mathbb{K}$ y $\mathbf{u}_j \in \mathbb{K}^n$,

$$f\left(\sum_{j=1}^{k} \lambda_j \mathbf{u}_j\right) = \sum_{j=1}^{k} \lambda_j f(\mathbf{u}_j). \tag{2.9}$$

Proposición 8 *Dada una aplicación lineal $f : \mathbb{K}^n \to \mathbb{K}^m$ y una base $\mathcal{B} = (\mathbf{u}_1, \ldots, \mathbf{u}_n)$ de $\mathbb{K}^n$, los vectores $f(\mathbf{u}_1), \ldots, f(\mathbf{u}_n)$ de $\mathbb{K}^m$ determinan completamente f, es decir, nos permiten conocer la imagen mediante f de cualquier vector del espacio inicial.*

En efecto, si $\mathbf{x} = \sum_{j=1}^{n} x_j \mathbf{u}_j$ con $x_j \in \mathbb{K}$, por la propiedad de linealidad (2.9) se tiene que

$$f(\mathbf{x}) = f\left(\sum_{j=1}^{n} x_j \mathbf{u}_j\right) = \sum_{j=1}^{n} x_j f(\mathbf{u}_j). \tag{2.10}$$

Como caso particular, si $\mathbf{e}_1, \ldots, \mathbf{e}_n$ son los vectores canónicos de $\mathbb{K}^n$, los n vectores $f(\mathbf{e}_1), \ldots, f(\mathbf{e}_n)$ determinan cualquier aplicación lineal f con origen en $\mathbb{K}^n$.

2.1.2. Matriz de una aplicación lineal

Para cualquier vector $\mathbf{x} = (x_1, \ldots, x_n)$ del espacio inicial, es decir, $\mathbf{x} = \sum_{j=1}^{n} x_j \mathbf{e}_j$, sabemos que su transformado mediante f es de la forma

$$\mathbf{y} = f(\mathbf{x}) = \sum_{j=1}^{n} x_j f(\mathbf{e}_j). \tag{2.11}$$

Si consideramos los n vectores $f(\mathbf{e}_1), \ldots, f(\mathbf{e}_n)$, pertenecientes a $\mathbb{K}^m$, cada uno de ellos podrá expresarse como

$$f(\mathbf{e}_j) = (a_{1j}, \ldots, a_{mj}) = \sum_{i=1}^{m} a_{ij} \tilde{\mathbf{e}}_i, \tag{2.12}$$

en donde $\tilde{\mathbf{e}}_1, \ldots, \tilde{\mathbf{e}}_m$ denotan los vectores canónicos de $\mathbb{K}^m$ y $a_{ij} \in \mathbb{K}$ para $1 \le i \le m,\ 1 \le j \le n$.

Si llevamos este resultado a la ecuación (2.11) y cambiamos el orden de sumación, obtenemos:

$$\mathbf{y} = f(\mathbf{x}) = \sum_{j=1}^{n} x_j \left(\sum_{i=1}^{m} a_{ij} \tilde{\mathbf{e}}_i \right) = \sum_{i=1}^{m} \left(\sum_{j=1}^{n} a_{ij} x_j \right) \tilde{\mathbf{e}}_i. \tag{2.13}$$

Si ahora identificamos cada componente en la ecuación vectorial anterior, podemos escribir las m ecuaciones

$$y_i = \sum_{j=1}^{n} a_{ij} x_j,\ 1 \le i \le m, \tag{2.14}$$

que escribimos desarrolladas para mayor claridad:

$$\begin{array}{rcl} y_1 & = & a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n \\ y_2 & = & a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n \\ \vdots & & \vdots \\ y_m & = & a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n \end{array} \tag{2.15}$$

Es decir, que los $m \cdot n$ escalares a_{ij} determinan completamente la aplicación lineal, permitiéndonos calcular las componentes del transformado de cualquier vector en función de las componentes de este.

Este resultado sugiere que la relación entre un vector genérico de $\mathbb{K}^n$ y su transformado mediante f se escriba en *forma matricial* como sigue:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \tag{2.16}$$

(Más adelante resultará evidente que el producto del segundo miembro corresponde al concepto de *producto matricial*.)

Y abreviadamente,

$$\mathbf{y} = \mathbf{A}\mathbf{x}. \tag{2.17}$$

Todo ello motiva la siguiente definición:

Definición 25 (matriz de una aplicación lineal) *Dada una aplicación lineal $f : \mathbb{K}^n \to \mathbb{K}^m$, se llama matriz de f a aquella cuyas columnas son, ordenadamente, las imágenes mediante f de los vectores canónicos de $\mathbb{K}^n$.*

Observación: Se trata, por consiguiente, de una matriz de m filas (= dimensión del espacio final) y n columnas (= dimensión del espacio inicial).

Existe, por tanto, una **correspondencia biunívoca** entre aplicaciones lineales de $\mathbb{K}^n$ a $\mathbb{K}^m$ y matrices de m filas y n columnas (cada matriz determina una aplicación lineal y viceversa).

Ejemplo: Calcúlese la matriz de la aplicación lineal f de $\mathbb{R}^2$ a $\mathbb{R}^3$ caracterizada por

$$f(1,-1) = (1,2,-1);\ f(1,2) = (0,1,-1). \tag{2.18}$$

Solución: Necesitamos conocer los transformados mediante f de los vectores canónicos de $\mathbb{R}^2$, para lo cual precisamos la expresión de estos últimos como combinaciones lineales de los dos vectores cuyas imágenes conocemos:

$$(1,0) = \alpha(1,-1) + \beta(1,2) \text{ (resolviendo)} \Rightarrow \alpha = \frac{2}{3}, \beta = \frac{1}{3}; \tag{2.19}$$

$$(0,1) = \alpha(1,-1) + \beta(1,2) \text{ (resolviendo)} \Rightarrow \alpha = -\frac{1}{3}, \beta = \frac{1}{3}. \tag{2.20}$$

Por lo tanto,

$$f(\mathbf{e}_1) = \frac{2}{3}f(1,-1) + \frac{1}{3}f(1,2) = \frac{1}{3}(2,5,-3); \tag{2.21}$$

$$f(\mathbf{e}_2) = -\frac{1}{3}f(1,-1) + \frac{1}{3}f(1,2) = \frac{1}{3}(-1,-1,0). \tag{2.22}$$

Así pues, la matriz de f es la que tiene por columnas los dos vectores obtenidos, es decir:

$$\mathbf{A} = \frac{1}{3}\begin{pmatrix} 2 & -1 \\ 5 & -1 \\ -3 & 0 \end{pmatrix}. \tag{2.23}$$

Compruébese la exactitud del resultado aplicando dicha matriz a los vectores $(1,-1), (1,2)$ escritos como columnas. $\square$

2.1.3. Composición de aplicaciones lineales y producto matricial

Para motivar la definición de producto matricial, consideramos la composición de dos aplicaciones lineales. Supongamos que tenemos sendas aplicaciones lineales f, g definidas según el siguiente esquema:

$$\mathbb{K}^r \xrightarrow{g} \mathbb{K}^n \xrightarrow{f} \mathbb{K}^m \tag{2.24}$$

Consideramos ahora la aplicación $f \circ g : \mathbb{K}^r \to \mathbb{K}^m$. Veamos en primer lugar que esta aplicación también es lineal:

Proposición 9 *Toda composición de aplicaciones lineales es lineal.*

Demostración: En las condiciones antes establecidas, $\forall \mathbf{u}, \mathbf{v} \in \mathbb{K}^r$, $\forall \lambda, \mu \in \mathbb{K}$,

$$f \circ g(\lambda\mathbf{u} + \mu\mathbf{v}) = f(g(\lambda\mathbf{u} + \mu\mathbf{v})) = f(\lambda g(\mathbf{u}) + \mu g(\mathbf{v})) = \lambda f(g(\mathbf{u})) + \mu f(g(\mathbf{v})) =$$

$$= \lambda f \circ g(\mathbf{u}) + \mu f \circ g(\mathbf{v}). \tag{2.25}$$

□

La aplicación f tendrá asociada una cierta matriz $\mathbf{A} \in \mathbb{K}^{m\times n}$, mientras que g se corresponderá con una cierta $\mathbf{B} \in \mathbb{K}^{n\times r}$. Es razonable denotar la matriz $\mathbf{C} \in \mathbb{K}^{m\times r}$ de $f \circ g$, resultante de aplicar primero $\mathbf{B}$ y a continuación $\mathbf{A}$, como $\mathbf{AB}$. ¿Podremos expresar la matriz $\mathbf{C}$ en términos de las dos anteriores?

Si $\mathbf{x}$ denota un vector arbitrario de $\mathbb{K}^r$, $\mathbf{y} = g(\mathbf{x})$, $\mathbf{z} = f(\mathbf{y}) = f \circ g(\mathbf{x})$, sabemos por la ecuación (2.14) que cada componente de $\mathbf{y}$ se obtiene como

$$1 \le k \le n, \quad y_k = \sum_{j=1}^{r} b_{kj} x_j, \tag{2.26}$$

mientras que

$$1 \le i \le m, \quad z_i = \sum_{k=1}^{n} a_{ik} y_k = \sum_{j=1}^{n} a_{ik} \left(\sum_{j=1}^{r} b_{kj} x_j \right) = \sum_{j=1}^{r} \left(\sum_{k=1}^{n} a_{ik} b_{kj} \right) x_j. \tag{2.27}$$

Por otro lado, los elementos de la matriz $\mathbf{C}$ que transforma $\mathbf{x}$ en $\mathbf{z}$ satisfacen, por la misma ecuación (2.14),

$$1 \le i \le m, \quad z_i = \sum_{j=1}^{r} c_{ij} x_j. \tag{2.28}$$

Así pues, las ecuaciones (2.27) y (2.28) han de ser equivalentes. Ello sugiere el tomar como matriz $\mathbf{C} = \mathbf{AB}$ aquella de dimensiones $m \times r$ cuyo elemento genérico se calcula mediante

$$1 \le i \le m, \ 1 \le j \le r, \quad c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}, \tag{2.29}$$

que será la definición que adoptaremos para el producto de matrices.

Definición 26 *Dadas dos matrices* $\mathbf{A} \in \mathbb{K}^{m\times n}, \mathbf{B} \in \mathbb{K}^{n\times r}$*, se define su* ***producto*** $\mathbf{AB}$ *como aquella matriz* $\mathbf{C} \in \mathbb{K}^{m\times r}$ *tal que* $c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}, \ i = 1, 2, \ldots, m \ , \ j = 1, 2, \ldots, r.$

- El elemento (i, j) del producto se obtiene multiplicando la fila i de la primera matriz por la columna j de la segunda. Es preciso que el número de columnas de la primera matriz iguale al número de filas de la segunda.

- El producto matricial NO es conmutativo: en general $\mathbf{AB} \neq \mathbf{BA}$ aunque $\mathbf{A}, \mathbf{B}$ sean matrices cuadradas del mismo orden.

- El producto de matrices es distributivo respecto de la suma a izquierda y derecha, es decir:

$$\begin{aligned} \mathbf{A}(\mathbf{B} + \mathbf{C}) &= \mathbf{AB} + \mathbf{AC} \\ (\mathbf{B} + \mathbf{C})\mathbf{A} &= \mathbf{BA} + \mathbf{CA} \end{aligned}$$

  siempre y cuando las matrices tengan las dimensiones adecuadas para que dichos productos puedan realizarse.

Proposición 10 *Sean* $\mathbf{A}, \mathbf{B}$ *dos matrices para las que el producto* $\mathbf{AB}$ *tenga sentido; entonces*

$$(\mathbf{AB})^t = \mathbf{B}^t\mathbf{A}^t. \tag{2.30}$$

$$(\mathbf{AB})^h = \mathbf{B}^h\mathbf{A}^h. \tag{2.31}$$

Demostración: Cada elemento (i,j) de la matriz $(\mathbf{AB})^t$ puede escribirse como

$$\left((\mathbf{AB})^t\right)_{ij} = (\mathbf{AB})_{ji} = \sum_{k=1}^{n} a_{jk}b_{ki} = \sum_{k=1}^{n} b_{ki}a_{jk} = \sum_{k=1}^{n} \left(\mathbf{B}^t\right)_{ik}\left(\mathbf{A}^t\right)_{kj} = \left(\mathbf{B}^t\mathbf{A}^t\right)_{ij}. \tag{2.32}$$

Así pues, las matrices $(\mathbf{AB})^t$ y $\mathbf{B}^t\mathbf{A}^t$ coinciden elemento a elemento, por tanto son iguales: $(\mathbf{AB})^t = \mathbf{B}^t\mathbf{A}^t$. Finalmente, para demostrar la igualdad en el caso de la traspuesta conjugada, basta conjugar la ecuación anterior.

□

Producto matricial y combinaciones lineales

Sean $\mathbf{A} \in \mathbb{K}^{m\times n}$, $\mathbf{B} \in \mathbb{K}^{n\times r}$, denotemos C_k la columna k-ésima y F_k la fila, entonces:

- Las columnas de $\mathbf{AB}$ son combinaciones lineales de las de $\mathbf{A}$: $C_k(\mathbf{AB}) = \mathbf{A}C_k(\mathbf{B})$.
- Las filas de $\mathbf{AB}$ son combinaciones lineales de las de $\mathbf{B}$: $F_k(\mathbf{AB}) = F_k(\mathbf{A})\mathbf{B}$.
- La matriz $\mathbf{AB}$ puede expresarse como suma de matrices "producto columna-fila":

$$\mathbf{AB} = \sum_{k=1}^{n} C_k(\mathbf{A})F_k(\mathbf{B}). \tag{2.33}$$

La demostración de esta última igualdad es inmediata, teniendo en cuenta que el elemento (i,j) de la matriz del segundo miembro es

$$\sum_{k=1}^{n} \left(C_k(\mathbf{A})F_k(\mathbf{B})\right)_{ij} = \sum_{k=1}^{n} a_{ik}b_{kj} = (\mathbf{AB})_{ij} \tag{2.34}$$

de donde se deduce que ambas matrices coinciden elemento a elemento, luego son iguales.

2.1.4. La inversa de una matriz

Definición 27 *Una matriz cuadrada* $\mathbf{A} \in \mathbb{K}^{n\times n}$*, se dice que es* ***invertible*** *(o* ***regular****) si* $\exists \mathbf{A}^{-1} \in \mathbb{K}^{n\times n}$ *tal que* $\mathbf{AA}^{-1} = \mathbf{I}$ *o bien* $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$*. La matriz* $\mathbf{A}^{-1}$ *se denomina* ***inversa*** *de* $\mathbf{A}$ *y es única.*

Observación: Al ser $\mathbf{A}$ cuadrada, la propiedad $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$ implica que $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$ y viceversa.

Definición 28 *Una matriz cuadrada que no es invertible se dice que es* ***singular****.*

Proposición 11 *Sean* $\mathbf{A}$ *y* $\mathbf{B}$ *dos matrices cuadradas del mismo orden* n*, ambas invertibles. Entonces* $\mathbf{AB}$ *también es invertible y*

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}. \tag{2.35}$$

Demostración: Basta comprobar que el producto de $\mathbf{AB}$ por la matriz $\mathbf{B}^{-1}\mathbf{A}^{-1}$ da como resultado la identidad: en efecto, $(\mathbf{AB})(\mathbf{B}^{-1}\mathbf{A}^{-1}) = \mathbf{A}\left(\mathbf{B}\mathbf{B}^{-1}\right)\mathbf{A}^{-1} = \mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$ y también $(\mathbf{B}^{-1}\mathbf{A}^{-1})(\mathbf{AB}) = \mathbf{B}^{-1}(\mathbf{A}^{-1}\mathbf{A})\mathbf{B} = \mathbf{B}^{-1}\mathbf{B} = \mathbf{I}$. □

2.2. Imagen y núcleo de una matriz

Recordemos que cualquier matriz $\mathbf{A} \in \mathbb{K}^{m\times n}$ transforma vectores columna de $\mathbb{K}^n$ en vectores columna de $\mathbb{K}^m$; es decir, para cada $\mathbf{x} \in \mathbb{K}^n$ se obtiene un vector

$$\mathbf{y} = \mathbf{A}\mathbf{x} \in \mathbb{K}^m \tag{2.36}$$

Además se dice que $\mathbf{y}$ es la imagen de $\mathbf{x}$ por $\mathbf{A}$. Si consideramos el conjunto de todos los vectores imagen, dicho conjunto se denomina **imagen** de $\mathbf{A}$. Así pues, toda matriz $\mathbf{A} \in\mathbb{K}^{m\times n}$ lleva asociados dos conjuntos: su *imagen* en $\mathbb{K}^m$, y su *núcleo* en $\mathbb{K}^n$. Veamos estas definiciones:

Definición 29 (imagen y núcleo) *Dada una matriz* $\mathbf{A} \in \mathbb{K}^{m\times n}$ *se define la* ***imagen de*** $\mathbf{A}$ *como el conjunto*

$$\operatorname{Im}\mathbf{A} := \{\mathbf{A}\mathbf{x} : \mathbf{x} \in \mathbb{K}^n\}. \tag{2.37}$$

Se define el ***núcleo de*** $\mathbf{A}$ *como el conjunto*

$$\ker\mathbf{A} := \{\mathbf{u} \in \mathbb{K}^n : \mathbf{A}\mathbf{u} = \mathbf{0}\}. \tag{2.38}$$

Tanto el núcleo como la imagen de una matriz son **subespacios vectoriales** (la demostración es sencilla y se deja al lector).

Dada la correspondencia biunívoca entre aplicaciones lineales y matrices, pueden definirse equivalentemente el núcleo y la imagen de una aplicación lineal:

Definición 30 *Dada una aplicación lineal* $f : \mathbb{K}^n \to \mathbb{K}^m$*, se define la* ***imagen de*** f *como el conjunto*

$$\operatorname{Im} f := \{f(\mathbf{x}) : \mathbf{x} \in \mathbb{K}^n\}. \tag{2.39}$$

Se define el ***núcleo de*** f *como el conjunto*

$$\ker f := \{\mathbf{u} \in \mathbb{K}^n : f(\mathbf{u}) = \mathbf{0}\}. \tag{2.40}$$

Resulta evidente que, si consideramos la matriz asociada a una determinada aplicación lineal, el núcleo y la imagen de aquella coinciden con los de esta.

Observación: El vector $\mathbf{Ax}$ puede escribirse como

$$\mathbf{Ax} = \mathbf{A} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = x_1 C_1 (\mathbf{A}) + ... + x_n C_n (\mathbf{A}) \tag{2.41}$$

donde $C_j (\mathbf{A})$ indica la columna j-ésima de $\mathbf{A}$. De ello se deduce fácilmente que un vector es de la forma $\mathbf{Ax}$ si y sólo si es combinación lineal de las columnas de $\mathbf{A}$. Por ello Im$\mathbf{A}$ es el subespacio generado por las columnas de $\mathbf{A}$:

$$\operatorname{Im} \mathbf{A} = L[C_1 (\mathbf{A}), .., C_n(\mathbf{A})] \tag{2.42}$$

Definición 31 *El* ***subespacio columna*** *de una matriz* $\mathbf{A} \in \mathbb{K}^{m\times n}$ *es el subespacio de* $\mathbb{K}^m$ *engendrado por las columnas de* $\mathbf{A}$*. De igual forma, su* ***subespacio fila*** *es el subespacio de* $\mathbb{K}^n$ *engendrado por las filas de* $\mathbf{A}$*.*

Observación: Como acabamos de ver, el **subespacio columna** de una matriz coincide con $\operatorname{Im} \mathbf{A}$; análogamente, el **subespacio fila** se denota $\operatorname{Im} \mathbf{A}^t$.

Proposición 12 *Todo subespacio es tanto el subespacio imagen de alguna matriz, como el núcleo de otra matriz.*

Demostración: Dado un subespacio M, si extraemos un sistema de generadores de M, y construimos la matriz $\mathbf{A}$ que los contiene como columnas, se tiene que

$$\operatorname{Im}\mathbf{A} = L[C_1 (\mathbf{A}), .., C_n(\mathbf{A})] = M. \tag{2.43}$$

De esta forma M es el subespacio imagen de alguna matriz.

Por otro lado, M es también el núcleo de alguna matriz: basta con observar las ecuaciones implícitas de M, que se escriben matricialmente como $\mathbf{Bx} = \mathbf{0}$; en otras palabras, un vector $\mathbf{x}$ pertenece a M si y sólo si $\mathbf{x} \in \ker \mathbf{B}$. Por tanto, $M = \ker \mathbf{B}$. □

Observación: Una aplicación lineal es suprayectiva si y sólo si su subespacio imagen coincide con el espacio final, es decir:

Dada $f : \mathbb{K}^n \to \mathbb{K}^m$ lineal, f es suprayectiva $\Leftrightarrow \dim(\operatorname{Im} f) = m$.

2.2.1. Núcleo e inyectividad

Sabemos que toda aplicación lineal transforma el vector nulo del espacio inicial en el nulo del espacio final. Por otra parte, una aplicación lineal inyectiva es aquella para la que ningún vector del espacio final es imagen de más de un vector del inicial. Así pues, una aplicación lineal cuyo núcleo contenga algún vector no nulo no puede ser inyectiva.

Además, como vamos a ver a continuación, la condición de que el núcleo se reduzca al vector nulo caracteriza la inyectividad de las aplicaciones lineales.

Por simplicidad, denotamos por $\mathbf{0}$ el vector nulo tanto de $\mathbb{K}^n$ como de $\mathbb{K}^m$.

Teorema 4 *Sea* $f : \mathbb{K}^n \to \mathbb{K}^m$ *lineal;* f *es inyectiva si y sólo si* $\ker f = \{\mathbf{0}\}$.

Demostración:

$\boxed{\Rightarrow}$ Si $\ker f \neq \{\mathbf{0}\}$, $\exists \mathbf{u} \in \ker f$, $\mathbf{u} \neq \mathbf{0}$. Como $\mathbf{0} = f(\mathbf{0}) = f(\mathbf{u})$, f no puede ser inyectiva.

$\boxed{\Leftarrow}$ Supongamos que $\ker f = \{\mathbf{0}\}$ y que existen $\mathbf{u}, \mathbf{v} \in \mathbb{K}^n$ tales que $f(\mathbf{u}) = f(\mathbf{v})$. Entonces, por linealidad,

$$f(\mathbf{u} - \mathbf{v}) = f(\mathbf{u}) - f(\mathbf{v}) = \mathbf{0}, \tag{2.44}$$

luego $\mathbf{u} = \mathbf{v}$ y por tanto f es inyectiva. $\square$

2.3. Rango de una matriz

Definición 32 *Dada una matriz* $\mathbf{A} \in \mathbb{K}^{m \times n}$, *su* ***rango por columnas*** *es la dimensión del subespacio columna de* $\mathbf{A}$, *a saber,*

$$\dim(\operatorname{Im} \mathbf{A}) = \dim\left(L[C_1(\mathbf{A}), .., C_n(\mathbf{A})]\right). \tag{2.45}$$

Es decir, el rango por columnas es el número máximo de columnas de $\mathbf{A}$ linealmente independientes. Por otro lado, el **rango por filas** de $\mathbf{A}$ es la dimensión del subespacio fila de $\mathbf{A}$; es decir, el número máximo de filas de $\mathbf{A}$ linealmente independientes.

Proposición 13 *Para cualquier matriz* $\mathbf{A} \in \mathbb{K}^{m \times n}$, ***el rango por filas y el rango por columnas*** *de* $\mathbf{A}$ *coinciden.*

Demostración: Sea $\mathbf{A} \in \mathbb{K}^{m \times n}$ de rango r por columnas. Consideramos una matriz $\mathbf{C} \in \mathbb{K}^{m \times r}$ formada por r columnas de $\mathbf{A}$ linealmente independientes.

Por ser las columnas de $\mathbf{A}$ combinaciones lineales de las de $\mathbf{C}$, existirá una matriz $\mathbf{F}$ (cuyos elementos serán precisamente los coeficientes de esas combinaciones lineales) tal que $\mathbf{A} = \mathbf{CF}$.

Pero la igualdad anterior significa también que las filas de $\mathbf{A}$ son combinaciones lineales de las de $\mathbf{F}$; y esta matriz es de dimensiones $r \times n$ (tiene r filas), así que el máximo número de filas de $\mathbf{A}$ linealmente independientes no puede exceder de r, es decir, el rango por filas es menor o igual que el rango por columnas.

Dado que el rango por columnas de una matriz es el rango por filas de su transpuesta, aplicando el resultado anterior a la matriz $\mathbf{A}^t$ se concluye el resultado. $\square$

Observación: Este resultado que hemos demostrado permite que hablemos del **rango** de una matriz, sin especificar si es por filas o por columnas. A dicho número se le denota por $r(\mathbf{A})$.

Observación: En esta proposición también se ha demostrado que una matriz $\mathbf{A}$ de rango r puede factorizarse en producto de dos matrices $\mathbf{A} = \mathbf{CF}$, ambas de rango máximo r ya que el número de columnas de $\mathbf{C}$ y el número de filas de $\mathbf{F}$ son ambos r. En particular, toda matriz de rango uno puede escribirse como el producto de un vector columna por un vector fila.

Ejemplo: La matriz $\mathbf{A} = \begin{pmatrix} 1 & -1 & 2 \\ 2 & 0 & 2 \\ 1 & 1 & 0 \\ 0 & -1 & 1 \end{pmatrix}$ es de rango 2; se ve fácilmente que dos columnas cualesquiera son linealmente independientes y $\mathbf{C}_1 - \mathbf{C}_2 - \mathbf{C}_3 = \mathbf{0}$; por tanto puede tomarse como matriz $\mathbf{C}$ la formada por las dos primeras columnas:

$$\mathbf{C} = \begin{pmatrix} 1 & -1 \\ 2 & 0 \\ 1 & 1 \\ 0 & -1 \end{pmatrix}.$$

En cuanto a la matriz $\mathbf{F}$, sus columnas reflejan las combinaciones lineales que hay que hacer con las columnas de $\mathbf{C}$ para obtener las de $\mathbf{A}$:

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

Obviamente no hay unicidad, ya que la matriz $\mathbf{C}$ puede formarse con una base *cualquiera* de $\operatorname{Im}\mathbf{A}$.

Proposición 14

- *Si* $\mathbf{A}, \mathbf{B} \in \mathbb{K}^{m\times n}$ *entonces* $r(\mathbf{A}+\mathbf{B}) \le r(\mathbf{A}) + r(\mathbf{B})$.
- *Dadas dos matrices* $\mathbf{A}, \mathbf{B}$ *para las que el producto* $\mathbf{AB}$ *tenga sentido, se tiene que*

$$r(\mathbf{AB}) \le \min(r(\mathbf{A}), r(\mathbf{B})). \tag{2.46}$$

Demostración: Para la primera desigualdad, obsérvese que la columna j-ésima de $(\mathbf{A}+\mathbf{B})$ es $C_j(\mathbf{A}) + C_j(\mathbf{B})$. Por ello, el subespacio columna de $(\mathbf{A}+\mathbf{B})$ es

$$\operatorname{Im}(\mathbf{A}+\mathbf{B}) = \operatorname{Im}[C_1(\mathbf{A}+\mathbf{B}), \cdots, C_n(\mathbf{A}+\mathbf{B})] \subset \operatorname{Im}\mathbf{A} + \operatorname{Im}\mathbf{B}. \tag{2.47}$$

Así pues, $r(\mathbf{A}+\mathbf{B})$, que es la dimensión del subespacio de la izquierda, no puede superar la dimensión del subespacio $\mathrm{Im}\mathbf{A}+\mathrm{Im}\mathbf{B}$, que como mucho es $r(\mathbf{A})+r(\mathbf{B})$.

Para la segunda desigualdad, basta recordar que la columnas de $\mathbf{AB}$ son combinación lineal de las columnas de $\mathbf{A}$, luego $\mathrm{Im}(\mathbf{AB}) \subset \mathrm{Im}(\mathbf{A})$, de donde $r(\mathbf{AB}) \leq r(\mathbf{A})$. Por otro lado, las filas de $\mathbf{AB}$ son combinación lineal de las filas de $\mathbf{B}$, de donde $r(\mathbf{AB}) \leq r(\mathbf{B})$, concluyendo la demostración. □

Proposición 15 *Si* $\mathbf{A}, \mathbf{B}, \mathbf{C}$ *son matrices cualesquiera para las que los productos siguientes tengan sentido, entonces:*

- *Si* $\mathbf{A}$ *tiene rango máximo por columnas, entonces* $r(\mathbf{AB}) = r(\mathbf{B})$.
- *Si* $\mathbf{A}$ *tiene rango máximo por filas, entonces* $r(\mathbf{CA}) = r(\mathbf{C})$.

Teorema 5 *Si* $\mathbf{A}$ *es una matriz de* n *columnas, entonces*

$$\dim(\ker \mathbf{A}) + r(\mathbf{A}) = n. \tag{2.48}$$

Equivalentemente: si $f : \mathbb{K}^n \to \mathbb{K}^m$ *es lineal, entonces*

$$\dim(\ker f) + \dim(\mathrm{Im}\, f) = n. \tag{2.49}$$

Demostración *(no exigible):* Como $\mathbf{A}$ posee n columnas, $\ker \mathbf{A}$ es un subespacio de $\mathbb{K}^n$. Tomemos una base de $\ker \mathbf{A}$, $(\mathbf{u}_1, ..., \mathbf{u}_k)$, y completémosla hasta obtener una base $(\mathbf{u}_1, ..., \mathbf{u}_k, \mathbf{u}_{k+1}, ..., \mathbf{u}_n)$ de $\mathbb{K}^n$.

Por la propiedad de linealidad, es fácil ver que $\mathrm{Im}\mathbf{A} = L[\mathbf{Au}_1, ..., \mathbf{Au}_k, \mathbf{Au}_{k+1}, ..., \mathbf{Au}_n]$ $= L[\mathbf{Au}_{k+1}, ..., \mathbf{Au}_n]$ ya que $\mathbf{Au}_1 = ... = \mathbf{Au}_k = \mathbf{0}$ por definición. Así, $\{\mathbf{Au}_{k+1}, ..., \mathbf{Au}_n\}$ es un sistema de generadores de $\mathrm{Im}\mathbf{A}$ constituido por $n-k$ vectores. Si demostramos que son linealmente independientes, entonces formarán base de $\mathrm{Im}\mathbf{A}$, y podremos garantizar el resultado buscado:

$$r(\mathbf{A}) = \dim(\mathrm{Im}\mathbf{A}) = n - k = n - \dim(\ker \mathbf{A}). \tag{2.50}$$

Basta, pues, demostrar que $\mathbf{Au}_{k+1}, ..., \mathbf{Au}_n$ son linealmente independientes. Apliquemos la definición de independencia lineal: se considera la igualdad $\alpha_1\mathbf{Au}_{k+1} + ... + \alpha_{n-k}\mathbf{Au}_n = \mathbf{0}$, y el objetivo es demostrar que $\alpha_1 = ... = \alpha_{n-k} = 0$. Escribimos esta igualdad como

$$\mathbf{A}(\alpha_1\mathbf{u}_{k+1} + ... + \alpha_{n-k}\mathbf{u}_n) = \mathbf{0} \tag{2.51}$$

lo cual significa que el vector $\alpha_1\mathbf{u}_{k+1} + ... + \alpha_{n-k}\mathbf{u}_n \in \ker \mathbf{A}$. Es decir $\alpha_1\mathbf{u}_{k+1} + ... + \alpha_{n-k}\mathbf{u}_n$ es combinación lineal de $\mathbf{u}_1, ..., \mathbf{u}_k$: $\alpha_1\mathbf{u}_{k+1} + ... + \alpha_{n-k}\mathbf{u}_n = \beta_1\mathbf{u}_1 + ... + \beta_k\mathbf{u}_k$. O, equivalentemente $\beta_1\mathbf{u}_1 + ... + \beta_k\mathbf{u}_k - \alpha_1\mathbf{u}_{k+1} + ... - \alpha_{n-k}\mathbf{u}_n = \mathbf{0}$ pero como $(\mathbf{u}_1, ..., \mathbf{u}_k, \mathbf{u}_{k+1}, ..., \mathbf{u}_n)$ son independientes, todos los coeficientes deben ser 0, en particular se tiene que $\alpha_1 = ... = \alpha_{n-k} = 0$. □

Del teorema anterior se desprende el siguiente resultado: para matrices cuadradas, el **rango máximo** caracteriza su **invertibilidad**:

Corolario: Sea $\mathbf{A}$ una matriz cuadrada de orden n. Entonces las siguientes afirmaciones son equivalentes:

$$\mathbf{A} \text{ es invertible } \iff r(\mathbf{A}) = n \iff \ker \mathbf{A} = \{\mathbf{0}\}. \tag{2.52}$$

Observación: Sea $f : \mathbb{K}^n \to \mathbb{K}^m$ lineal.

- Si $m < n$, f no puede ser inyectiva.
- Si $m > n$, f no puede ser suprayectiva.
- Si $m = n$, f es inyectiva $\Leftrightarrow$ f es suprayectiva.

2.3.1. Cálculo del rango mediante operaciones elementales

Operaciones elementales de reducción gaussiana

Definición 33 *Dada una matriz $\mathbf{A} \in \mathbb{K}^{m\times n}$, se definen las siguientes* ***operaciones elementales*** *por filas (columnas) sobre la matriz $\mathbf{A}$:*

1) Suma de una fila (columna) de la matriz, multiplicada por un escalar, a otra fila (columna) distinta.

2) Multiplicación de una fila (columna) de la matriz por un escalar no nulo.

3) Intercambio de dos filas (columnas).

Definición 34 *Se llama* ***matriz elemental*** *por filas (columnas) a una matriz cuadrada, resultado de efectuarle una operación elemental por filas (columnas) a la matriz identidad.*

Ejemplos: La matriz

$$\mathbf{E}_1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -2 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

es una matriz elemental por filas, resultante de restarle a la tercera fila de la identidad el doble de la primera.

Asimismo, es una matriz elemental por columnas, resultante de restarle a la primera columna de la identidad el doble de la tercera.

La matriz

$$\mathbf{E}_2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

es una matriz elemental por filas, resultante de intercambiar las filas 2 y 4 de la identidad.

También es matriz elemental por columnas, resultante de intercambiar las columnas 2 y 4 de la identidad.

Observación: Toda matriz elemental por filas lo es por columnas y viceversa (aunque no necesariamente asociadas a la misma operación por filas que por columnas).

Observación: Toda matriz elemental es invertible, y su inversa es la matriz elemental asociada a la operación inversa (la operación inversa de la suma es la resta, la operación inversa de multiplicar una fila o columna por un número no nulo es dividir por él, y la operación inversa de un intercambio de filas o columnas es ese mismo intercambio). Las matrices inversas de las de los ejemplos son:

$$\mathbf{E}_1^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 2 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \ \mathbf{E}_2 = \mathbf{E}_2^{-1}.$$

Algoritmo de cálculo del rango mediante operaciones de reducción gaussiana

Dada una matriz $\mathbf{A} \in \mathbb{K}^{m \times n}$, realizar operaciones elementales sobre sus filas equivale a premultiplicar por sucesivas matrices elementales:

$$\mathbf{E}_k \cdots \mathbf{E}_2 \mathbf{E}_1 \mathbf{A}. \tag{2.53}$$

Cada una de estas operaciones deja invariante el rango. Si mediante este proceso llegamos a una *matriz triangular* (cuyo rango se deduce inmediatamente) habremos conseguido calcular $r(\mathbf{A})$.

Igualmente puede procederse efectuando operaciones elementales sobre las columnas (posmultiplicación por matrices elementales), e incluso combinando ambos métodos.

2.3.2. Algoritmo de Gauss-Jordan

Permite calcular la inversa de una matriz usando operaciones elementales. Sea $\mathbf{A}$ una matriz cuadrada e invertible. Si, de igual modo que antes, mediante operaciones elementales sobre sus **filas**, conseguimos llegar a la matriz identidad, tendremos:

$$\mathbf{E}_k \cdots \mathbf{E}_2 \mathbf{E}_1 \mathbf{A} = \mathbf{I} \tag{2.54}$$

Es decir, que $\mathbf{A}^{-1} = \mathbf{E}_k \cdots \mathbf{E}_2 \mathbf{E}_1 \mathbf{I}$, que es el resultado de efectuar, sobre la matriz identidad, las mismas operaciones efectuadas sobre $\mathbf{A}$, lo cual nos permite calcular $\mathbf{A}^{-1}$ fácilmente.

También se puede proceder de igual modo por **columnas**, pero **sin mezclar** en ningún caso **ambos métodos**.

Ejemplo: Vamos a calcular la inversa de la matriz

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 1 \\ 2 & 4 & 3 \\ 1 & 1 & 1 \end{pmatrix}$$

aplicando este algoritmo; consideramos la matriz $(\mathbf{A} \mid \mathbf{I})$ y mediante transformaciones elementales de filas llegaremos a $(\mathbf{I} \mid \mathbf{A}^{-1})$

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 1 & 1 & 0 & 0 \\ 2 & 4 & 3 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 \end{array}\right) \overset{\substack{F_2-2F_1 \\ F_3-F_1}}{\longrightarrow} \left(\begin{array}{ccc|ccc} 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 4 & 1 & -2 & 1 & 0 \\ 0 & 1 & 0 & -1 & 0 & 1 \end{array}\right) \overset{F_2 \leftrightarrow F_3}{\longrightarrow}$$

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & 1 \\ 0 & 4 & 1 & -2 & 1 & 0 \end{array}\right) \overset{F_3-4F_2}{\longrightarrow} \left(\begin{array}{ccc|ccc} 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & 1 \\ 0 & 0 & 1 & 2 & 1 & -4 \end{array}\right) \overset{F_1-F_3}{\longrightarrow} \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & -1 & -1 & 4 \\ 0 & 1 & 0 & -1 & 0 & 1 \\ 0 & 0 & 1 & 2 & 1 & -4 \end{array}\right)$$

Concluimos que la inversa que buscábamos es

$$\mathbf{A}^{-1} = \begin{pmatrix} -1 & -1 & 4 \\ -1 & 0 & 1 \\ 2 & 1 & -4 \end{pmatrix}.$$

2.3.3. Matriz de cambio de base

Si $\mathcal{B} = (\mathbf{b}_1, \ldots, \mathbf{b}_n)$ y $\mathcal{V} = (\mathbf{v}_1, \ldots, \mathbf{v}_n)$ son dos bases de $\mathbb{K}^n$, entonces los vectores de cada una de ellas se pueden poner como combinaciones lineales de los de la otra, de la siguiente forma:

$$\begin{aligned} \mathbf{v}_1 &= p_{11}\mathbf{b}_1 + p_{21}\mathbf{b}_2 + \ldots + p_{n1}\mathbf{b}_n \\ \mathbf{v}_2 &= p_{12}\mathbf{b}_1 + p_{22}\mathbf{b}_2 + \ldots + p_{n2}\mathbf{b}_n \\ &\vdots \\ \mathbf{v}_n &= p_{1n}\mathbf{b}_1 + p_{2n}\mathbf{b}_2 + \ldots + p_{nn}\mathbf{b}_n. \end{aligned} \tag{2.55}$$

Estas expresiones dan lugar a la siguiente definición:

Definición 35 (matriz de cambio de base) *Se llama matriz de cambio de base (o matriz de paso) de $\mathcal{B}$ a $\mathcal{V}$ a la matriz*

$$\mathbf{P} = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{pmatrix}. \tag{2.56}$$

Es decir, es la matriz cuya columna j-ésima contiene las coordenadas del vector $\mathbf{v}_j$ en la base $\mathcal{B}$.

Observación: La matriz de paso de $\mathcal{B}$ a $\mathcal{V}$ es invertible y su inversa es la que hace el cambio inverso, es decir, la matriz de paso de $\mathcal{V}$ a $\mathcal{B}$.

Observación: Cada vector $\mathbf{u} \in \mathbb{K}^n$ lleva asociado un vector de coordenadas respecto a $\mathcal{B}$ (denotado $[\mathbf{u}]^{\mathcal{B}}$) y un vector de coordenadas respecto a $\mathcal{V}$ (denotado $[\mathbf{u}]^{\mathcal{V}}$). Pues bien, la matriz $\mathbf{P}$ de paso de $\mathcal{B}$ a $\mathcal{V}$ transforma $[\mathbf{u}]^{\mathcal{V}}$ en $[\mathbf{u}]^{\mathcal{B}}$ (*¡cuidado con el orden de los vectores!*). Matricialmente,

$$[\mathbf{u}]^{\mathcal{B}} = \mathbf{P}[\mathbf{u}]^{\mathcal{V}}. \tag{2.57}$$

Para demostrar esta última igualdad, simplemente reescribimos la ecuación (2.55) matricialmente como $\mathbf{V} = \mathbf{BP}$, donde $\mathbf{V}$ es la matriz cuyas n columnas son los vectores $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ y $\mathbf{B}$ es la matriz de columnas $\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n$. A partir de ello, todo vector $\mathbf{u} \in \mathbb{K}^n$ puede escribirse como

$$\mathbf{B}[\mathbf{u}]^{\mathcal{B}} = \mathbf{u} = \mathbf{V}[\mathbf{u}]^{\mathcal{V}} = \mathbf{BP}[\mathbf{u}]^{\mathcal{V}}. \tag{2.58}$$

Igualamos el primer y el último término, y multiplicamos por $\mathbf{B}^{-1}$ (que existe pues $\mathbf{B}$ es cuadrada de columnas independientes); de esta forma se llega a la identidad (2.57). Nótese que $\mathbf{P} = \mathbf{B}^{-\mathbf{1}}\mathbf{V}$, lo cual también permite calcular la matriz $\mathbf{P}$ de cambio de base.

2.4. Sistemas de ecuaciones lineales

Se trata de encontrar, si existen, todas las n-uplas de valores $x_1, x_2, \ldots, x_n$, en el cuerpo $\mathbb{K}$, que satisfagan simultáneamente las m ecuaciones

$$\left.\begin{array}{rcl} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n & = & b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n & = & b_2 \\ & & \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n & = & b_m \end{array}\right\} \tag{2.59}$$

en donde a_{ij} y b_i son elementos de $\mathbb{K}$ prefijados. Las ecuaciones anteriores (ecuaciones lineales) constituyen un **sistema lineal de m ecuaciones y n incógnitas** $x_1, \ldots, x_n$.

Llamando

$$\mathbf{A} = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \quad \mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \quad \mathbf{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix} \tag{2.60}$$

el sistema anterior puede expresarse matricialmente como

$$\mathbf{Ax} = \mathbf{b}. \tag{2.61}$$

Un vector $\mathbf{u} \in \mathbb{K}^n$ que satisfaga que $\mathbf{Au} = \mathbf{b}$ se denomina **vector solución** del sistema.

La matriz $\mathbf{A}$ se denomina **matriz de coeficientes**; el vector $\mathbf{b}$, **vector de términos independientes**; la yuxtaposición de ambos,

$$(\mathbf{A}|\mathbf{b}) \in \mathbb{K}^{m\times(n+1)} \tag{2.62}$$

se llama **matriz ampliada**, y contiene todos los datos que determinan el sistema lineal y, por tanto, sus soluciones.

Definición 36 *Un sistema lineal se dice **compatible** si admite alguna solución; si, además, es única, se llama compatible **determinado**; si admite más de una solución, compatible **indeterminado**; si no admite ninguna solución, **incompatible**.*

2.4.1. Estructura de las soluciones

Un sistema lineal se dice **homogéneo** si sus **términos independientes** son todos **nulos**. En otras palabras, si el sistema es de la forma $\mathbf{Ax} = \mathbf{0}$.

Observación: Todo sistema homogéneo admite como solución el vector nulo de $\mathbb{K}^n$, que constituye la llamada *solución trivial*. Pero el conjunto de todas las soluciones del sistema homogéneo $\mathbf{Ax} = \mathbf{0}$ es precisamente el núcleo de $\mathbf{A}$ (que se denota $\ker \mathbf{A}$, ya definido en la expresión (2.38)).

Para cada sistema lineal $\mathbf{Ax} = \mathbf{b}$, que llamaremos **sistema completo**, su **sistema homogéneo asociado** es $\mathbf{Ax} = \mathbf{0}$.

Se tiene que cualesquiera dos soluciones del sistema completo difieren en una solución del sistema homogéneo; en efecto: sean $\mathbf{u}_1, \mathbf{u}_2$ soluciones de $\mathbf{Ax} = \mathbf{b}$; entonces

$$\mathbf{A}(\mathbf{u}_1 - \mathbf{u}_2) = \mathbf{Au}_1 - \mathbf{Au}_2 = \mathbf{b} - \mathbf{b} = \mathbf{0}, \tag{2.63}$$

luego $\mathbf{u}_1 - \mathbf{u}_2$ es solución del sistema homogéneo. Por otro lado, la suma de una solución particular del sistema completo y una solución del sistema homogéneo da como resultado una solución del sistema completo (en efecto, si $\mathbf{Au} = \mathbf{b}$ y $\mathbf{Ax} = \mathbf{0}$, entonces $\mathbf{A}(\mathbf{u} + \mathbf{x}) = \mathbf{0} + \mathbf{b} = \mathbf{b}$). En resumen: fijada una solución particular de $\mathbf{Ax} = \mathbf{b}$, cada solución de ese sistema puede escribirse como suma de esa solución particular más una determinada solución del sistema homogéneo asociado.

Proposición 16 *Si el sistema $\mathbf{Ax} = \mathbf{b}$ es compatible y $r(\mathbf{A}) = r$, entonces todas sus soluciones vienen dadas en función de $n - r$ parámetros independientes (es decir, tantos como el número de incógnitas menos el rango de la matriz de coeficientes). Más aún, dichas soluciones se escriben como*

$$\mathbf{x} = \mathbf{x}_0 + \alpha_1\mathbf{u}_1 + \cdots + \alpha_{n-r}\mathbf{u}_{n-r} \tag{2.64}$$

donde $\mathbf{x}_0$ es una solución particular del sistema, $(\mathbf{u}_1, \cdots, \mathbf{u}_{n-r})$ es base de $\ker \mathbf{A}$, y $\alpha_1, \cdots, \alpha_{n-r}$ son los parámetros mencionados.

Demostración: Según la observación anterior, fijada una solución particular $\mathbf{x}_0$ del sistema $\mathbf{Ax} = \mathbf{b}$, cualquier vector $\mathbf{x}$ es solución del sistema si y sólo si $\mathbf{x} - \mathbf{x}_0 \in \ker \mathbf{A}$. Por otro lado, la igualdad (2.48) afirma que $\dim(\ker \mathbf{A}) = n - r$, luego cualquier base de $\ker \mathbf{A}$ debe poseer $n - r$ vectores independientes. Sea $(\mathbf{u}_1, \cdots, \mathbf{u}_{n-r})$ una base de $\ker \mathbf{A}$; el vector $\mathbf{x} - \mathbf{x}_0 \in \ker \mathbf{A}$ si y sólo si existen escalares $\alpha_1, \cdots, \alpha_{n-r}$ tales que

$$\mathbf{x} - \mathbf{x}_0 = \alpha_1\mathbf{u}_1 + \cdots + \alpha_{n-r}\mathbf{u}_{n-r} \tag{2.65}$$

y esta igualdad es equivalente a la expresión (2.64).

2.4.2. El teorema de Rouché-Frobenius

Recordando la igualdad (2.41), se tiene que todo sistema lineal $\mathbf{Ax} = \mathbf{b}$ puede reescribirse como

$$x_1C_1(\mathbf{A}) + x_2C_2(\mathbf{A}) + \ldots + x_nC_n(\mathbf{A}) = \mathbf{b}. \tag{2.66}$$

En otras palabras, el vector de términos independientes debe ser combinación lineal de las n columnas de la matriz $\mathbf{A}$, cuyos coeficientes son precisamente las incógnitas.

Así pues, el sistema admitirá solución cuando, y solo cuando, $\mathbf{b}$ sea combinación lineal de las columnas de $\mathbf{A}$, lo cual se expresa en el siguiente

Teorema 6 (Rouché-Frobenius) *El sistema lineal $\mathbf{Ax} = \mathbf{b}$ es compatible si y solo si $r(\mathbf{A}|\mathbf{b}) = r(\mathbf{A})$.*

Además, si el sistema lineal $\mathbf{Ax} = \mathbf{b}$ es de n incógnitas, y llamamos r al rango de $\mathbf{A}$, caben las siguientes posibilidades:

- $r(\mathbf{A}|\mathbf{b}) = r + 1 \iff$ *sistema incompatible.*
- $r(\mathbf{A}|\mathbf{b}) = r \iff$ *sistema compatible. Y en ese caso,*

- $r = n \iff$ *compatible determinado: solución única.*
- $r < n \iff$ *compatible indeterminado; la solución general depende de $n - r$ parámetros.*

Demostración: La primera parte del teorema (condición de compatibilidad) equivale a la observación previa a dicho teorema: **b** es combinación lineal de las columnas de **A** si y sólo si $r(\mathbf{A}|\mathbf{b}) = r(\mathbf{A}) = r$. En ese caso el sistema es compatible; en cualquier otro caso ($r(\mathbf{A}|\mathbf{b}) > r$) es incompatible. En los casos compatibles, las soluciones vienen dadas en función de $n - r$ parámetros (es lo que afirma la Proposición 5). En particular, el caso $r = n$ indica que la solución viene dada por $n - r = 0$ parámetros; ello quiere decir que la solución es única. (Obsérvese que esto se verifica cuando $r(\mathbf{A}) = n$, es decir, cuando y sólo cuando todas las columnas de **A** son independientes). □

2.4.3. Resolución de sistemas lineales por reducción gaussiana

Observación: Dado un sistema lineal $\mathbf{Ax} = \mathbf{b}$, de matriz de coeficientes $\mathbf{A} \in \mathbb{K}^{m\times n}$ y una matriz $\mathbf{Q} \in \mathbb{K}^{m\times m}$ **invertible**, el conjunto de soluciones del sistema anterior coincide con el de

$$\mathbf{QAx} = \mathbf{Qb} \tag{2.67}$$

y se dice que este último sistema es *equivalente* al $\mathbf{Ax} = \mathbf{b}$.

Esta observación permite pasar de un sistema inicial a otro que posee las **mismas soluciones** pero que quizá es más fácilmente resoluble; el objetivo es que la nueva matriz de coeficientes **QA** sea matriz triangular o escalonada. ¿Cómo conseguir que **QA** tenga forma escalonada? Recuérdese el concepto de operación elemental por filas, equivalente a premultiplicación por la correspondiente matriz elemental, que siempre es invertible. Pues bien, aplicando operaciones elementales de reducción gaussiana convertimos **A** en una matriz escalonada $\mathbf{E}_k \cdots \mathbf{E}_2\mathbf{E}_1\mathbf{A}$; esta nueva matriz es **QA**, donde $\mathbf{Q} = \mathbf{E}_k \cdots \mathbf{E}_2\mathbf{E}_1$.

Obsérvese que la matriz ampliada del sistema inicial es $(\mathbf{A}|\mathbf{b})$ y la del nuevo sistema es $(\mathbf{QA}|\mathbf{Qb})$. Ello significa que, aplicando dichas **operaciones elementales sobre las filas completas** de la matriz ampliada, obtenemos un sistema equivalente y fácilmente resoluble. En esto consiste el método de reducción gaussiana.

Veamos este procedimiento con más detalle: a la hora de resolver un sistema lineal, realizamos operaciones elementales **por filas** sobre la matriz ampliada $(\mathbf{A}|\mathbf{b})$, lo que equivale a premultiplicar **A** y **b** por una misma matriz invertible, obteniendo un sistema equivalente.

Nuestro propósito será llegar a una configuración de la forma (ejemplo de tamaño 4×6)

$$\left(\begin{array}{cccccc|c} \blacksquare & * & * & * & * & * & * \\ 0 & \blacksquare & * & * & * & * & * \\ 0 & 0 & \blacksquare & * & * & * & * \\ 0 & 0 & 0 & \blacksquare & * & * & * \end{array}\right) \tag{2.68}$$

en donde el símbolo $\blacksquare$ denota elementos no nulos, y el símbolo $*$ elementos cualesquiera.

En este ejemplo, el sistema será compatible indeterminado, dependiendo la solución general de $6 - 4 = 2$ parámetros; podemos dejar libres dos incógnitas, por ejemplo $x_5 = \lambda$, $x_6 = \mu$ $\forall \lambda, \mu \in \mathbb{K}$, despejando *de abajo a arriba* las cuatro restantes en función de éstas.

Si se llega a una situación de la forma

$$\left(\begin{array}{cccccc|c} \blacksquare & * & * & * & * & * & * \\ 0 & \blacksquare & * & * & * & * & * \\ 0 & 0 & \blacksquare & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & \blacksquare \end{array}\right), \tag{2.69}$$

entonces el sistema será incompatible ($r(\mathbf{A}) = 3$ y $r(\mathbf{A}|\mathbf{b}) = 4$).

Para llegar a estas configuraciones de **matrices escalonadas**, se sigue el siguiente algoritmo:

- Si es $a_{11} \neq 0$, se efectúa para cada fila $2, 3, \ldots, m$ la operación

  $$\mathbf{F}_i' = \mathbf{F}_i - \frac{a_{i1}}{a_{11}} \mathbf{F}_1, \tag{2.70}$$

  es decir, a cada fila le restamos el adecuado múltiplo de la primera para hacer ceros en la primera columna, obteniendo

  $$\left(\begin{array}{cccc|c} \blacksquare & * & \cdots & * & * \\ 0 & * & \cdots & * & * \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & * & \cdots & * & * \end{array}\right). \tag{2.71}$$

  Si es $a_{11} = 0$, se escoge una fila k tal que sea $a_{k1} \neq 0$, y se intercambian las filas 1 y k; a continuación, se procede como se explicó antes. Si todos los elementos de la primera columna son nulos, el valor x_1 de la primera incógnita es arbitrario y se pasa al siguiente paso.

El siguiente paso es proceder del mismo modo con la columna 2, para las ecuaciones $3, \ldots, n$. Si fuesen nulos todos los a'_{2k} , $k = 2, \ldots, n$, se tendría ya hecho el trabajo sobre la columna 2 y se pasaría a operar con la 3. En caso contrario:

Si es $a'_{22} \neq 0$, se hacen las operaciones

$$\mathbf{F}_i'' = \mathbf{F}_i' - \frac{a'_{i2}}{a'_{22}} \mathbf{F}_2' \ , \ \ i = 3, 4, \ldots, m, \tag{2.72}$$

y si es $a'_{22} = 0$, se intercambia la fila 2 con alguna posterior cuyo elemento a'_{k2} sea no nulo, haciéndose a continuación las operaciones dichas.

Reiterando el proceso, se llega a la matriz escalonada.

En cada paso no trivial hay que dividir por un elemento diagonal $a_{ii}^{(i-1)}$ que se denominará ***pivote*** o, eventualmente, intercambiar dos ecuaciones para obtener un pivote no nulo. Así pues, en muchos casos es imprescindible efectuar operaciones de intercambio de filas. Por ejemplo, si tuviéramos

$$\left(\begin{array}{cccccc|c} \blacksquare & * & * & * & * & * & * \\ 0 & 0 & \blacksquare & * & * & * & * \\ 0 & \blacksquare & * & * & * & * & * \\ 0 & 0 & * & * & * & * & * \end{array}\right) \tag{2.73}$$

sería imposible anular el elemento a_{32}, por lo que se procedería a intercambiar las filas 2 y 3.

En ocasiones puede llegarse a configuraciones como

$$\left(\begin{array}{cccccc|c} \blacksquare & * & * & * & * & * & * \\ 0 & \blacksquare & * & * & * & * & * \\ 0 & 0 & 0 & \blacksquare & * & * & * \\ 0 & 0 & 0 & 0 & 0 & \blacksquare & * \end{array}\right) \tag{2.74}$$

que nos obligan a tomar como variables libres algunas de las incógnitas intermedias (en el ejemplo mostrado, la 3ª y la 5ª).

2.5. Ejercicios

2.1.– Una matriz $\mathbf{A} \in \mathbb{R}^{3\times 2}$ verifica

$$\mathbf{A}\begin{pmatrix} 3 \\ 5 \end{pmatrix} = \begin{pmatrix} 2 \\ -4 \\ 1 \end{pmatrix} \quad \mathbf{A}\begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} -5 \\ 3 \\ 4 \end{pmatrix}.$$

Calcular la imagen por $\mathbf{A}$ del vector $\begin{pmatrix} 2 \\ -1 \end{pmatrix}$. Hallar asimismo $\mathbf{A}$.

2.2.– Hállese la matriz $\mathbf{A}$ que cumple:

$$\mathbf{A}\begin{pmatrix} 3 \\ 2 \end{pmatrix} = \begin{pmatrix} 9 \\ 11 \end{pmatrix}, \quad \mathbf{A}\begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 5 \\ 3 \end{pmatrix}.$$

2.3.– Para cada número real a se considera la matriz

$$\mathbf{A} = \begin{pmatrix} 1 & a & -1 & 2 \\ 2 & -1 & 3 & 5 \\ 1 & 10 & -6 & 1 \end{pmatrix}.$$

Hallar unas ecuaciones paramétricas del núcleo de $\mathbf{A}$ y una base del subespacio imagen de $\mathbf{A}$.

2.4.– Sea $\mathbf{A}$ una matriz que verifica

$$\mathbf{A}\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \quad \mathbf{A}\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \\ -2 \end{pmatrix}, \quad \mathbf{A}\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix}.$$

Se pide:

1. Escribir la matriz $\mathbf{A}$.
2. Determinar unas bases de $\ker \mathbf{A}$ e $\operatorname{Im} \mathbf{A}$, así como unas ecuaciones cartesianas de los mismos.
3. Dado el subespacio $M = L[(1,2,0),(2,1,1)]$ de $\mathbb{R}^3$, determinar una base del subespacio de los transformados de los vectores de M por la matriz $\mathbf{A}$; es decir, de $\{\mathbf{Ax},\ \mathbf{x} \in M\}$.

2.5.– De una matriz $\mathbf{A} \in \mathbb{C}^{4\times 4}$ se sabe que

$$\mathbf{A}\begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} = \begin{pmatrix} 4 \\ 3 \\ 2 \\ 1 \end{pmatrix}$$

y que $\operatorname{Im} \mathbf{A} \subsetneq \ker \mathbf{A}$.

1. Calcúlese el rango de $\mathbf{A}$.

2. Calcúlese el vector suma de las columnas de $\mathbf{A}$.

2.6.– Calcular el rango de la matriz

$$\mathbf{M} = \begin{pmatrix} a & b & b & \cdots & b & b \\ b & a & b & \cdots & b & b \\ b & b & a & \cdots & b & b \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ b & b & b & \cdots & a & b \\ b & b & b & \cdots & b & a \end{pmatrix}$$

y deducir para qué valores de a y b es invertible.

2.7.– Calcular, empleando el método de Gauss-Jordan, las inversas de las matrices

$$\begin{pmatrix} 1 & 4 & 1 \\ 2 & 9 & 2 \\ 5 & -2 & 6 \end{pmatrix}; \qquad \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & i & -1 & -i \\ 1 & -1 & 1 & -1 \\ 1 & -i & -1 & i \end{pmatrix}.$$

2.8.– Hallar, usando el método de Gauss-Jordan, la inversa de la matriz

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

Probar que una matriz triangular es invertible si y sólo si todos los elementos de su diagonal principal son distintos de cero y que, en ese caso, su inversa es una matriz triangular del mismo tipo.

2.9.– Calcular, mediante el método de Gauss-Jordan, la inversa de la matriz

$$\begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 & 0 \\ 0 & 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & 0 & -1 & 2 \end{pmatrix}.$$

Generalizar el resultado obtenido a una matriz de orden n que tenga la misma estructura. Este ejemplo pone de manifiesto que la inversa de una matriz dispersa no tiene por qué ser ella misma dispersa.

2.10.– Sea $\mathbf{A}$ una matriz compleja. Demostrar que $\ker \mathbf{A}^h\mathbf{A} = \ker \mathbf{A}$ y deducir la igualdad $r\left(\mathbf{A}^h\mathbf{A}\right) = r(\mathbf{A})$. Probar también que $\mathbf{A}^h\mathbf{A} = \mathbf{O}$ si, y solamente si, $\mathbf{A} = \mathbf{O}$.

2.11.– Resolver por el método de Gauss los siguientes sistemas de ecuaciones lineales:

$$\left\{\begin{aligned} 2x - y + 3z + u + 4v &= 2 \\ 4x + y + 8z + u + 12v &= 13 \\ -6x + 15y + 4z - 5u + v &= 27 \end{aligned}\right. \qquad \left\{\begin{aligned} 2x + 4y + 8z &= 6 \\ 4x + y + 9z &= 12 \\ x + 2y + 4z &= 3 \\ 5x - 4y + 6z &= 15 \\ 3x + 5y + 11z &= 9 \end{aligned}\right.$$

2.12.– Encontrar los vértices de un triángulo cuyos lados tienen por puntos medios los puntos de coordenadas

$$(4,1), (2,3), \ (6,2).$$

¿Quedan determinados los vértices de un cuadrilátero si se conocen los puntos medios de sus lados? Generalizar estos resultados al caso de un polígono con un número cualquiera de lados.

2.13.– Discutir según los valores de los parámetros a y b los siguientes sistemas de ecuaciones

$$\left\{\begin{aligned} ax + y + z &= 1 \\ x + ay + z &= b \\ x + y + az &= b^2 \end{aligned}\right. \qquad \left\{\begin{aligned} ax + y + bz &= 1 \\ x + ay + bz &= 1 \\ x + y + abz &= b \end{aligned}\right.$$

2.5.1. Cuestiones

2.14.– La matriz de cambio de la base B a la base V es $\begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 2 & 1 \end{pmatrix}$. Si un vector tiene coordenadas $(1,2,0)$ respecto a B, ¿qué coordenadas tiene respecto a V?:

○ (A) $(3,2,0)$

○ (B) $(1,3,5)$

○ (C) $(1,1,-3)$

○ (D) $(-1,2,0)$

2.15.– Indicar si las siguientes afirmaciones son verdaderas (V) o falsas (F).

	V	F
El producto de dos matrices invertibles es una matriz invertible		
El producto de dos matrices simétricas es una matriz simétrica		
Si $\mathbf{A}$ transforma vectores linealmente independientes en vectores linealmente independientes, entonces su rango por columnas es máximo.		
Si $\ker(\mathbf{A}) = \{\mathbf{0}\}$ entonces $\mathbf{A}$ transforma vectores linealmente independientes en vectores linealmente independientes.		
Si $\mathbf{A}$ es cuadrada, entonces el rango de $\mathbf{A}^2$ es mayor o igual que el rango de $\mathbf{A}$.		

2.16.– Sean $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ matrices no necesariamente cuadradas. Indicar si las siguientes afirmaciones son verdaderas (V) o falsas (F).

	V	F
Si $\mathbf{AB} = \mathbf{O}$, entonces o bien $\mathbf{A} = \mathbf{O}$ o bien $\mathbf{B} = \mathbf{O}$.		
Si $\mathbf{A}^2 = \mathbf{O}$, entonces $\mathbf{A}$ es necesariamente nula.		
Si $\mathbf{AB} = \mathbf{AC}$ entonces $\mathbf{B} = \mathbf{C}$.		
Si $\mathbf{AB} = \mathbf{AC}$ y $\mathbf{A}$ es cuadrada invertible, entonces $\mathbf{B} = \mathbf{C}$.		
Si $\mathbf{A}$ es rectangular, de columnas independientes, y $\mathbf{AB} = \mathbf{AC}$, entonces $\mathbf{B} = \mathbf{C}$.		
Si $\mathbf{A}^2 = \mathbf{B}^2$ entonces $\mathbf{A} = \pm\mathbf{B}$.		

2.17.– Indicar si las siguientes afirmaciones son verdaderas (V) o falsas (F).

	V	F
Si $\mathbf{Ax} = \mathbf{0}$ tiene solución única, entonces $\ker(\mathbf{A}) = \{\mathbf{0}\}$		
Si el sistema $\mathbf{Ax} = \mathbf{b}$ tiene solución única, entonces el sistema $\mathbf{Ax} = \mathbf{0}$ tiene como única solución la trivial.		
Si el sistema $\mathbf{Ax} = \mathbf{0}$ tiene solución única, entonces el sistema $\mathbf{Ax} = \mathbf{b}$ tiene solución única.		
Si $\mathbf{Ax} = \mathbf{b}$ tiene solución única, entonces $\mathbf{Ax} = \mathbf{v}$ tiene solución única.		
Si $\mathbf{b}, \mathbf{v} \in \mathrm{Im}(\mathbf{A})$ entonces $\mathbf{Ax} = \mathbf{b}$ es determinado $\Leftrightarrow$ $\mathbf{Ax} = \mathbf{v}$ es determinado.		

Capítulo 3

Producto escalar y ortogonalidad

3.1 Producto escalar y norma asociada en $\mathbb{R}^n$. Desigualdades de Cauchy-Schwarz y triangular.

3.2 Ortogonalidad. El suplementario ortogonal. El teorema de la proyección ortogonal. Familias ortogonales. Bases ortonormales. Matrices ortogonales. El método de ortonormalización de Gram–Schmidt. Factorización **QR**.

3.3 Extensión a $\mathbb{C}^n$.

3.4 Ejercicios. Cuestiones.

Introducción

En los temas anteriores hemos atendido a propiedades *cualitativas* de los vectores: pueden sumarse o multiplicarse por un número, etc. Ahora vamos a ver otras propiedades de tipo *cuantitativo*: qué entendemos por longitud de un vector, cómo medir el ángulo que forman dos vectores, cómo proyectar un vector (sobre un subespacio), etc. En el plano muchas de estas cuestiones pueden resolverse de forma bastante elemental, "con regla y compás". Queremos extender de manera natural las ideas intuitivas de la geometría elemental tanto a $\mathbb{R}^n$ como a $\mathbb{C}^n$.

Nota: Por razones de notación identificamos los vectores de $\mathbb{R}^n$ y $\mathbb{C}^n$ con vectores columna.

3.1. Producto escalar y norma

Definición 37 (producto escalar natural en $\mathbb{R}^n$) *Dados dos vectores* $\mathbf{x}, \mathbf{y}$ *de* $\mathbb{R}^n$ *su producto escalar es el número real*

$$\langle \mathbf{x}, \mathbf{y} \rangle := \mathbf{x}^t \mathbf{y} = \sum_{j=1}^{n} x_j y_j = \mathbf{y}^t \mathbf{x}. \tag{3.1}$$

De la definición se desprenden las siguientes propiedades:

1. Distributividad, es decir: $\forall\ \mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^n,\ \forall\ \lambda, \mu \in \mathbb{R}$

$$\langle \lambda\mathbf{u} + \mu\mathbf{v}, \mathbf{w}\rangle = \lambda\langle \mathbf{u}, \mathbf{w}\rangle + \mu\langle \mathbf{v}, \mathbf{w}\rangle;\ \langle \mathbf{u}, \lambda\mathbf{v} + \mu\mathbf{w}\rangle = \lambda\langle \mathbf{u}, \mathbf{v}\rangle + \mu\langle \mathbf{u}, \mathbf{w}\rangle \tag{3.2}$$

   Se dice también que el producto escalar es lineal en cada componente.

2. Simetría:

$$\forall\ \mathbf{u}, \mathbf{v} \in \mathbb{R}^n \quad \langle \mathbf{u}, \mathbf{v}\rangle = \langle \mathbf{v}, \mathbf{u}\rangle. \tag{3.3}$$

3. Positividad:

$$\forall\ \mathbf{u} \in \mathbb{R}^n,\ \mathbf{u} \neq \mathbf{0} \quad \langle \mathbf{u}, \mathbf{u}\rangle = \mathbf{u}^t\mathbf{u} = \sum_{j=1}^{n} u_j^2 > 0. \tag{3.4}$$

Esta última propiedad es muy importante y motiva la siguiente

Definición 38 (norma euclídea) *Para cada vector* $\mathbf{u}$ *de* $\mathbb{R}^n$ *el número*

$$\|\mathbf{u}\| := \sqrt{\langle \mathbf{u}, \mathbf{u}\rangle} \tag{3.5}$$

recibe el nombre de norma o longitud del vector.

Proposición 17 *La aplicación que a cada vector* $\mathbf{u}$ *le asocia su norma* $\|\mathbf{u}\|$ *se denomina* ***norma asociada al producto escalar*** *y cumple:*

1. $\forall\ \mathbf{u} \in \mathbb{R}^n,\ \|\mathbf{u}\| \geq 0,\ \|\mathbf{u}\| = 0 \iff \mathbf{u} = \mathbf{0}.$
2. $\forall\ \lambda \in \mathbb{R},\ \forall\ \mathbf{u} \in \mathbb{R}^n \quad \|\lambda\mathbf{u}\| = |\lambda|\ \|\mathbf{u}\|.$
3. $\forall\ \mathbf{u}, \mathbf{v} \in \mathbb{R}^n,\ \|\mathbf{u} + \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + 2\langle \mathbf{u}, \mathbf{v}\rangle + \|\mathbf{v}\|^2.$

Definición 39 *Dados dos vectores* $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ *se define la distancia entre ellos:*

$$d(\mathbf{u}, \mathbf{v}) := \|\mathbf{u} - \mathbf{v}\|. \tag{3.6}$$

Proposición 18 (desigualdades de Cauchy-Schwarz y triangular) *Sean* $\mathbf{u}, \mathbf{v}$ *dos vectores cualesquiera de* $\mathbb{R}^n$*. Entonces*

1. *La* ***desigualdad de Cauchy-Schwarz*** *nos da la relación entre el producto escalar de los dos vectores y sus longitudes:*

$$\forall\ \mathbf{u}, \mathbf{v} \in \mathbb{R}^n \quad |\langle \mathbf{u}, \mathbf{v}\rangle| \leq \|\mathbf{u}\|\ \|\mathbf{v}\|. \tag{3.7}$$

   La igualdad se da si, y solo si, los vectores son proporcionales.

2. *La* ***desigualdad triangular*** *relaciona la suma de las longitudes con la longitud de la suma:*

$$\forall\ \mathbf{u},\mathbf{v}\in\mathbb{R}^n,\ \|\mathbf{u}+\mathbf{v}\|\leq\|\mathbf{u}\|+\|\mathbf{v}\|. \tag{3.8}$$

La igualdad se da si, y solo si, los vectores son proporcionales con constante de proporcionalidad no negativa.

Demostración: En virtud de la distributividad y de la simetría:

$$\|\mathbf{u}+\lambda\mathbf{v}\|^2=\langle\mathbf{u}+\lambda\mathbf{v},\mathbf{u}+\lambda\mathbf{v}\rangle=\|\mathbf{u}\|^2+2\lambda\langle\mathbf{u},\mathbf{v}\rangle+\lambda^2\|\mathbf{v}\|^2 \tag{3.9}$$

cualquiera que sea $\lambda\in\mathbb{R}$. Podemos suponer $\mathbf{v}\neq\mathbf{0}$ (si $\mathbf{v}=\mathbf{0}$ ambos miembros de (3.7) son nulos). La expresión (3.9) es un polinomio de segundo grado en λ mayor o igual que cero para todo λ. Su discriminante tiene que ser menor o igual que cero:

$$\langle\mathbf{u},\mathbf{v}\rangle^2-\|\mathbf{u}\|^2\|\mathbf{v}\|^2\leq 0 \iff \langle\mathbf{u},\mathbf{v}\rangle^2\leq\|\mathbf{u}\|^2\|\mathbf{v}\|^2$$

Al tomar raíces cuadradas positivas en ambos miembros queda demostrada la desigualdad. Además, la igualdad en (3.7) equivale a $\|\mathbf{u}+\lambda\mathbf{v}\|^2=0$, es decir, $\mathbf{u}=-\lambda\mathbf{v}$, para un determinado valor de λ: los vectores son proporcionales.

Para probar la desigualdad triangular elevamos al cuadrado el primer miembro de la desigualdad, desarrollamos y aplicamos la desigualdad de Cauchy-Schwarz:

$$\|\mathbf{u}+\mathbf{v}\|^2=\|\mathbf{u}\|^2+2\langle\mathbf{u},\mathbf{v}\rangle+\|\mathbf{v}\|^2\leq\|\mathbf{u}\|^2+2\|\mathbf{u}\|\ \|\mathbf{v}\|+\|\mathbf{v}\|^2=(\|\mathbf{u}\|+\|\mathbf{v}\|)^2$$

Al extraer raíces cuadradas positivas se llega al resultado deseado.

La igualdad se da si, y solo si, $\langle\mathbf{u},\mathbf{v}\rangle=\|\mathbf{u}\|\ \|\mathbf{v}\|$, entonces $\mathbf{u}$ y $\mathbf{v}$ son proporcionales, $\mathbf{u}=-\lambda\mathbf{v}$ y además

$$0\leq\|\mathbf{u}\|\ \|\mathbf{v}\|=\langle\mathbf{u},\mathbf{v}\rangle=\langle-\lambda\mathbf{v},\mathbf{v}\rangle=-\lambda\|\mathbf{v}\|^2$$

por lo que la constante de proporcionalidad $(-\lambda)$ es no negativa. □

Observación: La desigualdad triangular dice que la longitud de una suma de vectores no puede exceder la suma de las longitudes de ambos vectores. Intuitivamente es muy natural; si pensamos geométricamente, los dos vectores definen un paralelogramo de lados de longitudes $\|\mathbf{u}\|$ y $\|\mathbf{v}\|$, mientras que las longitudes de las diagonales son $\|\mathbf{u}\pm\mathbf{v}\|$.

La desigualdad de Cauchy-Schwarz permite definir el ángulo entre dos vectores,

Definición 40 *Sean* $\mathbf{u},\mathbf{v}$ *dos vectores no nulos de* $\mathbb{R}^n$*, se define al ángulo* φ *que forman como aquel cuyo coseno es*

$$\cos\varphi=\frac{\langle\mathbf{u},\mathbf{v}\rangle}{\|\mathbf{u}\|\ \|\mathbf{v}\|},\quad \varphi\in[0,\pi]. \tag{3.10}$$

3.2. Ortogonalidad

Definición 41 *Dos vectores* $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ *son* ***ortogonales*** *(o perpendiculares) si su producto escalar es nulo:* $\langle \mathbf{u}, \mathbf{v} \rangle = 0$*. En ese caso es frecuente escribir* $\mathbf{u} \perp \mathbf{v}$*.*

Observación: Si un vector es ortogonal a sí mismo es nulo. Es consecuencia de la propiedad de positividad. A veces esta propiedad se enuncia diciendo que el único vector ortogonal a todos los del espacio es el nulo:

$$\forall\ \mathbf{v} \in \mathbb{R}^n\ \langle \mathbf{u}, \mathbf{v} \rangle = 0 \Rightarrow \mathbf{u} = \mathbf{0}. \tag{3.11}$$

Teorema 7 (Pitágoras) *Dos vectores* $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ *son* ***ortogonales*** *si, y solamente si*

$$\|\mathbf{u} + \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2. \tag{3.12}$$

Demostración: Es consecuencia inmediata del apartado 3. de la proposición 17. $\square$

Proposición 19 (suplementario ortogonal) *Sea* M *un subespacio de* $\mathbb{R}^n$*, el conjunto*

$$M^\perp := \{\mathbf{v} \in \mathbb{R}^n,\ \forall\ \mathbf{u} \in M,\ \langle \mathbf{u}, \mathbf{v} \rangle = 0\} \tag{3.13}$$

es un subespacio vectorial de $\mathbb{R}^n$ *que es suplementario de* M*. Es decir,* $\mathbb{R}^n = M \oplus M^\perp$*.*

Demostración: Sea m la dimensión de M. Identifiquemos $M = \operatorname{Im} \mathbf{A}$ siendo $\mathbf{A} \in \mathbb{R}^{n \times m}$ una matriz cuyas columnas son una base de M. Un vector $\mathbf{x}$ es de $M^\perp$ si, y solamente si, es ortogonal a los vectores de una base de M, es decir si $\mathbf{A}^t\,\mathbf{x} = \mathbf{0}$, así pues $M^\perp = \ker \mathbf{A}^t$. Como el rango de $\mathbf{A}$ es m, la dimensión del núcleo de $\mathbf{A}^t$ es $n - m$, luego para que M y $M^\perp$ sean suplementarios basta que su intersección sea nula; pero ello es evidente dado que el único vector ortogonal a sí mismo es el nulo, por lo que podemos concluir que $\mathbb{R}^n = M \oplus M^\perp$. $\square$

Observación: Notemos que este razonamiento puede hacerse también considerando una matriz $\mathbf{A}$ cuyas columnas sean un generador de M, no necesariamente libre, en este caso pueden variar sus dimensiones (habrá tantas columnas como vectores posea el generador) pero no su rango, que coincide con la dimensión.

Observación: De la relación de dimensiones y de la definición se desprende que $(M^\perp)^\perp = M$.

Corolario: Dada una matriz real cualquiera $\mathbf{A}$ se verifica:

$$(\mathrm{Im}\,\mathbf{A})^\perp = \ker \mathbf{A}^t, \qquad (\mathrm{Im}\,\mathbf{A}^t)^\perp = \ker \mathbf{A}. \tag{3.14}$$

Teorema 8 (de la proyección ortogonal) *Sea M un subespacio de $\mathbb{R}^n$. Entonces para cada vector $\mathbf{u} \in \mathbb{R}^n$ existen únicos $\mathbf{v} \in M$, $\mathbf{w} \in M^\perp$ de forma que $\mathbf{u} = \mathbf{v} + \mathbf{w}$; además, para cualquier $\mathbf{z} \in M$ se cumple $d(\mathbf{u}, \mathbf{z}) = \|\mathbf{u} - \mathbf{z}\| \geq \|\mathbf{u} - \mathbf{v}\| = d(\mathbf{u}, \mathbf{v})$.*

Demostración: De $\mathbb{R}^n = M \oplus M^\perp$ se desprende la descomposición $\mathbf{u} = \mathbf{v} + \mathbf{w}$, $\mathbf{v} \in M$, $\mathbf{w} \in M^\perp$ y su unicidad. En cuanto a la distancia, sea $\mathbf{z} \in M$ cualquiera,

$$\|\mathbf{u} - \mathbf{z}\|^2 = \|(\mathbf{u} - \mathbf{v}) + (\mathbf{v} - \mathbf{z})\|^2 \overset{(*)}{=} \|\mathbf{u} - \mathbf{v}\|^2 + \|\mathbf{v} - \mathbf{z}\|^2 \geq \|\mathbf{u} - \mathbf{v}\|^2.$$

La igualdad en $(*)$ se da en virtud del teorema de Pitágoras ya que $(\mathbf{u} - \mathbf{v})$ y $(\mathbf{v} - \mathbf{z})$ son ortogonales (el primero es un vector de $M^\perp$ y el segundo es de M). □

Los resultados de este teorema motivan las dos definiciones siguientes:

Definición 42 *El vector $\mathbf{v}$ recibe el nombre de **proyección ortogonal** de $\mathbf{u}$ sobre M; lo denotaremos $P_M\mathbf{u}$.*

Definición 43 *Dado un subespacio M de $\mathbb{R}^n$, para cada vector $\mathbf{u} \in \mathbb{R}^n$ su distancia a M se define como*

$$d(\mathbf{u}, M) := \min\{d(\mathbf{u}, \mathbf{v}),\ \mathbf{v} \in M\} = \|\mathbf{u} - P_M\mathbf{u}\|. \tag{3.15}$$

Observación: El teorema de la proyección ortogonal no solo garantiza la existencia del mínimo sino que establece que se alcanza precisamente en la proyección de $\mathbf{u}$ sobre M.

3.2.1. Familias ortogonales

Definición 44 *Un conjunto de vectores $\mathcal{F} = \{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ de $\mathbb{R}^n$ es **ortogonal** si $\langle \mathbf{u}_i, \mathbf{u}_j \rangle = 0$ para $i \neq j$. Se dice que los vectores de $\mathcal{F}$ son **ortogonales dos a dos**.*

Definición 45 *Un conjunto de vectores de $\mathbb{R}^n$ es **ortonormal** si sus vectores son ortogonales dos a dos y **unitarios** (es decir de norma 1).*

Proposición 20 *Una familia ortogonal que no contiene el vector nulo es libre. En particular, toda familia ortonormal es libre.*

Demostración: Sea $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_s\}$ un conjunto de vectores tales que $\mathbf{u}_j^t\mathbf{u}_k = 0,\ j \neq k$. Igualemos a cero una combinación lineal:

$$\lambda_1\mathbf{u}_1 + \lambda_2\mathbf{u}_2 + \cdots + \lambda_s\mathbf{u}_s = \mathbf{0} \tag{3.16}$$

y multipliquemos escalarmente por $\mathbf{u}_k$ para un índice k cualquiera,

$$\lambda_1 \mathbf{u}_k^t \mathbf{u}_1 + \lambda_2 \mathbf{u}_k^t \mathbf{u}_2 + \cdots + \lambda_s \mathbf{u}_k^t \mathbf{u}_s = 0 \tag{3.17}$$

la expresión queda reducida a

$$\lambda_k \mathbf{u}_k^t \mathbf{u}_k = \lambda_k \|\mathbf{u}_k\|^2 = 0 \tag{3.18}$$

como el vector $\mathbf{u}_k$ es no nulo, tiene que ser $\lambda_k = 0$, haciendo variar k desde 1 hasta s queda probada la propiedad. $\square$

Corolario: (generalización del teorema de Pitágoras) Si los vectores $\mathbf{u}_1, \ldots, \mathbf{u}_m$ son ortogonales dos a dos, entonces

$$\|\mathbf{u}_1 + \cdots + \mathbf{u}_m\|^2 = \|\mathbf{u}_1\|^2 + \cdots + \|\mathbf{u}_m\|^2. \tag{3.19}$$

Observación: El recíproco no es cierto para $m > 2$ como muestran los vectores

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

para los que se cumple $\|\mathbf{e}_1 + \mathbf{e}_2 + \mathbf{u}\|^2 = \|\mathbf{e}_1\|^2 + \|\mathbf{e}_2\|^2 + \|\mathbf{u}\|^2 = 4$ pero no son ortogonales dos a dos ya que por ejemplo $\langle \mathbf{e}_1, \mathbf{u} \rangle = 1$.

Definición 46 *Una base ortogonal de $\mathbb{R}^n$ es aquella cuyos vectores son ortogonales dos a dos; si además son unitarios la base se llama ortonormal.*

Ejemplo:

$$\mathcal{B}_1 = \left(\begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -2 \end{pmatrix} \right) \quad \mathcal{B}_o = \left(\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right)$$

$\mathcal{B}_1$ es una base ortogonal de $\mathbb{R}^2$ mientras que $\mathcal{B}_o$ es ortonormal.

Observación: Sea $\mathcal{B} = (\mathbf{u}_1, \ldots, \mathbf{u}_n)$ una base **ortogonal** de $\mathbb{R}^n$ y sea $\mathbf{v} \in \mathbb{R}^n$, entonces

$$\mathbf{v} = \sum_{j=1}^{n} \frac{\langle \mathbf{v}, \mathbf{u}_j \rangle}{\|\mathbf{u}_j\|^2} \mathbf{u}_j. \tag{3.20}$$

Si en la base $\mathcal{B}$ normalizamos cada vector $\mathbf{q}_j := \mathbf{u}_j/\|\mathbf{u}_j\|$ la nueva base $\mathcal{B}_o = (\mathbf{q}_1, \ldots, \mathbf{q}_n)$ es **ortonormal** y la expresión anterior se simplifica:

$$\mathbf{v} = \sum_{j=1}^{n} \langle \mathbf{v}, \mathbf{q}_j \rangle \mathbf{q}_j \tag{3.21}$$

y el vector de coordenadas de $\mathbf{v}$ en la base ortonormal es muy sencillo:

$$[\mathbf{v}]^{\mathcal{B}_o} = \begin{pmatrix} \langle \mathbf{v}, \mathbf{q}_1 \rangle \\ \vdots \\ \langle \mathbf{v}, \mathbf{q}_n \rangle \end{pmatrix} \tag{3.22}$$

Además, y conforme a la fórmula (3.19), podemos expresar la norma:

$$\|\mathbf{v}\| = \sqrt{\sum_{j=1}^{n} \langle \mathbf{v}, \mathbf{q}_j \rangle^2}. \tag{3.23}$$

Mientras que el producto escalar de dos vectores $\mathbf{v}, \mathbf{w}$ puede escribirse como:

$$\langle \mathbf{v}, \mathbf{w} \rangle = \sum_{j=1}^{n} \langle \mathbf{v}, \mathbf{q}_j \rangle \langle \mathbf{q}_j, \mathbf{w} \rangle. \tag{3.24}$$

Definición 47 *Una matriz* $\mathbf{Q} \in \mathbb{R}^{n\times n}$ *es* ***ortogonal*** *si* $\mathbf{Q}^t\mathbf{Q} = \mathbf{I}$*, es decir* $\mathbf{Q}^t = \mathbf{Q}^{-1}$.

Proposición 21 *Sea* $\mathbf{Q} \in \mathbb{R}^{n\times n}$*. Entonces son equivalentes:*

1. $\mathbf{Q}$ *es ortogonal.*
2. $\mathbf{Q}$ *conserva el producto escalar, es decir:*
$$\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n \quad \langle \mathbf{Q}\mathbf{x}, \mathbf{Q}\mathbf{y} \rangle = \langle \mathbf{x}, \mathbf{y} \rangle.$$
3. $\mathbf{Q}$ *conserva la norma, es decir:*
$$\forall \mathbf{x} \in \mathbb{R}^n \quad \|\mathbf{Q}\mathbf{x}\| = \|\mathbf{x}\|.$$

Demostración: Comprobaremos la triple equivalencia mediante un razonamiento circular: (1) $\Rightarrow$ (2) $\Rightarrow$ (3) $\Rightarrow$ (1).

(1) $\Rightarrow$ (2) Sean $\mathbf{x}, \mathbf{y}$ dos vectores cualesquiera, entonces:

$$\langle \mathbf{Q}\mathbf{x}, \mathbf{Q}\mathbf{y} \rangle = (\mathbf{Q}\mathbf{x})^t\mathbf{Q}\mathbf{y} = \mathbf{x}^t\mathbf{Q}^t\mathbf{Q}\mathbf{y} = \mathbf{x}^t\mathbf{I}\mathbf{y} = \mathbf{x}^t\mathbf{y} = \langle \mathbf{x}, \mathbf{y} \rangle.$$

(2) ⇒ (3) Basta tomar $\mathbf{x} = \mathbf{y}$.

(3) ⇒ (1) Sean $\mathbf{e}_i, \mathbf{e}_j$ dos vectores canónicos,

$$\|\mathbf{Q}(\mathbf{e}_i + \mathbf{e}_j)\|^2 = \|\mathbf{Q}\mathbf{e}_i\|^2 + 2\langle \mathbf{Q}\mathbf{e}_i, \mathbf{Q}\mathbf{e}_j\rangle + \|\mathbf{Q}\mathbf{e}_j\|^2$$

$$\|\mathbf{e}_i + \mathbf{e}_j\|^2 = \|\mathbf{e}_i\|^2 + 2\langle \mathbf{e}_i, \mathbf{e}_j\rangle + \|\mathbf{e}_j\|^2$$

Por hipótesis los primeros miembros son iguales $\|\mathbf{Q}(\mathbf{e}_i + \mathbf{e}_j)\|^2 = \|\mathbf{e}_i + \mathbf{e}_j\|^2$, igualemos los segundos:

$$\begin{aligned} \|\mathbf{Q}\mathbf{e}_i\|^2 + 2\langle \mathbf{Q}\mathbf{e}_i, \mathbf{Q}\mathbf{e}_j\rangle + \|\mathbf{Q}\mathbf{e}_j\|^2 &= \|\mathbf{e}_i\|^2 + 2\langle \mathbf{e}_i, \mathbf{e}_j\rangle + \|\mathbf{e}_j\|^2 \\ \iff \langle \mathbf{Q}\mathbf{e}_i, \mathbf{Q}\mathbf{e}_j\rangle = \langle \mathbf{e}_i, \mathbf{e}_j\rangle &\iff \mathbf{e}_i^t\mathbf{Q}^t\mathbf{Q}\mathbf{e}_j = \mathbf{e}_i^t\mathbf{e}_j \\ \iff (\mathbf{Q}^t\mathbf{Q})_{ij} &= \begin{cases} 1 & \text{si } i = j \\ 0 & \text{si } i \neq j \end{cases} \end{aligned}$$

Con lo que queda probado que $\mathbf{Q}^t\mathbf{Q} = \mathbf{I}$. □

Observación: Las columnas (y filas) de una matriz ortogonal $\mathbf{Q}$ constituyen una base ortonormal de $\mathbb{R}^n$ con el producto escalar natural. Además, el hecho de conservar el producto escalar y la norma se traduce en que también se conservan los ángulos.

3.2.2. Ortonormalización de Gram-Schmidt y factorización QR

La base canónica en $\mathbb{R}^n$ es ortonormal, pero pueden encontrarse otras bases ortonormales construidas con direcciones predeterminadas. En este apartado se demuestra este resultado y se estudian algunas consecuencias importantes de este hecho.

Teorema 9 (Gram-Schmidt) *Sea* $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\}$ *una familia libre en* $\mathbb{R}^n$*, entonces existe una familia **ortonormal*** $\{\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_m\}$ *tal que*

$$L[\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k] = L[\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_k], \quad k = 1, 2, \ldots, m. \tag{3.25}$$

Construcción:

$$\begin{aligned} &\mathbf{q}_1 = \frac{\mathbf{u}_1}{\|\mathbf{u}_1\|}, \\ &\tilde{\mathbf{q}}_2 = \mathbf{u}_2 - \langle \mathbf{u}_2, \mathbf{q}_1\rangle \mathbf{q}_1, \quad \mathbf{q}_2 = \frac{\tilde{\mathbf{q}}_2}{\|\tilde{\mathbf{q}}_2\|} \\ &\vdots \\ &\tilde{\mathbf{q}}_k = \mathbf{u}_k - \langle \mathbf{u}_k, \mathbf{q}_1\rangle \mathbf{q}_1 - \cdots - \langle \mathbf{u}_k, \mathbf{q}_{k-1}\rangle \mathbf{q}_{k-1}, \quad \mathbf{q}_k = \frac{\tilde{\mathbf{q}}_k}{\|\tilde{\mathbf{q}}_k\|} \qquad k = 3, \ldots, m. \end{aligned} \tag{3.26}$$

Observemos que el procedimiento seguido consiste simplemente en restar a cada vector su proyección ortogonal sobre el subespacio engendrado por los vectores anteriores; es decir, en la etapa k restamos al vector $\mathbf{u}_k$ su proyección sobre $L[\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{k-1}]$, como puede verse en la figura 3.1.

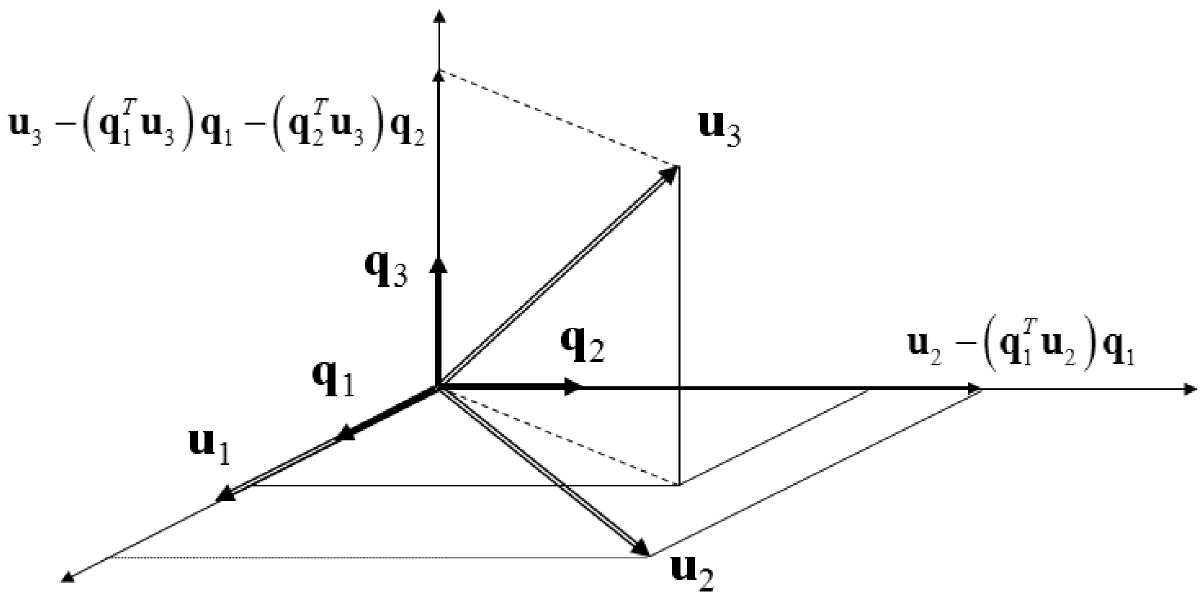

Figura 3.1: Gram-Schmidt para tres vectores

Consecuencia importante

Sea M un subespacio de $\mathbb{R}^n$ de dimensión m y $\mathcal{B} = (\mathbf{u}_1, \ldots, \mathbf{u}_m, \mathbf{u}_{m+1}, \ldots, \mathbf{u}_n)$ una base de $\mathbb{R}^n$ obtenida *completando* una base de M; al ortonormalizar por Gram-Schmidt obtenemos una base ortonormal $\mathcal{B}_o = (\mathbf{q}_1, \ldots, \mathbf{q}_m, \mathbf{q}_{m+1}, \ldots, \mathbf{q}_n)$ de $\mathbb{R}^n$. Los vectores de esta base pueden agruparse de modo que los m primeros $(\mathbf{q}_1, \ldots, \mathbf{q}_m)$ constituyen una base ortonormal de M y el resto $(\mathbf{q}_{m+1}, \ldots, \mathbf{q}_n)$ una base ortonormal de $M^\perp$.

Teorema 10 (factorización QR) *Sea* $\mathbf{A} \in \mathbb{R}^{n\times m}$ *de rango* m*, entonces existen matrices* $\mathbf{Q} \in \mathbb{R}^{n\times m}$ *y* $\mathbf{R} \in \mathbb{R}^{m\times m}$ *tales que* $\mathbf{Q}^t\mathbf{Q} = \mathbf{I}$*,* $\mathbf{R}$ *es triangular superior con elementos positivos en su diagonal y tales que* $\mathbf{A} = \mathbf{QR}$*. Además la factorización es única.*

Demostración: (no exigible) Para demostrar la existencia basta con aplicar el método de Gram-Schmidt a las columnas de $\mathbf{A}$ que son linealmente independientes (por hipótesis el rango de A es m), llamemos $\mathbf{u}_k$ a las columnas de $\mathbf{A}$ y denotemos por $\mathbf{q}_k$ los vectores obtenidos por Gram-Schmidt, despejemos en (3.26) los vectores $\mathbf{u}_k$:

$$\begin{aligned}
&\mathbf{u}_1 = \|\mathbf{u}_1\|\mathbf{q}_1,\\
&\mathbf{u}_2 = (\mathbf{u}_2^t\mathbf{q}_1)\ \mathbf{q}_1 + \|\mathbf{u}_2 - (\mathbf{u}_2^t\mathbf{q}_1)\ \mathbf{q}_1\|\ \mathbf{q}_2\\
&\vdots\\
&\mathbf{u}_k = (\mathbf{u}_k^t\mathbf{q}_1)\ \mathbf{q}_1 + (\mathbf{u}_k^t\mathbf{q}_2)\ \mathbf{q}_2 + \cdots + (\mathbf{u}_k^t\mathbf{q}_{k-1})\ \mathbf{q}_{k-1} + \|\mathbf{u}_k - \mathbf{u}_k^t\mathbf{q}_{k-1}\ \mathbf{q}_{k-1} - \cdots - \mathbf{u}_k^t\mathbf{q}_1\ \mathbf{q}_1\|\ \mathbf{q}_k\\
&\vdots
\end{aligned}$$

y escribamos matricialmente el resultado:

$$(\mathbf{u}_1 \quad \mathbf{u}_2 \quad \cdots \quad \mathbf{u}_m) = (\mathbf{q}_1 \quad \mathbf{q}_2 \quad \cdots \quad \mathbf{q}_m) \begin{pmatrix} * & * & \cdots & * \\ 0 & * & \cdots & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & * \end{pmatrix} = \mathbf{QR}.$$

Observemos que la matriz $\mathbf{R}$ puede contemplarse como la matriz de paso de la base de $\operatorname{Im}\mathbf{A}$ formada por las columnas de $\mathbf{Q}$ a la base formada por las columnas de $\mathbf{A}$. Es evidente que $\mathbf{R}$ es triangular superior con elementos positivos en la diagonal.

En cuanto a la unicidad, supongamos dos factorizaciones $\mathbf{A} = \mathbf{QR} = \mathbf{PT}$ que cumplan los requisitos, es decir $\mathbf{Q}^t\mathbf{Q} = \mathbf{I}$, $\mathbf{P}^t\mathbf{P} = \mathbf{I}$, $\mathbf{R}$ y $\mathbf{T}$ triangulares superiores con elementos diagonales positivos, y concluyamos que son iguales. Sígase atentamente la siguiente cadena:

$$\mathbf{I} = \mathbf{Q}^t\mathbf{Q} = (\mathbf{PTR}^{-1})^t\ \mathbf{PTR}^{-1} = (\mathbf{TR}^{-1})^t\ \mathbf{P}^t\ \mathbf{PTR}^{-1} = (\mathbf{TR}^{-1})^t\ \mathbf{TR}^{-1}$$

La matriz $\mathbf{TR}^{-1}$ es triangular superior y su inversa es su traspuesta, pero la inversa de una matriz triangular superior es triangular superior y su traspuesta es triangular inferior, ambas cosas simultáneamente solo son posibles si $\mathbf{TR}^{-1}$ es diagonal e igual a su inversa. Si llamamos $r_i > 0$ a los elementos diagonales de $\mathbf{R}$ y $t_i > 0$ a los de $\mathbf{T}$ tenemos:

$$\mathbf{TR}^{-1} = \operatorname{diag}\left(\frac{t_1}{r_1}, \ldots, \frac{t_m}{r_m}\right), \quad \mathbf{RT}^{-1} = \operatorname{diag}\left(\frac{r_1}{t_1}, \ldots, \frac{r_m}{t_m}\right)$$

$$\mathbf{TR}^{-1} = \mathbf{RT}^{-1} \Rightarrow \frac{t_i}{r_i} = \frac{r_i}{t_i} \Rightarrow t_i^2 = r_i^2 \Rightarrow t_i = r_i.$$

Se concluye que $\mathbf{TR}^{-1} = \mathbf{I}$ y por tanto $\mathbf{T} = \mathbf{R}$: la factorización es única. $\square$

3.3. Extensión a $\mathbb{C}^n$

Todo lo visto anteriormente puede extenderse al espacio $\mathbb{C}^n$; el producto escalar se define del siguiente modo:

Definición 48 (producto escalar usual) *Dados* $\mathbf{u}, \mathbf{v} \in \mathbb{C}^n$*, su producto escalar usual es* $\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}^t\overline{\mathbf{v}} = \mathbf{v}^h\mathbf{u}$.

Propiedades

De forma análoga a como ocurría en el caso real, se tienen las siguientes propiedades:

1. Linealidad en la primera componente, es decir:

$$\forall\ \mathbf{u},\mathbf{v},\mathbf{w}\in\mathbb{C}^n,\ \forall\ \lambda,\mu\in\mathbb{C}\quad \langle\lambda\mathbf{u}+\mu\mathbf{v},\mathbf{w}\rangle=\lambda\langle\mathbf{u},\mathbf{w}\rangle+\mu\langle\mathbf{v},\mathbf{w}\rangle. \tag{3.27}$$

2. Simetría *hermítica*:

$$\forall\ \mathbf{u},\mathbf{v}\in\mathbb{C}^n\quad \langle\mathbf{u},\mathbf{v}\rangle=\overline{\langle\mathbf{v},\mathbf{u}\rangle}. \tag{3.28}$$

3. Positividad:

$$\forall\ \mathbf{u}\in\mathbb{C}^n,\ \mathbf{u}\neq\mathbf{0}\quad \langle\mathbf{u},\mathbf{u}\rangle\in\mathbb{R},\quad \langle\mathbf{u},\mathbf{u}\rangle>0. \tag{3.29}$$

Observación: Las propiedades 1 y 2 conjuntamente implican una propiedad llamada *antilinealidad* en la segunda componente:

$$\forall\ \mathbf{u},\mathbf{v},\mathbf{w}\in\mathbb{C}^n,\ \forall\ \lambda,\mu\in\mathbb{C}\quad \langle\mathbf{u},\lambda\mathbf{v}+\mu\mathbf{w}\rangle=\overline{\lambda}\langle\mathbf{u},\mathbf{v}\rangle+\overline{\mu}\langle\mathbf{u},\mathbf{w}\rangle. \tag{3.30}$$

Cuando una aplicación sobre $\mathbb{C}^n\times\mathbb{C}^n$ es lineal en la primera componente y antilineal en la segunda se dice que es *sesquilineal*.

Ejemplo: El producto escalar de $\mathbf{u}=\begin{pmatrix}1+i\\2-i\end{pmatrix}$ por $\mathbf{v}=\begin{pmatrix}1-2i\\2+3i\end{pmatrix}$ es:

$$\langle\mathbf{u},\mathbf{v}\rangle=\mathbf{v}^h\ \mathbf{u}=(1+2i\ \ 2-3i)\begin{pmatrix}1+i\\2-i\end{pmatrix}=(1+2i)(1+i)+(2-3i)(2-i)=-5i.$$

Observemos que el <u>orden</u> de los vectores es importante ya que, en virtud de la simetría hermítica, $\langle\mathbf{v},\mathbf{u}\rangle=\overline{\langle\mathbf{u},\mathbf{v}\rangle}=5i$.

Norma y propiedades relacionadas

La definición de norma es idéntica a la dada en (3.5),

$$\mathbf{u}\in\mathbb{C}^n\quad \|\mathbf{u}\|=\sqrt{\langle\mathbf{u},\mathbf{u}\rangle}=\sqrt{\mathbf{u}^h\mathbf{u}}=\sqrt{\sum_{j=1}^{n}|u_j|^2}. \tag{3.31}$$

Ejemplo: La norma de $\mathbf{u}=\begin{pmatrix}1+i\\2-i\\-1\end{pmatrix}$ es $\|\mathbf{u}\|=\sqrt{|1+i|^2+|2-i|^2+1}=2\sqrt{2}$.

Observación: Las propiedades enunciadas en $\mathbb{R}^n$ generalmente son válidas en el espacio $\mathbb{C}^n$; por ejemplo, las desigualdades de Cauchy-Schwarz y triangular se cumplen en el caso complejo si bien las demostraciones que hemos dado solo son válidas en el caso real. El teorema de Pitágoras solo se cumple en su versión generalizada, es decir, la igualdad $\|\mathbf{u}+\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$ para vectores $\mathbf{u}, \mathbf{v} \in \mathbb{C}^n$ no implica la ortogonalidad (véase el ejercicio 3.16).

3.3.1. Matrices unitarias

Definición 49 (matriz unitaria) *Una matriz* $\mathbf{U} \in \mathbb{C}^{n\times n}$ *es* ***unitaria*** *si* $\mathbf{U}^h\mathbf{U} = \mathbf{I}$.

Enunciamos (sin demostración) el equivalente de la proposición 21 para matrices unitarias.

Proposición 22 *Sea* $\mathbf{U} \in \mathbb{C}^{n\times n}$. *Entonces son equivalentes:*

1. $\mathbf{U}$ *es unitaria.*
2. $\mathbf{U}$ *conserva el producto escalar, es decir:*
$$\forall \mathbf{x}, \mathbf{y} \in \mathbb{C}^n \quad \langle \mathbf{U}\mathbf{x}, \mathbf{U}\mathbf{y}\rangle = \langle \mathbf{x}, \mathbf{y}\rangle.$$
3. $\mathbf{U}$ *conserva la norma, es decir:*
$$\forall \mathbf{x} \in \mathbb{C}^n \quad \|\mathbf{U}\mathbf{x}\| = \|\mathbf{x}\|.$$

Observación: Como en el caso real, las columnas (y filas) de una matriz unitaria constituyen una base ortonormal de $\mathbb{C}^n$ con el producto escalar natural.

Ejemplo: La matriz
$$\frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ i & -i \end{pmatrix}$$
es unitaria. Sus columnas forman una base ortonormal de $\mathbb{C}^2$.

3.4. Ejercicios

3.1.– Dados los vectores $(2,-1,1)$ y $(1,0,1)$, hallar su producto escalar, sus normas y el ángulo que forman.

3.2.– Sean $\mathbf{v}$ y $\mathbf{w}$ dos vectores distintos de $\mathbb{R}^n$, $(n \geq 3)$. Utilizar la desigualdad triangular para demostrar que si se cumple la igualdad

$$\|\mathbf{x}-\mathbf{v}\| + \|\mathbf{x}-\mathbf{w}\| = \|\mathbf{v}-\mathbf{w}\|$$

entonces $\mathbf{x}$ pertenece al subespacio generado por $\mathbf{v}$ y $\mathbf{w}$.

3.3.– Hallar los valores de a, b y c para que sea ortogonal la familia

$$\{(1,-2,1,0),(1,a,3,-7),(b,4,-1,c)\}.$$

3.4.– Demostrar que dos vectores $\mathbf{v}$ y $\mathbf{w}$ tienen la misma norma si y solo si su suma $\mathbf{v}+\mathbf{w}$ y su diferencia $\mathbf{v}-\mathbf{w}$ son ortogonales.

3.5.– Sean F y G subespacios vectoriales de $\mathbb{R}^n$. Demostrar las siguientes igualdades:

1. $(F+G)^{\perp} = F^{\perp} \cap G^{\perp}$.
2. $(F \cap G)^{\perp} = F^{\perp} + G^{\perp}$.

3.6.– Hallar una base ortonormal del plano $x+y=5z$, de forma que su primer vector sea proporcional a $(1,-1,0)$.

3.7.– Probar que los siguientes subespacios son suplementarios ortogonales en $\mathbb{R}^n$:

$$L: \; x_1 + \cdots + x_n = 0 \qquad M: \; x_1 = \cdots = x_n.$$

Calcular la proyección ortogonal del vector $(1,0,0,\cdots,0)$ sobre L.

3.8.– Determinar la proyección ortogonal del vector $(1,-1,0,2)$ sobre el subespacio generado por los dos vectores $(1,1,0,1)$ y $(0,2,-1,1)$.

3.9.– En el espacio euclídeo $\mathbb{R}^4$ se considera el subespacio $\mathcal{M}$ generado por los vectores $(1,0,1,1), (1,1,0,1)$ y $(1,1,1,0)$. Se pide:

1. Calcular una base ortonormal del subespacio $\mathcal{M}^{\perp}$.
2. Descomponer el vector $\mathbf{b} = (7,7,7,7)$ como suma de sendos vectores de $\mathcal{M}$ y de $\mathcal{M}^{\perp}$.

3.10.– Escribir la distancia del vector $(1,3,2,0)$ al subespacio de $\mathbb{R}^4$ de ecuaciones $x_1 = x_2, \; x_3 = x_4$.

3.11.– Dados dos vectores $\mathbf{u}$ y $\mathbf{v} \neq \mathbf{0}$ de $\mathbb{R}^n$, se pide:

1. Determinar la proyección ortogonal de $\mathbf{u}$ sobre la recta generada por $\mathbf{v}$.
2. Expresar la proyección anterior como el producto de una matriz por el vector $\mathbf{u}$.
3. Sea $n = 3$. Escribir la matriz $\mathbf{P}$ del apartado anterior para el vector $\mathbf{v} = (1, 1, 1)^t$. ¿Cuál es su rango? Determinar bases ortogonales de $\ker \mathbf{P}$ e $\operatorname{Im} \mathbf{P}$.
4. Determinar una base ortonormal de $\mathbb{R}^3$ cuyo primer vector sea proporcional a $\mathbf{v} = (1, 1, 1)^t$.

3.12.– Sea a el menor entero positivo para el cual

$$\left(\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ a \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \right)$$

es una base de $\mathbb{R}^3$. Se pide ortonormalizarla por Gram-Schmidt.

3.13.– Determinar la factorización QR de la matriz:

$$\begin{pmatrix} -3 & 0 & 8 \\ 2 & -1 & -2 \\ 2 & 2 & -7 \\ 2 & 2 & -2 \\ 2 & 2 & -2 \end{pmatrix}.$$

3.14.– Calcular todos los posibles valores reales de a, b y c para que la siguiente matriz sea ortogonal:

$$\begin{pmatrix} 0 & 2b & c \\ a & b & -c \\ a & -b & c \end{pmatrix}.$$

3.15.– Sea $\mathbf{Q}$ una matriz ortogonal de segundo orden y determinante igual a 1.

1. Calcular su forma más general.
2. Interpretar el significado geométrico de la transformación $\mathbf{x} \mapsto \mathbf{Qx}$.

3.16.– Dados los vectores complejos $\mathbf{u} = \begin{pmatrix} 1 \\ i \end{pmatrix}$ y $\mathbf{v} = \begin{pmatrix} i \\ -1 \end{pmatrix}$, estúdiese si son ortogonales y compruébese que se cumple $\|\mathbf{u} + \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$.

3.17.– Sean $\mathbf{u}, \mathbf{v} \in \mathbb{C}^n$ tales que

$$\|\mathbf{u} + \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 \ \wedge \ \|\mathbf{u} + i\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$$

¿Puede concluirse que son ortogonales?

3.18.– Calcular todos los posibles valores reales de a y b para que la siguiente matriz sea unitaria:

$$\begin{pmatrix} a & bi \\ a & -bi \end{pmatrix}.$$

3.19.– Sea $\mathbf{U} = \mathbf{A} + i\mathbf{B}$, con $\mathbf{A}$ y $\mathbf{B}$ reales, una matriz compleja unitaria. Comprobar que la matriz

$$\mathbf{M} = \begin{pmatrix} \mathbf{A} & -\mathbf{B} \\ \mathbf{B} & \mathbf{A} \end{pmatrix}$$

es ortogonal.

3.4.1. Cuestiones

3.20.– En $\mathbb{C}^3$ se considera el plano $\mathcal{P}$ de ecuación $x + (1+i)y - 2z = 0$. Decidir sobre la veracidad o falsedad de las siguientes proposiciones.

1. Los vectores reales de $\mathcal{P}$ tienen segunda componente nula.
2. Si $\mathbf{u} \in \mathcal{P}$ entonces $\overline{\mathbf{u}} \in \mathcal{P}$.
3. El único vector real de $\mathcal{P}$ es el vector nulo.
4. El vector $(0, 1-i, 1)$ pertenece a $\mathcal{P}$.
5. El vector $(1, 1+i, -2)$ es ortogonal a $\mathcal{P}$.

3.21.– Sean F y G dos subespacios de $\mathbb{R}^n$. Decidir sobre la veracidad o falsedad de las siguientes proposiciones.

1. $F \subset G \Rightarrow F^\perp \subset G^\perp$
2. $F \subset G \Rightarrow G^\perp \subset F^\perp$
3. $\mathbb{R}^n = F \oplus G \Rightarrow \mathbb{R}^n = F^\perp \oplus G^\perp$
4. $\mathbb{R}^n = F \oplus G^\perp \Rightarrow F = G$

3.22.– Sea $\mathbf{A} \in \mathbb{R}^{n \times n}$ tal que $\langle \mathbf{u}, \mathbf{A}\mathbf{u} \rangle = 0$ cualquiera que sea $\mathbf{u} \in \mathbb{R}^n$. Decidir sobre la veracidad o falsedad de las siguientes afirmaciones.

1. $\mathbf{A}$ es nula.
2. Los elementos diagonales de $\mathbf{A}$ son nulos.
3. $\det \mathbf{A} = 0$.
4. $\forall i, j \quad a_{ij} = -a_{ji}$

5. $\mathbf{A}$ es antisimétrica.

6. $\mathbb{R}^n = \operatorname{Im} \mathbf{A} \oplus \ker \mathbf{A}$

3.23.– Sea $\mathbf{A} \in \mathbb{R}^{n\times n}$. Decidir sobre la veracidad o falsedad de las siguientes proposiciones.

1. $\mathbf{A}$ simétrica $\Rightarrow \mathbb{R}^n = \operatorname{Im} \mathbf{A} \oplus \ker \mathbf{A}$

2. $\mathbb{R}^n = \operatorname{Im} \mathbf{A} \oplus \ker \mathbf{A} \Rightarrow \mathbf{A}$ simétrica.

3. $\mathbf{A}$ antisimétrica $\Rightarrow \mathbb{R}^n = \operatorname{Im} \mathbf{A} \oplus \ker \mathbf{A}$

4. $\operatorname{Im} \mathbf{A} = (\ker \mathbf{A})^{\perp} \Rightarrow \mathbf{A}$ simétrica.

Capítulo 4

Proyecciones ortogonales y sus aplicaciones

4.1. Matriz de proyección ortogonal

En el capítulo anterior, se dio la definición de proyección ortogonal de un vector sobre un subespacio (Definición 42), que a su vez se basaba en el Teorema de la proyección ortogonal (Teorema 8). Recordémoslo aquí: Si M es un subespacio de $\mathbb{R}^n$, entonces para cada vector $\mathbf{u} \in \mathbb{R}^n$ existen únicos $\mathbf{v} \in M$, $\mathbf{w} \in M^\perp$ de forma que

$$\mathbf{u} = \mathbf{v} + \mathbf{w}.$$

El vector $\mathbf{v}$ recibe el nombre de **proyección ortogonal** de $\mathbf{u}$ sobre M. Asimismo, $\mathbf{w}$ es la **proyección ortogonal** de $\mathbf{u}$ sobre $M^\perp$.

A continuación vamos a deducir que este vector proyectado puede calcularse mediante una matriz, la llamada matriz de proyección ortogonal. Para ello, consideremos que M es un subespacio de $\mathbb{R}^n$ de dimensión m, y sea $\mathcal{B}_M = (\mathbf{u}_1, \ldots, \mathbf{u}_m)$ una base **ortogonal** de M; por otro lado, su subespacio ortogonal $M^\perp$ tiene dimensión $n - m$, y denotemos $\mathcal{B}_{M^\perp} = (\mathbf{u}_{m+1}, \ldots, \mathbf{u}_n)$ una base ortogonal de $M^\perp$. Entonces, la reunión de dichas bases es

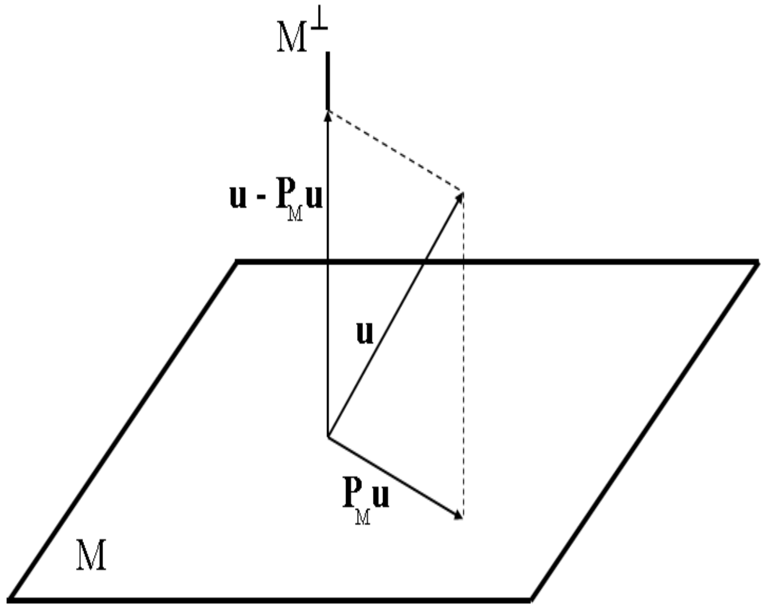

Figura 4.1: Proyección ortogonal de $\mathbf{u}$ sobre el subespacio M

una base ortogonal de $\mathbb{R}^n$, y según la ecuación (3.20), todo vector $\mathbf{u} \in \mathbb{R}^n$ puede escribirse como

$$\mathbf{u} = \sum_{j=1}^{m} \frac{\langle \mathbf{u}, \mathbf{u}_j \rangle}{\|\mathbf{u}_j\|^2} \mathbf{u}_j + \sum_{j=m+1}^{n} \frac{\langle \mathbf{u}, \mathbf{u}_j \rangle}{\|\mathbf{u}_j\|^2} \mathbf{u}_j.$$

Pues bien, esta expresión puede reescribirse como $\mathbf{u} = \mathbf{v} + \mathbf{w}$, donde el primer sumando $\mathbf{v}$ es un vector perteneciente a M (es combinación lineal de una base de M) y el segundo sumando $\mathbf{w}$ es un vector perteneciente a $M^{\perp}$ (es combinación lineal de una base de $M^{\perp}$). Por tanto, $\mathbf{v}$ es la proyección ortogonal de $\mathbf{u}$ sobre M (y $\mathbf{w}$ es la proyección ortogonal de $\mathbf{u}$ sobre $M^{\perp}$). En otras palabras, la proyección ortogonal de $\mathbf{u}$ sobre M puede calcularse de la siguiente forma: basta encontrar una base ortogonal de M, $(\mathbf{u}_1, \ldots, \mathbf{u}_m)$, y obtener el vector

$$\mathbf{v} = \sum_{j=1}^{m} \frac{\langle \mathbf{u}, \mathbf{u}_j \rangle}{\|\mathbf{u}_j\|^2} \mathbf{u}_j. \tag{4.1}$$

Paralelamente, la proyección ortogonal de $\mathbf{u}$ sobre $M^{\perp}$ es

$$\mathbf{w} = \sum_{j=m+1}^{n} \frac{\langle \mathbf{u}, \mathbf{u}_j \rangle}{\|\mathbf{u}_j\|^2} \mathbf{u}_j,$$

aunque también se puede calcular como $\mathbf{w} = \mathbf{u} - \mathbf{v}$.

Así pues, el vector proyectado puede expresarse como

$$\begin{aligned} \mathbf{v} &= \sum_{j=1}^{m} \mathbf{u}_j \frac{\langle \mathbf{u}, \mathbf{u}_j \rangle}{\|\mathbf{u}_j\|^2} = \sum_{j=1}^{m} \frac{\mathbf{u}_j (\mathbf{u}_j^t \mathbf{u})}{\|\mathbf{u}_j\|^2} = \\ &= \left(\sum_{j=1}^{m} \frac{\mathbf{u}_j \mathbf{u}_j^t}{\|\mathbf{u}_j\|^2} \right) \mathbf{u} \end{aligned}$$

donde hemos utilizado que $\langle \mathbf{u}, \mathbf{u}_j \rangle = \mathbf{u}_j^t \mathbf{u}$. Obsérvese que la última expresión es un producto de matriz (entre paréntesis) por el vector $\mathbf{u}$. Por tanto, hemos demostrado que existe una matriz, que denotaremos $\mathbf{P}_M$, con

$$\mathbf{P}_M = \sum_{j=1}^{m} \frac{\mathbf{u}_j \mathbf{u}_j^t}{\|\mathbf{u}_j\|^2} \tag{4.2}$$

tal que $\mathbf{v} = \mathbf{P}_M \mathbf{u}$. Esta matriz que proporciona el vector proyectado sobre M recibe el nombre de matriz de proyección ortogonal sobre M. Ahora que hemos demostrado su existencia, podemos definirla así:

Definición 50 (matriz de proyección ortogonal sobre un subespacio) *Sea M un subespacio de $\mathbb{R}^n$; se dice que $\mathbf{P}_M \in \mathbb{R}^{n \times n}$ es* ***matriz de proyección ortogonal sobre*** *M si, para cada $\mathbf{u} \in \mathbb{R}^n$, $\mathbf{P}_M \mathbf{u}$ es la proyección ortogonal de $\mathbf{u}$ sobre M.*

Una forma general de calcularla es mediante la expresión (4.2). En particular, la matriz de proyección ortogonal sobre la recta generada por un vector no nulo $\mathbf{v}$ es

$$\mathbf{P}_{L[\mathbf{v}]} = \frac{1}{\mathbf{v}^t \mathbf{v}} \mathbf{v} \mathbf{v}^t \tag{4.3}$$

En general, la ecuación (4.2) indica que la proyección ortogonal de un vector sobre un subespacio se obtiene sumando las proyecciones ortogonales de dicho vector sobre las rectas vectoriales generadas por los vectores de una base ortogonal de dicho subespacio.

A continuación, damos otras expresiones de la matriz de proyección ortogonal:

Proposición 23

- *Si $\mathbf{Q}$ es una matriz cuyas columnas forman* ***base ortonormal*** *del subespacio M, entonces la matriz de proyección ortogonal sobre M es*

$$\mathbf{P}_M = \mathbf{Q} \mathbf{Q}^t \tag{4.4}$$

- *Si $\mathbf{A}$ es una matriz cuyas columnas forman* ***base*** *del subespacio M, entonces la matriz de proyección ortogonal sobre M es*

$$\mathbf{P}_M = \mathbf{A} \left(\mathbf{A}^t \mathbf{A} \right)^{-1} \mathbf{A}^t. \tag{4.5}$$

Demostración: El primer apartado se deduce de la expresión (4.2), sin más que tomar

$$\mathbf{q}_j := \frac{\mathbf{u}_j}{\|\mathbf{u}_j\|}$$

con lo que los vectores $\mathbf{q}_j$ son ortogonales y unitarios:

$$\mathbf{P}_M = \sum_{j=1}^{m} \mathbf{q}_j \mathbf{q}_j^t = \begin{pmatrix} \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_m \end{pmatrix} \begin{pmatrix} \mathbf{q}_1^t \\ \mathbf{q}_2^t \\ \vdots \\ \mathbf{q}_m^t \end{pmatrix} = \mathbf{Q}\mathbf{Q}^t.$$

Para el segundo apartado: dado que las columnas de $\mathbf{A}$ forman base de M, recordemos que una forma de obtener una base ortonormal de M es la factorización QR de $\mathbf{A}$. En efecto, si $\mathbf{A} = \mathbf{QR}$ (con $\mathbf{Q}$ de columnas ortonormales y $\mathbf{R}$ triangular superior invertible) entonces las columnas de $\mathbf{Q}$ son base ortonormal de M, y por tanto, podemos aplicar la expresión (4.4) tomando $\mathbf{Q} = \mathbf{AR}^{-1}$:

$$\begin{aligned} \mathbf{P}_M &= \mathbf{QQ}^t = \left(\mathbf{AR}^{-1}\right)\left(\mathbf{AR}^{-1}\right)^t = \mathbf{AR}^{-1}\left(\mathbf{R}^{-1}\right)^t \mathbf{A}^t \\ &= \mathbf{A}\left(\mathbf{R}^t\mathbf{R}\right)^{-1}\mathbf{A}^t = \mathbf{A}\left(\mathbf{A}^t\mathbf{A}\right)^{-1}\mathbf{A}^t. \end{aligned}$$

En el último paso se ha utilizado que

$$\mathbf{A}^t\mathbf{A} = (\mathbf{QR})^t(\mathbf{QR}) = \mathbf{R}^t\mathbf{Q}^t\mathbf{QR} = \mathbf{R}^t\mathbf{R}$$

ya que $\mathbf{Q}^t\mathbf{Q} = \mathbf{I}$ al tener $\mathbf{Q}$ columnas ortonormales. Esto concluye la demostración. □

4.1.1. Propiedades de las matrices de proyección ortogonal

Se cumplen las siguientes propiedades:

- $\mathbf{P}_M$ es simétrica ($\mathbf{P}_M^t = \mathbf{P}_M$) e idempotente ($\mathbf{P}_M^2 = \mathbf{P}_M$), como se deduce de la anterior expresión (4.2) de $\mathbf{P}_M$.
- $\operatorname{Im}\mathbf{P}_M = M$ y $\ker\mathbf{P}_M = M^\perp$, por lo que $\operatorname{Im}\mathbf{P}_M$ y $\ker\mathbf{P}_M$ son subespacios suplementarios ortogonales.
- Si $\mathbf{P}_M$ es matriz de proyección ortogonal sobre M, $\mathbf{P}_{M^\perp} = \mathbf{I} - \mathbf{P}_M$ lo es sobre $M^\perp$.

Proposición 24 *Si una matriz $\mathbf{P} \in \mathbb{R}^{n\times n}$ es simétrica e idempotente, entonces es matriz de proyección ortogonal sobre su imagen.*

Demostración: En efecto, para cada $\mathbf{u} \in \mathbb{R}^n$ se tiene que $\mathbf{u} = \mathbf{Pu} + \mathbf{u} - \mathbf{Pu}$. Se cumplen $\mathbf{Pu} \in \operatorname{Im}\mathbf{P}$ y, por las propiedades de $\mathbf{P}$, $\mathbf{u} - \mathbf{Pu} \in \ker\mathbf{P} = (\operatorname{Im}\mathbf{P})^\perp$. Por tanto, $\mathbf{Pu}$ es la proyección ortogonal de $\mathbf{u}$ sobre $\operatorname{Im}\mathbf{P}$. □

Nótese que $\operatorname{Im}\mathbf{P} = \ker(\mathbf{I} - \mathbf{P})$.

Todo lo anterior es similar en el caso de $\mathbb{C}^n$ sustituyendo la trasposición por trasposición y conjugación, resultando la matriz de proyección ortogonal sobre un subespacio hermítica e idempotente.

4.2. El problema de mínimos cuadrados

Un modelo matemático es un esquema ideal que responde a un problema real. La búsqueda de solución de dicho problema conlleva la realización de observaciones, muchas veces con errores de medida. El planteamiento matemático de tal problema puede conducir, por ejemplo, a un sistema lineal incompatible. En esos casos se *corrige* el término independiente (que hace incompatible el sistema) y se sustituye por un término independiente para el cual el sistema tenga solución. A esta solución del sistema con término independiente modificado se le llama **pseudosolución** del sistema de partida.

4.2.1. Soluciones de mínimos cuadrados de un sistema

Sea $\mathbf{A} \in \mathbb{R}^{m\times n}$ y $\mathbf{b} \in \mathbb{R}^m$. En caso de que el sistema $\mathbf{Ax} = \mathbf{b}$ sea compatible, sus soluciones coinciden con aquellos valores de $\mathbf{x}$ para los cuales

$$\|\mathbf{Ax} - \mathbf{b}\| = 0.$$

Por otro lado, la compatibilidad del sistema equivale a que el término independiente $\mathbf{b}$ pertenezca a $\operatorname{Im}\mathbf{A}$. Si el sistema es incompatible ello se debe a que $\mathbf{b} \notin \operatorname{Im}\mathbf{A}$; hablando de forma intuitiva, la *distancia* de $\mathbf{b}$ a $\operatorname{Im}\mathbf{A}$ nos da una idea de cuan incompatible es el sistema. Recordemos que según la fórmula (3.15) dicha distancia es igual a

$$d(\mathbf{b}, \operatorname{Im}\mathbf{A}) = \|\mathbf{b} - \mathbf{P}_{\operatorname{Im}\mathbf{A}}\mathbf{b}\|.$$

Conforme a lo que indicábamos en la introducción a esta sección, cuando el sistema $\mathbf{Ax} = \mathbf{b}$ sea incompatible, vamos a *sustituir* el vector $\mathbf{b}$ por un nuevo término independiente para el cual el sistema sea compatible; pues bien, vamos a elegir como nuevo término independiente la proyección sobre la imagen de $\mathbf{A}$. De esta forma garantizamos, además, que los vectores $\mathbf{x}$ solución del sistema modificado son también los que hacen mínima la norma: $\|\mathbf{Ax} - \mathbf{b}\|$. Adoptaremos la siguiente:

Definición 51 (solución de mínimos cuadrados) *Dado un sistema lineal* $\mathbf{Ax} = \mathbf{b}$*, se denomina* ***solución de mínimos cuadrados*** *o* ***pseudosolución*** *a cualquier solución del sistema* $\mathbf{Ax} = \mathbf{P}_{\operatorname{Im}\mathbf{A}}\mathbf{b}$

Observación: Si el sistema $\mathbf{Ax} = \mathbf{b}$ es compatible, entonces sus soluciones coinciden con sus soluciones de mínimos cuadrados.

Definición 52 (vector residuo) *Se define el vector* ***residuo*** *del sistema* $\mathbf{Ax} = \mathbf{b}$ *como* $\mathbf{r} = \mathbf{b} - \mathbf{P}_{\operatorname{Im}\mathbf{A}}\mathbf{b}$.

Definición 53 (ecuaciones normales) *Dado el sistema* $\mathbf{Ax} = \mathbf{b}$*, sus* ***ecuaciones normales*** *vienen dadas por el sistema*

$$\mathbf{A}^t\mathbf{Ax} = \mathbf{A}^t\mathbf{b}. \tag{4.6}$$

Observación: El sistema de ecuaciones normales $\mathbf{A}^t\mathbf{A}\mathbf{x} = \mathbf{A}^t\mathbf{b}$ es siempre compatible, ya que $\operatorname{Im}\mathbf{A}^t = \operatorname{Im}(\mathbf{A}^t\mathbf{A})$.

Proposición 25 $\mathbf{x}_0$ *es solución de mínimos cuadrados del sistema* $\mathbf{A}\mathbf{x} = \mathbf{b}$ *si, y solo si,* $\mathbf{x}_0$ *es solución de las ecuaciones normales de dicho sistema.*

Demostración: Sabemos que el espacio $\mathbb{R}^m = \operatorname{Im}\mathbf{A} \oplus \ker\mathbf{A}^t$; descomponemos el vector $\mathbf{b}$ según esta suma directa,

$$\mathbf{b} = \mathbf{P}_{\operatorname{Im}\mathbf{A}}\mathbf{b} + \mathbf{w}, \quad \mathbf{w} \in \ker\mathbf{A}^t$$

Entonces:

$$\mathbf{A}^t\mathbf{A}\mathbf{x}_0 = \mathbf{A}^t\mathbf{b} \Leftrightarrow \mathbf{A}^t\mathbf{A}\mathbf{x}_0 = \mathbf{A}^t(\mathbf{P}_{\operatorname{Im}\mathbf{A}}\mathbf{b} + \mathbf{w}) = \mathbf{A}^t\mathbf{P}_{\operatorname{Im}\mathbf{A}}\mathbf{b}$$

$$\Leftrightarrow \mathbf{A}^t(\mathbf{A}\mathbf{x}_0 - \mathbf{P}_{\operatorname{Im}\mathbf{A}}\mathbf{b}) = \mathbf{0} \Leftrightarrow \mathbf{A}\mathbf{x}_0 - \mathbf{P}_{\operatorname{Im}\mathbf{A}}\mathbf{b} \in \operatorname{Im}\mathbf{A} \cap \ker\mathbf{A}^t = \{\mathbf{0}\}$$

$$\Leftrightarrow \mathbf{A}\mathbf{x}_0 = \mathbf{P}_{\operatorname{Im}\mathbf{A}}\mathbf{b}$$

$\square$

Observación: La solución $\mathbf{x}_0$ de mínimos cuadrados es única si y sólo si el rango de $\mathbf{A}$ es igual a n. Y, en este caso, $\mathbf{x}_0 = (\mathbf{A}^t\mathbf{A})^{-1}\mathbf{A}^t\mathbf{b}$.

4.2.2. Solución de mínima norma de un sistema compatible indeterminado

Sean $\mathbf{A} \in \mathbb{R}^{m\times n}$ y $\mathbf{A}\mathbf{x} = \mathbf{b}$ sistema compatible indeterminado. Se trata de encontrar, entre todas sus soluciones, aquella cuya norma sea mínima.

Sea $\mathbf{x}_p$ una solución de $\mathbf{A}\mathbf{x} = \mathbf{b}$. Todas las soluciones del sistema anterior constituyen el conjunto $\{\mathbf{x}_p + \mathbf{u} : \mathbf{u} \in \ker\mathbf{A}\}$. Por el teorema de la Proyección Ortogonal, visto en el capítulo 3, la solución de mínima norma será $\mathbf{x}_0 = \mathbf{x}_p - \mathbf{P}_{\ker\mathbf{A}}\mathbf{x}_p \in (\ker\mathbf{A})^{\perp} = \operatorname{Im}\mathbf{A}^t$, y es única.

La solución de mínima norma puede obtenerse por uno cualquiera de los siguientes métodos:

1. Proyectando ortogonalmente cualquiera de las soluciones del sistema sobre $\operatorname{Im}\mathbf{A}^t$.

2. Determinando aquella solución que pertenezca a $\operatorname{Im}\mathbf{A}^t$.

3. Eligiendo, entre todas las soluciones del sistema, aquella que sea ortogonal al $\ker\mathbf{A}$.

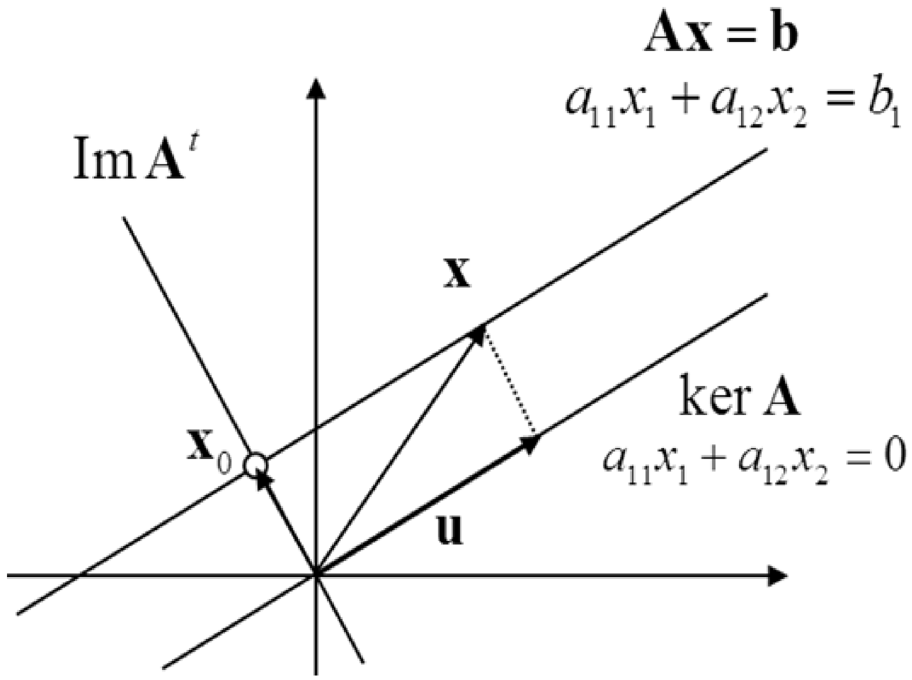

Figura 4.2: Solución $\mathbf{x}_0$ de mínima norma de un sistema compatible indeterminado

De los tres procedimientos, el más rápido suele ser el segundo, que se detalla a continuación:

Buscamos un vector $\mathbf{x}_0$ que satisfaga

$$\mathbf{A}\mathbf{x}_0 = \mathbf{b} \tag{4.7}$$

y tal que $\mathbf{x}_0 \in \mathrm{Im}\,\mathbf{A}^t$; es decir, será una solución de la forma $\mathbf{x}_0 = \mathbf{A}^t\mathbf{u}$ para un cierto $\mathbf{u} \in \mathbb{R}^m$. Sustituyendo esta expresión en (4.7), resulta el sistema $m \times m$

$$\mathbf{A}\mathbf{A}^t\mathbf{u} = \mathbf{b} \tag{4.8}$$

del que hallaremos una solución $\mathbf{u}_0$ que, finalmente, nos permitirá obtener la solución de mínima norma como $\mathbf{x}_0 = \mathbf{A}^t\mathbf{u}_0$.

Obsérvese que, en caso de ser indeterminado el sistema (4.8), solamente es necesario el cálculo de una cualquiera de sus soluciones.

Por otro lado, en los casos "típicos"de sistemas lineales indeterminados, es decir, cuando $m < n$, el sistema (4.8) reduce el número de incógnitas a m, facilitando la resolución.

Observación: Si $r(\mathbf{A}) = m < n$, la solución de mínima norma puede expresarse como $\mathbf{x}_0 = \mathbf{A}^t(\mathbf{A}\mathbf{A}^t)^{-1}\mathbf{b}$. Para llegar al resultado anterior, basta tener en cuenta que $\mathbf{x}_0 \in \mathrm{Im}\,\mathbf{A}^t$ conduce a que existe $\mathbf{v} \in \mathbb{R}^m$ tal que $\mathbf{A}^t\mathbf{v} = \mathbf{x}_0$; por otra parte, $\mathbf{A}\mathbf{x}_0 = \mathbf{b}$, de modo que $\mathbf{A}\mathbf{A}^t\mathbf{v} = \mathbf{b}$, resultando $\mathbf{v} = \left(\mathbf{A}\mathbf{A}^t\right)^{-1}\mathbf{b}$, ya que $\mathbf{A}\mathbf{A}^t$ tiene rango igual a m. Finalmente, $\mathbf{x}_0 = \mathbf{A}^t(\mathbf{A}\mathbf{A}^t)^{-1}\mathbf{b}$.

Hay que tener en cuenta que este resultado sólo es válido si el rango de $\mathbf{A}$ es igual a m. En caso contrario, la matriz $\mathbf{A}^t\mathbf{A}$, cuadrada de orden m, no es invertible, ya que su rango, igual al de $\mathbf{A}$, es menor que m.

4.2.3. Solución de mínimos cuadrados y mínima norma de un sistema

En el caso general, las soluciones de mínimos cuadrados del sistema $\mathbf{Ax} = \mathbf{b}$ son las soluciones de sus ecuaciones normales $\mathbf{A}^t\mathbf{Ax} = \mathbf{A}^t\mathbf{b}$, sistema, como se ha visto, compatible. Para determinar, entre ellas, la de mínima norma, que está univocamente determinada, basta tener en cuenta, además, el resultado ya visto en el apartado anterior. Por tanto, la solución de mínimos cuadrados y mínima norma del sistema $\mathbf{Ax} = \mathbf{b}$ es $\mathbf{x}_0 \in \mathbb{R}^n$ tal que $\mathbf{A}^t\mathbf{Ax}_0 = \mathbf{A}^t\mathbf{b}$ y $\mathbf{x}_0 \in (\ker \mathbf{A}^t\mathbf{A})^\perp = (\ker \mathbf{A})^\perp = \operatorname{Im} \mathbf{A}^t$.

Todo lo visto en este apartado es similar en el caso de $\mathbf{A} \in \mathbb{C}^{m\times n}$, sustituyendo la trasposición por trasposición y conjugación.

4.3. Matriz de simetría ortogonal

Definición 54 (matriz de simetría ortogonal respecto a un subespacio) *Sea M un subespacio de $\mathbb{R}^n$ y $\mathbf{P}_M$ la matriz de la proyección ortogonal sobre M. Entonces*

$$\mathbf{S}_M := 2\mathbf{P}_M - \mathbf{I} \tag{4.9}$$

es la matriz de simetría ortogonal respecto de M.

4.3.1. Propiedades de las matrices de simetría ortogonal

Se cumplen las siguientes propiedades:

- $\mathbf{S}_M$ es simétrica ($\mathbf{S}_M = \mathbf{S}_M^t$) y ortogonal ($\mathbf{S}_M\mathbf{S}_M^t = \mathbf{I}$). Es decir, $\mathbf{S}_M^{-1} = \mathbf{S}_M = \mathbf{S}_M^t$.
- $\mathbf{S}_M = \mathbf{P}_M - \mathbf{P}_{M^\perp}$
- Si $\mathbf{S}_M$ es matriz de simetría ortogonal respecto a M, entonces $-\mathbf{S}_M$ lo es respecto a $M^\perp$.
- $M = \ker(\mathbf{S}_M - \mathbf{I}) = \operatorname{Im}(\mathbf{S}_M + \mathbf{I})$
- $M^\perp = \ker(\mathbf{S}_M + \mathbf{I}) = \operatorname{Im}(\mathbf{S}_M - \mathbf{I})$
- En particular, si $M = L[\mathbf{v}]$ es la recta engendrada por un vector no nulo $\mathbf{v}$, la matriz de simetría ortogonal es:

$$\mathbf{S}_{L[\mathbf{v}]} = \frac{2}{\mathbf{v}^t\mathbf{v}}\mathbf{v}\mathbf{v}^t - \mathbf{I} \tag{4.10}$$

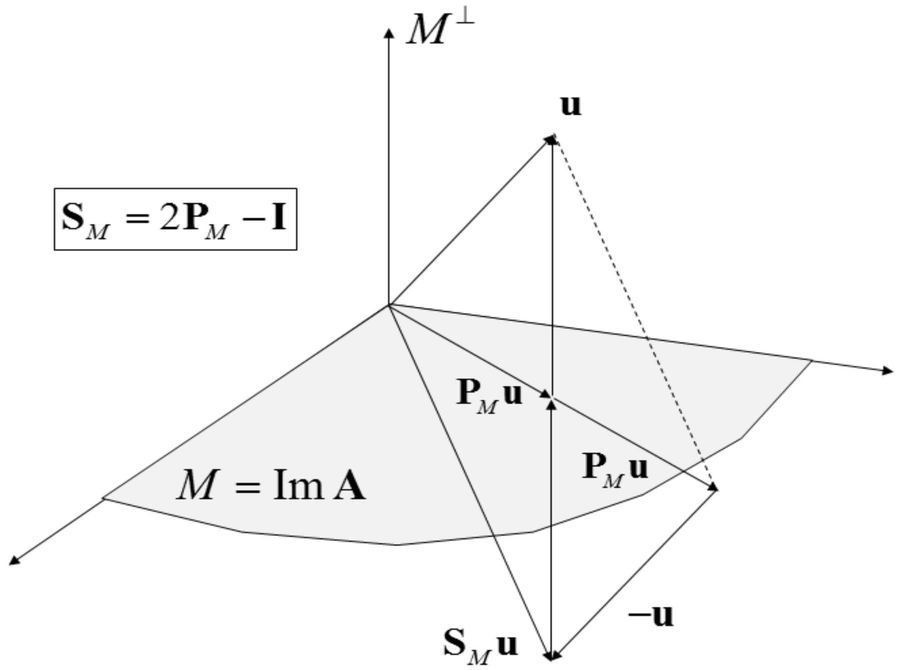

Figura 4.3: Matriz de simetría ortogonal

Proposición 26 *Si una matriz* $\mathbf{S}$ *es simétrica y ortogonal, entonces es matriz de simetría ortogonal respecto a* $\operatorname{Im}(\mathbf{S}+\mathbf{I})$.

Demostración: Para que $\mathbf{S}$ sea la matriz de una simetría ortogonal (véase la relación (4.9)) es necesario y suficiente que

$$\mathbf{P} := \frac{1}{2}(\mathbf{S}+\mathbf{I})$$

sea una proyección ortogonal; equivalentemente, por la proposición 24, $\mathbf{P} = \mathbf{P}^2 = \mathbf{P}^t$. De la simetría de $\mathbf{S}$ se deduce inmediatamente la simetría de $\mathbf{P}$, en cuanto a la idempotencia:

$$\mathbf{P}^2 = \frac{1}{4}(\mathbf{S}+\mathbf{I})^2 = \frac{1}{4}\left(\mathbf{S}^2 + 2\mathbf{S} + \mathbf{I}\right) = \frac{1}{2}(\mathbf{S}+\mathbf{I}) = \mathbf{P}.$$

Por último, $\mathbf{P}$ es matriz de proyección sobre su imagen, pero $\operatorname{Im}\mathbf{P} = \operatorname{Im}(\mathbf{S}+\mathbf{I})$. □

4.3.2. Matriz de Householder

Sea H un hiperplano de $\mathbb{R}^n$ (es decir, un subespacio de dimensión $n-1$), de ecuación $\mathbf{v}^t\mathbf{x} = 0$. La matriz de simetría ortogonal respecto a dicho hiperplano es, por lo visto anteriormente –fórmula (4.10)–, $\mathbf{S}_H = -\mathbf{S}_{L[\mathbf{v}]}$. Dicha matriz de simetría se llama **matriz de Householder**, y está definida por el vector $\mathbf{v}$, no nulo, ortogonal a H. Su expresión es

$$\mathbf{H}_\mathbf{v} = \mathbf{I} - \frac{2}{\mathbf{v}^t\mathbf{v}}\mathbf{v}\mathbf{v}^t \tag{4.11}$$

Dados dos vectores no nulos cualesquiera $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$, existe una matriz de Householder que transforma al primero en proporcional al segundo; es decir, existe $\mathbf{v} \in \mathbb{R}^n$ tal que $\mathbf{H_v a} = \frac{\|\mathbf{a}\|}{\|\mathbf{b}\|}\mathbf{b}$ (recuérdese que $\mathbf{a}$ y su simétrico deben tener normas iguales). Puesto que la simetría se realiza respecto al hiperplano ortogonal al vector diferencia entre $\mathbf{a}$ y su simétrico, debemos tomar $\mathbf{v} = \mathbf{a} - \frac{\|\mathbf{a}\|}{\|\mathbf{b}\|}\mathbf{b}$.

Del mismo modo, si tomamos $\mathbf{w} = \mathbf{a} + \frac{\|\mathbf{a}\|}{\|\mathbf{b}\|}\mathbf{b}$, obtendremos que $\mathbf{H_w a} = -\frac{\|\mathbf{a}\|}{\|\mathbf{b}\|}\mathbf{b}$.

Esta idea se aplica frecuentemente en Álgebra Numérica; por ejemplo, cuando se desea que un determinado vector $\mathbf{a} \neq \mathbf{0}$ se transforme en un vector con todas sus componentes nulas excepto la primera, es decir, en $\|\mathbf{a}\|\mathbf{e}_1$, o bien con todas nulas excepto las dos primeras, etc. Ello permite triangularizar una matriz mediante transformaciones ortogonales de sus columnas.

Todo lo visto en este apartado es similar en el caso de $\mathbb{C}^n$, sustituyendo la trasposición por trasposición y conjugación, resultando las matrices de simetría ortogonal hermíticas y unitarias.

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

Definición 55 (producto vectorial de dos vectores en $\mathbb{R}^3$) *Dados dos vectores* $\mathbf{u} = (u_1, u_2, u_3)$ *y* $\mathbf{v} = (v_1, v_2, v_3)$ *de* $\mathbb{R}^3$*, su producto vectorial* $\mathbf{u} \times \mathbf{v}$ *es*

$$\mathbf{u} \times \mathbf{v} = (u_2v_3 - u_3v_2, u_3v_1 - u_1v_3, u_1v_2 - u_2v_1).$$

Esta expresión quizá no sea fácil de recordar de memoria, por lo que suele calcularse desarrollando por la primera fila (formada por los vectores canónicos) el determinante *simbólico*

$$\begin{vmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix} = (u_2v_3 - u_3v_2)\mathbf{e}_1 + (u_3v_1 - u_1v_3)\mathbf{e}_2 + (u_1v_2 - u_2v_1)\mathbf{e}_3$$

4.4.1. Propiedades del producto vectorial

Sean $\mathbf{u}, \mathbf{v}, \mathbf{w}$ vectores cualesquiera de $\mathbb{R}^3$.

1. $\mathbf{u} \times \mathbf{v} = \mathbf{0} \iff \mathbf{u}$ y $\mathbf{v}$ son proporcionales

2. $\mathbf{u} \times \mathbf{v} = -\mathbf{v} \times \mathbf{u}$

3. $\mathbf{u} \perp \mathbf{v} \iff \|\mathbf{u} \times \mathbf{v}\| = \|\mathbf{u}\|\|\mathbf{v}\|$

4. Producto mixto: $\langle \mathbf{u} \times \mathbf{v}, \mathbf{w} \rangle = \mathbf{w}^t (\mathbf{u} \times \mathbf{v}) = \det\, (\mathbf{u}|\mathbf{v}|\mathbf{w})$

4.4.2. El producto vectorial en forma matricial

Sea $\mathbf{u} \in \mathbb{R}^3$. Se cumple que, para cada vector $\mathbf{v} \in \mathbb{R}^3$

$$\mathbf{u} \times \mathbf{v} = \begin{pmatrix} 0 & -u_3 & u_2 \\ u_3 & 0 & -u_1 \\ -u_2 & u_1 & 0 \end{pmatrix} \mathbf{v} \tag{4.12}$$

La matriz anterior se denotará $\widetilde{\mathbf{u}}$.

Observación: $\ker \widetilde{\mathbf{u}} = L[\mathbf{u}]$.

4.5. Giros en $\mathbb{R}^2$ y $\mathbb{R}^3$

Definición 56 (Matriz de giro) $\mathbf{G} \in \mathbb{R}^{n\times n}$ $(n = 2, 3)$ *es un una matriz de* ***giro*** *si es ortogonal y su determinante es igual a la unidad.*

4.5.1. Matriz de giro en $\mathbb{R}^2$

Sea $\mathbf{G} \in \mathbb{R}^{2\times 2}$, ortogonal y de determinante igual a 1. Conforme al ejercicio 3.15, dicha matriz será de la forma $\mathbf{G} = \begin{pmatrix} \cos\varphi & -\operatorname{sen}\varphi \\ \operatorname{sen}\varphi & \cos\varphi \end{pmatrix}$, donde $\varphi \in [0, 2\pi)$. Se tendrá, por tanto, que $\mathbf{G}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} \cos\varphi \\ \operatorname{sen}\varphi \end{pmatrix}$ y $\mathbf{G}\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} -\operatorname{sen}\varphi \\ \cos\varphi \end{pmatrix}$. Estos resultados tienen la siguiente interpretación geométrica: se trata de un giro, en sentido contrario a las agujas del reloj, alrededor del origen de $\mathbb{R}^2$ de ángulo de giro igual a φ, tal como se indica en la figura 4.4.

4.5.2. Matriz de giro en $\mathbb{R}^3$

Se comenzará utilizando el concepto geométrico de un giro de ángulo φ, alrededor de un eje orientado en uno de sus dos sentidos. En el caso de un vector, basta proyectar ortogonalmente dicho vector sobre el plano ortogonal al eje, efectuar el giro de ángulo φ de dicha proyección, en dicho plano, según la regla de avance del sacacorchos en el sentido del eje, y sumarle la proyección ortogonal del vector sobre el eje (véase la figura 4.5).

4.5.3. Cálculo de la matriz de giro en función de la proyección sobre el eje de giro (*no exigible*)

Si se considera $\mathbf{P} = \mathbf{u}\mathbf{u}^t$, matriz de proyección ortogonal sobre la recta $L[\mathbf{u}]$, donde $\mathbf{u}$ es unitario, todo vector $\mathbf{r}$ puede descomponerse como $\mathbf{r} = \mathbf{P}\mathbf{r} + (\mathbf{I} - \mathbf{P})\mathbf{r}$, en donde la primera componente permanece invariante mediante el giro de ángulo φ y eje de giro $L[\mathbf{u}]$ y la segunda, perteneciente al plano ortogonal al eje de giro, al girar se transforma en $\cos\varphi(\mathbf{I} - \mathbf{P})\mathbf{r} + \operatorname{sen}\varphi\ \mathbf{u} \times (\mathbf{I} - \mathbf{P})\mathbf{r}$.

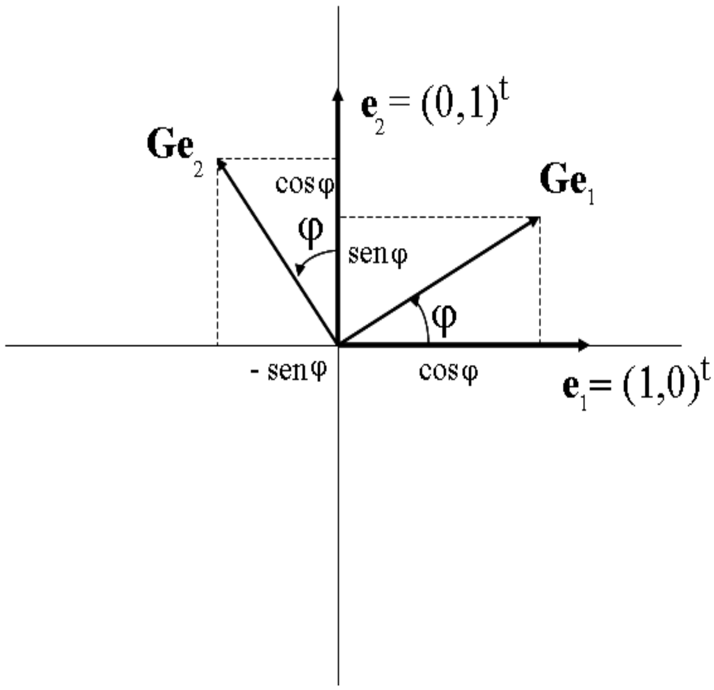

Figura 4.4: Giro en el plano

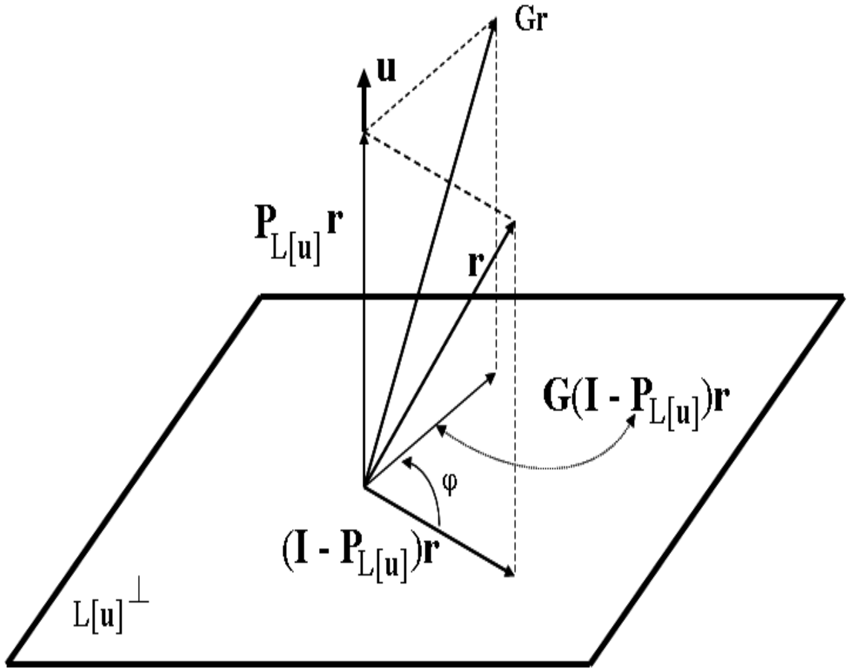

Figura 4.5: Giro en el espacio en función de la proyección sobre el eje de giro

Así, pues, la matriz del giro puede escribirse en la forma

$$\mathbf{G} = \mathbf{P} + \cos\varphi\ (\mathbf{I} - \mathbf{P}) + \operatorname{sen}\varphi\ \tilde{\mathbf{u}}(\mathbf{I} - \mathbf{P})$$

y, dado que $\tilde{\mathbf{u}}\mathbf{P} = \tilde{\mathbf{u}}\mathbf{u}\mathbf{u}^t = \mathbf{O}$, ya que $\mathbf{u} \in \ker \tilde{\mathbf{u}}$, se tiene, finalmente

$$\mathbf{G} = \mathbf{P} + \cos\varphi\ (\mathbf{I} - \mathbf{P}) + \operatorname{sen}\varphi \tilde{\mathbf{u}} \tag{4.13}$$

Para demostrar que $\mathbf{G}$ es ortogonal de determinante igual a 1, considérese una base ortonormal de $\mathbb{R}^3$, $(\mathbf{u}, \mathbf{v}, \mathbf{w})$, donde $\mathbf{u}$ es el vector unitario anterior, y $\mathbf{v}$ y $\mathbf{w}$ son dos vectores unitarios ortogonales pertenecientes al plano ortogonal a $L[\mathbf{u}]$, tales que $\det(\mathbf{u}|\mathbf{v}|\mathbf{w}) = 1$, lo cual equivale a que $\mathbf{v}\times\mathbf{w} = \mathbf{u}$.

Según la expresión 4.13 de $\mathbf{G}$ se verifica $\begin{cases} \mathbf{G}\mathbf{u} = \mathbf{u} \\ \mathbf{G}\mathbf{v} = \cos\varphi\mathbf{v} + \operatorname{sen}\varphi\mathbf{w} \\ \mathbf{G}\mathbf{w} = -\operatorname{sen}\varphi\mathbf{v} + \cos\varphi\mathbf{w} \end{cases}$, de modo que

$$\mathbf{G} = \begin{pmatrix} \mathbf{u} & \mathbf{v} & \mathbf{w} \end{pmatrix}^t \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\varphi & -\operatorname{sen}\varphi \\ 0 & \operatorname{sen}\varphi & \cos\varphi \end{pmatrix} \begin{pmatrix} \mathbf{u} & \mathbf{v} & \mathbf{w} \end{pmatrix}$$

de donde se deduce que $\mathbf{G}$ es ortogonal, al serlo las matrices factores, y $\det \mathbf{G} = 1$.

El ejercicio 6.8 demuestra el recíproco; es decir, que toda matriz ortogonal de orden 3 y determinante igual a 1, es un giro alrededor de un eje de ángulo de giro igual a $\varphi \in [0, 2\pi)$.

4.6. Ejercicios

4.1.– Hállese la matriz de proyección ortogonal sobre la recta $x = y = 2z$. Calcúlese asimismo la matriz de simetría ortogonal respecto a esa misma recta.

4.2.– Escríbanse la matriz de proyección ortogonal sobre el plano $2x + 2y + z = 0$, y la matriz de simetría ortogonal respecto a dicho plano.

4.3.– Hállese la matriz de proyección ortogonal sobre el subespacio de $\mathbb{R}^4$ determinado por las ecuaciones

$$\left\{ \begin{array}{rcl} x_1 + x_2 \qquad\qquad & = & 0, \\ x_3 + x_4 & = & 0. \end{array} \right.$$

4.4.– Sea $\mathbf{P} = a \begin{pmatrix} b & c & -2 \\ -4 & 5 & d \\ -2 & -2 & 8 \end{pmatrix}$ una matriz no nula.

1. Calcúlense los valores reales a, b, c y d para que $\mathbf{P}$ sea la matriz de una proyección ortogonal en $\mathbb{R}^3$.

2. Siendo $\mathbf{P}$ la matriz obtenida a partir de los resultados del apartado 1, determínese una base ortonormal del subespacio M de $\mathbb{R}^3$ sobre el que proyecta $\mathbf{P}$.

3. Siendo $\mathbf{P}$ la matriz obtenida a partir de los resultados del apartado 1, determínese la matriz de proyección ortogonal sobre $M^{\perp}$.

4.5.– Determínense las matrices de Householder tales que

1. $\mathbf{H_v}(2, 2, -1)^t = (3, 0, 0)^t$

2. $\mathbf{H_w}(2, 2, -1)^t = (-3, 0, 0)^t$.

Estúdiese la relación que existe entre los vectores $\mathbf{v}$ y $\mathbf{w}$.

4.6.– Calcúlese la solución de mínimos cuadrados del sistema

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}$$

4.7.– Determínese la solución de mínimos cuadrados del sistema

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

4.8.– Dados los puntos del plano de coordenadas (x_k, y_k) con $k = 1, 2, \ldots, n$, la recta de regresión de y sobre x es la recta de ecuación $y = mx + p$, cuyas pendiente m y ordenada en el origen p minimizan la expresión $\sum_{k=1}^{n}(y_k - mx_k - p)^2$. Dedúzcase la expresión explícita de los coeficientes m y p. Hállese la recta de regresión de los puntos $(2, 1), (4, 2), (6, 3), (8, 1), (10, 2)$.

4.9.– Determínese la solución de mínima norma del sistema de ecuaciones

$$\left\{ \begin{array}{l} x_1 + x_2 - x_3 + x_4 = 2 \\ x_1 - x_2 + x_3 - x_4 = 2 \end{array} \right. .$$

4.10.– Calcúlese la solución de mínimos cuadrados y mínima norma del sistema

$$\begin{pmatrix} 1 & 1 & -1 \\ 2 & 1 & -1 \\ 2 & -1 & 1 \\ 1 & -1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 3 \\ 2 \\ 0 \end{pmatrix}.$$

4.11.–

1. Calcúlese la proyección ortogonal del vector $(1, 1, 1)$ sobre el subespacio

$$F = L\left[(1, 1, 0), (1, 2, 1), (0, 1, 1)\right]$$

2. Calcúlese la solución de mínimos cuadrados y mínima norma del sistema

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

4.12.– Sea $\mathbf{S} = a \begin{pmatrix} 1 & b & -8 \\ c & 7 & -4 \\ d & -4 & 1 \end{pmatrix} \in \mathbb{R}^{3\times 3}$.

1. Calcúlense los valores reales a, b, c y d para que $\mathbf{S}$ sea la matriz de una simetría ortogonal en $\mathbb{R}^3$.

2. Siendo $\mathbf{S}$ la matriz obtenida a partir de los resultados del apartado 1 con $a > 0$, determínese una base ortonormal del subespacio L de $\mathbb{R}^3$ respecto al que $\mathbf{S}$ simetriza.

3. Siendo $\mathbf{S}$ la matriz obtenida a partir de los resultados del apartado 1 con $a < 0$, determínense unas ecuaciones cartesianas del subespacio M de $\mathbb{R}^3$ respecto al que $\mathbf{S}$ simetriza.

4. Estúdiese si M es el suplemento ortogonal de L en $\mathbb{R}^3$.

5. Estúdiese en cuál de los dos casos anteriores (2. o 3.) es $\mathbf{S}$ una matriz de Householder y determínese un vector $\mathbf{v}$ que la defina.

4.13.– Calcúlese la matriz de giro de ángulo $\frac{2\pi}{3}$ en torno al vector $(1, 1, 1)$.

4.6.1. Cuestiones

4.14.– Márquese la única proposición falsa.
Si $\mathbf{P} \in \mathbb{R}^{n\times n}$ es matriz de proyección ortogonal, entonces

1. $\mathbb{R}^n = \ker \mathbf{P} \oplus \operatorname{Im} \mathbf{P}$
2. $\mathbf{P}^t = \mathbf{P}^2$
3. $(\ker \mathbf{P})^\perp = \operatorname{Im}(\mathbf{I} - \mathbf{P})$
4. $\forall \mathbf{u} \in \operatorname{Im}(\mathbf{I} - \mathbf{P})$, $\mathbf{Pu} = \mathbf{0}$

4.15.– Indíquense las afirmaciones que son verdaderas y las que son falsas.
Si M es un subespacio vectorial de $\mathbb{R}^n$, entonces

1. $\mathbf{S}_M = \mathbf{I} - 2\mathbf{P}_M$
2. $\mathbf{P}_M + \mathbf{P}_{M^\perp} = \mathbf{I}$
3. $\mathbf{S}_{M^\perp}\mathbf{P}_M = -\mathbf{P}_M$
4. $\mathbf{P}_{M^\perp}\mathbf{P}_M = \mathbf{O}$
5. $\mathbf{S}_{M^\perp}\mathbf{S}_M = -\mathbf{I}$

4.16.– Demuéstrese la siguiente proposición si es verdadera o dese un contraejemplo si es falsa: "Si $\mathbf{P} \in \mathbb{R}^{n\times n}$ es matriz de proyección ortogonal y $\mathbf{b} \in \mathbb{R}^n$, entonces $\mathbf{Pb}$ es la solución de mínimos cuadrados y mínima norma del sistema $\mathbf{Px} = \mathbf{b}$".

4.17.– Estúdiese si la siguiente proposición es verdadera: "Sea $\mathbf{v} \in \mathbb{R}^n$, no nulo, y $\mathbf{M} = \frac{2}{\mathbf{v}^t\mathbf{v}}\mathbf{v}\mathbf{v}^t - \mathbf{I}$, entonces $\mathbf{M}$ es una matriz de Householder".

Capítulo 5

Reducción por semejanza de una matriz

5.1. Introducción

En este capítulo estudiaremos condiciones necesarias y suficientes para que una matriz sea semejante a una matriz diagonal (o dicho de otra forma, que pueda representarse por una matriz diagonal), lo que equivale a encontrar una base de vectores que se transforman en proporcionales a sí mismos. Un vector no nulo $\mathbf{x}$ es vector propio de $\mathbf{A}$ si se transforma en su propia dirección, es decir si se cumple que $\mathbf{Ax} = \lambda\mathbf{x}$. El escalar λ es el valor propio asociado a $\mathbf{x}$.

En gran número de aplicaciones los valores y vectores propios tienen un claro significado geométrico. Consideremos por ejemplo la **matriz de simetría** respecto a un plano en $\mathbb{R}^3$. Todos los vectores de ese plano coinciden con sus transformados, luego son vectores propios asociados con el valor propio 1 ($\mathbf{Ax} = \mathbf{x}$). Sin embargo, un vector perpendicular al plano de simetría mantiene su dirección y cambia de sentido ($\mathbf{Ax} = -\mathbf{x}$), luego es un vector propio asociado con el valor propio -1.

Consideremos ahora la **matriz de giro** que represente por ejemplo la apertura de una puerta en $\mathbb{R}^3$. ¿Hay algún vector asociado a la puerta que no cambie de dirección al girar la puerta 60^o? Sí, los del eje de la puerta, que serán vectores propios asociados con el valor propio 1. Es fácil ver que no hay ningún otro vector asociado a la puerta que no cambie de dirección con ese giro.

El cálculo de valores y vectores propios tiene una gran importancia en muchas áreas de la ingeniería, como por ejemplo al **análisis sísmico** de estructuras. Otro ejemplo interesante es el algoritmo original de **Google** para ordenar los resultados de las búsquedas. Este algoritmo, denominado **PageRank**, está basado en el cálculo del vector propio asociado con el máximo valor propio de una matriz con tantas filas y columnas como páginas web hay en Internet, y formada a partir de los enlaces de llegada y d salida que ralacionan cada página con las restantes.

Nota: Veremos a lo largo del tema que los valores propios están relacionados con las raíces de un cierto polinomio; en virtud del teorema fundamental del Álgebra (véase el apéndice 3) los polinomios complejos admiten tantas raíces como su grado indica. Para garantizar la existencia de valores propios contemplaremos únicamente matrices complejas, si bien en los ejemplos casi siempre serán reales pero vistas como complejas con parte imaginaria nula.

5.2. Matrices semejantes y matrices diagonalizables

Definición 57 *Dos matrices cuadradas* $\mathbf{A}$ *y* $\mathbf{B}$ *se dicen* ***semejantes*** *si existe una matriz invertible* $\mathbf{P}$ *tal que* $\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{B}$.

Definición 58 *Una matriz cuadrada* $\mathbf{A}$ *es* ***diagonalizable*** *cuando es semejante a una matriz diagonal, es decir cuando existe una matriz* $\mathbf{P}$ *invertible tal que* $\mathbf{P}^{-1}\mathbf{A}\mathbf{P}$ *es diagonal.*

Las matrices semejantes comparten muchas propiedades, como se verá en detalle a lo largo de este tema. Por ejemplo, a partir de la definición es fácil observar que tienen el mismo determinante.

5.3. Valores y vectores propios

Definición 59 (valor y vector propio) *Sea* $\mathbf{A}$ *una matriz cuadrada de orden* n.

- *Se dice que un escalar* λ *de* $\mathbb{C}$ *es un* ***valor propio*** *o autovalor de la matriz* $\mathbf{A}$ *si existe un vector* $\mathbf{v}$*, perteneciente a* $\mathbb{C}^n$*,* ***no nulo*** *y tal que* $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$.
- *Se dice que un vector* $\mathbf{v}$ *de* $\mathbb{C}^n$ *es un* ***vector propio*** *o autovector de la matriz* $\mathbf{A}$ *si* $\mathbf{v}$ *es* ***no nulo*** *y además existe un escalar* λ *en* $\mathbb{C}$ *tal que* $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$.

Observaciones:

- Dada una matriz $\mathbf{A}$ cuadrada cualquiera, la ecuación $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$ admite la solución nula, cualquiera que sea el valor de λ; por ese motivo en la definición de vector propio se excluye el vector nulo. Es decir, los valores de λ *significativos* son los que resuelven la ecuación para vectores *no nulos*.

- Sin embargo, $\lambda = 0$ sí puede ser un valor propio de $\mathbf{A}$, en concreto, asociado a los vectores (no nulos) de $\ker \mathbf{A}$, es decir, cuando $\dim \ker \mathbf{A} \geq 1$.

- Si $\mathbf{A}$ es real y admite un valor propio real, los vectores propios correspondientes pueden escogerse reales.

- A todo valor propio corresponde, al menos, un vector propio y todos los vectores proporcionales al mismo (con excepción del vector nulo), ya que los proporcionales no nulos de un vector propio son vectores propios asociados al mismo valor propio.

  De hecho, el conjunto de todos los vectores propios asociados a un determinado valor propio constituye, junto con el vector nulo, un subespacio vectorial al que se denominará **subespacio característico o propio** de la matriz asociado al susodicho valor.

  En efecto, basta observar que si λ es un valor propio, entonces los vectores propios $\mathbf{v}$ asociados a él verifican $\mathbf{Av} - \lambda\mathbf{v} = \mathbf{0}$. Es decir, son las soluciones no nulas del sistema lineal homogéneo $(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0}$ donde $\mathbf{I}$ representa la matriz identidad de orden n. Por consiguiente, el subespacio característico asociado a λ es el núcleo de la matriz $\mathbf{A} - \lambda\mathbf{I}$ que denotamos por $\ker(\mathbf{A} - \lambda\mathbf{I})$.

  Sin embargo, como se justificará más adelante, a cada vector propio le corresponde un **único** valor propio.

- Si la matriz $\mathbf{A}$ es una matriz **diagonal**, entonces los elementos diagonales son valores propios, mientras que los vectores de la base canónica son vectores propios.

- Si la matriz $\mathbf{A}$ es **triangular**, entonces los elementos diagonales son sus valores propios, pero los vectores canónicos ya no son, en general, sus autovectores, aunque el que algunos puedan serlo dependerá de la estructura de la matriz triangular.

Proposición 27 *Si $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r$ son r vectores propios de una matriz $\mathbf{A}$, asociados respectivamente a r valores propios $\lambda_1, \lambda_2, \ldots, \lambda_r$, distintos dos a dos, entonces dichos vectores son linealmente independientes.*

Demostración: Se demostrará por inducción sobre el número r de valores propios distintos. Está claro que el conjunto cuyo único vector propio es $\mathbf{v}_1$ es libre, porque es vector propio y, por tanto, no nulo. Supóngase ahora $r > 1$. Como hipótesis de inducción se establece que la proposición anterior es cierta para $r-1$ vectores propios en las condiciones del enunciado.

En base a dicha hipótesis, se parte de que los vectores $\mathbf{v}_2, \mathbf{v}_3, \ldots, \mathbf{v}_r$ son linealmente independientes.

Considérese la siguiente combinación lineal nula de los r vectores:

$$\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \ldots + \alpha_r\mathbf{v}_r = \mathbf{0} \tag{5.1}$$

Se ha de probar que los escalares $\alpha_j, j = 1, \ldots, r$ son nulos. Premultiplicando la expresión (5.1) por $\mathbf{A}$ y teniendo en cuenta que $\mathbf{A}\mathbf{v}_i = \lambda_i \mathbf{v}_i$ se obtiene:

$$\alpha_1 \lambda_1 \mathbf{v}_1 + \alpha_2 \lambda_2 \mathbf{v}_2 + \ldots + \alpha_r \lambda_r \mathbf{v}_r = \mathbf{0} \tag{5.2}$$

Si a la expresión (5.2) se le resta la expresión (5.1) multiplicada por λ_1 se obtiene:

$$\alpha_2(\lambda_2 - \lambda_1)\mathbf{v}_2 + \alpha_3(\lambda_3 - \lambda_1)\mathbf{v}_3 + \ldots + \alpha_r(\lambda_r - \lambda_1)\mathbf{v}_r = \mathbf{0} \tag{5.3}$$

Por la hipótesis de inducción en esta combinación lineal los vectores son linealmente independientes, luego los coeficientes deben ser nulos. Como $(\lambda_i - \lambda_1) \neq 0$, se sigue que $\alpha_k = 0, k = 2, \ldots, r$. Llevando este resultado a la expresión (5.1) y teniendo en cuenta que $\mathbf{v}_1$ es no nulo, se sigue que también $\alpha_1 = 0$ y por tanto los r vectores $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r$ son independientes. $\square$

Observación: En particular, si una matriz $\mathbf{A}$ de orden n tiene n valores propios distintos, entonces escogiendo sendos vectores propios asociados a cada uno de ellos se construye una base de $\mathbb{C}^n$. Si se forma una matriz $\mathbf{P}$ con los vectores propios de esa base, entonces la matriz $\mathbf{P}^{-1}\mathbf{A}\mathbf{P}$ es una matriz diagonal semejante a $\mathbf{A}$ y por tanto $\mathbf{A}$ es diagonalizable.

En efecto, las ecuaciones $\mathbf{A}\mathbf{v}_i = \lambda_i \mathbf{v}_i, i = 1, 2, \ldots, n$ se pueden escribir conjuntamente en la forma:

$$\mathbf{A}\begin{pmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{pmatrix} = \begin{pmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{pmatrix} \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix} \tag{5.4}$$

Si $\mathbf{D} = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ y $\mathbf{P} = \begin{pmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{pmatrix}$ la ecuación (5.4) se puede escribir abreviadamente en las formas:

$$\mathbf{A}\mathbf{P} = \mathbf{P}\mathbf{D} \quad (\det \mathbf{P} \neq 0) \tag{5.5}$$

$$\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{D} \tag{5.6}$$

Sin embargo, la existencia de n valores propios distintos es condición suficiente pero **no necesaria** para que exista tal base; por ejemplo, si se considera la matriz identidad, con "1" como único autovalor, toda base de $\mathbb{C}^n$ es de vectores propios de esta matriz. Se pueden encontrar otros ejemplos no tan triviales: matrices diagonales con elementos repetidos, etc.

Definición 60 (espectro y radio espectral) *Sea* $\mathbf{A}$ *una matriz cuadrada compleja.*

- *Se llama* ***espectro*** *de la matriz* $\mathbf{A}$ *al conjunto de todos sus valores propios.*
- *Se llama* ***radio espectral*** *de una matriz* $\mathbf{A}$ *al máximo de los módulos de los valores propios. Se suele denotar como* $\rho(\mathbf{A})$.

5.3.1. Polinomio característico

Ya se ha visto que si λ es un valor propio de una matriz $\mathbf{A}$, el subespacio característico asociado es precisamente $\ker(\mathbf{A} - \lambda\mathbf{I})$. Para que el sistema homogéneo admita solución no trivial es necesario que la matriz $\mathbf{A} - \lambda\mathbf{I}$ tenga rango inferior a su orden o, equivalentemente, que su determinante sea nulo. He aquí entonces, un **procedimiento para calcular los valores y vectores propios** de una matriz: una vez calculadas las raíces de la ecuación $\det(\lambda\mathbf{I} - \mathbf{A}) = 0$, se resuelven los sistemas lineales homogéneos $(\lambda\mathbf{I} - \mathbf{A})\mathbf{x} = \mathbf{0}$ para cada una de dichas raíces y las soluciones no nulas serán vectores propios.

Si consideramos

$$|\lambda\mathbf{I} - \mathbf{A}| = \begin{vmatrix} \lambda - a_{11} & -a_{12} & \cdots & -a_{1n} \\ -a_{21} & \lambda - a_{22} & \cdots & -a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ -a_{n1} & -a_{n2} & \cdots & \lambda - a_{nn} \end{vmatrix}, \tag{5.7}$$

observamos que el desarrollo de tal determinante nos da un polinomio en la indeterminada λ, cuyo término de mayor grado, que se obtiene a partir del producto de los elementos de la diagonal, es λ^n; se trata, pues, de un polinomio mónico de grado igual al orden de la matriz.

Observación: Desde un punto de vista práctico (numérico), el cálculo de valores propios es difícil ya que no existe *regla* para calcular *a mano* las raíces de un polinomio de grado superior a 4; incluso las fórmulas para calcular las raíces de polinomios de grados 3 y 4 –aunque conocidas– son poco prácticas. En este curso los ejercicios son sencillos porque lo que se pretende es que se comprendan las ideas teóricas. En cualquier texto básico de Álgebra Numérica se explican algoritmos para el cálculo aproximado de valores propios.

Definición 61 *Dada* $\mathbf{A} \in \mathbb{C}^{n\times n}$*, se llama **polinomio característico** de* $\mathbf{A}$*, y se denota por* $\chi_{\mathbf{A}}$*, al polinomio mónico de grado* n *definido por*

$$\chi_{\mathbf{A}}(\lambda) := |\lambda\mathbf{I} - \mathbf{A}|, \tag{5.8}$$

cuyas raíces son los autovalores de $\mathbf{A}$*.*

Definición 62 *Dada* $\mathbf{A} \in \mathbb{C}^{n\times n}$*, un autovalor* λ *de* $\mathbf{A}$ *se dice que tiene **multiplicidad algebraica** igual a* m *si tiene multiplicidad* m *como raíz de* $\chi_{\mathbf{A}}$*. La dimensión del subespacio* $\ker(\mathbf{A} - \lambda\mathbf{I})$ *se llama **multiplicidad geométrica** de* λ*.*

Proposición 28 *Dos matrices semejantes tienen el mismo polinomio característico.*

Demostración: Sean $\mathbf{A}, \mathbf{B}$ dos matrices semejantes, relacionadas por $\mathbf{B} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P}$, entonces:

$$\begin{aligned} \chi_{\mathbf{B}}(\lambda) &= \det\,(\lambda\mathbf{I} - \mathbf{B}) = \det\,(\lambda\mathbf{I} - \mathbf{P}^{-1}\mathbf{A}\mathbf{P}) \\ &= \det\,[\mathbf{P}^{-1}(\lambda\mathbf{I} - \mathbf{A})\mathbf{P}] = \det\,(\lambda\mathbf{I} - \mathbf{A}) = \chi_{\mathbf{A}}(\lambda) \end{aligned}$$

Como corolario se deduce que dos matrices semejantes tienen la misma traza (opuesto del coeficiente de λ^{n-1} en el polinomio característico), así como el mismo determinante ($(-1)^n$ por el término independiente). $\square$

Observaciones: Sea $\mathbf{A}$ una matriz compleja de orden n, entonces:

- En virtud del teorema 19, tiene n autovalores.
- El término independiente de $\chi_{\mathbf{A}}$ es $\chi_{\mathbf{A}}(0) = \det(-\mathbf{A}) = (-1)^n \det \mathbf{A}$. Por otro lado, el término independiente de un polinomio mónico es $(-1)^n$ multiplicado por el producto de sus raíces, concluimos que

$$\det \mathbf{A} = \lambda_1 \lambda_2 \dots \lambda_n, \tag{5.9}$$

  en donde los λ_j anteriores representan todos los autovalores de $\mathbf{A}$, multiplicados tantas veces como indiquen sus multiplicidades.

- El término de grado $n-1$ de $\chi_{\mathbf{A}}$ se obtiene también a partir de la suma de todos los elementos de la diagonal principal:

$$\chi_{\mathbf{A}}(\lambda) = (\lambda - a_{11})(\lambda - a_{22}) \dots (\lambda - a_{nn}) + \cdots$$
$$= \lambda^n - (a_{11} + a_{22} + \dots + a_{nn})\lambda^{n-1} + (\text{términos de grado menor que } n-1) \tag{5.10}$$

  Se observa, pues, que el coeficiente de grado $n-1$ de $\chi_{\mathbf{A}}$ es el opuesto de la traza de $\mathbf{A}$. Además dicho coeficiente coincide con el opuesto de la suma de las raíces y, en conclusión,

$$\mathrm{tr}\mathbf{A} = \lambda_1 + \lambda_2 + \dots + \lambda_n, \tag{5.11}$$

  siendo los λ_j todos los autovalores de $\mathbf{A}$, sumados tantas veces como indiquen sus multiplicidades algebraicas.

- Si la matriz $\mathbf{A}$ es **invertible**, los valores propios de la inversa son los inversos de los valores propios de $\mathbf{A}$. Para comprobar la relación entre valores propios basta observar las equivalencias:

$$\mathbf{A}\mathbf{v} = \lambda \mathbf{v} \quad \Longleftrightarrow \quad \mathbf{v} = \lambda \mathbf{A}^{-1}\mathbf{v} \quad \Longleftrightarrow \quad \mathbf{A}^{-1}\mathbf{v} = \frac{1}{\lambda}\mathbf{v} \tag{5.12}$$

  Está claro que todos los valores propios λ de esta matriz han de ser distintos de cero, puesto que si la matriz es invertible, el sistema homogéneo asociado $\mathbf{A}\mathbf{x} = \mathbf{0}$ solo admite solución trivial.

  Además, cada valor propio de $\mathbf{A}$ y su correspondiente inverso, valor propio de $\mathbf{A}^{-1}$, tienen igual multiplicidad algebraica.

- Las matrices reales de orden impar tienen, al menos, un valor propio real.

- La matriz

$$\mathbf{A} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \tag{5.13}$$

tiene como polinomio característico $\lambda^2 + 1$. Como matriz real no admite valores propios, mientras que como matriz compleja admite dos distintos i y $-i$: existe una base de $\mathbb{C}^2$ (no real) formada por vectores propios de $\mathbf{A}$.

- Si una matriz real (contemplada como matriz compleja) tiene un valor propio complejo α, con parte imaginaria no nula, entonces $\overline{\alpha}$ es también valor propio con la misma multiplicidad algebraica que α. Además, los vectores propios correspondientes pueden tomarse conjugados:

$$\mathbf{A}\mathbf{v} = \alpha\mathbf{v} \qquad \Longleftrightarrow \qquad \mathbf{A}\overline{\mathbf{v}} = \overline{\alpha}\,\overline{\mathbf{v}} \tag{5.14}$$

Como consecuencia, $\{\mathbf{v}, \overline{\mathbf{v}}\}$ es libre.

- La igualdad de los polinomios característicos es una condición necesaria, no suficiente de semejanza. Por ejemplo, la matriz

$$\mathbf{A} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \tag{5.15}$$

tiene por polinomio característico λ^2, al igual que la matriz nula, sin embargo $\mathbf{A}$ no es semejante a la nula ya que no tienen el mismo rango; además, la única matriz semejante a la matriz nula es ella misma.

- Dadas dos matrices $\mathbf{A} \in \mathbb{C}^{m\times n}$ y $\mathbf{B} \in \mathbb{C}^{n\times m}$, se verifica la siguiente relación:

$$\lambda^n \chi_{\mathbf{AB}}(\lambda) = \lambda^m \chi_{\mathbf{BA}}(\lambda) \tag{5.16}$$

Este resultado se demuestra parcialmente en el problema 5.15.

Ejemplo: Calcular los valores y vectores propios de la matriz

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 4 & -1 \\ -1 & 0 & 2 \end{pmatrix} \tag{5.17}$$

Calculemos en primer lugar su polinomio característico

$$\chi_{\mathbf{A}}(\lambda) = \begin{vmatrix} \lambda - 1 & 0 & 0 \\ -1 & \lambda - 4 & 1 \\ 1 & 0 & \lambda - 2 \end{vmatrix} = (\lambda - 1)(\lambda - 4)(\lambda - 2) \tag{5.18}$$

Para calcular los vectores propios asociados con $\lambda = 1$ resolvamos:

$$(\mathbf{A} - \mathbf{I})\mathbf{x} = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 3 & -1 \\ -1 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \tag{5.19}$$

Las soluciones son los múltiplos de $\mathbf{v}_1^t = (1, 0, 1)$. Análogamente, los vectores propios asociados a $\lambda = 4$:

$$(\mathbf{A} - 4\mathbf{I})\mathbf{x} = \begin{pmatrix} -3 & 0 & 0 \\ 1 & 0 & -1 \\ -1 & 0 & -2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \tag{5.20}$$

Las soluciones son los múltiplos de $\mathbf{v}_2^t = (0, 1, 0)$. Por último, para el valor propio $\lambda = 2$ se tiene

$$(\mathbf{A} - 2\mathbf{I})\mathbf{x} = \begin{pmatrix} -1 & 0 & 0 \\ 1 & 2 & -1 \\ -1 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \tag{5.21}$$

Las soluciones son los múltiplos de $\mathbf{v}_3^t = (0, 1, 2)$.

Los vectores $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ son linealmente independientes. Si $\mathbf{P}$ es la matriz cuyas columnas son dichos vectores,

$$\mathbf{P} = \begin{pmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_n \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 2 \end{pmatrix} \tag{5.22}$$

el producto $\mathbf{AP}$ tiene como primera columna $\mathbf{v}_1$, como segunda $4\mathbf{v}_2$ y como tercera $2\mathbf{v}_3$. Dicho de otra forma, se verifica $\mathbf{AP} = \mathbf{P}\mathrm{diag}\,(1, 4, 2)$. Equivalentemente, $\mathbf{P}^{-1}\mathbf{AP} = \mathrm{diag}\,(1, 4, 2)$.

Ejemplo: Calcular los valores y vectores propios de la matriz

$$\mathbf{A} = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix} \tag{5.23}$$

Calculemos en primer lugar su polinomio característico

$$\chi_{\mathbf{A}}(\lambda) = \begin{vmatrix} \lambda - 2 & -1 & -1 \\ -1 & \lambda - 2 & -1 \\ -1 & -1 & \lambda - 2 \end{vmatrix} = \lambda^3 - 6\lambda^2 + 9\lambda - 4 = (\lambda - 4)(\lambda - 1)^2 \tag{5.24}$$

Para calcular ahora los vectores propios asociados con $\lambda = 4$ resolvamos:

$$(\mathbf{A} - 4\mathbf{I})\mathbf{x} = \begin{pmatrix} -2 & 1 & 1 \\ 1 & -2 & 1 \\ 1 & 1 & -2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \tag{5.25}$$

Las soluciones son los múltiplos de $\mathbf{v}_1^t = (1, 1, 1)$. Los vectores propios asociados al valor propio doble $\lambda = 1$:

$$(\mathbf{A} - \mathbf{I})\mathbf{x} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \tag{5.26}$$

La matriz anterior tiene rango 1 y por tanto hay un subespacio propio de dimensión 2 asociado con $\lambda = 1$. Tomando x_2 y x_3 como variables libres a las que se da alternativamente valor 1 ó 0 se obtienen las soluciones $\mathbf{v}_2^t = (-1, 1, 0)$ y $\mathbf{v}_3^t = (-1, 0, 1)$. Los vectores propios correspondientes son todas las combinaciones lineales no nulas de estos dos vectores.

5.4. Diagonalización

En el primer ejemplo se ha encontrado una matriz diagonal **semejante** a la matriz original $\mathbf{A}$. Los elementos de la diagonal son los valores propios de $\mathbf{A}$, mientras que las columnas de la matriz $\mathbf{P}$ son vectores propios. En el segundo ejemplo también se han encontrado tres vectores propios linealmente independientes, a pesar de que uno de los valores propios es doble.

De una forma más general, si una matriz cuadrada $\mathbf{A}$ es semejante a una diagonal $\mathbf{D}$, entonces la relación $\mathbf{AP} = \mathbf{PD}$ significa que la columna j-ésima de $\mathbf{P}$ es proporcional a la j-ésima de $\mathbf{AP}$. Dicho con otras palabras, el elemento j de la diagonal $\mathbf{D}$ es un valor propio de $\mathbf{A}$ y tiene como vector propio asociado precisamente la columna j de $\mathbf{P}$. Se tiene que las columnas de $\mathbf{P}$ son una base de $\mathbb{C}^n$ formada por vectores propios de $\mathbf{A}$.

Proposición 29 *Sea* $\mathbf{A}$ *una matriz cuadrada de orden* n. *Supóngase que el polinomio característico de* $\mathbf{A}$ *admite la factorización:*

$$\chi_{\mathbf{A}}(\lambda) = (\lambda - \lambda_1)^{\sigma_1} \dots (\lambda - \lambda_r)^{\sigma_r} \tag{5.27}$$

donde $\lambda_j, j = 1, \dots, r$ *son los valores propios distintos (dos a dos) de* $\mathbf{A}$ *y* σ_j *sus multiplicidades respectivas. Entonces, para cada* $j = 1, \dots, r$ *se tiene*

$$1 \leq dim\ \ker(\mathbf{A} - \lambda_j \mathbf{I}) \leq \sigma_j \tag{5.28}$$

Es decir, la multiplicidad geométrica es menor o igual que la algebraica.

Demostración: La demostración se hará construyendo una matriz semejante a $\mathbf{A}$ y tal que admita λ_j como valor propio de multiplicidad al menos igual a la dimensión del subespacio característico asociado. Para ello, se considera una base $\mathcal{B}$ de $\mathbb{C}^n$ tal que los m primeros vectores formen una base del subespacio $\ker(\mathbf{A} - \lambda_j \mathbf{I})$, supuesto de dimensión igual a m. Es obvio que m es mayor o igual que 1 puesto que λ_j se calcula de modo que la matriz $(\mathbf{A} - \lambda_j \mathbf{I})$ tenga determinante nulo y, en consecuencia, rango menor que n.

Si $\mathbf{P}$ representa la matriz cuyas columnas son los vectores de la base $\mathcal{B}$ entonces $\mathbf{P}$ es invertible y se verifica la igualdad

$$\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \begin{pmatrix} \lambda_j \mathbf{I}_m & \mathbf{B} \\ \mathbf{0} & \mathbf{C} \end{pmatrix} \tag{5.29}$$

en la cual $\mathbf{I}_m$ representa la matriz identidad de orden m, $\mathbf{B}$ es una matriz $m \times (n-m)$ y $\mathbf{C}$ es cuadrada de orden $n - m$. Es claro que el polinomio característico de esta matriz coincide con el de $\mathbf{A}$ por semejanza y es igual a $(\lambda - \lambda_j)^m \chi_{\mathbf{C}}(\lambda)$. Se tiene entonces que λ_j es valor propio de $\mathbf{A}$ de multiplicidad por lo menos m. $\square$

Proposición 30 (caracterización de matriz diagonalizable) *Sea $\mathbf{A}$ una matriz compleja, cuadrada de orden n. Entonces $\mathbf{A}$ es diagonalizable en $\mathbb{C}$ si, y solamente si, para cada valor propio de $\mathbf{A}$ las multiplicidades algebraica y geométrica coinciden.*

Obsérvese que una matriz real puede ser no diagonalizable en $\mathbb{R}$, pero diagonalizable en $\mathbb{C}$, por ejemplo

$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \tag{5.30}$$

Observaciones: Todas las propiedades siguientes se enuncian para una matriz cuadrada de orden n:

1. Si $\mathbf{A}$ es **diagonalizable**, entonces $\mathbf{A}^t, \bar{\mathbf{A}}$ y $\bar{\mathbf{A}}^t$ también lo son (y la matriz de paso que la diagonaliza es la inversa de la transpuesta de la de $\mathbf{A}$).

   En efecto, si $\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{D}$ diagonal, entonces

   $$\mathbf{D} = \mathbf{D}^t = (\mathbf{P}^{-1}\mathbf{A}\mathbf{P})^t = \mathbf{P}^t\mathbf{A}^t(\mathbf{P}^{-1})^t = \mathbf{P}^t\mathbf{A}^t(\mathbf{P}^t)^{-1}. \tag{5.31}$$

2. Si $\mathbf{A}$ es **diagonalizable** e **invertible**, entonces $\mathbf{A}^{-1}$ también lo es (y la matriz de paso que la diagonaliza es la misma que la de $\mathbf{A}$).

   En efecto, si $\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{D} = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ con todos los λ_j no nulos, entonces

   $$\operatorname{diag}(\lambda_1^{-1}, \ldots, \lambda_n^{-1}) = \mathbf{D}^{-1} = (\mathbf{P}^{-1}\mathbf{A}\mathbf{P})^{-1} = \mathbf{P}^{-1}\mathbf{A}^{-1}\mathbf{P}. \tag{5.32}$$

3. Si $\mathbf{A}$ es **diagonalizable**, entonces cualquier potencia de $\mathbf{A}$ también lo es (y la matriz que la diagonaliza es la misma).

   En efecto, $\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{D} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ implica que

$$\begin{aligned} \mathrm{diag}\left(\lambda_1^k, ..., \lambda_n^k\right) = \mathbf{D}^k = \left(\mathbf{P}^{-1}\mathbf{A}\mathbf{P}\right)^k = \\ = \left(\mathbf{P}^{-1}\mathbf{A}\mathbf{P}\right)\left(\mathbf{P}^{-1}\mathbf{A}\mathbf{P}\right)\cdots\left(\mathbf{P}^{-1}\mathbf{A}\mathbf{P}\right) = \mathbf{P}^{-1}\mathbf{A}^k\mathbf{P} \end{aligned} \tag{5.33}$$

   El recíproco no es cierto. Una potencia diagonalizable no implica que $\mathbf{A}$ lo sea. Por ejemplo la matriz $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ no es diagonalizable, aunque su cuadrado es la matriz nula que es un caso particular de matriz diagonal.

5.4.1. Teorema de Cayley–Hamilton

Sea $\mathbf{A}$ una matriz compleja y sea

$$P(\lambda) = \sum_{j=0}^{m} a_j \lambda^j \tag{5.34}$$

un polinomio cualquiera. Se define la matriz $P(\mathbf{A})$ como la que se obtiene sustituyendo en la expresión del polinomio las potencias de λ por potencias de $\mathbf{A}$, es decir:

$$P(\mathbf{A}) := \sum_{j=0}^{m} a_j \mathbf{A}^j = a_0\mathbf{I} + a_1\mathbf{A} + a_2\mathbf{A}^2 + \cdots + a_m\mathbf{A}^m \tag{5.35}$$

donde se toma por convenio, $\mathbf{A}^0 = \mathbf{I}$.

Observaciones:

1. Si λ_0 es valor propio de una matriz $\mathbf{A}$, entonces para cualquier polinomio P, $P(\lambda_0)$ es valor propio de $P(\mathbf{A})$.

2. Los vectores propios de $\mathbf{A}$ lo son de $P(\mathbf{A})$ para cualquier polinomio P.

   ¡Atención! el recíproco no es, necesariamente, cierto. Considérese por ejemplo la matriz $\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$, cuyos valores propios son 1 y -1. Los correspondientes vectores propios son $\mathbf{v}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ y $\mathbf{v}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$. Considérese el caso particular $P(\mathbf{A}) = \mathbf{A}^2$ y obsérvese que $\mathbf{A}^2 = \mathbf{I}$. Así como todos los vectores propios de $\mathbf{A}$ son vectores propios de $\mathbf{I}$, el recíproco no es cierto porque cualquier vector no nulo en $\mathbb{R}^2$, por ejemplo el $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$, es vector propio de $\mathbf{I}$, pero no tiene por qué serlo de $\mathbf{A}$.

Definición 63 *Se dice que una matriz* $\mathbf{A}$ *es* ***raíz*** *de un polinomio* P *si la matriz* $P(\mathbf{A})$ *es la matriz nula.*

Teorema 11 (de Cayley-Hamilton) *Toda matriz es raíz de su polinomio característico. Es decir, si su polinomio característico es*

$$\chi_{\mathbf{A}}(\lambda) = b_0 + b_1\lambda + b_2\lambda^2 + \cdots + b_{n-1}\lambda^{n-1} + \lambda^n$$

entonces

$$b_0\mathbf{I} + b_1\mathbf{A} + b_2\mathbf{A}^2 + \cdots + \mathbf{A}^n \equiv \mathbf{O} \tag{5.36}$$

5.4.2. Aplicaciones

Sea $\mathbf{A}$ una matriz cuadrada, sea χ su polinomio característico y P un polinomio cualquiera.

1. En virtud del teorema de división euclídea de polinomios, $P(\mathbf{A}) = R(\mathbf{A})$ siendo R el resto de dividir P entre χ.

2. En particular, el resultado anterior puede utilizarse para calcular potencias de $\mathbf{A}$ ya que $\mathbf{A}^m = P(\mathbf{A})$ para $P(\lambda) = \lambda^m$.

3. Si $\mathbf{A}$ es diagonalizable, entonces $P(\mathbf{A})$ también lo es, cualquiera que sea el polinomio P. En efecto, todas las potencias de $\mathbf{A}$ son diagonalizadas por la misma matriz $\mathbf{P}$ y como consecuencia la matriz $P(\mathbf{A})$ también es diagonalizada por $\mathbf{P}$. Si $\mathbf{P}^{-1}\mathbf{AP} = \mathbf{D}$, con $\mathbf{D}$ diagonal, entonces

$$\mathbf{P}^{-1}P(\mathbf{A})\mathbf{P} = P(\mathbf{D}) \tag{5.37}$$

5.5. Ejercicios

5.1.– Calcular los valores y vectores propios de cada una de las siguientes matrices

$$\mathbf{A} = \begin{pmatrix} -3 & 2 \\ -2 & 2 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 3 & -1 \\ 4 & -1 \end{pmatrix}, \quad \mathbf{C} = \begin{pmatrix} -3 & -2 & -1 \\ 6 & 4 & 2 \\ -2 & -1 & -1 \end{pmatrix},$$

$$\mathbf{D} = \begin{pmatrix} 0 & 1 & 1 \\ -1 & -2 & -1 \\ 1 & 1 & 0 \end{pmatrix}, \quad \mathbf{E} = \begin{pmatrix} -1 & -1 & -1 \\ 4 & 3 & 2 \\ -2 & -1 & -1 \end{pmatrix}, \quad \mathbf{F} = \begin{pmatrix} 3 & 1 & 3 \\ 0 & 1 & -2 \\ -2 & -1 & -1 \end{pmatrix}.$$

analizando si son o no diagonalizables. Para aquellas que sean diagonalizables determinar una matriz que las diagonalice.

5.2.– Calcular los valores de a y b que hacen que $(1, 1, -1)^t$ sea un autovector de la matriz $\begin{pmatrix} -b & 2 & 0 \\ a & b & 1 \\ 0 & 0 & a \end{pmatrix}$.

5.3.– Indicar los valores de $a \in \mathbb{R}$ para que $\begin{pmatrix} 1 & -a+1 & -1 \\ 0 & -2 & -3 \\ 0 & -2 & -1 \end{pmatrix}$ sea semejante a diag $(1, -4, 1)$.

5.4.– Hallar el valor del parámetro a que hace que la matriz $\begin{pmatrix} -1 & -1 & 0 \\ -2 & -1 & a \\ 0 & 1 & -1 \end{pmatrix}$ no sea diagonalizable.

5.5.– Estudiar si la matriz

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & -6 \\ 1 & 0 & 2 \\ 1 & 0 & 2 \end{pmatrix}$$

es diagonalizable y diagonalizarla en caso afirmativo.

5.6.– Hallar los valores propios de la matriz

$$\tilde{\mathbf{w}} = \begin{pmatrix} 0 & -c & b \\ c & 0 & -a \\ -b & a & 0 \end{pmatrix}$$

Interpretar el resultado geométricamente considerando el operador $\mathbf{L}(\mathbf{r}) := \mathbf{w} \times \mathbf{r}$ siendo $\mathbf{w} = (a, b, c)^t$.

5.7.– Comprobar que una matriz y su traspuesta tienen el mismo polinomio característico y, por tanto, los mismos valores propios. Analizar lo que sucede con los vectores propios.

5.8.– Sea $\mathbf{A}$ una matriz cuyo polinomio característico es $\chi_{\mathbf{A}}(z) = (z-3)^{13}(z+4)^7$. Hallar el polinomio característico de $\mathbf{A} - 2\mathbf{I}$.

5.9.– Dados dos vectores $\mathbf{a}$ y $\mathbf{b}$ en $\mathbb{R}^n$, determínense los valores y vectores propios de la matriz de rango unidad $\mathbf{ab}^t$. Como aplicación, hállense los valores y vectores propios de la matriz

$$\begin{pmatrix} 1 & 1 & \cdots & 1 \\ 2 & 2 & \cdots & 2 \\ \vdots & \vdots & \ddots & \vdots \\ n & n & \cdots & n \end{pmatrix}$$

5.10.– Establecer la relación que existe entre los valores y vectores propios de una matriz compleja $\mathbf{M}$ y los de la matriz $a\mathbf{I} + b\mathbf{M}$ siendo a, b números complejos no nulos arbitrarios –se aconseja revisitar el problema 2.6–. Como aplicación, deducir los valores y vectores propios de la matriz

$$\begin{pmatrix} c & d & \cdots & d \\ d & c & \cdots & d \\ \vdots & \vdots & \ddots & \vdots \\ d & d & \cdots & c \end{pmatrix}.$$

5.11.– Sea $\mathbf{A} \in \mathbb{C}^{n\times n}$ con $r(\mathbf{A}) = tr(\mathbf{A}) = 1$. Hallar los valores propios de $\mathbf{A}$, así como sus multiplicidades algebraicas y geométricas. Deducir si $\mathbf{A}$ es o no diagonalizable.

5.12.– Se dice que una matriz $\mathbf{A}$ es nilpotente si existe $k \in \mathbb{N}$ tal que $\mathbf{A}^k = 0$. Sea $\mathbf{A}$ nilpotente.

1. Calcular los valores propios de $\mathbf{A}$ y de $\mathbf{A} - \mathbf{I}$
2. Deducir del apartado anterior la traza y el determinante de $\mathbf{A} - \mathbf{I}$.

5.13.– Sea $\mathbf{A}$ una matriz cuadrada invertible de orden n. Probar que la matriz inversa de $\mathbf{A}$ se puede expresar como un polinomio en $\mathbf{A}$. Obtener este polinomio (¿es único?) en el caso particular de la matriz

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

5.14.– Se considera el polinomio $p(z) = c_0 + c_1 z + c_2 z^2 + c_3 z^3$ y la matriz $\mathbf{S} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}$.

1. Determinar los valores y vectores propios de $\mathbf{S}$.

2. Obtener explícitamente la matriz que resulta al evaluar el polinomio $p(z)$ en la matriz $\mathbf{S}$.

3. Determinar los valores y vectores propios de la matriz resultante, comprobando que siempre es diagonalizable.

4. Determinar los coeficientes de $p(z)$ para que la matriz

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

sea $p(\mathbf{S})$. Deducir sus valores y vectores propios.

5.15.– Sean las matrices $\mathbf{A} \in \mathbb{C}^{m\times n}$ y $\mathbf{B} \in \mathbb{C}^{n\times m}$, demostrar que los valores propios no nulos de $\mathbf{AB}$ coinciden con los de $\mathbf{BA}$.

5.5.1. Cuestiones

5.16.– Dada la matriz $\begin{pmatrix} 0 & -2 & 2 \\ 0 & 3 & 0 \\ 3 & \alpha & 1 \end{pmatrix}$ señalar cuáles de las siguientes afirmaciones son verdaderas y cuáles son falsas:

1. $\mathbf{A}$ es diagonalizable para todo valor de α

2. Si $\alpha = 0$, entonces $\mathbf{A}$ tiene un subespacio propio de dimensión 2

3. Si $\mathbf{A}$ es diagonalizable, entonces $\alpha = 2$

4. Para algún valor de α, $\mathbf{A}$ tiene tres valores propios distintos

5. Si $\alpha = -2$, existen tres vectores propios de $\mathbf{A}$ linealmente independientes

5.17.– Dada la matriz $\mathbf{A} = \begin{pmatrix} -1 & 2 & 1 \\ 0 & 3 & 0 \\ 2 & -4 & a \end{pmatrix} \in \mathbb{R}^{3\times 3}$, determinar cuáles de las siguientes afirmaciones son ciertas:

1. $\lambda = 3$ es siempre un autovalor de $\mathbf{A}$.

2. Si $a = -2$ entonces $\lambda = 0$ no es un autovalor de $\mathbf{A}$.

3. Para $a = 0$ la matriz $\mathbf{A}$ no es diagonalizable.

4. Para $a = 0$, $\mathbf{A}$ es semejante a $\operatorname{diag}(1, 3, -2)$.

Capítulo 6

Matrices normales

6.1. Semejanza unitaria y diagonalización unitaria

Definición 64 (matrices semejantes unitariamente) *Se dice que dos matrices cuadradas complejas* $\mathbf{A}$ *y* $\mathbf{B}$ *son* ***semejantes unitariamente*** *si existe una matriz unitaria* $\mathbf{U}$ *tal que* $\mathbf{B} = \mathbf{U}^{-1}\mathbf{A}\mathbf{U}$.

Observación: Puesto que $\mathbf{U}^{-1} = \mathbf{U}^h$, podemos expresar también la relación de semejanza unitaria como $\mathbf{B} = \mathbf{U}^h\mathbf{A}\mathbf{U}$.

Definición 65 (matriz diagonalizable unitariamente y ortogonalmente)

- *Dada* $\mathbf{A} \in \mathbb{C}^{n\times n}$*, se dice que es* ***diagonalizable unitariamente*** *si es semejante unitariamente a una matriz diagonal, es decir, si existe una matriz unitaria* $\mathbf{U}$ *tal que* $\mathbf{D} = \mathbf{U}^{-1}\mathbf{A}\mathbf{U}$ *es diagonal.*

- *Como caso particular, dada* $\mathbf{A} \in \mathbb{R}^{n\times n}$*, se dice que es* ***diagonalizable ortogonalmente*** *si existe una matriz ortogonal* $\mathbf{Q} \in \mathbb{R}^{n\times n}$ *tal que* $\mathbf{D} = \mathbf{Q}^{-1}\mathbf{A}\mathbf{Q}$ *es diagonal.*

Observación: Los elementos diagonales de $\mathbf{D}$ son los autovalores de $\mathbf{A}$, mientras que las columnas de $\mathbf{U}$ constituyen una base ortonormal de $\mathbb{C}^n$ formada por autovectores de $\mathbf{A}$.

La anterior observación nos permite concluir una primera caracterización de las matrices diagonalizables unitariamente:

Proposición 31 $\mathbf{A} \in \mathbb{C}^{n\times n}$; $\mathbf{A}$ *es diagonalizable unitariamente* $\Leftrightarrow$ $\mathbf{A}$ *es diagonalizable y sus subespacios propios son ortogonales dos a dos.*

6.2. Matrices normales

Definición 66 (matriz normal) $\mathbf{A} \in \mathbb{C}^{n\times n}$ *se dice que es* ***normal*** *si* $\mathbf{A}\mathbf{A}^h = \mathbf{A}^h\mathbf{A}$.

Ejemplos: Algunos ejemplos de matrices normales son:

- Las matrices hermíticas complejas (y en particular las reales simétricas).
- Las matrices antihermíticas complejas (y en particular las reales antisimétricas).
- Las matrices unitarias (y en particular las ortogonales).
- Las matrices diagonales.

Ejemplo: Existen muchas otras matrices normales que no pertenecen a ninguno de los tipos anteriores; por ejemplo, la matriz $\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}$ es normal (compruébese).

Proposición 32 *Una matriz triangular es normal si y sólo si es diagonal.*

Demostración: Ya sabemos que toda matriz diagonal es normal. Vamos a demostrar que, dada una matriz triangular superior $\mathbf{T} \in \mathbb{C}^{n\times n}$, la hipótesis de normalidad implica que $t_{ij} = 0$ siempre que $i < j$ (para matrices triangulares inferiores, observamos que una matriz es normal si y sólo si su transpuesta lo es).

Sea $\mathbf{T} = \begin{pmatrix} t_{11} & t_{12} & \cdots & t_{nn} \\ 0 & t_{22} & \cdots & t_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & t_{nn} \end{pmatrix}$, tal que $\mathbf{T}^h\mathbf{T} = \mathbf{T}\mathbf{T}^h$.

Calculamos los elementos diagonales de los productos anteriores y les imponemos la igualdad:

$$(\mathbf{T}^h\mathbf{T})_{11} = |t_{11}|^2, \tag{6.1}$$

pero también

$$(\mathbf{T}\mathbf{T}^h)_{11} = |t_{11}|^2 + |t_{12}|^2 + \ldots + |t_{1n}|^2. \tag{6.2}$$

La igualdad de los dos resultados anteriores implica que $t_{12} = t_{13} = \ldots = t_{1n} = 0$.

Del mismo modo,

$$(\mathbf{T}^h\mathbf{T})_{22} = |t_{22}|^2, \tag{6.3}$$

mientras que

$$(\mathbf{T}\mathbf{T}^h)_{22} = |t_{22}|^2 + |t_{23}|^2 + \ldots + |t_{2n}|^2, \tag{6.4}$$

de cuya igualdad se desprende que $t_{23} = t_{24} = \ldots = t_{2n} = 0$.

En general (supuesto que $t_{ij} = 0$ si $i < j$ para $i = 1, \ldots, k-1$),

$$(\mathbf{T}^h\mathbf{T})_{kk} = |t_{kk}|^2, \tag{6.5}$$

mientras que

$$(\mathbf{T}\mathbf{T}^h)_{kk} = |t_{kk}|^2 + |t_{k,k+1}|^2 + \ldots + |t_{kn}|^2, \tag{6.6}$$

lo que garantiza que $t_{k,k+1} = \ldots = t_{kn} = 0$. □

Observación: Una matriz **compleja no real y simétrica** no es, en general, normal. Por ejemplo, si $\mathbf{A} = \begin{pmatrix} 1 & i \\ i & -1 \end{pmatrix}$, $\mathbf{A}^h\mathbf{A} = \begin{pmatrix} 2 & 2i \\ -2i & 2 \end{pmatrix}$, mientras que $\mathbf{A}\mathbf{A}^h = \begin{pmatrix} 2 & -2i \\ 2i & 2 \end{pmatrix}$.

Observación: Si una matriz $\mathbf{A}$ es normal, entonces toda matriz de la forma $\mathbf{A} + \alpha\mathbf{I}$, con $\alpha \in \mathbb{C}$, también lo es.

Observación: Si dos matrices $\mathbf{A}, \mathbf{B}$ son semejantes unitariamente, entonces $\mathbf{A}$ es normal si y sólo si $\mathbf{B}$ es normal.

6.3. Teorema espectral

El teorema espectral caracteriza de manera sencilla las matrices diagonalizables unitariamente. Éstas resultan ser, precisamente, las matrices normales.

Teorema 12 (Teorema espectral para matrices normales) *Una matriz $\mathbf{A} \in \mathbb{C}^{n\times n}$ es diagonalizable unitariamente si y sólo si es normal.*

Demostración:

$\Rightarrow$) Si $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{U}^h$, con $\mathbf{D} = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$, para una cierta $\mathbf{U}$ unitaria, entonces:

$$\mathbf{A}^h\mathbf{A} = \mathbf{U}\mathbf{D}^h\mathbf{U}^h\mathbf{U}\mathbf{D}\mathbf{U}^h = \mathbf{U}\mathbf{D}\mathbf{D}^h\mathbf{U}^h = \mathbf{U}\operatorname{diag}\left(|\lambda_1|^2, \ldots, |\lambda_n|^2\right)\mathbf{U}^h; \tag{6.7}$$

$$\mathbf{A}\mathbf{A}^h = \mathbf{U}\mathbf{D}\mathbf{U}^h\mathbf{U}\mathbf{D}^h\mathbf{U}^h = \mathbf{U}\mathbf{D}\mathbf{D}^h\mathbf{U}^h = \mathbf{U}\operatorname{diag}\left(|\lambda_1|^2, \ldots, |\lambda_n|^2\right)\mathbf{U}^h, \tag{6.8}$$

luego $\mathbf{A}$ es normal. □

$\Leftarrow$) *(no exigible)*

Procederemos por inducción en n, el orden de la matriz. Para $n = 1$ el resultado es trivialmente cierto. Supongamos que lo es para matrices de orden $n-1$, es decir, que toda matriz normal de orden $n - 1$ es diagonalizable unitariamente.

Sea $\mathbf{A} \in \mathbb{C}^{n\times n}$ una matriz normal. $\mathbf{A}$ tiene algún autovalor $\lambda_1 \in \mathbb{C}$; sea $\mathbf{u}_1$ un vector propio unitario asociado a λ_1, ampliamos hasta una base $(\mathbf{u}_1, \ldots, \mathbf{u}_n)$ ortonormal de $\mathbb{C}^n$ y consideramos la matriz unitaria $\tilde{\mathbf{U}} = (\mathbf{u}_1|\ldots|\mathbf{u}_n)$. Se tiene que

$$\tilde{\mathbf{U}}^h\mathbf{A}\tilde{\mathbf{U}} = \begin{pmatrix} \lambda_1 & \mathbf{v}^t \\ \mathbf{0} & \mathbf{A}_{n-1} \end{pmatrix} = \tilde{\mathbf{A}} \tag{6.9}$$

para unos ciertos $\mathbf{v} \in \mathbb{C}^{n-1}$ y $\mathbf{A}_{n-1} \in \mathbb{C}^{(n-1)\times(n-1)}$.

La matriz $\tilde{\mathbf{A}}$ es normal por ser semejante unitariamente a $\mathbf{A}$. Ahora bien,

$$\tilde{\mathbf{A}}^h = \begin{pmatrix} \overline{\lambda}_1 & \mathbf{0}^t \\ \overline{\mathbf{v}} & \mathbf{A}_{n-1}^h \end{pmatrix} \tag{6.10}$$

y, por consiguiente,

$$(\tilde{\mathbf{A}}^h\tilde{\mathbf{A}})_{11} = |\lambda_1|^2 \tag{6.11}$$

mientras que

$$(\tilde{\mathbf{A}}\tilde{\mathbf{A}}^h)_{11} = |\lambda_1|^2 + \mathbf{v}^t\overline{\mathbf{v}}. \tag{6.12}$$

Como ambos resultados han de ser iguales, concluimos que $0 = \mathbf{v}^t\overline{\mathbf{v}} = \mathbf{v}^h\mathbf{v} = \|\mathbf{v}\|^2$, es decir, $\mathbf{v} = \mathbf{0}$. Así pues,

$$\tilde{\mathbf{A}} = \begin{pmatrix} \lambda_1 & \mathbf{0}^t \\ \mathbf{0} & \mathbf{A}_{n-1} \end{pmatrix}. \tag{6.13}$$

Ahora resulta evidente que la normalidad de esta matriz implica la de $\mathbf{A}_{n-1}$, ya que

$$\tilde{\mathbf{A}}^h\tilde{\mathbf{A}} = \begin{pmatrix} |\lambda_1|^2 & \mathbf{0}^t \\ \mathbf{0} & \mathbf{A}_{n-1}^h\mathbf{A}_{n-1} \end{pmatrix} \tag{6.14}$$

mientras que

$$\tilde{\mathbf{A}}\tilde{\mathbf{A}}^h = \begin{pmatrix} |\lambda_1|^2 & \mathbf{0}^t \\ \mathbf{0} & \mathbf{A}_{n-1}\mathbf{A}_{n-1}^h \end{pmatrix}, \tag{6.15}$$

de donde deducimos que $\mathbf{A}_{n-1}^h \mathbf{A}_{n-1} = \mathbf{A}_{n-1} \mathbf{A}_{n-1}^h$.

Al ser $\mathbf{A}_{n-1}$ una matriz normal de orden $n-1$, podemos aplicarle la hipótesis de inducción: existe $\mathbf{U}_{n-1} \in \mathbb{C}^{(n-1)\times(n-1)}$ unitaria tal que

$$\mathbf{U}_{n-1}^h \mathbf{A}_{n-1} \mathbf{U}_{n-1} = \mathrm{diag}(\lambda_2, \ldots, \lambda_n) = \mathbf{D}_{n-1}. \tag{6.16}$$

Consideramos finalmente la matriz

$$\mathbf{U} = \tilde{\mathbf{U}} \begin{pmatrix} 1 & \mathbf{0}^t \\ \mathbf{0} & \mathbf{U}_{n-1} \end{pmatrix} \in \mathbb{C}^{n\times n}, \tag{6.17}$$

que es unitaria por ser producto de unitarias. Vamos a ver que esta matriz diagonaliza unitariamente a $\mathbf{A}$. En efecto,

$$\mathbf{U}^h \mathbf{A} \mathbf{U} = \begin{pmatrix} 1 & \mathbf{0}^t \\ \mathbf{0} & \mathbf{U}_{n-1}^h \end{pmatrix} \tilde{\mathbf{U}}^h \mathbf{A} \tilde{\mathbf{U}} \begin{pmatrix} 1 & \mathbf{0}^t \\ \mathbf{0} & \mathbf{U}_{n-1} \end{pmatrix} = \tag{6.18}$$

$$= \begin{pmatrix} 1 & \mathbf{0}^t \\ \mathbf{0} & \mathbf{U}_{n-1}^h \end{pmatrix} \begin{pmatrix} \lambda_1 & \mathbf{0}^t \\ \mathbf{0} & \mathbf{A}_{n-1} \end{pmatrix} \begin{pmatrix} 1 & \mathbf{0}^t \\ \mathbf{0} & \mathbf{U}_{n-1} \end{pmatrix} = \tag{6.19}$$

$$= \begin{pmatrix} \lambda_1 & \mathbf{0}^t \\ \mathbf{0} & \mathbf{U}_{n-1}^h \mathbf{A}_{n-1} \mathbf{U}_{n-1} \end{pmatrix} = \begin{pmatrix} \lambda_1 & \mathbf{0}^t \\ \mathbf{0} & \mathbf{D}_{n-1} \end{pmatrix} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n). \tag{6.20}$$

Lo cual prueba que $\mathbf{A}$ es diagonalizable unitariamente. □

A continuación particularizaremos este resultado en algunas clases de matrices normales de especial relevancia.

6.3.1. Aplicación a matrices hermíticas, antihermíticas y unitarias

Toda matriz $\mathbf{A}$ normal ($\mathbf{A}^h\mathbf{A} = \mathbf{A}\mathbf{A}^h$) es diagonalizable unitariamente, es decir: para una cierta matriz unitaria $\mathbf{U}$,

$$\mathbf{U}^h \mathbf{A} \mathbf{U} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n) = \mathbf{D} \text{ o, equivalentemente, } \mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{U}^h,$$

en donde $\lambda_1, \ldots, \lambda_n \in \mathbb{C}$ son los autovalores de $\mathbf{A}$. En los casos particulares anteriormente destacados ($\mathbf{A}$ hermítica, antihermítica o unitaria) la matriz $\mathbf{D}$ es del mismo tipo que $\mathbf{A}$.

Matrices hermíticas

En este caso, puesto que $\mathbf{D}$ es hermítica,

$$\mathrm{diag}(\overline{\lambda}_1, \ldots, \overline{\lambda}_n) = \mathbf{D}^h = \mathbf{D} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n). \tag{6.21}$$

Por tanto, los autovalores de una matriz hermítica son reales.

Recíprocamente, si una matriz $\mathbf{A}$ es diagonalizable unitariamente y sus autovalores son reales,

$$\mathbf{A} = \mathbf{U}\,\mathrm{diag}(\lambda_1, \ldots, \lambda_n)\mathbf{U}^h \text{ con } \mathbf{U} \text{ unitaria y } \lambda_1, \ldots, \lambda_n \in \mathbb{R}, \tag{6.22}$$

entonces $\mathbf{A}^h = \mathbf{U}\,\mathrm{diag}(\lambda_1, \ldots, \lambda_n)\mathbf{U}^h = \mathbf{A}$, es decir, $\mathbf{A}$ es hermítica.

Todo ello se sintetiza en el resultado siguiente:

Teorema 13 (Teorema espectral para matrices hermíticas) *Sea* $\mathbf{A} \in \mathbb{C}^{n\times n}$*;* $\mathbf{A}$ *es diagonalizable unitariamente con autovalores reales* $\Leftrightarrow$ $\mathbf{A}$ *es hermítica.*

Como caso particular, se deduce:

Corolario: (Teorema espectral para matrices reales simétricas) Una matriz real es diagonalizable ortogonalmente si y solo si es simétrica.

Matrices antihermíticas

Como en este caso $\mathbf{D}$ es antihermítica, tenemos que

$$\mathrm{diag}(\overline{\lambda}_1, \ldots, \overline{\lambda}_n) = \mathbf{D}^h = -\mathbf{D} = \mathrm{diag}(-\lambda_1, \ldots, -\lambda_n), \tag{6.23}$$

de donde deducimos que los autovalores de una matriz antihermítica son imaginarios puros.

Recíprocamente, si una matriz $\mathbf{A}$ es diagonalizable unitariamente siendo todos sus autovalores imaginarios puros, entonces

$$\mathbf{A}^h = (\mathbf{U}\mathbf{D}\mathbf{U}^h)^h = \mathbf{U}\mathbf{D}^h\mathbf{U}^h = -\mathbf{U}\mathbf{D}\mathbf{U}^h = -\mathbf{A}, \tag{6.24}$$

es decir, $\mathbf{A}$ es antihermítica.

Como caso particular, toda matriz real y antisimétrica es diagonalizable unitariamente (¡no ortogonalmente!) con autovalores imaginarios puros.

Matrices unitarias

Puesto que $\mathbf{D}$ es unitaria cuando $\mathbf{A}$ lo es, tenemos que $\mathbf{I} = \mathbf{D}^h\mathbf{D} = \mathrm{diag}(|\lambda_1|^2, \ldots, |\lambda_n|^2)$, lo que garantiza que todos los autovalores de cualquier matriz unitaria tienen módulo unidad.

Recíprocamente, si $\mathbf{A}$ es una matriz normal cuyos autovalores tienen módulo unidad, entonces

$$\mathbf{A}^h\mathbf{A} = (\mathbf{U}\mathbf{D}\mathbf{U}^h)^h\mathbf{U}\mathbf{D}\mathbf{U}^h = \mathbf{U}\mathbf{D}^h\mathbf{U}^h\mathbf{U}\mathbf{D}\mathbf{U}^h = \mathbf{U}\mathbf{D}^h\mathbf{D}\mathbf{U}^h = \tag{6.25}$$

$$= \mathbf{U}\,\mathrm{diag}(|\lambda_1|^2, \ldots, |\lambda_n|^2)\mathbf{U}^h = \mathbf{U}^h\mathbf{I}\mathbf{U} = \mathbf{I}, \tag{6.26}$$

luego $\mathbf{A}$ es unitaria.

En particular, toda matriz real ortogonal es diagonalizable unitariamente con autovalores de módulo unidad, pero éstos pueden ser no reales (recuérdense las matrices de giro), luego no podemos hablar en general de diagonalización ortogonal.

6.3.2. Descomposición espectral

Toda $\mathbf{A} \in \mathbb{C}^{n\times n}$ normal, diagonalizable unitariamente a través de $\mathbf{U} = (\mathbf{U}_1|\ldots|\mathbf{U}_n)$, puede expresarse de la siguiente manera:

$$\mathbf{A} = \mathbf{U}\,\mathrm{diag}(\lambda_1,\ldots,\lambda_n)\mathbf{U}^h = \sum_{j=1}^{n} \lambda_j \mathbf{U}_j \mathbf{U}_j^h, \tag{6.27}$$

es decir, como combinación lineal de matrices de rango 1, cada una de las cuales representa la matriz de proyección ortogonal sobre su correspondiente subespacio columna, siendo estos subespacios ortogonales entre sí y los autovalores de $\mathbf{A}$ los coeficientes de dicha combinación lineal. Nótese que el número de sumandos no nulos es igual al rango de $\mathbf{A}$.

Si ahora ordenamos los autovalores no nulos de $\mathbf{A}$ agrupando aquellos que sean iguales, es decir, consideramos $\alpha_1, \alpha_2, \ldots, \alpha_p$ autovalores de multiplicidades respectivas $m_1, m_2, \ldots, m_p$, con $m_1 + \ldots + m_p = r(\mathbf{A})$, los sumandos correspondientes al autovalor α_k en la expresión (6.27) se agrupan como sigue:

$$\alpha_k(\mathbf{U}_{k_1}\mathbf{U}_{k_1}^h + \ldots + \mathbf{U}_{k_{m_k}}\mathbf{U}_{k_{m_k}}^h) = \alpha_k \mathbf{P}_{\ker(\mathbf{A}-\alpha_k\mathbf{I})} \tag{6.28}$$

en donde el subespacio propio asociado al autovalor α_k se supone engendrado por las columnas $\mathbf{U}_{k_1}, \ldots, \mathbf{U}_{k_{m_k}}$ de $\mathbf{U}$.

En conclusión, se verifica la siguiente proposición:

Proposición 33 (Descomposición espectral de una matriz normal) *Toda matriz normal* $\mathbf{A}$ *puede expresarse como*

$$\mathbf{A} = \sum_{k=1}^{p} \alpha_k \mathbf{P}_k, \tag{6.29}$$

en donde α_k *son los autovalores (distintos dos a dos) no nulos de* $\mathbf{A}$ *y* $\mathbf{P}_k$ *representa la matriz de proyección ortogonal sobre el subespacio propio* $\ker(\mathbf{A} - \alpha_k\mathbf{I})$ *(estos subespacios son ortogonales dos a dos).*

6.4. Matrices hermíticas

6.4.1. Formas cuadráticas

Definición 67 (forma cuadrática) *Dada una matriz hermítica* $\mathbf{A} \in \mathbb{C}^{n\times n}$*, se llama forma cuadrática asociada a* $\mathbf{A}$ *a la función de* $\mathbb{C}^n$ *a* $\mathbb{R}$ *definida mediante*

$$\mathbf{x} \mapsto \mathbf{x}^h\mathbf{A}\mathbf{x}. \tag{6.30}$$

Por el Teorema espectral, toda matriz hermítica $\mathbf{A}$ es diagonalizable unitariamente, y sus autovalores son reales. Obsérvese que, si λ es un autovalor de $\mathbf{A}$ y $\mathbf{u}$ un autovector asociado a λ,

$$\mathbf{u}^h\mathbf{A}\mathbf{u} = \lambda\|\mathbf{u}\|^2. \tag{6.31}$$

Teorema 14 *Sea* $\mathbf{A} \in \mathbb{C}^{n\times n}$ *hermítica, y sean* $\lambda_{\text{mín}}$ *y* $\lambda_{\text{máx}}$ *sus autovalores mínimo y máximo, respectivamente. Entonces, para todo* $\mathbf{x} \in \mathbb{C}^n$*, se verifica que*

$$\lambda_{\text{mín}} \|\mathbf{x}\|^2 \leq \mathbf{x}^h \mathbf{A}\mathbf{x} \leq \lambda_{\text{máx}} \|\mathbf{x}\|^2 , \tag{6.32}$$

alcanzándose cada una de las igualdades en todos los vectores propios asociados, respectivamente, a $\lambda_{\text{mín}}$ *y* $\lambda_{\text{máx}}$.

Demostración: Por el teorema espectral para matrices hermíticas, existe una matriz unitaria $\mathbf{Q}$ tal que

$$\mathbf{Q}^h\mathbf{A}\mathbf{Q} = \mathbf{D} = \operatorname{diag}(\lambda_{\text{mín}}, \cdots, \lambda_{\text{máx}}) . \tag{6.33}$$

Denotando $\mathbf{x} = \mathbf{Q}\mathbf{y}$ se tiene que

$$\mathbf{x}^h\mathbf{A}\mathbf{x} = \mathbf{y}^h\mathbf{Q}^h\mathbf{A}\mathbf{Q}\mathbf{y} = \mathbf{y}^h \operatorname{diag}(\lambda_{\text{mín}}, \cdots, \lambda_{\text{máx}})\, \mathbf{y} = \lambda_{\text{mín}}|y_1|^2 + \cdots + \lambda_{\text{máx}}|y_n|^2 \tag{6.34}$$

pero dicha cantidad puede acotarse superiormente por

$$\lambda_{\text{máx}} \left(|y_1|^2 + \cdots + |y_n|^2\right) = \lambda_{\text{máx}} \|\mathbf{y}\|^2 = \lambda_{\text{máx}} \|\mathbf{x}\|^2 \tag{6.35}$$

(se ha usado que $\|\mathbf{y}\| = \|\mathbf{Q}\mathbf{y}\| = \|\mathbf{x}\|$ pues $\mathbf{Q}$ es unitaria). Así pues, $\mathbf{x}^h\mathbf{A}\mathbf{x} \leq \lambda_{\text{máx}} \|\mathbf{x}\|^2$.

Por la ecuación (6.31), sabemos que la igualdad se alcanza si $\mathbf{x}$ es un vector propio asociado a $\lambda_{\text{máx}}$.

Análogamente, para la acotación inferior, se observa que la expresión (6.34) es mayor o igual que

$$\lambda_{\text{mín}} \left(|y_1|^2 + \cdots + |y_n|^2\right) = \lambda_{\text{mín}} \|\mathbf{y}\|^2 = \lambda_{\text{mín}} \|\mathbf{x}\|^2 \tag{6.36}$$

luego $\lambda_{\text{mín}} \|\mathbf{x}\|^2 \leq \mathbf{x}^h\mathbf{A}\mathbf{x}$, concluyendo la demostración. □

6.4.2. Cociente de Rayleigh

Definición 68 (cociente de Rayleigh) *Dada* $\mathbf{A} \in \mathbb{C}^{n\times n}$ *hermítica, el cociente de Rayleigh es la función de* $\mathbb{C}^n \setminus \{\mathbf{0}\}$ *a* $\mathbb{R}$ *definida por el cociente*

$$R(\mathbf{x}) = \frac{\mathbf{x}^h\mathbf{A}\mathbf{x}}{\|\mathbf{x}\|^2} \qquad \forall \mathbf{x} \neq \mathbf{0}. \tag{6.37}$$

Obsérvese que si $\mathbf{u}$ es autovector de $\mathbf{A}$ asociado al autovalor λ, en virtud de la ecuación (6.31) se tiene que

$$R(\mathbf{u}) = \frac{\mathbf{u}^h\mathbf{A}\mathbf{u}}{\|\mathbf{u}\|^2} = \lambda. \tag{6.38}$$

Por otro lado, la ecuación (6.32) garantiza que

$$\lambda_{\text{mín}} \leq \frac{\mathbf{x}^h\mathbf{A}\mathbf{x}}{\|\mathbf{x}\|^2} \leq \lambda_{\text{máx}} \qquad \forall \mathbf{x} \neq \mathbf{0}, \tag{6.39}$$

y la continuidad de la función R en $\mathbb{C}^n \setminus \{\mathbf{0}\}$ garantiza que se alcanzan todos los valores del intervalo $[\lambda_{\text{mín}}, \lambda_{\text{máx}}]$.

Así pues, concluimos lo siguiente:

Proposición 34 *Sea* $\mathbf{A}$ *real y simétrica, sean* $\lambda_{\text{mín}}, \lambda_{\text{máx}}$ *los autovalores mínimo y máximo, respectivamente, de* $\mathbf{A}$*; entonces*

$$\min_{\mathbf{x}\neq\mathbf{0}} \frac{\mathbf{x}^h\mathbf{A}\mathbf{x}}{\|\mathbf{x}\|^2} = \lambda_{\text{mín}} \quad ; \quad \max_{\mathbf{x}\neq\mathbf{0}} \frac{\mathbf{x}^h\mathbf{A}\mathbf{x}}{\|\mathbf{x}\|^2} = \lambda_{\text{máx}}, \tag{6.40}$$

alcanzándose dichos valores en los correspondientes autovectores de $\mathbf{A}$ *asociados a* $\lambda_{\text{mín}}$ *y* $\lambda_{\text{máx}}$.

6.4.3. Clasificación de matrices hermíticas

Como hemos visto, los signos que puede tomar una forma cuadrática están determinados por los de los autovalores de la matriz hermítica que la define. Ello motiva la siguiente clasificación de las matrices hermíticas:

Sea $\mathbf{A}$ una matriz hermítica; se dice que:

- $\mathbf{A}$ es **semidefinida positiva** si $\mathbf{x}^h\mathbf{A}\mathbf{x} \geq 0\ \forall\mathbf{x}$, lo que equivale a que sus autovalores sean positivos o nulos ($\lambda_k \geq 0$).
  - En particular, $\mathbf{A}$ es **definida positiva** si $\mathbf{x}^h\mathbf{A}\mathbf{x} > 0\ \forall\mathbf{x} \neq \mathbf{0}$, es decir, si sus autovalores son positivos ($\lambda_k > 0$).
- $\mathbf{A}$ es **semidefinida negativa** si $\mathbf{x}^h\mathbf{A}\mathbf{x} \leq 0\ \forall\mathbf{x}$; equivalentemente, si sus autovalores son negativos o nulos ($\lambda_k \leq 0$).
  - En particular, $\mathbf{A}$ es **definida negativa** si $\mathbf{x}^h\mathbf{A}\mathbf{x} < 0\ \forall\mathbf{x} \neq \mathbf{0}$ o, lo que es lo mismo, si sus autovalores son negativos ($\lambda_k < 0$).
- $\mathbf{A}$ es **indefinida** si existen vectores $\mathbf{x}, \mathbf{y}$ tales que $\mathbf{x}^h\mathbf{A}\mathbf{x} > 0$ e $\mathbf{y}^h\mathbf{A}\mathbf{y} < 0$; ello equivale a que $\mathbf{A}$ posea algún autovalor positivo y alguno negativo.

Definición 69 (signatura) *Se define la signatura de una matriz hermítica* $\mathbf{A}$*, y se denota por* $\sigma(\mathbf{A})$*, como el número de sus autovalores positivos menos el número de sus autovalores negativos.*

Observación: Obsérvese que la signatura, en conjunción con el rango de la matriz (que coincide con el número de sus autovalores no nulos) nos permite clasificar la matriz.

La siguiente proposición cobrará importancia cuando estudiemos, en el último tema, la descomposición en valores singulares. Obsérvese que cualquier matriz de la forma $\mathbf{A}^h\mathbf{A}$ es hermítica y, por tanto, diagonalizable unitariamante con autovalores reales.

Proposición 35 *Para cualquier matriz* $\mathbf{A} \in \mathbb{C}^{m\times n}$*, los autovalores de* $\mathbf{A}^h\mathbf{A}$ *son no negativos. En particular, si* $r(\mathbf{A}) = n$ *(columnas linealmente independientes), los autovalores de* $\mathbf{A}^h\mathbf{A}$ *son todos positivos.*

Demostración: Sea $\lambda \in \mathbb{C}$ un autovalor de $\mathbf{A}^h\mathbf{A}$ de vector propio unitario $\mathbf{u}$. Entonces

$$0 \leq \|\mathbf{A}\mathbf{u}\|^2 = (\mathbf{A}\mathbf{u})^h\mathbf{A}\mathbf{u} = \mathbf{u}^h\mathbf{A}^h\mathbf{A}\mathbf{u} = \mathbf{u}^h\lambda\mathbf{u} = \lambda\|\mathbf{u}\|^2 = \lambda. \tag{6.41}$$

En particular, si $r(\mathbf{A}) = n$, $\mathbf{u} \neq 0 \Rightarrow \mathbf{A}\mathbf{u} \neq \mathbf{0}$, luego $\lambda = \|\mathbf{A}\mathbf{u}\|^2$ es positivo.

Corolario: Si $\mathbf{A} \in \mathbb{C}^{m\times n}$, $\mathbf{A}^h\mathbf{A}$ es semidefinida positiva (definida positiva en caso de ser $r(\mathbf{A}) = n$).

6.4.4. Matrices reales simétricas

Toda matriz real simétrica es en particular una matriz hermítica. Por tanto, se le puede aplicar todo lo visto en esta sección. Pero además, existe un procedimiento más sencillo para clasificarla, sin necesidad de calcular sus autovalores. Nótese que, a estos efectos, no nos interesan las magnitudes de los autovalores, sino tan solo sus signos.

Definición 70 (Congruencia de matrices) *Dadas dos matrices* $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n\times n}$*, se dicen congruentes entre sí si existe* $\mathbf{P} \in \mathbb{R}^{n\times n}$ ***regular*** *tal que* $\mathbf{B} = \mathbf{P}^t\mathbf{A}\mathbf{P}$.

Observación: Obsérvese que la semejanza ortogonal es un caso particular (muy restrictivo) de congruencia: dos matrices reales simétricas son semejantes ortogonalmente si son congruentes y, además, la matriz de paso $\mathbf{P}$ es ortogonal ($\mathbf{P}^t = \mathbf{P}^{-1}$).

Teorema 15 (Ley de inercia de Sylvester) *Dos matrices reales simétricas, congruentes entre sí, tienen el mismo número de autovalores positivos, el mismo número de negativos y el mismo número de nulos.*

En virtud de este teorema, cuya demostración omitimos, podemos establecer el siguiente procedimiento, denominado **diagonalización por congruencia**, para encontrar los signos de los autovalores de $\mathbf{A}$ real y simétrica, que consistirá simplemente en hallar una matriz congruente con $\mathbf{A}$ que sea diagonal. Sus elementos diagonales no serán, en general, los autovalores de $\mathbf{A}$, pero tendrán los mismos signos que éstos.

Observamos primero que si realizamos una operación elemental sobre las filas de $\mathbf{A}$ y a continuación la misma operación sobre sus columnas, se conserva la simetría de la matriz.

Ello equivale a la post- y premultiplicación por una matriz elemental y su transpuesta: $\mathbf{E}_1^t\mathbf{A}\mathbf{E}_1$, obteniéndose una matriz congruente con la primera.

Reiteramos este proceso hasta obtener una matriz diagonal (lo cual siempre es posible porque equivale a triangularizar $\mathbf{A}$ mediante operaciones elementales de filas), y llegamos a:

$$\mathbf{E}_k^t \cdots \mathbf{E}_2^t\mathbf{E}_1^t\mathbf{A}\mathbf{E}_1\mathbf{E}_2\cdots\mathbf{E}_k = \mathbf{D}, \text{ es decir, } (\mathbf{E}_1\mathbf{E}_2\cdots\mathbf{E}_k)^t\mathbf{A}\mathbf{E}_1\mathbf{E}_2\cdots\mathbf{E}_k = \mathbf{D}. \quad (6.42)$$

Así pues, tomando $\mathbf{P} = \mathbf{E}_1\mathbf{E}_2\cdots\mathbf{E}_k$, esto es, la matriz resultante de realizar sobre la identidad las mismas operaciones **de columnas** efectuadas sobre $\mathbf{A}$, tenemos que $\mathbf{P}^t\mathbf{A}\mathbf{P} = \mathbf{D}$.

Ejemplo: Clasifíquese la matriz real y simétrica $\begin{pmatrix} 1 & -2 & 1 \\ -2 & 2 & 0 \\ 1 & 0 & -1 \end{pmatrix}$.

Solución: Diagonalizamos por congruencia realizando operaciones gaussianas:

$$\left\{ \begin{array}{rcl} F_2' & = & F_2 + 2F_1 \\ F_3' & = & F_3 - F_1 \end{array} \right\} \begin{pmatrix} 1 & -2 & 1 \\ 0 & -2 & 2 \\ 0 & 2 & -2 \end{pmatrix} \left\{ \begin{array}{rcl} C_2' & = & C_2 + 2C_1 \\ C_3' & = & C_3 - C_1 \end{array} \right\} \begin{pmatrix} 1 & 0 & 0 \\ 0 & -2 & 2 \\ 0 & 2 & -2 \end{pmatrix}$$

$$\left\{ \begin{array}{rcl} F_3' & = & F_3 + F_2 \\ C_3' & = & C_3 + C_2 \end{array} \right\} \begin{pmatrix} 1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Así pues, la matriz es indefinida, siendo su rango igual a 2 y su signatura igual a 0.

Nota: Si además deseamos conocer la matriz $\mathbf{P}$ tal que $\mathbf{P}^t\mathbf{A}\mathbf{P} = \mathbf{D}$, en donde $\mathbf{A}$ es la matriz de partida y $\mathbf{D}$ la diagonal congruente con ella, debemos realizar las mismas operaciones sólo por columnas sobre la matriz identidad, obteniéndose en este caso:

$$\mathbf{P} = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$

6.5. Ejercicios

6.1.– Compruébese que las dos matrices

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

son semejantes pero no son unitariamente semejantes.

6.2.– Determínese una base ortonormal de vectores propios de la matriz

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

6.3.– Diagonalícese unitariamente la matriz

$$\begin{pmatrix} 0 & -1 & 2 \\ 1 & 0 & -2 \\ -2 & 2 & 0 \end{pmatrix}.$$

6.4.– Diagonalícese unitariamente la matriz

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

Indicación: Utilícense los resultados del ejercicio anterior.

6.5.– Diagonalícese ortogonalmente la matriz

$$\begin{pmatrix} 1 & -1 & 1 & 1 \\ -1 & 1 & -1 & -1 \\ 1 & -1 & -1 & -1 \\ 1 & -1 & -1 & -1 \end{pmatrix}.$$

6.6.– Diagonalícese unitariamente la matriz

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

6.7.– *(Ejercicio avanzado)* Demuéstrese que una matriz compleja $\mathbf{A}$ es normal si y solo si cumple que

$$\forall \mathbf{x} \in \mathbb{C}^n, \quad \|\mathbf{A}^h\mathbf{x}\| = \|\mathbf{A}\mathbf{x}\|.$$

6.8.– Pruébese que toda matriz ortogonal del espacio tridimensional euclídeo ordinario se reduce a una de las tres posibilidades:

a) Un giro respecto de un eje.

b) Una simetría respecto de un plano.

c) La composición de dos transformaciones de los tipos anteriores.

6.9.– Determínense los valores del parámetro a que hacen definida positiva la matriz

$$\begin{pmatrix} a & 1 & 1 \\ 1 & a & 1 \\ 1 & 1 & a \end{pmatrix}.$$

6.10.– Clasifíquese la siguiente matriz real y simétrica

$$\begin{pmatrix} a+b & -a & -b \\ -a & a+c & -c \\ -b & -c & b+c \end{pmatrix}, \quad a,b,c>0.$$

6.11.– Calcúlense los valores máximo y mínimo de la función R definida en $\mathbb{R}^3 \backslash \{(0,0,0)\}$ mediante

$$R(x,y,z) = \frac{3(x^2+y^2+z^2)+2(xy+yz+zx)}{x^2+y^2+z^2},$$

así como los vectores en los que se alcanzan dichos valores.

6.12.– Calcúlense los valores máximo y mínimo que alcanza sobre $\mathbb{R}^2 \backslash \{(0,0)\}$ el cociente

$$\frac{\begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} 5 & 2 \\ 0 & 5 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}}{\|(x,y)\|^2},$$

indicando sobre qué vectores toma dichos valores.

6.13.– Diagonalícese por congruencia cada una de las siguientes matrices, indicando el rango y la signatura:

$$a) \begin{pmatrix} 1 & -2 & 2 \\ -2 & 4 & -4 \\ 2 & -4 & 4 \end{pmatrix}; \quad b) \begin{pmatrix} 2 & -3 & -1 \\ -3 & 3 & 1 \\ -1 & 1 & 1 \end{pmatrix}; \quad c) \begin{pmatrix} 3 & 1 & -5 & -2 \\ 1 & 0 & -2 & -1 \\ -5 & -2 & 8 & 3 \\ -2 & -1 & 3 & 1 \end{pmatrix}.$$

6.5.1. Cuestiones

6.14.– Sean a, b, c números reales no nulos. Se considera la matriz $\mathbf{A} = \begin{pmatrix} 13 & a & b \\ -a & 13 & c \\ -b & -c & 13 \end{pmatrix}$.

Determínese la veracidad o falsedad de las siguientes afirmaciones:

1. $\mathbf{A}$ no es diagonalizable unitariamente.
2. $\mathbf{A}$ es diagonalizable unitariamente y todos sus autovalores son reales.
3. $\mathbf{A}$ es diagonalizable unitariamente y todos sus autovalores son imaginarios puros.
4. $\mathbf{A}$ es diagonalizable unitariamente y todos sus autovalores tienen parte real igual a 13.
5. $\mathbf{A}$ es normal.
6. Ningún autovalor de $\mathbf{A}$ es real.
7. Los subespacios propios de $\mathbf{A}$ son ortogonales dos a dos.

6.15.– Sea $\mathbf{A} \in \mathbb{C}^{n \times n}$. Determínese la veracidad o falsedad de las siguientes afirmaciones:

1. Si todos los autovalores de $\mathbf{A}$ son reales, entonces $\mathbf{A}$ es hermítica.
2. Si $\mathbf{A}^h\mathbf{A} = \mathbf{A}\mathbf{A}^h$ y todos sus autovalores tienen módulo unidad, entonces $\mathbf{A}$ es unitaria.
3. $\mathbf{A}$ es diagonalizable unitariamente $\Leftrightarrow \mathbf{A}^h\mathbf{A} = \mathbf{A}\mathbf{A}^h$.
4. Si $\mathbf{A}$ es normal y todos sus autovalores son reales, entonces $\mathbf{A}$ es real y simétrica.
5. $\mathbf{A}$ es simétrica $\Rightarrow \mathbf{A}$ es diagonalizable unitariamente.
6. $\mathbf{A}$ es normal $\Rightarrow$ sus subespacios propios son ortogonales dos a dos.

Capítulo 7

Descomposición en valores singulares

Introducción

Con este tema concluye la asignatura de Álgebra. La Descomposición en Valores Singulares (DVS) utiliza, compendia y generaliza el contenido de los seis temas anteriores. Así, en la DVS aparecen bases ortogonales de la imagen y el núcleo de una matriz, así como sus suplementarios ortogonales (temas 1 y 3). A partir de la DVS se calcula con facilidad la pseudoinversa de una matriz, que puede considerarse como una generalización del concepto de inversa (tema 2) a cualquier matriz, no necesariamente cuadrada. Cuando la matriz es de rango máximo, la pseudoinversa conduce a la solución de mínimos cuadrados si el sistema es incompatible y a la de mínima norma si es indeterminado. En el caso más general permite hallar la solución de mínima norma de entre las que minimizan el error cuadrático. La DVS permite hallar fácilmente matrices de proyección ortogonal (tema 4).

Por otra parte, la DVS puede verse también como una generalización de la diagonalización por semejanza (tema 5) o de la descomposición espectral de una matriz normal (tema 6), con la ventaja de que existe siempre. El cálculo de la DVS se apoya en el cálculo de los valores y vectores propios de una matriz normal (tema 6), en el teorema de compleción de la base (tema 1) y en la ortogonalización de Gram-Schmidt (tema 3).

La DVS tiene numerosas aplicaciones teóricas y prácticas en diversas ramas de la ingeniería. Sin embargo, su generalidad no la convierte en la panacea general, en el método universal siempre aplicable. Cuando el tamaño crece, resulta mucho más costosa que los métodos alternativos basados en la eliminación de Gauss o en la factorización QR. Por este motivo, su uso debe limitarse a los casos en que sea imprescindible.

7.1. Descomposición en valores singulares (DVS) de una matriz

Teorema 16 *(Existencia de la DVS). Para cualquier matriz $\mathbf{A} \in \mathbb{C}^{m\times n}$, de rango igual a r, existen $\mathbf{U} \in \mathbb{C}^{m\times m}$, unitaria, $\mathbf{V} \in \mathbb{C}^{n\times n}$, unitaria, y $\mathbf{\Sigma} = diag(\sigma_1, \sigma_2, \ldots, \sigma_p) \in \mathbb{R}^{m\times n}$, donde $p = \text{mín}\{m, n\}$ tales que*

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^h \tag{7.1}$$

y $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > \sigma_{r+1} = \cdots = \sigma_p = 0$.

Demostración: Por la importancia del tema se ofrece a continuación una demostración de tipo constructivo. Es importante porque es el método que se sigue en la práctica para calcular una DVS de una matriz.

Se va a demostrar la existencia de la DVS detallando un procedimiento bien definido para calcular las matrices $\mathbf{U}$, $\mathbf{\Sigma}$ y $\mathbf{V}$, tales que la ec. (7.1) se cumple para cualquier matriz compleja $\mathbf{A}$. Sin pérdida de generalidad se supondrá $m \geq n$. Si $m < n$ el procedimiento que sigue se puede aplicar a $\mathbf{A}^h$.

La matriz $\mathbf{A}^h\mathbf{A}$ es hermítica, con valores propios no negativos. Suponiendo que la factorización (7.1) existe, esta matriz resulta ser:

$$\mathbf{A}^h\mathbf{A} = \mathbf{V}\mathbf{\Sigma}^t\mathbf{U}^h\mathbf{U}\mathbf{\Sigma}\mathbf{V}^h = \mathbf{V}\mathbf{\Sigma}^t\mathbf{\Sigma}\mathbf{V}^h = \mathbf{V}\mathbf{D}_n\mathbf{V}^h \tag{7.2}$$

de donde se deduce que $\mathbf{V}$ es una matriz unitaria que diagonaliza a $\mathbf{A}^h\mathbf{A}$, y $\mathbf{D}_n$ es la correspondiente matriz diagonal de valores propios, reales y no negativos, que se suponen ordenados de mayor a menor,

$$\mathbf{D}_n = \text{diag}(\lambda_1, \ldots, \lambda_n), \qquad \lambda_1 \geq \cdots \geq \lambda_r > \lambda_{r+1} = \cdots = \lambda_n = 0 \tag{7.3}$$

Teniendo en cuenta que según la ecuación (7.2), $\mathbf{D}_n = \mathbf{\Sigma}^t\mathbf{\Sigma}$, la matriz rectangular y diagonal $\mathbf{\Sigma}$, con elementos diagonales no negativos, queda unívocamente determinada en la forma,

$$\mathbf{\Sigma} = \text{diag}(\sigma_1, \ldots, \sigma_n) \in \mathbb{R}^{m\times n}, \quad \sigma_1 \geq \cdots \geq \sigma_r > \sigma_{r+1} = \cdots = \sigma_n = 0 \tag{7.4}$$

cuyos σ_i son:

$$\sigma_i = \sqrt{\lambda_i(\mathbf{A}^h\mathbf{A})}, \qquad i = 1, 2, \ldots, n \tag{7.5}$$

Las expresiones (7.2) y (7.4) permiten calcular $\mathbf{V}$ y $\mathbf{\Sigma}$. Queda por determinar la matriz $\mathbf{U}$ de modo que sea unitaria y satisfaga la ec. (7.1). A partir de dicha ecuación se tendrá,

$$\mathbf{A}\mathbf{V} = \mathbf{U}\mathbf{\Sigma} \tag{7.6}$$

que en forma desarrollada se puede escribir,

$$\mathbf{A}\begin{pmatrix} \mathbf{V}_1 & \mathbf{V}_2 & \cdots & \mathbf{V}_n \end{pmatrix} = \begin{pmatrix} \sigma_1\mathbf{U}_1 & \sigma_2\mathbf{U}_2 & \cdots & \sigma_n\mathbf{U}_n \end{pmatrix} \tag{7.7}$$

De donde se deduce que

$$\mathbf{U}_j = \frac{\mathbf{A}\mathbf{V}_j}{\sigma_j}, \qquad j = 1, 2, \ldots, r \tag{7.8}$$

La ecuación (7.8) sólo permite calcular las r columnas de $\mathbf{U}$ asociadas con valores singulares no nulos. Como esta matriz debe ser unitaria, hay que demostrar que las columnas obtenidas mediante la ec. (7.8) constituyen un sistema ortonormal. En efecto,

$$\mathbf{U}_i^h\mathbf{U}_j = \frac{1}{\sigma_i\sigma_j}\mathbf{V}_i^h\mathbf{A}^h\mathbf{A}\mathbf{V}_j = \frac{1}{\sigma_i\sigma_j}\lambda_j\delta_{ij} = \delta_{ij}, \qquad j = 1, 2, \ldots, r \tag{7.9}$$

Las restantes columnas de $\mathbf{U}$ se pueden calcular ampliando el sistema ortonormal $\{\mathbf{U}_1, \mathbf{U}_2, \ldots, \mathbf{U}_r\}$ a una base ortonormal $(\mathbf{U}_1, \ldots, \mathbf{U}_r, \mathbf{U}_{r+1}, \ldots, \mathbf{U}_m)$ de $\mathbb{C}^m$ de modo que $\mathbf{U} = (\mathbf{U}_1|\cdots|\mathbf{U}_m)$ resulte unitaria.

Para $j = r+1, \ldots, n$, se cumple $\mathbf{A}^h\mathbf{A}\mathbf{V}_j = 0$, y, al ser $\ker\mathbf{A} = \ker\big(\mathbf{A}^h\mathbf{A}\big)$, se tiene que $\mathbf{A}\mathbf{V}_j = \mathbf{0}$; con lo cual queda totalmente demostrada la fórmula (7.1). $\square$

Definición 71 (valores singulares) *Si* $\mathbf{A} \in \mathbb{C}^{m\times n}$, *los*

$$\sigma_j = \sqrt{\lambda_j(\mathbf{A}^h\mathbf{A})}, \; j = 1, 2, \ldots, p \tag{7.10}$$

donde $p = \text{mín}\{m, n\}$, *se denominan* ***valores singulares*** *de* $\mathbf{A}$.

Definición 72 (descomposición en valores singulares) *La fórmula* $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^h$ *se denomina* ***descomposición en valores singulares (DVS) de*** $\mathbf{A}$.

Nótese que $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^h = \sum_{j=1}^r \sigma_j\mathbf{U}_j\mathbf{V}_j^h$. Es decir, toda $\mathbf{A}$ de rango r se puede expresar como suma de r matrices de rango 1 utilizando la DVS. Esta observación da lugar a la siguiente definición:

Definición 73 (forma reducida de la DVS) *Si* $\{\mathbf{U}_j, j = 1, ..., r\}$ *es una familia ortonormal de vectores de* $\mathbb{C}^m$, $\{\mathbf{V}_j, j = 1, ..., r\}$ *es una familia ortonormal de vectores de* $\mathbb{C}^n$, *y* $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$, *entonces la expresión*

$$\mathbf{A} = \sum_{j=1}^{r} \sigma_j\mathbf{U}_j\mathbf{V}_j^h$$

se denomina ***forma reducida de la descomposición en valores singulares de*** $\mathbf{A}$.

Observación: La DVS no es única, ya que $\mathbf{V}$ no es única y $\mathbf{U}$, en general, tampoco lo es. Según el método de cálculo que se acaba de describir, la indeterminación está en el cálculo de los vectores propios de $\mathbf{A}^h\mathbf{A}$, que constituyen las columnas de la matriz unitaria $\mathbf{V}$ (para valores propios simples de $\mathbf{A}^h\mathbf{A}$ podrían tomarse vectores propios unitarios relacionados por medio del factor $e^{i\varphi}$; para valores propios múltiples podrían elegirse diferentes bases ortonormales del correspondiente subespacio propio). Tampoco el cálculo de las columnas de $\mathbf{U}$ correspondientes a los valores singulares nulos está determinado (la ampliación de la familia ortonormal $\{\mathbf{U}_1, \mathbf{U}_2, \ldots, \mathbf{U}_r\}$, cuando $r < m$, a una base ortonormal de $\mathbb{C}^m$ no es única; sólo queda unívocamente determinada la matriz $\mathbf{U}$ cuando todos los valores singulares son no nulos).

Observación: Las columnas de $\mathbf{U}$ son vectores propios de la matriz $\mathbf{A}\mathbf{A}^h$. En efecto, sustituyendo $\mathbf{A}$ por su DVS, dada por (7.1),

$$\mathbf{A}\mathbf{A}^h = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^h\mathbf{V}\mathbf{\Sigma}^t\mathbf{U}^h = \mathbf{U}\mathbf{\Sigma}\mathbf{\Sigma}^t\mathbf{U}^h = \mathbf{U}\mathbf{D}_m\mathbf{U}^h \tag{7.11}$$

Sin embargo, no es correcto calcular de modo independiente la matriz $\mathbf{U}$ según la expresión (7.11) y la matriz $\mathbf{V}$ según la expresión (7.2). Una vez calculada una de ellas, la segunda debe ser calculada según la expresión (7.1), tal como se ha hecho para $\mathbf{U}$ a partir de la expresión (7.8).

Observación: Las matrices $\mathbf{A}^h\mathbf{A}$ y $\mathbf{A}\mathbf{A}^h$ tienen los mismos valores propios no nulos y con las mismas multiplicidades. Si $m > n$ o $n > m$, el valor propio nulo tendrá multiplicidad algebraica aumentada en $m - n$ en $\mathbf{A}\mathbf{A}^h$ o en $n - m$ en $\mathbf{A}^h\mathbf{A}$, respectivamente. De hecho, se sabe que

$$\lambda^m \chi_{\mathbf{A}^h\mathbf{A}}(\lambda) = \lambda^n \chi_{\mathbf{A}\mathbf{A}^h}(\lambda) \tag{7.12}$$

Este resultado permite calcular los valores singulares de $\mathbf{A}$ (véase (7.10)) a partir de los valores propios de $\mathbf{A}^h\mathbf{A}$ o de $\mathbf{A}\mathbf{A}^h$, siendo conveniente elegir la de menor orden. En el caso de efectuar la DVS a partir de $\mathbf{A}\mathbf{A}^h$, hay que tener en cuenta que la DVS será la de $\mathbf{A}^h = \mathbf{V}\mathbf{\Sigma}^t\mathbf{U}^h$.

7.1.1. Propiedades de la DVS

Se cumplen las siguientes propiedades:

1. Los vectores columna de las matrices $\mathbf{U}$ y $\mathbf{V}$ proporcionan bases ortonormales de los cuatro subespacios asociados a la matriz $\mathbf{A}$, de la siguiente forma:

$$\begin{aligned}
\mathrm{Im}\mathbf{A} &= \mathrm{Im}\left(\mathbf{A}\mathbf{A}^h\right) = L[\mathbf{U}_1, \ldots, \mathbf{U}_r] \\
\ker\mathbf{A} &= \ker\left(\mathbf{A}^h\mathbf{A}\right) = L[\mathbf{V}_{r+1}, \ldots, \mathbf{V}_n] \\
\ker\mathbf{A}^h &= L[\mathbf{U}_{r+1}, \ldots, \mathbf{U}_m] \\
\mathrm{Im}\mathbf{A}^h &= L[\mathbf{V}_1, \ldots, \mathbf{V}_r].
\end{aligned}$$

donde r es el rango de $\mathbf{A}$, y se ha utilizado que $\ker\mathbf{A}^h = (\mathrm{Im}\,\mathbf{A})^\perp$.

2. Si $\mathbf{A} \in \mathbb{C}^{n\times n}$ es normal, entonces sus valores singulares son los módulos de sus valores propios, comprobémoslo:

   Sea$\mathbf{W}$ una matriz unitaria que diagonalice $\mathbf{A}$, es decir:

   $$\mathbf{W}^h\mathbf{A}\mathbf{W} = \mathbf{D} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$$

   y podemos suponer los valores propios ordenados en módulo de mayor a menor $|\lambda_1| \geq \ldots \geq |\lambda_n| \geq 0$, entonces:

   $$\mathbf{W}^h\mathbf{A}^h\mathbf{A}\mathbf{W} = \mathbf{W}^h\mathbf{A}^h\mathbf{W}\mathbf{W}^h\mathbf{A}\mathbf{W} = \mathbf{D}^h\mathbf{D} = \mathrm{diag}(|\lambda_1|^2, \ldots, |\lambda_n|^2) = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_n^n),$$

   de donde se deduce la relación entre los valores propios y los valores singulares: $\sigma_j = |\lambda_j|$.

Representación esquemática de la DVS

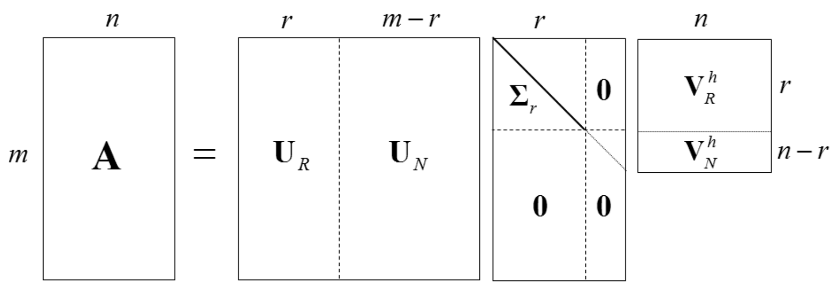

Figura 7.1: Representación gráfica de la DVS de una matriz de rango r.

La Figura 7.1 representa la DVS de una matriz rectangular con $r < n < m$. Las matrices $\mathbf{U}_R$ y $\mathbf{V}_R$ corresponden con las r primeras columnas de $\mathbf{U}$ y $\mathbf{V}$, respectivamente. La DVS se puede escribir de las siguientes formas (forma general y forma reducida):

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^h = \mathbf{U}_R\mathbf{\Sigma}_r\mathbf{V}_R^h = \sum_{j=1}^{r}\sigma_j\mathbf{U}_j\mathbf{V}_j^h \tag{7.13}$$

Obsérvense, en primer lugar, los tamaños de las distintas matrices: $\mathbf{U}$ y $\mathbf{V}$ son cuadradas (unitarias), mientras que $\mathbf{\Sigma}$ tiene el mismo tamaño que $\mathbf{A}$. Obsérvese, también, que las submatrices $\mathbf{U}_N$ y $\mathbf{V}_N$ no intervienen en el resultado de la DVS, al estar multiplicadas por submatrices nulas de la matriz $\mathbf{\Sigma}$. Recuérdese que las columnas de $\mathbf{U}_R$ son una base ortonormal de Im$\mathbf{A}$ y que las columnas de $\mathbf{V}_R$ lo son de Im$\mathbf{A}^h$. Asimismo las columnas de $\mathbf{V}_N$ y de $\mathbf{U}_N$ son respectivamente bases ortonormales de ker$\mathbf{A}$ y de ker$\mathbf{A}^h$.

7.1.2. Expresiones de los valores singulares máximo y mínimo de una matriz

Sea $\mathbf{A} \in \mathbb{C}^{m\times n}$, donde $m \geq n$. Teniendo en cuenta el concepto de norma euclídea de un vector, definida en el capítulo 3, se tienen

$$\sigma_{máx} = \sqrt{\lambda_{máx}(\mathbf{A}^h\mathbf{A})} = \max_{\mathbf{x}\neq\mathbf{0}} \frac{\|\mathbf{Ax}\|_2}{\|\mathbf{x}\|_2} = \max_{\|\mathbf{x}\|_2=1} \|\mathbf{Ax}\|_2 \tag{7.14}$$

$$\sigma_{mín} = \sqrt{\lambda_{mín}(\mathbf{A}^h\mathbf{A})} = \min_{\mathbf{x}\neq\mathbf{0}} \frac{\|\mathbf{Ax}\|_2}{\|\mathbf{x}\|_2} = \min_{\|\mathbf{x}\|_2=1} \|\mathbf{Ax}\|_2 \tag{7.15}$$

donde $\sigma_{máx}$ y $\sigma_{mín}$ son, respectivamente, los valores singulares máximo y mínimo de $\mathbf{A}$ y las normas euclídeas se han denotado con el subíndice 2.

La demostración puede realizarse de forma similar a la del teorema 13 del capítulo 6, considerando la diagonalización unitaria de la matriz hermítica $\mathbf{A}^h\mathbf{A}$, cuyos valores propios máximo y mínimo son, respectivamente, $\sigma^2_{máx}$ y $\sigma^2_{\sigma_{mín}}$.

En el caso de que $m < n$, bastaría utilizar las fórmulas (7.14) y (7.15) con la matriz $\mathbf{A}^h$ en vez de $\mathbf{A}$.

7.1.3. Matriz pseudoinversa

Definición 74 (matriz pseudoinversa) *La matriz pseudoinversa de $\mathbf{A} \in \mathbb{C}^{m\times n}$, siendo $\mathbf{U\Sigma V}^h = \sum_{j=1}^r \sigma_j \mathbf{U}_j \mathbf{V}_j^h$, una de sus DVS, se define*[1] *como*

$$\mathbf{A}^+ = \sum_{j=1}^r \sigma_j^{-1} \mathbf{V}_j \mathbf{U}_j^h \tag{7.16}$$

A partir de la definición anterior, si $\mathbf{\Sigma} = \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_r, 0, \ldots, 0) \in \mathbb{R}^{m\times n}$, con $\sigma_1, \sigma_2, \ldots, \sigma_r$ no nulos, es fácil comprobar que

$$\mathbf{\Sigma}^+ = \left(\begin{array}{ccc|c} \sigma_1^{-1} & \cdots & 0 & \\ \vdots & \ddots & \vdots & \mathbf{O} \\ 0 & \cdots & \sigma_r^{-1} & \\ \hline & \mathbf{O} & & \mathbf{O} \end{array}\right) \in \mathbb{R}^{n\times m}$$

Por tanto, la expresión anterior de la pseudoinversa de $\mathbf{A}$ puede escribirse, tambien, como

$$\mathbf{A}^+ = \mathbf{V\Sigma}^+\mathbf{U}^h$$

Proposición 36 *Dado el sistema $\mathbf{Ax} = \mathbf{b}$, donde $\mathbf{A} \in \mathbb{C}^{m\times n}$ y $\mathbf{b} \in \mathbb{C}^m$, la solución de mínimos cuadrados y mínima norma de dicho sistema es $\mathbf{x}_0 = \mathbf{A}^+\mathbf{b}$.*

[1] La unicidad de la matriz pseudoinversa será justificada más adelante.

Demostración: Por una parte, $\mathbf{x}_0 = \mathbf{A}^+\mathbf{b}$ es solución de mínimos cuadrados:
$\mathbf{A}\mathbf{x}_0 = \mathbf{A}\mathbf{A}^+\mathbf{b} = \mathbf{A}\left(\sum_{j=1}^r \sigma_j^{-1}\mathbf{V}_j\mathbf{U}_j^h\right)\mathbf{b} = \left(\sum_{j=1}^r \sigma_j^{-1}\mathbf{A}\mathbf{V}_j\mathbf{U}_j^h\right)\mathbf{b} =$
$\left(\sum_{j=1}^r \sigma_j^{-1}\sigma_j\mathbf{U}_j\mathbf{U}_j^h\right)\mathbf{b} = \left(\sum_{j=1}^r \mathbf{U}_j\mathbf{U}_j^h\right)\mathbf{b} = \mathbf{P}_{\mathrm{Im}\,\mathbf{A}}\mathbf{b}$.

Por otra parte, es solución de mínima norma:
$\mathbf{x}_0 = \mathbf{A}^+\mathbf{b} = \left(\sum_{j=1}^r \sigma_j^{-1}\mathbf{V}_j\mathbf{U}_j^h\right)\mathbf{b} = \sum_{j=1}^r \sigma_j^{-1}\mathbf{U}_j^h\mathbf{b}\mathbf{V}_j$; es decir, $\mathbf{x}_0 \in L\left[\{\mathbf{V}_1, \ldots, \mathbf{V}_r\}\right] = \mathrm{Im}\mathbf{A}^h = (\ker\mathbf{A})^\perp$. □

Proposición 37 *La pseudoinversa de $\mathbf{A} \in \mathbb{C}^{m\times n}$ está univocamente determinada.*

Demostración: Se cumple, por la proposición anterior, que $\forall\mathbf{b} \in \mathbb{C}^m$, el sistema $\mathbf{A}\mathbf{x} = \mathbf{b}$ tiene una **única** solución de mínimos cuadrados y mínima norma, dada por $\mathbf{A}^+\mathbf{b}$. Supóngase que existiese otra matriz $\mathbf{B} \in \mathbb{C}^{n\times m}$, tal que $\forall\mathbf{b} \in \mathbb{C}^m$ dicha solución fuese $\mathbf{B}\mathbf{b}$. Se tendría, pues, $\forall\mathbf{b} \in \mathbb{C}^m$, $\mathbf{B}\mathbf{b} = \mathbf{A}^+\mathbf{b}$, lo cual implica que $\mathbf{B} = \mathbf{A}^+$. □

7.2. Número de condición espectral de una matriz

En primer lugar se motivará el concepto de número de condición espectral de una matriz.

Perturbación del término independiente de un sistema lineal

Sean $\mathbf{A} \in \mathbb{C}^{m\times n}$, de rango igual a n, y el sistema compatible y determinado $\mathbf{A}\mathbf{x} = \mathbf{b} \neq \mathbf{0}$. Se desea estudiar la variación $\Delta\mathbf{x}_0$ del vector solución $\mathbf{x}_0$ del sistema inicial ante variaciones $\Delta\mathbf{b} \neq \mathbf{0}$ del vector $\mathbf{b}$, en el supuesto de que el sistema $\mathbf{A}(\mathbf{x} + \Delta\mathbf{x}) = \mathbf{b} + \Delta\mathbf{b}$ siga siendo compatible y determinado.

Se tiene

$$\begin{aligned} \mathbf{A}\mathbf{x}_0 &= \mathbf{b} \\ \mathbf{A}(\mathbf{x}_0 + \Delta\mathbf{x}_0) &= \mathbf{b} + \Delta\mathbf{b} \end{aligned}$$

que equivale a

$$\mathbf{A}\mathbf{x}_0 = \mathbf{b} \tag{7.17}$$
$$\mathbf{A}\Delta\mathbf{x}_0 = \Delta\mathbf{b} \tag{7.18}$$

Teniendo en cuenta (7.14) y (7.15) se obtiene

$$\sigma_{m\acute{a}x}(\mathbf{A}) \geq \frac{\|\mathbf{A}\mathbf{x}_0\|_2}{\|\mathbf{x}_0\|_2} = \frac{\|\mathbf{b}\|_2}{\|\mathbf{x}_0\|_2} \tag{7.19}$$
$$\sigma_{m\acute{\imath}n}(\mathbf{A}) \leq \frac{\|\mathbf{A}\Delta\mathbf{x}_0\|_2}{\|\Delta\mathbf{x}_0\|_2} = \frac{\|\Delta\mathbf{b}\|_2}{\|\Delta\mathbf{x}_0\|_2} \tag{7.20}$$

Y, a partir de lo anterior

$$\begin{aligned} \|\Delta \mathbf{x}_0\|_2 &\leq \frac{1}{\sigma_{mín}(\mathbf{A})} \|\Delta \mathbf{b}\|_2 \\ \frac{1}{\|\mathbf{x}_0\|_2} &\leq \sigma_{máx}(\mathbf{A}) \frac{1}{\|\mathbf{b}\|_2} \end{aligned}$$

Multiplicando miembro a miembro se llega a

$$\frac{\|\Delta \mathbf{x}_0\|_2}{\|\mathbf{x}_0\|_2} \leq \frac{\sigma_{máx}(\mathbf{A})}{\sigma_{mín}(\mathbf{A})} \frac{\|\Delta \mathbf{b}\|_2}{\|\mathbf{b}\|_2} \tag{7.21}$$

Definición 75 (número de condición espectral) *El número de condición espectral de* $\mathbf{A} \in \mathbb{C}^{m \times n}$*, de rango igual a* n*, que se denota* $c_2(\mathbf{A})$*, se define como*

$$c_2(\mathbf{A}) = \sqrt{\frac{\lambda_{máx}(\mathbf{A}^h \mathbf{A})}{\lambda_{mín}(\mathbf{A}^h \mathbf{A})}} = \frac{\sigma_{máx}(\mathbf{A})}{\sigma_{mín}(\mathbf{A})}.$$

Si $\mathbf{A} \in \mathbb{C}^{n \times n}$ es normal e invertible, teniendo en cuenta que los valores singulares de $\mathbf{A}$ coinciden con los módulos de los valores propios de $\mathbf{A}$, se obtiene

$$c_2(\mathbf{A}) = \frac{\sigma_{máx}(\mathbf{A})}{\sigma_{mín}(\mathbf{A})} = \frac{|\lambda_1(\mathbf{A})|}{|\lambda_n(\mathbf{A})|} \tag{7.22}$$

en el supuesto de que los valores propios de $\mathbf{A}$ cumplen $|\lambda_1(\mathbf{A})| \geq \cdots \geq |\lambda_n(\mathbf{A})| > 0$.

Es inmediato ver que $c_2(\mathbf{A}) \geq 1$.

Definición 76 (error relativo del vector solución) *El cociente* $\frac{\|\Delta \mathbf{x}_0\|_2}{\|\mathbf{x}_0\|_2}$*, que aparece en (7.21), es el error relativo, medido por la norma euclidea, del vector solución del sistema cuando el vector* $\mathbf{b}$ *se modifica.*

El resultado (7.21) queda, por tanto, como

$$\frac{\|\Delta \mathbf{x}_0\|_2}{\|\mathbf{x}_0\|_2} \leq c_2(\mathbf{A}) \frac{\|\Delta \mathbf{b}\|_2}{\|\mathbf{b}\|_2}$$

Por tanto, el número de condición proporciona una cota superior para el posible aumento del error relativo, cuando se modifica el término independiente del sistema. Esta cota superior es óptima (es decir, alcanzable).

Pueden obtenerse expresiones que relacionan $\Delta \mathbf{x}_0$, $\mathbf{x}_0$ y $c_2(\mathbf{A})$, cuando la matriz $\mathbf{A}$ pasa a ser $\mathbf{A} + \Delta \mathbf{A}$ o cuando se modifican conjuntamente la matriz $\mathbf{A}$ y el vector $\mathbf{b}$.

7.3. Normas vectoriales y matriciales

En esta sección se generaliza el concepto de norma de un vector y se extiende a matrices de cualquier tipo. De esta manera se puede cuantificar un error definido como diferencia entre dos matrices. Las normas de vectores y matrices sirven para estudiar cómo se propagan a la solución de los sistemas de ecuaciones lineales los pequeños errores o perturbaciones que se introduzcan en los datos del problema: la matriz el término independiente. Se define también el número de condición de una matriz, que es un número real mayor o igual que 1 que cuantifica la sensibilidad que esta matriz tiene ante dichos errores. Los valores singulares de una matriz ofrecen la mejor caracterización de dicha sensibilidad. ¿Por qué definir distintos tipos de normas? En espacios de dimensión finita todas las normas son equivalentes lo que significa que si dos vectores o dos matrices son suficientemente parecidos según una norma, también lo serán según cualquier otra norma. La pluralidad de normas se justifica entonces por la mayor o menor dificultad con la que permiten establecer ciertos resultados teóricos y por el muy distinto esfuerzo de cálculo que requieren.

7.3.1. Normas vectoriales

En todo lo que sigue $E = \mathbb{K}^{m\times n}$ o $E = \mathbb{K}^n$, donde $\mathbb{K} = \mathbb{R}$ o $\mathbb{K} = \mathbb{R}$.

Definición 77 *Una norma en E es una aplicación*

$$\|\cdot\| \;:\; E \;\longrightarrow\; \mathbb{R} \tag{7.23}$$

que verifique las siguientes propiedades:

1)

$$\forall \mathbf{z} \in E,\ \|\mathbf{z}\| \geq 0,\ \text{ y } \qquad \|\mathbf{z}\| = 0 \Leftrightarrow \mathbf{z} = \mathbf{0}. \tag{7.24}$$

2)

$$\forall \mathbf{z} \in E,\ \forall \lambda \in \mathbb{K},\ \|\lambda \mathbf{z}\| = |\lambda|\|\mathbf{z}\| \tag{7.25}$$

3)

$$\forall \mathbf{z}, \mathbf{w} \in E,\ \|\mathbf{z}+\mathbf{w}\| \leq \|\mathbf{z}\| + \|\mathbf{w}\| \qquad \textit{(Desigualdad triangular).} \tag{7.26}$$

Ejemplo 1. Tres ejemplos clásicos de norma en $E = \mathbb{C}^n$ son:

– **norma 1:**

$$\|\mathbf{z}\|_1 = |z_1| + \cdots + |z_n| \tag{7.27}$$

– **norma 2** o norma euclídea (norma inducida por el producto escalar definido en el capítulo 3, a la que ya se ha hecho referencia al considerar las expresiones de los valores singulares máximo y mínimo):

$$\|\mathbf{z}\|_2 = \sqrt{|z_1|^2 + \cdots + |z_n|^2} \tag{7.28}$$

– **norma** ∞ o norma del supremo:

$$\|\mathbf{z}\|_\infty = \text{máx}\{|z_1|, \ldots, |z_n|\} \tag{7.29}$$

Se puede comprobar con relativa facilidad que cada una de las expresiones anteriores satisface las tres condiciones de la definición de norma.

Observación: Toda norma $\|\cdot\|$ en E induce una distancia $d(\mathbf{z}, \mathbf{w}) := \|\mathbf{z} - \mathbf{w}\|$ en E. En particular, la norma euclídea en $\mathbb{R}^n$ induce la distancia habitual.

Definición 78 (normas equivalentes) *Se dice que dos normas $\|\cdot\|_i$, $\|\cdot\|_j$ en E son* **equivalentes** *si existen unas constantes reales $\alpha, \beta > 0$ tales que*

$$\forall \mathbf{z} \in E, \qquad \alpha\|\mathbf{z}\|_i \leq \|\mathbf{z}\|_j \leq \beta\|\mathbf{z}\|_i \tag{7.30}$$

Proposición 38 *En los espacios vectoriales de dimensión finita todas las normas son equivalentes.*

Ejemplo 2. La **norma de Frobenius** viene definida por

$$\|\mathbf{A}\|_F = \sqrt{\text{tr}(\mathbf{A}^h\mathbf{A})} = \sqrt{\sum_{i,j} |a_{ij}|^2} = \sqrt{\sum_{j=1}^{r} \sigma_j^2} \tag{7.31}$$

donde los σ_j son los valores singulares no nulos de $\mathbf{A}$.

7.3.2. Normas matriciales inducidas por normas vectoriales

Definición 79 *Sean $\mathbf{A} \in \mathbb{C}^{m\times n}$ y $\|\cdot\|_p$, $p \in \{1, 2, \infty\}$, una norma vectorial definida tanto en $\mathbb{C}^m$ como en $\mathbb{C}^n$. Se denomina **norma matricial** $\|\cdot\|_p$ en $\mathbb{C}^{m\times n}$ **inducida por** dicha **norma vectorial** a*

$$\|\mathbf{A}\|_p := \max_{\mathbf{x}\neq\mathbf{0}} \frac{\|\mathbf{A}\mathbf{x}\|_p}{\|\mathbf{x}\|_p} \tag{7.32}$$

Se puede demostrar de un modo sencillo que la definición anterior verifica las propiedades de norma. En lo sucesivo con frecuencia se omitirá el subíndice p y se entenderá que la norma es una de las normas vectoriales $p = \{1, 2, \infty\}$, y la correspondiente norma matricial inducida.

Como consecuencia de la definición,

$$\forall \mathbf{x} \in \mathbb{C}^n, \qquad \|\mathbf{A}\mathbf{x}\| \leq \|\mathbf{A}\|\|\mathbf{x}\| \tag{7.33}$$

Se verifica

$$\|\mathbf{A}\| = \max_{\mathbf{x}\neq\mathbf{0}} \frac{\|\mathbf{A}\mathbf{x}\|}{\|\mathbf{x}\|} = \max_{\mathbf{x}\neq\mathbf{0}} \left\|\mathbf{A}\frac{\mathbf{x}}{\|\mathbf{x}\|}\right\| = \max_{\|\mathbf{x}\|=1} \|\mathbf{A}\mathbf{x}\| \tag{7.34}$$

Si $\mathbf{A}$ es regular se cumple

$$\|\mathbf{A}^{-1}\| = \max_{\mathbf{x}\neq\mathbf{0}} \frac{\|\mathbf{A}^{-1}\mathbf{x}\|}{\|\mathbf{x}\|} = \left(\min_{\|\mathbf{x}\|=1} \|\mathbf{A}\mathbf{x}\|\right)^{-1} \tag{7.35}$$

Proposición 39 *Si $\|\cdot\|$ es una norma matricial, inducida por una norma vectorial,*

$$\forall\mathbf{A}\in\mathbb{C}^{m\times n}, \forall\mathbf{B}\in\mathbb{C}^{n\times p}, \qquad \|\mathbf{AB}\| \leq \|\mathbf{A}\|\|\mathbf{B}\| \tag{7.36}$$

Demostración: *Si $\mathbf{y}$ es un vector unitario tal que $\|\mathbf{AB}\| = \|(\mathbf{AB})\,\mathbf{y}\|$, se tiene:*

$$\|\mathbf{AB}\| = \|(\mathbf{AB})\,\mathbf{y}\| = \|\mathbf{A}\,(\mathbf{By})\| \leq \|\mathbf{A}\|\,\|\mathbf{By}\| \leq \|\mathbf{A}\|\,\|\mathbf{B}\| \tag{7.37}$$

Ejemplos de normas matriciales inducidas

Ejemplos relevantes de normas matriciales inducidas por normas vectoriales son los siguientes:

- **Norma 1:** Utilizando la norma $\|\cdot\|_1$ en $\mathbb{C}^m$ y $\mathbb{C}^n$,

$$\|\mathbf{A}\|_1 = \max_{\|\mathbf{u}\|_1=1} \|\mathbf{Au}\|_1 = \max_k \|\mathbf{A}_k\|_1 \tag{7.38}$$

  es decir, coincide con el máximo de las normas 1 de los vectores columna $\mathbf{A}_k$ de $\mathbf{A}$.

- **Norma 2 o norma espectral:** Utilizando la norma $\|\cdot\|_2$ en $\mathbb{C}^m$ y $\mathbb{C}^n$,

$$\|\mathbf{A}\|_2 = \max_{\|\mathbf{u}\|_2=1} \|\mathbf{Au}\|_2 = \sqrt{\rho(\mathbf{A}^h\mathbf{A})} = \sigma_{\text{máx}} \tag{7.39}$$

  donde $\rho(\mathbf{A}^h\mathbf{A})$ es el radio espectral de dicha matriz y $\sigma_{\text{máx}}$ es el valor singular máximo de $\mathbf{A}$.

- **Norma ∞:** Utilizando la norma $\|\cdot\|_\infty$ en $\mathbb{C}^m$ y $\mathbb{C}^n$,

$$\|\mathbf{A}\|_\infty = \max_{\|\mathbf{u}\|_\infty=1} \|\mathbf{Au}\|_\infty = \max_j \|\mathbf{A}^j\|_1 \tag{7.40}$$

  es decir, coincide con el máximo de las normas 1 de los vectores fila $\mathbf{A}^j$ de $\mathbf{A}$.

Se satisfacen las relaciones,

$$\|\mathbf{A}\|_\infty = \|\mathbf{A}^t\|_1, \qquad \|\mathbf{A}\|_2 = \|\mathbf{A}^t\|_2 = \|\mathbf{A}^h\|_2 \tag{7.41}$$

y, para cualquiera de las normas 1, 2, ∞ se cumple

$$\|\mathbf{A}\| = \|\overline{\mathbf{A}}\| \tag{7.42}$$

Observación: La norma de Frobenius no está inducida por ninguna norma vectorial, como se comprueba considerando la matriz identidad de orden n, ya que $\|\mathbf{I}_n\|_F = \sqrt{n}$.

Proposición 40 *Cualquier norma matricial $\|\cdot\|$ en $\mathbb{C}^{n\times n}$ inducida por una norma vectorial verifica*

$$\rho(\mathbf{A}) \leq \|\mathbf{A}\| \tag{7.43}$$

Demostración: Si $\mathbf{x}$ es un vector propio de $\mathbf{A}$ y $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$, se tiene $\|\lambda\mathbf{x}\| = |\lambda|\,\|\mathbf{x}\| = \|\mathbf{A}\mathbf{x}\| \leq \|\mathbf{A}\|\,\|\mathbf{x}\|$, de donde $|\lambda| \leq \|\mathbf{A}\|$, lo cual implica que el radio espectral de $\mathbf{A}$ está acotado superiormente por la norma de $\mathbf{A}$. $\square$

7.4. Ejercicios

7.1.– Determínese una descomposición en valores singulares de las siguientes matrices:

$$\mathbf{A} = \begin{pmatrix} 8 & 4 & -8 \\ 6 & 3 & -6 \end{pmatrix}, \ \mathbf{B} = \begin{pmatrix} 7 & 4 \\ 4 & -8 \\ 4 & 1 \end{pmatrix}.$$

7.2.– Hállese una descomposición en valores singulares (DVS) de la matriz

$$\mathbf{A} = \begin{pmatrix} 1 & i \\ i & -1 \end{pmatrix}.$$

7.3.– Sea $\mathbf{A} = \begin{pmatrix} -1 & 2 & 0 \\ 2 & 0 & 2 \\ 0 & 2 & 1 \end{pmatrix}$.

1. Calcúlese una matriz $\mathbf{Q}$ ortogonal tal que $\mathbf{Q}^t\mathbf{AQ}$ sea diagonal.
2. Dedúzcase una descomposición en valores singulares $\mathbf{U\Sigma V}^t$ de $\mathbf{A}$, expresando las matrices $\mathbf{U}$ y $\mathbf{V}$ en función de $\mathbf{Q}$.

7.4.– Sea $\mathbf{A}$ una matriz compleja. Demuéstrense las siguientes propiedades:

1. $\mathbf{A}, \overline{\mathbf{A}}, \mathbf{A}^t, \mathbf{A}^h$ tienen los mismos valores singulares.
2. Si las filas (o columnas) de $\mathbf{A}$ son ortonormales, entonces los valores singulares de dicha matriz son todos iguales a la unidad. En particular, los valores singulares de una matriz unitaria son 1.
3. Si $\mathbf{A}$ es cuadrada entonces el producto de sus valores singulares coincide con el módulo del determinante de la matriz.

7.5.– Hállese una descomposición en valores singulares (DVS) de la matriz

$$\mathbf{M} = \begin{pmatrix} 5 & 3 & 5 & 3 \\ 4 & 0 & 4 & 0 \\ 2 & 6 & 2 & 6 \end{pmatrix}$$

así como su pseudoinversa. **Indicación:** Se recomienda diagonalizar (ortogonalmente) en primer lugar $\mathbf{M}^t\mathbf{M}$.

7.6.– Sea $\mathbf{A} \in \mathbb{C}^{m\times n}$ y $\mathbf{A}^+$ su matriz pseudoinversa:

1. Si $m = n$ y $\mathbf{A}$ es invertible, entonces $\mathbf{A}^+ = \mathbf{A}^{-1}$.
2. Si el rango de $\mathbf{A}$ es igual a n, entonces $\mathbf{A}^+ = \left(\mathbf{A}^h\mathbf{A}\right)^{-1}\mathbf{A}^h$.

3. Si el rango de $\mathbf{A}$ es igual a m, entonces $\mathbf{A}^+ = \mathbf{A}^h \left(\mathbf{A}\mathbf{A}^h\right)^{-1}$.

7.7.– Dada la matriz

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

1. Calcúlense sus valores singulares.

2. Hállese una descomposición en valores singulares $\mathbf{U\Sigma V}^t$ de $\mathbf{A}$, de modo que la primera columna de $\mathbf{V}$ sea un vector canónico.

3. Calcúlese la pseudoinversa de $\mathbf{A}$ como suma de r matrices de rango 1, siendo r el rango de la matriz $\mathbf{A}$.

4. Calcúlese la solución de mínimos cuadrados y mínima norma del sistema $\mathbf{Ax} = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 1 \end{pmatrix}$.

7.8.– Determínese el número de condición espectral de la matriz $\begin{pmatrix} 5 & 10 \\ -11 & -2 \end{pmatrix}$.

7.9.– Hállense las normas 1, 2 e ∞ de los siguientes vectores:

$$\mathbf{u} = (1, -2, 0, 2)\,, \qquad \mathbf{v} = (1, -i, 3+4i, -2+i)\,.$$

7.10.– Represéntese los conjuntos del plano $\mathbb{R}^2$ y del espacio $\mathbb{R}^3$ determinados por las siguientes condiciones:

$$\|\mathbf{x}\|_1 \leq 1, \qquad \|\mathbf{x}\|_2 \leq 1, \qquad \|\mathbf{x}\|_\infty \leq 1$$

7.11.– Hállese la norma 2 de cada una de las siguientes matrices:

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & 3 \\ -2 & 2 & 0 \end{pmatrix}, \ \mathbf{B} = \begin{pmatrix} 1 & -2 \\ 2 & -1 \end{pmatrix}, \ \mathbf{C} = \begin{pmatrix} -5 & 4 \\ 4 & -5 \end{pmatrix}, \ \mathbf{D} = \begin{pmatrix} 1 \\ 2 \\ 5 \end{pmatrix}.$$

7.12.– Estúdiese la relación entre el radio y la norma espectrales de la matriz $\begin{pmatrix} 2 & 6 \\ 6 & -3 \end{pmatrix}$. Compruébese que el resultado se extiende a cualquier matriz normal.

7.4.1. Cuestiones

7.13.– Se considera la matriz

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

Indíquese la única proposición falsa:

1. $\mathbf{A}$ sólo tiene dos valores singulares distintos
2. La norma espectral de $\mathbf{A}$ vale 28
3. $\min_{\mathbf{x}\neq\mathbf{0}} \frac{\|\mathbf{Ax}\|_2}{\|\mathbf{x}\|_2} = \sqrt{14}$
4. El número de condición espectral de $\mathbf{A}$ vale $\sqrt{2}$

7.14.– Se considera la matriz

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

Indíquese la única proposición falsa:

1. $\|\mathbf{A}\|_1 = 6$
2. El número de condición espectral de $\mathbf{A}$ vale $\sqrt{5}$
3. $\forall \mathbf{x} \in \mathbb{R}^3 \setminus \{\mathbf{0}\}, \ \frac{\|\mathbf{Ax}\|_2}{\|\mathbf{x}\|_2} \leq \sqrt{5}$
4. $\|\mathbf{A}\|_\infty = 3$

7.15.– Se considera la matriz

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

Indíquense las proposiciones verdaderas y las falsas:

1. Para todo vector $\mathbf{u}$ unitario de $\mathbb{R}^3$, $\|\mathbf{Au}\|_2 \geq \sqrt{2}$
2. $\|\mathbf{A}^t\|_2 = \|\mathbf{A}\|_2$

3. Todos los autovalores de $\mathbf{A}\mathbf{A}^t$ son reales y no negativos

4. El número de condición espectral de $\mathbf{A}$ vale 5

5. $\|\mathbf{A}\|_2 = \sqrt{10}$

7.16.– Si $\mathbf{M}^+$ es la matriz pseudoinversa de $\mathbf{M}$, indíquense las proposiciones verdaderas y las falsas:

1. Si $\mathbf{P}$ es una matriz de proyección ortogonal, entonces $\mathbf{P}^+ = \mathbf{P}$

2. Si $\mathbf{S}$ es una matriz de simetría ortogonal, entonces $\mathbf{S}^+ = \mathbf{S}$

3. Si $\mathbf{A} \in \mathbb{C}^{m\times n}$ de rango igual a n, entonces $\mathbf{A}^+ = \left(\mathbf{A}^h\mathbf{A}\right)^{-1}\mathbf{A}^h$

4. Si $\mathbf{A} \in \mathbb{C}^{m\times n}$ de rango igual a m, entonces $\mathbf{A}^+ = \mathbf{A}^h\left(\mathbf{A}\mathbf{A}^h\right)^{-1}$

5. Si $\mathbf{A} \in \mathbb{C}^{n\times n}$ tal que $\mathbf{A}^+ = \mathbf{A}$, entonces $\mathbf{A}$ es invertible.

Apéndice 1: Nociones de teoría de conjuntos

1.1 Definiciones y propiedades. Pertenencia e inclusión. Los cuantificadores. Lógica formal.

1.2 Unión e intersección de conjuntos. Conjuntos complementarios. Producto cartesiano.

1.3 Aplicaciones. Composición. Inversa.

1.1. Conjuntos y lógica formal

1.1.1. Definiciones básicas

- Un **conjunto** es una colección de objetos, que se denominan **elementos** del conjunto.
- Los elementos de un conjunto siempre son distintos dos a dos, y no se presupone ningún orden entre ellos. Por ejemplo los conjuntos $\{a, a, b\}$ y $\{b, a\}$ son iguales.
- Se llama **conjunto vacío** y se denota por $\varnothing$ el conjunto que no tiene ningún elemento.
- Un conjunto se dice **finito** si contiene un número finito de elementos, que en ese caso podrán numerarse y, por tanto, todo conjunto E finito y no vacío puede expresarse como $E = \{a_1, a_2, \cdots, a_n\}$, siendo n (entero positivo) el número de sus elementos.
- En el caso anterior, diremos que E tiene **cardinal** n, y lo denotaremos por $\text{Card }(E) = n$.
- Un conjunto se dice **infinito** si tiene infinitos elementos.

1.1.2. Relaciones de pertenencia e inclusión

- **Relación de pertenencia:** Si E es un conjunto y a es un elemento del mismo, se denota por $a \in E$ ("a pertenece a E"). En caso contrario, escribimos $a \notin E$.

- **Relación de inclusión:** Sean E y F dos conjuntos, se dice que E está contenido en F (y se denota por $E \subset F$) si todo elemento de E pertenece a F. También se puede escribir $F \supset E$ ("F contiene a E").
- Si $E \subset F$, también se dice que E es un **subconjunto** de F.
- Cualquiera que sea el conjunto E, se tiene que $\varnothing \subset E$.
- **Igualdad de conjuntos.** Dos conjuntos E y F son iguales ($E = F$) si y sólo si $E \subset F$ y $F \subset E$.
- Así pues, $E \neq F$ significa que existe algún elemento de E que no pertenece a F o algún elemento de F que no pertenece a E.

1.1.3. Los cuantificadores

- **Cuantificador universal:** Se denota por el símbolo $\forall$ (*para todo*). Sea $\mathcal{P}$ una proposición que depende de los elementos a del conjunto E. Si la proposición es verdadera para todos los elementos del conjunto, escribiremos

$$\forall a \in E \quad , \quad \mathcal{P}(a).$$

  Por ejemplo, $\forall a \in \mathbb{Z},\ \ a \in \mathbb{Q}$; es decir, todo número entero es racional.
- **Cuantificador existencial:** Se denota por el símbolo $\exists$ (*existe*). Sea $\mathcal{P}$ una proposición que depende de los elementos a del conjunto E. Si la proposición es verdadera para algún elemento del conjunto, escribiremos

$$\exists a \in E \quad , \quad \mathcal{P}(a).$$

  Por ejemplo, $\exists a \in \mathbb{Q},\ \ a \in \mathbb{N}$; es decir, existe un número racional que es natural.
- Cuando existe un único elemento puede escribirse $\exists! a \in E$.
- Denotaremos por $\neg\mathcal{P}$ la **negación** de la proposición $\mathcal{P}$ (es decir, la proposición que se cumple únicamente cuando no se cumple $\mathcal{P}$, y viceversa.
- Negación de proposiciones con cuantificadores: se comprueba fácilmente que:
  - $\neg(\forall a \in E,\ \ \mathcal{P}(a))$ si y sólo si $\exists a \in E,\ \ \neg\mathcal{P}(a)$
  - $\neg(\exists a \in E,\ \ \mathcal{P}(a))$ si y sólo si $\forall a \in E,\ \ \neg\mathcal{P}(a)$
- Como ejemplo de la utilización de los cuantificadores anteriores, puede expresarse la relación de inclusión de la forma siguiente:

$$E \subset F \text{ si y sólo si } \forall a \in E,\ a \in F$$

La negación de dicha inclusión se expresa:

$$\neg(E \subset F) \text{ si y sólo si } \exists a \in E,\ a \notin F$$

1.1.4. Inclusión y lógica formal

- Si una proposición $\mathcal{P}_2$ se cumple siempre que se cumpla otra cierta proposición $\mathcal{P}_1$, escribiremos

$$\mathcal{P}_1 \Rightarrow \mathcal{P}_2$$

  y diremos que "$\mathcal{P}_1$ implica $\mathcal{P}_2$", o bien que $\mathcal{P}_1$ es una **condición suficiente** para $\mathcal{P}_2$, o bien que $\mathcal{P}_2$ es una **condición necesaria** para $\mathcal{P}_1$. También diremos "si $\mathcal{P}_1$ entonces $\mathcal{P}_2$".

- Si $\mathcal{P}_1 \Rightarrow \mathcal{P}_2$ y $\mathcal{P}_2 \Rightarrow \mathcal{P}_1$, escribiremos

$$\mathcal{P}_1 \Leftrightarrow \mathcal{P}_2$$

  (léase "$\mathcal{P}_1$ si y sólo si $\mathcal{P}_2$") y diremos que $\mathcal{P}_1$ y $\mathcal{P}_2$ son proposiciones **equivalentes**.

- La proposición $\mathcal{P}_2 \Rightarrow \mathcal{P}_1$ se llama la **recíproca** de $\mathcal{P}_1 \Rightarrow \mathcal{P}_2$

- Si E es el conjunto de los elementos que satisfacen una proposición $\mathcal{P}_1$, y F es el conjunto de los elementos que satisfacen otra proposición $\mathcal{P}_2$, entonces

$$(\mathcal{P}_1 \Rightarrow \mathcal{P}_2) \Leftrightarrow E \subset F.$$

1.2. Operaciones con conjuntos

1.2.1. Unión e intersección de conjuntos

- Dados dos conjuntos E y F, se define su **unión** como $E \cup F$.

$$E \cup F := \{a : a \in E \ \vee \ a \in F\}$$

  El símbolo $\vee$ equivale a la disyunción "o".

  Esta definición se extiende naturalmente a la unión de una familia finita de conjuntos $\{A_1, A_2, \ldots, A_n\}$:

$$A_1 \cup A_2 \cup \ldots \cup A_n = \{a : \exists\ j,\ a \in A_j\}.$$

- Obsérvese que, para cualquier conjunto A, se tiene que $A \cup \varnothing = A$.

- Se define la intersección de dos conjuntos E y F como

$$E \cap F := \{a : a \in E \ \wedge \ a \in F\}.$$

El símbolo $\wedge$ equivale a la conjunción "y".

Asimismo,

$$A_1 \cap A_2 \cap \ldots \cap A_n$$

denota la intersección de una familia finita $\{A_1, A_2, \ldots, A_n\}$.

- Obsérvese que, para cualquier conjunto A, se tiene que $A \cap \varnothing = \varnothing$.

- Dos conjuntos A y B se dicen **disjuntos** (entre sí) si $A \cap B = \varnothing$.

Propiedades: (En todo lo que sigue, A, B, C son conjuntos).

  - $A \cap B \subset A \subset A \cup B$. (Igualmente para B). También es válida para uniones e intersecciones infinitas.
  - $A \subset B \Leftrightarrow A \cup B = B \Leftrightarrow A \cap B = A$.
  - $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ (propiedad distributiva de la intersección respecto a la unión).
  - $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ (propiedad distributiva de la unión respecto a la intersección).

1.2.2. Conjuntos complementarios

- Dados E y F conjuntos, se define la **diferencia** de E y F como

$$E \setminus F := \{a \in E : a \notin F\}.$$

- Dados A y E conjuntos, con $A \subset E$ se define el **complementario** de A respecto a E, y se denota A^c o bien A' o también $c(A)$, como

$$A^c := E \setminus A = \{a \in E : a \notin A\}.$$

Propiedades: En todo lo que sigue, suponemos $A, B \subset E$.

  - $(A^c)^c = A$.
  - $A \cup A^c = E$.
  - $A \cap A^c = \varnothing$.
  - $A \subset B \Leftrightarrow B^c \subset A^c$ $(*)$; veremos después la importante traducción de esta propiedad a la lógica formal.
  - $A \setminus B = A \cap B^c$.

- $(A \cup B)^c = A^c \cap B^c$.
- $(A \cap B)^c = A^c \cup B^c$.

  Estas dos últimas propiedades se conocen como *Leyes de Morgan*.

- Si A, subconjunto de E, está constituido por los elementos de E que satisfacen una determinada propiedad $\mathcal{P}$, entonces A^c es el conjunto de los elementos de E para los cuales $\mathcal{P}$ es falsa, es decir,

$$A^c = \{a \in E : \neg\mathcal{P}(a)\}.$$

- La propiedad $(*)$ antes vista, y traduciendo los conjuntos A, B a proposiciones lógicas $\mathcal{P}_1, \mathcal{P}_2$, se convierte en

$$(\mathcal{P}_1 \Rightarrow \mathcal{P}_2) \Leftrightarrow (\neg\mathcal{P}_2 \Rightarrow \neg\mathcal{P}_1).$$

  La proposición que se encuentra a la derecha de la doble implicación se denomina proposición **contrarrecíproca** de la de la izquierda. Ambas son equivalentes y, a veces, es más fácil demostrar la proposición contrarrecíproca de una dada que la proposición original. Este *truco* se emplea a menudo en razonamientos matemáticos, y es la herramienta conocida como **paso al contrarrecíproco**.

- No debe confundirse lo anterior con el método de razonamiento conocido como **reducción al absurdo** (o por contradicción); en este caso, para demostrar $(\mathcal{P}_1 \Rightarrow \mathcal{P}_2)$ se suponen simultáneamente ciertas ($\mathcal{P}_1$ y $\neg\mathcal{P}_2$), y se llega a una contradicción.

1.2.3. Producto cartesiano

- Dados dos conjuntos E y F, se define su **producto cartesiano** $E \times F$ como

$$E \times F := \{(x, y) : x \in E,\ y \in F\}.$$

  Los elementos (x, y) de $E \times F$ se llaman **pares ordenados**.

- Recursivamente, puede extenderse la definición anterior a cualquier número finito de conjuntos: dados $E_1, E_2, \ldots, E_n$, se define

$$E_1 \times E_2 \times \cdots \times E_n := \{(x_1, x_2, \ldots, x_n) : x_j \in E_j,\ j = 1, 2, \ldots, n\}.$$

  Los elementos $(x_1, x_2, \ldots, x_n)$ se denominan n-**uplas**.

- Dos n-uplas formadas por los mismos elementos son distintas si aquéllos aparecen en cada una en un orden diferente: De otra forma, dos n-uplas son iguales si:

$$(x_1, x_2, \ldots, x_n) = (y_1, y_2, \ldots, y_n) \Leftrightarrow \forall i = 1, 2, \ldots, n,\ \ x_i = y_i$$

- Cuando $E_1 = E_2 = \ldots = E_n = E$, se denota $E_1 \times E_2 \times \ldots \times E_n$ por E^n.

1.3. Aplicaciones entre conjuntos

1.3.1. Definiciones

- Sean E y F conjuntos no vacíos. Una **aplicación** f de E en F es un subconjunto de $E \times F$ tal que:

$$\forall x \in E, \text{ existe un único } y \in F \text{ tal que } (x,y) \in f.$$

  Se denota por

$$f : E \to F.$$

- Cuando $(x,y) \in f$, con $x \in E,\ y \in F$, se denota por

$$f(x) = y$$

  y se dice que y es la **imagen** por f de x. También se dice que x es una **contraimagen** de y mediante f.

- Obsérvese que, a la vista de la notación anterior, una aplicación de E a F no es más que una *regla de asignación* que, a cada elemento de E, le asigna un **único** elemento de F.

- E se llama **dominio** o **conjunto inicial** de f; F se denomina **conjunto final** de f. Obsérvese que cada elemento del conjunto inicial tiene una única imagen.

- Dados una aplicación $f : E \to F$, y un subconjunto $A \subset E$, se define la **imagen** de A por f como

$$f(A) := \{f(x) : x \in A\},$$

  que es un subconjunto de F.

  Dado un subconjunto $B \subset F$, se define la **contraimagen** o imagen recíproca de B mediante f como

$$f^{-1}(B) = \{x \in E : f(x) \in B\},$$

  que es un subconjunto de A.

  Nota: La notación $f^{-1}(B)$ **no** supone la existencia de la aplicación inversa de f, que se definirá más adelante; $f^{-1}(B)$ existe cualesquiera que sean la aplicación f y el subconjunto B del conjunto final.

- Dadas dos aplicaciones f y g, definidas entre los mismos conjuntos E y F, decimos que $f = g$, si y sólo si

$$\forall x \in E,\ f(x) = g(x).$$

- Una aplicación $f : E \to F$ se dice **inyectiva** si

$$\forall x, y \in E, \text{ si } f(x) = f(y), \text{ entonces } x = y,$$

  es decir, que cada elemento del conjunto final tiene, como máximo, una única contraimagen.

- Una aplicación $f : E \to F$ se dice **suprayectiva** si $f(E) = F$ o, lo que es lo mismo,

$$\forall y \in F, \ \exists x \in E \text{ tal que } f(x) = y,$$

  es decir, si cada elemento del conjunto final tiene, como mínimo, una contraimagen.

- Una aplicación se dice **biyectiva** si es a la vez inyectiva y suprayectiva, es decir, si cada elemento del conjunto final tiene una única contraimagen.

- Una aplicación biyectiva de E a F también se denomina una **biyección** entre E y F.

- Obsérvese que dos conjuntos **finitos** tienen el mismo número de elementos si y sólo si existe una aplicación biyectiva entre ellos.

- Dado un conjunto E, se define la aplicación identidad de E, y se denota por I_E, como aquella

$$I_E : E \to E$$

  tal que, $\forall x \in E, \ I_E(x) = x$, que, obviamente, es una aplicación biyectiva.

1.3.2. Composición de aplicaciones

- Dados tres conjuntos E, F, G y dos aplicaciones

$$f : E \to F \text{ y } g : F \to G,$$

  se define la **composición** de f y g como la aplicación $g \circ f$ (léase "f compuesta con g") como

$$g \circ f : E \to G$$

  tal que, $\forall x \in E$,

$$g \circ f(x) = g(f(x)).$$

  Es decir, el resultado de aplicar sucesivamente f y g. Obsérvese que la imagen de E por f ha de estar contenida en el conjunto inicial de g.

- **Propiedades de la composición:**

- Dadas tres aplicaciones $f : E \to F,\ g : F \to G$ y $h : G \to H$, se tiene que

$$(h \circ g) \circ f = h \circ (g \circ f),$$

  es decir, la composición de aplicaciones es **asociativa**.
- Para toda aplicación $f : E \to F$, se tiene que $f \circ I_E = f$ y $I_F \circ f = f$.
- La composición de dos aplicaciones inyectivas es inyectiva.
- La composición de dos aplicaciones suprayectivas es suprayectiva.
- Como consecuencia de las dos anteriores, la composición de dos aplicaciones biyectivas es biyectiva.

- Sean $f : E \to F$ y $g : F \to G$ dos aplicaciones; se cumple:
  - $g \circ f$ inyectiva $\Rightarrow f$ inyectiva.
  - $g \circ f$ suprayectiva $\Rightarrow g$ suprayectiva.
  - $g \circ f$ biyectiva $\Rightarrow f$ inyectiva y g suprayectiva.

1.3.3. La aplicación inversa

- **Construcción de la aplicación inversa:** Si $f : E \to F$ es biyectiva, quiere decir que:

$$\forall y \in F, \exists! x \in E \text{ tal que } f(x) = y.$$

  Así pues, podemos definir una nueva aplicación $f^{-1} : F \to E$ de la siguiente manera:

$$\forall y \in F,\ f^{-1}(y) = x \text{ tal que } f(x) = y.$$

  Es obvio que, $\forall x \in E,\ f^{-1}(f(x)) = f^{-1}(y) = x$, y que $\forall y \in F,\ f(f^{-1}(y)) = f(x) = y$.

- Dada una aplicación **biyectiva** $f : E \to F$, existe una única aplicación $g : F \to E$ tal que $g \circ f = I_E$ y $f \circ g = I_F$; esta aplicación se llama la inversa de f, y se denota por f^{-1}.

- Por otro lado, es obvio que, dada una aplicación cualquiera f, si existe una aplicación f^{-1} que satisfaga la condición anterior, entonces f es biyectiva.

- Para una aplicación $f : E \to E$, y para cada número natural n, se usa la notación:

$$f^0 = I_E;$$
$$f^n = f \circ \overset{n}{\ldots} \circ f \text{ si } n \geq 1.$$

  Si además f es biyectiva, se denota $f^{-n} = f^{-1} \circ \overset{n}{\ldots} \circ f^{-1}$.

- **Propiedades:**

  Dadas dos aplicaciones **biyectivas** $f : E \to F$ y $g : F \to G$, se cumple que:

  - $(f^{-1})^{-1} = f$.
  - $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$.

Apéndice 2: Los números complejos

2.1 Definición axiomática. Operaciones con números complejos.

2.2 La conjugación y el módulo: propiedades. Representación trigonométrica. Módulo del producto de dos complejos. La desigualdad triangular. Argumento del producto de dos complejos.

2.3 Potencias n–ésimas. Fórmula de De Moivre. Raíces n–ésimas.

2.1. Definición y propiedades

Se pretende construir un conjunto que contenga naturalmente a los números reales, cuyas operaciones sean coherentes con las del cuerpo real y en el cual todo elemento admita raíz cuadrada. Para ello:

Definición 80 *El conjunto de los números complejos, que se denota por $\mathbb{C}$, es:*

$$\mathbb{C} := \{x + yi,\ \ x, y \in \mathbb{R},\ i^2 = -1\}$$

Observación: El *número i* se denomina *unidad imaginaria.* Un número complejo $z := x + yi$ queda determinado unívocamente por los dos números reales (x, y) que lo definen, así puede identificarse geométricamente el conjunto de los números complejos con el plano $\mathbb{R}^2$. Esta expresión del número complejo z se denomina *forma binómica.*

Definición 81 *Dado $z = a + ib \in \mathbb{C}$, con $a, b \in \mathbb{R}$, el número real a se llama **parte real** de z, mientras que el número real b se denomina **parte imaginaria** de z. Se denotan por*

$$Re\, z = a \ \ , \ \ \mathrm{Im}\, z = b.$$

Observación: Los números reales son los que tienen parte imaginaria nula; los números complejos con parte real nula se llaman **imaginarios puros**.

Los números complejos se suman agrupando sus partes reales e imaginarias y se multiplican operando formalmente y teniendo presente que $i^2 = -1$. Esto es:

Definición 82 *Dados dos números complejos $z := a + bi$, $w := c + di$ se define su* ***suma*** *como*

$$z + w := (a + c) + (b + d)i.$$

y su ***producto***

$$zw := (ac - bd) + (ad + bc)i.$$

Geométricamente, la suma de vectores se corresponde con la conocida *regla del paralelogramo.*

La suma y el producto *heredan* de los números reales las propiedades que se enuncian a continuación.

Propiedades:

1. Asociativas: $\forall z, w, u \in \mathbb{C},\ (z + w) + u = z + (w + u)$, $(zw)u = z(wu)$.

2. Para la suma existe un elemento neutro que es el 0 y para el producto un elemento unidad que es el 1.

3. Elemento opuesto (respecto de la suma): $\forall z = a + bi \in \mathbb{C},\ \exists (-z) = -a + (-b)i \in \mathbb{C}$ tal que $z + (-z) = 0$.

4. Elemento inverso (respecto del producto): $\forall z \in \mathbb{C} \setminus \{0\},\ \exists z^{-1} \in \mathbb{C}$ tal que $zz^{-1} = z^{-1}z = 1$.

5. Conmutativas: $\forall z, w \in \mathbb{C},\ z + w = w + z$, $zw = wz$.

6. Propiedad distributiva del producto respecto a la suma: $\forall z, w, u \in \mathbb{C}$, se tiene que $z(w + u) = zw + zu$.

Este conjunto de propiedades nos permite decir que la terna $(\mathbb{C}, +, \cdot)$ tiene estructura de **cuerpo** (conmutativo).

2.2. Conjugación, módulo y argumento

Definición 83 *Dado $z = a + ib \in \mathbb{C}$, con $a, b \in \mathbb{R}$, definimos su* ***conjugado****, y lo denotamos por $\overline{z}$, como el complejo $a - ib$.*

La interpretación geométrica de la conjugación es que los vectores correspondientes a complejos conjugados entre sí son simétricos respecto al eje de abscisas.

Propiedades de la conjugación:

- $\forall z \in \mathbb{C},\ \overline{\overline{z}} = z$.
- $\forall z, w \in \mathbb{C},\ \overline{z+w} = \overline{z} + \overline{w}$.
- $\forall z, w \in \mathbb{C},\ \overline{zw} = \overline{z}\,\overline{w}$.
- $\forall z \in \mathbb{C},\ \operatorname{Re} z = \dfrac{1}{2}(z+\overline{z})\ ;\ \operatorname{Im} z = \dfrac{-i}{2}(z-\overline{z}) = \dfrac{z-\overline{z}}{2i}$.

Definición 84 *Dado $z = x + iy \in \mathbb{C}, x, y \in \mathbb{R}$, definimos su **módulo** como*

$$|z| := \sqrt{z\overline{z}} = \sqrt{x^2+y^2}$$

Geométricamente, mide la longitud del vector que representa z en el plano. También se cumple, evidentemente, que $|z| \geq 0$ y $(|z| = 0 \Leftrightarrow z = 0)$.

Propiedades del módulo:

- $\forall z \in \mathbb{C},\ |\operatorname{Re} z|, |\operatorname{Im} z| \leq |z| \leq |\operatorname{Re} z| + |\operatorname{Im} z|$.
- $\forall z \in \mathbb{C},\ |z| = |\overline{z}|$.
- $\forall z, w \in \mathbb{C},\ |zw| = |z||w|$.
- **Desigualdad triangular:** $\forall z, w \in \mathbb{C},\ |z+w| \leq |z|+|w|$; además, si $w \neq 0$, entonces $|z+w| = |z| + |w| \Leftrightarrow z = \alpha w$ para algún $\alpha \in \mathbb{R}, \alpha \geq 0$.

Observación: Si $|z| = 1$, entonces $z\overline{z} = 1$, es decir: $z^{-1} = \overline{z}$.

2.2.1. Argumento de un complejo

Dados $x, y \in \mathbb{R}$ no simultáneamente nulos, se pueden encontrar valores de $\theta \in \mathbb{R}$ tales que

$$\cos\theta = \frac{x}{\sqrt{x^2+y^2}}\ ;\ \operatorname{sen}\theta = \frac{y}{\sqrt{x^2+y^2}}.$$

Por ejemplo, si $x > 0$, se puede tomar $\theta = \arctan\frac{y}{x}$; si $x = 0$, se puede tomar $\theta = \frac{\pi}{2}$ si $y > 0$ y $\theta = -\frac{\pi}{2}$ si $y < 0$; si $x < 0$, puede tomarse $\theta = \pi - \arctan\frac{y}{x}$.

Definición 85 *Dado $z \in \mathbb{C} \setminus \{0\}$, el **argumento** de z es el conjunto*

$$\mathrm{Arg\ (z)} := \{\theta \in \mathbb{R}\ /\ \mathrm{z} = |\mathrm{z}|(\cos\theta + \mathrm{i}\operatorname{sen}\theta)\}. \tag{2.44}$$

Habitualmente diremos *un valor del argumento* de z para indicar cualquier elemento del conjunto Arg (z). Por abuso de lenguaje se simplifica y se dice simplemente *un argumento* de z. En particular, se llama **argumento principal** de z al único de los valores anteriores que pertenece al intervalo $(-\pi, \pi]$. En cada intervalo de longitud 2π hay un único valor del argumento. El que está en el intervalo $[0, 2\pi)$ se llama **ángulo polar**. Geométricamente, el ángulo polar de z corresponde al ángulo, medido en sentido positivo (esto es, contrario a las agujas del reloj), que forma con el eje de abscisas el vector que representa a z.

La expresión $z = |z|(\cos\theta + i \operatorname{sen}\theta)$ se llama *forma trigonométrica* de z.

Observación: Cualesquiera dos valores del argumento de un mismo número complejo difieren en un múltiplo entero de 2π.

Si θ es un valor del argumento de un número complejo z, entonces

- $-\theta$ es un valor del argumento de $\overline{z}$.
- $\pi + \theta$ es un valor del argumento de $-z$.

2.3. Potencias. Fórmula de De Moivre

Proposición 41 *Dados dos complejos no nulos* $z = |z|(\cos\theta + i \operatorname{sen}\theta)$ *y* $|w|(\cos\varphi + i \operatorname{sen}\varphi)$*, se cumple que*

$$zw = |z||w| \left(\cos(\theta + \varphi) + i \operatorname{sen}(\theta + \varphi)\right).$$

Es decir, que ***el argumento de un producto es la suma de los argumentos****.*

El resultado anterior motiva la siguiente notación:

Identidad de Euler

$$\forall \theta \in \mathbb{R}, \quad e^{i\theta} := \cos\theta + i \operatorname{sen}\theta \tag{2.45}$$

Así pues, la expresión general de cualquier número complejo no nulo es $re^{i\theta}$, $\theta, r \in \mathbb{R}$ con $r > 0$.

Cuando un número complejo se escribe según la expresión anterior, se dice que está en *forma exponencial.*

De modo que, si tenemos dos números complejos de módulo unidad $e^{i\theta}$ y $e^{i\varphi}$, su producto vale $e^{i(\theta+\varphi)} = e^{i\theta}e^{i\varphi}$; además, la unidad, que es el número de módulo 1 y argumento 0, se puede expresar como $1 = e^{i0}$; es decir, que esta notación es coherente con las propiedades de la exponencial.

Nota: En el contexto presente, la identidad de Euler no es más que una notación, pero en realidad responde al concepto de exponencial de un número complejo que se estudia en matemáticas más avanzadas.

Observación: El producto de un número z por otro de módulo unidad $e^{i\theta}$, se interpreta geométricamente como el resultado de girar el vector que representa z un ángulo θ en sentido positivo (que es el contrario a las agujas del reloj).

2.3.1. Fórmula de De Moivre

Dados $\theta \in \mathbb{R}$ y $n \in \mathbb{Z}$, se tiene que

$$(\cos\theta + i \operatorname{sen}\theta)^n = \cos n\theta + i \operatorname{sen} n\theta. \tag{2.46}$$

(Donde para n negativo se entiende potencia $-n$ del inverso).

Así pues, la potenciación de números complejos (para potencias enteras) consiste en elevar el módulo a la correspondiente potencia y multiplicar el argumento por dicha potencia:

$$\left(re^{i\theta}\right)^n = r^n e^{in\theta}. \tag{2.47}$$

Apéndice 3: Polinomios

3.1 Polinomios con coeficientes en $\mathbb{R}$ o $\mathbb{C}$. Operaciones con polinomios. Grado de un polinomio.

3.2 División euclídea de polinomios. Teorema del resto. Máximo común divisor y mínimo común múltiplo de dos polinomios.

3.3 Raíces de un polinomio: simples y múltiples. El teorema fundamental del Álgebra.

3.1. Polinomios con coeficientes en un cuerpo

Definición 86 *Dado $(\mathbb{K}, +, \cdot)$ cuerpo, un polinomio con coeficientes en $\mathbb{K}$ es una expresión de la forma*

$$P(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n,$$

en donde $a_j \in \mathbb{K}$ y $n \in \mathbb{N}$. Los a_j son los coeficientes del polinomio P. Si $a_n \neq 0$ en la expresión anterior, se dice que P tiene grado n; a_n se dice el coeficiente principal de P, si $a_n = 1$ se dice que el polinomio es mónico. El coeficiente a_0 se denomina término independiente.

Nota: En la práctica, el cuerpo será siempre el real o el complejo.

Definición 87 *El conjunto de todos los polinomios con coeficientes en $\mathbb{K}$ se denota por $\mathbb{K}[x]$; el conjunto de los polinomios de $\mathbb{K}[x]$ cuyo grado es* ***menor o igual*** *que $n \in \mathbb{N}$ se denota por $\mathbb{K}_n[x]$.*

Observación: Es importante notar que lo que determina un polinomio son sus coeficientes a_j; la indeterminada x puede nombrarse mediante cualquier otra letra o símbolo. Cuando el cuerpo $\mathbb{K}$ es el de los complejos, a menudo se utiliza la letra z para la indeterminada.

3.1.1. Operaciones con polinomios

Definición 88 *Dados dos polinomios* $P(x) = \sum_{j=0}^{n} a_j x^j$ *y* $Q(x) = \sum_{j=0}^{m} b_j x^j$*, se define su* ***suma****:*

$$(P+Q)(x) := P(x) + Q(x) = \sum_{j=0}^{\max(m,n)} (a_j + b_j) x^j,$$

(en donde suponemos que $a_j = 0$ *si* $j > n$ *y* $b_j = 0$ *si* $j > m$*) y su* ***producto****:*

$$(PQ)(x) := P(x)Q(x) = \sum_{k=0}^{m+n} \sum_{i+j=k} a_i b_j x^k.$$

Es decir, que para sumar polinomios sumamos los coeficientes de términos de igual grado. Para multiplicarlos, multiplicamos todos los términos del primero por todos los del segundo y sumamos, agrupando los coeficientes de igual grado. El grado del producto es la suma de los grados de los factores.

3.2. División euclídea de polinomios

Definición 89 *Dados dos polinomios no nulos* P, Q *de* $\mathbb{K}[x]$*, se dice que* Q ***divide*** *a* P *(se denota por* $Q|P$*) si existe un polinomio* $C \in \mathbb{K}[x]$ *tal que* $P = CQ$ *(también se dice que* Q *es un* ***divisor*** *de* P *y que* P *es un* ***múltiplo*** *de* Q*).*

Definición 90 *Dados dos polinomios no nulos* P, Q *de* $\mathbb{K}[x]$*, se define un máximo común divisor de* P *y* Q *como un polinomio* D*, de grado máximo, tal que* $D|P$ *y* $D|Q$*.*

Definición 91 *Dados dos polinomios no nulos* P, Q *de* $\mathbb{K}[x]$*, se define un mínimo común múltiplo de* P *y* Q *como un polinomio* M*, de grado mínimo, tal que:* $P|M$ *y* $Q|M$*.*

Teorema 17 (de división de Euclides) *Dados* $P, Q \in \mathbb{K}[x]$*, con* $Q \neq 0$*, existen únicos* $C, R \in \mathbb{K}[x]$ *tales que*

$$P = CQ + R$$

y grado$(R) <$ *grado*(Q)*.*

Observación: $R = 0$ si y solo si Q divide a P.

Proposición 42 *Si* $P = CQ + R$*, entonces m.c.d.*$(P, Q) =$*m.c.d.*(Q, R)*.*

De esta proposición se deduce el **Algoritmo de Euclides** para el cálculo del máximo común divisor de dos polinomios:

Dados P, Q, con grado$(P) \geq$ grado(Q), hacemos la división euclídea $P = C_1 Q + R_1$; si R_1 es no nulo, volvemos a dividir: $Q = C_2 R_1 + R_2$; de nuevo, si R_2 es no nulo, volvemos a dividir $R_1 = C_3R_2 + R_3$, y reiteramos el proceso hasta que lleguemos a un resto $R_{k+1} = 0$. Como $R_{k-1} = C_{k+1}R_k$, en virtud de la proposición anterior tenemos que

$$\text{m.c.d.}(P, Q) = \text{m.c.d.}(Q, R_1) = \text{m.c.d.}(R_1, R_2) = \ldots = \text{m.c.d.}(R_{k-1}, R_k) = R_k$$

es decir, que un **máximo común divisor de** P **y** Q **es el último resto no nulo** obtenido en el proceso anterior.

Teorema 18 (del resto) *Para cualquier polinomio* $P \neq 0$*, y para cualquier* $a \in \mathbb{K}$*, el resto de dividir* $P(x)$ *entre* $x - a$ *vale* $P(a)$*.*

3.3. Raíces de un polinomio

Dado un polinomio de segundo grado, $ax^2 + bx + c$, la fórmula para calcular sus raíces

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

es conocida desde la antigüedad; sabemos también que si el discriminante $\Delta = b^2 - 4ac$ no es nulo, el polinomio admite dos raíces distintas y si es nulo tiene una raíz doble. ¿Qué ocurre para polinomios de grado mayor que 2? En el siglo XVI se encontraron fórmulas para obtener explícitamente las raíces de polinomios de tercer y cuarto grado y desde entonces hasta el siglo XIX fueron muchos los esfuerzos que se hicieron por encontrar una fórmula general. Fue un jovencísimo matemético francés, Évariste Galois (1811-1832), quien demostró la inexistencia de tal fórmula.

En este epígrafe indicamos algunos resultados relacionados con raíces de polinomios pero nos parece adecuado indicar la diferencia sustancial entre la existencia *teórica* de raíces y el cálculo *explícito* de las mismas.

Definición 92 *Dado* $P \in \mathbb{K}[x]$*, se dice que* $a \in \mathbb{K}$ *es una raíz de* P *si* $P(a) = 0$ *(i.e. si* P *es divisible entre* $x - a$*).*

Se dice que a *es una raíz múltiple de orden* $m \geq 2$ *si* $P(x) = (x - a)^m Q(x)$ *con* $Q(a) \neq 0$*; para* $m = 1$*, se dice que* a *es una raíz simple.*

Proposición 43 (Caracterización de las raíces múltiples) *Dados* $P \in \mathbb{K}[x]$ *y* $a \in \mathbb{K}$*,* a *es raíz de* P *múltiple de orden* m *si y solo si*

$$P(a) = P'(a) = \ldots = P^{(m-1)}(a) = 0 \text{ pero } P^{(m)}(a) \neq 0.$$

Teorema 19 (Fundamental del Álgebra) *Todo polinomio complejo no constante posee alguna raíz.*

Si P es un polinomio de grado $n > 1$, tiene alguna raíz $a_1 \in \mathbb{C}$; entonces $P(z) = (z - a_1)Q(z)$, donde Q es un polinomio de grado $n - 1$ que también tendrá alguna raíz $a_2 \in \mathbb{C}$; podemos aplicar el teorema recursivamente hasta llegar a un polinomio de grado 0 (es decir, constante), lo que nos permite concluir el siguiente resultado:

Corolario: **(factorización en $\mathbb{C}[z]$)** Todo polinomio complejo de grado n puede escribirse como producto de factores n lineales, es decir:

$$P(z) = \alpha(z - z_1)(z - z_2) \dots (z - z_n),$$

en donde $\alpha \in \mathbb{C}$ es el coeficiente principal de P y $z_1, z_2, \dots, z_n \in \mathbb{C}$ son sus raíces.

Observación: Otra forma de expresar la factorización es

$$P(z) = \alpha \prod_{k=1}^{p} (z - \alpha_k)^{m_k}$$

en donde α_k son las raíces distintas y m_k sus respectivas multiplicidades y $\displaystyle\sum_{k=1}^{p} m_k = n$.

Observación: **(Raíces de polinomios reales.)** Sea $P(z) = a_0 + a_1 z + \ldots + a_n z^n \in \mathbb{C}[z]$; Entonces $\overline{P(z)} = \overline{a}_0 + \overline{a}_1 \overline{z} + \ldots + \overline{a}_n \overline{z}^n$, con lo que $z_0 \in \mathbb{C}$ será raíz de P si y solo si su conjugada $\overline{z}_0$ lo es del polinomio $\overline{P}$ (el polinomio cuyos coeficientes son los conjugados de los de P), y ambas tienen la misma multiplicidad.

En particular, si $P \in \mathbb{R}[x]$, z_0 será raíz de P si y solo si $\overline{z}_0$ lo es. Así, las raíces de un polinomio con coeficientes reales, o bien son reales, o bien aparecen como pares de complejos conjugados, ambos de la misma multiplicidad.

Como consecuencia, todo polinomio real y de grado impar tiene alguna raíz real.

3.3.1. Raíces de polinomios con coeficientes enteros

Proposición 44 *Sea $P(x) = a_0 + a_1 x + \ldots a_{n-1} x^{n-1} + a_n x^n$, con $a_j \in \mathbb{Z}\ \forall j$ y $a_0 \neq 0$; si $\frac{t}{s}$ (fracción irreducible) es una raíz racional de P, entonces $t | a_0$ y $s | a_n$.*

En particular, si P es mónico, sus raíces racionales tienen que ser números enteros divisores de $a_0 \neq 0$.

Ejercicios

1.– Dados los números complejos $z = 3 + 2i$ y $w = 4 - i$, exprésense en forma binómica los siguientes números complejos:

$$z + w,\ 3z - 4w,\ iz + w,\ zw,\ z/w.$$

2.– Exprésese en forma binómica el número complejo

$$\frac{2 + 11i}{3 + 4i}.$$

3.– Determínese el argumento principal y la forma exponencial de los números complejos

$$1,\ i,\ 1 + i,\ \sqrt{3} + i,\ -1 - \sqrt{3}i.$$

4.– Utilícese la fórmula de De Moivre para expresar $\cos 3\theta$ como un polinomio en $\cos\theta$ y $\operatorname{sen} 3\theta$ en función de potencias de $\operatorname{sen}\theta$. Dedúzcanse las igualdades:

$$\begin{aligned} \cos^3\theta &= \frac{3}{4}\cos\theta + \frac{1}{4}\cos 3\theta, \\ \operatorname{sen}^3\theta &= \frac{3}{4}\text{sen } \theta - \frac{1}{4}\text{sen } 3\theta. \end{aligned}$$

5.– Calcúlese un polinomio sabiendo que el resto de dividirlo por $x+2$ es -2, el cociente de dividirlo entre $x^2 - 4$ es $3x$ y su derivada admite 1 como raíz.

6.– Calcúlese un polinomio $p(z)$ de grado 2 que verifique

$$p\,(1) = p\,(2) = -1 + 3i, \quad p\,(i) = 0.$$

¿Es único?, ¿existe algún polinomio de grado 3 que cumpla esas condiciones?, ¿es único?

7.– Factorícese el polinomio $P(z) = 2z^4 - z^3 - 7z^2 + 3z + 3$.
Indicación: Utilícese la proposición 4 para establecer las raíces racionales.

8.– Factorícese el polinomio $P(z) = 4z^4 - 12z^3 + 13z^2 - 12z + 9$ sabiendo que alguna de sus raíces es imaginaria pura.

9.– Calcúlense las dos raíces cuadradas del número complejo $-80 - 18i$.
Indicación: Resuélvase el sistema de ecuaciones que resulta al identificar las partes real e imaginaria en la ecuación $(x + iy)^2 = -80 - 18i$.

Como aplicación, resuélvase la ecuación de segundo grado

$$z^2 - (5 + i)z + (26 + 7i) = 0.$$

10.– Calcúlense las raíces de la ecuación $z^2 + |z|^2 = 1 + i$.

Bibliografía incompleta

512-BUR A J. de Burgos *Álgebra Lineal y geometría cartesiana.* McGraw–Hill. Madrid, 2006.

512-BUR B J. de Burgos *Álgebra Lineal: definiciones, teoremas y resultados.* García-Maroto. Madrid, 2007.

512-GRO S. I. Grossmann, *Álgebra Lineal.* McGraw–Hill. México, 2007.

512-LAY D. C. Lay, *Álgebra Lineal y sus aplicaciones.* 3a. ed., Pearson Addison Wesley, México 2007.

¿? C. D. Meyer, *Matrix Analysis and Applied Linear Algebra.* SIAM[2]

512-NOB B. Noble B., J. W. Daniel, *Álgebra lineal aplicada.* Prentice-Hall Hispanoamericana. México, 1989.

512-ROJ J. Rojo *Álgebra Lineal.* McGraw–Hill. Madrid, 2001.

512-VIL–P De la Villa A. *Problemas de Álgebra.* Clagsa. Madrid, 1994.

Nota: El número 512 y las siglas que anteceden a cada libro se refieren a la signatura de registro de los libros en la biblioteca de la E.T.S.I. Industriales.

[2]Este libro aún no está en la biblioteca, pero se puede consultar online