uc3m | Universidad Carlos III de Madrid

Grado en Ingeniería Informática

2018-2019

Apuntes

Algebra Lineal

Jorge Rodríguez Fraile[1]

[1] Universidad: 100405951@alumnos.uc3m.es | Personal: jrf1616@gmail.com

ÍNDICE GENERAL

Parte I

Información

Cronograma Álgebra Lineal: Grupo 84 (06/Septiembre/2018)

Semana	Sesión	Día	Actividad
0	0	Mie 05/Sep 17:00	Introducción
1	1	Vie 07/Sep 15:00	Magistral 1: Matrices
	2	Mie 12/Sep 17:00	Sesión de problemas 1: matrices
2	3	Vie 14/Sep 15:00	Magistral 2: Sistemas de Ecuaciones Lineales
	4	Mie 19/Sep 17:00	Sesión de problemas 2: sistemas de ecuaciones lineales
3	5	Vie 21/Sep 15:00	Magistral 3: Espacios Vectoriales
	6	Mie 26/Sep 17:00	Sesión de problemas 3: espacios vectoriales
4	7	Vie 28/Sep 15:00	Magistral 4: Bases y dimensión
	8	Mie 03/Oct 17:00	Sesión de problemas 4: bases y dimensión
5	9	Vie 05/Oct 15:00	Magistral 5: Transformaciones Lineales
	10	Mie 10/Oct 17:00	Sesión de problemas 5: transformaciones lineales
6	11	Mar 16/Oct 19:00	Magistral 6: TL y matrices
	12	Mie 17/Oct 17:00	Sesión de problemas 6: TL y matrices
7	13	Vie 19/Oct 15:00	Magistral 7: Forma Normal
	14	Mie 24/Oct 17:00	Sesión de problemas 7: forma normal
8	15	Vie 26/Oct 15:00	Magistral 8: Autovalores y Autovectores
	16	Mie 31/Oct 17:00	Sesión de problemas 8: autovalores y autovectores
9	17	Mar 06/Nov 19:00	Magistral 9: Producto Interno
	18	Mie 07/Nov 17:00	Sesión de problemas 9: producto interno
10	19	Vie 09/Nov 15:00	Magistral 10: Bases Ortogonales
	20	Mar 13/Nov 19:00	**Control Temas 1–9**
	21	Mie 14/Nov 17:00	Sesión de problemas 10: bases ortogonales
11	22	Vie 16/Nov 15:00	Magistral 11: Teorema Espectral
	23	Mie 21/Nov 17:00	Sesión de problemas 11: teorema espectral
12	24	Vie 23/Nov 15:00	Magistral 12: Geometría y TL
	25	Mie 28/Nov 17:00	Sesión de problemas 12: geometría y TL
13	26	Vie 30/Nov 15:00	Magistral 13: Mínimos Cuadrados
	27	Mie 05/Dic 17:00	Sesión de problemas 13: mínimos cuadrados
14	28	Mar 11/Dic 19:00	Magistral 14: Pseudoinversa
	29	Mie 12/Dic 17:00	Sesión de problemas 14: pseudoinversa

Referencias:

Véase Ficha de la asignatura

Parte II

Tema 1. Matrices y determinantes

Tema 1

Matrices

1.1. Algunas definiciones básicas

En este tema de introducción a la asignatura de Álgebra Lineal definimos algunas nociones y enunciamos algunas propiedades de las matrices, en su mayoría conocidas de cursos anteriores.

- Una **matriz de dimensión** $m \times n$ es una disposición rectangular en m filas y n columnas de números, reales o complejos, denominados *elementos* o entradas. Las matrices se suelen indicar con letras mayúsculas, mientras que los elementos de las mismas se denotan con letras minúsculas:

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

  Así, el elemento (i, j) de la matriz, es decir, el que ocupa la posición dada por la i-ésima fila y la j-ésima columna, se representa mediante a_{ij}.

NOTA: en algunas ocasiones es más sencillo representar los elementos de la matriz A usando también letras mayúsculas; por ejemplo, $A = (A_{ij})$.

El conjunto de todas las matrices $m \times n$ con elementos reales se denota por $\mathbb{R}^{m\times n}$; el conjunto de todas las matrices $m \times n$ con elementos complejos se denota por $\mathbb{C}^{m\times n}$. Para referirnos en general a cualquiera de estos conjuntos escribiremos $\mathbb{K}^{m\times n}$.

También usaremos el término **escalar** para referirnos a un elemento del conjunto $\mathbb{K}$, ya sea éste $\mathbb{R}$ ó $\mathbb{C}$. (El símbolo $\mathbb{K}$ representa en general un conjunto con la estructura algebraica de cuerpo. En este curso nos centraremos fundamentalmente en el cuerpo de los reales. La definición precisa de cuerpo se verá en el Tema 3).

- Una matriz de dimensión $m \times m$ se denomina **matriz cuadrada**.

- Una matriz de dimensión $m \times 1$, es decir, que consta de una única columna, a menudo se denomina *vector columna*; del mismo modo, una matriz de dimensión $1 \times m$, es decir, que consta de una única fila, se denomina *vector fila*. En ocasiones se simplifica la notación y escribiremos simplemente $A \in \mathbb{K}^m$.

- La **diagonal principal** de una matriz A de dimensión $m \times n$ está formada por los elementos de la forma a_{ii}, para $1 \leq i \leq \text{mín}\{m, n\}$.

- Una matriz cuadrada en la que todas las entradas que no están en la diagonal principal son iguales a cero se denomina **matriz diagonal**. Es decir, $a_{ij} = 0$ para todo $i \neq j$.

- Una matriz cuadrada cuyas entradas por encima de su diagonal principal son cero se llama **matriz triangular inferior**. Es decir, $a_{ij} = 0$ para todo $j > i$. Una matriz cuadrada cuyas entradas por debajo de su diagonal principal son cero se llama **matriz triangular superior**. Es decir, $a_{ij} = 0$ para todo $j < i$.

- Una matriz de dimensión $m \times n$ con todas las entradas iguales a cero se denomina **matriz cero** y la representamos por $0_{m\times n}$ (ó simplemente por 0 si las dimensiones se sobreentienden).

- Una matriz cuadrada de dimensión $n \times n$ con todas las entradas iguales a cero excepto las de la diagonal principal, que valen 1, se denomina **matriz identidad** y se representa mediante I_n (ó por I si las dimensiones se sobreentienden).

- Se llama **traza** de una matriz cuadrada A (y la denotamos por $\mathrm{tr}(A)$) a la suma de los elementos de la diagonal principal:

$$\boxed{\mathrm{tr}(A) = \sum_{i=1}^{m} a_{ii}\ .}$$

- Dados los vectores columna $v_1, \ldots, v_p$, llamamos **combinación lineal** de estos vectores a una expresión de la forma: $\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n$, donde $\alpha_1, \alpha_2, \ldots, \alpha_n$ son escalares. Decimos que los vectores $v_1, \ldots, v_p$ son **linealmente independientes** si, de la expresión $\alpha_1 v_1 + \cdots + \alpha_p v_p = 0$, se deduce que necesariamente $\alpha_1 = \alpha_2 = \cdots = \alpha_p = 0$.

  Entonces, llamamos **rango de una matriz** A (y lo denotamos por $\mathrm{rg}(A)$) al número de columnas linealmente independientes de la matriz. Este número coincide con el número de filas linealmente independientes de A.

1.2. Operaciones con matrices

A continuación describimos distintas operaciones que involucran matrices y enunciamos algunas de sus propiedades más importantes.

1.2.1. Suma de matrices

Dadas las matrices $A, B \in \mathbb{K}^{m\times n}$, definimos la **suma** de A y B, y lo representamos por $A + B$, como la matriz cuyo elemento (i, j) es $(A + B)_{ij} = a_{ij} + b_{ij}$.

Propiedades de la suma de matrices:

Si $A, B \in \mathbb{K}^{m\times n}$, entonces la suma de matrices satisface las propiedades:

1. Asociativa: $(A + B) + C = A + (B + C)$.
2. Conmutativa: $A + B = B + A$.
3. Elemento neutro: $A + 0_{m\times n} = 0_{m\times n} + A = A$.
4. Elemento inverso: Para toda matrix $A \in \mathbb{K}^{m\times n}$ existe el elemento inverso $-A$, dado por $-A = (-a_{ij})$, tal que $A + (-A) = (-A) + A = 0_{m\times n}$.

Por estas propiedades, se dice que $\mathbb{K}^{m\times n}$ con la operación suma tiene *estructura algebraica* de grupo conmutativo.

1.2.2. Producto por escalares

Dados el escalar (real o complejo) α y la matriz $A \in \mathbb{K}^{m\times n}$, definimos el **producto** del escalar α por A, y lo representamos por αA, mediante la matriz cuyos elemento (i, j) es $(\alpha A)_{ij} = \alpha a_{ij}$.

Propiedades del producto de matrices por escalares:

Si $A, B \in \mathbb{K}^{m\times n}$ y $\alpha, \beta \in \mathbb{K}$, entoces el producto de matrices por escalares en $\mathbb{K}^{m\times n}$ satisface las siguientes propiedades:

1. (Pseudo)asociativa: $\alpha(AB) = (\alpha A)B = A(\alpha B)$.

2. Distributiva respecto a la suma de matrices: $\alpha(A+B) = \alpha A + \alpha B$.

3. Distributiva respecto a la suma de escalares: $(\alpha+\beta)A = \alpha A + \beta A$.

4. Identidad: dado $1 \in \mathbb{K}$, para toda matriz $A \in \mathbb{R}^{m\times n}$, se tiene que $1A = A$.

1.2.3. Producto de matrices

Dadas las matrices $A \in \mathbb{K}^{m\times n}$ y $B \in \mathbb{K}^{n\times p}$, donde el número n de columnas de A es igual al número de filas de B, el **producto** $C = A\,B$ existe y tiene dimensión $m \times p$ (aunque $B\,A$ puede existir o no). Obsérvese que este producto normalmente se escribe sin el signo de multiplicación. Si $C = (c_{ij})$ escribimos:

$$c_{ij} = \sum_{k=1}^{n} a_{ik}b_{kj} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj}\,.$$

Es decir el elemento c_{ij} del producto es el **producto escalar** de la i-ésima fila de A (a_{i1}, $a_{i2}, \cdots, a_{in}$ y la j-ésima columna de B $(b_{1j}, b_{2j}, \cdots, b_{nj})$.

Propiedades del producto de matrices:

- Asociativa: $(A\,B)\,C = A\,(B\,C)$.
- Elemento neutro (o unidad): $I_m\,A = A$ y $A\,I_n = A$ para toda matriz $A \in \mathbb{K}^{m\times n}$, siendo I_m, I_n las matrices identidad de dimensión $m \times m$ y $n \times n$ respectivamente.
- La multiplicación de matrices no es conmutativa en general, es decir *normalmente* $A\,B \neq B\,A$, incluso si las matrices A y B son cuadradas.
- No toda matriz tiene inversa. Para tener inversa, una matriz tiene que ser cuadrada, pero no toda matriz cuadrada tiene inversa.
- Propiedad distributiva respecto de la suma: $A\,(B+C) = A\,B + A\,C$ (por la izquierda) y $(B+C)\,A = B\,A + C\,A$ (por la derecha).

- $A\,O_{n\times p} = O_{m\times p}$ y $O_{p\times m}\,A = O_{p\times n}$ para toda matriz $A \in \mathbb{K}^{m\times n}$.

Una matriz A de dimensión $n \times n$ que verifica $A\,A = A$ (escrito también como $A^2 = A$) se denomina **idempotente**. Estas matrices están relacionadas con las aplicaciones lineales que se denominan **proyectores** (ver Tema 6).

1.3. Traspuesta de una matriz

Cualquier matriz $A = (a_{ij}) \in \mathbb{K}^{m\times n}$ tiene una única matriz **traspuesta**, representada por A^t, de tamaño $n \times m$, cuyo elemento (i, j) se define por $(A^t)_{ij} = a_{ji}$.

Obviamente, la primera fila de A es la primera columna de A^t; la segunda fila de A es la segunda columna de A^t y así sucesivamente.

Propiedades de la traspuesta:

- $(A^t)^t = A$.
- $(A + B)^t = A^t + B^t$.
- $(\alpha A)^t = \alpha A^t$.
- $\boxed{(A\,B)^t = B^t\,A^t.}$

La matriz A se dice **simétrica** si $A = A^t$. La matriz A se dice **antisimétrica** si $A = -A^t$. Obviamente, las matrices simétricas y antisimétricas son cuadradas.

Dada la matriz $A \in \mathbb{K}^{m\times n}$, podemos calcular otras matrices relacionadas con ella.

Las matrices $A^t\,A$ y $A\,A^t$ son simétricas.

Demostración.

$$\begin{aligned} (A^t A)^t &= A^t (A^t)^t = A^t A\,, \\ (A A^t)^t &= (A^t)^t A^t = A A^t\,. \end{aligned}$$

□

La matriz $A^t A$ será de gran importancia en los Temas 13 y 14 del curso.

Además dada la matriz *cuadrada* $A \in \mathbb{K}^{n\times n}$, se tiene que:

La matriz $A + A^t$ es simétrica y la matriz $A - A^t$ es antisimétrica.

Demostración.

$$\begin{aligned} (A + A^t)^t &= A^t + (A^t)^t = A^t + A = A + A^t\,, \\ (A - A^t)^t &= A^t - (A^t)^t = A^t - A = -(A - A^t)\,. \end{aligned}$$

□

NOTA: A menudo escribiremos los vectores columna $\begin{pmatrix} v_1 \\ \vdots \\ v_m \end{pmatrix}$ en la forma $(v_1, \ldots, v_m)^t$ para ahorrar espacio.

1.4. Inversa de una matriz cuadrada

Consideremos una matriz cuadrada A de dimensión $n \times n$. Si A tiene rango n, se dice que A es **no singular**; si su rango es menor que n, la llamaremos matriz **singular**. En el caso de que A sea no singular, podemos asociarle una matriz especial denominada *matriz inversa*.

Toda matriz cuadrada $A \in \mathbb{K}^{n \times n}$ no singular (i.e., de rango n) tiene una matriz **inversa**, de dimensión $n \times n$, denotada por A^{-1}, que satisface:

$$A\,A^{-1} = A^{-1}\,A = I_n\,.$$

Una matriz con inversa se dice que es **invertible**.

Propiedades de la matriz inversa:

- Si la matriz A tiene inversa A^{-1}, ésta es única.
- Si A y B son matrices invertibles de dimensión $n \times n$, entonces $A\,B$ es invertible y $\boxed{(A\,B)^{-1} = B^{-1}\,A^{-1}.}$
- Si A es invertible entonces A^{-1} es invertible y $(A^{-1})^{-1} = A$.
- Si A es invertible entonces A^t es invertible y $(A^t)^{-1} = (A^{-1})^t$.
- Si A es invertible y α es un escalar no nulo, entonces αA es invertible y $(\alpha A)^{-1} = \alpha^{-1}\,A^{-1}$ donde α^{-1} es el único elemento inverso (multiplicativo) de $\alpha \in \mathbb{K}$.
- A es no singular si y sólo si A es invertible.
- Si A y B son matrices cuadradas de dimensión $n \times n$, el producto $A\,B$ es no singular si y sólo si A y B son las dos no singulares.

Veamos algunas definiciones que necesitaremos en temas posteriores:

- Si A es una matriz de dimensión $m \times n$ tal que $A^t\,A = I_n$, decimos que A es **ortogonal**. Como consecuencia, si A es una matriz ortogonal cuadrada de dimensión $n \times n$, entonces A es no singular y $A^{-1} = A^t$.

- Dos matrices $A, B \in \mathbb{K}^{m\times n}$ son **equivalentes** si existen matrices invertibles P y Q tales que $B = Q^{-1} A P$.

- Dos matrices cuadradas $A, B \in \mathbb{K}^{n\times n}$ se dicen **semejantes** si existe una matriz invertible P tal que $B = P^{-1} A P$.

1.5. Determinantes de matrices cuadradas

Toda matriz *cuadrada* $A \in \mathbb{K}^{n\times n}$ tiene asociado un determinante, denotado por $\det(A)$ ó $|A|$, que es un elemento de $\mathbb{K}$ calculado a partir de los elementos de la matriz. El valor numérico del mismo se obtiene por medio de un algoritmo recursivo; esto quiere decir que el cálculo del determinante de una matriz de dimensión $n\times n$ está definida en función de los determinantes de matrices de dimensión $(n-1)\times(n-1)$ y éstos en función de los determinantes de matrices de dimensión $(n-2)\times(n-2)$ y así sucesivamente hasta alcanzar matrices de dimensión 2×2, para las que el determinante se calcula con una fórmula. Para describir con detalle el cálculo, necesitamos las siguientes definiciones.

Dada una matriz $A = (a_{ij}) \in \mathbb{K}^{n\times n}$ el **menor** (i, j) **de** A, denotado por M^A_{ij}, es la matriz de $\mathbb{K}^{(n-1)\times(n-1)}$ que se obtiene de A al eliminar la i-ésima fila y la j-ésima columna. El **cofactor de** a_{ij}, denotado por C^A_{ij}, se define mediante la fórmula

$$C^A_{ij} = (-1)^{i+j} \det\left(M^A_{ij}\right) .$$

Desarrollo de Laplace

Dada una matriz $A = (a_{ij}) \in \mathbb{K}^{n\times n}$ se tiene:

1. Para cualquier fila $1 \leq i \leq n$, $\det(A) = \sum_{j=1}^{n} a_{ij} C_{ij}^A$.

2. Para cualquier columna $1 \leq j \leq n$, $\det(A) = \sum_{i=1}^{n} a_{ij} C_{ij}^A$.

Este desarrollo nos proporciona el siguiente método para el cálculo del determinante.

- Para una matriz $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ de dimensión 2×2 el determinante se calcula mediante:

$$\det(A) = ad - bc\,.$$

- Para una matriz $A = (a_{ij})$ de dimensión 3×3 el determinante se puede calcular mediante:

$$\det(A) = a_{11}\, C_{11}^A + a_{12}\, C_{12}^A + a_{13}\, C_{13}^A\,.$$

  Pero también puede calcularse con el mismo proceso aplicado a cualquier fila o columna: por ejemplo mediante $\det(A) = a_{21}\, C_{21}^A + a_{22}\, C_{22}^A + a_{23}\, C_{23}^A$ ó mediante $\det(A) = a_{13}\, C_{13}^A + a_{23}\, C_{23}^A + a_{33}\, C_{33}^A$.

- Para matrices de dimensión mayor se utiliza la misma fórmula de manera recursiva. Por ejemplo, si $A = (a_{ij})$ es 4×4, evaluando con la primera fila tendremos:

$$\det(A) = a_{11}\, C_{11}^A + a_{12}\, C_{12}^A + a_{13}\, C_{13}^A + a_{14}\, C_{14}^A\,,$$

  donde cada uno de los cofactores $C_{11}^A, C_{12}^A, C_{13}^A, C_{14}^A$ se obtiene con la definición del determinante de una matriz 3×3, etc.

- Este método no es muy eficiente. Un método mejor consiste en utilizar las siguientes propiedades.

Propiedades de los determinantes:

Dadas la matriz cuadrada $A \in \mathbb{K}^{n\times n}$ y los escalares $\alpha, \beta \in \mathbb{K}$, se tiene:

1. Si A tiene dos filas o dos columnas iguales, $\det(A) = 0$.

2. $\det(I_n) = 1$.

3. Para todo $1 \le i \le n$, el determinante es una **función lineal** de la i-ésima columna. Es decir

$$\begin{vmatrix} a_{11} & \cdots & \alpha a_{1i} + \beta b_{1i} & \cdots & a_{1n} \\ a_{21} & \cdots & \alpha a_{2i} + \beta b_{2i} & \cdots & a_{2n} \\ \vdots & & \vdots & & \vdots \\ a_{n1} & \cdots & \alpha a_{ni} + \beta b_{ni} & \cdots & a_{nn} \end{vmatrix} = \alpha \begin{vmatrix} a_{11} & \cdots & a_{1i} & \cdots & a_{1n} \\ a_{21} & \cdots & a_{2i} & \cdots & a_{2n} \\ \vdots & & \vdots & & \vdots \\ a_{n1} & \cdots & a_{ni} & \cdots & a_{nn} \end{vmatrix}$$
$$+\beta \begin{vmatrix} a_{11} & \cdots & b_{1i} & \cdots & a_{1n} \\ a_{21} & \cdots & b_{2i} & \cdots & a_{2n} \\ \vdots & & \vdots & & \vdots \\ a_{n1} & \cdots & b_{ni} & \cdots & a_{nn} \end{vmatrix}$$

   Igualmente, el determinante es una **función lineal** de la i-ésima fila.

De estas propiedades básicas se pueden probar muchas más: si $A, B \in \mathbb{K}^{n\times n}$ y $\alpha \in \mathbb{K}$, entonces:

1. $\det(A) = 0$ si y sólo si A es singular.

2. $\det(A\,B) = \det(A)\det(B)$.

3. $\det(A^t) = \det(A)$.

4. Si la matriz D se obtiene al intercambiar entre sí dos filas o dos columnas de A, entonces $\mathrm{det}(D) = -\mathrm{det}(A)$.

5. Si la matriz D se obtiene al multiplicar una fila o columna de A por un escalar $\alpha \in \mathbb{K}$, entonces $\mathrm{det}(D) = \alpha \,\mathrm{det}(A)$. Luego $\mathrm{det}(\alpha A) = \alpha^n \,\mathrm{det}(A)$.

6. Si la matriz D se obtiene a partir de la matriz A sumando a una fila cualquiera de A una combinación lineal del resto de las filas de A, entonces $\mathrm{det}(D) = \mathrm{det}(A)$. Lo mismo ocurre si consideramos las columnas de A.

7. Si A^{-1} existe entonces $\mathrm{det}(A^{-1}) = \dfrac{1}{\mathrm{det}(A)}$.

8. Si A^{-1} existe entonces $A^{-1} = \dfrac{1}{\mathrm{det}(A)} \times \mathrm{adj}(A)$, donde $\mathrm{adj}(A)$ es la **matriz adjunta de** A que se obtiene sustituyendo cada entrada en A por su cofactor y después trasponiendo la matriz resultante.

1.6. Matrices por bloques

En ocasiones, es particularmente útil "descomponer" matrices con un gran número de filas o columnas en otras más pequeñas, a veces solamente para ahorrar espacio y a veces por la importante ventaja de que permite resolver problemas más pequeños que el original de manera más sencilla. Estas submatrices se denominan **bloques**, se denotan por A_{ij} y se construyen trazando rectas verticales y horizontales imaginarias entre las filas y columnas de A:

$$A = \left(\begin{array}{c|c|c|c} A_{11} & A_{12} & \dots & A_{1s} \\ \hline \vdots & \vdots & & \vdots \\ \hline A_{r1} & A_{r2} & \dots & A_{rs} \end{array}\right).$$

NOTA: Se ha usado la misma notación para los bloques A_{ij} que para los elementos de matriz (que se pueden considerar bloques de dimensión 1×1). Si el significado de la notación A_{ij} no está claro por el contexto, éste se hará explícito en cada caso (elementos o bloques de la matriz A).

Ejemplo

Podemos descomponer la matriz A de dimensión 5×4, por ejemplo, en 2×2 bloques, como sigue:

$$A = \begin{pmatrix} 1 & 2 & -1 & 3 \\ 2 & 4 & 0 & 7 \\ 3 & 5 & 3 & 8 \\ 4 & 2 & 6 & 3 \\ 5 & 6 & -3 & 3 \end{pmatrix} = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}.$$

Esto permite, por ejemplo, realizar operaciones entre matrices descompuestas en bloques de dimensiones adecuadas:

- **Suma**: para sumar las matrices A y B, descompuestas en bloques A_{ij} y B_{ij} y cuyas dimensiones son iguales para todo i, j, entonces

  $$(A + B)_{ij} = A_{ij} + B_{ij} .$$

- **Producto**: el producto de las matrices A y B, descompuestas en bloques, se puede efectuar siguiendo la regla usual de su multiplicación, considerando a las submatrices como elementos:

  $$(A\, B)_{ij} = A_{i1}\, B_{1j} + A_{i2}\, B_{2j} + \cdots + A_{in}\, B_{nj} ,$$

  siempre y cuando las operaciones estén bien definidas, es decir, cuando los bloques tengan las dimensiones adecuadas.

Ejemplo

Podemos multiplicar las matrices

$$A = \left(\begin{array}{ccc|c} 0 & 2 & -3 & 4 \\ \hline 8 & -4 & 0 & 1 \\ 3 & 2 & -3 & 0 \end{array}\right) = \left(\begin{array}{c|c} A_{11} & A_{12} \\ \hline A_{21} & A_{22} \end{array}\right),$$

$$B = \left(\begin{array}{c|ccc} 4 & 2 & 6 & 3 \\ 0 & 2 & -1 & -3 \\ -2 & 2 & 0 & 1 \\ \hline 2 & -1 & 3 & 6 \end{array}\right) = \left(\begin{array}{c|c} B_{11} & B_{12} \\ \hline B_{21} & B_{22} \end{array}\right)$$

por bloques como sigue:

$$A\,B = \left(\begin{array}{c|c} A_{11}B_{11} + A_{12}B_{21} & A_{11}B_{12} + A_{12}B_{22} \\ \hline A_{21}B_{11} + A_{22}B_{21} & A_{21}B_{12} + A_{22}B_{22} \end{array}\right) = \left(\begin{array}{c|ccc} 14 & -6 & 10 & 15 \\ \hline 34 & 7 & 55 & 42 \\ 18 & 4 & 16 & 0 \end{array}\right).$$

Obviamente el resultado es el mismo que el que se obtiene sin usar bloques, pero se opera con matrices más pequeñas, lo cual supone una ventaja, por ejemplo, desde un punto de vista computacional.

De la misma manera, se pueden obtener la traspuesta o la inversa de una matriz por bloques o definir conceptos muy útiles en el álgebra lineal como las matrices diagonales por bloques.

- Una matriz cuadrada descompuesta en bloques en la que todas los bloques que no están en la diagonal principal son iguales a cero se denomina **matriz diagonal por**

bloques. Es decir, $A_{ij} = 0$ para todo $i \neq j$.

$$A = \left(\begin{array}{c|c|c|c} A_{11} & 0 & \dots & 0 \\ \hline 0 & A_{22} & \dots & 0 \\ \hline \vdots & \vdots & & \vdots \\ \hline 0 & 0 & \dots & A_{rr} \end{array}\right).$$

En esta ecuación 0 representa bloques con la dimensión adecuada y cuyas entradas son todas iguales al escalar $0 \in \mathbb{K}$.

- Una matriz cuadrada descompuesta en bloques en la que los bloques por encima de su diagonal principal son cero se llama **matriz triangular inferior por bloques**. Es decir, $A_{ij} = 0$ para todo $j > i$. Una matriz cuadrada descompuesta en bloques en la que los bloques por debajo de su diagonal principal son cero se llama **matriz triangular superior por bloques**. Es decir, $B_{ij} = 0$ para todo $j < i$.

$$A = \left(\begin{array}{c|c|c|c} A_{11} & 0 & \dots & 0 \\ \hline A_{21} & A_{22} & \dots & 0 \\ \hline \vdots & \vdots & & \vdots \\ \hline A_{r1} & A_{r2} & \dots & A_{rr} \end{array}\right), \qquad B = \left(\begin{array}{c|c|c|c} B_{11} & B_{12} & \dots & B_{1r} \\ \hline 0 & B_{22} & \dots & B_{2r} \\ \hline \vdots & \vdots & & \vdots \\ \hline 0 & 0 & \dots & B_{rr} \end{array}\right).$$

- Para las matrices triangulares (superior e inferior) por bloques A, así como para todas las matrices diagonales por bloques, se cumple que:

$$\det(A) = \prod_{k=1}^{r} \det(A_{kk}), \qquad \operatorname{tr}(A) = \sum_{k=1}^{r} \operatorname{tr}(A_{kk}).$$

1.7. Conjuntos inducidos por una matriz

En esta sección vamos a estudiar cuatro importantes conjuntos asociados a cualquier matriz $A \in \mathbb{K}^{m \times n}$; dos de ellos, son subconjuntos de $\mathbb{K}^m$ y los otros dos son subconjuntos de $\mathbb{K}^n$. En temas posteriores, veremos muchas propiedades interesantes de estos conjuntos; de momento nos limitamos a aprender a calcularlos.

Espacio nulo

Dada una matriz A de dimensión $m \times n$, su **espacio nulo** $N(A)$ es el conjunto de todos los elementos v de $\mathbb{K}^n$ que verifican que el producto $A\,v$ es el elemento 0 de $\mathbb{K}^m$.

NOTA: Obsérvese que las dimensiones de las matrices/vectores mencionados en la definición son las necesarias para que la multiplicación tenga sentido.

Espacio columna

Dada la matriz $A \in \mathbb{K}^{m\times n}$, con columnas denotadas por $(A_1, A_2, A_3, \ldots, A_n)$, su **espacio columna**, denotado por $\mathcal{C}(A)$, es el subconjunto de $\mathbb{K}^m$ que contiene todas las combinaciones lineales de las columnas de A.

Espacio fila

El **espacio fila** de $A \in \mathbb{K}^{m\times n}$ es el conjunto de $\mathbb{K}^n$ formado por todas las combinaciones lineales de las filas de A. Lo denotamos por $\mathcal{C}(A^t)$.

NOTA: Obsérvese que usamos la notación anterior para enfatizar que el espacio fila de A es el espacio columna de su traspuesta. Finalmente:

Espacio nulo de la traspuesta

El **espacio nulo de la traspuesta** es, obviamente, el espacio nulo de la matriz traspuesta A^t y lo denotamos por $N(A^t)$. A veces se denomina *espacio nulo izquierdo*.

Si resolvemos $A^t\,v = 0$, podemos trasponer ambos miembros y obtener $v^t\,A = 0^t$. Esto indica que el espacio nulo de la traspuesta es el conjunto de vectores (fila) que al multiplicar a la matriz A por la izquierda producen el vector fila con todas las componentes nulas.

Ejemplo

Sea la matriz

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

Vamos a determinar los cuatro espacios asociados.

1) **Espacio Nulo**: $N(A) = \{v \in \mathbb{R}^3 \colon A\,v = 0\}$. Si escribimos $v = (v_1, v_2, v_3)^t$, la condición $A\,v = 0$, es decir,

$$\begin{pmatrix} 1 & -2 & 0 \\ 2 & -4 & 1 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} ,$$

representa un sistema de dos ecuaciones y tres incógnitas v_1, v_2 y v_3 tal y como resulta al multiplicar las matrices e igualar componente a componente:

$$\left.\begin{aligned} v_1 - 2v_2 &= 0 \\ 2v_1 - 4v_2 + v_3 &= 0 \end{aligned}\right\} .$$

Al resolver el sistema nos queda $v_3 = 0$ y $v_1 = -2v_2$, donde v_2 actúa como parámetro. Por tanto, el espacio nulo de A es:

$$N(A) = \left\{ v \in \mathbb{R}^3 \colon v = v_2\,(-2, 1, 0)^t\,, \quad v_2 \in \mathbb{R} \right\} ,$$

es decir, los elementos de $N(A)$ son aquellos elementos de $\mathbb{R}^3$ en los que la primera coordenada es igual al doble de la segunda y cambiada de signo y la tercera coordenada es cero.

2) **Espacio columna**: $\mathcal{C}(A) = \{v \in \mathbb{R}^2 \colon v = \alpha_1 A_1 + \alpha_2 A_2 + \alpha_3 A_3\,, \alpha_i \in \mathbb{R}\}$ es el conjunto de vectores que son combinación lineal de las columnas A_i de la

matriz A. Si observamos las columnas 1 y 2, vemos que una es múltiplo de la otra, por tanto, una combinación lineal de todas las columnas puede ser reducida a una combinación lineal de sólo A_1 y A_3 (ó de sólo A_2 y A_3):

$$\begin{aligned}
\mathcal{C}(A) &= \{v \in \mathbb{R}^2 \colon v = \alpha_1 A_1 + \alpha_3 A_3\,, \alpha_1, \alpha_3 \in \mathbb{R}\} \\
&= \left\{v \in \mathbb{R}^2 \colon v = \alpha_1(1,2)^t + \alpha_3(0,1)^t\,, \alpha_1, \alpha_3 \in \mathbb{R}\right\} \\
&= \left\{v \in \mathbb{R}^2 \colon v = (\alpha_1, 2\alpha_1 + \alpha_3)^t\,, \alpha_1, \alpha_3 \in \mathbb{R}\right\} .
\end{aligned}$$

Si observamos cada una de las componentes de los elementos de $\mathcal{C}(A)$ descubrimos que obviamente la primera, α_1, puede ser cualquier número real, pero también la segunda coordenada puede ser cualquier real. Por tanto, es fácil ver que $\mathcal{C}(A)$ coincide con $\mathbb{R}^2$.

3) **Espacio fila**: Los elementos de $\mathcal{C}(A^t)$ se describen como combinaciones lineales de las dos filas de la matriz A, ya que éstas no son una múltiplo de la otra:

$$\begin{aligned}
\mathcal{C}(A^t) &= \left\{v \in \mathbb{R}^3 \colon v = \beta_1(1,-2,0)^t + \beta_2(2,-4,1)^t\,, \beta_1, \beta_2 \in \mathbb{R}\right\} \\
&= \left\{v \in \mathbb{R}^3 \colon v = (\beta_1 + 2\beta_2, -2(\beta_1 + 2\beta_2), \beta_2)^t\,, \beta_1, \beta_2 \in \mathbb{R}\right\} .
\end{aligned}$$

Obsérvese que todos los elementos de este conjunto cumplen que la segunda coordenada es igual a la primera multiplicada por -2 y que la tercera coordenada puede ser cualquier valor. Esto significa que también podemos escribir:

$$\mathcal{C}(A^t) = \left\{v \in \mathbb{R}^3 \colon v = (\beta_1, -2\beta_1, \beta_2)^t\,, \beta_1, \beta_2 \in \mathbb{R}\right\} .$$

4) **Espacio nulo de la traspuesta**: $N(A^t)$ es el conjunto de vectores de la forma

$v = (v_1, v_2)^t$ que satisfacen $A^t\, v = 0$. Por tanto

$$\begin{aligned}
N(A^t) &= \left\{ (v_1, v_2)^t \in \mathbb{R}^2 \colon \begin{pmatrix} 1 & 2 \\ -2 & -4 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \right\} \\
&= \left\{ (v_1, v_2)^t \in \mathbb{R}^2 \colon v_1 + 2v_2 = 0\,, v_2 = 0 \right\} \\
&= \left\{ (v_1, v_2)^t \in \mathbb{R}^2 \colon v_1 = v_2 = 0 \right\} = \left\{ (0, 0)^t \right\}\,,
\end{aligned}$$

esto es, el espacio nulo de la traspuesta sólo contiene un elemento: el vector 0 de $\mathbb{R}^2$.

En capítulos posteriores veremos formas sistemáticas para calcular estos conjuntos y para dar expresiones compactas de los mismos.

Tema 2

Estructuras algebraicas básicas

2.1. Operación interna

Definición 29. *Dados tres conjuntos A, B y C, se llama* ***ley de composición*** *en los conjuntos A y B y resultado en el conjunto C, y se denota por "$\oplus$", a una aplicación*[1]*:*

$$\begin{aligned} \oplus : A \times B &\longrightarrow C \\ (a,b) &\longrightarrow f(a,b) = a \oplus b = c \in C \end{aligned}$$

Definición 30. *Dada $\oplus : A \times B \to C$, se dirá que la ley de composición $\oplus$ es* ***interna*** *si $A = B = C$.*

Por lo tanto, una ley de composición $\oplus : A \times A \to A$ es una ley de composición interna.[2]

Ejemplo 23. *La suma y el producto ordinarios en $\mathbb{R}$, denotados respectivamente por "$+$" y "$\cdot$", son leyes de composición interna:*

$$\begin{aligned} + \ : \mathbb{R} \times \mathbb{R} &\longrightarrow \mathbb{R} \\ (x,y) &\longrightarrow z = x + y \end{aligned} \qquad\qquad \begin{aligned} \cdot \ : \mathbb{R} \times \mathbb{R} &\longrightarrow \mathbb{R} \\ (x,y) &\longrightarrow z = x \cdot y \end{aligned}$$

Ejemplo 24. *La operación* resta *no es una operación interna en $\mathbb{N}$, ya que el resultado de restar entre sí números naturales puede producir números negativos, que no están en $\mathbb{N}$. Por ejemplo: $1, 2 \in \mathbb{N}$ pero $1 - 2 = -1 \notin \mathbb{N}$.*

[1] A una ley de composición de $A \times B \to C$ se le conoce, también, con el nombre de **operación binaria**. La notación $\oplus$ es arbitraria, y podría elegirse cualquier otra. A lo largo de este capítulo se utilizarán diversos símbolos para representar este tipo de leyes de composición: $\oplus$, $\otimes$, $\odot$, $*$, $\cdot$, $+$, $\cdot$.

[2] Una ley de composición interna se llama, también, **operación binaria interna** u operación interna.

De la misma forma, la operación división no es interna en ninguno de los conjuntos numéricos habituales $\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}$ o $\mathbb{C}$, ya que el resultado que se obtendría al dividir entre 0 no está definido en ninguno de estos conjuntos.

Visto desde otra perspectiva: sólo determinadas operaciones resultan internas en cada conjunto. Así pues, parece necesario definir las leyes que no son internas:

Definición 31. *Dados dos conjuntos A y B, se llama* ***ley de composición externa*** *a una aplicación $A \times A \to B$, que a todo par de elementos de A asocia un elemento de B.*

La resta entre números naturales, del ejemplo anterior, resulta así una operación externa de la forma $\mathbb{N} \times \mathbb{N} \to \mathbb{Z}$.

Pero existe otra forma de ley de composición externa, con dos variantes:

Definición 32. *Dados dos conjuntos A y B, se dice que una aplicación de la forma:*

$$\begin{aligned} \odot\ : A \times B &\longrightarrow A \\ (a, b) &\longrightarrow c = a * b \end{aligned}$$

es una ***ley de composición externa por la derecha****, y a los elementos del conjunto B se les llama* ***multiplicadores*** *o* ***escalares*** *de la operación*[3].

Si la aplicación es de la forma:

$$\begin{aligned} \odot\ : B \times A &\longrightarrow A \\ (b, a) &\longrightarrow c = b \odot a \end{aligned}$$

se dirá que es una ***ley de composición externa por la izquierda****.*

Ejemplo 25. *Un ejemplo típico de operación externa es el producto de un escalar por vector en un espacio vectorial, que se verá con detalle en la sección siguiente. Si llamamos K al conjunto de escalares (cuerpo) sobre el que definiremos el espacio vectorial V, un producto de un* escalar *de K por un* vector *de V será de la forma:*

$$\begin{aligned} \bullet\ : K \times V &\longrightarrow V \\ (\lambda, v) &\longrightarrow u = \lambda \cdot v \end{aligned}$$

Ejemplo 26. *Un caso particular del ejemplo anterior es el producto de escalares por funciones reales de variable real. Si A es el conjunto de las funciones reales de variable real, $f \in A$ es una función de A, $\mathbb{R}$ el conjunto de los números reales y $k \in \mathbb{R}$ un número real, la aplicación*

$$\begin{aligned} \bullet\ : \mathbb{R} \times A &\longrightarrow A \\ \big(k, f\big) &\longrightarrow (k \cdot f)(x) = kf(x),\ \forall x \in \mathbb{R} \end{aligned}$$

resulta ser una operación externa en A.

[3]Esta ley se llama, también, **operación externa** por la derecha.

Propiedades 8. *Las leyes de composición no han de satisfacer, en general, ningún requisito en especial. Sin embargo, sólo serán interesantes aquellas que, en cada caso, verifiquen ciertas propiedades. De ellas, se exponen aquí las más interesantes, comenzando por las que se refieren a leyes de composición interna:*

Asociativa $a_1 \oplus (a_2 \oplus a_3) = (a_1 \oplus a_2) \oplus a_3, \ \forall\, a_1, a_2, a_3 \in A$

Conmutativa $a_1 \oplus a_2 = a_2 \oplus a_1, \ \forall\, a_1, a_2 \in A$

Distributiva *Dado el conjunto A y las leyes $\oplus$ y $\odot$, se dice que $\odot$ es **distributiva por la izquierda** respecto a $\oplus$ si:*

$$a_1 \odot (a_2 \oplus a_3) = (a_1 \odot a_2) \oplus (a_1 \odot a_3)\,, \ \forall\, a_1, a_2, a_3 \in A$$

*de la misma forma, se dice que $\odot$ es **distributiva por la derecha** respecto a $\oplus$ si:*

$$(a_2 \oplus a_3) \odot a_1 = (a_2 \odot a_1) \oplus (a_3 \odot a_1)\,, \ \forall\, a_1, a_2, a_3 \in A$$

*finalmente, $\odot$ es **distributiva** respecto de $\oplus$, si lo es por la izquierda y por la derecha, esto es:*

$$\begin{aligned}(a_1 \oplus a_2) \odot (a_3 \oplus a_4) = {} & (a_1 \odot a_3) \oplus (a_1 \odot a_4) \oplus \\ & \oplus (a_2 \odot a_3) \oplus (a_2 \odot a_4)\,, \ \forall\, a_1, a_2, a_3, a_4 \in A\end{aligned}$$

Elemento neutro *Se dice que una ley de composición interna en A tiene **elemento neutro** si:*

$$\exists\, e \in A \Big/ e \oplus a = a \oplus e = a, \ \forall\, a \in A$$

El elemento neutro[4] se denotará por e.

Elemento simétrico *Dada una ley de composición interna en A con elemento neutro e, se llama **elemento simétrico**, si existe, del elemento $a \in A$, a un elemento $\bar{a}$ tal que[5]:*

$$a \oplus \bar{a} = a \oplus \bar{a} = e$$

[4] Para la operación suma ordinaria en $\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}$, el elemento neutro es el “0” para la suma, y para el producto ordinario el “1”.

[5] Para la suma ordinaria, el elemento simétrico de un elemento x es $-x$, y se llama **elemento opuesto**. Para el producto ordinario entre números reales o racionales no nulos, el elemento simétrico es $\frac{1}{x}$, y se denomina **elemento inverso o recíproco**. Por otra parte, para el producto ordinario de matrices, que se estudiará más adelante, el elemento simétrico se llama **matriz inversa** y se representan por $\mathbf{A}^{-1}$.

Elemento regular o simplificable *Se dice que el elemento $a \in A$ es* ***regular o simplificable*** *para la ley de composición interna "$\oplus$" si se verifica:*

$$Si \ \ a \oplus a_1 = a \oplus a_2 \ \Rightarrow \ a_1 = a_2, \ \forall \, a_1, a_2 \in A$$

y:

$$Si \ \ a_1 \oplus a = a_2 \oplus a \ \Rightarrow \ a_1 = a_2, \ \forall \, a_1, a_2 \in A$$

La estructura de espacio vectorial, entre otras, constituye la base sobre la que se apoya el Álgebra Lineal. Los espacios vectoriales son estructuras matemáticas que cumplen unas determinadas propiedades. Estas propiedades son poco restrictivas, de forma que numerosos problemas reales pueden modelizarse mediante espacios vectoriales.

Para abordar el estudio de los espacios vectoriales recordaremos previamente una serie de definiciones de conceptos que son la base sobre la que se apoya la definición de espacio vectorial.

Hemos utilizado como ejemplo de estructuras los conjuntos de números $\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}$ y $\mathbb{C}$ que se estudiarán con detalle en el capítulo siguiente.

Definición 33. *Se llama* ***estructura algebraica*** *a un conjunto A y unas operaciones $\oplus, \otimes, \odot, \ldots$ –internas o externas– definidas en él, de forma que se verifican ciertas propiedades. Se denota por $(A, \oplus, \otimes, \odot, \ldots)$.*

Se exponen a continuación las estructuras más habituales, y las necesarias para llegar, finalmente, al espacio vectorial.

2.2. Grupos

Definición 34. *Se llama* ***grupo*** *a una estructura $(G, \otimes)$ que verifica las propiedades:*

1. $(x \otimes y) \otimes z = x \otimes (y \otimes z)\,, \ \forall \, x, y, z \in G$ *(Propiedad Asociativa)*
2. $\exists \, e \in G \ \big/ \ x \otimes e = e \otimes x = x, \ \forall \, x \in G$ *(Existencia de Elemento Neutro)*
3. $\forall x \in G \ \ \exists \, y \in G \ \big/ \ y \otimes x = x \otimes y = e$ *(Existencia de Elemento Simétrico)*

Definición 35. *Se llama* ***grupo conmutativo ó Abeliano*** *a un grupo $(G, \otimes)$ que verifica:*

4. $x \otimes y = y \otimes x, \ \forall \, x, y \in G$ *(Propiedad conmutativa)*

Ejemplo 1. *En el grupo $(\mathbb{R}, +)$ el elemento neutro es el "0" y el elemento simétrico es el elemento opuesto $(-x)$. De la misma forma, en el grupo $(\mathbb{R}_0 = \mathbb{R} - \{0\}, \cdot)$ el elemento neutro es el "1" y el elemento simétrico es el inverso*[6] *$(1/x)$. Ambos son grupos conmutativos. Además:*

[6] Quizás con más precisión, a éste elemento se le denomina también *recíproco*.

1. *Los conjuntos $\mathbb{Z}, \mathbb{Q}, \mathbb{R}$ y $\mathbb{C}$ son grupos abelianos respecto a la suma ordinaria*

2. *Los conjuntos $\mathbb{Q}_0, \mathbb{R}_0$ y $\mathbb{C}_0$ son grupos abelianos respecto al producto (sin el "0")*

3. *El conjunto $\mathcal{M}_{m\times n}(\mathbb{R})$ de matrices con m filas y n columnas y coeficientes en $\mathbb{R}$ es un grupo conmutativo respecto a la suma matricial.*[7]

2.3. Anillos

Se analizarán a continuación los anillos, un tipo de estructuras con dos operaciones relacionadas entre sí. Estructuras algebraicas de este tipo son los conjuntos numéricos $\mathbb{Z}, \mathbb{Q}$ y $\mathbb{R}$, de forma que estas estructuras resultan relativamente familiares; esto, no obstante, puede resultar un inconveniente porque anima a generalizar las propiedades a las que se está acostumbrado al manejar números. Esto, como se verá, no es siempre acertado.

Definición 36. *Se llama **anillo**, y se denota por $(A, \oplus, \odot)$, a un conjunto A dotado de dos operaciones "$\oplus$" y "$\odot$" que verifica las propiedades siguientes:*

1. *$(A, \oplus)$ es un grupo abeliano. Su elemento neutro lo denotaremos como* 0.

2. $(x \odot y) \odot z = x \odot (y \odot z)\,,\ \forall\, x, y, z \in A$ *(propiedad asociativa)*

3. $\left.\begin{array}{l} x \odot (y \oplus z) = (x \odot y) \oplus (x \odot z) \\ (x \oplus y) \odot z = (x \odot z) \oplus (y \odot z) \end{array}\right\} \quad \forall x, y, z \in A$ *(prop. distributiva)*

Definición 37. *$(A, \oplus, \odot)$ se llamará **anillo unitario** si verifica:*

4. $\exists^{\bullet}\, \bar{e} \in A \;/\; x \odot \bar{e} = \bar{e} \odot x = x,\ \forall\, x \in A$

Definición 38. *$(A, \oplus, \odot)$ se llamará **anillo conmutativo** si verifica:*

5. $x \odot y = y \odot x,\ \forall\, x, y \in A$ *(propiedad conmutativa)*

Definición 39. *Un elemento x de un anillo unitario $(A, \oplus, \odot)$ se dice **inversible** si posee simétrico respecto de la segunda operación, "$\odot$", es decir existe $y \in A$ tal que*

$$x \odot y = y \odot x = \bar{e}$$

Ejemplo 27. *En el anillo $(\mathbb{Z}, +, \cdot)$ los únicos elementos inversibles son el 1 y el -1, de forma que $(-1) \cdot (-1) = 1 \cdot 1 = 1$ (el 1 es el elemento neutro para la operación "$\cdot$"). Esto choca directamente con lo que sucede en el anillo*

[7] Si $A = (a_{ij}), B = (B_{ij}) \in \mathcal{M}_{m\times n}(\mathbb{R})$ son dos matrices reales, la suma $A + B = (a_{ij} + b_{ij})$

$(\mathbb{R}, +, \cdot)$*, en el que el único elemento no inversible es el* 0 *(precisamente, el elemento neutro para la operación "+").*

Con respecto a los conjuntos numéricos, se cumple que $(\mathbb{Z}, +, \cdot)$ *es un anillo conmutativo unitario. Los conjuntos* $(\mathbb{Q}, +, \cdot)$, $(\mathbb{R}, +, \cdot)$ *y* $(\mathbb{C}, +, \cdot)$ *son anillos conmutativos unitarios, en los que todos los elementos, salvo el nulo, son inversibles.*

Otro ejemplo muy utilizado de anillo es el conjunto $\mathbb{Z}_m$ *con las operaciones:*

$$[a] + [b] = [a + b] \quad [a][b] = [ab]$$

siendo $a, b \in \{0, 1, \ldots, m-1\}$.

Por analogía con los conjuntos numéricos, es habitual denotar al elemento neutro, respecto de $\odot$, de un anillo unitario $(A, \oplus, \odot)$ como 1.

En determinados anillos es posible encontrar dos elementos no nulos que, al operarlos entre sí mediante la segunda operación –a la que habitualmente denominamos con el producto–, se obtiene el elemento neutro de la primera –el 0, habitualmente–. En otras palabras, es posible multiplicar dos elementos no nulos y que el resultado sea cero. A estos elementos se les conoce como *divisores de cero*:

Definición 40. *En un anillo* $(A, \oplus, \odot)$ *se dice que un elemento* $a \in A$ *no nulo es un* ***divisor de cero*** *si existe otro elemento no nulo* $b \in A$ *tal que* $a \odot b = 0$.

Un ejemplo claro de este tipo de comportamiento se puede observar en el anillo $(\mathcal{M}_n(\mathbb{R}), +, \cdot)$, donde $\mathcal{M}_n(\mathbb{R})$ es el conjunto de las matrices cuadradas de tamaño n con coeficientes en $\mathbb{R}$ y "+","·" son las operaciones suma y producto habituales entre matrices.[8] Si tomamos las matrices no nulas $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right)$ y $\left(\begin{smallmatrix} 0 & 0 \\ 0 & 1 \end{smallmatrix}\right)$, su producto es:

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

Un caso especial de anillos son aquellos en los que no existen divisores de cero, es decir:

$$\forall\, x, y \in A \quad \big(x \odot y = 0 \ \Leftrightarrow\ (x = 0 \ \vee\ y = 0)\big)$$

a estos anillos se les denomina *anillos íntegros* o *dominios de integridad.*

Ejemplo 28. *Los anillos* $(\mathbb{Z}, +, \cdot)$, $(\mathbb{Q}, +, \cdot)$, $(\mathbb{R}, +, \cdot)$ *y* $(\mathbb{C}, +, \cdot)$*, con las operaciones* $+$ *y* $\cdot$ *habituales en ellos, son dominios de integridad.*

---

[8] Si $A = (a_{ij}), B = (b_{ij}) \in \mathcal{M}_n(\mathbb{R})$, la matriz producto AB tiene como coeficiente ij el resultado de multiplicar la fila i-ésima de A por la columna j-ésima de B, es decir

$$(AB)_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj} = (a_{i1}, \ldots, a_{in}) \begin{pmatrix} b_{1j} \\ \vdots \\ b_{nj} \end{pmatrix}.$$

2.4. Cuerpos

Definición 41. *Se llama* ***cuerpo*** *a un anillo unitario* $(K, \oplus, \odot)$*, tal que* $(K - \{0\}, \odot)$ *es un grupo, es decir todo elemento* $x \in K$ *distinto de* 0 *es inversible respecto de* $\odot$*.*

Si el anillo $(K, \oplus, \odot)$ es conmutativo, se dice que el cuerpo K es conmutativo.

Ejemplo 2. *Los conjuntos* $\mathbb{Q}$*,* $\mathbb{R}$ *y* $\mathbb{C}$ *son cuerpos conmutativos respecto a las operaciones suma y producto ordinarias.*

El anillo $\mathbb{Z}_m$ *es un cuerpo si* m *es un número primo y son la base para la construcción de cualquier cuerpo finito cuyo cardinal es siempre de la forma* p^n*, siendo* p *un primo y* n *un número natural. Estos cuerpos se utilizan en Criptografía, Teoría de Códigos, etc.*

2.5. Espacios Vectoriales

Definición 42. *Se dice que un conjunto* V *tiene estructura de* ***espacio vectorial*** *sobre un cuerpo* K *si:*

- *En* V *hay definida una operación interna (suma) que confiere a* V *estructura de grupo abeliano*

- *En* V *hay definida una operación externa (producto)*

$$: K \times V \to V$$

que verifica las siguientes operaciones:

1. $\lambda\,(u + v) = \lambda u + \lambda v,\ \forall\, u, v \in V,\ \forall\, \lambda \in K$
2. $(\lambda + \mu)\, u = \lambda u + \mu u,\ \forall\, u \in V,\ \forall\, \lambda, \mu \in K$
3. $\lambda\,(\mu u) = (\lambda\mu)\, u,\ \forall\, u \in V,\ \forall\, \lambda, \mu \in K$
4. $1 \cdot u = u,\ \forall\, u \in V$

Los elementos del espacio vectorial reciben el nombre de *vectores*, y los elementos del cuerpo K *escalares*.

Ejemplo 29. *Son ejemplos de espacios vectoriales* $\mathbb{R}^n$*,* $\mathbb{C}^n$*,* $\mathbb{Z}_m^n$*, para cualquier natural* n*, (en general* K^n*, si* K *es un cuerpo) y* $\mathcal{M}_{m\times n}(\mathbb{R})$*.*

Sea V un K-espacio vectorial. Si $\emptyset \neq U \subseteq V$ es un subconjunto no vacío de V, se dice que U es un *subespacio vectorial de* V, si U es un espacio vectorial sobre K, considerando en U las mismas operaciones definidas en V.

Proposición 2. *Sea $\emptyset \neq U \subseteq V$ un subconjunto no vacío de V, son equivalentes:*

1. *U es un subespacio vectorial de V*
2. *Dados $\vec{u}, \vec{u'} \in U$ y $\alpha \in K$, se tiene que $\vec{u} + \vec{u'} \in U$ y $\alpha\vec{u} \in U$*
3. *Dados $\vec{u}, \vec{u'} \in U$ y $\alpha, \beta \in K$, se tiene que $\alpha\vec{u} + \beta\vec{u'} \in U$*

Sea V un K-espacio vectorial y sea $S = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_p\}$ un subconjunto cualquiera de vectores de V. Una *combinación lineal* de los vectores de S es un vector $\vec{v}$ que se escribe como

$$\vec{v} = \alpha_1\vec{v}_1 + \alpha_2\vec{v}_2 + \cdots \alpha_p\vec{v}_p = \sum_{i=1}^{p} \alpha_i\vec{v}_i.$$

donde $\alpha_i \in K$.

Ejemplos 2.
1. *$\vec{0}$ es combinación lineal de cualquier conjunto de vectores sin más que tomar $\alpha_i = 0$ para todo i.*
2. *En $\mathbb{R}^3$, se tiene que*

$$(1, 1, 0) = (-1)(2, 1, -1) + 1(3, 2, -1)$$

Proposición 3. *El conjunto $\{$combinaciones lineales de $S\}$ es un subespacio vectorial de V llamado* subespacio generado o engendrado *por S y se denota $< S >$*

Definición 43. *Sea $S = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_p\}$ un conjunto de vectores de V. Se dice que S es un conjunto* libre *o* un conjunto de vectores linealmente independientes *si, para toda combinación lineal de los vectores de S:*

$$\sum_{i=1}^{p} \alpha_i\vec{v}_i = \alpha_1\vec{v}_1 + \cdots + \alpha_p\vec{v}_p = \vec{0},$$

se tiene que $\alpha_i = 0$, para todo i. Equivalentemente, S es libre si, ningún vector de S es combinación lineal de los demás.

Análogamente, se define:

Definición 44. *Sea $S = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_p\}$ un conjunto de vectores de V. Se dice que S es un conjunto* ligado *o* un conjunto de vectores linealmente dependientes *si, existe una combinación lineal de los vectores de S:*

$$\sum_{i=1}^{p} \alpha_i\vec{v}_i = \alpha_1\vec{v}_1 + \cdots + \alpha_p\vec{v}_p = \vec{0},$$

donde algún $\alpha_i \neq 0$. Equivalentemente, S es ligado si, algún vector de S es combinación lineal de los demás.

Ejemplo 30. *En* $\mathbb{R}^3$*, los vectores* $\{(1,1,0),(2,1,-1),(3,2,-1)\}$ *forman un conjunto ligado.*

Sea V un K-espacio vectorial.

Definición 45. *Se dice que* V *es un espacio vectorial* de tipo finito *(o finitamente generado) si existe* $S \subseteq V$ *finito tal que* $V =< S >$*.* S *es un* sistema de generadores de V.

Definición 46. *Un conjunto de vectores* $B = \{\vec{v}_1, \dots, \vec{v}_p\}$ *de* V *es una* base *de* V *si:*

- B *es un sistema de generadores de* V *(*$V =< B >$*)*
- B *es un conjunto libre.*

Ejemplo 31. *En* K^n*, el conjunto*

$$\{(1,0,\dots,0),(0,1,\dots,0),\dots,(0,0,\dots,1)\}$$

es una base (llamada base canónica)

Teorema 1. *Sea* $B = \{\vec{v}_1, \dots, \vec{v}_p\} \subseteq V$*. Son equivalentes:*

1. B *es una base de* V*.*
2. *Cualquier vector de* V *se escribe de* manera única *como combinación lineal de los vectores de la base de* B*.*

Si $\vec{v} = \alpha_1\vec{v}_1 + \alpha_2\vec{v}_2 + \cdots + \alpha_p\vec{v}_p$, los escalares $\alpha_1, \alpha_2, \dots, \alpha_p$ se llaman *coordenadas del vector* $\vec{v}$ respecto de la base B.

Proposición 4. *Sea* $V \neq \{\vec{0}\}$ *un espacio vectorial de tipo finito.* V *siempre admite una base.*

Proposición 5. *Sean* B *y* B' *dos bases de* V*. Entonces* $|B| = |B'|$ *y a este cardinal común se le llama* dimensión *de* V*.*

Sea S un conjunto de vectores de un espacio vectorial V de tipo finito. Se denomina *rango* de S a $dim(< S >)$, luego el rango de S es el mayor número de vectores linealmente independientes que hay dentro de S.

Parte III

Tema 2. Sistemas de Ecuaciones Lineales

Tema 2

Sistemas de ecuaciones lineales

2.1. Dos interpretaciones gráficas de los sistemas lineales

Muchos problemas prácticos que aparecen en diversas disciplinas, tales como la biología, la química, la economía o la ingeniería, pueden ser frecuentemente reducidos a la resolución de ecuaciones lineales, concepto conocido de cursos anteriores. En los sistemas lineales, las ecuaciones sólo involucran la primera potencia de cada incógnita y éstas no van multiplicadas entre sí. En particular cada incógnita sólo aparece multiplicada por algún escalar.

El álgebra lineal surgió como resultado de los intentos de encontrar métodos sistemáticos para resolver tales sistemas. En otras palabras, el problema central del álgebra lineal es resolver sistemas de n ecuaciones lineales. Recordemos algunas nociones básicas.

Sea la combinación lineal de las variables $x_1, \dots x_n$ dada por la expresión

$$a_1\, x_1 + a_2\, x_2 + \dots a_n\, x_n\,.$$

Los números reales $a_1, \dots, a_n \in \mathbb{R}$ se denominan **coeficientes de la combinación**.

Una **ecuación lineal** en las incógnitas o variables $x_1, \dots x_n$ es una expresión de la forma

$$a_1\, x_1 + a_2\, x_2 + \dots a_n\, x_n = b\,,$$

donde $b \in \mathbb{R}$ se denomina **término constante**.

Una n-tupla $(s_1, \dots, s_n)$ es una **solución** de la ecuación lineal

$$a_1\, x_1 + a_2\, x_2 + \dots a_n\, x_n = b\,,$$

si al sustituir las variables $x_1, x_2, \dots, x_n$ por los números $s_1, \dots, s_n$ se obtiene una identidad.

Un *sistema de* m *ecuaciones lineales y* n *incógnitas*

$$\left.\begin{array}{r} a_{11}\, x_1 + a_{12}\, x_2 + \dots + a_{1n}\, x_n = b_1 \\ a_{21}\, x_1 + a_{22}\, x_2 + \dots + a_{2n}\, x_n = b_2 \\ \vdots \\ a_{m1}\, x_1 + a_{m2}\, x_2 + \dots + a_{mn}\, x_n = b_m \end{array}\right\}$$

tiene **solución** $(s_1, \dots, s_n)$ si dicha n-tupla es una solución de <u>todas</u> las ecuaciones del sistema.

La solución de un sistema lineal no existe necesariamente y, si lo hace, no es necesariamente única.

Si un sistema de ecuaciones tiene solución o soluciones se denomina **sistema compatible**; en caso contrario diremos que es un **sistema incompatible**.

El conjunto de todas las soluciones de un sistema de ecuaciones lineales dado se denomina **conjunto de soluciones**.

Cuando el sistema no tiene solución, el conjunto de soluciones se dice *vacío*, lo que representamos con el símbolo $\emptyset$ (que denota al conjunto vacío en teoría de conjuntos).

Si un sistema de ecuaciones lineales tiene solución única se dice que es **compatible determinado**; si tiene más de una solución se dice que es **compatible indeterminado**.

Finalmente recordemos que:

Dos sistemas son **equivalentes** si ambos tienen el mismo conjunto de soluciones.

A continuación veremos dos maneras distintas de *interpretar* los sistemas lineales.

Ejemplo

Sea el siguiente sistema de dos ecuaciones con dos incógnitas

$$\left.\begin{aligned} x_1 - x_2 &= 0 \\ x_1 + 2x_2 &= 4 \end{aligned}\right\}.$$

La primera de las interpretaciones geométricas es bien conocida: cada par de números reales (x_1, x_2) que verifiquen $x_1 - x_2 = 0$, e.g., $(0,0)$, $(-3,-3)$ ó $(\sqrt{2}, \sqrt{2})$, es solución de la primera ecuación. Si representamos tales puntos en el plano obtenemos una recta, como se ve en la figura 2.1. En general, si representamos cada ecuación (fila) por una recta en el plano, obtendremos que la intersección de todas ellas, si existe, es la solución del sistema. En particular, el par $\left(\frac{4}{3}, \frac{4}{3}\right)$ es la solución del sistema anterior.

Obsérvese que podemos usar este procedimiento gráfico para ***resolver*** sistemas con dos incógnitas.

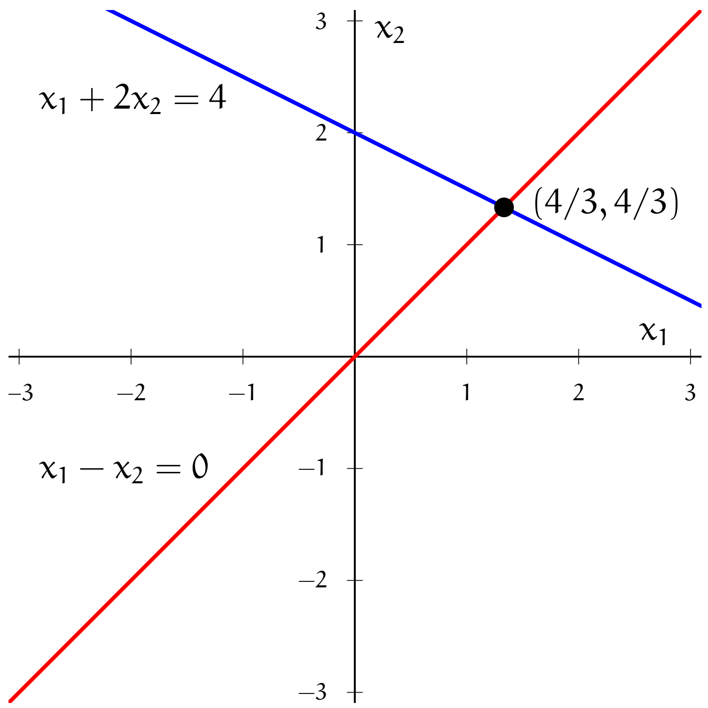


Figura 2.1: Representación gráfica de un sistema lineal por ***filas***.

De manera alternativa, podemos escribir el sistema usando notación vectorial de la siguiente manera: $x_1\,(1,1)^t + x_2\,(-1,2)^t = (0,4)^t$. Utilizando el concepto de combinación lineal, esta expresión nos dice que la $\boxed{\text{solución } (x_1, x_2)}$ del sistema lineal está dado por los $\boxed{\text{coeficientes}}$ (escalares x_1 y x_2) que hacen que la combinación lineal de los vectores columna $(1,1)^t$ y $(-1,2)^t$ sea igual al vector $(0,4)^t$.
Podemos observar en la figura 2.2 que la elección correcta para tales escalares es obviamente la dada por el punto intersección de la figura 2.1: $x_1 = x_2 = \dfrac{4}{3}$.

Nótese que esta representación $\boxed{\text{NO}}$ debe usarse como procedimiento sistemático para resolver sistemas (pues habría que hacerlo mediante prueba y error), pero es una manera de comprender el significado de la solución en términos de combinaciones lineales de ciertos vectores, concepto que será de importancia fundamental en futuros temas. La

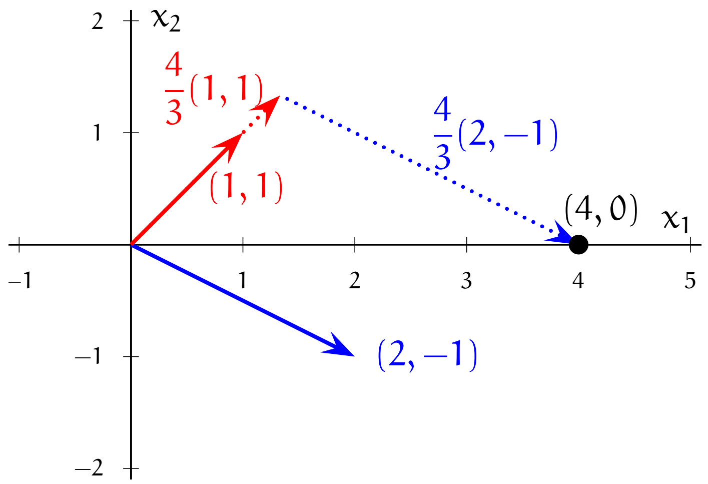


Figura 2.2: Representación gráfica de un sistema lineal por ***columnas***.

primera representación del sistema sistema lineal como líneas que se cortan (o no ...) en el plano (bidimensional) o como planos que se cortan en un espacio tridimensional, nos resulta más familiar que la que usa combinaciones lineales de vectores; sin embargo, para dimensiones mayores, es más sencillo "imaginar" una combinación lineal de, digamos, cuatro vectores en un espacio de dimensión 4, que "imaginar" la intersección de cuatro hiperplanos en dimensión 4.

Finalmente, recordemos que también podemos representar el sistema

$$\left.\begin{aligned} x_1 - x_2 &= 0 \\ x_1 + 2x_2 &= 4 \end{aligned}\right\}$$

como un problema matricial $A\,x = b$, escribiendo:

$$\overbrace{\begin{pmatrix} 1 & -1 \\ 1 & 2 \end{pmatrix}}^{A} \overbrace{\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}}^{x} = \overbrace{\begin{pmatrix} 0 \\ 4 \end{pmatrix}}^{b},$$

donde A es la matriz de coeficientes, x el vector (columna) de incógnitas y b el vector de constantes.

2.2. Resolución de sistemas lineales

Resolver un sistema consiste en encontrar su conjunto de soluciones.

Existen diferentes técnicas para encontrar el conjunto de soluciones de un sistema. En este tema nos centraremos en el método de Gauss, un procedimiento sistemático para resolver sistemas lineales, cuyo objetivo es obtener una secuencia de sistemas, cada uno de ellos equivalente al anterior, hasta obtener un sistema "escalonado, triangular superior" mediante la utilización de "operaciones elementales" de sus *filas*, para posteriormente obtener la solución mediante *sustitución hacia atrás*. Las operaciones entre filas permitidas están descritas en el siguiente teorema:

Teorema: Método de Gauss

Supongamos que un sistema lineal es modificado mediante alguna de las siguientes operaciones:

1) *Intercambio*: una ecuación intercambia su posición con cualquier otra.

2) *Re-escalado*: se multiplican los dos miembros de una ecuación por una constante no nula.

3) *Combinación de filas*: una ecuación es sustituida por la suma de ella y algún múltiplo de otra.

Entonces los dos sistemas tienen el mismo conjunto de soluciones.

Ejemplo

Consideremos otra vez el sistema:

$$\left.\begin{aligned} x_1 - x_2 &= 0 \\ x_1 + 2x_2 &= 4 \end{aligned}\right\},$$

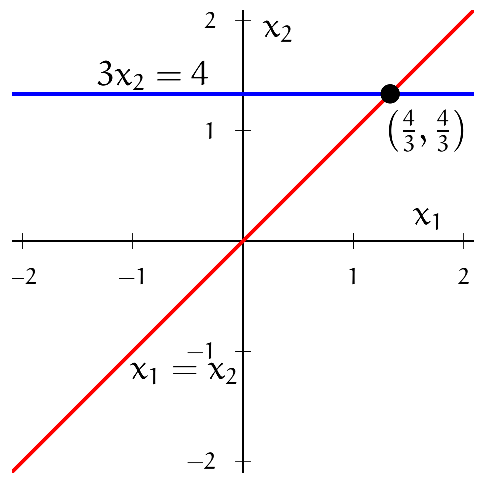


Figura 2.3: La aplicación de operaciones elementales de Gauss no cambia el conjunto de soluciones de un sistema lineal.

cuya solución está representada en la figura 2.1. Si sustituimos la segunda ecuación, R_2, por la combinación de filas $R_2 - R_1$ obtenemos el sistema:

$$\left.\begin{aligned} x_1 - x_2 &= 0 \\ 3x_2 &= 4 \end{aligned}\right\} .$$

La figura 2.3 ilustra que ambos sistemas son equivalentes (tienen la misma solución).

En cada fila de un sistema, la primera variable con coeficiente no nulo es la **variable pivote** de la fila. Un sistema está en **forma escalonada** si cada variable pivote se encuentra a la derecha de la variable pivote de la fila inmediatamente superior (exceptuando la variable pivote de la primera fila). Además, diremos que un sistema está en **forma escalonada reducida** si está escrito en forma escalonada, si los coeficientes de todas las variables pivote son iguales a 1 y si éste es el único coeficiente no nulo de la correspondiente ecuación.

Ejemplo

Si resolvemos el siguiente sistema mediante el método de Gauss, tenemos que realizar las siguientes operaciones:

$$\left.\begin{aligned} x_1+x_2&=0\\ 2x_1-x_2+3x_3&=3\\ x_1-2x_2-x_3&=3 \end{aligned}\right\} \xrightarrow[R_3-R_1]{R_2-2R_1} \left.\begin{aligned} x_1+x_2&=0\\ -3x_2+3x_3&=3\\ -3x_2-x_3&=3 \end{aligned}\right\} \xrightarrow[R_3-R_2]{} \left.\begin{aligned} x_1+x_2&=0\\ -3x_2+3x_3&=3\\ -4x_3&=0 \end{aligned}\right\}.$$

Observemos que este tercer sistema está en $\boxed{\text{forma escalonada}}$. Después de los pasos de eliminación, obtenemos de la tercera fila que $x_3=0$; sustituyendo en la segunda fila, resulta $x_2=-1$ y sustituyendo en la primera fila obtenemos $x_1=1$. Así, el conjunto de soluciones sólo consta de un elemento y lo podemos escribir como $\{(1,-1,0)\}$. También podemos obtener la $\boxed{\text{forma escalonada reducida}}$ del sistema con las siguientes operaciones:

$$\left.\begin{aligned} x_1+x_2&=0\\ -3x_2+3x_3&=3\\ -4x_3&=0 \end{aligned}\right\} \xrightarrow[-\frac{1}{4}R_3]{-\frac{1}{3}R_2} \left.\begin{aligned} x_1+x_2&=0\\ x_2-x_3&=-1\\ x_3&=0 \end{aligned}\right\} \xrightarrow{R2+R3} \left.\begin{aligned} x_1+x_2&=0\\ x_2&=-1\\ x_3&=0 \end{aligned}\right\}$$

$$\xrightarrow{R1-R2} \left.\begin{aligned} x_1&=1\\ x_2&=-1\\ x_3&=0 \end{aligned}\right\}.$$

que proporcionan de manera directa la solución del sistema (obviamente, la misma que hemos encontrado anteriormente). Este método se conoce como **método de Gauss-Jordan**.

Ejemplo

Consideremos el siguiente SEL (sistema de ecuaciones lineales) y las operaciones indicadas:

$$\left.\begin{aligned} x_1 + 3x_2 &= 1 \\ 2x_1 + x_2 &= -3 \\ 2x_1 + 2x_2 &= -2 \end{aligned}\right\} \xrightarrow[R_3-2R_1]{R_2-2R_1} \left.\begin{aligned} x_1 + 3x_2 &= 1 \\ -5x_2 &= -5 \\ -4x_2 &= -4 \end{aligned}\right\} \xrightarrow[R_3-\frac{4}{5}R_2]{} \left.\begin{aligned} x_1 + 3x_2 &= 1 \\ -5x_2 &= -5 \\ 0 &= 0 \end{aligned}\right\}.$$

Esto muestra que una de las ecuaciones es redundante. De la segunda fila obtenemos $x_2 = 1$ y sustituyendo en la primera ecuación resulta $x_1 = -2$, por lo que el conjunto de soluciones es $\{(-2, 1)\}$.

Ejemplo

Ahora consideremos el siguiente sistema, donde llevamos a cabo las siguientes operaciones:

$$\left.\begin{aligned} x_1 + 3x_2 &= 1 \\ 2x_1 + x_2 &= -3 \\ 2x_1 + 2x_2 &= 0 \end{aligned}\right\} \xrightarrow[R_3-2R_1]{R_2-2R_1} \left.\begin{aligned} x_1 + 3x_2 &= 1 \\ -5x_2 &= -5 \\ -4x_2 &= -2 \end{aligned}\right\}.$$

La tercera fila muestra que $x_2 = 1/2$ y la segunda que $x_2 = 1$, es decir, el sistema es incompatible o, en otras palabras, el conjunto de soluciones es vacío.

Así:

- Un sistema lineal con solución única tiene un conjunto de soluciones con un elemento.
- Un sistema lineal sin solución tiene un conjunto de soluciones vacío.

En estos casos, el conjunto de soluciones es fácil de describir. Los conjuntos de soluciones más complicados de expresar son aquellos que contienen "muchos" elementos.

Cuando esto ocurre, el número de posibles soluciones es infinito. La manera de identificar esta situación es la siguiente: *cuando un SEL tiene un número infinito de soluciones y lo escribimos en forma escalonada, entonces no todas las variables son variables pivote*.

Ejemplo

Consideremos el siguiente sistema:

$$\left.\begin{aligned} x_1 + x_2 &= 4 \\ 2x_1 + 2x_2 &= 8 \end{aligned}\right\} \xrightarrow{R_2 - 2R_1} \left.\begin{aligned} x_1 + x_2 &= 4 \\ 0 &= 0 \end{aligned}\right\}.$$

Esto muestra que cualquier par de números reales (x_1, x_2) que satisfaga la primera ecuación también satisface la segunda: hay un número infinito de soluciones del sistema.

El conjunto de soluciones puede describirse de la siguiente manera:

$$\begin{aligned} \{(x_1, x_2)\colon x_1 + x_2 = 4\,, \quad x_1, x_2 \in \mathbb{R}\} &= \{(x_1, x_2)\colon x_2 = 4 - x_1\,, \quad x_1 \in \mathbb{R}\} \\ &= \{(x_1, x_2) = (x_1, 4 - x_1)\,, \quad x_1 \in \mathbb{R}\} \\ &= \{(0, 4) + x_1(1, -1)\,, \quad x_1 \in \mathbb{R}\}. \end{aligned}$$

Se pueden obtener soluciones particulares asignando valores arbitrarios a x_1. Por ejemplo, los pares $(0, 4)$, $(1, 3)$, $(-2, 6)$ y $(\sqrt{3}, 4 - \sqrt{3})$ son soluciones particulares del sistema.

Llamamos **parámetro** a una variable que es utilizada para describir una familia de soluciones.

En resumen, para cualquier sistema de ecuaciones lineales existen exactamente tres posibilidades, que identificamos fácilmente si está escrito en forma escalonada:

1. No hay solución; esto ocurre cuando hay alguna inconsistencia entre las ecuaciones.

2. Hay solución única; esto ocurre cuando cada variable es una variable pivote.

3. Hay un número infinito de soluciones; esto ocurre cuando hay al menos una variable que no es pivote y, por tanto, existe al menos un parámetro.

2.2.1. Matrices equivalentes

Según hemos visto, las operaciones gaussianas producen *sistemas equivalentes*. También vimos anteriormente que los sistemas lineales pueden ser escritos de manera más compacta por medio de su expresión matricial, $A\,x = b$. Cuando resolvemos un sistema, esta notación es más adecuada, pues permite realizar exactamente las mismas operaciones que haríamos escribiendo el sistema completo, pero con una escritura más breve. Para ello, construimos lo que se conoce como la **matriz aumentada del sistema**, escribiendo como una sola matriz la matriz de coeficientes y el vector de constantes (separados por una barra vertical): $\left(A \,\middle|\, b\right)$.

Ejemplo

Consideremos una vez más el sistema:

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

Si realizamos operaciones elementales en la matriz aumentada del sistema, podemos obtener una forma escalonada (por filas) de dicha matriz como sigue:

$$\left(\begin{array}{rrr|r} 1 & 1 & 0 & 0 \\ 2 & -1 & 3 & 3 \\ 1 & -2 & -1 & 3 \end{array}\right) \xrightarrow[R_3-R_1]{R_2-2R_1} \left(\begin{array}{rrr|r} 1 & 1 & 0 & 0 \\ 0 & -3 & 3 & 3 \\ 0 & -3 & -1 & 3 \end{array}\right) \xrightarrow[R_3-R_2]{} \left(\begin{array}{rrr|r} 1 & 1 & 0 & 0 \\ 0 & -3 & 3 & 3 \\ 0 & 0 & -4 & 0 \end{array}\right)$$

y de nuevo resulta, por sustitución hacia atrás, el conjunto de soluciones $\{(1, -1, 0)\}$.

Ejemplo

Sea el sistema:

$$\left.\begin{aligned} x_1 + x_2 &= 0 \\ x_2 - x_3 &= 1 \\ x_3 &= 3 \end{aligned}\right\} .$$

La matriz aumentada asociada es

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

Usando notación matricial, escribimos el sistema de la forma $A\,x = b$, donde

$$\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix}, \quad x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, \quad b = \begin{pmatrix} 0 \\ 1 \\ 3 \end{pmatrix}.$$

Resolviendo hacia atrás determinamos el conjunto de soluciones: $\{(-4, 4, 3)\}$.

Ejemplo

Consideremos el sistema:

$$\left.\begin{aligned} x_1 &= 3 \\ x_1 + x_2 &= 1 \\ 2x_1 - x_2 + x_3 &= 3 \end{aligned}\right\} .$$

Si usamos notación matricial, describiremos dicho sistema de la forma $A\,x = b$, con

$$\begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 2 & -1 & 1 \end{pmatrix}, \quad x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, \quad b = \begin{pmatrix} 3 \\ 1 \\ 3 \end{pmatrix}.$$

Aunque la matriz aumentada

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

no está en forma escalonada, es triangular inferior y es evidente que podemos resolver el sistema mediante *sustitución hacia delante* (e.g., sustituyendo de arriba a abajo). El conjunto de soluciones es $\{(3, -2, -5)\}$.

Es muy sencillo observar que hacer operaciones gaussianas sobre los elementos de la correspondiente matriz ampliada de un SEL equivale a realizar ciertos productos entre matrices. En efecto:

1. El *intercambio* de las filas i y j de una matriz se obtiene multiplicándola (por la izquierda) por la matriz que resulta al intercambiar en la matriz identidad las filas i-ésima y j-ésima.

Ejemplo

Para intercambiar las filas 1 y 3 de A

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

la multiplicamos por la izquierda por la matriz

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

Así obtenemos

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

2. El *re-escalado* de las fila i de una matriz por un factor α se obtiene multiplicándola (por la izquierda) por la matriz que resulta al multiplicar la fila i-ésima de la matriz identidad por α.

Ejemplo

Para multiplicar la segunda fila de A

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

por el escalar 5, la multiplicamos por la izquierda por la matriz

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

Así obtenemos

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

3. La sustitución de la fila i por la *combinación* de la fila i y la fila j multiplicada por α se obtiene multiplicándo A (por la izquierda) por la matriz que resulta al sustituir

el elemento (i, j) de la matriz identidad por a.

Ejemplo

Dada la matriz A,

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

para restar a la fila primera de A su fila tercera multiplicada por 4, la multiplicamos por la izquierda por la matriz

$$\begin{pmatrix} 1 & 0 & -4 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Así obtenemos

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

Es evidente que cualquiera de las operaciones elementales anteriores sobre una matriz A produce una matriz B que es equivalente a A. El siguiente resultado es fundamental y nos facilitará muchos razonamientos:

Si las matrices A y B son equivalentes, entonces $\text{rg}(A) = \text{rg}(B)$.

2.2.2. Inversa de una matriz (mediante eliminación gaussiana)

Supongamos que queremos resolver un sistema lineal, escrito en forma matricial como $Ax = b$. Sabemos que si A es cuadrada ($n \times n$) y no singular ($\det(A) \neq 0$), podemos realizar la operación

$$\overbrace{A^{-1}\,A}^{I_n}\,x = A^{-1}\,b \quad \Rightarrow \quad x = A^{-1}\,b\,.$$

El método de Gauss-Jordan se puede usar para determinar fácilmente la inversa de una matriz no singular A como sigue.

Supongamos que el sistema que queremos resolver es $A\,x = b$, con $b = (1, 0, 0, \ldots, 0)^t$. Si multiplicamos $A^{-1}\,b$, observamos que el resultado de esta operación es la primera columna de la matriz A^{-1}. Si hacemos la misma consideración con el vector $b = (0, 1, 0, \ldots, 0)^t$, el producto $A^{-1}\,b$ proporciona la segunda columna de la inversa. Si repetimos este procedimiento hasta que multiplicamos $A^{-1}\,b$, con $b = (0, 0, \ldots, 0, 1)^t$ para obtener la última columna de A^{-1}, obtendremos la matriz inversa completa. Todos estos cálculos pueden ser realizados de manera eficiente, resolviendo todos los sistemas a la vez, si usamos eliminación gaussiana para obtener la forma escalonada reducida de la matriz

$$\left(A \,\middle|\, I_n\right) = \left(\begin{array}{ccc|ccc} a_{11} & \ldots & a_{1n} & 1 & \ldots & 0 \\ \vdots & & \vdots & \vdots & & \vdots \\ a_{n1} & \ldots & a_{nn} & 0 & \ldots & 1 \end{array}\right)$$

que es una matriz con n filas y $2n$ columnas, con A a la izquierda y la matriz identidad I_n de dimensión n a la derecha. Veámoslo con un ejemplo:

Ejemplo

Para calcular la inversa de la matriz

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

escribimos:

$$\left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \end{array}\right) \xrightarrow[R_3-R_1]{R_2-R_1} \left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & -1 & 0 & -1 & 1 & 0 \\ 0 & -1 & -1 & -1 & 0 & 1 \end{array}\right) \xrightarrow{-R_2}$$

$$\left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & -1 & 0 \\ 0 & -1 & -1 & -1 & 0 & 1 \end{array}\right) \xrightarrow[R_3+R_2]{R_1-R_2} \left(\begin{array}{ccc|ccc} 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & -1 & 0 \\ 0 & 0 & -1 & 0 & -1 & 1 \end{array}\right) \xrightarrow{-R_3}$$

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & 0 & 1 & -1 \end{array}\right) \xrightarrow{R_1-R_3} \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & 0 & 1 & -1 \end{array}\right).$$

Entonces, la inversa de A es la matriz de la derecha:

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

2.3. Teoría de los sistemas lineales

Como hemos visto, podemos interpretar el sistema de ecuaciones $A\,x = b$ de dos maneras. Si nos centramos en las filas, s es una solución del sistema si está situado sobre la intersección de todos los hiperplanos definidos por el sistema de ecuaciones. Si nos centramos en las columnas, la solución s es un conjunto de escalares que proporciona una representación del vector b como una combinación lineal de las $\boxed{\text{columnas de la matriz } A}$. Observemos que esto está estrechamente relacionado con el *espacio columna* de la matriz A

(que desempeñará un papel muy importante en futuros capítulos). Cuando tal solución existe, decimos que el sistema es *compatible*; en caso contrario, decimos que es *incompatible*. En particular, el sistema lineal $A\,x = b$ es compatible si y sólo si $b \in \mathcal{C}(A)$. Como ya sabemos, es fácil reconocer los sistemas incompatibles a partir de cualquiera de sus formas escalonadas: un sistema es incompatible cuando hay una ecuación escrita en la forma $0x_1 + \cdots + 0x_n = c$, para algún escalar c no nulo. En tal situación, ***es imposible expresar*** b ***como combinación lineal de las columnas de*** A.

Una cuestión más difícil es la siguiente: ¿bajo qué condiciones un vector $b \in \mathbb{K}^n$ puede ser representado como una combinación lineal de las columnas de A? Para responder a esta pregunta, se usa el *rango* r de la matriz A, que será igual al número de filas no nulas en la correspondiente matriz escalonada o, en otras palabras, el número de variables pivote del sistema. Aunque el siguiente resultado es bien conocido de cursos anteriores, recordamos que:

Teorema de Rouché-Frobenius

El sistema lineal $A\,x = b$ es compatible si y sólo si el rango de la matriz A es igual al rango de la matriz ampliada $\left(A \mid b\right)$. El sistema lineal $A\,x = b$ es compatible determinado si y sólo si el rango de la matriz A (igual al rango de la matriz ampliada) coincide con el número de incógnitas.

NOTA. En temas posteriores veremos una interpretación geométrica de este hecho, basándonos en el espacio columna de A.

2.3.1. Unicidad de soluciones

A continuación nos centramos en la pregunta de cuántas soluciones tiene un sistema de ecuaciones compatible dado. Como ya hemos mencionado, un sistema compatible tiene solución única si y sólo si el rango de A es igual al número de incógnitas. Además, como veremos más adelante, las soluciones de un sistema lineal compatible de la forma $Ax = b$ están íntimamente relacionados con las soluciones del sistema $Ax = 0$.

Diremos que una ecuación lineal es **homogénea** si su constante es igual a 0:

$$a_1 x_1 + \cdots + a_n x_n = 0 .$$

Si todas las ecuaciones de un sistema lineal son homogéneas, éste se denomina **sistema homogéneo**.

Claramente la solución dada por $x_1 = \cdots = x_n = 0$ es solución de dicho sistema y se denomina solución **trivial**. Cualquier solución en la que al menos una variable tiene valor no nulo es denominada **solución no trivial**.

Dado un sistema lineal, siempre podremos asociarle un sistema homogéneo haciendo cero las constantes. Así se verifica el siguiente resultado:

Proposición

Si un sistema de ecuaciones con n incógnitas $Ax = b$ es compatible, tiene solución única si y sólo si el sistema homogéneo asociado $Ax = 0$ sólo tiene solución trivial. Esto ocurre exactamente cuando $\mathrm{rg}(A) = n$.

Estos resultados también pueden relacionarse de la siguiente manera:

Proposición

Sea A una matriz $n \times n$. Los siguientes resultados son equivalentes:

1. A es no singular (i.e., $\det(A) \neq 0$).
2. $Ax = 0$ sólo tiene la solución trivial.
3. Para cada $b \in \mathbb{R}^n$, la ecuación $Ax = b$ tiene solución única.

Ejemplo

Consideremos el sistema

$$\left.\begin{aligned} x_1 + x_2 + x_3 &= 0 \\ x_1 + 2x_2 - x_3 &= 1 \\ 2x_1 + x_2 + 4x_3 &= -1 \end{aligned}\right\} .$$

Vamos a $\boxed{\text{ILUSTRAR}}$ la equivalencia de los tres resultados anteriores.

La matriz aumentada puede escribirse de la forma

$$\left(\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 1 & 2 & -1 & 1 \\ 2 & 1 & 4 & -1 \end{array}\right) \xrightarrow[R_3-2R_1]{R_2-R_1} \left(\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 0 & 1 & -2 & 1 \\ 0 & -1 & 2 & -1 \end{array}\right) \xrightarrow{R_3+R_1} \left(\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 0 & 1 & -2 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right).$$

Esto muestra que el rango de A y el de la matriz aumentada es 2 (inferior al número de incógnitas); por tanto, el sistema es compatible indeterminado.

Por otro lado, el sistema homogéneo asociado está dado por

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

y su conjunto de soluciones se describe como

$$\{(-3x_3, 2x_3, x_3) \colon x_3 \in \mathbb{R}\},$$

es decir, tiene un número infinito de soluciones. Finalmente, es inmediato comprobar que $\det(A) = 0$.

Además, los sistemas lineales homogéneos verifican las siguientes propiedades:

Sean s_1 y s_2 dos soluciones del sistema lineal homogéneo $A\,x = 0$ y sean $\alpha_1, \alpha_2 \in \mathbb{R}$ dos escalares arbitrarios; entonces $s_3 = \alpha_1\, s_1 + \alpha_2\, s_2$ es también solución del sistema homogéneo.

NOTA. Como veremos en el próximo capítulo, estas propiedades convierten al conjunto de soluciones de un sistema lineal homogéneo en lo que denominaremos un *espacio vectorial* (o más precisamente, en un *subespacio vectorial de* $\mathbb{R}^n$).

Finalmente, daremos el siguiente resultado para sistemas lineales no homogéneos:

Teorema

Sea el sistema de ecuaciones lineales compatible $A\,x = b$ donde A es una matriz de dimensión $m \times n$. Consideremos el sistema lineal homogéneo asociado $A\,x = 0$. Entonces cualquier solución del sistema homogéneo se puede escribir como una soluciión del sistema homogéneo más una una solución particular del sistema no homogéneo. En otras palabras, si $S_h = \{s_h \in \mathbb{R}^n \colon A\, s_h = 0\}$ es el conjunto de soluciones del sistema homogéneo y s_p es UNA SOLUCIÓN CUALQUIERA del sistema no homogéneo, entonces el conjunto de soluciones S del sistema no homogéneo se describe de la forma

$$S = \{s_p + s_h \colon s_h \in S_h\}.$$

NOTA. El conjunto de soluciones de un sistema lineal no homogéneo tiene la estructura de un *espacio afín*. Esto quiere decir que dadas dos soluciones cualquiera del sistema no homogéneo s_1 y s_2, su diferencia $s_1 - s_2$ es un elemento de un espacio vectorial (en este caso, es el definido por las soluciones del correspondiente sistema lineal homogéneo).

Ejemplo

Sea el sistema no homogéneo

$$\left.\begin{aligned} 2x_1 - x_2 + x_3 &= 0 \\ x_1 + 4x_2 + x_3 &= 3 \\ -x_1 + 5x_2 &= 3 \end{aligned}\right\} .$$

Si consideramos el sistema homogéneo asociado, comprobamos fácilmente que su conjunto de soluciones viene dado por:

$$S_h = \left\{ x_3 \left(-\frac{5}{9}, -\frac{1}{9}, 1 \right) : x_3 \in \mathbb{R} \right\} ,$$

que como vemos depende de un único parámetro $x_3 \in \mathbb{R}$.

Consideremos a continuación las ternas $(5, 1, -9)$, $(2, 1, -3)$ y $(-3, 0, 6)$. Es inmediato verificar que la primera de ellas NO es solución del sistema, pues al sustituir estos valores en las ecuaciones, las dos últimas no dan lugar a una identidad. En cambio, la segunda de ellas sí lo es. Según el teorema anterior, conocer esta segunda terna (o cualquier otra que sea solución) es suficiente para escribir el conjunto de soluciones:

$$S = \left\{ \left(2 - \frac{5}{9} s_3, 1 - \frac{1}{9} s_3, s_3 - 3 \right) : s_3 \in \mathbb{R} \right\} .$$

Finalmente, podemos comprobar que la tercera terna también es solución del sistema no homogéneo, por lo que de nuevo, por el teorema anterior, podemos escribir el conjunto de soluciones de la forma

$$S = \left\{ \left(-3 - \frac{5}{9} s_3, -\frac{1}{9} s_3, 6 + s_3 \right) : s_3 \in \mathbb{R} \right\} .$$

Obsérvese que aunque ambas descripciones del conjunto S son diferentes, corresponden al mismo conjunto. Por ejemplo, en el primer caso, si hacemos $s_3 = 0$ obtenemos la solución particular $(2, 1, -3)$. Esta misma solución resulta de la segunda descripción haciendo $s_3 = -9$. Se deja como ejercicio encontrar los valores de s_3 en la primera y en la segunda descripción que proporcionan la solución particular $(-3, 0, 6)$.

Parte IV

Tema 3. Espacios Vectoriales

Tema 3

Espacios vectoriales

3.1. Definición y propiedades

En este capítulo presentamos la definición formal de *espacio vectorial*. Los espacios vectoriales más elementales son los conocidos espacios $\mathbb{R}^n$, $n = 1, 2, \ldots$, que ilustran claramente el concepto. No obstante ya estamos familiarizados con otros conjuntos que, como veremos, también cumplen las condiciones que caracterizan a los espacios vectoriales.

Comenzamos considerando el plano $\mathbb{R}^2$. Un vector no nulo $v = (v_1, v_2)^t \in \mathbb{R}^2$ puede ser representado geométricamente en el plano mediante un *segmento dirigido* que va de $(0, 0)^t$ a $(v_1, v_2)^t$.

Esta representación nos ayuda a visualizar cómo funciona la bien conocida operación de *suma o adición*:

$$\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} + \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} v_1 + u_1 \\ v_2 + u_2 \end{pmatrix},$$

y la multiplicación por *escalares* (números reales, en este caso):

$$\alpha \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} \alpha v_1 \\ \alpha v_2 \end{pmatrix}.$$

La figura 3.1 ilustra ambas operaciones.

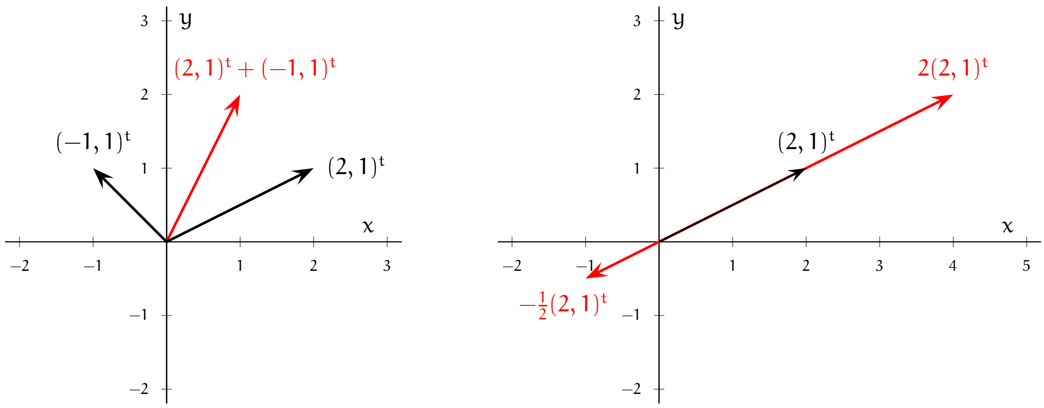


Figura 3.1: Suma de vectores y multiplicación por escalares.

En general, la suma y la multiplicación por escalares en $\mathbb{R}^n$ se definen respectivamente mediante

$$\begin{aligned} (v_1, v_2, \dots, v_n)^t + (u_1, u_2, \dots, u_n)^t &= (v_1 + u_1, v_2 + u_2, \dots, v_n + u_n)^t \\ \alpha \cdot (v_1, v_2, \dots, v_n)^t &= (\alpha \cdot v_1, \alpha \cdot v_2, \dots, \alpha \cdot v_n)^t \end{aligned}$$

donde el escalar α puede ser cualquier número real.

Comenzamos describiendo la *estructura* que poseen los escalares. Dicha estructura se denomina **cuerpo**.

Un **cuerpo** es un conjunto no vacío $\mathbb{K}$ en el que se definen dos operaciones: la *suma o adición*, denotada por $+$ y la *multiplicación o producto*, denotada por $\cdot$, y tales que, para todo $a, b, c \in \mathbb{K}$, se verifican los siguientes axiomas:

- Grupo aditivo conmutativo
  - $a + b \in \mathbb{K}$, [clausura].
  - $a + b = b + a$, [conmutatividad].
  - $(a + b) + c = a + (b + c)$, [asociatividad].

- $\exists\, 0 \in \mathbb{K}$ tal que $a + 0 = a$, [elemento neutro].
- $\forall a\,,\ \exists(-a) \in \mathbb{K}$ tal que $a + (-a) = 0$, [elemento inverso].

- Grupo multiplicativo conmutativo
  - $a \cdot b \in \mathbb{K}$, [clausura].
  - $a \cdot b = b \cdot a$, [conmutatividad].
  - $(a \cdot b) \cdot c = a \cdot (b \cdot c)$, [asociatividad].
  - $\exists\, 1 \in \mathbb{K}$ tal que $a \cdot 1 = a$, [elemento neutro].
  - $\forall a \neq 0\,,\ \exists a^{-1} \in \mathbb{K}$ tal que $a \cdot (a^{-1}) = 1$, [elemento inverso].
- Anillo
  - $a \cdot (b + c) = a \cdot b + a \cdot c$, [distributiva].

Obsérvese que consideramos que los elementos neutros son distintos: $0 \neq 1$.

NOTA: Los elementos inversos respecto de la suma y el producto son únicos en un cuerpo; por ello, les denotamos por $-a$ y a^{-1} respectivamente.

Ejemplo

Si examinamos los distintos conjuntos de números, es trivial observar cuáles de las propiedades anteriores, de sobra conocidas, se verifican en cada uno de ellos:

- Los conjuntos $\mathbb{N}$ y $\mathbb{Z}$ no son cuerpos (ya que carecen de inversos multiplicativos).
- Los conjuntos $\mathbb{Q}$, $\mathbb{R}$ y $\mathbb{C}$ son cuerpos.

NOTA: En lo que resta de curso, nos centraremos en el cuerpo de los reales $\mathbb{R}$ (salvo que se indique lo contrario).

Ejemplo

Sea el conjunto $A = \{0, 1\}$.

Vamos a $\boxed{\text{definir}}$ las siguientes operaciones entre los elementos del conjunto A.

1. **Suma**: $0 + 0 = 0;\ 0 + 1 = 1;\ 1 + 0 = 1$ y $1 + 1 = 0$.

   Es inmediato verificar que la operación así definida verifica las propiedades de grupo conmutativo. Por ejemplo:

   - Clausura: el resultado de sumar dos elementos cualesquiera de A es un elemento de A.
   - Conmutatividad: el único caso no trivial es $0 + 1 = 1 + 0 = 1$.
   - Elemento neutro: obviamente, el elemento neutro es el 0.

2. **Producto**: $0 \cdot 0 = 0 \cdot 1 = 1 \cdot 0 = 0$ y $1 \cdot 1 = 1$.

   De la misma manera, esta operación verifica las propiedades de grupo conmutativo. Veamos algunas:

   - Clausura: el resultado de multiplicar dos elementos cualesquiera de A es un elemento de A.
   - Conmutatividad: el único caso no trivial es $0 \cdot 1 = 1 \cdot 0 = 0$.
   - Elemento neutro: éste es el 1.

Puesto que además se cumple la propiedad distributiva, el conjunto A tiene estructura de cuerpo. Habitualmente se utiliza la notación $\mathbb{Z}_2$ para referirse al conjunto A. Intuitivamente, el elemento 0 representa a los enteros pares, mientras que el elemento 1 representa a los enteros impares, con lo que las definiciones dadas para la suma y el producto cobran sentido. Este ejemplo se generalizará en la asignatura de Matemática Discreta.

Una vez definida la estructura de cuerpo, pasamos a definir la estructura de *espacio vectorial*.

Dado un cuerpo $\mathbb{K}$, que denominaremos *conjunto de escalares*, consideremos un conjunto V no vacío en el que se definen las operaciones de suma y multiplicación por escalares. Si ambas operaciones satisfacen la propiedad de clausura, entonces diremos que el conjunto V con dichas operaciones es un **espacio vectorial** si se cumplen las siguientes propiedades para todo $u, v, w \in V$ y para todo $\alpha, \beta \in \mathbb{K}$:

1. $u + v = v + u$, [conmutatividad de la suma].
2. $(u + v) + w = u + (v + w)$, [asociatividad de la suma].
3. $\exists\, 0 \in V$ tal que $u + 0 = u$, [elemento neutro para la suma].
4. $\exists\, (-u) \in V$ tal que $u + (-u) = 0$, [elemento inverso para la suma].
5. $\alpha\,(u + v) = \alpha\, u + \alpha\, v$, [distributiva respecto a la suma de vectores].
6. $(\alpha + \beta)\, u = \alpha\, u + \beta\, u$, [distributiva respecto a la suma de escalares].
7. $(\alpha \cdot \beta)\, u = \alpha(\beta\, u)$, [asociatividad de la multiplicación por escalares].
8. $1 \cdot u = u$, [identidad].

Los elementos de un espacio vectorial V definido de la forma anterior se denominan **vectores**. Este término engloba mucho más que lo que hasta ahora hemos denominado vectores fila o columna, como un caso particular de matrices. Como veremos, los elementos de V pueden ser números, n-tuplas de números, matrices, polinomios, funciones, etc., ya que la definición de las operaciones de suma y multiplicación por escalares es lo que nos permitirá referirnos a ellos como “vectores”. En estas notas los vectores se denotarán por letras normales (p.e., a, v, etc.), aunque en otros libros se usan notaciones distintas (p.e., $\boldsymbol{a}$, $\vec{a}$, etc.).

Ejemplo

Es inmediato verificar que $\mathbb{R}^2$, con la suma ordinaria y la multiplicación por los escalares de $\mathbb{R}$, es un espacio vectorial, ya que se verifican todas las propiedades de la definición anterior.

Ejemplo

Según vimos en el tema 1, el conjunto $\mathbb{K}^{m\times n}$ de todas las matrices de dimensión $m\times n$ sobre el cuerpo $\mathbb{K}$, con la suma de matrices y la multiplicación por los escalares de $\mathbb{K}$, es un espacio vectorial.

Ejemplo

Sea el sistema de ecuaciones lineales homogéneo escrito en forma matricial como $A\,x = b$, con A de dimensión $m \times n$ y elementos reales. Si el sistema es compatible, su conjunto de soluciones está formado por n-tuplas (de números reales). Podemos concluir que el conjunto de soluciones de dicho sistema de ecuaciones lineales homogéneo es un espacio vectorial sin más que tener en cuenta las siguientes dos observaciones:

- La suma y la multiplicación por escalares ($\mathbb{R}$) en $\mathbb{R}^n$ verifica los ocho axiomas de la definición de espacio vectorial.

- Como vimos en el tema 2, CUALQUIER COMBINACIÓN LINEAL DE DOS SOLUCIONES DE DICHO SISTEMA CON COEFICIENTES REALES ES TAMBIÉN SOLUCIÓN DEL SISTEMA [propiedades de clausura].

Ejemplo

Las soluciones de los sistemas de ecuaciones lineales no homogéneos $A\,v = b \neq 0$ **NO** forman un espacio vectorial, ya que no contienen el vector cero.

Ejemplo

El conjunto $\mathbb{P}_n$ formado por todos los polinomios de coeficientes reales de grado menor o igual que n

$$\mathbb{P}_n = \left\{ p\colon \mathbb{R} \to \mathbb{R} \;\middle|\; p(x) = \sum_{a=0}^{n} \alpha_a\, x^a\,, \quad \alpha_a \in \mathbb{R} \right\}$$

es un espacio vectorial, pues la suma de polinomios y la multiplicación por escalares (reales) obviamente verifican todas las propiedades requeridas.

El siguiente teorema establece algunas propiedades fundamentales de los espacios vectoriales.

Teorema

Si V es un espacio vectorial sobre $\mathbb{K}$, entonces para todo $u \in V$ y $\alpha \in \mathbb{K}$, se cumple que:

i) $0\,u = 0$.

ii) $\alpha\,0 = 0$.

iii) $u + v = 0$ implica que $v = -u$ (i.e., el inverso respecto de la suma de u es único).

iv) $(-1)\,u = -u$.

NOTA: a pesar de que usaremos la misma notación, obsérvese la distinta naturaleza de los términos de la expresión: $\underbrace{0}_{\text{escalar}}\, u = \underbrace{0}_{\text{vector}}$.

3.2. Subespacios vectoriales

Dado un espacio vectorial V, a menudo es posible formar otro espacio vectorial tomando un subconjunto (propio) S de V que hereda las operaciones de V. Puesto que V es un espacio vectorial, las operaciones de suma y multiplicación por escalares siempre producirán un vector de V. Para que el nuevo subconjunto $S \subseteq V$ sea también un espacio vectorial, el conjunto S debe ser cerrado con respecto a dichas operaciones, i.e., la suma de dos elementos de S debe ser un elemento de S y el producto de un escalar y un elemento de S debe ser un elemento de S. Si esto ocurre, dado que el resto de propiedades que definen un espacio vectorial se satisfarán automáticamente en S, tendremos que:

Un subconjunto no vacío S del espacio vectorial V es un **subespacio** de V si satisface las propiedades de clausura $u + v \in S$ y $\alpha u \in S$ para todo $u, v \in S$ y todo $\alpha \in \mathbb{K}$.

En otras palabras, S es un subespacio de V si es cerrado bajo las operaciones de V. Obsérvese que esta definición implica que $0 \in S$.

Ejemplo

El subconjunto de $\mathbb{R}^2$ dado por

$$S = \left\{(x, 1)^t \colon x \in \mathbb{R}\right\}$$

NO verifica las propiedades de clausura:

$$(1, 1)^t + (3, 1)^t = (4, 2)^t \notin S, \qquad 3 \cdot (1, 1)^t = (3, 3)^t \notin S\,,$$

por lo que no es un subespacio vectorial de V.

Ejemplo

El subconjunto de $\mathbb{R}^2$ dado por

$$S = \{(x_1, x_2)^t \colon x_2 = 2x_1\}$$

verifica las propiedades de clausura:

$$\begin{aligned} (x, 2x)^t + (y, 2y)^t &= (x+y, 2x+2y)^t = (x+y, 2(x+y))^t \in S\,, \\ \alpha \cdot (x, 2x)^t &= (\alpha x, \alpha 2x)^t = (\alpha x, 2(\alpha x))^t \in S\,, \end{aligned}$$

por lo que S es subespacio de $\mathbb{R}^2$. En efecto, se ve fácilmente que S con las operaciones de $\mathbb{R}^2$ es un espacio vectorial.

Los siguientes ejemplos establecen importantes resultados teóricos:

Ejemplo

Dada una matriz $A \in \mathbb{K}^{m \times n}$, su espacio nulo $N(A)$ es un subespacio de $\mathbb{K}^n$.

Demostración. Claramente el espacio nulo es no vacío, puesto que el vector $v = 0$ de $\mathbb{K}^n$ es solución de $A\,v = 0$.

Si u e v son soluciones, entonces $A(u+v) = Au + Av = 0 + 0 = 0$, por lo que el espacio nulo es cerrado bajo la suma.

Si α es un escalar, se tiene que $A(\alpha v) = \alpha(Av) = \alpha 0 = 0$. Por ello, el espacio nulo es cerrado bajo la multiplicación por escalares, y en consecuencia, es un subespacio. □

Obviamente $N(A^t)$ es subespacio de $\mathbb{K}^m$.

Ejemplo

Dada una matriz $A \in \mathbb{K}^{m \times n}$, su espacio columna $\mathcal{C}(A)$ es un subespacio de $\mathbb{K}^m$.

Demostración. Claramente el conjunto es no vacío, pues contiene a todas las columnas de

A; también, si consideramos la combinación lineal:

$$0 \cdot A_1 + 0 \cdot A_2 + \cdots + 0 \cdot A_n = 0\,,$$

el cero pertenece a $\mathcal{C}(A)$.

Además es cerrado bajo la suma y la multiplicación por escalares:

$$\begin{aligned} (\alpha_1 A_1 + \cdots + \alpha_n A_n) + (\beta_1 A_1 + \cdots + \beta_n A_n) &= (\alpha_1 + \beta_1)A_1 + \cdots + (\alpha_n + \beta_n)A_n\,, \\ k(\alpha_1 A_1 + \cdots + \alpha_n A_n) &= k\alpha_1 A_1 + \cdots + k\alpha_n A_n\,, \end{aligned}$$

que son combinaciones lineales de las n columnas de la matriz. Por tanto $\mathcal{C}(A)$ es un subespacio. □

Obviamente $\mathcal{C}(A^t)$ es subespacio de $\mathbb{K}^n$.

3.3. Operaciones entre subespacios

Consideremos un espacio vectorial V y dos subespacios suyos S_1 y S_2. Además de la intersección y la unión de conjuntos arbitrarios, es posible definir entre S_1 y S_2 la suma de ellos, de la siguiente manera:

Dados los subespacios S_1 y S_2 del espacio vectorial V, definimos su **suma** mediante:

$$S_1 + S_2 = \{s_1 + s_2 \colon s_1 \in S_1, s_2 \in S_2\}\ .$$

Si la intersección entre S_1 y S_2 es $\{0\}$, entonces la suma se denomina **suma directa** y la representamos mediante $S_1 \oplus S_2$.

El siguiente resultado es fácil de demostrar:

Proposición

Dados los subespacios S_1 y S_2 de V (sobre $\mathbb{K}$), se tiene que:

1. $S_1 \cap S_2$ es un subespacio vectorial.

2. $S_1 \cup S_2$ no es en general un subespacio vectorial.

3. $S_1 + S_2$ es un subespacio vectorial.

Demostración.

1. Sean $s_1, s_2 \in S_1 \cap S_2$. Tendremos que $s_1, s_2 \in S_1$ y también que $s_1, s_2 \in S_2$. De la primera condición se deduce que, puesto que S_1 es un espacio vectorial, $s_1 + s_2 \in S_1$; de la segunda condición resulta que $s_1 + s_2 \in S_2$, al ser S_2 un espacio vectorial. En consecuencia:
$$s_1 + s_2 \in S_1 \cap S_2 \,.$$
Por otro lado, para todo $\alpha \in \mathbb{K}$, tendremos que $\alpha s_1 \in S_1$ y $\alpha s_1 \in S_2$. Por consiguiente:
$$\alpha s_1 \in S_1 \cap S_2$$
y concluimos que $S_1 \cap S_2$ es un subespacio vectorial de V.

2. Para probar que $S_1 \cup S_2$ no es en general un espacio vectorial consideramos por ejemplo $V = \mathbb{R}^2$ y los subconjuntos $S_1 = \{(v_1, 0)^t : v_1 \in \mathbb{R}\}$ y $S_2 = \{(0, v_2)^t : v_2 \in \mathbb{R}\}$. Es fácil ver que S_1 y S_2 son subespacios de $\mathbb{R}^2$. Si tomamos los vectores $(1, 0)^t, (0, 1)^t \in S_1 \cup S_2$, su suma:
$$(1, 0)^t + (0, 1)^t = (1, 1)^t$$
no pertenece ni a S_1 ni a S_2, por lo que tampoco pertenece a su unión y, por tanto, la suma no verifica la propiedad de clausura.

3. Por último, consideremos los vectores $s_1, s_2 \in S_1 + S_2$. Tendremos que $s_1 = v_1 + v_2$, con $v_1 \in S_1$ y $v_2 \in S_2$ y también que $s_2 = w_1 + w_2$, con $w_1 \in S_1$ y $w_2 \in S_2$.

   Por un lado, podemos escribir la suma como sigue:

$$s_1 + s_2 = (v_1 + v_2) + (w_1 + w_2) = \underbrace{(v_1 + w_1)}_{\in S_1} + \underbrace{(v_2 + w_2)}_{\in S_2} \in S_1 + S_2$$

   y por otro, tenemos que:

$$\alpha s_1 = \alpha(v_1 + v_2) = \underbrace{\alpha v_1}_{\in S_1} + \underbrace{\alpha v_2}_{\in S_1} \in S_1 + S_2\,.$$

   En consecuencia, $S_1 + S_2$ es subespacio vectorial. □

Ejemplo

Consideremos el espacio vectorial $\mathbb{R}^3$, y sus subespacios: $S_1 = \{(v_1, v_2, v_3)^t \colon v_1 + v_2 + v_3 = 0\}$ y $S_2 = \{(v_1, v_2, v_3)^t \colon v_1 - v_3 = 0\}$. Representados geométricamente S_1 y S_2 son dos planos que pasan por el origen. La intersección de ambos espacios se calcula resolviendo el sistema:

$$\left.\begin{aligned} v_1 + v_2 + v_3 &= 0 \\ v_1 - v_3 &= 0 \end{aligned}\right\}\,,$$

cuya solución es $v_1 = v_3$, $v_2 = -2v_3$ (una recta). Por tanto, podemos escribir:

$$S_1 \cap S_2 = \left\{(v_3, 2v_3, v_3)^t \colon v_3 \in \mathbb{R}\right\}\,.$$

Como esta intersección contiene más elementos que el vector cero, el espacio suma $S_1 + S_2$ no será suma directa. Para calcular dicho espacio suma, observamos que los elementos de S_1 son vectores $(v_1, v_2, v_3)^t$ cuyas coordenadas cumplen que $v_1 = -v_2 - v3$; esto es lo mismo que escribir:

$$(v_1, v_2, v_3)^t = (-v_2 - v_3, v_2, v_3)^t = v_2(-1, 1, 0)^t + v_3(-1, 0, 1)^t\,.$$

También, un elemento de S_2 podrá ser descrito mediante:

$$(v'_1, v'_2, v'_3)^t = (v'_3, v'_2, v'_3)^t = v'_2(0,1,0)^t + v'_3(1,0,1)^t \,.$$

Entonces, calculamos el espacio suma mediante:

$$\begin{aligned} S_1 + S_2 &= \{s_1 + s_2 \colon s_1 \in S_1, s_2 \in S_2\} \\ &= \left\{v_2(-1,1,0)^t + v_3(-1,0,1)^t + v'_2(0,1,0)^t + v'_3(1,0,1)^t \colon v_2, v_3, v'_2, v'_3 \in \mathbb{R}\right\} \\ &= \left\{(-v_2 - v_3 + v'_3, v_2 + v'_2, v_3 + v'_3)^t \colon v_2, v_3, v'_2, v'_3 \in \mathbb{R}\right\} \,. \end{aligned}$$

Como puede observarse, no existe ninguna relación entre las tres coordenadas de los vectores en $S_1 + S_2$; por ello, podemos escribir simplemente:

$$S_1 + S_2 = \left\{(\alpha, \beta, \gamma)^t \colon \alpha, \beta, \gamma \in \mathbb{R}\right\} = \mathbb{R}^3 \,.$$

ESPACIOS VECTORIALES

I. DEFINICIONES. PRIMERAS PROPIEDADES.

1. Estructura. Ejemplos:

a) Definición.
b) Notaciones.
c) Ejemplos.

2. Reglas de cálculo:

a) Distributividad externa respecto a la resta vectorial.
b) $\lambda.0 = 0$.
c) $\lambda(-x) = -\lambda x$
d) Distributividad externa respecto a la resta de escalares.
e) $0x = 0$
f) $(-\lambda)x = -\lambda x$
g) Si el producto de un escalar por un vector es el vector nulo, entonces el escalar es o ó el vector es nulo.

II. SUBESPACIOS VECTORIALES.

3. Definición. Ejemplos:

Teorema 1.- *"Toda parte no vacía F de un espacio vectorial E sobre K, que sea estable para la adición interna de E y la multiplicación externa, es un subespacio vectorial de E"*

Teorema 2.- *"Para que una parte no vacía F de un espacio vectorial E sobre K, sea un subespacio vectorial de E, es necesario y suficiente que:*

$$(1)\ (\forall x, y \in E) \quad x + y \in F$$
$$(2)\ (\forall x \in E) \quad (\forall \lambda \in K) \quad \lambda x \in F \text{ ''}$$

4. Intersección de subespacios vectoriales. Subespacio vectorial generado por una parte A de un espacio vectorial.:

"Toda intersección de subespacios vectoriales de E es un subespacio vectorial de E"

Definición de combinación lineal.

Teorema y definición:

"El menor subespacio vectorial F conteniendo p elementos de E, $x_1, x_2, \ldots, x_p$,es decir, el subespacio generado por { $x_1, x_2, \ldots, x_p$} es el subespacio de las combinaciones lineales de $x_1, x_2, \ldots, x_p$. Se dice que A = { $x_1, x_2, \ldots, x_p$} es un sistema generador de F.

Cuando el mismo espacio vectorial E está generado por un número finito de sus elementos se dice que es de dimensión finita"

IV. INDEPENDENCIA LINEAL. BASES.

5. Independencia lineal.

a) **Partes contenidas y que contienen un sistema generador.**

¿Se puede hallar un sistema generador G de E al añadir un elemento (vector) cualquiera, g_i, tal que el nuevo sistema ya no genere E?

¿Se puede hallar un sistema generador G de E tal que si prescindimos de cualquiera de sus elementos (vectores), g_i, el nuevo sistema ya no genere E?

Definición de sistema generador mínimo

Definición.- *"Un conjunto finito { $x_1, x_2, \dots, x_p$} de elementos de un espacio vectorial E es linealmente independiente (sistema libre) si :*

$$\lambda_1 x_1 + \lambda_2 x_2 + \dots + \lambda_p x_p = 0 \quad \Rightarrow \quad \lambda_1 = \lambda_2 = \dots = \lambda_p = 0$$

Un conjunto cualquiera que no es linealmente independiente es linealmente dependiente (sistema ligado)"

b) **Propiedades inmediatas:**

1. *Todo subsistema de un sistema libre es libre*
2. *Todo supersistema de un sistema ligado es ligado.*
3. *Los elementos de un sistema libre son todos distintos.*
4. *Los elementos de un sistema libre son no nulos.*

Como consecuencia: **"Un sistema cualquiera de vectores es linealmente independiente si lo son todos sus subsistemas finitos" y "Un sistema de vectores es linealmente dependiente si hay un subsistema finito ligado"**

c) **Teorema 1.-** *"Un sistema es ligado si y sólo si existe un elemento en el sistema que sea combinación lineal finita de los otros elementos del sistema"*

Corolario.- *" Si {* $x_1, x_2, \dots, x_p$} es un sistema libre de p elementos de E y *{* $x_1, x_2, \dots, x_p, x$*} es un sistema ligado, entonces x pertenece al subespacio generado por {* $x_1, x_2, \dots, x_p$*} y se tiene que x =* $\lambda_1 x_1 + \lambda_2 x_2 + \dots + \lambda_p x_p$ *de una manera única"*

Teorema 2.- *"Si L = {*$a_1, a_2, \dots, a_m$*}es un sistema libre con m elementos de un espacio vectorial E, y G = {* $g_1, g_2, \dots, g_p$*} es un sistema generador de E con p elementos, entonces* $m \leq p$ *y G' ={* $a_1, a_2, \dots, a_m, g_{m+1}, g_{m+2}, \dots, g_p$*} (cambiando eventualmente la numeración de los g) es también un sistema generador de E"*

Corolario.- *"Si G es un sistema generador, con p elementos, de un espacio vectorial E, todo conjunto de vectores de E que tenga estrictamente más de p elementos es ligado"*

6. Bases de un espacio vectorial de dimensión finita.

a) Teorema 3 y definición.- *"Para un sistema B no vacío de un espacio vectorial E sobre K, de dimensión finita, las tres propiedades siguientes son equivalentes:*

1. *B es un sistema generador libre de E.*
2. *B es un sistema generador mínimo.*
3. *B es un sistema libre máximo.*

Todo sistema B = { a_1, a_2, ... , a_n } que posee una de estas propiedades se le llama una ***base de E****. Y Para todo vector x de E existe una única combinación de escalares tales que:* $x = \sum_{i=1}^{n} \mu_i a_i$ *, esos escalares reciben el nombre de coordenadas de x respecto de la base B"*

b) Existencia de bases:

Teorema 4.- *"Todo espacio vectorial E de dimensión finita, no reducido a {0} admite base. De una manera más precisa, si G es un sistema generador de E y L es un sistema libre contenido en G, existe una base de E tal que* $L \subset B \subset G$ *"*

Corolario (teorema de la base incompleta).- *"Si L y G son, respectivamente, un sistema libre y un sistema generador de E, existe un subconjunto H de G tal que* $L \cup H$ *sea una base de E"*

7. Dimensión de un espacio vectorial .

Teorema 5 y definición.- *"En un espacio vectorial E de dimensión finita sobre el cuerpo K, todas las bases tienen el mismo número de elementos. Este número común se llama* ***dimensión del espacio vectorial****, se le designa por* $\mathbf{dim_K E}$*"*

Teorema 6.- *"En un espacio vectorial E de dimensión n sobre el cuerpo K:*

1. *Todo sistema libre tiene a lo sumo n elementos.*
2. *Todo sistema que tenga al menos n+1 elementos es ligado"*

Corolario.- *"Todo sistema B de un espacio vectorial E de dimensión n sobre K que posee dos de las tres propiedades siguientes (n>0):*

1. *B tiene n elementos.*
2. *B es libre.*
3. *B es sistema generador de E.*

es una base de E"

Dimensión de un subespacio vectorial.

8. Rango de un sistema de vectores.

Definición.- *"Se llama rango de un sistema de vectores de un espacio vectorial E sobre K, la dimensión del subespacio F, supuesto de dimensión finita, engendrado por este sistema de vectores. Se le representa por* ***rg(S)****"*

Parte V

Tema 4. Bases y Dimensiones

Tema 4

Base y dimensión

4.1. Conjuntos generadores. Conjuntos generados

Sean $v_1, v_2, \dots, v_n$ vectores del espacio vectorial V. Recordemos que la suma de la forma

$$\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n,$$

donde $\alpha_1, \alpha_2, \dots, \alpha_n$ son escalares, se denomina *combinación lineal* de $v_1, v_2, \dots, v_n$. Este concepto es esencial en la teoría de los espacios vectoriales y en él se apoyan numerosas ideas. Comenzamos el tema dando la siguiente definición:

El conjunto de todas las combinaciones lineales de $v_1, v_2, \dots, v_n$ se denomina **conjunto generado** por $v_1, v_2, \dots, v_n$. Lo denotaremos por $\mathrm{Gen}(v_1, \dots, v_n)$.

Ejemplo

Consideremos los vectores $(1,1,1)^t$ y $(1,1,0)^t$ de $\mathbb{R}^3$. El conjunto generado por dichos vectores viene dado por el conjunto de todas las combinaciones lineales de la forma:

$$\begin{aligned}\{\alpha(1,1,1)^t+\beta(1,1,0)^t\colon \alpha,\beta\in\mathbb{R}\} &= \{(\alpha+\beta,\alpha+\beta,\alpha)^t\colon \alpha,\beta\in\mathbb{R}\}\\ &= \{(\alpha^*,\alpha^*,\beta^*)^t\colon \alpha^*,\beta^*\in\mathbb{R}\}.\end{aligned}$$

Ahora consideremos los vectores $(-1,-1,2)^t$, $(1,1,1)^t$ y $(1,1,0)^t$. El conjunto generado por estos vectores se describe por

$$\begin{aligned}&\{\alpha(-1,-1,2)^t+\beta(1,1,1)^t+\gamma(1,1,0)^t\colon \alpha,\beta,\gamma\in\mathbb{R}\}\\ &\qquad\qquad=\{(-\alpha+\beta+\gamma,-\alpha+\beta+\gamma,2\alpha+\beta)^t\colon \alpha,\beta,\gamma\in\mathbb{R}\}.\end{aligned}$$

Sin embargo, si observamos la primera y la segunda coordenadas, vemos que éste también puede ser descrito por:

$$\{(\alpha^*,\alpha^*,\beta^*)^t\colon \alpha^*,\beta^*\in\mathbb{R}\}.$$

Así: $\quad \mathrm{Gen}((1,1,1)^t,(1,1,0)^t)=\mathrm{Gen}((-1,-1,2)^t,(1,1,1)^t,(1,1,0)^t).$

El siguiente resultado es esencial en temas posteriores.

Teorema

Si $v_1,\ldots,v_n$ son elementos de un espacio vectorial V, entonces $\mathrm{Gen}(v_1,\ldots,v_n)$ es un subespacio de V.

Demostración. Sea $v=\alpha_1v_1+\cdots+\alpha_nv_n$ un elemento arbitrario de $\mathrm{Gen}(v_1,\ldots,v_n)$ y λ cualquier escalar. Puesto que

$$\lambda v=\lambda(\alpha_1v_1+\cdots+\alpha_nv_n)=(\lambda\cdot\alpha_1)v_1+\cdots+(\lambda\cdot\alpha_n)v_n\,,$$

se deduce que $\lambda v\in\mathrm{Gen}(v_1,\ldots,v_n)$.

Ahora consideremos $v,w\in\mathrm{Gen}(v_1,\ldots,v_n)$, $v=\alpha_1v_1+\cdots+\alpha_nv_n$ y $w=\beta_1v_1+\cdots+$

$\beta_n v_n$. Puesto que

$$v + w = (\alpha_1 + \beta_1)v_1 + \cdots + (\alpha_n + \beta_n)v_n \in \mathrm{Gen}(v_1, \ldots, v_n)$$

hemos probado que $\mathrm{Gen}(v_1, \ldots, v_n)$ es un subespacio de V. $\square$

Además, el teorema anterior motiva la siguiente definición:

El conjunto $\{v_1, \ldots, v_n\}$ es un **conjunto generador de** V si cada vector de V puede escribirse como una combinación lineal de $v_1, \ldots, v_n$.

Ejemplo

Consideremos los siguientes vectores de $\mathbb{R}^3$: $v_1 = (1, -1, 2)^t$, $v_2 = (-2, 3, 1)^t$ y $v_3 = (-1, 3, 8)^t$. Sea S el subespacio de $\mathbb{R}^3$ generado por v_1, v_2, v_3. Como $v_3 = 3v_1 + 2v_2$, cualquier combinación lineal de v_1, v_2 y v_3 puede reducirse a una combinación lineal de v_1 y v_2:

$$\alpha_1 v_1 + \alpha_2 v_2 + \alpha_3 v_3 = \alpha_1 v_1 + \alpha_2 v_2 + \alpha_3(3v_1 + 2v_2) = (\alpha_1 + 3\alpha_3)v_1 + (\alpha_2 + 2\alpha_3)v_2\,.$$

Así, $\quad S = \mathrm{Gen}(v_1, v_2, v_3) = \mathrm{Gen}(v_1, v_2)$.

Este ejemplo ilustra las siguientes proposiciones:

Proposición

Si $\{v_1, \ldots, v_n\}$ genera un espacio vectorial V y uno de estos vectores puede escribirse como combinación lineal de los otros $n-1$ vectores, entonces esos $n-1$ vectores generan V.

Demostración. Supongamos que v_n puede escribirse como combinación lineal de v_1, ..., v_{n-1}:

$$v_n = \alpha_1 v_1 + \cdots + \alpha_{n-1} v_{n-1}\,.$$

Sea $v \in V$. Puesto que $v_1, \ldots, v_n$ genera V podemos escribir

$$\begin{aligned} v &= \lambda_1 v_1 + \cdots + \lambda_{n-1} v_{n-1} + \lambda_n v_n \\ &= \lambda_1 v_1 + \cdots + \lambda_{n-1} v_{n-1} + \lambda_n (\alpha_1 v_1 + \cdots + \alpha_{n-1} v_{n-1}) \\ &= (\lambda_1 + \lambda_n \alpha_1) v_1 + \cdots + (\lambda_{n-1} + \lambda_n \alpha_{n-1}) v_{n-1} \end{aligned}$$

y cualquier vector $v \in V$ puede escribirse como combinación lineal de $v_1, \ldots, v_{n-1}$, i.e., $V = \mathrm{Gen}(v_1, \ldots, v_{n-1})$. □

Proposición

Dados n vectores $v_1, \ldots, v_n$, es posible escribir uno de los vectores como combinación lineal de los otros $n-1$ si y sólo si existen escalares $c_1, \ldots, c_n$, no todos cero, tales que

$$c_1 v_1 + \cdots + c_n v_n = 0 .$$

Demostración. Supongamos que uno de estos vectores, digamos v_n, puede escribirse como combinación lineal de los otros:

$$v_n = \alpha_1 v_1 + \cdots + \alpha_{n-1} v_{n-1} .$$

Restando v_n a ambos lados de la ecuación, obtenemos

$$\alpha_1 v_1 + \cdots + \alpha_{n-1} v_{n-1} - v_n = 0 .$$

Si hacemos $c_1 = \alpha_1, \ldots, c_{n-1} = \alpha_{n-1}, c_n = -1$, se deduce que

$$c_1 v_1 + \cdots + c_n v_n = 0 .$$

Inversamente, si

$$c_1 v_1 + \cdots + c_n v_n = 0$$

y al menos uno de los c_i's, digamos c_n, es no nulo, entonces

$$v_n = \frac{-c_1}{c_n} v_1 + \cdots + \frac{-c_{n-1}}{c_n} v_{n-1}\,.$$

□

En virtud de las propiedades anteriores podemos dar las siguientes definiciones (ya conocidas):

Los vectores $v_1, \ldots, v_n$ del espacio vectorial V se dicen **linealmente independientes** si

$$c_1 v_1 + \cdots + c_n v_n = 0 \tag{4.1}$$

implica que todos los escalares $c_1, \ldots, c_n$ son iguales a 0. Inversamente, los vectores $v_1, \ldots, v_n$ del espacio vectorial V se denominan **linealmente dependientes** si existen ciertos escalares $c_1, \ldots, c_n$, no todos nulos, tales que

$$c_1 v_1 + \cdots + c_n v_n = 0\,.$$

Obsérvese que la expresión (4.1) representa un sistema lineal homogéneo, que siempre tiene solución trivial. Los vectores serán linealmente independientes si dicha solución es *única*.

El siguiente teorema se da sin demostración.

Teorema

Sean $v_1, \ldots, v_n$ vectores de un espacio vectorial V. Un vector v de $\mathrm{Gen}(v_1, \ldots, v_n)$ puede escribirse de manera única como combinación lineal de $v_1, \ldots, v_n$ si y sólo si $v_1, \ldots, v_n$ son linealmente independientes.

4.2. Bases y dimensión

Los vectores $v_1, \ldots, v_n$ forman una **base** del espacio vectorial V si y sólo si

1) $v_1, \ldots, v_n$ son linealmente independientes, y

2) $v_1, \ldots, v_n$ generan V.

NOTA: es importante observar que las bases son conjuntos ordenados de vectores. Por lo tanto denotaremos una base como $B = (v_1, v_2, \ldots, v_n)$.

Ejemplo

La base canónica de $\mathbb{R}^2$ es $B_0 = (e_1, e_2)$ con $e_1 = (1, 0)^t$ y $e_2 = (0, 1)^t$. Verifiquemos que B_0 es efectivamente una base.

Claramente, e_1 y e_2 son linealmente independientes, puesto que $\alpha_1 e_1 + \alpha_2 e_2 = 0$ implica que $\alpha_1 = \alpha_2 = 0$.

Además todo vector $v = (v_1, v_2)^t$ de $\mathbb{R}^2$ puede representarse como una combinación lineal de estos dos vectores de la forma

$$v = v_1 e_1 + e_2 v_2$$

y así $\mathbb{R}^2 = \mathrm{Gen}(e_1, e_2)$. Por tanto, $B_0 = (e_1, e_2)$ es una base de $\mathbb{R}^2$.

Veamos otras propiedades de los conjuntos generadores y de las bases de un espacio vectorial.

Teorema

Si $(v_1, \ldots, v_n)$ es un conjunto generador de un espacio vectorial V, cualquier colección de m vectores de V con $m > n$ es linealmente dependiente.

Como consecuencia, si $(v_1, \ldots, v_n)$ y $(u_1, \ldots, u_m)$ son bases de un espacio vectorial V, entonces $n = m$. A la vista de este resultado, podemos referirnos al número de elementos de cualquier base de un espacio vectorial dado V como la **dimensión** de dicho espacio vectorial. Si existe un conjunto finito de vectores que genera V, diremos que V es de dimensión finita. En caso contrario, diremos que V es de dimensión infinita. En este curso sólo nos ocuparemos de espacios vectoriales de dimensión finita.

Teorema

Si V es un espacio vectorial de dimensión finita $n > 0$:

- Cualquier conjunto de n vectores linealmente independientes genera V.
- Si n vectores generan V, son linealmente independientes.

Finalmente:

Proposición

Sea $B = (v_1, \dots, v_n)$ un conjunto de vectores de V (sobre $\mathbb{K}$). Las siguientes propiedades son equivalentes:

1. B es una *base* de V.
2. B es un conjunto *linealmente independiente maximal* (es decir, si añadimos cualquier vector al conjunto, éste deja de ser linealmente independiente).
3. B es un *conjunto generador minimal* de V (es decir, si eliminamos cualquier vector de B, éste deja de ser generador de V).

Veamos el caso de algunos espacios vectoriales muy utilizados. Es fácil demostrar que:

- $\dim(\mathbb{K}^n) = n$.
- $\dim(\mathbb{K}^{m\times n}) = mn$.

- $\dim(\mathbb{P}_n) = n + 1$.

Conocidos estos resultados, podremos determinar si un conjunto de vectores forma o no una base en el correspondiente espacio vectorial, observando si su cardinalidad coincide con la dimensión del espacio y, en tal caso, estudiando simplemente si son linealmente independientes.

Ejemplo

El espacio $\mathbb{R}^3$ de dimensión 3 es generado por los vectores linealmente independientes $e_1 = (1, 0, 0)^t$, $e_2 = (0, 1, 0)^t$ y $e_3 = (0, 0, 1)^t$, los cuales forman la base canónica $B_0 = (e_1, e_2, e_3)$ de $\mathbb{R}^3$.

El conjunto ordenado $((1, 1, 1)^t, (1, 1, 0)^t, (1, 0, 0)^t)$ también es una base de $\mathbb{R}^3$. Para probar esto, igualamos a cero una combinación lineal arbitraria de todos ellos

$$\alpha_1(1, 1, 1)^t + \alpha_2(1, 1, 0)^t + \alpha_3(1, 0, 0)^t = 0$$

que da lugar al sistema de ecuaciones

$$\left.\begin{aligned} \alpha_1 + \alpha_2 + \alpha_3 &= 0 \\ \alpha_1 + \alpha_2 &= 0 \\ \alpha_1 &= 0 \end{aligned}\right\} \Rightarrow \alpha_1 = \alpha_2 = \alpha_3 = 0\,,$$

lo que demuestra que los tres vectores son linealmente independientes y, por tanto, forman una base de $\mathbb{R}^3$.

Teorema Fundamental del Álgebra Lineal

Dada la matriz $A \in \mathbb{K}^{m\times n}$, los espacios fila y columna de A tienen la misma dimensión y ésta es igual al rango de la matriz $r = \mathrm{rg}(A)$. Se pueden encontrar bases para estos espacios entre las filas y columnas de la propia A. Los espacios nulos de A y

A^t tienen dimensiones $n - r$ y $m - r$.

Espacio Vectorial	Dimensión	Subespacio de
$N(A)$, espacio nulo	$n - rg(A)$	$\mathbb{K}^m$
$\mathcal{C}(A)$, espacio columna	$rg(A)$	$\mathbb{K}^n$
$\mathcal{C}(A^t)$, espacio fila	$rg(A)$	$\mathbb{K}^m$
$N(A^t)$, espacio nulo de A^t	$m - rg(A)$	$\mathbb{K}^n$

Para encontrar una base del espacio nulo de A, resolvemos el sistema $A\,x = 0$; si aplicamos el método de Gauss, tendremos que reducir A a forma escalonada.

Para encontrar una base del espacio columna, la forma escalonada de A nos permitirá identificar las columnas que forman un sistema linealmente independiente (las que corresponden a las variables pivote).

De igual modo, para encontrar una base del espacio fila, reducimos A a su forma escalonada; las filas no nulas corresponden a las que forman $\boxed{\text{(en la matriz } A)}$ un conjunto linealmente independiente.

Para el caso del espacio nulo de A^t, podemos encontrar una base resolviendo el sistema $A^t\,x = 0$. Sin embargo hay una alternativa más práctica: aplicar el método de Gauss a la matriz $\left(\,A \mid I\,\right)$ para obtener la matriz equivalente $\left(\,U \mid J\,\right)$. Las $m - rg(A)$ últimas filas de J forman una base del espacio nulo de la traspuesta de A. Esto se debe a que al realizar operaciones elementales sobre $\left(\,A \mid I\,\right)$, la matriz J, invertible, refleja las operaciones que se han realizado para obtener U, es decir, $J\,A = U$. Al realizar este producto, obtenemos que las últimas $m - rg(A)$ filas de A son vectores fila 0; es decir, las últimas $m - rg(A)$ filas de J, multiplicadas por A, producen el vector fila 0 y, por ello, forman una base del espacio nulo de A^t. Veámoslo con un ejemplo:

Ejemplo

Sea la matriz

$$A = \begin{pmatrix} 3 & 5 & 3 & 6 \\ 0 & 1 & 4 & -1 \\ 6 & 11 & 10 & 11 \end{pmatrix}.$$

Vamos a determinar sus cuatro espacios vectoriales asociados.

1) **Espacio Nulo**: $N(A) = \{v \in \mathbb{R}^4 : A\,v = 0\}$. La condición $A\,v = 0$ se transforma en:

$$\begin{pmatrix} 3 & 5 & 3 & 6 \\ 0 & 1 & 4 & -1 \\ 6 & 11 & 10 & 11 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix},$$

que representa un sistema lineal con tres ecuaciones y cuatro incógnitas v_1, v_2, v_3 y v_4. Si obtenemos una forma escalonada de la matriz, resulta:

$$\begin{pmatrix} 3 & 5 & 3 & 6 \\ 0 & 1 & 4 & -1 \\ 0 & 1 & 4 & -1 \end{pmatrix} \equiv \begin{pmatrix} \boxed{3} & 5 & 3 & 6 \\ 0 & \boxed{1} & 4 & -1 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Se tiene que v_1 y v_2 serán variables pivote y v_3 y v_4 los parámetros. Por tanto el espacio nulo será

$$N(A) = \left\{ v = (v_1, v_2, v_3, v_4)^t \in \mathbb{R}^4 \colon v_1 = \frac{17}{3} v_3 - \frac{11}{3} v_4,\ v_2 = -4v_3 + v_4 \right\}$$

y una base vendrá dada por

$$B = \left((17, -12, 3, 0)^t, (-11, 3, 0, 3)^t \right).$$

Obsérvese que de este resultado también se deduce que el rango de A es 2.

2) **Espacio columna**: $\mathcal{C}(A) = \{v \in \mathbb{R}^3 \colon v = a_1A_1 + a_2A_2 + a_3A_3 + a_4A_4,\ a_i \in \mathbb{R}\}$ es el conjunto de vectores que son combinación lineal de las columnas de la matriz A. Para determinar una base, podemos reducir la matriz A a forma escalonada (como hicimos antes):

$$\begin{pmatrix} 3 & 5 & 3 & 6 \\ 0 & 1 & 4 & -1 \\ 6 & 11 & 10 & 11 \end{pmatrix} \equiv \begin{array}{c} \begin{array}{cccc} \downarrow & \downarrow & & \end{array} \\ \begin{pmatrix} \boxed{3} & 5 & 3 & 6 \\ 0 & \boxed{1} & 4 & -1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \end{array}.$$

Las variables pivote se corresponden con las columnas 1 y 2, por tanto, una base del espacio columna estará formada por las $\boxed{\text{correspondientes columnas de la matriz A}}$

$$B = \left((3,0,6)^t, (5,1,11)^t\right).$$

La dimensión del espacio columna es, obviamente, 2.

3) **Espacio fila**: $\mathcal{C}(A^t)$ es generado por dos filas de la matriz A, las que corresponden a las filas no nulas en la forma escalonada obtenida o, equivalentemente, las que contienen los coeficientes de las variables pivote en la forma escalonada:

$$\begin{pmatrix} 3 & 5 & 3 & 6 \\ 0 & 1 & 4 & -1 \\ 6 & 11 & 10 & 11 \end{pmatrix} \equiv \begin{array}{c} \rightarrow \\ \rightarrow \\ \end{array} \begin{pmatrix} \boxed{3} & 5 & 3 & 6 \\ 0 & \boxed{1} & 4 & -1 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Así $\mathcal{C}(A^t) = \{v \in \mathbb{R}^4 \colon v = b_1(3,5,3,6)^t + b_2(0,1,4,-1)^t, b_1, b_2 \in \mathbb{R}\}$. La dimensión del espacio fila es también 2.

4) **Espacio nulo de la traspuesta**: $N(A^t)$ es el conjunto de vectores de la forma

$v = (v_1, v_2, v_3)^t$ que satisfacen $A^t x = 0$ ó también

$$(v_1, v_2, v_3) \begin{pmatrix} 3 & 5 & 3 & 6 \\ 0 & 1 & 4 & -1 \\ 6 & 11 & 10 & 11 \end{pmatrix} = (0, 0, 0, 0)\,.$$

Si escribimos la matriz ampliada del sistema lineal $A^t x = 0$ y hallamos su forma escalonada, resulta:

$$\left(\begin{array}{ccc|c} 3 & 0 & 6 & 0 \\ 5 & 1 & 11 & 0 \\ 3 & 4 & 10 & 0 \\ 6 & -1 & 11 & 0 \end{array}\right) \equiv \begin{pmatrix} 3 & 0 & 6 & 0 \\ 0 & 3 & 3 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Por tanto:

$$\begin{aligned} N(A^t) &= \left\{v = (v_1, v_2, v_3)^t \colon v_1 = -2v_3, v_2 = -v_3\right\} \\ &= \{v \in \mathbb{R}^3 \colon v = v_3(-2, -1, 1)^t, v_3 \in \mathbb{R}\}. \end{aligned}$$

Obviamente, la dimensión del espacio nulo de la traspuesta es $3 - 2 = 1$.

Una manera alternativa para obtener el espacio nulo de la traspuesta simultáneamente a los otros espacios $\boxed{\text{(con menos cálculos)}}$ consiste en escribir

$$(A|I) = \left(\begin{array}{cccc|ccc} 3 & 5 & 3 & 6 & 1 & 0 & 0 \\ 0 & 1 & 4 & -1 & 0 & 1 & 0 \\ 6 & 11 & 10 & 11 & 0 & 0 & 1 \end{array}\right)$$

y reducirla a forma escalonada:

$$(U|J) = \left(\begin{array}{cccc|ccc} 3 & 5 & 3 & 6 & 1 & 0 & 0 \\ 0 & 1 & 4 & -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 2 & 1 & -1 \end{array}\right).$$

La última fila de J, $(2, 1, -1)^t$ proporciona una base del espacio nulo de la traspuesta, como ya sabíamos.

4.2.1. Coordenadas

Sea V un espacio vectorial de dimensión n sobre $\mathbb{K}$ y sea $B = (v_1, \ldots, v_n)$ una base de V. El siguiente teorema es clave en muchos de los resultados que veremos en el curso:

Teorema de representación única

Sea $B = (b_1, \ldots, b_n)$ una base del espacio vectorial V. Todo vector $v \in V$ se puede expresar $\boxed{\text{de forma única}}$ como combinación lineal de los vectores de B:

$$v = c_1\, b_1 + \cdots + c_n\, b_n\,.$$

Como establece el teorema de representación única, si $v \in V$, podremos escribir v de la forma

$$v = v_1\, b_1 + \cdots + v_n\, b_n\,,$$

donde $v_1, \ldots, v_n$ son escalares ($\in \mathbb{K}$). Así, podemos asociar a cada vector $v \in V$ un $\boxed{\text{vector único con respecto a la base } B}$ $[v]_B \in \mathbb{K}^n$ dado por $[v]_B = (v_1, \ldots, v_n)^t$. Este vector se denomina **vector de coordenadas** de v $\boxed{\text{con respecto a la base } B}$. Nos referiremos a los escalares v_i $(i = 1, \ldots, n)$ como **coordenadas de v con respecto a** B.

Ejemplo

Consideremos el espacio vectorial $\mathbb{P}_1$. Es sencillo demostrar que $B = (1, x)$ es una base de este espacio. Cualquier polinomio $p(x) = a_0 + a_1 x$ puede representarse mediante el vector de coordenadas con respecto a la base B:

$$[p]_B = \begin{pmatrix} a_0 \\ a_1 \end{pmatrix} .$$

Obsérvese que $p \neq [p]_B$; p es un polinomio y $[p]_B$ es un vector de $\mathbb{R}^2$: el segundo es una útil *representación* del primero.

Sea un espacio vectorial V de dimensión n. En él definimos una base $B = (b_1, b_2, \dots, b_n)$. Dado un vector cualquiera $a \in V$, existe una única combinación lineal

$$a = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n ,$$

que definen las coordenadas $[a]_B = (a_1, a_2, \dots, a_n)^t$ de a en la base B. Podemos recuperar el vector original escribiéndolo en forma de producto matricial:

$$a = B\,[a]_B = (b_1, b_2, \dots, b_n)\,[a]_B = (b_1, b_2, \dots, b_n) \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} .$$

Nótese que, aunque definimos en el Tema 1 que las entradas de las matrices eran escalares, en la fórmula anterior los elementos del vector fila $(b_1, b_2, \dots, b_n)$ no son escalares, sino vectores. Sin embargo, la regla del producto sigue siendo la misma; pero el resultado es un vector, no un escalar.

4.3. Cambio de base

En el espacio vectorial $\mathbb{R}^n$ existe una base canónica natural que viene dada por

$$B_0 = (e_1, e_2, \dots, e_n) = \left((1, 0, 0, \dots, 0, 0)^t, (0, 1, 0, \dots, 0, 0)^t, \dots, (0, 0, 0, \dots, 0, 1)^t\right) .$$

Sin embargo, en muchas ocasiones es más eficiente trabajar en otras bases en $\mathbb{R}^n$. En otros espacios vectoriales no existe una base canónica natural y, en muchos cálculos prácticos, es conveniente elegir una base adecuada para realizarlos.

Es importante recordar que un vector $v \in V$ es un objeto abstracto que existe sin necesidad de establecer una base para el espacio vectorial V. Al fijar una base B en V podemos obtener las coordenadas de v con respecto a dicha base $[v]_B$. Estas coordenadas dependen obviamente de la base escogida: si elegimos otra base B', las coordenadas del vector v serán $[v]_{B'}$ y, en general, serán diferentes de $[v]_B$.

Ejemplo

Sea el espacio vectorial V formado por los polinomios $p(x)$ con coeficientes reales de grado menor que 5 y que sólo contienen potencias pares de x

$$V = \left\{p \colon \mathbb{R} \to \mathbb{R} \mid p(x) = a_0 + a_1 x^2 + a_2 x^4 \,,\, a_0, a_1, a_2 \in \mathbb{R}\right\} .$$

Sea ahora el vector p dado por la expresión $p(x) = x^4 - 2x^2$. La gráfica de este polinomio está representada en la figura 4.1. Una base posible de V es la siguiente

$$B_1 = (p_1, p_2, p_3) = (1, x^2, x^4) .$$

y las coordenadas de p en esta base son obviamente

$$[p]_{B_1} = (0, -2, 1)^t ,$$

ya que

$$p = (p_1, p_2, p_3)\,[p]_{B_1} = (p_1, p_2, p_3) \begin{pmatrix} 0 \\ -2 \\ 1 \end{pmatrix} = -2p_2 + p_3 = x^4 - 2x^2 .$$

Otra base posible de V es la siguiente

$$B_2 = (q_1, q_2, q_3) = \left(1, x^2 - 1, (x^2 - 1)^2\right) .$$

y las coordenadas de p en esta base son (después de un poco de álgebra)

$$[p]_{B_2} = (-1, 0, 1)^t \,,$$

ya que

$$p = (q_1, q_2, q_3)\,[p]_{B_2} = (q_1, q_2, q_3) \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} = -q_1 + q_3 = x^4 - 2x^2 \,.$$

Como se puede observar, el vector p no depende de la base elegida para representarlo (en otras palabras, la gráfica del polinomio $p(x)$ no depende de la base). Sin embargo, las coordenadas de dicho vector sí dependen de la base elegida.

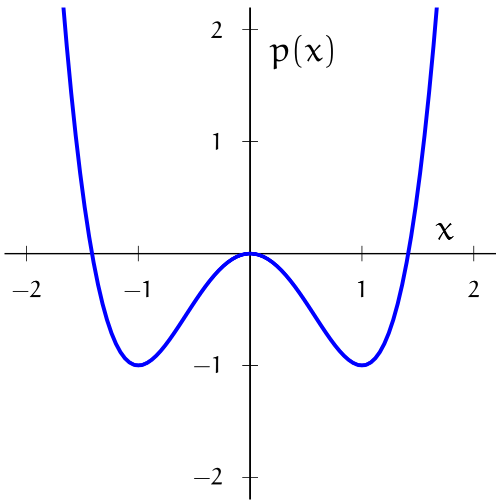


Figura 4.1: Gráfica del polinomio $p(x) = x^4 - 2x^2$.

Vamos a estudiar cómo varían las coordenadas de un vector $v \in V$ respecto a una base B_1 cuando usamos una nueva base B_2. Para ello definimos primero la **matriz de cambio de base** $T_{B_1 B_2}$. La definición de esta matriz **no** es universal, ya que en algunos libros se usa, en vez de $T = T_{B_1 B_2}$, su inversa T^{-1}.

Sea un espacio vectorial V de dimensión n. En él definimos dos bases distintas $B = (b_1, b_2, \dots, b_n)$ y $B' = (b'_1, b'_2, \dots, b'_n)$. La **matriz de cambio de base** $T_{BB'}$ (de dimensión $n \times n$) se define mediante la ecuación

$$B' = B\, T_{BB'} \quad \Leftrightarrow \quad (b'_1, b'_2, \dots, b'_n) = (b_1, b_2, \dots, b_n)\, T_{BB'}\,. \tag{4.2}$$

En otras palabras, la columna i-ésima T_i de la matriz $T = T_{BB'}$ contiene las coordenadas del vector b'_i de la "nueva" base B' respecto de la base "antigua" B: $\quad T_i = [b'_i]_B$. De otra manera:

$$T_{BB'} = \left([b'_1]_B, [b'_2]_B, \dots, [b'_n]_B\right)\,.$$

La matriz $T_{BB'}$ se denotará simplemente por T cuando las bases B y B' estén bien definidas en el contexto del problema.

La matriz de cambio de base $T_{BB'}$ es una matriz no singular ($\det(T_{BB'}) \neq 0$), ya que sus columnas son linealmente independientes.

Ejemplo

Supongamos que queremos usar la base $B = (u_1, u_2)$ de $\mathbb{R}^2$, donde las coordenadas de los vectores u_1, u_2, respecto de la base canónica $B_0 = (e_1, e_2)$ de $\mathbb{R}^2$, son

$$[u_1]_{B_0} = (3, 2)^t\,, \qquad [u_2]_{B_0} = (1, 1)^t\,.$$

La matriz asociada al cambio de base $B_0 \to B_1$ viene dada por la fórmula (4.2): cada columna contiene las coordenadas del vector correspondiente de la base B_1 en la base B_0

$$(u_1, u_2) = (e_1, e_2)\, T_{B_0 B_1} = (e_1, e_2) \begin{pmatrix} 3 & 1 \\ 2 & 1 \end{pmatrix}\,.$$

Si desarrollamos esta ecuación, vemos que es equivalente a:

$$\begin{aligned} u_1 &= 3e_1 + 2e_2 \\ u_2 &= e_1 + e_2 \,, \end{aligned}$$

tal y como ya sabíamos.

Ejemplo

Sea el espacio vectorial $\mathbb{P}_1$ de los polinomios de grado 1 con coeficientes reales. Tenemos la base de este espacio

$$B_1 = (1, x)$$

y queremos expresar los polinomios de $\mathbb{P}_1$ en función de la base

$$B_2 = (1 + x, 1 - x) \,,$$

donde la demostración de que B_2 es efectivamente una base de $\mathbb{P}_1$ de deja como ejercicio.

La matriz asociada al cambio de base $B_1 \to B_2$ viene dada por la fórmula (4.2):

$$(1 + x, 1 - x) = (1, x)\, T_{B_1 B_2} = (1, x) \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} .$$

Si desarrollamos esta ecuación, vemos que es equivalente a las tautologías

$$\begin{aligned} 1 + x &= 1 + x \\ 1 - x &= 1 - x \,, \end{aligned}$$

tal y como debe ser.

Si tenemos un vector $x \in V$ en un espacio vectorial V de dimensión n y calculamos sus coordenadas $[x]_B = (x_1, \dots, x_n)^t$ y $[x]_{B'} = (x'_1, \dots, x'_n)^t$ respecto a dos bases distintas $B = (b_1, \dots, b_n)$ y $B' = (b'_1, \dots, b'_n)$ de V y que están relacionadas mediante la matriz de cambio de base $T_{BB'}$, entonces:

$$\begin{aligned} x &= B\,[x]_B = (b_1, b_2, \dots, b_n)\ [x]_B \\ &= B'\,[x]_{B'} = (b'_1, b'_2, \dots, b'_n)\ [x]_{B'} \,. \end{aligned}$$

Dado que $(b'_1, b'_2, \dots, b'_n) = (b_1, b_2, \dots, b_n)\, T_{BB'}$, entonces

$$(b_1, b_2, \dots, b_n)\,[x]_B = (b'_1, b'_2, \dots, b'_n)\,[x]_{B'} = (b_1, b_2, \dots, b_n)\, T_{BB'}\,[x]_{B'} \,.$$

De esta ecuación se ve cómo se transforman las coordenadas del vector x al cambiar de base $B \to B'$:

$$[x]_B = T_{BB'}\,[x]_{B'} \,.$$

Aunque la fórmula fundamental es (4.2), que define la matriz de cambio de base $B \to B'$, esta última fórmula nos dice cómo se transforman las coordenadas de cualquier vector $x \in V$ bajo dicho cambio de coordenadas. Si queremos escribir las coordenadas del vector x en la base "nueva" B' en función de sus coordenadas en la base "antigua" B, basta con multiplicar (por la izquierda) la ecuación anterior por $T^{-1}_{BB'}$:

$$[x]_{B'} = T^{-1}_{BB'}\,[x]_B \,.$$

Nótese que la matriz $T_{BB'}$ define el cambio de base (4.2), pero su inversa $T^{-1}_{BB'}$ es la que define el cambio de coordenadas.

Ejemplo

Supongamos que queremos cambiar de la base canónica $B_0 = (e_1, e_2)$ de $\mathbb{R}^2$ a una nueva base $B = (b_1, b_2)$ donde las coordenadas de los vectores b_1, b_2, respecto a B_0, son

$$[b_1]_{B_0} = (3, 2)^t \,, \qquad [b_2]_{B_0} = (1, 1)^t \,.$$

La matriz de cambio de base asociada a $B_0 \to B_1$ es, como ya vimos,

$$T_{B_0B_1} = \begin{pmatrix} 3 & 1 \\ 2 & 1 \end{pmatrix}.$$

Esta matriz es no singular y su inversa es

$$T^{-1}_{B_0B_1} = \begin{pmatrix} 1 & -1 \\ -2 & 3 \end{pmatrix}.$$

Si tenemos el vector x de coordenadas $[x]_{B_0} = (2,-1)^t$ en la base canónica B_0, sus coordenadas en la nueva base B serán

$$[x]_B = T^{-1}_{B_0B_1}\,[x]_{B_0} = \begin{pmatrix} 1 & -1 \\ -2 & 3 \end{pmatrix} \begin{pmatrix} 2 \\ -1 \end{pmatrix} = \begin{pmatrix} 3 \\ -7 \end{pmatrix}$$

Nótese que

$$x = (b_1, b_2)\,[x]_B = (b_1, b_2) \begin{pmatrix} 3 \\ -7 \end{pmatrix} = 2e_1 - e_2 = (e_1, e_2) \begin{pmatrix} 2 \\ -1 \end{pmatrix} = (e_1, e_2)\,[x]_{B_0}.$$

Ejemplo

Sea el espacio vectorial $\mathbb{P}_1$ de los polinomios de grado 1 con coeficientes reales. Queremos hacer el cambio de base desde la base $B_1 = (1, x)$ a la nueva base $B_2 = (1+x, 1-x)$.

La matriz de cambio de base $T_{B_1B_2}$ viene dada por

$$T_{B_1B_2} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

Su inversa es

$$T^{-1}_{B_1B_2} = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

Si tenemos el vector $p(x) = 2 - 3x \in \mathbb{P}_1$, sus coordenadas en la base B_1 son obviamente $[p]_{B_1} = (2, -3)^t$. Sus coordenadas en la base B_2 vendrán dadas por:

$$[p]_{B_2} = T^{-1}_{B_1B_2}\,[p]_{B_1} = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} 2 \\ -3 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} -1 \\ 5 \end{pmatrix}$$

La comprobación de este resultado es trivial:

$$p = (1+x, 1-x)\,[p]_{B_2} = (1+x, 1-x)\,\frac{1}{2}\begin{pmatrix} -1 \\ 5 \end{pmatrix} = 2-3x = (1, x)\begin{pmatrix} 2 \\ -3 \end{pmatrix} = (1, x)\,[p]_{B_1}.$$

Hay dos propiedades importantes que cumplen las matrices de cambio de base:

1. Supongamos que hacemos el cambio de base $B \to B'$, lo que define la matriz de cambio de base $T_{BB'}$. Si hiciésemos el cambio de base inverso $B' \to B$, entonces la matriz de este cambio de base es simplemente

$$\boxed{T_{B'B} = T^{-1}_{BB'}.}$$

   Este resultado se obtiene de (4.2) multiplicando por la derecha por la matriz $T^{-1}_{BB'}$ (que siempre existe ya que $T_{BB'}$ es no singular). Esto implica que

$$[x]_{B'} = T^{-1}_{BB'}\,[x]_B = T_{B'B}\,[x]_B.$$

2. Supongamos que hacemos dos cambios sucesivos de base. Primero vamos de la base $B = (b_1, \ldots, b_n)$ a la base $B' = (b'_1, \ldots, b'_n)$ con la matriz de cambio de base $T_{BB'}$:

$$(b'_1, b'_2, \ldots, b'_n) = (b_1, b_2, \ldots, b_n)\,T_{BB'}.$$

Después cambiamos de la base B' a otra base $B'' = (b_1'', \ldots, b_n'')$ con la matriz $T_{B'B''}$:

$$(b_1'', b_2'', \ldots, b_n'') = (b_1', b_2', \ldots, b_n')\, T_{B'B''}\,.$$

Evidentemente podemos hacer este cambio de base en un solo paso $B \to B''$. La matriz de cambio de base $T_{BB''}$ vendría dada por:

$$(b_1'', b_2'', \ldots, b_n'') = (b_1', b_2', \ldots, b_n')\, T_{B'B''} = (b_1, b_2, \ldots, b_n)\, T_{BB'}\, T_{B'B''}\,.$$

Es decir,

$$\boxed{T_{BB''} = T_{BB'}\, T_{B'B''}\,.}$$

Ejemplo

Sea el espacio vectorial $\mathbb{P}_1$ de los polinomios de grado 1 con coeficientes reales. Queremos hacer el cambio de base desde la base $B_1 = (1, x)$ a la nueva base $B_2 = (1 + x, 1 - x)$.

La matriz de cambio de base $T_{B_1B_2}$ y su inversa vienen dadas por

$$T_{B_1B_2} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \qquad T_{B_1B_2}^{-1} = T_{B_2B_1} = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

Sea ahora una nueva base $B_3 = (1+2x, 1-2x)$. Para calcular la matriz $T_{B_2B_3}$ debemos calcular las coordenadas de $1 \pm 2x$ en la base B_2. Para ello hay que resolver los sistemas de ecuaciones lineales

$$\begin{aligned}
1 + 2x &= \alpha(1+x) + \beta(1-x) = (\alpha+\beta) + x(\alpha-\beta) \Rightarrow \left.\begin{array}{c} \alpha+\beta = 1 \\ \alpha-\beta = 2 \end{array}\right\} \\
&\Rightarrow \alpha = \frac{3}{2}, \quad \beta = -\frac{1}{2} \\
1 - 2x &= \alpha(1+x) + \beta(1-x) = (\alpha+\beta) + x(\alpha-\beta) \Rightarrow \left.\begin{array}{c} \alpha+\beta = 1 \\ \alpha-\beta = -2 \end{array}\right\} \\
&\Rightarrow \alpha = -\frac{1}{2}, \quad \beta = \frac{3}{2}
\end{aligned}$$

Luego la matriz de cambio de base $T_{B_2B_3}$ y su inversa vienen dadas por

$$T_{B_2B_3} = \frac{1}{2}\begin{pmatrix} 3 & -1 \\ -1 & 3 \end{pmatrix}, \qquad T^{-1}_{B_2B_3} = T_{B_3B_2} = \frac{1}{4}\begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}.$$

Ya hemos visto que el vector $p(x) = 2-3x \in \mathbb{P}_1$ tiene coordenadas $[p]_{B_1} = (2,-3)^t$ en la base B_1 y coordenadas $(-1/2, 5/2)^t$ en B_2. Sus coordenadas en la base B_3 vendrán dadas por:

$$[p]_{B_3} = T_{B_3B_2}\,[p]_{B_2} = \frac{1}{4}\begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}\frac{1}{2}\begin{pmatrix} -1 \\ 5 \end{pmatrix} = \frac{1}{4}\begin{pmatrix} 1 \\ 7 \end{pmatrix}.$$

Por supuesto este resultado es el correcto:

$$p = (1+2x, 1-2x)\,[p]_{B_3} = (1+2x, 1-2x)\,\frac{1}{4}\begin{pmatrix} 1 \\ 7 \end{pmatrix} = 2-3x\,.$$

Si queremos hacer el cambio de base $B_1 \to B_3$ directamente, basta calcular

$$T_{B_1B_3} = T_{B_1B_2}\,T_{B_2B_3} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\frac{1}{2}\begin{pmatrix} 3 & -1 \\ -1 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 2 & -2 \end{pmatrix},$$

que es el resultado esperado, ya que las columnas de $T_{B_1B_3}$ contienen las coordenadas de los vectores de B_3 en la base B_1. La inversa de esta matriz es

$$T^{-1}_{B_1B_3} = T_{B_3B_1} = \frac{1}{4}\begin{pmatrix} 2 & 1 \\ 2 & -1 \end{pmatrix}.$$

Las coordenadas de p en la base B_3 vienen dadas por

$$[p]_{B_3} = T_{B_3B_1}\,[p]_{B_1} = \frac{1}{4}\begin{pmatrix} 2 & 1 \\ 2 & -1 \end{pmatrix}\frac{1}{2}\begin{pmatrix} -1 \\ 5 \end{pmatrix} = \frac{1}{4}\begin{pmatrix} 1 \\ 7 \end{pmatrix}.$$

La comprobación es simple:

$$p = (1+2x, 1-2x)\,[p]_{B_3} = (1+2x, 1-2x)\,\frac{1}{4}\begin{pmatrix} 1 \\ 7 \end{pmatrix} = 2-3x\,.$$

4.3.1. Cambio de coordenadas usando la base canónica

Frecuentemente, una de las bases del espacio vectorial con las que trabajamos es la canónica, B_0. En este caso, la matriz de cambio de base $T_{B_0 B_1}$ a otra base B_1 es particularmente sencilla, como hemos visto. En cambio, si trabajamos con bases arbitrarias B_1 y B_2, el cálculo de las coordenadas de los vectores de una base con respecto a la otra puede ser engorroso. Una alternativa habitual consiste en lo siguiente. Dada la base B_1, podemos calcular la matriz de cambio de base $T_{B_0 B_1}$ de B_0 a B_1. De igual manera, dada la base B_2, podemos hallar su correspondiente matriz de cambio de base $T_{B_0 B_2}$ de B_0 a B_2. La idea consiste en, en vez de hacer directamente el cambio de base $B_1 \to B_2$, hacer el cambio de base pasando por la base canónica B_0: $B_1 \to B_0 \to B_2$. Entonces

$$\boxed{T_{B_1 B_2} = T_{B_1 B_0}\, T_{B_0 B_2} = T_{B_0 B_1}^{-1}\, T_{B_0 B_2}\,,}$$

de manera que ahora sólo tenemos que trabajar con las matrices $T_{B_0 B_1}$ y $T_{B_0 B_2}$ cuya obtención es sencilla. El precio a pagar es que hay que invertir una de ellas y efectuar una multiplicación entre dos matrices.

Parte VI

Tema 5. Transformaciones Lineales

Tema 5

Transformaciones lineales

5.1. Definiciones y propiedades

Comencemos el tema recordando terminología y notación conocidas: una *función* real de una variable real x, $f : A \to \mathbb{R}$, transforma cada número x del *dominio* de f (un subconjunto propio $A \subset \mathbb{R}$ o el propio $\mathbb{R}$) en exactamente un número $f(x) \in \mathbb{R}$. En otras palabras, podríamos decir que la función f transforma un vector $\mathbf{x}$ (puesto que $\mathbb{R}$ es un espacio vectorial) en otro vector $\mathbf{y} = f(\mathbf{x})$ de la *imagen* de f y este significado se refleja mediante el uso de la notación:

$$\begin{array}{rccc} f: & \mathbb{R} & \to & \mathbb{R} \\ & x & \mapsto & f(x)\,. \end{array}$$

De manera similar, una función real f de dos variables reales transforma vectores $(x, y)^t$ del dominio ($\mathbb{R}^2$ o un subconjunto de $\mathbb{R}^2$) en vectores de la imagen ($\mathbb{R}$ o un subconjunto de $\mathbb{R}$) y escribiremos:

$$\begin{array}{rccc} f: & \mathbb{R}^2 & \to & \mathbb{R} \\ & (x, y)^t & \mapsto & f(x, y)\,. \end{array}$$

Sin embargo, genericamente las funciones que hemos estudiado en Cálculo no respetan las operaciones naturales de los espacios vectoriales. Por ejemplo, la función $x \mapsto f(x) = x^2$ no respeta las combinaciones lineales ya que (si $x_1\, x_2 \neq 0$):

$$f(x_1 + x_2) = (x_1 + x_2)^2 = x_1^2 + x_2^2 + 2x_1x_2 \neq f(x_1) + f(x_2) = x_1^2 + x_2^2 \,.$$

Como veremos un poco más adelante, las funciones que respetan las operaciones naturales de los espacios vectoriales reciben el nombre de transformaciones lineales.

En general, dados dos espacios vectoriales V y W sobre un mismo cuerpo $\mathbb{K}$, una función T de V en W es una regla que asigna a cada vector $\boldsymbol{v} \in V$ un único vector $\boldsymbol{w} \in W$. Representamos esta función mediante $T\colon V \to W$.

Si una función $T\colon V \to W$ satisface la siguiente propiedad:

(PL) $T(\alpha_1 \boldsymbol{v}_1 + \alpha_2 \boldsymbol{v}_2) = \alpha_1 T(\boldsymbol{v}_1) + \alpha_2 T(\boldsymbol{v}_2)$ para todo $\boldsymbol{v}_1, \boldsymbol{v}_2 \in V$ y todo $\alpha_1, \alpha_2 \in \mathbb{K}$

se denomina **transformación lineal** (o función lineal u **homomorfismo**).

Equivalentemente, es directo comprobar que una función T es lineal si y sólo si satisface las siguientes propiedades:

i) $T(\boldsymbol{v}_1 + \boldsymbol{v}_2) = T(\boldsymbol{v}_1) + T(\boldsymbol{v}_2)$ para todo $\boldsymbol{v}_1, \boldsymbol{v}_2 \in V$.

ii) $T(\alpha \boldsymbol{v}_1) = \alpha T(\boldsymbol{v}_1)$ para todo $\boldsymbol{v}_1 \in V$ y todo $\alpha \in \mathbb{K}$.

Como en el caso de las bien conocidas funciones reales, dada $T : V \to W$, usamos los siguientes términos:

El espacio vectorial V es el **dominio** de T.

El conjunto de todos los valores $T(\boldsymbol{v})$ se denomina **imagen** de T. Denotaremos este conjunto por $\mathrm{Im}(T)$.

En ocasiones se utiliza el término **operador lineal** para referirse a una transformación lineal T, especialmente cuando $V = W$.

Ejemplo

Consideremos las siguientes transformaciones:

- $T_1 : \mathbb{P}_2 \to \mathbb{P}_2$, definida por $T_1(a_0 + a_1x + a_2x^2) = a_2\, x^2$.

  T_1 es lineal; para demostrarlo, consideremos los polinomios $p, q \in \mathbb{P}_2$ y el escalar α. Es inmediato ver que se cumplen las propiedades de linealidad. En efecto:

$$\begin{aligned} T_1(\underbrace{a_0 + a_1x + a_2x^2}_{p(x)} + \underbrace{b_0 + b_1x + b_2x^2}_{q(x)}) &= T_1(\underbrace{a_0 + b_0 + (a_1 + b_1)x + (a_2 + b_2)x^2}_{(p+q)(x)}) \\ &= (a_2 + b_2)x^2 = a_2x^2 + b_2x^2 \\ &= T_1(p(x)) + T_1(q(x)) \,. \end{aligned}$$

$$\begin{aligned} T_1(\alpha\, p(x)) &= T_1(\alpha(a_0 + a_1x + a_2x^2)) = T_1(\alpha\, a_0 + \alpha\, a_1x + \alpha\, a_2x^2) \\ &= \alpha\, a_2x^2 = \alpha\, T_1(p(x)) \,. \end{aligned}$$

  Además, es fácil entender que la imagen de T está formada por los polinomios de grado exactamente dos con un solo término junto al polinomio 0.

- $T_2 : \mathbb{P}_2 \to \mathbb{R}$, definida por $T_2(a_0+a_1x+a_2x^2) = a_M$ donde $a_M = \text{máx}(a_0, a_1, a_2)$.

  T_2 no es lineal, lo cual es de esperar ya que intuitivamente, la función que nos devuelve el máximo de un conjunto de números no lo es. Para comprobarlo, consideremos por ejemplo los polinomios $p(x) = x^2$ y $q(x) = x^2 + 2x$; tenemos que:

$$T_2(p(x) + q(x)) = T_2(2x^2 + 2x) = 2\,,$$

  mientras que

$$T_2(p(x)) + T_2(q(x)) = 1 + 2 = 3\,.$$

La siguiente definición introduce un conjunto de gran importancia en el estudio de las transformaciones lineales.

Si $T : V \to W$ es una transformación lineal y 0 es el vector cero de W, entonces el conjunto de vectores $v \in V$ para los que $T(v) = 0$ se denomina **núcleo de** T o *espacio nulo de* T o *kernel de* T y lo denotamos por $\ker(T)$.

Obsérvese que, dada una transformación lineal $T : V \to W$, hemos dado "nombre propio" a un subconjunto de V -el núcleo- y a un subconjunto de W -la imagen-. Esto no es casual, ya que, como veremos, ambos juegan un papel fundamental en temas posteriores.

Las siguientes definiciones, conocidas de cursos previos, nos permiten identificar ciertos tipos de transformaciones lineales.

Si $T : V \to W$ es una transformación lineal entonces T se dirá **inyectiva** si para todo par de vectores diferentes v_1, v_2 ($v_1 \neq v_2$) de V los vectores $T(v_1)$ y $T(v_2)$ de W también son distintos.

Si la imagen de T es W diremos que T es **suprayectiva**.

Una transformación lineal T inyectiva y suprayectiva se denomina **isomorfismo** entre los espacios V y W; en tal caso, se dice que V y W son **isomorfos**.

Un **automorfismo** es un isomorfismo de un espacio vectorial en sí mismo.

Obsérvese que si T es lineal y no inyectiva, existen $v_1 \neq v_2$ tales que $T(v_1) = T(v_2)$ o, lo que es lo mismo, $T(v_1 - v_2) = 0$. Como $v_1 \neq v_2$, $v = v_1 - v_2 \neq 0$ por lo que una $\boxed{\text{transformación lineal es inyectiva si y sólo si } \ker(T) = \{0\}.}$

Además, dada una transformación lineal, se pueden definir otras transformaciones lineales asociadas con ella. En este tema introducimos la siguiente:

Si $T : V \to W$ es inyectiva entonces la **transformación inversa** $T^{-1} : W' \to V$, donde $W' = \mathrm{Im}(T) \subset W$, se define como sigue: para cada par $\boldsymbol{v} \in V$ y $\boldsymbol{w} \in W'$ con $\boldsymbol{w} = T(\boldsymbol{v})$, es

$$T^{-1}(\boldsymbol{w}) = \boldsymbol{v}\,.$$

Veamos algunas propiedades sencillas derivadas de las definiciones anteriores:

Teorema

La imagen de la transformación lineal $T : V \to W$ es un espacio vectorial.

Demostración. Recordemos que, cuando S es un subconjunto de un espacio vectorial, para demostrar que S es un subespacio basta con probar que es cerrado bajo las operaciones de suma y producto por escalares. En efecto, en este caso tenemos que para todo $\boldsymbol{w}_1, \boldsymbol{w}_2 \in \mathrm{Im}(T)$ existen $\boldsymbol{v}_1, \boldsymbol{v}_2 \in V$ tales que

$$\boldsymbol{w}_1 + \boldsymbol{w}_2 = T(\boldsymbol{v}_1) + T(\boldsymbol{v}_2) \underset{\substack{\uparrow \\ \text{prop. lineal}}}{=} T(\boldsymbol{v}_1 + \boldsymbol{v}_2)$$

y, puesto que V es un espacio vectorial, $\boldsymbol{v}_1 + \boldsymbol{v}_2 \in V$; por tanto:

$$T(\boldsymbol{v}_1 + \boldsymbol{v}_2) \in \mathrm{Im}(T) \quad \Rightarrow \quad \boldsymbol{w}_1 + \boldsymbol{w}_2 \in \mathrm{Im}(T) \quad \text{(cerrado bajo la suma).}$$

Por otro lado, $\forall\, \boldsymbol{w} \in \mathrm{Im}(T)$, existe algún $\boldsymbol{v} \in V$ tal que $T(\boldsymbol{v}) = \boldsymbol{w}$ y $\forall\, \alpha \in \mathbb{R}$ tenemos:

$$\alpha\,\boldsymbol{w} = \alpha\,T(\boldsymbol{v}) \underset{\substack{\uparrow \\ \text{prop. lineal}}}{=} T(\alpha\,\boldsymbol{v})$$

y como $\alpha\,\boldsymbol{v} \in V$ resulta que

$$T(\alpha\,\boldsymbol{v}) \in \mathrm{Im}(T) \quad \Rightarrow \quad \alpha\,\boldsymbol{w} \in \mathrm{Im}(T) \quad \text{(cerrado bajo el producto por escalares).}$$

En consecuencia, $\mathrm{Im}(T)$ es un espacio vectorial. $\square$

Teorema

El núcleo de una transformación lineal $T : V \to W$ es un espacio vectorial.

Demostración. Puesto que $\ker(T)$ es un subconjunto del espacio vectorial V basta con probar que es cerrado con respecto a la suma y al producto por escalares para garantizar que es un espacio vectorial (puesto que sería subespacio de V).

Consideremos $\mathbf{u}, \mathbf{v} \in \ker(T)$ y $\alpha \in \mathbb{R}$. Tenemos que probar que $\mathbf{u} + \mathbf{v} \in \ker(T)$ y que $\alpha\mathbf{u} \in \ker(T)$. En efecto,

$$\begin{aligned}
T(\mathbf{u}+\mathbf{v}) &\underset{\substack{\uparrow \\ T \text{ lineal}}}{=} T(\mathbf{u}) + T(\mathbf{v}) \underset{\substack{\uparrow \\ \mathbf{u},\mathbf{v}\in\ker(T)}}{=} 0 + 0 = 0 \quad \Rightarrow \quad \mathbf{u}+\mathbf{v} \in \ker(T)\,, \\
T(\alpha\,\mathbf{u}) &\underset{\substack{\uparrow \\ T \text{ lineal}}}{=} \alpha\, T(\mathbf{u}) \underset{\substack{\uparrow \\ \mathbf{u}\in\ker(T)}}{=} \alpha\, 0 = 0 \quad \Rightarrow \quad \alpha\,\mathbf{u} \in \ker(T)
\end{aligned}$$

y esto completa la demostración. □

Los teoremas anteriores permiten formular las siguientes definiciones, que utilizaremos con frecuencia en el futuro.

La dimensión del espacio vectorial “imagen de la transformación lineal T” se denomina **rango de** T y se denota por $\mathrm{rg}(T)$.

La dimensión del espacio vectorial “núcleo de la transformación lineal T” se denomina **nulidad de** T y se representa por $\mathrm{nul}(T)$.

Obsérvese que la elección del término “rango” no es casual; en temas posteriores veremos la relación de éste con el rango de ciertas matrices.

El siguiente teorema, que se da sin demostración, relaciona ambas dimensiones.

Teorema del rango

Si $T : V \to W$ es una transformación lineal y el espacio vectorial V tiene dimensión

n, entonces:

$$\boxed{\mathrm{rg}(T) + \mathrm{nul}(T) = n.}$$

Nótese que $\mathrm{rg}(T)$ es la dimensión de la imagen de T (un subespacio vectorial en W), mientras que $\mathrm{nul}(T)$ es la dimensión del núcleo de T (un subespacio vectorial en V).

Ejemplo

La transformación $T_1 : \mathbb{P}_2 \to \mathbb{P}_2$, definida por $T_1(a_0 + a_1x + a_2x^2) = a_2x^2$, es lineal como vimos anteriormente. Es inmediato observar que la imagen de T es el espacio generado por el vector (polinomio) x^2, lo cual escribimos:

$$\mathrm{Im}(T) = \mathrm{Gen}(x^2)\,.$$

Por otro lado, para calcular el núcleo, tenemos que:

$$\begin{aligned}
\ker(T) &= \{p(x) \in \mathbb{P}_2 \colon T(p) = 0\} = \left\{a_0 + a_1x + a_2x^2 \in \mathbb{P}_2 \colon T(a_0 + a_1x + a_2x^2) = 0\right\} \\
&= \left\{a_0 + a_1x + a_2x^2 \in \mathbb{P}_2 \colon a_2x^2 = 0\right\} = \{a_0 + a_1x \in \mathbb{P}_2 \colon a_0, a_1 \in \mathbb{R}\} \\
&= \mathrm{Gen}(1, x)\,.
\end{aligned}$$

Obviamente $\mathrm{rg}(T) = 1$, $\mathrm{nul}(T) = 2$ y

$$\mathrm{nul}(T) + \mathrm{rg}(T) = 2 + 1 = 3 = \dim(\mathbb{P}_2)\,.$$

También es fácil ver que T no es inyectiva, ya que, por ejemplo, los polinomios x^2+1 y $x^2 - 1$ son distintos y sin embargo tienen la misma imagen x^2. Tampoco es suprayectiva, pues $\mathrm{Im}(T) \neq \mathbb{P}_2$.

Ejemplo

Sea $T : \mathbb{R}^3 \to \mathbb{R}^2$ la aplicación definida por:

$$T(\,(v_1, v_2, v_3)^t\,) = (v_1, v_2)^t\,.$$

Vamos a demostrar que T es una transformación lineal:

Consideremos $\mathbf{u}, \mathbf{v} \in \mathbb{R}^3$ y $\alpha \in \mathbb{R}$. Tenemos:

$$\begin{aligned} T(\mathbf{u}+\mathbf{v}) &= T((u_1+v_1, u_2+v_2, u_3+v_3)^t) = (u_1+v_1, u_2+v_2)^t \\ &= (u_1, u_2)^t + (v_1, v_2)^t = T(\mathbf{u}) + T(\mathbf{v})\,. \end{aligned}$$

También

$$\begin{aligned} T(\alpha\mathbf{u}) &= T(\alpha(u_1, u_2, u_3)^t) = T((\alpha u_1, \alpha u_2, \alpha u_3)^t) \\ &= (\alpha u_1, \alpha u_2)^t = \alpha(u_1, u_2)^t = \alpha T(\mathbf{u})\,, \end{aligned}$$

por lo que T es lineal.

Vamos a calcular ahora bases para los espacios vectoriales imagen y núcleo. Obviamente, $\mathrm{Im}(T) = \mathbb{R}^2$, $\mathrm{rg}(T) = 2$ y $B_0 = ((1,0)^t, (0,1)^t)$ es una base de la imagen de T.

Por otra parte, podemos calcular la nulidad de T mediante

$$n - \mathrm{rg}(T) = \mathrm{nul}(T) = 3 - 2 = 1\,.$$

Para determinar el núcleo, buscamos los vectores $(v_1, v_2, v_3)^t \in \mathbb{R}^3$ tales que $T((v_1, v_2, v_3)^t) = (v_1, v_2)^t = (0,0)^t$, es decir, el núcleo es el conjunto de vectores de la forma $(0, 0, v_3)^t, v_3 \in \mathbb{R}$. Una base de este subespacio viene dada por $B_1 = ((0,0,1)^t)$.

Finalmente, veamos si T puede ser un isomorfismo. Puesto que, por ejemplo, $(1,1,1)^t \neq (1,1,-1)^t$ y $T((1,1,1)^t) = (1,1)^t = T((1,1,-1)^t)$, concluimos que T no es inyectiva y por tanto no es isomorfismo; pero sí es suprayectiva, puesto que para cada $(u_1, u_2)^t \in \mathbb{R}^2$ hay al menos un vector de $\mathbb{R}^3$, por ejemplo $(u_1, u_2, 0)^t$, tal que $T(\,(u_1, u_2, 0)^t\,) = (u_1, u_2)^t$.

Las siguientes propiedades, muy intuitivas, son consecuencia directa de la linealidad.

Teorema

Sea T una transformación lineal $T : V \to W$.

a) Si $\mathbf{0} \in V$ es el vector cero de V entonces $T(\mathbf{0})$ es el vector cero de W.

b) Si $\boldsymbol{v} \in V$ y $-\boldsymbol{v}$ es el opuesto de $\boldsymbol{v}$ en V, entonces $T(-\boldsymbol{v})$ es el opuesto de $T(\boldsymbol{v})$ en W.

Demostración.

a) Sea $\boldsymbol{v} \in V$. Entonces

$$T(\mathbf{0}) \underset{\uparrow \atop \text{V espacio vectorial}}{=} T(0\,\boldsymbol{v}) \underset{\uparrow \atop \text{T lineal}}{=} 0\,T(\boldsymbol{v}) \underset{\uparrow \atop \text{Im(T) espacio vectorial}}{=} \mathbf{0}\,.$$

b) Sea $\boldsymbol{v} \in V$. Tenemos

$$T(\boldsymbol{v}) + T(-\boldsymbol{v}) \underset{\uparrow \atop \text{T lineal}}{=} T(\boldsymbol{v} + (-\boldsymbol{v})) \underset{\uparrow \atop \text{V esp. vectorial}}{=} T(\mathbf{0}) \underset{\uparrow \atop \text{prop. (a))}}{=} \mathbf{0}\,.$$

□

Teorema

Si $T : V \to W$ es una transformación lineal inyectiva y la imagen de T es W' (un subespacio de W), entonces T^{-1} es una transformación lineal de W' a V.

Demostración. Sean $\boldsymbol{w}_1, \boldsymbol{w}_2 \in W'$ y $\alpha \in \mathbb{K}$. Entonces, existen $\boldsymbol{v}_1, \boldsymbol{v}_2 \in V$ tales que $T(\boldsymbol{v}_1) =$

w_1, $T^{-1}(w_1) = v_1$ y $T(v_2) = w_2$, $T^{-1}(w_2) = v_2$. Primero, observemos que

$$\begin{aligned} T^{-1}(T(v_1)) &= T^{-1}(w_1) = v_1\,, \\ T(T^{-1}(w_1)) &= T(v_1) = w_1\,. \end{aligned}$$

Por tanto tenemos:

$$\begin{aligned} T^{-1}(w_1 + w_2) &\overset{w_1,w_2 \in W'}{\underset{\downarrow}{=}} T^{-1}(T(v_1) + T(v(2))) \overset{T \text{ lineal}}{\underset{\downarrow}{=}} T^{-1}(T(v_1 + v_2)) \\ &= v_1 + v_2 \underset{\substack{\uparrow \\ \text{def. de } T^{-1}}}{=} T^{-1}(w_1) + T^{-1}(w_2) \end{aligned}$$

y

$$T^{-1}(\alpha w_1) = T^{-1}(\alpha T(v_1)) \underset{\substack{\uparrow \\ T \text{ lineal}}}{=} T^{-1}(T(\alpha v_1)) = \alpha v_1 = \alpha T^{-1}(w_1)$$

y por tanto T^{-1} es lineal. $\square$

Teorema

Si $T : V \to W$ es un isomorfismo entre V y W y $(v_1, v_2, \ldots, v_n)$ es una base del espacio vectorial V, entonces $(T(v_1), T(v_2), \ldots, T(v_n))$ es una base del espacio vectorial W.

Demostración. Primero igualamos a cero una combinación lineal arbitraria de los vectores $T(v_1), \ldots, T(v_n)$:

$$0 = \alpha_1 T(v_1) + \cdots + \alpha_n T(v_n) \underset{\substack{\uparrow \\ T \text{ lineal}}}{=} T(\alpha_1 v_1 + \cdots + \alpha_n v_n)\,.$$

Puesto que $T(0) = 0$ y T es un isomorfismo:

$$\alpha_1 v_1 + \cdots + \alpha_n v_n = 0 \underset{\substack{\uparrow \\ v_1,\ldots,v_n \text{ lin. ind.}}}{\Rightarrow} \alpha_1 = \cdots = \alpha_n = 0$$

y $T(v_1), \ldots, T(v_n)$ son linealmente independientes. Además, si consideramos un $w \in W$ arbitrario, tenemos para algún $v \in V$:

$$w = T(v) \underset{\substack{\uparrow \\ v_1,\ldots,v_n \text{ base}}}{=} T(\beta_1 v_1 + \cdots + \beta_n v_n) \underset{\substack{\uparrow \\ T \text{ lineal}}}{=} \beta_1 T(v_1) + \cdots + \beta_n T(v_n)\,,$$

por lo que w puede expresarse como combinación lineal de los vectores $T(v_1), \ldots, T(v_n)$; por tanto, $(T(v_1), T(v_2), \ldots, T(v_n))$ es una base de W. $\square$

De aquí se deduce que dos espacios vectoriales isomorfos tienen la misma dimensión.

5.1.1. Álgebra de las transformaciones lineales

A continuación pasamos a describir algunas transformaciones lineales especialmente importantes y cómo realizar ciertas operaciones.

Proposición

- La aplicación IDENTIDAD I, definida de V en V mediante $I(v) = v$ es una aplicación lineal.
- La aplicación NULA 0, definida de V en W mediante $0(v) = 0$ es una aplicación lineal.

Proposición

Sean T_1, T_2 dos aplicaciones lineales de V en W y sea $\alpha \in \mathbb{K}$ un escalar.

- La aplicación SUMA, definida de V en W mediante:
$$(T_1 + T_2)(v) = T_1(v) + T_2(v)$$
es una aplicación lineal.
- La aplicación PRODUCTO POR EL ESCALAR $\alpha \in \mathbb{K}$, definida de V en W mediante:
$$(\alpha T_1)(v) = \alpha\, T_1(v)$$
es una aplicación lineal.

Es fácil ver que:

Proposición

Dados dos espacios vectoriales, V y W, el conjunto de todas las transformaciones lineales $T : V \to W$, con la suma y el producto por escalares de $\mathbb{K}$, es un espacio vectorial sobre $\mathbb{K}$.

Adicionalmente, necesitaremos la siguiente noción, que es conocida de cursos anteriores.

Si $T_1 : U \to V$ y $T_2 : V \to W$ son transformaciones lineales, respectivamente, de U en V y de V en W, entonces la **composición**, denotada por $T_2 \circ T_1$, es una transformación $U \to W$, definida para $\mathbf{u} \in U$ mediante:

$$(T_2 \circ T_1)(\mathbf{u}) = T_2(T_1(\mathbf{u})) .$$

Obsérvese que $T_1(\mathbf{u})$ es un elemento de V por lo que T_2 hace corresponder a éste el valor $T_2(T_1(\mathbf{u}))$ de W.

Teorema

Dadas dos transformaciones lineales $T_1 : U \to V$ y $T_2 : V \to W$:

a) La composición $T_2 \circ T_1$ es una transformación lineal de $U \to W$.

b) Si T_2 y T_1 son inyectivas entonces $T_2 \circ T_1$ es inyectiva.

c) Si T_2 y T_1 son suprayectivas entonces $T_2 \circ T_1$ es suprayectiva.

Ejemplo

Consideremos de nuevo $T : \mathbb{R}^3 \to \mathbb{R}^2$ definida por:

$$T(\,(v_1, v_2, v_3)^t\,) = (v_1, v_2)^t$$

y sea $S : \mathbb{R}^2 \to \mathbb{R}^3$ la transformación definida por

$$S((w_1, w_2)^t) = (w_1, w_2, 0)^t .$$

Vamos a comprobar que S es lineal y a describir las propiedades de $S \circ T$ (núcleo y base, nulidad, imagen y base, rango y si es o no inyectiva, suprayectiva o isomorfismo).

- Consideremos $\mathbf{u}, \mathbf{v} \in \mathbb{R}^2$ y $\alpha \in \mathbb{R}$:

$$\begin{aligned} S(\mathbf{u}+\mathbf{v}) &= T((u_1+v_1, u_2+v_2)^t) = (u_1+v_1, u_2+v_2, 0)^t \\ &= (u_1, u_2, 0)^t + (v_1, v_2, 0)^t = S(\mathbf{u}) + S(\mathbf{v})\,, \\ S(\alpha u) &= S(\alpha(u_1, u_2)^t) = S((\alpha u_1, \alpha u_2)^t) \\ &= (\alpha u_1, \alpha u_2, 0)^t = \alpha(u_1, u_2, 0)^t = \alpha S(\mathbf{u})\,, \end{aligned}$$

  por lo que S es lineal.

- La composición $S \circ T$ es

$$(S \circ T)((u_1, u_2, u_3)^t) = (u_1, u_2, 0)^t\,.$$

  Entonces, $rg(S \circ T) = 2$ y $B_0 = ((1,0,0)^t, (0,1,0)^t)$ es una base de la imagen de la composición.

- La nulidad de $S \circ T$ se calcula con

$$n - rg(S \circ T) = nul(S \circ T) = 3 - 2 = 1\,.$$

  El núcleo está formado por los vectores $(v_1, v_2, v_3)^t \in \mathbb{R}^3$ tales que

$$(S \circ T)((v_1, v_2, v_3)^t) = (v_1, v_2, 0)^t = (0,0,0)^t\,.$$

  Es decir, el núcleo es el conjunto de vectores de la forma $(0, 0, v_3)^t, v_3 \in \mathbb{R}$. Una base de este subespacio viene dada por $B_1 = ((0,0,1)^t)$.

- Finalmente tenemos de nuevo que $S \circ T$ no puede ser un isomorfismo porque $\dim(\mathbb{R}^3) \neq \mathrm{rg}(S \circ T)$. Si consideramos los vectores $(1,1,1)^t \neq (1,1,-1)^t$, resulta que $(S \circ T)((1,1,1)^t) = (1,1,0)^t = (S \circ T)((1,1,-1)^t)$ y de aquí se deduce que $S \circ T$ no es inyectiva, pero sí es suprayectiva, puesto que para cada $(u_1,u_2,0)^t \in \mathbb{R}^3$ existe al menos un vector de $\mathbb{R}^3$, por ejemplo el propio $(u_1,u_2,0)^t$, tal que $(S \circ T)((u_1,u_2,0)^t) = (u_1,u_2,0)^t$.

TRANSFORMACIONES (aplicaciones) LINEALES

1. Definiciones. Ejemplos.

Definición.- *"Dados dos espacios vectoriales E y F sobre el mismo cuerpo conmutativo K, se llama transformación lineal de E en F, toda aplicación T que verifica las dos condiciones siguientes:*

(1) $\forall\ x, y \in E, T(x+y) = T(x) + T(y)$

(2) $\forall\ x \in E, \forall\ \alpha \in K, T(\alpha x) = \alpha T(x)$

2. Propiedades fundamentales de las transformaciones lineales.

Teorema 1.- *"Sea T una transformación lineal de E y F; se verifica:*

a) La imagen del vector nulo de E es el vector nulo de F: $T(o_E) = o_F$.

b) $\forall\ x \in E,\ T(-x) = -T(x)$

c) $\forall\ x, y \in E, T(x-y) = T(x) - T(y)$"

Teorema 2.- *"Para que una transformación T entre espacios vectoriales E y F sobre el mismo cuerpo conmutativo K sea lineal, es necesario y suficiente que:*

$$\forall\ x, y \in E, \quad \forall\ \alpha, \beta \in K; T(\alpha x + \beta y) = \alpha T(x) + \beta T(y)$$

Definición y teorema 3.- *"Sea T una transformación lineal de E en F, y A un subconjunto de E entre espacios y B un subconjunto de F, se llama* ***conjunto imagen de A por T*** *y se escribe T(A) el subconjunto de F que está descrito por T(x) cuando x describe A:*

$$T(A) = \{T(x) : x \in A\} \;.$$

Se llama ***imagen recíproca de B por T,*** *y se escribe* $T^{-1}(B)$ *al subconjunto de E definido por:*

$$T^{-1}(B) = \{x \in E : T(x) \in B\}\,.$$

a) Si A es un s.e.v. de E, T(A) es un s.e.v. de F.

b) Si B es un s.e.v. de F, $T^{-1}(B)$ *es un s.e.v. de E"* .

3. Núcleo e imagen de una transformación lineal

Definición .- *"Sea T una transformación lineal de E en F, se define* ***núcleo de la transformación T*** *y se le representa por KerT, al subespacio:*

$$\{x \in E \;/\; T(x) = 0 \in F\} = T^{-1}(0)\,.$$

Se define ***imagen de la la transformación lineal T*** *y se representa por ImT, al subespacio:*

$$\{x' \in F \;/\; \exists\; x \in E, T(x) = x'\} = T(E)$$

Corolario .- *"Si T es transformación lineal de E en F:*

1. T es inyectiva $\Leftrightarrow$ *KerT = {0}*

2. T es suprayectiva $\Leftrightarrow$ *ImT = F"*

4. Imágenes de partes de E

Teorema 4.- *"Si T es una transformación lineal de E en F, si G es un sistema generador de E, entonces T(G) es sistema generador de ImT"*

5. Rango de una transformación lineal

Teorema 5 y definición.- *"Si T es una transformación lineal de E en F de dimensiones finitas sobre K, ImT es de dimensión finita sobre K y:*

$$\dim_K \operatorname{Im} T \leq \dim_K E \,.$$

La dimensión de ImT es el rango de la transformación T, se representa por rg(T), y además:

$$rg(T)=dim_K E - dim_K KerT$$
ó bien
$$dim_K E = dim_K KerT + dim_K ImT"$$

6. Matriz asociada a una transformación lineal

Lema y definición.- *"Si T es una transformación lineal de E en F de dimensiones finitas, n y m, sobre . Sean* $B_n = \{v_1,\ldots,v_n\}$ *y* $B_m = \{u_1,\ldots,u_m\}$ *bases, respectivamente, de E y F, entonces la matriz:*

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} = \left(a_{ij}\right);\ i = 1,\ldots,m \quad j = 1,\ldots,n$$

$$A = M(T)_{B_m \leftarrow B_n} = \left[\left[T(v_1)\right]_{B_m}, \ldots, \left[T(v_n)\right]_{B_m} \right] \in \mathbb{R}^{m\times n}$$

O bien, de otra manera escrito:

$$\left[T(v_1) \quad \cdots \quad T(v_n)\right] = [u_1 \quad \cdots \quad u_m] \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}$$

Verifica que: $\left[T(x)\right]_{B_m} = A[x]_{B_n} \quad \forall x \in E$. *Puesto que si* : $x = x_1v_1 + \cdots + x_nv_n = [v_1 \quad \cdots \quad v_n] \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ *entonces, se verificará:*

Por una parte: $[y]_{B_m} = \left[T(x)\right]_{B_m} = \left[T(v_1) \quad \cdots \quad T(v_n)\right] \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = [u_1 \quad \cdots \quad u_m] A \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$.

Por otra: $[y]_{B_m} = [u_1 \quad \cdots \quad u_m] \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}$. Identificando igualdades se obtiene esa identidad.

Corolario.- Casos especiales

- *Dimensiones de espacios de salida y llegada iguales.* $T : \mathbb{R}^n \to \mathbb{R}^n$. $M(T)_{B_n} \equiv M(T)_{B_n \leftarrow B_n}$
- *Matriz canónica o estándar: Las dos bases son las canónicas, entonces se verifica* $T(x) = Ax$
- *La aplicación identidad* $id : E \to E$ *entre el mismo espacio vectorial es una transformación lineal. La matriz asociada cuando las base de los espacios de partida y llegada coinciden es la matriz unidad de orden la dimensión de E. Si las bases no*

coinciden, la matriz se denomina MATRIZ CAMBIO DE BASE: $P_{B_2 \leftarrow B_1} = M(id)_{B_2 \leftarrow B_1}$

Ejemplo 1: Dada la transformación lineal definida como:

$$T:\mathbb{R}^2 \to \mathbb{R}^3; \quad T\left(\begin{bmatrix} x_1 & x_2 \end{bmatrix}^T\right) = \begin{bmatrix} 2x_1 - x_2 & x_2 & x_1 + x_2 \end{bmatrix}^T.$$

Calcular la matriz asociada respecto de las bases $B_2 = \left\{\begin{bmatrix} 1 & 2 \end{bmatrix}^T, \begin{bmatrix} 1 & 1 \end{bmatrix}^T\right\}$ y $B_3 = \left\{\begin{bmatrix} -1 & 1 & 1 \end{bmatrix}^T, \begin{bmatrix} 2 & -1 & -1 \end{bmatrix}^T, \begin{bmatrix} -1 & 0 & 1 \end{bmatrix}^T\right\}$

$$\begin{bmatrix} T(v_1) & T(v_2) \end{bmatrix} = \begin{bmatrix} 2 & -1 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 2 & 1 \\ 3 & 2 \end{bmatrix}$$

respecto de la base canónica del espacio de llegada. Ahora:

$$\begin{bmatrix} M_{B_3^c \leftarrow B_3} \vdots T(v_1) & T(v_2) \end{bmatrix} = \begin{bmatrix} -1 & 2 & -1 & 0 & 1 \\ 1 & -1 & 0 & 2 & 1 \\ 1 & -1 & 1 & 3 & 2 \end{bmatrix} \sim \begin{bmatrix} -1 & 2 & -1 & 0 & 1 \\ 0 & 1 & -1 & 2 & 2 \\ 0 & 1 & 0 & 3 & 3 \end{bmatrix} \sim$$

$$\begin{bmatrix} -1 & 0 & 1 & -4 & -3 \\ 0 & 1 & -1 & 2 & 2 \\ 0 & 0 & -1 & -1 & -1 \end{bmatrix} \sim \begin{bmatrix} -1 & 0 & 0 & -5 & -4 \\ 0 & 1 & 0 & 3 & 3 \\ 0 & 0 & -1 & -1 & -1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 & 5 & 4 \\ 0 & 1 & 0 & 3 & 3 \\ 0 & 0 & 1 & 1 & 1 \end{bmatrix} \sim$$

$$\sim \begin{bmatrix} I_3 \vdots \begin{bmatrix} T(v_1) \end{bmatrix}_{B_3}, \begin{bmatrix} T(v_1) \end{bmatrix}_{B_3} \end{bmatrix} \sim \begin{bmatrix} I_3 \vdots M(T)_{B_3 \leftarrow B_2} \end{bmatrix}$$

Así pues, la matriz pedida es

$$A = M(T)_{B_3 \leftarrow B_2} = \begin{bmatrix} 5 & 4 \\ 3 & 3 \\ 1 & 1 \end{bmatrix} \in \mathbb{R}^{3\times 2}$$

Ejemplo 2: En el espacio vectorial $P_2[x]$ de los polinomios de grado menor o igual a 2 con indeterminada x $\left(\dim P_2[x] = 3\right)$, con bases $B_1 = \left\{1, x, x^2\right\}$ canónica, y $B_2 = \left\{1+x, x+x^2, 1+x^2\right\}$, la pregunta es: "Hállense las coordenadas de un vector (polinomio) $p(x) = a_0 + a_1 x + a_2 x^2$ respecto de la segunda base hallando la matriz del cambio”

$$id : P_2[x] \to P_2[x]; \quad \left[id(p) \right]_{B_1} = \left[id\left(\begin{bmatrix} a_0 & a_1 & a_2 \end{bmatrix}^T \right) \right]_{B_1} = \left[\begin{bmatrix} a_0' & a_1' & a_2' \end{bmatrix}^T \right]_{B_2} = [p]_{B_2}$$

.

Si la base del espacio de llegada es la base estándar, el problema es directo. Puesto que: $[1+x]_{B_1} = \begin{bmatrix} 1 & 1 & 0 \end{bmatrix}^T$; $\left[x + x^2 \right]_{B_1} = \begin{bmatrix} 0 & 1 & 1 \end{bmatrix}^T$; $\left[1 + x^2 \right]_{B_1} = \begin{bmatrix} 1 & 0 & 1 \end{bmatrix}^T$; se obtiene directamente

$$P_{B_1 \leftarrow B_2} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$$

Ahora, la matriz pedida, que es $P_{B_2 \leftarrow B_1}$ será la inversa de la calculada:

$$\left[P_{B_1 \leftarrow B_2} \; \vdots \; I_3 \right] \equiv \begin{bmatrix} 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & -1 & -1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix} \sim$$

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

$$\sim \begin{bmatrix} 1 & 0 & 0 & 1/2 & 1/2 & -1/2 \\ 0 & 1 & 0 & -1/2 & 1/2 & 1/2 \\ 0 & 0 & 1 & 1/2 & -1/2 & 1/2 \end{bmatrix} \equiv \left[I_3 \; \vdots \; P_{B_2 \leftarrow B_1} \right]$$

Luego la matriz del cambio es $P_{B_2 \leftarrow B_1} = \dfrac{1}{2} \begin{bmatrix} 1 & 1 & -1 \\ -1 & 1 & 1 \\ 1 & -1 & 1 \end{bmatrix}$

Ahora, si p(x) = 1 + 2x – x^2, las coordenadas de p respecto de la base B_2, serán:

$$\left[p(x)\right]_{B_1}=\begin{bmatrix}1\\2\\-1\end{bmatrix};\quad \left[p(x)\right]_{B_2}=P_{B_2\leftarrow B_1}\cdot\left[p(x)\right]_{B_1}=$$

$$=\frac{1}{2}\begin{bmatrix}1&1&-1\\-1&1&1\\1&-1&1\end{bmatrix}\begin{bmatrix}1\\2\\-1\end{bmatrix}=\begin{bmatrix}2\\0\\-1\end{bmatrix}$$

7. Espacios nulos, espacios columna de una matriz

Definición 1 (Espacio nulo y Nulidad).- *"Si A es una matriz de $K^{m\times n}$, asociada a una transformación lineal T de E en F de dimensiones finitas, n y m, sobre K . Se define* ***espacio nulo de A*** *como:*

$$NulA = KerA = \left\{Soluciones\ de\ Ax = 0\right\} = KerT$$

y ***nulidad de A*** *a* $\dim NulA$ *"*

Definición 2 y teorema (Espacio columna).- *"Si A es una matriz de $K^{m\times n}$, asociada a una transformación lineal T de E en F de dimensiones finitas, n y m, sobre K . Se define:* $ColA = Gen\left\{a_1,\ldots,a_n\right\} = \operatorname{Im} T$ *s.e.v. de F. Luego se verifica:*

- *dim ColA = rg(T) = rg(A)*
- *dim ColA = dim K^n – dim Nul A*
- $n = rg(A) + \dim NulA$ *"*

Definición 3 (Espacio fila).- *"Si A es una matriz de $K^{m\times n}$, asociada a una transformación lineal T de E en F de dimensiones finitas, n y m, sobre K . Se define:* $FilA = ColA^T$ *s.e.v. de K^m"*

OBERVACIONES:

- Si $A \sim B$,entonces, evidentemente, $FilA = FilB$
- No hay una relación simple entre $FilA = NulA$, aunque son ambos s.e.v. de E.
- $rg\left(A^T\right) = \dim ColA^T = \dim FilA$ y $m = rg\left(A^T\right) + \dim NulA^T$

EJEMPLO: Cálculo de bases del núcleo (espacio nulo), espacio columna y espacio fila de una matriz

$$A = \begin{bmatrix} -2 & -5 & 8 & 0 & -17 \\ 1 & 3 & -5 & 1 & 5 \\ 3 & 11 & -19 & 7 & 1 \\ 1 & 7 & -13 & 5 & -3 \end{bmatrix} \sim \begin{bmatrix} 1 & 3 & -5 & 1 & 5 \\ 0 & 1 & -2 & 2 & -7 \\ 0 & 0 & 0 & -4 & 20 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & -2 & 0 & 3 \\ 0 & 0 & 0 & 1 & -5 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Tengamos en cuenta que:

$$A = \begin{bmatrix} c_1 & c_2 & c_3 & c_4 & c_5 \end{bmatrix} = \begin{bmatrix} f_1 \\ f_2 \\ f_3 \\ f_4 \end{bmatrix}$$

Luego:

$ColA = Gen\{c_1, c_2, c_4\}$, $FilA = Gen\{f_1, f_2, f_3\}$ $rg(A) = rg\left(A^T\right) = 3$

En cuanto al núcleo, el sistema homogéneo equivalente es:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} -x_3 - x_5 \\ 2x_3 - 3x_5 \\ x_3 \\ 5x_5 \\ x_5 \end{bmatrix} = x_3 \begin{bmatrix} -1 \\ 2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} -1 \\ -3 \\ 0 \\ 5 \\ 1 \end{bmatrix} = x_1 u_1 + x_2 u_2$$, luego:

$NulA = Gen\{u_1, u_2\}$ con dimensión 2.

Parte VII

Tema 6. Transformaciones Lineales y Matrices

Tema 6

Transformaciones lineales y matrices

6.1. Transformaciones lineales de $\mathbb{R}^n$ en $\mathbb{R}^m$

Comenzamos el tema estudiando un caso de transformaciones lineales particularmente sencillas: las que van del espacio vectorial $\mathbb{R}^n$ en $\mathbb{R}^m$. Vamos a describir algunas propiedades básicas que en secciones posteriores generalizaremos a transformaciones en otros espacios vectoriales. Recordemos que los vectores $v \in \mathbb{R}^n$ son de la forma

$$\begin{aligned} v &= (v_1, v_2, \ldots, v_n)^t = v_1(1,0,\ldots,0)^t + v_2(0,1,\ldots,0)^t + \cdots + v_n(0,0,\ldots,1)^t \\ &= v_1 e_1 + v_2 e_2 + \cdots + v_n e_n \,, \end{aligned}$$

donde e_i son los vectores de la base canónica. De manera natural se deduce el siguiente resultado.

Teorema

Si A es una matriz $m \times n$ y $T : \mathbb{R}^n \to \mathbb{R}^m$ está definida mediante $T(v) = A\,v$ para todo $v \in \mathbb{R}^n$; entonces T es una aplicación lineal.

Demostración. Puesto que A tiene dimensión $m \times n$, para cualquier vector $v \in \mathbb{R}^n$ tenemos que el producto $A\,v$ existe y es un vector m-dimensional, es decir, $T(v) = A\,v \in \mathbb{R}^m$.

Sean $v_1, v_2 \in \mathbb{R}^n$ y λ un escalar. Como

$$\begin{aligned} T(v_1+v_2) &= A(v_1+v_2) = A(v_1)+A(v_2) \Rightarrow \boxed{T(v_1+v_2) = T(v_1)+T(v_2)\,,} \\ T(\lambda\, v_1) &= A(\lambda\, v_1) = \lambda\, A v_1 \Rightarrow \boxed{T(\lambda\, v_1) = \lambda\, T(v_1)\,,} \end{aligned}$$

entonces T es lineal. □

Vamos a denotar como T_A la transformación lineal T basada en la matriz A, definida mediante $T(v) = A\,v$. Adicionalmente, podemos probar el resultado inverso: para cada transformación lineal $T : \mathbb{R}^n \to \mathbb{R}^m$, existe una matriz A tal que $T(v) = A\,v$.

Teorema

Supongamos que $T : \mathbb{R}^n \to \mathbb{R}^m$ es una transformación lineal. Sea $(e_1, \ldots, e_n)$ la base canónica (ordenada) de $\mathbb{R}^n$ y sea

$$A = (T(e_1), \ldots, T(e_n))$$

la matriz $m \times n$ cuyas columnas son los vectores $T(e_1), \ldots, T(e_n)$ de $\mathbb{R}^m$. Entonces, para cada $u \in \mathbb{R}^n$:

$$T(u) = A\,u\,.$$

Demostración. Sea $u = (u_1, \ldots, u_n)^t \in \mathbb{R}^n$. Se tiene que

$$u = u_1 e_1 + \cdots + u_n e_n\,.$$

Por las propiedades de linealidad de T resulta que

$$T(u) = u_1 T(e_1) + \cdots + u_n T(e_n)\,,$$

que es una combinación lineal de las columnas de A, lo que podemos expresar de la forma

$$T(u) = A\,u\,,$$

como queríamos demostrar. □

Así, **fijadas las bases canónicas** $B_{0,n}$ y $B_{0,m}$ de $\mathbb{R}^n$ y $\mathbb{R}^m$, respectivamente, a cada transformación lineal T le corresponde una única matriz, que denotamos por A_T y a cada matriz A le corresponde una única transformación lineal T_A.

En ciertas ocasiones, aún trabajando con aplicaciones de $\mathbb{R}^n$ en $\mathbb{R}^m$, nos interesará usar bases distintas de las canónicas. Supongamos, por ejemplo, que necesitamos usar las bases $B = (b_1, \dots, b_n)$ y $B' = (b'_1, \dots, b'_m)$ de $\mathbb{R}^n$ y $\mathbb{R}^m$, repectivamente. En ese caso, si el vector $u \in \mathbb{R}^n$ tiene coordenadas $[u]_B = (u_1, \dots, u_n)^t$ respecto de la base $B = (b_1, \dots, b_n)$, se cumple que

$$u = u_1 b_1 + \cdots + u_n b_n = (b_1, \dots, b_n)\,[u]_B \quad \Rightarrow \quad T(u) = (T(b_1), \dots, T(b_n))\,[u]_B\,.$$

De esta manera, si expandimos los vectores $T(b_i) \in \mathbb{R}^m$ en la base B' se tiene que

$$T(b_i) = t_{1i} b'_1 + t_{2i} b'_2 + \cdots + t_{mi} b'_m = (b'_1, b'_2, \dots, b'_m) \begin{pmatrix} t_{1i} \\ t_{2i} \\ \vdots \\ t_{mi} \end{pmatrix}$$

y, por tanto,

$$T(u) = (b'_1, b'_2, \dots, b'_m) \begin{pmatrix} t_{11} & \dots & t_{1n} \\ \vdots & & \vdots \\ t_{m1} & \dots & t_{mn} \end{pmatrix} \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}.$$

La ecuación anterior implica que las coordenadas de $T(u)$ en la base B' se obtienen sin más que multiplicar las coordendas de u en la base B por la matriz

$$A_{T,B'B} = \begin{pmatrix} t_{11} & \dots & t_{1n} \\ \vdots & & \vdots \\ t_{m1} & \dots & t_{mn} \end{pmatrix}.$$

El razonamiento anterior nos garantiza que, **fijadas las bases** B y B' de $\mathbb{R}^n$ y $\mathbb{R}^m$, a cada transformación lineal le corresponde una única matriz $A_{T,B'B}$. Además, según la demostración anterior, para determinar la matriz $A_{T,B'B}$ que representa la transformación T, $\boxed{\text{es suficiente con conocer la imagen de los vectores de } B \text{ en la base } B',}$ sean cuales sean dichas bases, pues con ello podremos determinar la imagen de cualquier vector. Nótese también que bases distintas de $\mathbb{R}^n$ y $\mathbb{R}^m$ producirán **distintas** matrices que REPRESENTAN una misma transformación T. Cuando sea claro y no sea necesario especificar con respecto a qué bases estamos calculando la matriz, denotaremos ésta simplemente como A_T.

Ejemplo

Si $O_{m\times n}$ es la matriz cero de dimensión $m \times n$ entonces $T_{O_{m\times n}}$ es la transformación lineal *nula* 0 que asigna a cada $v \in \mathbb{R}^n$ el vector cero de $\mathbb{R}^m$ (ver Tema 5).

Ejemplo

Si $n = m$ e I_n es la matriz identidad de dimensión $n \times n$, entonces T_{I_n} es la transformación lineal *identidad* $I : \mathbb{R}^n \to \mathbb{R}^n$ que le asigna a cada vector él mismo (ver Tema 5).

Ejemplo

Consideremos la matriz

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

y la transformación lineal asociada $T_A : \mathbb{R}^3 \to \mathbb{R}^2$

$$\begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} \mapsto \begin{pmatrix} 2 & 0 & -2 \\ 4 & 1 & -1 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} = \begin{pmatrix} 2u_1 - 2u_3 \\ 4u_1 + u_2 - u_3 \end{pmatrix} .$$

El núcleo de T_A viene dado por

$$\ker(T_A) = \left\{ u = (u_1, u_2, u_3)^t \in \mathbb{R}^3 \colon A\, u = 0 \right\} .$$

Este conjunto incluye todos los vectores que satisfacen $2u_1 - 2u_3 = 0$ y $4u_1 + u_2 - u_3 = 0$. Al resolver este sistema (compatible indeterminado) resulta que el núcleo es

$$\ker(T_A) = \left\{\alpha(1,-3,1)^t \colon \alpha \in \mathbb{R}\right\} ,$$

que, obviamente, es un espacio vectorial de dimension 1, generado por el vector $(1,-3,1)^t$.

La imagen de T_A es

$$\mathrm{Im}(T_A) = \left\{v = (v_1, v_2)^t \in \mathbb{R}^2 \colon v_1 = 2u_1 - 2u_3,\ v_2 = 4u_1 + u_2 - u_3,\ u_1, u_2, u_3 \in \mathbb{R}\right\}$$

y su dimensión se calcula fácilmente como

$$\mathrm{rg}(T_A) = n - \mathrm{nul}(T_A) = 3 - 1 = 2 .$$

Como $\mathrm{Im}(T_A) \subset \mathbb{R}^2$ y tiene la misma dimensión que éste, se tiene que

$$\mathrm{Im}(T_A) = \mathbb{R}^2 .$$

Finalmente, obsérvese que el rango de la matriz A es también 2.

Ejemplo

Sea $T : \mathbb{R}^3 \to \mathbb{R}^3$ la transformación lineal dada por

$$T((u_1, u_2, u_3)^t) = (u_1 + u_2 + u_3, u_1 - u_3, u_1 + 2u_2 - 3u_3)^t .$$

Puesto que

$$T(e_1) = (1,1,1)^t ; \quad T(e_2) = (1,-1,2)^t ; \quad T(e_3) = (1,0,-3)^t$$

respecto a la base canónica de $\mathbb{R}^3$, se tiene que

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

Obsérvese que las componentes de la matriz A_T, por columnas, son los coeficientes de u_1, u_2, u_3 en la definición de T.

El ejemplo anterior ilustra el siguiente resultado:

Teorema

Si $T_A : \mathbb{R}^n \to \mathbb{R}^m$ es la transformación lineal asociada a la matriz A de dimensión $m \times n$, entonces:

1. $\ker(T_A) = N(A)$.
2. $\operatorname{Im}(T_A) = \mathcal{C}(A)$.
3. $\operatorname{rg}(T_A) = \operatorname{rg}(A)$.
4. El máximo valor posible para $\operatorname{rg}(T_A)$ es igual al $\operatorname{mín}(m, n)$.

Teorema

Si A es una matriz $n \times n$ y $T_A : \mathbb{R}^n \to \mathbb{R}^n$ es la correspondiente transformación lineal, las siguientes propiedades son equivalentes:

1. A es no-singular (tiene inversa).
2. La imagen de T_A es $\mathbb{R}^n$.
3. T_A es inyectiva.
4. La transformación lineal T_A^{-1} existe y su matriz es A^{-1}.

Obsérvese que toda transformación lineal inyectiva corresponde exactamente a una matriz no-singular.

A continuación, generalizamos lo que hemos visto a espacios vectoriales arbitrarios; con ello se hará evidente que, para trabajar con transformaciones lineales, será suficiente con realizar cálculos con las matrices que las representan *respecto a las bases escogidas*.

6.2. Representación matricial de una transformación lineal arbitraria

Sean los espacios vectoriales U y V sobre $\mathbb{R}$, de dimensiones respectivas n y m, y bases (ordenadas) respectivas $B_U = (u_1, \ldots, u_n)$ y $B_V = (v_1, \ldots, v_m)$. Sabemos que todo vector de U puede ser representado de manera única mediante un vector columna de $\mathbb{R}^n$ cuyas componentes son las coordenadas de dicho vector con respecto a la base B_U. Análogamente, los vectores de V tienen representación única en la base B_V como un vector de $\mathbb{R}^m$.

Consideremos la aplicación lineal $T : U \to V$. Para cada vector u_i de la base B_U, su imagen con respecto a T será un vector $T(u_i) \in V$, que puede ser representado con respecto a la base B_V mediante

$$T(u_i) = t_{1i}v_1 + t_{2i}v_2 + \cdots + t_{mi}v_m = (v_1, v_2, \ldots, v_m)\begin{pmatrix} t_{1i} \\ t_{2i} \\ \vdots \\ t_{mi} \end{pmatrix}$$

o como el vector de coordenadas con respecto a B_V dado por $[T(u_i]_{B_V} = (t_{1i}, t_{2i}, \cdots, t_{mi})^t$. De esta manera, la transformación lineal T tiene una REPRESENTACIÓN MATRICIAL relativa a las bases B_U y B_V dada por la siguiente matriz $n \times m$:

$$A_{T,B_VB_U} = \begin{pmatrix} t_{11} & \dots & t_{1n} \\ \vdots & & \vdots \\ t_{m1} & \dots & t_{mn} \end{pmatrix}.$$

Obsérvese que igual que sucedía en el caso de los espacios $\mathbb{R}^n$ y $\mathbb{R}^m$, la i-ésima columna de la matriz A_{T,B_VB_U} viene dada por las coordenadas respecto a B_V de la imagen por T de u_i.

Si $U = V$ y usamos la misma base B para ambos espacios vectoriales, entonces podemos aligerar la notación definiendo $A_{T,B} = A_{T,BB}$.

Con esto en mente, el siguiente resultado es inmediato:

Proposición

Sean $T : U \to V$ una transformación lineal y A_{T,B_VB_U} su representación matricial relativa a las bases B_U y B_V. Si representamos los vectores de U y V como vectores columna respecto a dichas bases se tiene que

$$\overbrace{[T(u)]_{B_V}}^{\text{rep. coordenadas}} = \underbrace{A_{T,B_VB_U}}_{\text{rep. matricial}} \overbrace{[u]_{B_U}}^{\text{rep. coordenadas}}, \quad \forall u \in U.$$

Ejemplo

Consideremos la transformación lineal $T : \mathbb{P}_2 \to \mathbb{P}_1$ dada por $T(a_0 + a_1x + a_2x^2) = a_1 + a_2x$. Consideremos las bases $B_1 = (1, x, x^2)$ y $B_1' = (1, 1+x, 1+x+x^2)$ de $\mathbb{P}_2$ y $B_2 = (1, x)$ y $B_2' = (1-x, 1+x)$ de $\mathbb{P}_1$. La representación matricial de T respecto a las bases (canónicas) B_1 y B_2 es

$$A_{T,B_2B_1} = \left([T(1)]_{B_2}, [T(x)]_{B_2}, [T(x^2)]_{B_2}\right) = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Por otro lado, podemos escribir las imágenes de los vectores de la base B_1' respecto a la base B_2' como sigue:

$$\begin{aligned} T(1) &= \underbrace{0}_{\in \mathbb{P}_1} = \underbrace{0}_{\text{escalar}}(1-x) + \underbrace{0}_{\text{escalar}}(1+x)\,, \\ T(1+x) &= \underbrace{1}_{\in \mathbb{P}_1} = \frac{1}{2}(1-x) + \frac{1}{2}(1+x)\,, \\ T(1+x+x^2) &= \underbrace{1+x}_{\in \mathbb{P}_1} = \underbrace{0}_{\text{escalar}}(1-x) + \underbrace{1}_{\text{escalar}}(1+x)\,. \end{aligned}$$

Es decir:

$$[T(1)]_{B_2'} = (0,0)^t\,, \quad [T(1+x)]_{B_2'} = \frac{1}{2}\,(1,1)^t\,, \quad [T(1+x+x^2)]_{B_2'} = (0,1)^t$$

y, en consecuencia, la representación matricial de T respecto a las bases B_1' y B_2', está dada por

$$A_{T,B_2'B_1'} = \left(\, [T(1)]_{B_2'}, [T(1+x)]_{B_2'}, [T(1+x+x^2)]_{B_2'} \,\right) = \frac{1}{2}\begin{pmatrix} 0 & 1 & 0 \\ 0 & 1 & 2 \end{pmatrix}.$$

Finalmente, como adelantamos al final de la sección anterior, enunciamos las siguientes propiedades, que nos permitirán realizar ciertas operaciones con las transformaciones lineales utilizando simplemente operaciones equivalentes entre las matrices que las representan.

Proposición

- Si T_1 y T_2 son transformaciones lineales representadas (respecto a las mismas bases) por las matrices A_{T_1} y A_{T_2} respectivamente, entonces la transformación lineal $T_1 + T_2$ está representada por $A_{T_1} + A_{T_2}$.

- Si T es una transformación lineal representada por la matriz A_T y $\alpha \in \mathbb{K}$ es un escalar, entonces la transformación lineal $\alpha\, T$ está representada por $\alpha\, A_T$.

- Si $T_1 : U \to V$ y $T_2 : V \to W$ son transformaciones lineales representadas por A_{T_1} y A_{T_2}, entonces $T_2 \circ T_1$ está representada por el producto $A_{T_2} A_{T_1}$.
- Una transformación lineal T es invertible si y sólo si la matriz asociada A_T es invertible (independientemente de las bases escogidas para representar T). En tal caso, la matriz asociada a T^{-1} en las mismas bases es A_T^{-1}.

Ejemplo

Consideremos las transformaciones lineales $T_1 : \mathbb{R}^2 \to \mathbb{R}^2$ y $T_2 : \mathbb{R}^2 \to \mathbb{R}^2$ definidas por

$$\begin{aligned} T_1\left((v_1, v_2)^t\right) &= (v_1 + v_2, v_1 - v_2)^t, \\ T_2\left((v_1, v_2)^t\right) &= (2v_1 - v_2, 3v_1 + v_2)^t. \end{aligned}$$

La aplicación suma $S = T_1 + T_2$ verifica

$$\begin{aligned} S\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = (T_1 + T_2)\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} &= T_1\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} + T_2\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \\ &= \begin{pmatrix} v_1 + v_2 \\ v_1 - v_2 \end{pmatrix} + \begin{pmatrix} 2v_1 - v_2 \\ 3v_1 + v_2 \end{pmatrix} = \begin{pmatrix} 3v_1 \\ 4v_1 \end{pmatrix}. \end{aligned}$$

Las matrices asociadas a cada una de estas transformaciones respecto a la base canónica son

$$A_{T_1} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}; \quad A_{T_2} = \begin{pmatrix} 2 & -1 \\ 3 & 1 \end{pmatrix}; \quad A_S = \begin{pmatrix} 3 & 0 \\ 4 & 0 \end{pmatrix}$$

y obviamente $A_{T_1} + A_{T_2} = A_S$.

Además:

Dos matrices son equivalentes si y sólo si representan a la misma transformación lineal con respecto a dos bases distintas.

Ejemplo

Consideremos de nuevo la transformación lineal $T : \mathbb{P}_2 \to \mathbb{P}_1$ dada por $T(a_0 + a_1x + a_2x^2) = a_1 + a_2x$ y las bases $B_1 = (1, x, x^2)$ y $B_1' = (1, 1+x, 1+x+x^2)$ de $\mathbb{P}_2$ y $B_2 = (1, x)$ y $B_2' = (1-x, 1+x)$ de $\mathbb{P}_1$. Sabemos que las matrices

$$A_{T,B_2B_1} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad A_{T,B_2'B_1'} = \frac{1}{2}\begin{pmatrix} 0 & 1 & 0 \\ 0 & 1 & 2 \end{pmatrix}$$

representan a T con respecto a las bases indicadas. Según el resultado anterior, ambas matrices son equivalentes, es decir $A_{T,B_2B_1} = P\,A_{T,B_2'B_1'}\,Q^{-1}$ con P y Q invertibles. Para verlo, basta con construir una matriz P usando transformaciones de Gauss y elegir Q como la identidad. Hacemos las siguientes operaciones:

i) En la matriz $A_{T,B_2'B_1'}$ multiplicamos la primera fila por 2; es decir, multiplicamos por la izquierda por la matriz

$$P_1 = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}.$$

ii) En la matriz que resulta, sustituimos la segunda fila por la resta de la segunda fila menos la primera fila multiplicada por $1/2$; es decir, multiplicamos por la izquierda por la matriz

$$P_2 = \frac{1}{2}\begin{pmatrix} 2 & 0 \\ -1 & 2 \end{pmatrix}.$$

Uniendo ambas operaciones resulta la matriz

$$P = P_2\,P_1 = \begin{pmatrix} 2 & 0 \\ -1 & 1 \end{pmatrix}.$$

Así con las matrices invertibles P y $Q = I_3$ resulta que

$$A_{T,B_2B_1} = P\, A_{T,B_2'B_1'}\, Q^{-1},$$

por lo que efectivamente son equivalentes.

Dos matrices (cuadradas) son semejantes si y sólo si representan a la misma transformación lineal con respecto a dos bases distintas.

Ejemplo

Consideremos las matrices

$$A_1 = \begin{pmatrix} 2 & 1 \\ 4 & -3 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 2 & 4 \\ 1 & -3 \end{pmatrix}.$$

La matriz

$$P = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

verifica que

$$A_1 = P\, A_2\, P^{-1} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 2 & 4 \\ 1 & -3 \end{pmatrix} \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix},$$

por lo que A_1 y A_2 son semejantes.

Si consideramos A_1 como la matriz que representa a una cierta transformación lineal $T : \mathbb{P}_1 \to \mathbb{P}_1$ respecto a la base canónica $B_0 = (1, x)$, se tiene que

$$[T(a_0 + a_1 x)]_{B_0} = \underbrace{\begin{pmatrix} 2 & 1 \\ 4 & -3 \end{pmatrix}}_{A_{T,B_0}} \begin{pmatrix} a_0 \\ a_1 \end{pmatrix} = \begin{pmatrix} 2a_0 + a_1 \\ 4a_0 - 3a_1 \end{pmatrix},$$

es decir

$$T(a_0 + a_1 x) = 2a_0 + a_1 + (4a_0 - 3a_1)x.$$

Vamos a comprobar que, según indica el resultado anterior, A_2 representa la misma transformación T respecto a una base diferente. Consideremos la base alternativa $B = (1 + x, 1 - x)$; es claro que

$$a_0 + a_1 x = \frac{1}{2}(a_0 + a_1)(1 + x) + \frac{1}{2}(a_0 - a_1)(1 - x)\,,$$

por lo que las coordenadas del polinomio $a_0 + a_1 x$ en dicha base son:

$$[a_0 + a_1 x]_B = \frac{1}{2}\begin{pmatrix} a_0 + a_1 \\ a_0 - a_1 \end{pmatrix}.$$

La imagen de $a_0 + a_1 x$ también puede escribirse en función de esta base de la forma:

$$T(a_0 + a_1 x) = 2a_0 + a_1 + (4a_0 - 3a_1)x = (3a_0 - a_1)(1 + x) + (-a_0 + 2a_1)(1 - x)\,,$$

es decir

$$[T(a_0 + a_1 x)]_B = \begin{pmatrix} 3a_0 - a_1 \\ -a_0 + 2a_1 \end{pmatrix}.$$

Si multiplicamos la matriz A_2 por $[a_0 + a_1 x]_B$ obtenemos:

$$\begin{pmatrix} 2 & 4 \\ 1 & -3 \end{pmatrix} \frac{1}{2}\begin{pmatrix} a_0 + a_1 \\ a_0 - a_1 \end{pmatrix} = \begin{pmatrix} 3a_0 - a_1 \\ -a_0 + 2a_1 \end{pmatrix},$$

que coincide con las coordenadas de $T(a_0 + a_1 x)$ respecto a la base B, por lo que efectivamente $A_2 = A_{T,B}$.

NOTA: en el ejemplo anterior la matriz P y la base B han sido dadas, sin entrar en detalles de por qué son las que necesitamos utilizar para ilustrar el resultado sobre matrices semejantes; en el tema siguiente aprenderemos la relación que existe entre una y otra y a cómo calcularlas de manera explícita.

Parte VIII

Tema 7. Forma normal de una transformación lineal

Tema 7

Forma normal de una transformación lineal

7.1. Recordatorio sobre el cambio de base

En el Tema 4 estudiamos cómo cambian las coordenadas $[\mathbf{u}]_B$ de un vector $\mathbf{u} \in U$ respecto a una base B cuando cambiamos dicha base por una nueva base B'. En este apartado vamos a recordar brevemente las fórmulas básicas.

Supongamos que tenemos un espacio vectorial U de dimensión n. En dicho espacio tenemos una base $B = (\mathbf{b}_1, \ldots, \mathbf{b}_n)$, de manera que a cada vector $\mathbf{u} \in U$ le podemos asociar un vector columna con sus coordenadas en dicha base:

$$\mathbf{u} = \sum_{j=1}^{n} u_i \, \mathbf{b}_i \quad \Leftrightarrow \quad [\mathbf{u}]_B = (u_1, \ldots, u_n)^t \,,$$

de manera que podemos escribir

$$\mathbf{u} = (\mathbf{b}_1, \ldots, \mathbf{b}_n) \, [\mathbf{u}]_B = B \, [\mathbf{u}]_B \,.$$

Definimos ahora una nueva base $B' = (\mathbf{b}'_1, \ldots, \mathbf{b}'_n)$ en U. En esta base cada vector

$\mathbf{u} \in U$ tendrá una ciertas coordenadas $[\mathbf{u}]_{B'}$ dadas por

$$\mathbf{u} = \sum_{j=1}^{n} u_i'\,\mathbf{b}_i' \quad \Leftrightarrow \quad [\mathbf{u}]_{B'} = (u_1', u_2', \ldots, u_n')^t$$

y también se cumple que

$$\mathbf{u} = (\mathbf{b}_1', \ldots, \mathbf{b}_n')\,[\mathbf{u}]_{B'} = B'\,[\mathbf{u}]_{B'}\,.$$

La matriz de cambio de base $T_{BB'}$ se define de la manera habitual

$$T_{BB'} = ([\mathbf{b}_1']_B, [\mathbf{b}_2']_B, \ldots, [\mathbf{b}_n']_B)\ ,$$

de manera que

$$B' = (\mathbf{b}_1', \ldots, \mathbf{b}_n') = (\mathbf{b}_1, \ldots, \mathbf{b}_n)\,T_{BB'} = B\,T_{BB'}\,.$$

Con esta definición se tiene que las coordenadas cambian como sigue:

$$\begin{aligned} [\mathbf{u}]_B &= T_{BB'}\,[\mathbf{u}]_{B'}\,, \\ [\mathbf{u}]_{B'} &= T_{B'B}\,[\mathbf{u}]_B = T_{BB'}^{-1}\,[\mathbf{u}]_B\,. \end{aligned}$$

La matriz $T_{BB'}$ es siempre invertible. Si hacemos dos cambios sucesivos de base $B \to B' \to B''$, entonces se cumple que

$$T_{BB''} = T_{BB'}\,T_{B'B''}\,.$$

7.2. Forma normal de una transformación lineal

En el Tema 6 estudiamos cómo expresar una transformación lineal $T : U \to V$ por medio de una matriz referida a unas bases dadas de U y V. Si disponemos de otras bases distintas, es posible calcular la nueva matriz asociada a T, referida a éstas bases, con un cálculo sencillo, como veremos a continuación. Adicionalmente, dada la transformación lineal T, vamos a estudiar un procedimiento que nos permitirá encontrar ciertas bases con respecto a las cuales la matriz asociada tiene una forma por *bloques* particularmente sencilla.

7.2.1. Cambio de base como transformación lineal

Consideremos un espacio vectorial V de dimensión n y sean $B_1 = (\boldsymbol{v}_1, \dots, \boldsymbol{v}_n)$ y $B_2 = (\boldsymbol{w}_1, \dots, \boldsymbol{w}_n)$ dos bases de V. Para cualquier transformación lineal $T : V \to V$ y fijadas las bases B_1 y B_2, podemos encontrar una representación matricial respecto a dichas bases dada por

$$A_{T,B_2B_1} = ([T(\boldsymbol{v}_1)]_{B_2}, [T(\boldsymbol{v}_2)]_{B_2}, \dots, [T(\boldsymbol{v}_n)]_{B_2}) \ ,$$

de manera que

$$[T(\boldsymbol{v})]_{B_2} = A_{T,B_2B_1} [\boldsymbol{v}]_{B_1} \, .$$

Si consideramos ahora la *transformación identidad* $T = I$, que asocia a cada vector él mismo (es decir, $I(v) = v$), su representación respecto a las bases anteriores se obtiene usando $I(\boldsymbol{v}_i) = \boldsymbol{v}_i$ para todo $1 \leqslant i \leqslant n$:

$$A_{I,B_2B_1} = ([\boldsymbol{v}_1]_{B_2}, [\boldsymbol{v}_2]_{B_2}, \dots, [\boldsymbol{v}_n]_{B_2}) = T_{B_2B_1} \, ,$$

donde $T_{B_2B_1}$ es la matriz de cambio de base de B_2 a B_1. Entonces

$$[\boldsymbol{v}]_{B_2} = A_{I,B_2B_1} [\boldsymbol{v}]_{B_1} = T_{B_2B_1} [\boldsymbol{v}]_{B_1} \, .$$

Esta ecuación es equivalente a que la imagen por $T = I$ de cualquier vector $\boldsymbol{v} \in V$ es $I(\boldsymbol{v}) = \boldsymbol{v}$, como ya sabíamos.

Ya sabemos que toda matriz de cambio de base, equivalente a la matriz identidad, es invertible. Inversamente, toda matriz invertible se puede considerar como una matriz de cambio de base para pasar de una base B a una nueva base B', en la que sus columnas son las coordenadas de los vectores de la nueva base B' con respecto a la base B.

Finalmente, si usamos las matrices de cambio de base para definir una transformación lineal, es inmediata la relación que existe entre las diferentes matrices asociadas a una misma transformación lineal $T : U \to V$, referidas a bases distintas.

En efecto, si consideramos las bases B_1 y B_1' del espacio vectorial U y las bases B_2 y B_2' del espacio V, cualquier vector $\boldsymbol{v} \in U$ se transforma en $T(\boldsymbol{v})$ como se indica en el siguiente diagrama:

$$A_{T,B_2'B_1'} \;=\; T_{B_2'B_2}\, A_{T,B_2B_1}\, T_{B_1B_1'}$$

$$\begin{tikzcd}
{[\boldsymbol{v}]_{B_1}} \arrow[r, "A_{T,B_2B_1}"] & {[T(\boldsymbol{v})]_{B_2}} \arrow[d, "T_{B_2'B_2}"'] \\
{[\boldsymbol{v}]_{B_1'}} \arrow[u, "T_{B_1B_1'}"'] \arrow[r, "A_{T,B_2'B_1'}"'] & {[T(\boldsymbol{v})]_{B_2'}}
\end{tikzcd}$$

Es claro entonces que:

$$A_{T,B_2'B_1'} = T_{B_2'B_2}\, A_{T,B_2B_1}\, T_{B_1B_1'} = T^{-1}_{B_2B_2'}\, A_{T,B_2B_1}\, T_{B_1B_1'} \,.$$

Ejemplo

Consideremos la transformación lineal $T : \mathbb{P}_2 \to \mathbb{P}_1$ dada por $T(a_0 + a_1x + a_2x^2) = a_1 + a_2x$, que obviamente es lineal. Consideremos las bases $B_1 = (1, x, x^2)$ y $B_1' = (1, 1 + x, 1 + x + x^2)$ de $\mathbb{P}_2$ y $B_2 = (1, x)$ y $B_2' = (1 - x, 1 + x)$ de $\mathbb{P}_1$. La representación matricial de T respecto a las bases (canónicas) B_1 y B_2 es:

$$A_{T,B_2B_1} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} .$$

Por otro lado, las matrices de cambio de base son:

$$T_{B_1 B_1'} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

$$T_{B_2' B_2} = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}^{-1} = \frac{1}{2}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}.$$

Finalmente, la representación matricial de T respecto a las bases B_1' y B_2' puede escribirse de la forma:

$$A_{T,B_2'B_1'} = \frac{1}{2}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 0 & 1 & 0 \\ 0 & 1 & 2 \end{pmatrix}.$$

7.2.2. Cálculo de la forma normal

Consideremos la transformación lineal $T : U \to V$, donde los espacios vectoriales U y V tienen dimensiones respectivas m y n. Si denotamos el rango de la transformación por $r = rg(T)$, el procedimiento descrito a continuación nos permitirá encontrar una base B_1 de U y una base B_2 de V tales que la representación matricial de T con respecto a B_1 y B_2 sea de la forma:

$$A_{T,B_2B_1} = \left(\begin{array}{c|c} I_r & 0_{r\times(m-r)} \\ \hline 0_{(n-r)\times r} & 0_{(n-r)\times(m-r)} \end{array}\right),$$

conocida como **forma normal** de la representación (*matriz*) de la transformación lineal. Obsérvese que esta matriz es diagonal por bloques.

Por una parte, sean B_1' y B_2' bases cualesquiera de U y V, respectivamente. Consideremos la matriz asociada a T con respecto a dichas bases $A_{T,B_2'B_1'}$.

Por otra parte, construimos las bases B_1 y B_2 de la siguiente manera.

- Obtenemos una base de $\ker(T)$ (de dimensión $m-r$). Denotamos los $m-r$ vectores que la forman por $\mathbf{u}_{r+1}, \ldots, \mathbf{u}_m$. A continuación, extendemos ésta a una base de U añadiendo r vectores linealmente independientes, $\mathbf{u}_1, \ldots, \mathbf{u}_r$. Es decir:

$$B_1 = (\mathbf{u}_1, \ldots, \mathbf{u}_r, \underbrace{\mathbf{u}_{r+1}, \ldots, \mathbf{u}_m}_{\text{base de } \ker(T)}) .$$

- A continuación obtenemos una base de $\operatorname{Im}(T)$ de la forma $\mathbf{v}_i = T(\mathbf{u}_i)$ (para todo $1 \leqslant i \leqslant r$) y la extendemos a una base de V añadiendo $n-r$ vectores linealmente independientes:

$$B_2 = (\underbrace{\mathbf{v}_1, \ldots, \mathbf{v}_r}_{\text{base de } \operatorname{Im}(T)}, \mathbf{v}_{r+1}, \ldots, \mathbf{v}_n) .$$

Entonces, la matriz asociada a T respecto a las bases B_1 y B_2 toma la forma

$$A_{T,B_2B_1} = T_{B_2B_2'} A_{T,B_2'B_1'} T_{B_1'B_1} ,$$

que es la representación normal de la transformación T.

Ejemplo

Consideremos la transformación lineal $T : \mathbb{R}^3 \to \mathbb{R}^2$ definida por

$$T((u_1, u_2, u_3)^t) = (u_1 + u_2 + u_3, u_1 - u_2)^t .$$

El núcleo de T será

$$\ker(T) = \left\{(1, 1, -2)^t u_1 \colon u_1 \in \mathbb{R}\right\} = \operatorname{Gen}\left((-1, -1, 2)^t\right) .$$

Completamos una base de $\mathbb{R}^3$ con dos vectores linealmente independientes de $\mathbf{v}_3 = (-1, -1, 2)^t$: por ejemplo, $\mathbf{v}_1 = (1, 0, 1)^t$ y $\mathbf{v}_2 = (0, 1, 1)^t$.
A continuación hallamos una base de $\mathbb{R}^2$ mediante $T(\mathbf{v}_1) = (2, 1)^t$ y $T(\mathbf{v}_2) = (2, -1)^t$. Obsérvese que en este caso no es necesario extender la base, ya que $\operatorname{Im}(T) = \mathbb{R}^2$.

Para simplificar la escritura, vamos a *abusar de la notación* y vamos a representar por B_0 la base canónica tanto de $\mathbb{R}^3$ como de $\mathbb{R}^2$. La matriz asociada a T con respecto a las bases canónicas de $\mathbb{R}^3$ y $\mathbb{R}^2$ (eliminando la referencia a éstas en el subíndice) viene dada por

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

Entonces las matrices de cambio de base entre las bases calculadas y las respectivas bases canónicas son:

$$T_{B_0 B_1} = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 1 & 1 & 2 \end{pmatrix}, \quad T_{B_0 B_2} = \begin{pmatrix} 2 & 2 \\ 1 & -1 \end{pmatrix}.$$

Finalmente la representación normal de T será la matriz asociada a T referida a las bases B_1 y B_2, que calculamos mediante

$$\begin{aligned} A_{T,B_2B_1} = T_{B_2B_0}\, A_T\, T_{B_0B_1} &= \begin{pmatrix} 2 & 2 \\ 1 & -1 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 1 & 1 & 2 \end{pmatrix} \\ &= \frac{1}{4} \begin{pmatrix} 1 & 2 \\ 1 & -2 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 1 & 1 & 2 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}. \end{aligned}$$

Parte IX

Tema 8. Autovalores y autovectores. Diagonalización

Tema 8

Valores y vectores propios

8.1. Introducción

En este capítulo continuamos con el estudio de matrices cuadradas y las transformaciones lineales que representan. Los principales conceptos que abordaremos, valores y vectores propios, poseen muchas aplicaciones prácticas y desempeñan un papel fundamental en muchos campos de las matemáticas.

Una transformación $x \mapsto A\,x$, con A cuadrada, puede transformar los vectores x de muy diversas maneras, pero para cada matriz existen algunos vectores "especiales" en los que la "acción" de la transformación es particularmente simple.

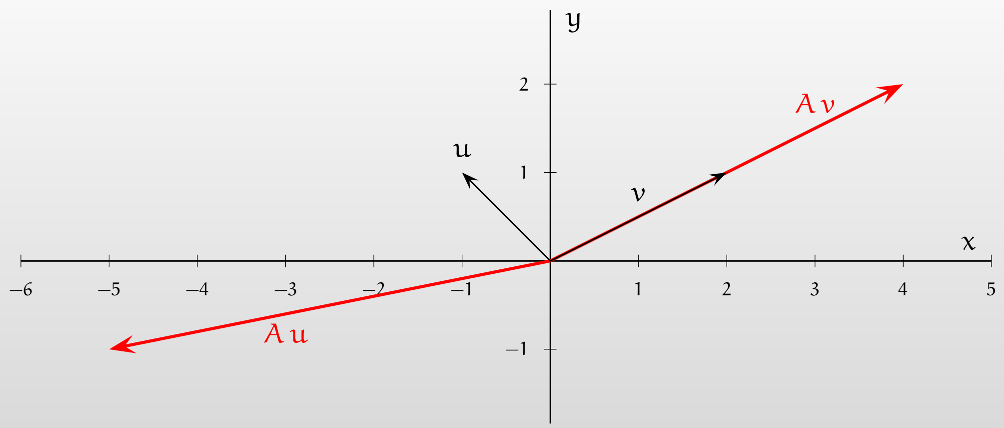
Ejemplo

Consideremos la matriz

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

y los vectores $u = (-1, 1)^t$ y $v = (2, 1)^t$. Sus imágenes son $T(u) = A\,u = (-5, -1)^t$ y

$T(v) = A\,v = (4,2)^t = 2v$; es decir, a diferencia con u, A dilata el vector v un factor 2.



Este ejemplo ilustra el concepto de *vector propio*.

Diremos que v es un **vector propio** (o *autovector*) de la matriz cuadrada $A \in \mathbb{K}^{n\times n}$ si es un $\boxed{\text{vector no nulo}}$ tal que

$$A\,v = \lambda\,v\,, \quad \text{para algún escalar } \lambda \in \mathbb{R}.$$

Dicho escalar se denomina **valor propio** (o *autovalor*) de A.

El conjunto de valores propios de una matriz A recibe el nombre de **espectro de** A y se representa por $\sigma(A)$.

En otros términos, los valores propios de la matriz A son los escalares para los que existe una solución no trivial v de $A\,v = \lambda v$ y nos referiremos a v como el *vector propio correspondiente al valor propio* λ. Obsérvese que v proporciona las “direcciones” en las que la acción de la matriz produce un “estiramiento puro” y que nos estamos restringiendo al caso en el que $\lambda \in \mathbb{R}$.

Es fácil VERIFICAR si un vector dado es o no un vector propio de A: basta con calcular $A\,v$ y comprobar si es un múltiplo de v. En cambio, para CALCULAR todos los posibles vectores propios de una matriz dada A, tendremos que resolver el siguiente sistema:

$$A\,v = \lambda v \quad \Rightarrow \quad A\,v = \lambda\,I_n\,v \quad \Rightarrow \quad A\,v - \lambda\,I_n\,v = 0 \quad \Rightarrow \quad \boxed{(A - \lambda\,I_n)\,v = 0.}$$

Así, λ es un valor propio de A si y sólo si el sistema homogéneo $(A - \lambda\,I_n)\,v = 0$ tiene una solución no trivial. El conjunto de todas las soluciones de este sistema es simplemente el espacio nulo de la matriz $A - \lambda\,I_n$, que es un subespacio de $\mathbb{R}^n$, y se denomina **espacio propio** (o *autoespacio*) de A asociado a λ. Por tanto, el espacio propio está formado por el vector cero y el conjunto de todos los vectores propios correspondientes a λ.

Ejemplo

Consideremos de nuevo la matriz

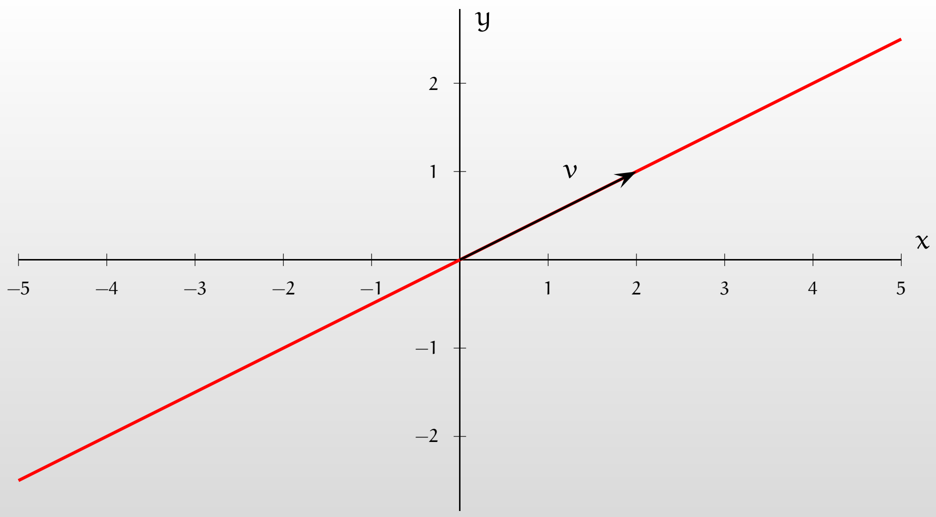
$$A = \begin{pmatrix} 3 & -2 \\ 1 & 0 \end{pmatrix}.$$

Ya sabemos que $\lambda_1 = 2$ es un valor propio de A; el correspondiente espacio propio es

$$N(A - 2\,I_2) = \left\{(v_1, v_2)^t \in \mathbb{R}^2 \colon v_1 - 2v_2 = 0\right\} = \left\{(2v_2, v_2)^t, v_2 \in \mathbb{R}\right\} = \mathrm{Gen}\left((2,1)^t\right).$$

Es decir, obviamente, $\boxed{\text{cualquier múltiplo no nulo de } v = (2,1)^t \text{ es vector propio de } A}$ asociado al mismo valor propio $\lambda_1 = 2$.

También sabemos que $(-1,1)^t$ (o cualquier múltiplo suyo) no es un vector propio de A.



Ejemplo

Sea A la matriz

$$A = \begin{pmatrix} 4 & -1 & 6 \\ 2 & 1 & 6 \\ 2 & -1 & 8 \end{pmatrix}.$$

Verifiquemos que $\lambda_1 = 2$ es también un valor propio de A y encontremos una base para el correspondiente espacio propio. Nótese que, a diferencia con el ejemplo anterior, en este caso NO conocemos ningún vector propio. Entonces, para comprobar que $\lambda_1 = 2$ es un valor propio, tendremos que comprobar que efectivamente $(A - \lambda_1 I_3)v = 0$. Es decir, el sistema

$$\begin{pmatrix} 2 & -1 & 6 \\ 2 & -1 & 6 \\ 2 & -1 & 6 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix},$$

debe tener solución no trivial, lo cual implica que $rg(A - 2\, I_3) < 3$, como obviamente es el caso. Por tanto, concluimos que $\lambda_1 = 2$ es un valor propio.

Por otra parte, toda solución del sistema cumple $2v_1 - v_2 + 6v_3 = 0$, es decir:

$$
\begin{aligned}
N(A-2\,I_3) &= \left\{ v = \left(\frac{1}{2}v_2 - 3v_3, v_2, v_3\right)^t \in \mathbb{R}^3 \colon v_2, v_3 \in \mathbb{R} \right\} \\
&= \left\{ v = v_2 \left(\frac{1}{2}, 1, 0\right)^t + v_3\,(-3,0,1)^t \in \mathbb{R}^3 \colon v_2, v_3 \in \mathbb{R} \right\} \\
&= \operatorname{Gen}\left(\left(\frac{1}{2}, 1, 0\right)^t, (-3,0,1)^t \right).
\end{aligned}
$$

Tras estos ejemplos es inmediato formularse la pregunta de cuántos valores propios tiene una matriz dada. En la siguiente sección responderemos a esta cuestión y aprenderemos a calcularlos todos.

8.2. La ecuación característica

Acabamos de ver que los valores propios son aquellos escalares que hacen que el sistema homogéneo $(A - \lambda I_n)v = 0$ tenga solución no trivial. Para ello, la matriz $A - \lambda I_n$ debe ser singular y consecuentemente, su determinante ha de ser 0; esto conduce a la siguiente definición:

Llamaremos **ecuación característica** de $A \in \mathbb{R}^{n\times n}$ a la ecuación:

$$\boxed{\det(A - \lambda\, I_n) = 0.}$$

Obsérvese que $\det(A - \lambda\, I_n)$ es un polinomio (en λ) de grado n, cuyo coeficiente del término en λ^n es $(-1)^n$. En cambio, el polinomio $\det(\lambda\, I_n - A)$ tiene el coeficiente del término en λ^n igual a 1.

Llamamos **polinomio característico** de A al polinomio $c_A(\lambda) = \det(\lambda I_n - A)$.

Obviamente, las soluciones de la ecuación característica (o las raíces del polinomio característico) son los valores propios de A y su multiplicidad recibe el nombre de **multiplicidad algebraica del valor propio**.

Ejemplo

Sea

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

Vamos a encontrar los valores propios y los correspondientes espacios propios.

En primer lugar, calculamos la ecuación característica:

$$\det(A - \lambda I_3) = 0 \quad \Rightarrow \begin{vmatrix} 2-\lambda & -3 & 1 \\ 1 & -2-\lambda & 1 \\ 1 & -3 & 2-\lambda \end{vmatrix} = -\lambda(\lambda-1)^2 = 0\,,$$

cuyas raíces son:

$$\lambda_1 = 0, \text{ con multiplicidad algebraica 1, y}$$

$$\lambda_2 = 1, \text{ con multiplicidad algebraica 2.}$$

El espacio propio correspondiente a $\lambda_1 = 0$ coincide con el espacio nulo de A, por lo que si calculamos la forma escalonada de A resulta

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

y el espacio propio será:

$$N(A) = \left\{ x = x_3(1,1,1)^t \in \mathbb{R} \colon x_3 \in \mathbb{R} \right\}\,,$$

que es un espacio vectorial de dimensión 1. Finalmente, para obtener el espacio propio correspondiente a $\lambda_2 = 1$ calculamos el espacio nulo de la matriz

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

que es

$$\left\{x = x_2(3,1,0)^t + x_3(-1,0,1)^t \in \mathbb{R}^3 \colon x_2, x_3 \in \mathbb{R}\right\},$$

y tiene dimensión 2.

Obsérvese que los tres vectores propios encontrados en el ejemplo anterior son linealmente independientes y por ello forman una base de $\mathbb{R}^3$.

El siguiente resultado es fundamental:

Dada la matriz cuadrada $A \in \mathbb{R}^{n\times n}$, su traza es igual a la suma de sus valores propios, $\lambda_i, i = 1, \ldots, n$ (incluyendo multiplicidades algebraicas mayores que 1):

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

Ejemplo

Es inmediato comprobar que en el ejemplo anterior, la traza de

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

coincide con la suma de sus valores propios: $0 + 1 + 1 = 2$.

Ejemplo

Consideremos la matriz

$$A = \begin{pmatrix} 1 & 4 & 6 \\ 0 & 2 & 5 \\ 0 & 0 & 3 \end{pmatrix}.$$

Los valores propios son las raíces de

$$\det \begin{pmatrix} 1-\lambda & 4 & 6 \\ 0 & 2-\lambda & 5 \\ 0 & 0 & 3-\lambda \end{pmatrix} = 0 \quad \Rightarrow \quad (1-\lambda)(2-\lambda)(3-\lambda) = 0\,.$$

Por tanto, los valores propios son $\lambda_1 = 1$, $\lambda_2 = 2$ y $\lambda_3 = 3$, todos con multiplicidad algebraica 1. Se observa que los valores propios de la matriz A vienen dados por los términos de la diagonal principal (por lo que obviamente la traza coincide con la suma de éstos). Es inmediato probar que este hecho es cierto para todas las matrices triangulares.

Ejemplo

Consideremos ahora la matriz

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

Sus valores propios son las raíces de la ecuación característica:

$$\det \begin{pmatrix} 2-\lambda & -1 & 0 \\ 1 & -\lambda & 0 \\ 0 & 0 & 3-\lambda \end{pmatrix} = 0 \quad \Rightarrow \quad (3-\lambda)(\lambda-1)^2 = 0\,.$$

Así los valores propios son $\lambda_1 = 3$, con multiplicidad algebraica 1, y $\lambda_2 = 1$, con multiplicidad algebraica 2. Obsérvese que la traza de A coincide con la suma de los valores propios. Vamos a determinar los espacios propios.

Para $\boxed{\lambda = 3}$ tenemos:

$$A - 3\,I_3 = \begin{pmatrix} -1 & -1 & 0 \\ 1 & -3 & 0 \\ 0 & 0 & 0 \end{pmatrix} \underset{\substack{\uparrow \\ \text{Gauss}}}{\equiv} \begin{pmatrix} -1 & -1 & 0 \\ 0 & -4 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

y por tanto:

$$N(A - 3\,I_3) = \left\{ v = (0, 0, v_3)^t \in \mathbb{R}^3 \colon v_3 \in \mathbb{R} \right\},$$

es decir, de dimensión 1 y con base $B_{\lambda=3} = ((0, 0, 1)^t)$.

Para $\boxed{\lambda = 1}$ calculamos el espacio nulo de

$$A - I_3 = \begin{pmatrix} 1 & -1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 2 \end{pmatrix} \underset{\substack{\uparrow \\ \text{Gauss}}}{\equiv} \begin{pmatrix} 1 & -1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix},$$

que es

$$\left\{ v = (v_1, v_1, 0)^t \in \mathbb{R}^3 \colon v_1 \in \mathbb{R} \right\},$$

que tiene una base dada por $B_{\lambda=1} = ((1, 1, 0)^t)$, también de dimensión 1.
En resumen, ambos espacios propios son de dimensión 1 y por tanto sólo podemos encontrar dos vectores propios linealmente independientes, por lo que $\mathbb{R}^3$ no tendrá una base formada por vectores propios linealmente independientes de la matriz A.

Además tenemos el siguiente resultado, que relaciona los valores propios con otra característica de las matrices cuadradas: su determinante.

Dada una matriz cuadrada A, de dimensión $n \times n$, se tiene que el producto de sus valores propios coincide con su determinante (incluyendo multiplicidades algebraicas mayores que 1):

$$\boxed{\det(A) = \prod_{i=1}^{n} \lambda_i.}$$

Ejemplo

Consideremos de nuevo la matriz

$$A = \begin{pmatrix} 1 & 4 & 6 \\ 0 & 2 & 5 \\ 0 & 0 & 3 \end{pmatrix},$$

con valores propios $\lambda_1 = 1$, $\lambda_2 = 2$ y $\lambda_3 = 3$. Es trivial comprobar que su determinante es 6, que coincide con el producto de sus valores propios.

Los resultados que enunciamos a continuación (sin demostración) son fundamentales para comprender en profundidad las secciones siguientes.

Dadas las matrices semejantes $A, B \in \mathbb{R}^{n\times n}$, se tiene que:

1. A y B tienen los mismos valores propios. En consecuencia:
   - $\mathrm{tr}(A) = \mathrm{tr}(B)$.
   - $\det(A) = \det(B)$.
2. A y B tienen el mismo polinomio característico.
3. A^k y B^k son semejantes.

8.3. El problema de la diagonalización

Hemos visto que existen casos en los que el número de vectores propios linealmente independientes no coincide con la multiplicidad algebraica del correspondiente valor propio. Nos referiremos al número de vectores propios linealmente independientes asociados a un valor propio como la **multiplicidad geométrica** de tal valor propio. Si la multiplicidad geométrica de algún valor propio es menor que su multiplicidad algebraica, diremos que la matriz es **defectiva**.

Ejemplo

Consideremos de nuevo la matriz

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

Sabemos que un valor propio es $\lambda_1 = 2$, con vector propio $v_1 = (2,1)^t$, y fácilmente se obtiene que $\lambda_2 = 1$ es el otro valor propio, con vector propio $v_2 = (1,1)^t$. También podemos probar de manera sencilla que $B = ((2,1)^t, (1,1)^t)$ es una base de $\mathbb{R}^2$. Entonces, cada vector $u \in \mathbb{R}^2$ puede ser escrito como combinación lineal de v_1 y v_2 de la forma

$$c_1v_1 + c_2v_2 = \underbrace{(v_1, v_2)\begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = u}_{\text{sistema lineal}}, \quad c_1, c_2 \in \mathbb{R}$$

y

$$Au = A\,(c_1v_1 + c_2v_2) = c_1A\,v_1 + c_2A\,v_2 = 2c_1v_1 + c_2v_2 = (v_1, v_2)\begin{pmatrix} 2c_1 \\ c_2 \end{pmatrix}$$

$$= (v_1, v_2)\begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} c_1 \\ c_2 \end{pmatrix}.$$

Esto implica que

$$A\,(v_1, v_2)\begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = (v_1, v_2)\begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} c_1 \\ c_2 \end{pmatrix}, \quad \forall c_1, c_2 \in \mathbb{R}$$

o equivalentemente

$$\begin{gathered} A\,(v_1, v_2)\begin{pmatrix} c_1 \\ c_2 \end{pmatrix} - (v_1, v_2)\begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = 0\,, \quad \forall c_1, c_2 \in \mathbb{R}\,, \\ \left[A\,(v_1, v_2) - (v_1, v_2)\begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}\right]\begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = 0\,, \quad \forall c_1, c_2 \in \mathbb{R}\,. \end{gathered}$$

Por tanto, debe ser

$$A\,(v_1, v_2) - (v_1, v_2)\begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} = 0 \quad \Rightarrow \quad A\,(v_1, v_2) = (v_1, v_2)\begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}.$$

Si escribimos $P = (v_1, v_2)$, puesto que v_1 y v_2 forman una base, las columnas de P son linealmente independientes, P tiene rango 2 y es invertible. Así, podemos escribir

$$D = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} = P^{-1}\,A\,P\,.$$

Obsérvese que los elementos de la diagonal principal de D son los valores propios de A y que las columnas de P son los correspondientes vectores propios asociados.

El ejemplo anterior ilustra el siguiente teorema:

Teorema

Si A es una matriz $n \times n$, con valores propios $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$, no necesariamente distintos, y con vectores propios correspondientes $v_1, \ldots, v_n \in \mathbb{R}^n$ linealmente independientes, entonces

$$P^{-1} A P = D\,,$$

donde

$$P = (v_1, \ldots, v_n)\,, \qquad D = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}.$$

Este resultado motiva la siguiente definición:

Una matriz A de dimensión $n \times n$ es **diagonalizable** si existe una matriz invertible P tal que $P^{-1} A P$ es una matriz diagonal.

En otras palabras, si A es diagonalizable es semejante a una matriz diagonal.

Teorema

Si $A \in \mathbb{R}^{n \times n}$ es diagonalizable, entonces A tiene n vectores propios linealmente independientes en $\mathbb{R}^n$.

Teorema

Si $A \in \mathbb{R}^{n \times n}$ tiene valores propios distintos $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ y vectores propios correspondientes $v_1, \ldots, v_n \in \mathbb{R}^n$, entonces $v_1, \ldots, v_n$ son linealmente independientes.

Además, tenemos el siguiente resultado:

Si $A \in \mathbb{R}^{n \times n}$ es diagonalizable de la forma $A = P\,D\,P^{-1}$ y $B \in \mathbb{R}^{n \times n}$ es una matriz semejante a A, entonces B es diagonalizable de la forma $B = Q\,D\,Q^{-1}$.

Demostración. Si B es semejante a A, existe una matriz invertible R tal que $B = R\,A\,R^{-1}$;

además, si $A = P\,D\,P^{-1}$, tendremos:

$$B = R\,P\,D\,P^{-1}\,R^{-1} = \underbrace{(R\,P)}_{Q}\,D\,(R\,P)^{-1} = Q\,D\,Q^{-1}\,.$$

$\square$

Así, aunque dos matrices semejantes tienen los mismos valores propios, en general, sus vectores propios no son los mismos.

8.3.1. El polinomio característico

Sea A una matriz $n \times n$. Como sabemos, el polinomio característico de A es el polinomio de grado n dado por

$$c_A(x) = \det(x\,I_n - A)$$

y las raíces del polinomio característico de A son los valores propios de A. (Nótese que hemos cambiado la variable de λ a x.) Además de permitir el cálculo de valores propios, el polinomio característico tiene otras aplicaciones. El siguiente teorema indica una de éstas.

Teorema de Caley-Hamilton

Dada una matriz A, de dimensión $n \times n$ y polinomio característico $c_A(x)$, se tiene que

$$\boxed{c_A(A) = 0_{n\times n}.}$$

Ilustremos este resultado con el siguiente ejemplo:

Ejemplo

Consideremos la matriz

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Su polinomio característico es

$$c_A(x) = \det \begin{pmatrix} x-a & -b \\ -c & x-d \end{pmatrix} = x^2 - (a+d)x + ad - bc\,.$$

Si “evaluamos” c_A en A –introduciendo convenientemente la matriz identidad para el término independiente– resulta:

$$\begin{aligned} c_A(A) &= A^2 - (a+d)\,A + (ad-bc)\,I_2 \\ &= \begin{pmatrix} a & b \\ c & d \end{pmatrix}^2 - (a+d)\begin{pmatrix} a & b \\ c & d \end{pmatrix} + (ad-bc)\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} a^2+bc & ab+bd \\ ac+cd & bc+d^2 \end{pmatrix} - \begin{pmatrix} a^2+ad & ab+bd \\ ac+cd & ad+d^2 \end{pmatrix} \\ &\qquad + \begin{pmatrix} ad-bc & 0 \\ 0 & ad-bc \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} = 0_{2\times 2}\,, \end{aligned}$$

tal y como indica el teorema de Caley-Hamilton.

También se puede usar el teorema para calcular A^2, pues sabemos que

$$A^2 - (a+d)\,A + (ad-bc)\,I_2 = 0$$

y por tanto

$$A^2 = (a+d)\,A - (ad-bc)\,I_2 = \begin{pmatrix} a^2+ad & ab+bd \\ ac+cd & ad+d^2 \end{pmatrix}.$$

Del mismo modo, podemos usar el teorema de Caley-Hamilton para calcular la inversa de A, ya que si multiplicamos ambos miembros de $A^2-(a+d)\,A+(ad-bc)\,I_2 =$

$0_{2\times 2}$ por A^{-1}, resulta:

$$A^2\,A^{-1} - (a+d)\,A\,A^{-1} + (ad-bc)\,A^{-1} = 0$$

$$\Rightarrow \quad A^{-1} = \frac{1}{ad-bc}[(a+d)\,I - A] = \frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix},$$

como es bien sabido.

8.4. Valores y vectores propios de aplicaciones lineales

Hasta ahora nos hemos centrado en el estudio de los valores y vectores propios de una matriz $A \in \mathbb{R}^{n\times n}$. En esta sección vamos a relacionar los resultados que hemos obtenido con las aplicaciones lineales.

Como las matrices semejantes comparten, entre otras características, el polinomio característico y los valores propios, podemos trasladar las definiciones y los resultados anteriores a las transformaciones lineales.

Dada una trasformación lineal T del espacio vectorial V en sí mismo, diremos que un vector no nulo v de V es un **vector propio** de T, con **valor propio** λ si $T(v) = \lambda\, v$. El conjunto de los vectores propios de T asociados a λ, junto con el vector cero, forman el **espacio propio** de T asociado a λ.

Obviamente, los valores propios de la transformación T son los de cualquier matriz cuadrada A_T que la represente. Del mismo modo:

La **traza** $\mathrm{tr}(T)$ **de una transformación lineal** $T : V \to V$ es la traza de cualquier matriz A_T que la represente.

El **determinante** $\det(T)$ **de una transformación lineal** $T : V \to V$ es el determinante de cualquier matriz A_T que la represente.

El **polinomio característico** $c_T(x)$ **de una transformación lineal** T es el polinomio característico de cualquier matriz A_T que la represente.

Una transformación lineal T sobre el espacio vectorial V se dice **diagonalizable** si existe una base de V relativa a la cual la matriz que representa a T es una matriz diagonal.

Obsérvese que aunque dos matrices semejantes, que representan la misma transformación T, en general tienen vectores propios diferentes, los vectores propios de T se obtienen de forma única a partir de cualquiera de sus representaciones. Veámoslo en el siguiente ejemplo:

Ejemplo

Consideremos la transformación lineal $T\colon \mathbb{P}_2 \to \mathbb{P}_2$ dada por

$$T(p(x)) = T(a_0 + a_1x + a_2x^2) = a_2 + 2a_1x + (a_0 + a_1)x^2 .$$

Si fijamos, por ejemplo, la base $B = (1, x, x^2)$, la matriz que representa dicha transformación respecto a la base B es:

$$A_{T,B} = \big([T(1)]_B\ ,\ [T(x)]_B\ ,\ [T(x^2)]_B\big) = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 1 & 0 \end{pmatrix} .$$

Los valores propios de $A_{T,B}$ (y por tanto los de T) se calculan mediante la ecuación característica:

$$\det\begin{pmatrix} -\lambda & 0 & 1 \\ 0 & 2-\lambda & 0 \\ 1 & 1 & -\lambda \end{pmatrix} = 0 \quad \Rightarrow \quad (2-\lambda)(\lambda-1)(\lambda+1) = 0\,.$$

Así, el espectro de T será

$$\sigma(T) = \{-1, 1, 2\}$$

y su determinante viene dado por

$$\det(T) = -2\,.$$

Calculemos los vectores propios de $A_{T,B}$.

Para $\lambda = -1$:

$$N(A_{T,B} - \lambda\, I) = N\left[\begin{pmatrix} 1 & 0 & 1 \\ 0 & 3 & 0 \\ 1 & 1 & 1 \end{pmatrix}\right] = Gen\left((-1,0,1)^t\right)\,.$$

Para $\lambda = 1$:

$$N(A_{T,B} - \lambda\, I) = N\left[\begin{pmatrix} -1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & -1 \end{pmatrix}\right] = Gen\left((1,0,1)^t\right)\,.$$

Para $\lambda = 2$:

$$N(A_{T,B} - \lambda\, I) = N\left[\begin{pmatrix} -2 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 1 & -2 \end{pmatrix}\right] = Gen\left((1,3,2)^t\right)\,.$$

Cada uno de estos vectores propios de la matriz $A_{T,B}$ corresponde a un vector propio de T. Las coordenadas $(-1,0,1)^t$ en la base B hacen referencia al vector

$$p_1(x) = x^2 - 1; \quad \text{cuya imagen es} \quad T(p_1) = -(x^2-1)\,;$$

análogamente, para $(1,0,1)^t$ tenemos

$$p_2(x) = x^2 + 1; \quad \text{con} \quad T(p_2) = x^2 + 1$$

y para $(1,3,2)^t$ será:

$$p_3(x) = 2x^2 + 3x + 1; \quad \text{con} \quad T(p_3) = 4x^2 + 6x + 2 = 2(2x^2 + 3x + 1)\,.$$

Si consideramos una base distinta de $\mathbb{P}_2$, por ejemplo $B_1 = (1, 1+x, 1+x+x^2)$, la matriz que representa a T respecto a esta base será:

$$A_{T,B_1} = \left([T(1)]_{B_1}\,,\ [T(1+x)]_{B_1}\,,\ [T(1+x+x^2)]_{B_1}\right) = \begin{pmatrix} 0 & -2 & -1 \\ -1 & 0 & 0 \\ 1 & 2 & 2 \end{pmatrix},$$

como es fácil comprobar. Esta matriz también tiene valores propios -1, 1 y 2. Si calculamos los vectores propios (de la matriz) asociados, resulta que:

$$\begin{aligned} N(A_{T,B_1} + I) &= \mathrm{Gen}\left((-1,-1,1)^t\right)\,, \\ N(A_{T,B_1} - I) &= \mathrm{Gen}\left((1,-1,1)^t\right)\,, \\ N(A_{T,B_1} - 2\,I) &= \mathrm{Gen}\left((-2,1,2)^t\right) \end{aligned}$$

y se observa que los espacios propios de A_{T,B_1} no coinciden con los de $A_{T,B}$. No obstante, sí representan los mismos vectores propios de T, ya que de las correspondientes coordenadas se obtiene

$$[p_1]_{B_1} = (-1,-1,1)^t\,, \quad \text{pues} \quad -1 - (1+x) + (1+x+x^2) = x^2 - 1 = p_1(x)$$

y también

$$[p_2]_{B_1} = (1,-1,1)^t\,, \qquad [p_3]_{B_1} = (-2,1,2)^t\,.$$

Obviamente T es diagonalizable, ya que $B' = (p_1, p_2, p_3)$ es una base de $\mathbb{P}_2$ y la matriz que representa a T con respecto a dicha base es

$$A_{T,B'} = ([T(p_1)]_{B'}, [T(p_2)]_{B'}, [T(p_3)]_{B'}) = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

Ahora, el siguiente resultado es fácil de entender:

Proposición

Una transformación lineal T sobre V es diagonalizable si y sólo si existe una base de V que consiste en vectores propios de T.

Así, diagonalizar una transformación lineal $T : V \to V$ consiste en encontrar una base B_d de V en la que la matriz asociada A_{T,B_d} sea diagonal. A_{T,B_d} será *semejante* a cualquier otra matriz A_{T,B_1} que represente a T. En efecto, si T es diagonalizable existe una base B_d de V formada por vectores propios de T.

Para cualquier otra base B_1, la matriz A_{T,B_1} puede escribirse como $A_{T,B_1} = P\,D\,P^{-1}$, donde las columnas de $P = T_{B_1 B_d}$ *representan* tales vectores propios en la base B_1. Así, $P = T_{B_1 B_d}$ es la matriz de cambio de base para pasar de la base B_1 a la base B_d y $D = A_{T,B_d}$ es la matriz de la transformación T cuando los vectores de V están expresados en términos de la base B_d, como se indica en el siguiente diagrama:

$$D \;=\; A_{T,B_d} = T_{B_1 B_d}^{-1}\, A_{T,B_1}\, T_{B_1 B_d}$$

$$\begin{tikzcd}
{[v]_{B_1}} \arrow[rr, "A_{T,B_1}"] & & {[T(v)]_{B_1}} \arrow[d, "T_{B_d B_1}"] \\
{[v]_{B_d}} \arrow[u, "T_{B_1 B_d} = P"] \arrow[rr, "D"'] & & {[T(v)]_{B_d}}
\end{tikzcd}$$

Ejemplo

Sabemos que la matriz

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

(que representa una transformación lineal S respecto a la base canónica B_0 de $\mathbb{R}^2$) puede ser diagonalizada de la forma:

$$A = P\,D\,P^{-1}$$

donde

$$D = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}, \quad P = T_{B_0 B_d} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}, \quad P^{-1} = T_{B_d B_0} = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}$$

donde B_d es la base de $\mathbb{R}^2$ con respecto a la cual la matriz asociada a S es D.

Si consideramos, por ejemplo, el vector $u = (5, -3)^t$, la transformación lineal S lo transforma en $w = S(u)$, cuyas coordenadas son $w = (21, 5)^t$.

Por otra parte, sabemos que las columnas de P son vectores propios linealmente independientes de A y que forman una base B_d de $\mathbb{R}^2$. Si expresamos u y $S(u)$ en términos de tales vectores se obtiene:

$$[u]_{B_d} = (8, -11)^t\,, \quad [S(u)]_{B_d} = (16, -11)^t\,.$$

Es fácil comprobar que

$$
\begin{aligned}
[u]_{B_d} &= T_{B_d B_0}\, u = P^{-1}\, u\,, \\
[S(u)]_{B_d} &= T_{B_d B_0}\, [S(u)]_{B_0} = P^{-1}\, [S(u)]_{B_0}\,, \\
[S(u)]_{B_d} &= D\, [u]_{B_d}\,.
\end{aligned}
$$

Teorema

Una transformación lineal T sobre V es diagonalizable si y sólo T se puede escribir de la forma

$$T = \lambda_1\, T_1 + \cdots + \lambda_n\, T_n$$

donde λ_i son los valores propios de T y las T_i son transformaciones que verifican $T_i \circ T_i = T_i$, $T_1 + \cdots + T_n = I$ (aplicación identidad) y $T_i \circ T_j = 0$ (aplicación nula) para todo $i \neq j$.

Autovalores y autovectores

AUTOVECTORES (VECTORES PROPIOS) Y AUTOVALORES (VALORES PROPIOS) DE UNA MATRIZ CUADRADA.

a) **Definición 1.-** *Sea A una matriz cuadrada con elementos de un cuerpo conmutativo K, se llama autovector (vector propio) de A todo vector x de* K^n *tal que existe un escalar* λ *de K verificando:*

$$Ax = \lambda x$$

AUTOVECTORES (VECTORES PROPIOS) Y AUTOVALORES (VALORES PROPIOS) DE UNA MATRIZ CUADRADA.

Definición 2.- "Si A es una matriz cuadrada con elementos de un cuerpo conmutativo K, se llama autovalor (valor propio) de A todo elemento de K tal que existe un vector x de K^n verificando:

$$Ax = \lambda x$$

Al conjunto de todos los autovalores de A se le llama espectro de A y se escribe

$$\sigma(A)$$

AUTOVECTORES (VECTORES PROPIOS) Y AUTOVALORES (VALORES PROPIOS) DE UNA MATRIZ CUADRADA.

Teorema 1.- Dada una matriz cuadrada A sobre K:

- A todo autovector de A corresponde un autovalor único, llamado autovalor asociado a x.

- A todo autovalor de f corresponde un subespacio vectorial $V(\lambda)$ de K^n, que está descrito por los vectores x de K^n que verifican $Ax = \lambda x$, se le llama el subespacio propio asociado a λ".

AUTOVECTORES (VECTORES PROPIOS) Y AUTOVALORES (VALORES PROPIOS) DE UNA MATRIZ CUADRADA.

b) *Propiedades de los autovalores y de los autovectores de una matriz cuadrada. Polinomio característico y ecuación característica.*

Teorema 2.- *"Dado una matriz cuadrada A sobre K, para de K, las propiedades siguientes son equivalentes:*

1.- λ *es un autovalor de A.*

2.- $A-\lambda I_n$ *no es invertible.*

3.- $\det\left(A-\lambda I_n\right)=0$

AUTOVECTORES (VECTORES PROPIOS) Y AUTOVALORES (VALORES PROPIOS) DE UNA MATRIZ CUADRADA.

Definición 3.- *Si A es una matriz cuadrada de orden n con elementos de un cuerpo conmutativo K, se llama polinomio característico de A al polinomio en* λ *de grado n*

$$P_A(\lambda) = \det(A - \lambda I_n)$$

Y se llama ecuación característica a la ecuación

$$\det(A - \lambda I_n) = 0$$

AUTOVECTORES (VECTORES PROPIOS) Y AUTOVALORES (VALORES PROPIOS) DE UNA MATRIZ CUADRADA.

Teorema 3.- *"Los autovalores de una matriz triangular son los elementos de la diagonal principal."*

Teorema 4.- *"Una matriz cuadrada A es invertible si y sólo si 0 no es autovalor de A."*

Teorema 5.- *"Si λ es autovalor de A y x un autovector asociado, entonces:*

a) *Para n natural, λ^n es autovalor de A^n con autovector x.*

b) *Si A es invertible, entonces es λ^{-1} autovalor de A^{-1} con autovector x*

c) *Para n entero, λ^n es autovalor de A^n con autovector x.*

AUTOVECTORES (VECTORES PROPIOS) Y AUTOVALORES (VALORES PROPIOS) DE UNA MATRIZ CUADRADA.

Teorema 6.- *Los subespacios vectoriales* $V(\lambda_1)$ *y* $V(\lambda_2)$ *asociados a dos autovalores distintos de una matriz cuadrada A, sólo tienen en común el vector nulo.*

Teorema 7.- *Siendo A una matriz cuadrada, que admite m autovalores distintos dos a dos,* $\lambda_1, \lambda_2, \cdots, \lambda_m$, *el sistema de vectores* $\{x_i\}$ *(i = 1, ..., m), de autovalores no nulos asociados a* λ_i *es libre*

Corolario.- *"Toda matriz cuadrada de orden n tiene como máximo n autovalores distintos dos a dos"*

Teorema 8.- *"Si A es una matriz cuadrada de orden n, los autovalores de A son las raíces de su polinomio característico. Existen n como máximo. Si K es algebraicamente cerrado, A posee n autovalores distintos o confundidos"*

SEMEJANZA Y DIAGONALIZACIÓN

Definición 4 (matrices semejantes).- *Se dice que dos matrices A y B cuadradas de orden n son semejantes, si están asociadas a la misma aplicación lineal f pero respecto a dos bases diferentes. Es decir, existe una matriz P de cambio de base que verifica:*

$$AP = PB \qquad \textit{equivalente a} \qquad P^{-1}AP = B$$

Propiedades (inmediatas).-

(1) La relación de semejanza de matrices es de equivalencia.

(2) det(A) = det(B).

(3) A es invertible si y sólo si B es invertible.

(4) $rg(A) = rg(B)$.

(5) $P_A(\lambda) = P_B(\lambda)$.

(6) $\sigma(A) = \sigma(B)$.

SEMEJANZA Y DIAGONALIZACIÓN

Definición 5 (matrices diagonalizables).- *"Se dice que una matriz cuadrada A es diagonalizable si existe una matriz diagonal D, del mismo orden, emejante a A. Es decir, existe una matriz P de cambio de base que verifica:*

$$AP = PD \qquad \text{equivalente a} \qquad P^{-1}AP = D\text{"}$$

Teorema 9.- *"A es diagonalizable si y sólo si, A tiene n autovectores linealmente independientes, es decir , se puede formar una base del espacio vectorial de referencia formado por autovectores de A"*

Demostración.- Sea A diagonalizable, es decir, existe una base respecto de la cual, la matriz A se puede escribir:

$$A = \begin{bmatrix} \mu_1 & & & & \\ & \ddots & & & \\ & & \mu_i & & \\ & & & \ddots & \\ & & & & \mu_n \end{bmatrix}$$

SEMEJANZA Y DIAGONALIZACIÓN

con polinomio característico $p_A(x) = (\mu_1 - x)\cdots(\mu_i - x)\cdots(\mu_n - x)$,

luego $\mu_i, i = 1,\ldots,n$ son los autovalores de A, y son las raíces de p,

que están en el cuerpo K. Además, existen $a_i, i = 1,\ldots,n$, tales que:

$$Aa_i = \mu_i a_i, i = 1,\ldots,n$$

o bien, en forma matricial:

$$\left[Aa_1\ Aa_2 \cdots Aa_n\right] = \left[a_1\ a_2 \cdots a_n\right] \begin{bmatrix} \mu_1 & & & \\ & \mu_2 & & \\ & & \ddots & \\ & & & \mu_n \end{bmatrix} \qquad (\ *\ )$$

Es decir son n vectores propios de A (base, pues, de E)

SEMEJANZA Y DIAGONALIZACIÓN

Recíprocamente, supongamos que se puede hallar una base de E formada por vectores propios de A, es decir que se verifique

$$Aa_i = \mu_i a_i \,, \; i = 1, \ldots, n$$

lo que es equivalente a la expresión (*) anterior, con matriz asociada diagonal.

Teorema 10.- *"Sea A una matriz cuadrada de orden n. Si* $\lambda_1, \lambda_2, \ldots, \lambda_k$ son k autovalores distintos ($k \leq n$) de *A, y* B_i *es una base del subespacio propio , entonces* $B = B_1 \cup B_2 \cup \cdots \cup B_k$ es un sistema libre*"*

Teorema 11 y definición 6 .- *"Si A es una matriz cuadrada de orden n y si su polinomio característico admite una raíz múltiple* λ *de orden k, se tiene* $1 \leq \dim V(\lambda) \leq k$ *. El valor* $\dim V(\lambda)$ *recibe el nombre de multiplicidad geométrica del autovalor "*

SEMEJANZA Y DIAGONALIZACIÓN

Demostración.- Supongamos que $\lambda \in K$ es una raíz del polinomio característico asociado, es decir $p_A(\lambda)=0$ y x un vector propio asociado ($Ax=\lambda x$). Sea $B=\{a_i\}_{i=1}^{n}$ una base del espacio vectorial de referencia E. Es decir: $x=\sum_{i=1}^{n} x_i a_i$ donde las x_i son las coordenadas de x respecto de la base B.

Así que x_i serán las soluciones del sistema homogéneo:

$$(A-\lambda I_n)x=0$$

con soluciones no triviales, es decir que:

$$r=rg(A-\lambda I_n)<n \quad \text{y} \quad \dim V(\lambda)=n-r$$

Como $V(\lambda)\neq\{0\}$, entonces $1\leq \dim V(\lambda)$

SEMEJANZA Y DIAGONALIZACIÓN

Supongamos que $\dim V(\lambda) = h$ (multiplicidad geométrica); sea $\{a_1, \ldots, a_h\}$ una base de $V(\lambda)$, que serán, por tanto, vectores propios asociados al valor propio λ.

Luego la matriz asociada a f respecto de la base será:

$$A = \begin{bmatrix} \lambda I_h & A' \\ 0 & A'' \end{bmatrix}$$

escrita por bloques, donde A' es de orden h x (n-h) , y A'' es cuadrada de orden (n-h) .

Luego el polinomio característico será:

$$p_A(x) = (\lambda - x)^h \det(A'' - xI_{n-h})$$

SEMEJANZA Y DIAGONALIZACIÓN

Luego $\quad h \leq k \quad \Rightarrow \quad \dim V(\lambda) \leq k$

Es decir la multiplicidad geométrica asociada a cada autovalor es menor o igual que la multiplicidad algebraica de ese autovalor.

Teorema 12 (teorema de la diagonalización).- *"Los enunciados siguientes son equivalentes*

(1) Una matriz cuadrada A de orden n, es diagonalizable.

(2) La unión B de las bases de los subespacios propios de A contiene n vectores.

(3) Las multiplicidades algebraica y geométrica coinciden para cada autovalor, es decir:

$$m_{alg}(\lambda_i) = m_{geom}(\lambda_i)$$

SEMEJANZA Y DIAGONALIZACIÓN

Teorema 13.- *"Si una matriz cuadrada de orden n sobre K, posee n autovalores distintos en K, A es diagonalizable".*

Corolario.- *"Siendo A una matriz cuadrada de orden n con elementos de K, si K es algebraicamente cerrado, para que A sea diagonalizable es suficiente que todas las raíces de su polinomio característico sean simples"*

Parte X

Tema 9. Producto interno. Ortogonalidad

Tema 9

Producto interno y ortogonalidad en espacios vectoriales sobre $\mathbb{R}$

9.1. Longitud, ángulos y ortogonalidad

Vamos a introducir el concepto de ortogonalidad en espacios vectoriales con un ejemplo sencillo. Consideremos un vector v de $\mathbb{R}^2$. De cursos anteriores, sabemos que si utilizamos la representación geométrica de v como un segmento dirigido, podemos usar el teorema de Pitágoras para DEFINIR la **longitud** de v de la forma

$$\|v\| = \sqrt{v_1^2 + v_2^2},$$

donde v_1 y v_2 son las coordenadas de v con respecto a la base canónica $B_0 = (e_1, e_2)$. Ver la figura 9.1(a). Obsérvese que, al representar los ejes cartesianos, entendemos que estamos trabajando con la base canónica de $\mathbb{R}^2$. Un vector de longitud 1 se denomina **vector unitario**.

Sean u, v dos vectores de $\mathbb{R}^2$ y sea $\theta \in [0, \pi]$ el ángulo entre ellos (ver la figura 9.1(b)). Si $0 < \theta < \pi$, los vectores u, v y $w = v - u$ forman un triángulo, donde w es el lado opuesto al ángulo θ.

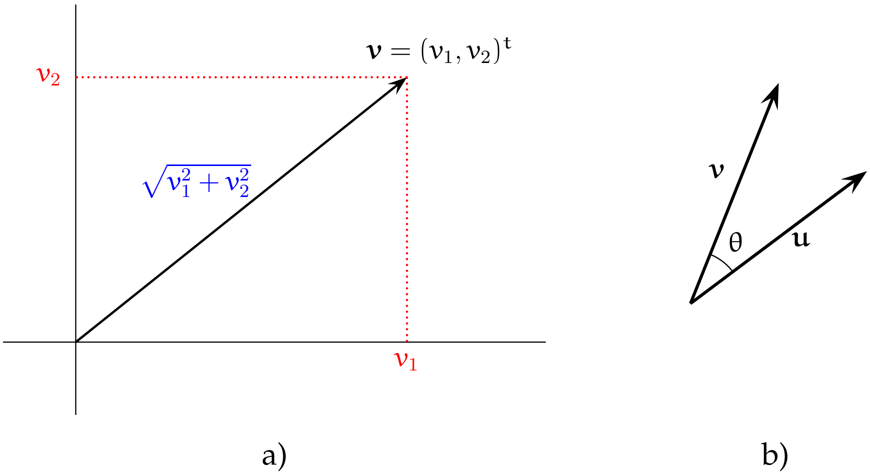


Figura 9.1: a) Representación de la longitud de un vector. b) Representación del ángulo $\theta \in [0, \pi]$ entre dos vectores.

Para determinar el ángulo θ entre u y v utilizamos el *producto escalar usual*:

$$\boxed{\langle u, v \rangle = \underbrace{u^t v}_{\substack{\text{producto}\\ \text{matricial}}}} = u_1 v_1 + u_2 v_2$$

y hacemos uso de la expresión

$$\langle u, v \rangle = \|u\| \, \|v\| \, \cos(\theta)\,.$$

NOTA IMPORTANTE: en la fórmula anterior $u = [u]_{B_0}$; pero si cambiamos la base de $\mathbb{R}^2$ de B_0 a B, entonces hay que tener cuidado porque en general $[u]_B^t \, [v]_B \neq \langle u, v \rangle$.

También sabemos que dos vectores NO NULOS u y v son **ortogonales** (o perpendiculares o normales) si el ángulo entre ellos es $\theta = \dfrac{\pi}{2}$. Si dos vectores unitarios son ortogonales, diremos que son **ortonormales**.

La generalización de estos conceptos a $\mathbb{R}^n$ es sencilla y conocida de cursos anteriores. A continuación abordamos estas ideas en espacios vectoriales arbitrarios sobre el cuerpo $\mathbb{R}$.

9.2. Producto interno y norma sobre $\mathbb{R}$

Sea V un espacio vectorial $\boxed{\text{sobre el cuerpo } \mathbb{R} \text{ de los reales.}}$

Un **producto interno o escalar** definido sobre V es una aplicación entre el conjunto de todos los pares de vectores (u, v) y $\mathbb{R}$, cuyo resultado es un número real denotado por $\langle u, v\rangle$, que satisface las siguientes propiedades para todo $u, v, w \in V$ y todo escalar $\alpha \in \mathbb{R}$:

1. $\langle u, v\rangle = \langle v, u\rangle$.
2. $\alpha \langle u, v\rangle = \langle(\alpha\, u), v\rangle = \langle u, (\alpha\, v)\rangle$.
3. $\langle u + v, w\rangle = \langle u, w\rangle + \langle v, w\rangle$.
4. $\langle u, u\rangle \geqslant 0$ y $\langle u, u\rangle = 0$ si y sólo si $u = 0$.

Ejemplo

El producto escalar *usual* en $\mathbb{R}^2$ es un producto interno. Sean $u = (u_1, u_2)^t, v = (v_1, v_2)^t, w = (w_1, w_2)^t \in \mathbb{R}^2$ y sea $\alpha \in \mathbb{R}$. Entonces:

1. $\langle u, v\rangle = u_1v_1 + u_2v_2 = v_1u_1 + v_2u_2 = \langle v, u\rangle$.
2. $\alpha \langle u, v\rangle = \alpha\,(u_1v_1 + u_2v_2) = (\alpha u_1)v_1 + (\alpha u_2)v_2 = \langle \alpha\, v, u\rangle$ y análogamente para la otra igualdad $\alpha \langle u, v\rangle = \langle v, \alpha\, u\rangle$.
3. $\langle u + v, w\rangle = (u_1 + v_1)w_1 + (u_2 + v_2)w_2 = u_1w_1 + v_1w_1 + u_2w_2 + v_2w_2 = (u_1w_1 + u_2w_2) + (v_1w_1 + v_2w_2) = \langle u, w\rangle + \langle v, w\rangle$.
4. $\langle u, u\rangle = u_1^2 + u_2^2 \geqslant 0$ y además $\langle u, u\rangle = 0 \iff u_1^2 + u_2^2 = 0 \iff u_1 = u_2 = 0 \iff u = 0$.

Ejemplo

Consideremos el espacio vectorial $\mathbb{P}_1$ sobre $\mathbb{R}$ y definamos en él la operación para vectores arbitrarios $p(x) = a_0 + a_1x$ y $q(x) = b_0 + b_1x$ dada por:

$$\langle p, q\rangle = a_0b_0 + 2a_1b_1 \,.$$

Veamos que es un producto interno: para todo $p(x) = a_0+a_1x$, $q(x) = b_0+b_1x$, $t(x) = c_0 + c_1x \in \mathbb{P}_1$ y todo $\alpha \in \mathbb{R}$, se cumple que:

1. $\langle p, q\rangle = a_0b_0 + 2a_1b_1 = b_0a_0 + 2b_1a_1 = \langle q, p\rangle$.
2. $\langle \alpha\, p, q\rangle = \alpha\,(a_0b_0 + 2a_1b_1) = (\alpha\, a_0)b_0 + 2(\alpha\, a_1)b_1 = \langle \alpha\, p, q\rangle$; del mismo modo se prueba la otra igualdad: $\alpha\,\langle p, q\rangle = \langle p, \alpha\, q\rangle$.
3. $\langle p + q, t\rangle = (a_0 + b_0)c_0 + 2(a_1 + b_1)c_1 = a_0c_0 + b_0c_0 + 2a_1c_1 + 2b_1c_1 = (a_0c_0 + 2a_1c_1) + (b_0c_0 + 2b_1c_1) = \langle p, t\rangle + \langle q, t\rangle$.
4. $\langle p, p\rangle = a_0^2 + 2a_1^2 \geqslant 0$ y además $\langle p, q\rangle = 0 \iff a_0^2 + 2a_1^2 = 0 \iff a_1 = a_2 = 0 \iff p(x) = 0$.

Por tanto es un producto interno.

Como vimos en el caso de $\mathbb{R}^2$ (o de $\mathbb{R}^n$ en general), tenemos la posibilidad de definir, a partir del producto escalar, conceptos geométricos tales como la longitud de un vector y la distancia y el ángulo entre dos vectores de dicho espacio. Tales nociones pueden ser generalizadas a cualquier espacio vectorial con producto interno fácilmente. En particular:

Definimos la *longitud* o **norma** $\|u\|$ de un vector $u \in V$ como el número real:

$$\|u\| = \sqrt{\langle u, u\rangle}\,.$$

Ejemplo

En el espacio vectorial $\mathbb{P}_1$ sobre $\mathbb{R}$ con el producto interno

$$\langle a_0 + a_1 x, b_0 + b_1 x \rangle = a_0 b_0 + 2a_1 b_1 \,,$$

vamos a calcular la norma del polinomio $p(x) = 4 - 5x \in \mathbb{P}_1$:

$$\|p\| = \sqrt{\langle p, p \rangle} = \sqrt{\langle 4-5x, 4-5x \rangle} = \sqrt{4 \cdot 4 + 2 \cdot (-5) \cdot (-5)} = \sqrt{66} \,.$$

Proposición

Toda norma definida en V a partir de un producto interno verifica las siguientes propiedades: para todo $u, v \in V$ y todo $\alpha \in \mathbb{K}$, se cumple que

- $\|u\| \geqslant 0$ y $\|u\| = 0 \iff u = 0$ [positividad].
- $\|\alpha u\| = |\alpha| \, \|u\|$ para todo $u \in V$ [homogeneidad].
- $\|u + v\| \leqslant \|u\| + \|v\|$ [desigualdad triangular].

Desigualdad de Cauchy-Schwarz

Dado el espacio vectorial V dotado de producto interno, para todo $u, v \in V$ se tiene:

$$\boxed{|\langle u, v \rangle| \leqslant \|u\| \, \|v\|.}$$

Demostración. Consideremos

$$\|u + \lambda v\|^2 = \langle u + \lambda v, u + \lambda v \rangle = \|u\|^2 + \lambda^2 \|v\|^2 + 2\lambda \langle u, v \rangle \geqslant 0 \,.$$

Por tanto, considerando la expresión $p(\lambda) = \|u\|^2 + \lambda^2 \|v\|^2 + 2\lambda \langle u, v \rangle$, la ecuación $p(\lambda) = 0$ tendrá discriminante $\Delta \leqslant 0$. Dicho discriminante es

$$\Delta = 4 \left(\langle u, v \rangle \right)^2 - 4\|u\|^2 \|v\|^2 \leqslant 0 \quad \Rightarrow \quad \left(\langle u, v \rangle \right)^2 \leqslant \|u\|^2 \, \|v\|^2 .$$

$\square$

También es posible medir la *distancia* entre vectores utilizando la siguiente definición:

Dados los vectores u y v del espacio vectorial V sobre $\mathbb{R}$, dotado de producto interno, se llama **distancia** entre u y v al número real:

$$d(u,v) = \|u - v\| .$$

Este concepto de distancia nos va a permitir en los Temas 13 y 14 abordar el problema de mínimos cuadrados, en el que nuestro objetivo será minimizar la distancia entre ciertos puntos, en espacios vectoriales con producto interno arbitrarios.

Las siguientes propiedades son intuitivas y fáciles de demostrar:

Proposición

Para todo $u, v, w \in V$, se tiene que:

i) $d(u,v) \geqslant 0$.

ii) $d(u,v) = 0 \iff u = v$.

iii) $d(u,v) = d(v,u)$.

iv) Desigualdad triangular: $d(u,w) \leqslant d(u,v) + d(v,w)$.

Ejemplo

Consideremos de nuevo el espacio vectorial $\mathbb{P}_1$ sobre $\mathbb{R}$ y el producto interno:

$$\langle a_0 + a_1 x,\ b_0 + b_1 x\rangle = a_0 b_0 + 2a_1 b_1 .$$

Vamos a calcular la distancia entre los polinomios $p_1(x) = 1$ y $p_2(x) = 1 + 2x$. Como $p_1(x) - p_2(x) = -2x$,

$$d(p_1, p_2) = \|p_1 - p_2\| = \|-2x\| = \sqrt{0^2 + 2 \cdot (-2)^2} = \sqrt{8} .$$

Obviamente,

$$d(p_2,p_1) = \|p_2 - p_1\| = \|2x\| = \sqrt{0^2 + 2\cdot 2^2} = \sqrt{8}\,.$$

Si además consideramos el vector $p_3(x) = 2{-}3x$, podemos comprobar que se verifica la desigualdad triangular. En efecto:

$$\begin{aligned} d(p_2,p_3) &= \|p_2(x) - p_3(x)\| = \|-1+5x\| = \sqrt{(-1)^2 + 2\cdot 5^2} = \sqrt{51}\,, \\ d(p_1,p_3) &= \|p_1(x) - p_3(x)\| = \|-1+3x\| = \sqrt{(-1)^2 + 2\cdot 3^2} = \sqrt{19} \end{aligned}$$

y claramente

$$\sqrt{19} \leqslant \sqrt{8} + \sqrt{51}\,.$$

Obsérvese que en el ejemplo anterior hemos *comprobado* que se cumple la desigualdad triangular para una terna particular de polinomios, lo cual no puede considerarse una *demostración*.

El concepto de **ángulo** θ entre dos vectores u y v en un espacio vectorial V dotado de un producto interno se generaliza mediante la expresión

$$\cos(\theta) = \frac{\langle u, v\rangle}{\|u\|\,\|v\|}\,.$$

Nótese que la desigualdad de Cauchy-Schwarz garantiza que $-1 \leqslant \dfrac{\langle u, v\rangle}{\|u\|\,\|v\|} \leqslant 1$, lo que nos permite escribir $\cos(\theta) = \dfrac{\langle u, v\rangle}{\|u\|\,\|v\|}$ con $\theta \in [0,\pi]$.

Obviamente si dos vectores u y v verifican que $\langle u, v\rangle = 0$, el ángulo comprendido entre ellos, según el producto interno dado, es de $\dfrac{\pi}{2}$; en tal caso, se dice que los vectores son **ortogonales** o *perpendiculares* **con respecto a dicho producto interno**. Cuando el producto interno se sobreentienda, diremos simplemente que son ortogonales.

Ejemplo

Consideremos una vez más el espacio vectorial $\mathbb{P}_1$ sobre $\mathbb{R}$ y el producto interno:

$$\langle a_0 + a_1 x, b_0 + b_1 x\rangle = a_0 b_0 + 2a_1 b_1 \,.$$

Vamos a calcular ahora el ángulo entre los polinomios $p_1(x) = 1$ y $p_2(x) = 1 + 2x$. Tenemos que

$$\begin{aligned} \|p_1\| &= \sqrt{1^2 + 2\cdot 0^2} = 1\,, \\ \|p_2\| &= \sqrt{1^2 + 2\cdot 2^2} = 3\,, \\ \langle p_1, p_2\rangle &= 1\cdot 1 + 2\cdot 0\cdot 2 = 1\,. \end{aligned}$$

Por tanto,

$$\cos(\theta) = \frac{1}{3} \quad \Rightarrow \quad \theta = \arccos\left(\frac{1}{3}\right) \neq 0\,.$$

Obviamente, no son ortogonales según el producto interno dado.

9.2.1. Proyección ortogonal sobre un vector

Como hemos visto, la noción de producto interno permite medir ángulos y, en particular, decidir si dos vectores son o no ortogonales. Para ello, basta con comprobar que su producto interno es cero, ya que en tal caso, el ángulo que formen será de $\frac{\pi}{2}$.

Con frecuencia es necesario obtener (en un cierto espacio vectorial con producto interno) la "proyección ortogonal" de un vector sobre otro. En $\mathbb{R}^2$, esta idea es muy intuitiva. Consideremos dos vectores u y v, representados geométricamente como en la figura 9.2.

La proyección ortogonal del vector v sobre el vector u es otro *vector* que representaremos por $P_u(v)$. En la figura está representado en azul y es perpendicular al segmento punteado rojo.

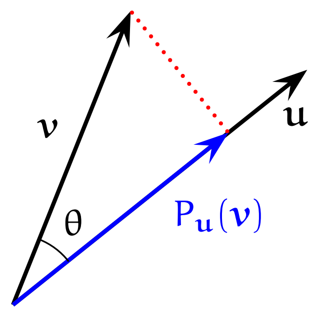


Figura 9.2: Proyección ortogonal de v sobre u.

Utilizando el concepto de producto interno, la longitud de dicha proyección se puede calcular fácilmente mediante

$$\|P_u(v)\| = \frac{\langle u, v\rangle}{\|u\|} .$$

Puesto que la proyección tiene la misma dirección que u, podemos escribir

$$\boxed{P_u(v) = \frac{\langle u, v\rangle}{\|u\|}\,\frac{u}{\|u\|} = \frac{\langle u, v\rangle}{\langle u, u\rangle}u.}$$

Si u es un vector unitario ($\|u\| = 1$), escribiremos simplemente

$$P_u(v) = \langle u, v\rangle\, u\,.$$

Ejemplo

Vamos a encontrar ahora la proyección del vector $p_1(x) = 1$ sobre el vector $p_2(x) = 1 + 2x$ del espacio vectorial $\mathbb{P}_1$ con el producto interno utilizado anteriormente. Tenemos que

$$P_{p_2}(p_1) = \frac{\langle p_1, p_2\rangle}{\langle p_2, p_2\rangle}\,p_2(x) = \frac{1}{9}\,(1 + 2\,x)\ .$$

En general, si V es un espacio vectorial dotado de un producto interno y consideramos cualquier vector unitario $u \in V$, podemos definir una aplicación P_u de V en V mediante

$$P_u(v) = \overbrace{\langle u, v \rangle}^{\text{escalar}} \underbrace{u}_{\text{vector}}$$

para todo $v \in V$. Tal aplicación se denomina **proyección ortogonal sobre** u y, para cada vector v de V, llamaremos al vector $P_u(v)$ la *proyección ortogonal de* v *sobre* u.

En temas posteriores definiremos la proyección de un vector, no sobre otro vector, sino sobre un subespacio vectorial.

9.3. Complemento ortogonal

Consideremos el espacio vectorial V con producto interno.

Dos subespacios S_1 y S_2 de V se denominan **subespacios ortogonales** si $\langle s_1, s_2 \rangle = 0$ para cada $s_1 \in S_1$ y para cada $s_2 \in S_2$.

Si S_1 y S_2 son ortogonales, escribiremos $S_1 \perp S_2$.

Obsérvese que el concepto anterior difiere del que definimos a continuación:

Sea S un subespacio de V. El conjunto de todos los vectores de V que son ortogonales a cada vector de S se denota por $S^\perp$ y se denomina **complemento ortogonal de** S. Así:

$$S^\perp = \{x \in V \colon \langle x, s \rangle = 0, \text{ para cada } s \in S\} .$$

Ejemplo

Sean $S_1 = \mathrm{Gen}(e_1)$ y $S_2 = \mathrm{Gen}(e_2)$ dos subespacios de $\mathbb{R}^3$. Es fácil ver que S_1 y S_2 son ortogonales $\boxed{\text{con respecto al producto escalar usual.}}$ Si $s_1 \in S_1$, tenemos que $s_1 = (\alpha, 0, 0)^t$ y, si $s_2 \in S_2$, será $s_2 = (0, \beta, 0)^t$; por tanto

$$\langle s_1, s_2 \rangle = \alpha \cdot 0 + 0 \cdot \beta + 0 \cdot 0 = 0$$

y $S_1 \perp S_2$. No obstante, S_1 y S_2 no son complementos ortogonales, pues e_3 es perpendicular a cualquier vector de S_1 y sin embargo no está en S_2.

Ejemplo

Vamos a calcular el complemento ortogonal del subespacio $S_1 = \mathrm{Gen}(e_1, e_2)$ de $\mathbb{R}^3$ $\boxed{\text{con respecto al producto escalar usual.}}$ Éste estará formado por todos los vectores que son perpendiculares a TODOS los vectores de S_1, es decir:

$$S_1^{\perp} = \left\{ v \in \mathbb{R}^3 \colon \langle v, s \rangle = v^t\, s = 0\,,\ \forall\, s \in S_1 \right\}\ .$$

Para determinar qué vectores pertenecen a $S_1^{\perp}$, observemos que los vectores de S_1 se escriben de la forma $\alpha_1 e_1 + \alpha_2 e_2 = (\alpha_1, \alpha_2, 0)^t$ con $\alpha_1, \alpha_2 \in \mathbb{R}$. Para que un vector $v = (v_1, v_2, v_3)^t \in \mathbb{R}^3$ sea perpendicular a todos los vectores $w \in S_1$ debe cumplirse que:

$$\langle v, w \rangle = 0 \quad \Rightarrow \quad \alpha_1 v_1 + \alpha_2 v_2 = 0\,,\ \text{ para todo } \alpha_1, \alpha_2 \in \mathbb{R}\,.$$

Esto sólo es posible si $v_1 = v_2 = 0$. Por tanto, los vectores de $S_1^{\perp}$ son de la forma $(0, 0, v_3)^t$ con $v_3 \in \mathbb{R}$ ó en otras palabras:

$$S_1^{\perp} = \mathrm{Gen}(e_3)\,.$$

Ejemplo

Consideremos el espacio vectorial $\mathbb{P}_1$ sobre $\mathbb{R}$ con el producto interno:

$$\langle a_0 + a_1 x\ ,\ b_0 + b_1 x \rangle = a_0 b_0 + 2 a_1 b_1\,.$$

Vamos a calcular el complemento ortogonal del subespacio $S_1 = \mathrm{Gen}(x)$ con

$\boxed{\text{respecto a este producto interno.}}$ Tenemos que

$$\begin{aligned} S_1^{\perp} &= \{p(x) \in \mathbb{P}_1 \colon \langle \alpha x, p(x) \rangle = 0\,, \forall\, \alpha \in \mathbb{R}\} \\ &= \{p(x) = a_0 + a_1 x \colon 2\alpha a_1 = 0\,, \forall\, \alpha \in \mathbb{R}\} \\ &= \{p(x) = a_0 + a_1 x \colon a_1 = 0\} = \mathrm{Gen}(1) = \mathbb{P}_0\,. \end{aligned}$$

En palabras, $S_1^{\perp}$ está formado por los polinomios que sólo constan del término independiente.

Finalmente, enunciamos a continuación algunos resultados interesantes:

Si S_1 y S_2 son subespacios ortogonales de V, entonces $S_1 \cap S_2 = \{0\}$.

Si S es un subespacio de V, entonces $S^{\perp}$ también es un subespacio de V.

Si S es un subespacio de V, entonces

$$\dim(S) + \dim(S^{\perp}) = \dim(V)\,.$$

Además, si $(v_1, \ldots, v_r)$ es una base de S y $(v_{r+1}, \ldots, v_n)$ es una base de $S^{\perp}$, entonces $(v_1, \ldots, v_r, v_{r+1}, \ldots, v_n)$ es una base de V.

Obsérvese que los resultados anteriores nos dicen que $S \oplus S^{\perp} = V$.

Ejemplo

Consideremos el espacio vectorial $\mathbb{P}_1$, con el producto interno definido anteriormente: $\langle a_0 + a_1 x, b_0 + b_1 x \rangle = a_0 b_0 + 2 a_1 b_1$ y sean los subespacios $S_1 = \mathrm{Gen}(x)$ y

$S_1^{\perp} = \mathrm{Gen}(1)$. Evidentemente:

$$\begin{aligned} S_1 \cap S_1^{\perp} &= \{0\}, \\ 2 &= \dim(\mathbb{P}_1) = \dim(S_1) + \dim(S_1^{\perp}) = 1 + 1 \end{aligned}$$

y puede formarse una base de $\mathbb{P}_1$ uniendo las bases $B_{S_1} = (x)$ y $B_{S_1^{\perp}} = (1)$.

Sea $A \in \mathbb{R}^{m \times n}$. El espacio nulo de A es el complemento ortogonal de su espacio fila. Análogamente, el espacio columna de A es el complemento ortogonal del espacio nulo de la traspuesta de A.

Ejemplo

Sea la matriz

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

Es fácil encontrar cada uno de los subespacios asociados a esta matriz:

$$\begin{aligned} N(A) &= \mathrm{Gen}\left((-1,1,1)^t\right), \\ \mathcal{C}(A^t) &= \mathrm{Gen}\left((1,3,-2)^t, (0,1,-1)^t\right), \\ N(A^t) &= \mathrm{Gen}\left((2,0,1)^t\right), \\ \mathcal{C}(A) &= \mathrm{Gen}\left((1,0,-2)^t, (3,1,-6)^t\right). \end{aligned}$$

Obviamente, todos los vectores de $N(A)$ son ortogonales a todos los vectores de $\mathcal{C}(A^t)$, al serlo los vectores que generan ambos subespacios, y todos los vectores de $N(A^t)$ son ortogonales a todos los vectores de $\mathcal{C}(A)$.

También es claro que

$$
\begin{aligned}
\mathsf{N}(A) \cap \mathcal{C}(A^{t}) &= \{0\}, \\
\mathsf{N}(A^{t}) \cap \mathcal{C}(A) &= \{0\}, \\
\mathsf{N}(A)^{\perp} &= \mathcal{C}(A^{t}), \\
\mathsf{N}(A^{t})^{\perp} &= \mathcal{C}(A), \\
\mathbb{R}^{3} &= \mathsf{N}(A) \oplus \mathcal{C}(A^{t}) = \mathsf{N}(A^{t}) \oplus \mathcal{C}(A).
\end{aligned}
$$

I. PRODUCTO ESCALAR Y NORMA.

1. **Definición (Producto escalar usual en $\mathbb{R}^n$).-** *"Sean $x, y \in \mathbb{R}^n$, escritos en forma matricial:*

$$x = [\, x_1 \; x_2 \cdots x_n \,]^T \;, \quad y = [\, y_1 \; y_2 \cdots y_n \,]^T$$

Se define su producto escalar, y se escribe $x \cdot y$ o bien $\langle x, y \rangle$, al escalar:

$$x \cdot y = x^T y = \sum_{i=1}^{n} x_i y_i \text{ ''}$$

2. **Definición (Producto escalar usual en $\mathbb{C}^n$).-** *"Sean $x, y \in \mathbb{C}^n$, escritos en forma matricial:*

$$x = [\, x_1 \; x_2 \cdots x_n \,]^T \;, \quad y = [\, y_1 \; y_2 \cdots y_n \,]^T$$

Se define su producto escalar, y se escribe $x \cdot y$ o bien $\langle x, y \rangle$, al escalar:

$$x \cdot y = x^H y = \sum_{i=1}^{n} \overline{x_i} y_i \text{ ''}$$

3. **Propiedades del producto escalar**.

En $\mathbb{R}^n$: $\forall\, x, y, z \in \mathbb{R}^n,\ \forall\, \alpha \in \mathbb{R}$

1. $x \cdot y = y \cdot x$
2. $x \cdot (y + z) = x \cdot y + x \cdot z$
3. $x \cdot (\alpha\, y) = \alpha\, (x \cdot y)$
4. $(\alpha\, x) \cdot y = \alpha\, (x \cdot y)$
5. $x \cdot x \geq 0$, además: $x \cdot x = 0 \Leftrightarrow x = 0$

En $\mathbb{C}^n$: $\forall\, x, y, z \in \mathbb{C}^n,\ \forall\, \alpha \in \mathbb{C}$

1. $x \cdot y = \overline{y \cdot x}$
2. $x \cdot (y + z) = x \cdot y + x \cdot z$
3. $x \cdot (\alpha\, y) = \alpha\, (x \cdot y)$
4. $(\alpha\, x) \cdot y = \overline{\alpha}\, (x \cdot y)$
5. $x \cdot x \geq 0$, además: $x \cdot x = 0 \Leftrightarrow x = 0$

4. **Definición (Norma en** $\mathbb{R}^n$**).-** *"Sea* $x \in \mathbb{R}^n$, escrito en forma matricial:

$$x = \left[\, x_1\ x_2\ \cdots\ x_n \,\right]^T$$

Se define norma de x, y se escribe $\|x\|$*, al escalar:*

$$\|x\| = +\sqrt{x^T x} = +\sqrt{\sum_{i=1}^{n} |x_i|^2}\ \text{"}$$

5. **Definición (Norma en $\mathbb{C}^n$).-** *"Sea* $x \in \mathbb{C}^n$, escrito en forma matricial:

$$x = \left[\, x_1 \; x_2 \; \cdots \; x_n \,\right]^T$$

Se define norma de x, y se escribe $\|x\|$*, al escalar:*

$$\|x\| = +\sqrt{x^H x} = +\sqrt{\sum_{i=1}^{n} |x_i|^2} \text{ "}$$

6. **Propiedades de la norma**.

En K^n: $\forall\, x, y \in K^n$, $\forall\, \alpha \in K$

1. $\|x\| \geq 0, \quad \|x\| = 0 \Leftrightarrow x = 0$
2. $\|\alpha\, x\| = |\alpha|\,\|x\|$
3. $|x \cdot y| \leq \|x\|\,\|y\|$
4. $\|x + y\| \leq \|x\| + \|y\|$

7. **Definición (**Vectores unitarios).- *"Si* $x \in K^n$*, se dice que x es unitario, si verifica* $\|x\| = 1$".

8. **Definición (**Distancia entre vectores).- *"Si* $x, y \in K^n$*, se define distancia entre x e y, al escalar* $d(x, y) = \|x - y\|$".

9. **Definición (**Vectores ortogonales).- *"Si* $x, y \in K^n$*, se dice que x e y son ortogonales, si verifican* $x \cdot y = 0$".

10. **Teorema de Pitágoras**.- *"Dos vectores* $x, y \in K^n$*, son ortogonales si y sólo si* $\|x+y\|^2 = \|x\|^2 + \|y\|^2$".

11. **Definición (**Complemento ortogonal).- *"Sea V un subespacio vectorial de* K^n*, se define complemento ortogonal de V, se escribe* $V^{\perp}$*, al conjunto de vectores* $V^{\perp} = \left\{ x \in K^n : x \cdot y = 0, \forall y \in V \right\}$".

12. **Teorema**.- *"El complemento ortogonal de un s.e.v. es un subespacio vectorial*".

13. **Teorema**.- *"Sea A una matriz mxn. Entonces el complemento ortogonal del espacio de filas de A es el espacio nulo de A, y el complemento ortogonal del espacio de la columnas es el espacio nulo de* A^T*;* $(FilA)^{\perp} = NulA, \quad (ColA)^{\perp} = NulA^T$".

II. CONJUNTOS ORTOGONALES.

1. **Definición**.- *"Sea* $S = \left\{ v_1, v_2, \ldots, v_p \right\} \subset K^n$*, se dice que S es un conjunto ortogonal si cada par de vectores distintos en el conjunto es ortogonal, es*

decir, $v_i \cdot v_j = 0 \quad \forall\, i \neq j$*. Si además S es una base, entonces se denominará base ortogonal*".

2. **Definición**.- *"Sea* $R = \{u_1, u_2, \ldots, u_p\} \subset K^n$*, se dice que R es un conjunto ortonornal si es ortogonal y todos sus vectores unitarios. Si además R es una base, entonces se denominará base ortonornal*"

3. **Definición**.- *"Sea* $A = [\, a_1\ a_2 \ldots a_n \,] \in K^{m\times n}$ una matriz, se dice que A tiene las columnas ortogonales (u ortonormales) si el conjunto de vectores columna de K^m forman un conjunto ortogonal (ortonormal)".

4. **Definición**.- *"Sea* Q una matriz cuadrada de orden n, se dice que Q es **ortogonal** (si está construida sobre $\mathbb{R}$) o **unitaria** (si está construida sobre $\mathbb{C}$) si sus vectores columna forman un conjunto ortonornal, es decir, se verifica $Q^T Q = I_n$ *sobre* $\mathbb{R}$, o bien $Q^H Q = I_n$ *sobre* $\mathbb{C}$".

5. **Teorema**.- *"Si Q es una matriz con columnas ortonormales, y* $x, y \in K^n$ *dos vectores. Entonces se verifican las siguientes propiedades:*
   a) $\|Qx\| = \|x\|$.
   b) $(Qx).(Qy) = x.y$
   c) $(Qx).(Qy) = 0 \quad \Leftrightarrow \quad x.y = 0$".

6. **Teorema**.- *"Si* $S = \{v_1, v_2, \ldots, v_p\} \subset K^n$ *es un conjunto ortogonal de vectores no nulos, entonces S es un sistema libre, y por lo tanto es una base del subespacio generado por S*".

7. **Teorema**.- *"Sea* $B = \{u_1, u_2, \ldots, u_p\} \subset K^n$ *una base ortogonal de un subespacio W. Entonces todo vector* $x \in W$ *puede escribirse:*

$$x = \left(\frac{x.u_1}{u_1.u_1}\right)u_1 + \left(\frac{x.u_2}{u_2.u_2}\right)u_2 + \cdots + \left(\frac{x.u_p}{u_p.u_p}\right)u_p = \sum_{i=1}^{p}\left(\frac{x.u_i}{u_i.u_i}\right)u_i$$

En el caso en que la base sea ortonormal esa expresión queda reducida a:

$$x = \sum_{i=1}^{p}\left(x.u_i\right)u_i \text{''}.$$

III. PROYECCIONES ORTOGONALES.

1. **Definición**.- *" Sea W un subespacio de* K^n*,* $B = \{u_1, u_2, \ldots, u_p\} \subset K^n$ *una base ortogonal de W y* $x \in K^n$ *un vector cualquiera , entonces, se define:*

$$x_W = proy_W(x) = \sum_{i=1}^{p}\left(\frac{u_i.x}{u_i.u_i}\right)u_i = \sum_{i=1}^{p} proy_{u_i}\left(x\right)$$

como la proyección ortogonal del vector x sobre el subespacio W, y cada sumando de la expresión es *la proyección ortogonal del vector x sobre los subespacios generados por cada vector de la base.*

Si la base es ortonormal, esa expresión queda reducida a:

$$x_W = proy_W(x) = \sum_{i=1}^{p} \left(u_i . x \right) u_i = \sum_{i=1}^{p} proy_{u_i}\left(x \right) \text{”}.$$

2. **Teorema de la descomposición ortogonal**.-

“Sea W un subespacio de K^n*,* $B = \left\{ u_1, u_2, \ldots, u_p \right\} \subset K^n$ *una base ortogonal de W y* $x \in K^n$ *un vector cualquiera , entonces, se puede escribir de forma única:*

$$x = x_W + y\text{, siendo}$$

$$x_W = proy_W(x) = \sum_{i=1}^{p} \left(\frac{u_i . x}{u_i . u_i} \right) u_i = \sum_{i=1}^{p} proy_{u_i}\left(x \right)$$

e $y = x - x_W$ es un vector del complemento ortogonal de W”

3. **Teorema (propiedades de las proyecciones ortogonales)**.-

   (1) $proy_W(x) = x \Leftrightarrow x \in W$

   (2) *Linealidad:* $proy_W(\alpha x + \beta y) = \alpha\, proy_W(x) + \beta\, proy_W(y)$

   (3) $proy_{W^\perp}(x) = x - proy_W\left(x \right)$

4. **Teorema (propiedades del complemento ortogonal)**.-

   (1) $\dim W + \dim W^\perp = n$

   (2) $\left(W^\perp \right)^\perp = W$

(3) $B = B_W \cup B_{W^\perp}$ *es base de* K^n

5. **Teorema (Matriz canónica de la proyección ortogonal).-**
"Sea W un subespacio de K^n, $B = \{u_1, u_2, \ldots, u_p\} \subset K^n$ una base ortonormal de W y $x \in K^n$ un vector cualquiera , entonces:

$$proy_W(x) = QQ^T x, \text{ siendo}$$

$$Q = [u_1\ u_2 \ldots u_p] \in K^{n \times p} \text{ que verifica } Q^T Q = I_p\text{"}$$

6. **Teorema de la aproximación óptima.-**
"Sea W un subespacio de K^n, y $x \in K^n$ un vector cualquiera, entonces:

$$\|x - proy_W(x)\| \le \|x - y\| \quad \forall y \in W.$$

La desigualdad es estricta si $y \neq proy_W(x)$"

III. EL PROCESO DE GRAM-SCHMIDT.

1. **Teorema de Gram-Schmidt.-**

"*Sea B =* $\{v_1, v_2, \cdots, v_p\}$ *base de W subespacio vectorial de* K^n*, definiendo la serie de vectores:*

$$\left.\begin{aligned} w_1 &= v_1 \\ w_2 &= v_2 - \frac{w_1.v_2}{w_1.w_1} w_1 \\ w_3 &= v_3 - \frac{w_1.v_3}{w_1.w_1} w_1 - \frac{w_2.v_3}{w_2.w_2} w_2 \\ &\vdots \\ w_p &= v_p - \sum_{i=1}^{p-1} \frac{w_i v_p}{w_i.w_i} w_i \end{aligned}\right\} w_j = v_j - \sum_{i=1}^{j-1} \frac{w_i.v_j}{w_i.w_i} w_i \ ; \text{j} = 1,\ldots,\text{p}$$

a) El sistema $\{w_1, w_2, \cdots, w_p\}$ *es una base ortogonal de W.*

b) *Gen*$\{w_1, w_2, \cdots, w_j\}$ = *Gen*$\{v_1, v_2, \cdots, v_j\}$ *para j =1, ..., p"*

Demostración.-

a) La construcción de los vectores asegura que w_j es ortogonal a $Gen\{w_1, w_2, \cdots, w_{j-1}\}$, razonando por inducción se llega a que $\{w_1, w_2, \cdots, w_p\}$ es base ortogonal de W.

Veamos como se realiza la inducción:

El hecho de que $\{v_1, v_2, \cdots, v_p\}$ sea base, asegura independencia lineal y no nulos.

Para j=1 y 2 , $w_1 = v_1$ y $w_2 = v_2 - proy_{w_1}(v_2)$ y v_2 no pertenece a Gen{w_1}, entonces w_2 pertenece al complemento ortogonal de Gen{w_1}, luego es ortogonal a w_1.

Supongámoslo cierto para j = k, entonces para j = k+1, $w_{k+1} = v_{k+1} - proy_{W_k}(v_{k+1})$ y $v_{k+1} \notin Gen\{w_1, w_2, ..., w_k\}$, entonces w_{k+1} pertenece al complemento ortogonal de $Gen\{w_1, w_2, ..., w_k\}$, luego es ortogonal a todo $w_1, w_2, ..., w_k$.

q. e. d.

b) INDUCCIÓN:

Claramente cierto para j = 1 ya que $w_1 = v_1$.

Supongamos que es cierto para j = k, entonces $Gen\{w_1, w_2, \cdots, w_k\} = Gen\{v_1, v_2, \cdots, v_k\} = W_k$.

Como $w_{k+1} = v_{k+1} - proy_{W_k}(v_{k+1})$, y además $\begin{cases} v_{k+1} \notin W_k \\ v_{k+1} \neq 0 \end{cases}$ ya que es linealmente independiente a los anteriores, por lo tanto, por el teorema de la base se puede sustituir v_{k+1} por w_{k+1}.

q. e. d.

ORTOGONALIDAD

Parte XI

Tema 10. Bases ortogonales

Tema 10

Bases ortogonales

En el Tema 9 derivamos conceptos tales como la distancia y el ángulo entre vectores (y con éste el de ortogonalidad) o la longitud de vectores a partir de la noción de producto interno. Si dotamos a un espacio vectorial de tal producto, también podemos aplicar el término "ortogonal" a ciertos subespacios vectoriales. En lo que sigue, vamos a aplicar dicho término a otros conceptos.

10.1. Conjuntos ortogonales y bases ortogonales

Un conjunto de vectores $\{u_1, \ldots, u_p\}$ de V se denomina **conjunto ortogonal** si cada par de vectores diferentes del conjunto es ortogonal, es decir, si $\langle u_i, u_j \rangle = 0$ siempre que $i \neq j$.

Ejemplo

Vamos a demostrar que $\{u_1, u_2, u_3\} \subset \mathbb{R}^3$ con $u_1 = (3, 1, 1)^t$, $u_2 = (-1, 2, 1)^t$ y $u_3 = \frac{1}{2}(-1, -4, 7)^t$ es un conjunto ortogonal $\boxed{\text{respecto al producto escalar usual.}}$
Debemos considerar los tres posibles pares de vectores diferentes:

$$
\begin{aligned}
\langle u_1, u_2\rangle &= 3\cdot(-1)+1\cdot 2+1\cdot 1=0\,,\\
\langle u_1, u_3\rangle &= 3\cdot(-1)\frac{1}{2}+1\cdot(-2)+1\cdot\frac{7}{2}=0\,,\\
\langle u_2, u_3\rangle &= -1\cdot(-1)\frac{1}{2}+2\cdot(-2)+1\cdot\frac{7}{2}=0\,.
\end{aligned}
$$

Por tanto, $\{u_1, u_2, u_3\}$ es un conjunto ortogonal.

Teorema

Si $S = \{u_1, \ldots, u_p\}$ es un conjunto ortogonal de vectores no nulos de V, entonces S es linealmente independiente.

Obsérvese que la implicación inversa no es cierta en general.

Ejemplo

Ya sabemos que $\left\{(3,1,1)^t, (-1,2,1)^t, \frac{1}{2}(-1,-4,7)^t\right\}$ es un conjunto ortogonal. Comprobemos que, efectivamente, sus vectores son linealmente independientes; si escribimos la combinación lineal:

$$\alpha(3,1,1)^t + \beta(-1,2,1) + \gamma\frac{1}{2}\,(-1,-4,7)^t$$

y la igualamos a cero, es trivial ver que ha de ser $\alpha = \beta = \gamma = 0$. En cambio, el conjunto $\{(1,1,1)^t, (1,0,1)^t\}$ es linealmente independiente y, obviamente, no es ortogonal.

Un conjunto $\{u_1, \ldots, u_p\}$ es un **conjunto ortonormal** si es un conjunto ortogonal de vectores unitarios.

Ejemplo

Podemos obtener un conjunto ortonormal a partir del conjunto ortogonal $\left\{(3,1,1)^t, (-1,2,1)^t, \frac{1}{2}(-1,-4,7)^t\right\}$ simplemente dividiendo cada vector entre su norma. Así, tendremos que

$$\left\{\frac{1}{\sqrt{11}}\,(3,1,1)^t\,,\frac{1}{\sqrt{6}}\,(-1,2,1)^t\,,\frac{1}{\sqrt{66}}\,(-1,-4,7)^t\,,\right\}$$

es un conjunto ortonormal, como se comprueba de manera trivial.

Obviamente, si S es un conjunto ortogonal de vectores no nulos de V, como es linealmente independiente, será una base del subespacio W generado por S. Puesto que S es tanto un conjunto ortogonal como una base de W la denominaremos **base ortogonal** de W. Si además los vectores de S son unitarios, diremos que S es una **base ortonormal** de W. Si hay $n = \dim(V)$ vectores en S, entonces $W = V$ y S es una base ortogonal (ortonormal) de V.

Veamos una consecuencia importante de esta idea:

Proposición

Sea $(w_1, \ldots, w_p)$ una base ortogonal del subespacio W de V. Entonces, cada w de W tiene una representación única como combinación lineal de $w_1, \ldots, w_p$. De hecho, si

$$w = c_1 w_1 + \cdots + c_p w_p\,,$$

entonces

$$c_j = \frac{\langle w, w_j\rangle}{\langle w_j, w_j\rangle}\,,\quad \text{para todo } j = 1, \ldots, p.$$

Ejemplo

Los vectores $u_1 = (3,1,1)^t$, $u_2 = (-1,2,1)^t$ y $u_3 = \dfrac{1}{2}(-1,-4,7)^t$ forman un conjunto ortogonal S; así, son linealmente independientes y, en consecuencia, forman una base ortogonal de $\mathbb{R}^3$. Vamos a expresar el vector $v = (6,1,-8)^t$ como combinación lineal de estos vectores. Se tiene que

$$\begin{aligned} \langle v, u_1 \rangle &= 11\,, \quad \langle v, u_2 \rangle = -12\,, \quad \langle v, u_3 \rangle = -33\,, \\ \langle u_1, u_1 \rangle &= 11\,, \quad \langle u_2, u_2 \rangle = 6\,, \quad \langle u_3, u_3 \rangle = \frac{33}{2}\,. \end{aligned}$$

Entonces

$$v = \frac{11}{11}\,u_1 - \frac{12}{6}\,u_2 - \frac{33}{33/2}\,u_3 = u_1 - 2u_2 - 2u_3\,.$$

En otras palabras, las coordenadas de v con respecto a la base S son:

$$[v]_S = (1,-2,-2)^t\,.$$

El siguiente resultado relaciona el concepto de conjuntos ortonormales con las matrices ortogonales.

Teorema

Una matriz A $(m \times n)$ es ORTOGONAL (es decir, $A^t A = I$) si y sólo si tiene columnas ORTONORMALES.

Además, como veremos en el Tema 12, dada una matriz ortogonal cuadrada Q, ésta representa, respecto a alguna base ortogonal, un cierto tipo de transformaciones lineales con la propiedad de transformar bases ortonormales en bases ortonormales.

A continuación estudiamos un método que nos permite obtener una base ortogonal para cualquier subespacio W de un espacio vectorial dado V, a partir de cualquier base de W. Obviamente, si queremos obtener una base ortonormal, bastará con normalizar los vectores obtenidos con dicho método.

10.1.1. Procedimiento de Gram-Schmidt

El procedimiento de Gram-Schmidt es un algoritmo simple para obtener una base ortogonal (u ortonormal, normalizando los nuevos vectores) a partir de otra que no lo es. Para ilustrar cómo funciona, consideremos el siguiente ejemplo:

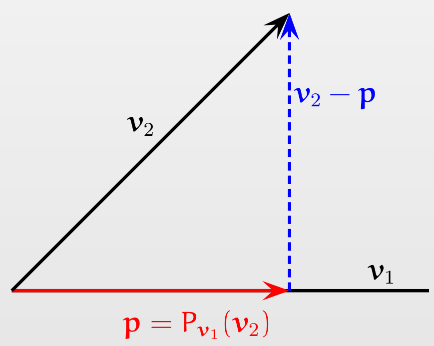
Ejemplo

Sea S el subespacio de $\mathbb{R}^2$ con el producto escalar usual y definido por $S = \text{Gen}(v_1, v_2)$, con $v_1 = (3, 6, 0)^t$ y $v_2 = (1, 2, 2)^t$, linealmente independientes y no ortogonales (su producto interno no es nulo).

Si queremos construir una base ortogonal de S, podemos considerar

$$p = P_{v_1}(v_2) = \frac{15}{45}(3, 6, 0)^t ,$$

de manera que $v_2 - p$ sea ortogonal a v_1 (ver la figura).



Obviamente, $w_2 = v_2 - p = (0, 0, 2)^t$ pertenece a S, puesto que es una combinación lineal de v_1 (con coeficiente $-\|P_{v_1}(v_2)\|$) y de v_2 (con coeficiente 1). Además es ortogonal a v_1. De esta manera, el par (v_1, w_2) constituye una base ortogonal de S.

Finalmente obtenemos una base ortonormal B_S de S dividiendo cada vector por su norma:

$$B_S = \left(\frac{1}{\sqrt{5}}(1, 2, 0)^t , (0, 0, 1)^t \right) .$$

Generalicemos esta idea.

Método de Gram-Schmidt.

Sea una base $(v_1, \dots, v_p)$ del subespacio S de V. Definimos:

$$\begin{aligned}
w_1 &= v_1\,, \\
w_2 &= v_2 - \frac{\langle v_2, w_1\rangle}{\langle w_1, w_1\rangle} w_1\,, \\
w_3 &= v_3 - \frac{\langle v_3, w_1\rangle}{\langle w_1, w_1\rangle} w_1 - \frac{\langle v_3, w_2\rangle}{\langle w_2, w_2\rangle} w_2\,, \\
&\vdots \\
w_p &= v_p - \frac{\langle v_p, w_1\rangle}{\langle w_1, w_1\rangle} w_1 - \frac{\langle v_p, w_2\rangle}{\langle w_2, w_2\rangle} w_2 - \cdots - \frac{\langle v_p, w_{p-1}\rangle}{\langle w_{p-1}, w_{p-1}\rangle} w_{p-1}\,.
\end{aligned}$$

Entonces $(w_1, \dots, w_p)$ es una base ortogonal de S. Además

$$\mathrm{Gen}(v_1, \dots, v_k) = \mathrm{Gen}(w_1, \dots, w_k), \quad 1 \leqslant k \leqslant p\,.$$

Ejemplo

Consideremos el espacio vectorial $\mathbb{P}_1$ sobre $\mathbb{R}$ con el producto interno:

$$\langle a_0 + a_1 x, b_0 + b_1 x\rangle = a_0 b_0 + 2a_1 b_1\,.$$

Consideremos la base $B_1 = (p_1, p_2) = (1 + x, 1 - x)$. Claramente, los vectores de esta base no son ortogonales, ya que $\langle p_1, p_2\rangle = 1 \cdot 1 + 2 \cdot 1 \cdot (-1) = -1 \neq 0$. Vamos a utilizar el método de Gram-Schmidt para obtener una base ortogonal de $\mathbb{P}_1$. En primer lugar, consideramos el vector $q_1 = p_1$; el segundo vector de la base q_2 lo calculamos mediante:

$$q_2 = p_2 - \frac{\langle p_2, q_1\rangle}{\langle q_1, q_1\rangle} q_1 = p_2 + \frac{1}{3} q_1 = \frac{2}{3}(2 - x)\,.$$

Por tanto, la base $B_2 = (q_1, q_2)$ es ortogonal, como se comprueba fácilmente.

Si estamos interesados en utilizar una base ortonormal, dividimos cada vector de B_2

por su norma:

$$q_1' = \frac{1}{\sqrt{3}}(1+x), \quad q_2' = \frac{2}{\sqrt{3}}(2-x).$$

10.1.2. Factorización QR

Una aplicación adicional del método de Gram-Schmidt es la siguiente. Si $\{A_1, \dots, A_n\}$ son las columnas de una matriz A de dimensión $m \times n$ con columnas linealmente independientes, podemos aplicar Gram-Schmidt a $A_1, \dots, A_n$. Normalizando, es posible obtener una descomposición de A de la forma descrita en el siguiente teorema. Esta factorización es muy útil para simplificar diversos problemas, como la resolución de sistemas de ecuaciones o el cálculo de la inversa de una matriz.

Teorema: La factorización QR

Si A es una matriz $m \times n$ con columnas linealmente independientes, entonces A puede ser factorizada de la forma $A = QR$, donde Q es una matriz $m \times n$ cuyas columnas forman una base ortonormal de $\mathcal{C}(A)$ y R es una matriz $n \times n$ triangular superior e invertible con componentes positivas en su diagonal principal.

Ejemplo

Vamos a encontrar la factorización QR de

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

Las columnas de A son linealmente independientes, por lo que podemos encontrar una base ortonormal para $\mathrm{Gen}(A_1, A_2, A_3) = \mathcal{C}(A)$. En primer lugar, hacemos $w_1 =$

$A_1 = (1,1,1,1)^t$. A continuación hacemos

$$w_2 = A_2 - \frac{\langle A_2, w_1\rangle}{\langle w_1, w_1\rangle}\, w_1 = \left((0,1,1,1)^t - \frac{3}{4}(1,1,1,1)^t\right) = \frac{1}{4}\,(-3,1,1,1)^t\,.$$

Para simplificar los cálculos, escalamos w_2 usando un factor de 4, es decir:

$$w_2' = (-3,1,1,1)^t\,.$$

Finalmente, obtenemos el tercer vector de la base ortogonal de la forma:

$$\begin{aligned} w_3 &= A_3 - \frac{\langle A_3, w_1\rangle}{\langle w_1, w_1\rangle} w_1 - \frac{\langle A_3, w_2'\rangle}{\langle w_2', w_2'\rangle} w_2' \\ &= (0,0,1,1)^t - \frac{2}{4}(1,1,1,1)^t - \frac{2}{12}(-3,1,1,1)^t \\ &= \frac{1}{3}(0,-2,1,1)^t\,. \end{aligned}$$

Es decir, una base ortogonal sencilla de $\mathcal{C}(A)$ sería:

$$\left((1,1,1,1)^t, (-3,1,1,1)^t, (0,-2,1,1)^t\right)\,.$$

Ahora podemos normalizar la base; los correspondientes vectores formarán las columnas de Q en la factorización:

$$Q = \begin{pmatrix} 1/2 & -3/\sqrt{12} & 0 \\ 1/2 & 1/\sqrt{12} & -2/\sqrt{6} \\ 1/2 & 1/\sqrt{12} & 1/\sqrt{6} \\ 1/2 & 1/\sqrt{12} & 1/\sqrt{6} \end{pmatrix}\,.$$

Puesto que las columnas de Q son ortonormales, tenemos que $Q^t\,Q = I_3$. Ahora necesitamos encontrar la matriz triangular superior R que verifica $A = Q\,R$. Si multiplicamos ambos miembros de esta expresión por Q^t resulta

$$Q^t\,A = Q^t\,Q\,R = I_3\,R = R\,.$$

Así:

$$
\begin{aligned}
R = Q^t\,A &= \begin{pmatrix} 1/2 & 1/2 & 1/2 & 1/2 \\ -3/\sqrt{12} & 1/\sqrt{12} & 1/\sqrt{12} & 1/\sqrt{12} \\ 0 & -2/\sqrt{6} & 1/\sqrt{6} & 1/\sqrt{6} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \\
&= \begin{pmatrix} 2 & 3/2 & 1 \\ 0 & 3/\sqrt{12} & 2/\sqrt{12} \\ 0 & 0 & 2/\sqrt{6} \end{pmatrix}.
\end{aligned}
$$

Parte XII

Tema 11.- Diagonalización ortogonal. Teorema de la descomposición espectral

Tema 11

El teorema espectral en $\mathbb{R}$

En este tema concluimos el estudio de los valores y vectores propios de una matriz cuadrada analizando un tipo particular de matrices: las simétricas de coeficientes reales. Aunque de apariencia simple, éstas aparecen en multitud de campos como la física, la estadística, la geometría o la ingeniería, lo que las dota de especial interés. En este tema estudiaremos las características del proceso de diagonalización de dichas matrices, es decir, el cálculo de las matrices P y D tales que $A = P\,D\,P^{-1}$.

11.1. Diagonalización de matrices simétricas reales

Ejemplo

Consideremos la siguiente matriz cuadrada y simétrica:

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

Veamos cuáles son los valores propios y las correspondientes multiplicidades, para ver si es diagonalizable. La ecuación característica es

$$-(\lambda-8)(\lambda-6)(\lambda-3)=0$$

y sus raíces son $\lambda_1=8$, $\lambda_2=6$ y $\lambda_3=3$. Obviamente, las multiplicidades algebraicas y geométricas de todas ellas coinciden y por tanto A es diagonalizable.
Los espacios propios asociados tienen bases dadas por

$$B_{\lambda_1}=\left((-1,1,0)^t\right)\ ;\quad B_{\lambda_2}=\left((-1,-1,2)^t\right)\ ;\quad B_{\lambda_3}=\left((1,1,1)^t\right)\ .$$

Estos tres vectores forman una base de $\mathbb{R}^3$; además, es fácil comprobar que también forman un conjunto ortogonal. Si en lugar de usar estos vectores para obtener la matriz P en el proceso de diagonalización, usamos los vectores unitarios correspondientes, P será una matriz con columnas ortonormales, por lo que será una matriz ortogonal, cuya inversa se calcula simplemente como $P^{-1}=P^t$. Por tanto, el proceso de diagonalización da

$$A=P\,D\,P^t=\begin{pmatrix}-1/\sqrt{2} & -1/\sqrt{6} & 1/\sqrt{3}\\ 1/\sqrt{2} & -1/\sqrt{6} & 1/\sqrt{3}\\ 0 & 2/\sqrt{6} & 1/\sqrt{3}\end{pmatrix}\begin{pmatrix}8&0&0\\0&6&0\\0&0&3\end{pmatrix}\begin{pmatrix}-1/\sqrt{2} & 1/\sqrt{2} & 0\\ -1/\sqrt{6} & -1/\sqrt{6} & 2/\sqrt{6}\\ 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3}\end{pmatrix}.$$

Las observaciones que hemos hecho en el ejemplo anterior son generales. Tenemos los siguientes importantes resultados:

Teorema espectral

Una matriz A de dimensión $n\times n$, de entradas $\boxed{\text{reales}}$ y $\boxed{\text{simétrica}}$ verifica las siguientes propiedades:

1. A tiene n $\boxed{\text{valores propios reales}}$ (incluyendo multiplicidades).

2. Si λ es un valor propio de A con multiplicidad k, entonces el espacio propio asociado a λ es k-dimensional.

3. Los espacios propios son mutualmente ortogonales, es decir: dos vectores propios que corresponden a dos valores propios distintos son ortogonales.

4. Existen una matriz ortogonal P y una matriz diagonal D tales que $A = P\,D\,P^t$.

El teorema espectral nos dice que en el caso de las matrices simétricas (de coeficientes reales) los espacios propios forman una "descomposición" ortogonal de $\mathbb{R}^n$ y es posible encontrar una base ortonormal de $\mathbb{R}^n$ formada por vectores propios de A.

El teorema anterior justifica la siguiente definición:

Una matriz $A \in \mathbb{R}^{n\times n}$ se dice que es **ortogonalmente diagonalizable** si y sólo si es diagonalizable mediante una matriz de diagonalización P ortogonal, es decir, si existen P ortogonal y D diagonal tales que $A = P\,D\,P^t$.

Ejemplo

Consideremos la siguiente matriz cuadrada y simétrica:

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

Por ser simétrica y real será ortogonalmente diagonalizable. Vamos a encontrar sus valores propios (que sabemos que han de ser reales) y los correspondientes espacios propios (que sabemos que serán ortogonales), junto con una terna de vectores propios ortogonales.

La ecuación característica es

$$-(\lambda-1)^2(\lambda+2)=0,$$

con raíces: $\lambda_1 = \lambda_2 = 1$ y $\lambda_3 = -2$. Las multiplicidades algebraicas y geométricas de todos ellos han de coincidir (por ser A simétrica). Si determinamos los espacios propios asociados, se comprueba que éstos tienen bases dadas por

$$B_{\lambda_1} = (v_1, v_2) = \left((-1,1,0)^t, (-1,0,1)^t\right); \quad B_{\lambda_3} = (v_3) = \left((1,1,1)^t\right).$$

Es claro que v_1 y v_2 son ortogonales a v_3, pero en cambio v_1 y v_2 no son ortogonales. Esto significa que, si buscamos una matriz P que sea ortogonal para diagonalizar A, no podremos hacer uso de estos vectores. En su lugar habrá que buscar dos vectores ortonormales que generen el mismo espacio propio. Claramente, conseguiremos este propósito usando el método de Gram-Schmidt y dividiendo por la norma los vectores obtenidos. Así, las columnas de P serán:

$$w_1 = \frac{1}{\sqrt{2}}(-1,1,0)^t, w_2 = \frac{1}{\sqrt{6}}(-1,-1,2)^t, w_3 = \frac{1}{\sqrt{3}}(1,1,1)^t$$

y tendremos

$$\begin{aligned} A &= P\,D\,P^t \\ &= \begin{pmatrix} -1/\sqrt{2} & -1/\sqrt{6} & 1/\sqrt{3} \\ 1/\sqrt{2} & -1/\sqrt{6} & 1/\sqrt{3} \\ 0 & 2/\sqrt{6} & 1/\sqrt{3} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix} \begin{pmatrix} -1/\sqrt{2} & 1/\sqrt{2} & 0 \\ -1/\sqrt{6} & -1/\sqrt{6} & 2/\sqrt{6} \\ 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \end{pmatrix}. \end{aligned}$$

Además, tenemos el siguiente resultado fundamental:

Teorema

Una matriz $A \in \mathbb{R}^{n\times n}$ es ortogonalmente diagonalizable si y sólo si A es simétrica.

11.2. Teorema de descomposición espectral

Además del teorema espectral y del teorema de caracterización de las matrices ortogonalmente diagonalizables, las matrices reales simétricas de dimensión $n \times n$ satisfacen lo que se conoce como teorema de descomposición espectral, que permite escribir (*descomponer*) éstas como una suma de n matrices derivadas a partir de sus valores y vectores propios. Antes de abordarlo, introducimos en esta sección algunas definiciones y resultados.

Dado un espacio vectorial V sobre $\mathbb{R}$ dotado de producto interno, diremos que una transformación lineal $T : V \to V$ es una **proyección** (o un *proyector*) si

$$T \circ T = T\,.$$

Es frecuente utilizar la notación Π para referirse a proyectores. Recordemos que:

Una matriz cuadrada A se dice **idempotente** si $A^2 = A$.

Ambas definiciones se relacionan de la siguiente manera:

Proposición

Si la proyección $\Pi \colon V \to V$ se representa por la matriz A_Π relativa a una base ortonormal de V, entonces A_Π es idempotente.

Además introducimos la siguiente definición:

Sea un espacio vectorial V sobre $\mathbb{R}$ dotado de producto interno. Dada la transformación lineal $T \colon V \to V$, definimos la **transformación adjunta de** T como la transformación lineal $T^* \colon V \to V$ tal que para todo $v_1, v_2 \in V$:

$$\langle v_1, T^*(v_2)\rangle = \langle T(v_1), v_2\rangle\,.$$

Es importante observar que para toda transformación lineal T la transformación adjunta existe y es única.

Proposición

Si la transformación lineal $T\colon V \to V$ se representa por la matriz A_T relativa a una base ortonormal de V, entonces T^* se representa mediante A_T^t.

Ambos conceptos se emplean para definir un nuevo tipo de transformación lineal (desde el punto de visto de la geometría, como veremos en el Tema 12).

Dado un espacio vectorial V sobre $\mathbb{R}$ dotado de producto interno, diremos que una transformación lineal $\Pi\colon V \to V$ es una **proyección ortogonal** (o un *proyector ortogonal*) si

1. $\Pi \circ \Pi = \Pi$ (es decir, Π es un proyector).
2. $\Pi^* = \Pi$ (es decir, Π es *autoadjunta*).

Y obviamente, tenemos el siguiente resultado:

Dado un espacio vectorial V sobre $\mathbb{R}$ dotado de producto interno, si la proyección ortogonal $\Pi\colon V \to V$ se representa por la matriz A_Π relativa a una base ortonormal de V, entonces:

1. $A_\Pi^2 = A_\Pi$ (es decir, A_Π es idempotente).
2. $A_\Pi^t = A_\Pi$ (es decir, A_Π es *simétrica*).

Ejemplo

Consideremos el espacio $\mathbb{P}_1$ sobre $\mathbb{R}$ dotado del producto interno

$$\langle a_0 + a_1 x, b_0 + b_1 x \rangle = a_0 b_0 + a_1 b_1 \,.$$

La prueba de que efectivamente es producto interno es inmediata. Sea la transformación lineal Π que asocia a cada polinomio $a_0 + a_1 x$ el polinomio $\Pi(a_0 + a_1 x) = 2a_0 + a_1 - (2a_0 + a_1)x$. Con respecto a la base ortonormal de $\mathbb{P}_1$ dada por $B = (1, x)$, la transformación Π se representa mediante la matriz

$$A_{\Pi,B} = \begin{pmatrix} 2 & 1 \\ -2 & -1 \end{pmatrix} .$$

Evidentemente, tenemos que $A_{\Pi,B}$ es idempotente:

$$A_{\Pi}^2 = \begin{pmatrix} 2 & 1 \\ -2 & -1 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ -2 & -1 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ -2 & -1 \end{pmatrix} = A_{\Pi} ,$$

por lo que Π es una proyección. Sin embargo, A_{Π} no es simétrica, por lo que Π no es proyección ortogonal.

La transformación adjunta Π^* de Π la calcularíamos como sigue: consideremos la imagen por Π^* de los vectores de la base B (suficiente para describir Π^*). Se debe cumplir, por una parte, que:

$$\begin{aligned} \langle a_0 + a_1 x, \underbrace{\alpha + \beta x}_{\Pi^*(1)} \rangle &= \langle \underbrace{2a_0 + a_1 - (2a_0 + a_1)x}_{\Pi(a_0 + a_1 x)}, 1 \rangle \\ \Rightarrow \quad a_0\alpha + a_1\beta = 2a_0 + a_1 \,, &\quad \Rightarrow \quad \alpha = 2 \,, \beta = 1 \end{aligned}$$

y por otra que:

$$\begin{aligned} \langle a_0 + a_1 x, \underbrace{\gamma + \delta x}_{\Pi^*(x)} \rangle &= \langle \underbrace{2a_0 + a_1 - (2a_0 + a_1)x}_{\Pi(a_0 + a_1 x)}, x \rangle \\ \Rightarrow \quad a_0\gamma + a_1\delta = -2a_0 - a_1 \,, &\quad \Rightarrow \quad \gamma = -2, \delta = -1 \,. \end{aligned}$$

Así, la imagen por Π^* de un polinomio arbitrario $a_0 + a_1 x$ vendrá dada por:

$$\begin{aligned} \Pi^*(a_0 + a_1 x) &= a_0 \, \Pi^*(1) + a_1 \, \Pi^*(x) = a_0(2 + x) + a_1(-2 - x) \\ &= 2a_0 - 2a_1 + (a_0 - a_1)x \,. \end{aligned}$$

Si representamos Π^* mediante una matriz respecto a la base B, se obtiene que

$$A_{\Pi^*,B} = \begin{pmatrix} 2 & -2 \\ 1 & -1 \end{pmatrix} = A^t_{\Pi,B} \,.$$

Proposición

Si Π es una proyección ortogonal entonces $\ker(\Pi) = \mathrm{Im}(\Pi)^\perp$.

Ejemplo

Consideremos en $\mathbb{R}^3$ con el producto escalar ordinario la transformación Π que asocia a cada vector $v = (v_1, v_2, v_3)^t$ la imagen dada por

$$\Pi\left((v_1, v_2, v_3)^t\right) = \frac{1}{9}\left(v_1 + 2(v_2 + v_3), 2v_1 + 4(v_2 + v_3), 2v_1 + 4(v_2 + v_3)\right)^t \,.$$

Obviamente, respecto a la base canónica, la representación de Π viene dada por la matriz simétrica

$$A_\Pi = \frac{1}{9}\begin{pmatrix} 1 & 2 & 2 \\ 2 & 4 & 4 \\ 2 & 4 & 4 \end{pmatrix},$$

por lo que Π es autoadjunta. También es fácil ver que $A_\Pi^2 = A_\Pi$, por lo que Π será una proyección ortogonal. Es inmediato observar que $\mathrm{Im}(\Pi) = \mathrm{Gen}\left((1, 2, 2)^t\right)$.

El kernel de Π es

$$\ker(\Pi) = \{v \in \mathbb{R}^3 \colon \Pi(v) = 0\} = \mathrm{Gen}\left((-2, 1, 0)^t, (-2, 0, 1)^t\right),$$

que obviamente es el complemento ortogonal de $\mathrm{Im}(\Pi)$.

Teorema

Sean $\Pi_1, \ldots, \Pi_r$ proyecciones ortogonales sobre un espacio vectorial real V con producto interno que satisfacen:

$$\begin{aligned} \Pi_1 + \cdots + \Pi_r &= I \quad \text{(aplicación identidad)}, \\ \Pi_i \circ \Pi_j &= 0 \quad \text{(aplicación nula)}, \quad \text{para todo } i \neq j. \end{aligned}$$

Sea $U_i = \mathrm{Im}(\Pi_i)$. Entonces

$$V = U_1 \oplus U_2 \oplus \cdots \oplus U_r$$

y si $u_i \in U_i$ y $u_j \in U_j$ para todo $i \neq j$, entonces $\langle u_i, u_j \rangle = 0$.

Una descomposición de un espacio en forma de suma directa que satisface las condiciones del teorema anterior se denomina una **descomposición ortogonal** de V. Dicho teorema es el escenario general del teorema de descomposición espectral de las matrices reales simétricas, que enunciamos a continuación.

Teorema de descomposición espectral

Sea A una matriz real simétrica de dimensión $n \times n$ que podemos escribir en la forma $A = P\,D\,P^t$, donde las columnas de P son las coordenadas de los vectores propios *ortonormales* $v_1, \ldots, v_n$ de A y los correspondientes valores propios $\lambda_1, \ldots, \lambda_n$ están en la diagonal principal de la matriz D. Entonces

$$A = \sum_{i=1}^{n} \lambda_i\, A_i = \lambda_1 A_1 + \cdots + \lambda_n A_n\,,$$

donde las matrices $A_i = v_i\, v_i^t$ $(1 \leqslant i \leqslant n)$ son simétricas, idempotentes y verifican que

$$\begin{aligned}
\sum_{i=1}^{n} A_i &= \sum_{i=1}^{n} v_i\, v_i^t = I_n\,, \\
A_i A_j &= \left(v_i\, v_i^t\right)\,\left(v_j\, v_j^t\right) = 0_{n\times n}\,, \quad \forall\, i \neq j\,.
\end{aligned}$$

Obsérvese que las matrices A_i representan proyectores ortogonales Π_i, que verifican las condiciones del teorema de descomposición ortogonal. En estas condiciones, la transformación lineal $T = \sum_{i=1}^{n} \lambda_i \Pi_i$ está representada por la matriz simétrica real A y obviamente tiene valores propios $\lambda_1, \ldots, \lambda_n$.

Ejemplo

Consideremos de nuevo la matriz

$$\begin{aligned}
A &= \begin{pmatrix} 6 & -2 & -1 \\ -2 & 6 & -1 \\ -1 & -1 & 5 \end{pmatrix} \\
&= \underbrace{\begin{pmatrix} -1/\sqrt{2} & -1/\sqrt{6} & 1/\sqrt{3} \\ 1/\sqrt{2} & -1/\sqrt{6} & 1/\sqrt{3} \\ 0 & 2/\sqrt{6} & 1/\sqrt{3} \end{pmatrix}}_{P} \underbrace{\begin{pmatrix} 8 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 3 \end{pmatrix}}_{D} \underbrace{\begin{pmatrix} -1/\sqrt{2} & 1/\sqrt{2} & 0 \\ -1/\sqrt{6} & -1/\sqrt{6} & 2/\sqrt{6} \\ 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \end{pmatrix}}_{P^t}.
\end{aligned}$$

Utilizando la descomposición espectral podemos escribir:

$$\begin{aligned}
A &= 8\,\frac{1}{\sqrt{2}}\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}\frac{1}{\sqrt{2}}\,(-1,1,0) + 6\,\frac{1}{\sqrt{6}}\begin{pmatrix} -1 \\ -1 \\ 2 \end{pmatrix}\frac{1}{\sqrt{6}}\,(-1,1,2) \\
&\quad +3\,\frac{1}{\sqrt{3}}\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}\frac{1}{\sqrt{3}}\,(1,1,1) \\
&= 8\underbrace{\left[\frac{1}{2}\begin{pmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}\right]}_{A_1} + 6\underbrace{\left[\frac{1}{6}\begin{pmatrix} 1 & 1 & -2 \\ 1 & 1 & -2 \\ -2 & -2 & 4 \end{pmatrix}\right]}_{A_2} + 3\underbrace{\left[\frac{1}{3}\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}\right]}_{A_3}
\end{aligned}$$

cuya suma, efectivamente, es A.

Es trivial comprobar que

$$\begin{aligned}
A_1\,A_1 &= A_1\,, \quad A_2\,A_2 = A_2\,, \quad A_3\,A_3 = A_3 \quad \text{(idempotentes)}\,, \\
A_1^t &= A_1\,, \quad A_2^t = A_2\,, \quad A_3^t = A_3 \quad \text{(simétricas)}\,, \\
I_3 &= A_1 + A_2 + A_3\,, \\
0_{3\times 3} &= A_1\,A_2 = A_1\,A_3 = A_2\,A_3\,.
\end{aligned}$$

Ejemplo

Consideremos la matriz

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

con valores propios $\lambda_1 = 5$, $\lambda_2 = 2$ y $\lambda_3 = -1$ y vectores propios $v_1 = (2,2,1)^t$, $v_2 = (2,-1,-2)^t$ y $v_3 = (1,-2,2)^t$. El teorema de descomposición espectral nos dice que podemos escribir A como la suma de las tres matrices:

$$M_1 = \frac{\lambda_1}{\|v_1\|^2} v_1\, v_1^t = \frac{1}{9}\begin{pmatrix} 20 & 20 & 10 \\ 20 & 20 & 10 \\ 10 & 10 & 5 \end{pmatrix},$$

$$M_2 = \frac{\lambda_2}{\|v_2\|^2} v_2\, v_2^t = \frac{1}{9}\begin{pmatrix} 8 & -4 & -8 \\ -4 & 2 & 4 \\ -8 & 4 & 8 \end{pmatrix},$$

$$M_3 = \frac{\lambda_3}{\|v_3\|^2} v_3\, v_3^t = \frac{1}{9}\begin{pmatrix} -1 & 2 & -2 \\ 2 & -4 & 4 \\ -2 & 4 & -4 \end{pmatrix}.$$

Es fácil ver que las matrices

$$A_1 = \frac{1}{\|v_1\|^2} v_1\, {v_1}^t = \frac{1}{5} M_1\,, \quad A_2 = \frac{1}{\|v_2\|^2} v_2\, {v_2}^t = \frac{1}{2} M_2\,, \quad A_3 = \frac{1}{\|v_3\|^2} v_3\, {v_3}^t = -M_3$$

son idempotentes (representan proyecciones Π_i) y simétricas (las proyecciones son ortogonales). También es inmediato comprobar que

$$A_1 + A_2 + A_3 = I_3 \quad \text{(matriz identidad)}$$

o equivalentemente que $\Pi_1 + \Pi_2 + \Pi_3 = I_3$ (aplicación identidad) y que

$$A_i\, A_j = 0_{3\times 3} \quad \text{(matriz nula) para todo } i \neq j$$

o equivalentemente que $\Pi_i \circ \Pi_j = 0$ (aplicación nula) para todo $i \neq j$. Finalmente, se observa que

$$\mathrm{Im}(\Pi_1) = \mathrm{Gen}(v_1)\,, \quad \mathrm{Im}(\Pi_2) = \mathrm{Gen}(v_2)\,, \quad \mathrm{Im}(\Pi_3) = \mathrm{Gen}(v_3)\,,$$

por lo que $\mathbb{R}^3$ puede descomponerse como suma directa en la forma:

$$\mathbb{R}^3 = \mathrm{Im}(\Pi_1) \oplus \mathrm{Im}(\Pi_2) \oplus \mathrm{Im}(\Pi_3)\,.$$

Esto indica que, como sabemos, se puede formar una base de $\mathbb{R}^3$ con vectores propios de la matriz simétrica A.

DESCOMPOSICIÓN ESPECTRAL

Teorema.- Sea A una matriz cuadrada de orden n sobre el cuerpo de los números reales:

A es simétrica si, y sólo si, es ortogonalmente diagonalizable, ésto es, $D = P^T A P$, con P ortogonal y con los autovalores de A (repetidos de acuerdo a sus multiplicidades) en la diagonal principal de D

***Demostración de ⇒ (inducción en el orden de la matriz)*.-**

Las dos partes del teorema son claramente ciertas cuando el orden $k = 1$.

Supongamos el teorema válido para $k = n$.

Sea A, ahora, de orden $n + 1$.

Sea λ_1 un autovalor de A asociado al autovector v_1 normalizado; v_1 se puede usar para completar una base ortonormal.

Sea $\{v_1, w_1, w_2, \dots, w_n\}$ ortonormal, luego la matriz $U = [v_1\ w_1\ w_2 \dots w_n] = [v_1\ \ W]$ es ortogonal.

Sea, ahora, $A' = U^T A U = [v_1\ W]^T A\ [v_1\ W] =$

$= [v_1\ \ W]^T\ [Av_1\ \ AW\ ] = [v_1\ \ W]^T\ [\lambda_1 v_1\ \ AW\ ] =$

$$= \begin{bmatrix} v_1^T \\ W^T \end{bmatrix} \begin{bmatrix} \lambda_1 v_1 & AW \end{bmatrix} = \begin{bmatrix} v_1^T \lambda_1 v_1 & v_1^T AW \\ W^T \lambda_1 v_1 & W^T AW \end{bmatrix} = \begin{bmatrix} \lambda_1 & v_1^T AW \\ 0 & C \end{bmatrix} =$$

Como A' es semejante a A, tiene los mismos autovalores, incluyendo sus multiplicidades.

Luego

$$\det(A' - \lambda I) = (\lambda_1 - \lambda).\det(C - \lambda I)$$

Por lo tanto el espectro de A es el mismo de C, además de λ_1, por lo que la hipótesis inductiva es válida, ya que C es de orden n, luego existirá V, ortogonal tal que

$$V^T C V = D' \text{ que será diagonal.}$$

Sea ahora

$$P = U\begin{bmatrix} 1 & 0 \\ 0 & V \end{bmatrix}$$

$$P^T AP = \begin{bmatrix} 1 & 0 \\ 0 & V^T \end{bmatrix} U^T AU \begin{bmatrix} 1 & 0 \\ 0 & V \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & V^T \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 \\ 0 & C \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & V \end{bmatrix} =$$

$$= \begin{bmatrix} 1 & 0 \\ 0 & V^T \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 \\ 0 & CV \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & V^T CV \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & D' \end{bmatrix}$$

que es diagonal.

Parte XIII

Tema 12. Geometría y Transformaciones Lineales

Tema 12

Geometría de las transformaciones lineales en $\mathbb{R}$

12.1. Matrices y transformaciones ortogonales reales

Por lo aprendido en temas anteriores, sabemos que, además de a vectores y a subespacios vectoriales, el término "ortogonal" se aplica a ciertas matrices $A \in \mathbb{R}^{m \times n}$: las que verifican $A^t A = I_n$. En este tema vamos a ver que podemos aplicarlo también a ciertas transformaciones lineales, siempre relacionado con algún producto interno, y también a cómo se relaciona, en estos casos, con el concepto geométrico de ortogonalidad.

Comenzamos introduciendo algunos resultados.

Teorema

Sea A una matriz $m \times n$ con columnas ortonormales y sean v y w elementos de $\mathbb{R}^n$. Entonces:

1) $\|A\,v\| = \|v\|$.

2) $\langle A\,v, A\,w\rangle = \langle v, w\rangle$.

3) $\langle A\,v, A\,w\rangle = 0$ si y sólo si $\langle v, w\rangle = 0$.

Las propiedades 1) y 3) dicen que la transformación lineal asociada a A, $v \mapsto A\,v$, conserva la longitud y la ortogonalidad; además, la propiedad 2) garantiza que también se conserva el ángulo.

Veamos ahora qué significado e implicaciones tiene la ortogonalidad en las transformaciones lineales.

Una **transformación lineal** T de U en V se dice que es **ortogonal** si preserva el producto interno de los vectores; es decir, para todo $u_1, u_2 \in U$ se cumple que

$$\boxed{\langle T(u_1), T(u_2)\rangle = \langle u_1, u_2\rangle.}$$

Proposición

Si $T : U \to V$ es ortogonal entonces preserva la longitud de los vectores:

$$\|T(u)\| = \|u\|\,, \quad \text{para todo } u \in U.$$

En consecuencia, a menudo se denomina a las transformaciones ortogonales **isometrías**.

El siguiente resultado es fundamental para entender una propiedad importante de las transformaciones lineales ortogonales.

Proposición

Si V es un espacio vectorial de dimensión finita con producto interno sobre $\mathbb{R}$, toda isometría lineal $T\colon V \to V$ es un isomorfismo.

De aquí resulta que T transforma cualquier base de V en otra base de V; como además preserva la longitud y el ángulo, es claro que:

Proposición

Si V es un espacio vectorial de dimensión finita con producto interno sobre $\mathbb{R}$ y el isomorfismo $T : V \to V$ es ortogonal (isometría), entonces T transforma bases ortonormales en bases ortonormales.

Además, las matrices ortogonales están íntimamente relacionadas con las transformaciones ortogonales.

Proposición

Una transformación lineal $T \colon V \to V$ es ortogonal si y sólo si las columnas de la matriz A_T que representa a T, *relativa a cualquier base ortogonal*, forman una base ortonormal de V.

Proposición

Sea A_T una matriz cuadrada que representa una transformación lineal ortogonal $T \colon V \to V$, entonces $\det(A_T) = \pm 1$ y los valores propios de A_T satisfacen $|\lambda_i| = 1$.

Otras propiedades interesantes que comparten las matrices ortogonales y las transformaciones lineales que representan son las siguientes:

- El producto de dos matrices ortogonales es ortogonal. Equivalentemente, la composición de dos transformaciones ortogonales es ortogonal.
- La inversa de una matriz ortogonal es ortogonal; equivalentemente, la inversa de una transformación ortogonal es una transformación ortogonal.

Ejemplo

Consideremos en $\mathbb{R}^2$ la transformación que consiste en rotar los vectores un ángulo de $\dfrac{\pi}{2}$ en sentido contrario a las agujas del reloj. Sabemos que podemos construir una matriz asociada a T, respecto a la base canónica, de la forma:

$$A_T = (T(e_1), T(e_2)) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

Obviamente, esta matriz es ortogonal (podemos comprobarlo verificando que $A^t\, A = I$ o también observando que está formada por columnas ortonormales). La transformación T también es ortogonal, pues obviamente conserva la longitud de los vectores (basta comprobar que $\|v\| = \|A_T\, v\|$ para un vector arbitrario v).

Además, T es invertible y su inversa es la transformación S que rota los vectores de $\mathbb{R}^2$ un ángulo de $\dfrac{\pi}{2}$ en sentido horario. En este caso:

$$A_S = (S(e_1), S(e_2)) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

Es inmediato comprobar que

$$A_T^{-1} = A_S$$

y también que

$$A_T^t = A_T^{-1}.$$

Obsérvese que esta sencilla matriz no es diagonalizable en $\mathbb{R}$ ya que sus valores propios $\lambda = \pm i$ no pertenecen a este conjunto de escalares. La situación cambia si consideramos el cuerpo $\mathbb{C}$.

12.2. Geometría de las aplicaciones lineales

Los gráficos de la figura 12.1 muestran aplicaciones de $\mathbb{R}$ en $\mathbb{R}$; dos de ellas, $f_1(x) = e^x$ y $f_2(x) = x^2$, no son lineales y se puede ver que distorsionan el dominio cuando lo trans-

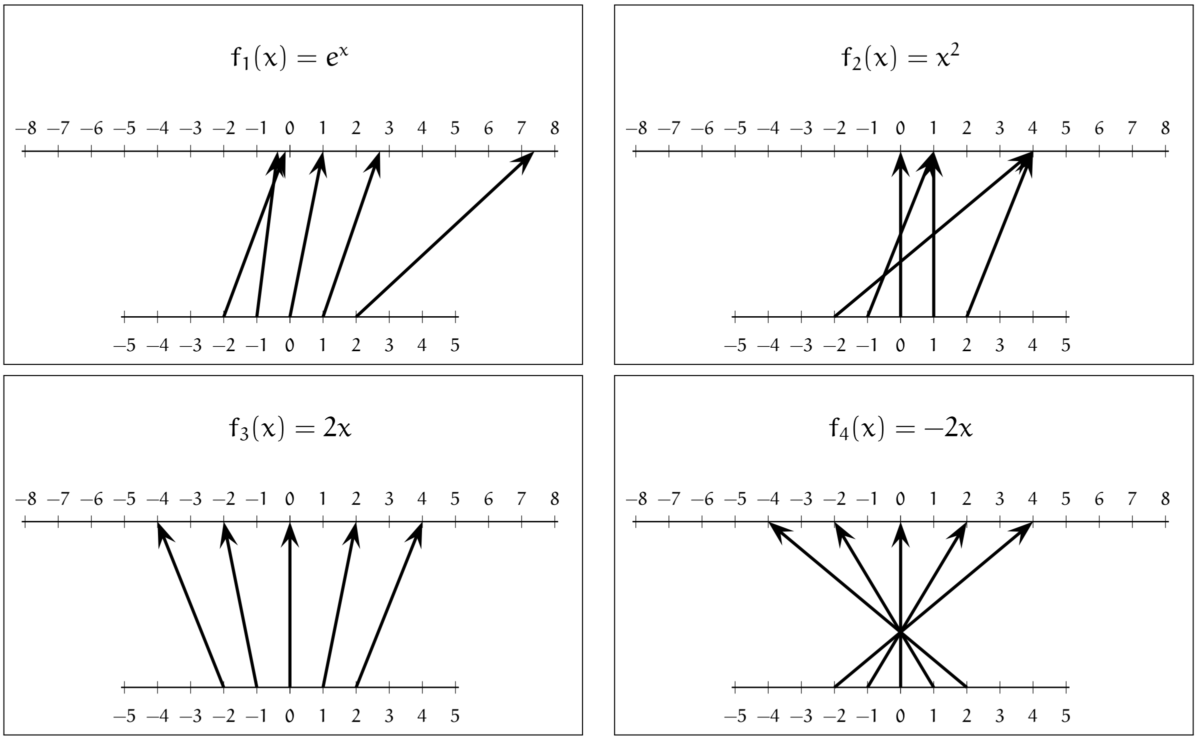


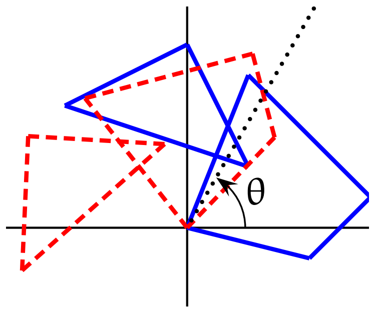
Figura 12.1: Transformaciones lineales y no lineales

forman en la imagen. Por otro lado, $f_3(x) = 2x$ y $f_4(x) = -2x$ son aplicaciones lineales y claramente expanden el dominio de manera uniforme, multiplicando siempre por el mismo factor.

Las únicas aplicaciones lineales de $\mathbb{R}$ en $\mathbb{R}$ son multiplicaciones por escalares, pero en dimensiones mayores existe una variedad mayor de situaciones. Por ejemplo, la transformación lineal de $\mathbb{R}^2$ en $\mathbb{R}^2$ definida por:

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto \begin{pmatrix} \cos(\theta) & \operatorname{sen}(\theta) \\ -\operatorname{sen}(\theta) & \cos(\theta) \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$$

rota los vectores en sentido antihorario un ángulo θ.



En las siguientes secciones vamos a ver algunas transformaciones lineales de gran importancia geométrica. Aunque pueden ser generalizadas a otros espacios con producto interno, nos centraremos en los espacios $\mathbb{R}^2$ y $\mathbb{R}^3$ con el producto escalar habitual, y analizaremos cuáles de ellas son transformaciones ortogonales.

12.2.1. Reflexiones

Una **reflexión** es una transformación lineal T de un espacio vectorial V en sí mismo, en el que existe un hiperplano de puntos fijos, es decir, de puntos cuyas imágenes por T coinciden con ellos mismos; tal conjunto se denomina **hiperplano de reflexión** (o bien **eje de reflexión** si V tiene dimensión 2 ó **plano de reflexión** en dimensión 3).

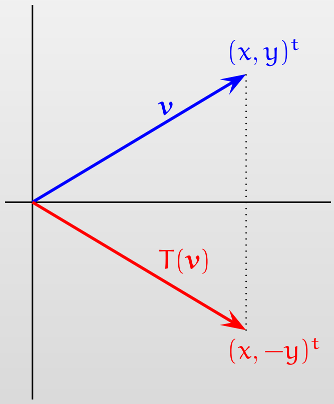
Intuitivamente, podemos visualizar la imagen de un vector por una reflexión mediante su imagen especular en el hiperplano de reflexión.

Veamos algunas reflexiones sencillas, donde todas las coordenadas se refieren a las bases canónicas:

Reflexión respecto al eje X en $\boxed{\mathbb{R}^2}$

Esta reflexión transforma el vector de $\mathbb{R}^2$ con coordenadas $(x, y)^t$ en el vector $(x, -y)^t$. Todos los puntos del eje X son puntos fijos. La matriz asociada a esta transformación lineal viene dada por $A_T = (T(e_1), T(e_2))$. Claramente,

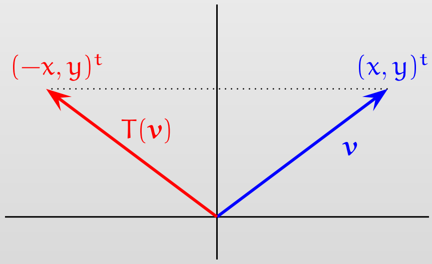
$$A_T = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$



Reflexión respecto al eje Y en $\boxed{\mathbb{R}^2}$

Esta reflexión transforma el vector de $\mathbb{R}^2$ con coordenadas $(x, y)^t$ en el $(-x, y)^t$. Los puntos fijos son los del eje Y. La matriz asociada a esta transformación lineal en esta situación es

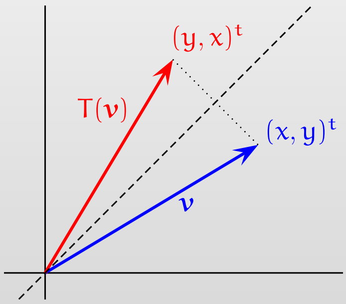
$$A_T = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}.$$



Reflexión respecto a la bisectriz en $\mathbb{R}^2$

Esta reflexión transforma el vector de $\mathbb{R}^2$ con coordenadas $(x, y)^t$ en el vector $(y, x)^t$, por lo que los puntos fijos son los de la recta $y = x$. La matriz asociada es

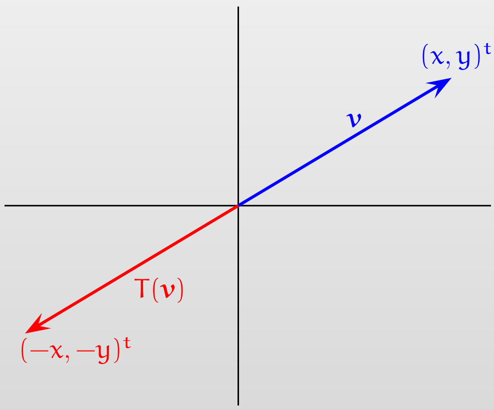
$$A_T = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$



Reflexión respecto al origen en $\mathbb{R}^2$

Esta reflexión transforma el vector de $\mathbb{R}^2$ con coordenadas $(x, y)^t$ en el vector $(-x, -y)^t$; de manera que el conjunto de puntos fijos sólo contiene al origen. La matriz asociada es

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



Reflexión respecto al plano XY en $\boxed{\mathbb{R}^3}$

Los vectores "sobre" el plano XY se "mueven" al semiespacio bajo dicho plano y viceversa, cambiando el signo de la coordenada z. La matriz asociada es

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

Reflexión respecto al plano YZ en $\boxed{\mathbb{R}^3}$

Con un razonamiento semejante, se deduce que la matriz asociada es

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

Reflexión respecto al plano XZ en $\boxed{\mathbb{R}^3}$

Por último, la matriz asociada a esta transformación es

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

Puede comprobarse fácilmente que todas las matrices anteriores representan transformaciones lineales ortogonales. Veámoslo con una de ellas:

Ejemplo

Consideremos la transformación lineal dada por la reflexión respecto al plano XY en $\mathbb{R}^3$, con el producto escalar usual.

Sean dos vectores arbitrarios $u = (u_1, u_2, u_3)^t$ y $v = (v_1, v_2, v_3)^t$, con imágenes respectivas

$$T(u) = (u_1, u_2, -u_3)^t \ , \quad T(v) = (v_1, v_2, -v_3)^t \ .$$

Tenemos que

$$\langle T(u),\ T(v)\rangle = u_1v_1 + u_2v_2 + (-u_3)(-v_3) = u_1v_1 + u_2v_2 + u_3v_3 = \langle u, v\rangle \ ,$$

por lo que es una transformación ortogonal.

Alternativamente, podemos observar que la matriz A_T (respecto a la base ortonormal canónica) es ortogonal, ya que tiene columnas ortonormales y también porque verifica que $A_T^t\, A_T = I_3$.

12.2.2. Contracciones y dilataciones

Estas transformaciones corresponden a multiplicaciones por escalares de la forma: $T(x) = c\,x$ donde c es un escalar no negativo. La transformación se denomina **contracción** si $0 \leqslant c < 1$ y **dilatación** si $c \geqslant 1$.

Contracciones y dilataciones en $\boxed{\mathbb{R}^2}$ y $\boxed{\mathbb{R}^3}$

Obviamente, las coordenadas de un vector $(x, y)^t$ de $\mathbb{R}^2$ se transforman en $(cx, cy)^t$. Así, la matriz asociada a esta transformación lineal viene dada por:

$$A_T = \begin{pmatrix} c & 0 \\ 0 & c \end{pmatrix} .$$

Análogamente, en $\mathbb{R}^3$ la matriz asociada a la dilatación/contracción es:

$$A_T = \begin{pmatrix} c & 0 & 0 \\ 0 & c & 0 \\ 0 & 0 & c \end{pmatrix}.$$

Es evidente que, para $c \neq 1$, las contracciones no son transformaciones ortogonales.

Ejemplo

Consideremos la transformación lineal T dada por una dilatación por un factor $c \neq 1$ en $\mathbb{R}^2$ con el producto escalar usual. Consideremos dos vectores arbitrarios

$$u = (u_1, u_2)^t \ , \qquad v = (v_1, v_2)^t \ ,$$

cuyas imágenes respectivas son

$$T(u) = (cu_1, cu_2)^t \ , \qquad T(v) = (cv_1, cv_2)^t \ .$$

Tenemos que

$$\langle T(u),\ T(v)\rangle = c^2 u_1 v_1 + c^2 u_2 v_2 = c^2(u_1 v_1 + u_2 v_2) \neq u_1 v_1 + u_2 v_2 = \langle u, v\rangle \ ,$$

por lo que no es una transformación ortogonal.

12.2.3. Rotaciones

Rotación en sentido antihorario en $\boxed{\mathbb{R}^2}$ (respecto al origen)

Dado un vector v de coordenadas $(v_1, v_2)^t$ y dada la transformación lineal T que rota el espacio en sentido antihorario un ángulo θ, la matriz asociada está caracterizada por $T(e_1) = (\cos(\theta), \operatorname{sen}(\theta))^t$ y $T(e_2) = (-\operatorname{sen}(\theta), \cos(\theta))^t$; así:

$$A_T = \begin{pmatrix} \cos(\theta) & -\operatorname{sen}(\theta) \\ \operatorname{sen}(\theta) & \cos(\theta) \end{pmatrix}.$$

Rotación en sentido antihorario respecto al eje X en $\boxed{\mathbb{R}^3}$

Es sencillo observar que la matriz asociada viene dada por:

$$A_T = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta) & -\operatorname{sen}(\theta) \\ 0 & \operatorname{sen}(\theta) & \cos(\theta) \end{pmatrix}.$$

Rotación en sentido antihorario respecto al eje Y en $\boxed{\mathbb{R}^3}$

Del mismo modo, la matriz asociada viene dada por:

$$A_T = \begin{pmatrix} \cos(\theta) & 0 & \operatorname{sen}(\theta) \\ 0 & 1 & 0 \\ -\operatorname{sen}(\theta) & 0 & \cos(\theta) \end{pmatrix}.$$

Rotación en sentido antihorario respecto al eje Z en $\boxed{\mathbb{R}^3}$

Finalmente, la matriz asociada en este caso viene dada por:

$$A_T = \begin{pmatrix} \cos(\theta) & -\operatorname{sen}(\theta) & 0 \\ \operatorname{sen}(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Las rotaciones en el espacio $\mathbb{R}^n$ con el producto escalar usual son transformaciones lineales ortogonales (preservan dicho producto escalar).

Ejemplo

Consideremos la transformación lineal dada por la rotación en sentido horario en $\mathbb{R}^2$, con el producto escalar usual, un ángulo θ. Las imagenes de los vectores $u = (u_1, u_2)^t$ y $v = (v_1, v_2)^t$ son:

$$\begin{aligned} T(u) &= (u_1 \cos(\theta) - u_2 \operatorname{sen}(\theta), u_1 \operatorname{sen}(\theta) + u_2 \cos(\theta))^t, \\ T(v) &= (v_1 \cos(\theta) - v_2 \operatorname{sen}(\theta), v_1 \operatorname{sen}(\theta) + v_2 \cos(\theta))^t. \end{aligned}$$

Tenemos que

$$\begin{aligned} \langle T(u), T(v) \rangle &= (u_1 \cos(\theta) - u_2 \operatorname{sen}(\theta))(v_1 \cos(\theta) - v_2 \operatorname{sen}(\theta)) \\ &\quad + (u_1 \operatorname{sen}(\theta) + u_2 \cos(\theta))(v_1 \operatorname{sen}(\theta) + v_2 \cos(\theta)) \\ &= u_1 v_1 \cos^2(\theta) + u_2 v_2 \operatorname{sen}^2(\theta) + u_1 v_1 \operatorname{sen}^2(\theta) + u_2 v_2 \cos^2(\theta) \\ &= u_1 v_1 + u_2 v_2 = \langle u, v \rangle, \end{aligned}$$

por lo que es una transformación ortogonal.

Se deja como ejercicio comprobar que la rotación dada preserva la longitud de los vectores.

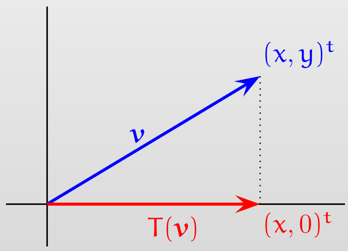
12.2.4. Proyecciones ortogonales

En el Tema 11 introdujimos un tipo especial de transformaciones lineales, los proyectores ortogonales, y estudiamos algunas de sus propiedades. En esta sección vamos a ver algunos ejemplos de proyecciones ortogonales en los espacios $\mathbb{R}^2$ y $\mathbb{R}^3$.

Proyección sobre el eje X en $\boxed{\mathbb{R}^2}$

Esta proyección transforma cada vector de $\mathbb{R}^2$ con coordenadas $(x, y)^t$ en el vector $(x, 0)^t$. La matriz asociada a esta transformación lineal viene dada por:

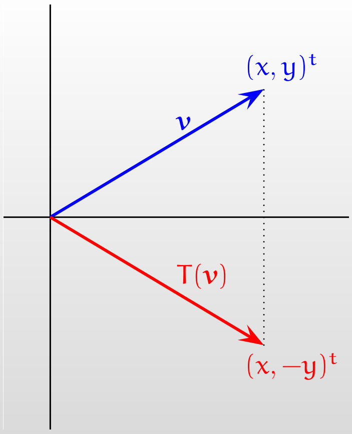
$$A_T = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$



Proyección sobre el eje Y en $\boxed{\mathbb{R}^2}$

Esta proyección transforma cada vector de $\mathbb{R}^2$ con coordenadas $(x, y)^t$ en el vector $(0, y)^t$ y tiene por matriz asociada

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



Proyección sobre el eje X en $\boxed{\mathbb{R}^3}$

Esta proyección transforma el vector de $\mathbb{R}^3$ con coordenadas $(x, y, z)^t$ en el vector $(x, 0, 0)^t$. La matriz asociada será

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

Proyección sobre el eje Y en $\boxed{\mathbb{R}^3}$

Del mismo modo, esta proyección transforma el vector de $\mathbb{R}^3$ con coordenadas $(x, y, z)^t$ en el vector $(0, y, 0)^t$. Ahora, la matriz asociada a la transformación lineal será:

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

Proyección sobre el eje Z en $\boxed{\mathbb{R}^3}$

Por último, esta proyección transforma el vector de $\mathbb{R}^3$ con coordenadas $(x, y, z)^t$ en el vector $(0, 0, z)^t$ y tendrá por matriz asociada:

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

Proyección sobre el plano XY en $\boxed{\mathbb{R}^3}$

Es claro que esta proyección sobre el plano transforma el vector de $\mathbb{R}^3$ con coordenadas $(x, y, z)^t$ en el vector $(x, y, 0)^t$. Por tanto, la matriz asociada a esta transformación lineal será:

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

Proyección sobre el plano XZ en $\boxed{\mathbb{R}^3}$

Ahora la proyección transforma cada vector de $\mathbb{R}^3$ con coordenadas $(x, y, z)^t$ en el vector $(x, 0, z)^t$. Su matriz asociada vendrá dada por:

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

Proyección sobre el plano YZ en $\boxed{\mathbb{R}^3}$

Esta proyección transforma cada vector de $\mathbb{R}^3$ con coordenadas $(x, y, z)^t$ en el vector $(0, y, z)^t$. La matriz asociada a esta transformación lineal será:

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

Es inmediato comprobar que todas las transformaciones descritas verifican las condiciones de proyección ortogonal y que las correspondientes matrices que las representan son idempotentes. No obstante, las proyecciones ortogonales *no son transformaciones lineales ortogonales*, es decir, no son isometrías, como ilustra el siguiente ejemplo.

Ejemplo

Consideremos en $\mathbb{R}^3$, con el producto escalar usual, la proyección ortogonal T que proyecta ortogonalmente sobre el plano XY. T transforma los vectores $u_1 = (1, 1, 1)^t$ y $u_2 = (0, 0, 1)^t$ en los vectores $T(u_1) = (1, 1, 0)^t$ y $T(u_2) = (0, 0, 0)^t$, respectivamente. Es inmediato comprobar que T no es ortogonal pues no se verifica la definición:

$$\langle T(u_1), T(u_2) \rangle = 0 \neq 1 = \langle u_1, u_2 \rangle .$$

Alternativamente, también podríamos haber observado que T no preserva la longitud:

$$\|T(u_1)\| = \sqrt{2} \neq \sqrt{3} = \|u_1\| .$$

Otras opciones para probar que no es transformación ortogonal serían calcular el determinante de A_T (que vale 0) u observar que las columnas de A_T no forman una base ortonormal de $\mathbb{R}^3$.

Parte XIV

Tema 13. Mínimos cuadrados

Tema 13

Mínimos cuadrados

13.1. El problema de los mínimos cuadrados

En temas anteriores hemos estudiado el problema de resolver sistemas lineales de la forma $A\,x = b$; en términos de "distancia", cuando intentamos resolver tales sistemas, lo que buscamos es el conjunto de vectores x que hacen que la distancia entre $A\,x$ y b sea igual a 0:

$$\|A\,x - b\| = 0$$

(o, equivalentemente para evitar raíces cuadradas, $\|A\,x - b\|^2 = 0$). Si el conjunto de soluciones de $A\,x = b$ es vacío, con frecuencia estamos interesados en encontrar un vector x_0 que haga $\|A\,x_0 - b\|^2$ tan pequeño como sea posible (en una cierta norma). Tal vector lo denominaremos *solución de mínimos cuadrados* del sistema $A\,x = b$. Este término se deriva del hecho de que $\|A\,x - b\|$ es la raíz cuadrada de una suma de términos al cuadrado.

Dados una matriz A de dimensión $m \times n$ y un vector $b \in \mathbb{R}^m$, un vector $x_0 \in \mathbb{R}^n$ se denomina una **solución de mínimos cuadrados de** $A\,x = b$ si y sólo si para todo

$x \in \mathbb{R}^n$:

$$\|A\,x_0 - b\|^2 \leqslant \|A\,x - b\|^2\,.$$

Obsérvese que, si x_0 resuelve el sistema $A\,x = b$, entonces es una solución de mínimos cuadrados.

Los problemas de mínimos cuadrados aparecen en muchas situaciones; la más típica es la siguiente. Supongamos que un cierto experimento físico proporciona la colección de datos experimentales $(x_1, y_1), \ldots, (x_m, y_m)$, correspondientes a las magnitudes físicas X e Y, que teóricamente satisfacen la ley

$$Y = \alpha + \beta\, X$$

(i.e., el gráfico de la relación entre ambas es una recta). Los datos medidos proporcionan un sistema lineal de la forma

$$\begin{cases} y_1 = \alpha + \beta\, x_1\,, \\ \quad\vdots \\ y_m = \alpha + \beta\, x_m\,, \end{cases}$$

que puede escribirse usando la matriz ampliada como sigue:

$$\left(\begin{array}{cc|c} 1 & x_1 & y_1 \\ \vdots & \vdots & \vdots \\ 1 & x_m & y_m \end{array}\right)$$

o en formal matricial como

$$A\,x = b\,,$$

donde

$$A = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_m \end{pmatrix}\;;\quad x = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}\;;\quad b = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix}\,.$$

Los datos experimentales (que de manera inevitable contienen "ruido"), como los mostrados en la figura 13.1, conducen a un sistema incompatible (i.e., no existe una línea recta que pase por todos ellos), por lo que buscaremos una solución de mínimos cuadrados. La siguiente sección estudia cómo hacerlo.

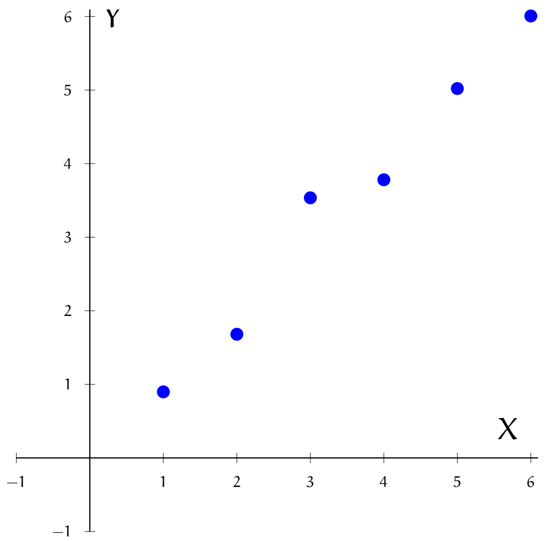
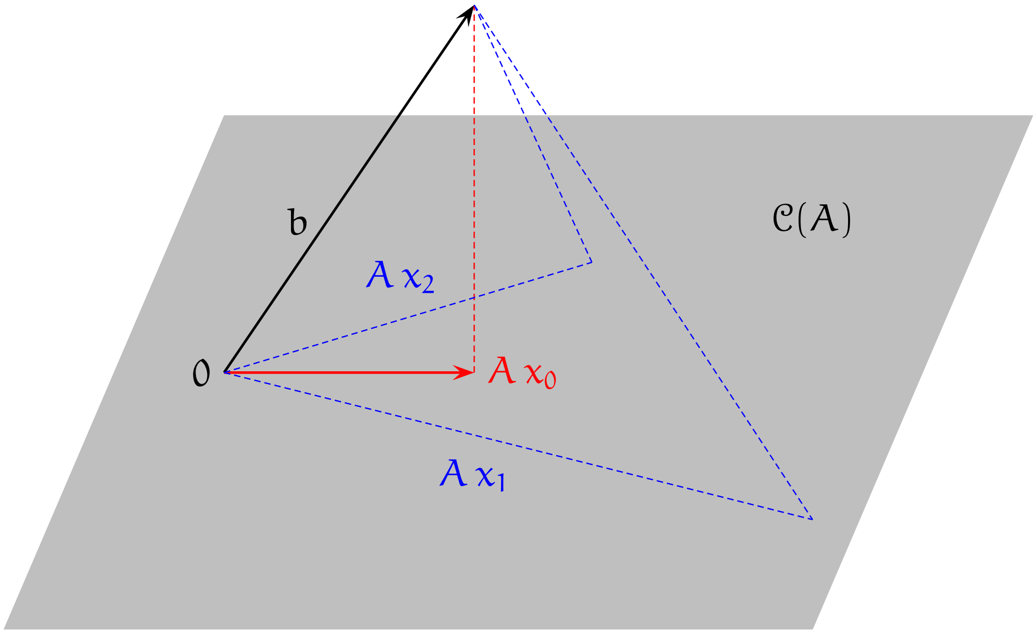


Figura 13.1: Gráfico de seis puntos experimentales, no alineados debido al ruido.

13.1.1. Interpretación geométrica de la solución de mínimos cuadrados

Es evidente que la solución de mínimos cuadrados x_0 del sistema $A\,x = b$ satisfará que el vector $A\,x_0$ pertenece al espacio columna de A, $\mathcal{C}(A)$, y además será el vector que haga que la distancia de $A\,x_0$ al vector b sea mínima.



Obviamente, la proyección ortogonal de b sobre el espacio columna de A proporcionará $A\,x_0$. Supongamos que x_0 satisface $A\,x_0 = P_{\mathcal{C}(A)}(b)$. Tal proyección cumple que $b - P_{\mathcal{C}(A)}(b)$ es ortogonal a $\mathcal{C}(A)$ y así $b - A\,x_0$ es ortogonal a cada columna A_j de A:

$$A_j^t\,(b - A\,x_0) = 0\,, \quad j = 1, \ldots, m\,.$$

Estas ecuaciones pueden ser escritas en una única expresión matricial de la forma

$$A^t\,(b - A\,x_0) = 0 \quad \Rightarrow \quad A^t\,A\,x_0 = A^t\,b\,.$$

Esto prueba el siguiente teorema:

Teorema

El conjunto de soluciones de mínimos cuadrados de $A\,x = b$ coincide con el conjunto de soluciones *no vacío* del sistema $\boxed{A^t\,A\,x = A^t\,b.}$

El sistema escrito en forma matricial $A^t\,A\,x = A^t\,b$ es denominado sistema de **ecuaciones normales para** x_0. Para encontrar la solución de mínimos cuadrados de $A\,x = b$, el primer paso consiste en multiplicar por la izquierda ambos miembros por A^t. Así, obtenemos

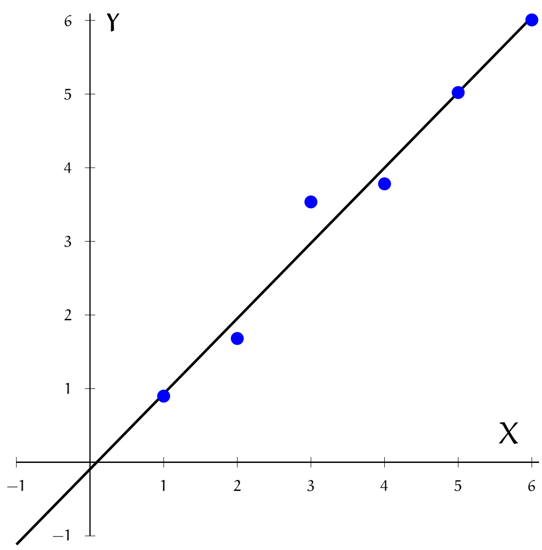
$$\begin{pmatrix} 1 & \dots & 1 \\ x_1 & \dots & x_n \end{pmatrix} \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} 1 & \dots & 1 \\ x_1 & \dots & x_n \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix},$$

$$\underbrace{\begin{pmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{pmatrix}}_{A^t\,A} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \underbrace{\begin{pmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i y_i \end{pmatrix}}_{A^t\,b}.$$

Si la matriz $A^t\,A$ *es invertible*, el sistema $A^t\,A\,x = A^t\,b$ tiene *solución única* dada por:

$$\boxed{x_0 = (A^t\,A)^{-1}\,A^t\,b\,.}$$

Para los puntos experimentales anteriormente citados, la línea recta $Y = \alpha + \beta X$ cuyos coeficientes se encuentran por mínimos cuadrados es la que se muestra en la siguiente figura:



Si la matriz A^tA *no es invertible*, entonces el sistema $A^t\,A\,x = A^t\,b$ tendrá infinitas soluciones (que encontraremos, por ejemplo, utilizando el método de Gauss).

Ejemplo

Vamos a obtener la solución de mínimos cuadrados del sistema $A\,x = b$, donde

$$A = \begin{pmatrix} 1 & -6 \\ 1 & -2 \\ 1 & 1 \\ 1 & 7 \end{pmatrix}, \quad b = \begin{pmatrix} -1 \\ 2 \\ 1 \\ 6 \end{pmatrix}.$$

Calculamos la solución de la forma:

$$\begin{aligned} x_0 &= (A^t\,A)^{-1}\,A^t\,b \\ &= \left[\begin{pmatrix} 1 & 1 & 1 & 1 \\ -6 & -2 & 1 & 7 \end{pmatrix} \begin{pmatrix} 1 & -6 \\ 1 & -2 \\ 1 & 1 \\ 1 & 7 \end{pmatrix} \right]^{-1} \begin{pmatrix} 1 & 1 & 1 & 1 \\ -6 & -2 & 1 & 7 \end{pmatrix} \begin{pmatrix} -1 \\ 2 \\ 1 \\ 6 \end{pmatrix} \\ &= \underbrace{\begin{pmatrix} 4 & 0 \\ 0 & 90 \end{pmatrix}}_{A^t\,A}^{-1} \underbrace{\begin{pmatrix} 8 \\ 45 \end{pmatrix}}_{A^t\,b} = \frac{1}{2} \begin{pmatrix} 4 \\ 1 \end{pmatrix}, \end{aligned}$$

que es la solución de mínimos cuadrados del problema.

En general, si las columnas de $A \in \mathbb{R}^{m\times n}$ son linealmente independientes, la proyección ortogonal de un vector $b \in \mathbb{R}^n$ en el espacio columna de A está dado por

$$P_{\mathcal{C}(A)}(b) = A\,x_0 = A\,(A^t\,A)^{-1}\,A^t\,b\,.$$

Ejemplo

Consideremos la matriz (con columnas linealmente independientes)

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix}.$$

Su espacio columna tiene dimensión 2 (un plano en $\mathbb{R}^3$). La proyección ortogonal de cualquier vector b en este subespacio se calcula usando:

$$P_{\mathcal{C}(A)}(b) = A\,(A^t\,A)^{-1}\,A^t\,b = \frac{1}{6}\begin{pmatrix} 5 & 2 & -1 \\ 2 & 2 & 2 \\ -1 & 2 & 5 \end{pmatrix} b\,.$$

Sea ahora

$$b = (1,0,0)^t \notin \mathcal{C}(A)\,.$$

Tenemos que

$$P_{\mathcal{C}(A)}(b) = \frac{1}{6}\begin{pmatrix} 5 & 2 & -1 \\ 2 & 2 & 2 \\ -1 & 2 & 5 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \frac{1}{6}\begin{pmatrix} 5 \\ 2 \\ -1 \end{pmatrix}.$$

Resolviendo el sistema

$$A\,x = P_{\mathcal{C}(A)}(b)\,,$$

vemos fácilmente que:

$$P_{\mathcal{C}(A)}(b) = \frac{1}{6}\,(5,2,1)^t \in \mathcal{C}(A)\,.$$

Ahora consideremos

$$b = (1,1,1)^t \in \mathcal{C}(A)\,.$$

Su proyección sobre $\mathcal{C}(A)$ es lógicamente

$$P_{\mathcal{C}(A)}(b) = \frac{1}{6}\begin{pmatrix} 5 & 2 & -1 \\ 2 & 2 & 2 \\ -1 & 2 & 5 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = b\,.$$

Ejemplo

Consideremos el subespacio S de $\mathbb{R}^3$ descrito por:

$$S = \left\{(x, y, z)^t \in \mathbb{R}^3 \colon x + y + z = 0\right\}\,.$$

Claramente es de dimensión 2 (un plano). Podemos calcular la proyección ortogonal de cualquier vector $b \in \mathbb{R}^3$ en este subespacio como sigue. En primer lugar, hallamos una matriz cuyo espacio columna coincida con S. Para ello, usamos dos vectores linealmente independientes v_1, v_2 de S y construimos la matriz

$$A = (v_1, v_2)\ ,$$

puesto que es claro que el espacio columna de tal matriz es S. Una manera sencilla de hacerlo consiste en escribir

$$\begin{aligned} S &= \left\{(x, y, z)^t \in \mathbb{R}^3 \colon z = -x - y\right\} = \left\{(x, y, -x-y)^t \in \mathbb{R}^3\right\} \\ &= \left\{x(1, 0, -1)^t + y(0, 1, -1)^t \colon x, y \in \mathbb{R}\right\} \\ &= \mathrm{Gen}\left((1, 0, -1)^t, (0, 1, -1)^t\right)\,. \end{aligned}$$

Por tanto

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

Ahora es inmediato calcular las proyecciones de cualquier vector b en S usando:

$$\begin{aligned} P_S(b) &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{pmatrix} \underbrace{\left[\frac{1}{3} \begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \end{pmatrix} \right]}_{(A^t A)^{-1} A^t} b \\ &= \frac{1}{3} \begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix} b\,. \end{aligned}$$

Es fundamental observar que

$$\boxed{N(A^t A) = N(A)}$$

y

$$\boxed{\mathcal{C}(A^t A) = \mathcal{C}(A)\,.}$$

Nótese también el siguiente resultado:

Teorema

Dada una matriz A con columnas linealmente independientes, sea $A = Q\,R$ su factorización QR. La ecuación $A\,x = b$ tiene una solución de mínimos cuadrados única dada por

$$x_0 = R^{-1} Q^t\, b\,.$$

Obviamente, el siguiente resultado también se verifica:

Teorema

Dada una matriz A de dimensión $m \times n$ cuyas columnas forman un conjunto ortonormal de vectores de $\mathbb{R}^m$, la solución al problema de mínimos cuadrados es

$$x_0 = A^t\, b\,.$$

13.1.2. Proyección ortogonal y factorización QR

Sea A una matriz de columnas linealmente independientes. Si conocemos su factorización QR, podemos obtener una manera alternativa de calcular la proyección de un vector b en el espacio columna de A como sigue:

$$\begin{aligned} A\,(A^t\,A)^{-1}\,A^t &= Q\,R\,((Q\,R)^t\,Q\,R)^{-1}\,(Q\,R)^t = Q\,R(R^t\,Q^t\,Q\,R)^{-1}\,R^t\,Q^t \\ &= Q\,R\,(R^t\,I_n\,R)^{-1}\,R^t\,Q^t = Q\,R\,(R^t\,R)^{-1}\,R^t\,Q^t = Q\,R\,R^{-1}\,(R^t)^{-1}\,R^t\,Q^t \\ &= Q\,I_n\,I_n\,Q^t = Q\,Q^t\,. \end{aligned}$$

Por tanto:

$$\boxed{P_S(b) = A\,(A^t\,A)^{-1}\,A^t\,b = Q\,Q^t\,b\,.}$$

Ejemplo

Consideremos de nuevo el subespacio de $\mathbb{R}^3$

$$S = \left\{(x, y, z)^t \in \mathbb{R}^3 \colon x + y + z = 0\right\}\,.$$

Vamos a obtener la matriz asociada a la proyección sobre S usando la factorización QR de la matriz

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

Usando Gram-Schmidt, tenemos

$$w_1 = v_1 = (1, 0, -1)^t\,, \quad w_2 = v_2 - \frac{\langle w_1, v_2\rangle}{\langle w_1, w_1\rangle}\,w_1 = \frac{1}{2}\,(-1, 2, -1)^t\,.$$

Normalizando, resulta

$$w_1' = \frac{1}{\sqrt{2}}\,(1, 0, -1)^t\,, \quad w_2' = \frac{1}{\sqrt{6}}\,(-1, 2, -1)^t\,.$$

Por tanto

$$Q = \frac{1}{\sqrt{6}} \begin{pmatrix} \sqrt{3} & -1 \\ 0 & 2 \\ -\sqrt{3} & -1 \end{pmatrix}$$

y

$$\begin{aligned} A\,(A^t\,A)^{-1}\,A^t = Q\,Q^t &= \frac{1}{\sqrt{6}} \begin{pmatrix} \sqrt{3} & -1 \\ 0 & 2 \\ -\sqrt{3} & -1 \end{pmatrix} \frac{1}{\sqrt{6}} \begin{pmatrix} \sqrt{3} & 0 & -\sqrt{3} \\ -1 & 2 & -1 \end{pmatrix} \\ &= \frac{1}{3} \begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix}, \end{aligned}$$

como ya sabíamos.

13.2. Aproximación de funciones mediante polinomios

En muchas aplicaciones es necesario aproximar una función continua en términos de funciones de algún tipo especial. El caso más común es el de aproximar por medio de un polinomio de grado $\leqslant n$ usando bases ortonormales.

Ejemplo

Consideremos la función $f(x) = e^x$ en el intervalo $[0, 1]$. Vamos a aproximarla por medio de un polinomio perteneciente al espacio $\mathbb{P}_1$. Recordemos que el concepto de ortogonalidad se basa en el producto interno. Es fácil demostrar que la siguiente DEFINICIÓN proporciona un producto interno en el conjunto $\mathcal{C}[0, 1]$ de las funciones continuas definidas en $[0,1]$:

$$\langle f, g\rangle = \int_0^1 f(x)\, g(x)\, dx\,.$$

En primer lugar, vamos a determinar una base ORTONORMAL para el espacio vectorial $\mathbb{P}_1$. La norma (al cuadrado) de $p_0(x) = 1$ es

$$\|p_0\|^2 = \int_0^1 dx = 1\,,$$

es decir, p_0 es unitario. Ahora consideremos un polinomio $p_1(x) = ax + b$ tal que $a \neq 0$, sea ortogonal a p_0 y también tenga norma 1. Por una parte tendremos

$$\langle p_0, p_1\rangle = \int_0^1 (ax+b)\, dx = \frac{a}{2} + b = 0 \quad \Rightarrow \quad a = -2b\,.$$

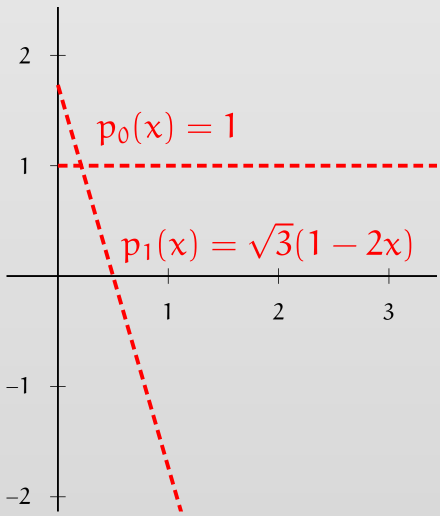
Por otra, el cuadrado de la norma de p_1 debería ser 1, luego

$$\|p_1\|^2 = \int_0^1 (b\,(-2x+1))^2\, dx = \frac{b^2}{3} = 1 \quad \Rightarrow \quad b = \sqrt{3}\,, \quad a = -2\sqrt{3}\,.$$

Así, tenemos la base ortonormal:

$$B = \left(1, \sqrt{3}(1-2x)\right)\,.$$

Obviamente no podemos escribir $f(x) = e^x = c_0\, p_0(x) + c_1\, p_1(x)$ para todo $x \in [0,1]$ (ya que la función exponencial no es un polinomio); pero sí podemos APROXIMAR $f(x) = e^x$ por un polinomio en $\mathbb{P}_1$ (i.e., encontrar la solución de mínimos cuadrados).



Tal aproximación vendrá dada por la proyección ortogonal del vector $f(x) = e^x$ sobre $\mathbb{P}_1$:

$$P_{\mathbb{P}_1}(f) = c_0 p_0 + c_1 p_1 = \|P_{p_0}(f)\| \ p_0 + \|P_{p_1}(f)\| \ p_1 \,.$$

Puesto que p_0 y p_1 son vectores unitarios:

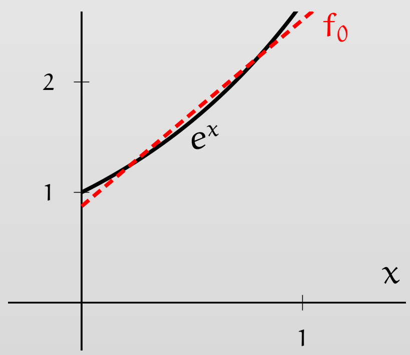
$$\|P_{p_0}(f)\| = \langle p_0, f\rangle = \int_0^1 e^x \, dx = e^x \Big|_0^1 = e - 1$$

y

$$\|P_{p_1}(f)\| = \langle p_1, f\rangle = \int_0^1 \sqrt{3}\,(1-2x)\, e^x \, dx \underset{\substack{\uparrow \\ \text{por partes}}}{=} \sqrt{3}\,(e-3)\,.$$

Así, la mejor aproximación f_0 en el espacio vectorial $\mathbb{P}_1$ de $f(x) = e^x$ usando el producto interno dado anteriormente es

$$f_0(x) = (e-1) + \sqrt{3}\,(e-3)\,\sqrt{3}\,(1-2x) = 4e - 10 + (18-6e)\,x\,.$$



Universidad Carlos III de Madrid

Mínimos cuadrados y análisis de regresión

Para entender el problema de la regresión por medio del ajuste por mínimos cuadrados, veamos un ejemplo concreto y muy sencillo, para luego extender el resultado a una situación más general.

Dados los puntos con coordenadas (1, 2) ; (2, 4) y (3, 4), que, evidentemente, no están alineados, pero se pretenden "trazar" una recta que se ajuste lo mejor posible a ellos, cuya ecuación será:

$$y = a_0 + a_1 x$$

y de la que se pretende hallar el valor de los parámetros a_0 y a_1.

El planteamiento del problema mediante un sistema de ecuaciones, sería el siguiente:

$$\begin{cases} a_0 + 1a_1 = 2 \\ a_0 + 2a_1 = 4 \\ a_0 + 3a_1 = 4 \end{cases}$$

que escrito en forma matricial y vectorial será:

$$\begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \\ 4 \end{bmatrix} \Leftrightarrow a_0 c_1 + a_1 c_2 = b \Leftrightarrow Ax = b$$

Evidentemente este sistema será incompatible (en caso general también), a no ser que los tres puntos estuvieran alineados (pero en ese caso no tendría sentido la regresión).

Desde un punto de vista puramente algebraico, podemos entender el problema, diciendo que el vector columna de los términos independientes, b, de $\mathbb{R}^3$, no pertenece al subespacio vectorial generado por los dos vectores columna de la matriz A, que suele designarse por *ColA*:

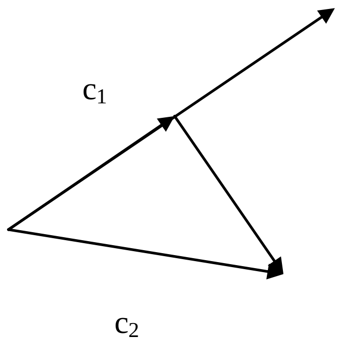
$$b \notin Gen\left(c_1, c_2\right) = ColA$$

La búsqueda de la "mejor" solución (aproximada) del problema, se basa en la consideración de que "mejor" significará proyectar el vector b sobre el espacio *ColA*, que designaremos por $\tilde{b}$, es decir:

$$\tilde{b} = proy_{ColA}(b)$$

Y eso dará lugar a nuevo sistema compatible con una solución aproximada x, pero la mejor aproximación, en el sentido "*mínimos cuadrados*", esta última frase se denomina así, por minimizar la norma habitual (euclídea), que como se conoce desde siempre es la raíz de una suma de cuadrados (pero no tendría por qué ser así necesariamente).

Así que el problema se reduce a hallar proyecciones, pero no se trata de un problema directo ya que la proyección sobre un subespacio requiere la obtención previa de una base ortogonal. En nuestro caso los dos vectores columna no son ortogonales($\{c_1, c_2\}$ no es ortogonal, es decir, así se necesita una ortogonalización previa, que puede ser tal y como muestra la figura siguiente, dando lugar a los vectores:



$\{e_1, e_2\}$ que tendrán la siguiente relación con $\{c_1, c_2\}$

$$\begin{aligned} e_1 &= c_1 \\ e_2 &= c_2 - proy_{c_1}\{c_2\} \end{aligned}$$

Si denominamos por B a la matriz formada por las columnas de componentes los nuevos vectores, tendremos una base ortogonal, para luego proyectar el vector b sobre ella y sumando resultados obtener $\tilde{b}$, es decir:

$$\tilde{b} = proy_{ColB}(b) = proy_{e_1}(b) + proy_{e_2}(b)$$

Vayamos a los cálculos en nuestro caso concreto:

$$proy_{c_1}\{c_2\} = \left(\frac{c_1.c_2}{c_1.c_1}\right)c_1 = \frac{1.1+1.2+1.3}{1.1+1.1+1.1}\begin{bmatrix}1\\1\\1\end{bmatrix} = \begin{bmatrix}2\\2\\2\end{bmatrix}$$

$$e_2 = c_2 - proy_{c_1}\{c_2\} = c_2 - e_1 = \begin{bmatrix}1\\2\\3\end{bmatrix} - \begin{bmatrix}2\\2\\2\end{bmatrix} = \begin{bmatrix}-1\\0\\1\end{bmatrix}$$

$$proy_{e_1}\{b\} = \left(\frac{e_1.b}{e_1.e_1}\right)e_1 = \frac{1.2+1.4+1.4}{1.1+1.1+1.1}\begin{bmatrix}1\\1\\1\end{bmatrix} = \begin{bmatrix}10/3\\10/3\\10/3\end{bmatrix}$$

$$proy_{e_2}\{b\} = \left(\frac{e_2.b}{e_2.e_2}\right)e_2 = \frac{-1.2+0.4+1.4}{2}\begin{bmatrix}-1\\0\\1\end{bmatrix} = \begin{bmatrix}-1\\0\\-1\end{bmatrix} \text{ luego}$$

$$\tilde{b} = proy_{ColB}(b) = proy_{e_1}(b) + proy_{e_2}(b) = \begin{bmatrix}7/3\\10/3\\13/3\end{bmatrix}$$

Así que el sistema (ahora compatible) que debemos resolver será:

$$\begin{cases} a_0 + 1a_1 = 7/3 \\ a_0 + 2a_1 = 10/3 \\ a_0 + 3a_1 = 13/3 \end{cases}$$

Cuya solución es $a_0 = 4/3$, $a_1 = 1$, lo que da lugar a la recta de regresión siguiente:

$$y = \frac{4}{3} + x$$

Conviene aclarar inmediatamente, que este no es el procedimiento estándar, por ser demasiado laborioso. Después de una carga teórica importante, se puede deducir que ese resultado se obtiene de resolver el sistema siguiente (que será compatible):

$$\left(A^T A\right) x = A^T b$$

Y, por supuesto, ésto también es válido para cualquier otra regresión posible, por ejemplo, el ajuste de una serie de datos (n) a una parábola:

$$y = a_0 + a_1 x + a_2 x^2$$

con planteamiento similar, pero:

$$\begin{pmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_n^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{pmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

Parte XV

Tema 14. La descomposición en valores singulares (DVS). Pseudoinversa

Tema 14

Pseudo-inversa y descomposición en valores singulares

14.1. Pseudo-inversa

El concepto de inversa de una matriz cuadrada cuyo rango coincide con el número de filas (o columnas) es bien conocido. A^{-1} es la inversa de A si y sólo si $A\,A^{-1} = A^{-1}\,A = I$. Podríamos decir que A^{-1} es la inversa "bilateral" de A. También sabemos que cuando una matriz es invertible, el sistema de ecuaciones $A\,x = b$ tiene solución única dada por $x = A^{-1}\,b$.

En este tema cerramos el curso generalizando el concepto de inversa a otras matrices, lo cual nos permitirá, por ejemplo, proporcionar una alternativa a resolver el sistema $A\,x = b$, cuando A no tiene inversa, tal y como hicimos cuando abordamos el problema de mínimos cuadrados. Para ello, se introduce la matriz conocida como **pseudo-inversa de Moore-Penrose** (o simplemente como *pseudo-inversa*) de la siguiente manera:

Si $A \in \mathbb{R}^{m\times n}$, entonces existe una *<u>única</u>* matriz $A^+ \in \mathbb{R}^{n\times m}$ que satisface las siguientes cuatro condiciones:

1. $A\,A^+\,A = A$.

2. $A^+\,A\,A^+ = A^+$.

3. $A^+\,A = (A^+A)^t$.

4. $A\,A^+ = (A\,A^+)^t$.

Estas condiciones se denominan **condiciones de Penrose**.

Obsérvese que para cualquier matriz A la transformación lineal $T_{A\,A^+}$ asociada a $A\,A^+$ es siempre una proyección, pues

$$(T_{A\,A^+} \circ T_{A\,A^+})(b) = T_{A\,A^+}(A\,A^+\,b) = A\,A^+\,A\,A^+\,b \underset{\substack{\uparrow \\ A\,A^+\,A=A}}{=} A\,A^+\,b = T_{A\,A^+}(b)\,.$$

Si $m = n$ y A es no singular, es evidente que A^{-1} satisface de manera trivial las cuatro condiciones de Penrose, es decir, la pseudo-inversa de A coincide con la inversa de A.

En caso de que $m > n$ y que las $\boxed{\text{columnas de } A \text{ sean linealmente independientes}}$ (lo que llamaremos de rango columna completo), su pseudo-inversa se obtiene con las siguiente expresión:

$$\boxed{A^+ = (A^t\,A)^{-1}\,A^t.}$$

Dicha expresión ha de resultarnos familiar del Tema 13, en el que abordamos el problema de mínimos cuadrados. De hecho, al final de este capítulo estudiaremos la relación entre la pseudo-inversa y la resolución de un sistema por mínimos cuadrados.

Ejemplo

Consideremos la matriz

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix}.$$

Es evidente que las columnas de A son linealmente independientes. Por tanto, calculamos su pseudo-inversa mediante

$$\begin{aligned}
A^{+} = (A^{t}A)^{-1}A^{t} \quad &= \left[\begin{pmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix}\right]^{-1}\begin{pmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{pmatrix} \\
&= \begin{pmatrix} 35 & 44 \\ 44 & 56 \end{pmatrix}^{-1}\begin{pmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{pmatrix} \\
&= \frac{1}{24}\begin{pmatrix} 56 & -44 \\ -44 & 35 \end{pmatrix}\begin{pmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{pmatrix} \\
&= \frac{1}{12}\begin{pmatrix} -16 & -4 & 8 \\ 13 & 4 & -5 \end{pmatrix}.
\end{aligned}$$

Podemos verificar que se cumplen las condiciones de Penrose:

1. $A\,A^{+}\,A = A$:

$$\begin{aligned}
A\,A^{+}\,A \quad &= \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix}\frac{1}{12}\begin{pmatrix} 16 & -4 & 8 \\ 13 & 4 & -5 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} \\
&= \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = A\,.
\end{aligned}$$

2. $A^+ A A^+ = A^+$:

$$A^+AA^+ = \frac{1}{12}\begin{pmatrix} -16 & -4 & 8 \\ 13 & 4 & -5 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix}\frac{1}{12}\begin{pmatrix} -16 & -4 & 8 \\ 13 & 4 & -5 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\frac{1}{12}\begin{pmatrix} -16 & -4 & 8 \\ 13 & 4 & -5 \end{pmatrix} = A^+ .$$

3. $A^+ A = (A^+ A)^t$:

$$A^+ A = \frac{1}{12}\begin{pmatrix} -16 & -4 & 8 \\ 13 & 4 & -5 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = (A^+ A)^t .$$

4. $A A^+ = (A A^+)^t$:

$$A A^+ = \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix}\frac{1}{12}\begin{pmatrix} -16 & -4 & 8 \\ 13 & 4 & -5 \end{pmatrix} = \frac{1}{6}\begin{pmatrix} 5 & 2 & -1 \\ 2 & 2 & 2 \\ -1 & 2 & 5 \end{pmatrix}$$

$$= (A A^+)^t .$$

De manera análoga, si $m < n$ y las $\boxed{\text{filas de } A \text{ son linealmente independientes}}$ (lo que llamaremos de rango fila completo), su pseudo-inversa se obtiene con las siguiente expresión:

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

No obstante, en el caso de que A no sea de rango completo no se pueden usar estas fórmulas.

Ejemplo

Sea la matriz:

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 6 \end{pmatrix}.$$

Obviamente $\det(A) = 0$ y por tanto no tiene inversa: las columnas son linealmente dependientes (o no tiene rango completo). Si intentamos usar la expresión $A^+ = (A^t A)^{-1} A^t$ obtenemos que la matriz

$$A^t A = \begin{pmatrix} 1 & 3 \\ 2 & 6 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 6 \end{pmatrix} = \begin{pmatrix} 10 & 20 \\ 20 & 40 \end{pmatrix}$$

también es singular y no tiene inversa.

Para abordar este caso necesitamos introducir el concepto de valores singulares de una matriz.

14.2. Valores singulares

Consideremos una matriz cuadrada $A \in \mathbb{R}^{n \times n}$. Sabemos que, desde un punto de vista geométrico, los vectores propios de A indican las direcciones de "estiramiento puro" de la transformación lineal asociada, mientras que los valores propios corresponden a la magnitud de tal estiramiento. Como vimos anteriormente, una clase particularmente importante de matrices es la de las matrices simétricas reales, con cuyos vectores propios se puede formar una base ortogonal de $\mathbb{R}^n$.

En el caso opuesto, si A es una matriz rectangular, ésta no tiene valores propios. Sin embargo, sí podemos calcular los valores propios de la matriz cuadrada (simétrica) asociada $K = A^t A$; sus correspondientes raíces cuadradas, de gran utilidad en diversos campos como la compresión de imágenes o la recuperación de información, se denominan *valores singulares* de A. Veamos algunas propiedades de la matriz K.

Como ya sabemos, K es simétrica (ver Tema 1) y, en consecuencia, la matriz es diagonalizable y tendrá valores propios reales. Además:

Los valores propios de $A^t A$ son no negativos.

Demostración. En efecto, si $\lambda \in \sigma(A^t A)$ tiene vector propio asociado v_1 (que podemos considerar unitario respecto al producto escalar usual, es decir, $\|v_1\| = 1$), se obtiene que

$$\lambda = \lambda\|v_1\|^2 = \lambda \langle v_1, v_1\rangle = \langle v_1, \lambda v_1\rangle = \langle v_1, A^t A v_1\rangle = \langle A v_1, A v_1\rangle = \|A v_1\|^2 \geqslant 0\,.$$

□

Así, podemos definir los **valores singulares** de una matriz arbitraria $A \in \mathbb{R}^{m\times n}$ como las n raíces cuadradas (positivas) de los n valores propios no negativos de $A^t A$. Estos valores singulares se denotan de manera habitual por $\sigma_i \left(= \sqrt{\lambda_i}\right)$ $(1 \leqslant i \leqslant n)$ y se ordenan de manera decreciente, teniendo en cuenta sus multiplicidades algebraicas:

$$\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_n\,.$$

Ejemplo

Consideremos la matriz

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

Para determinar sus valores singulares, calculamos

$$A^t A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$$

y sus correspondientes valores propios:

$$\sigma(A^t A) = \{1, 3\}.$$

De este modo, los valores singulares de A son $\sigma_1 = \sqrt{3}$ y $\sigma_2 = 1$.

14.2.1. Descomposición en valores singulares

El cálculo de los valores singulares permite además encontrar una factorización especialmente útil para *cualquier* matriz $A \in \mathbb{R}^{m\times n}$. Esta descomposición, a la que nos referiremos como SVD, es de la forma:

$$\boxed{A = U\,S\,V^t}$$

siendo U ortogonal, S diagonal y V ortogonal.

Descomposición en valores singulares

Sea la matriz $A \in \mathbb{R}^{m\times n}$. Entonces existen matrices ortogonales $U \in \mathbb{R}^{m\times m}$ y $V \in \mathbb{R}^{n\times n}$, de manera que la matriz puede descomponerse como $A = U\,S\,V^t$, siendo S una matriz $m \times n$ de la forma:

$$S = \left(\begin{array}{cccc|c} \sigma_1 & 0 & \cdots & 0 & 0 \\ 0 & \sigma_2 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & \sigma_r & 0 \\ \hline 0 & 0 & \cdots & 0 & \underbrace{0}_{n-r \text{ columnas}} \end{array}\right) \Big\}\, {\scriptstyle m-r \text{ filas}}$$

con $\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_r > 0$ y siendo $r \leqslant \text{mín}(m, n)$ el número de valores singulares no nulos.

Las $m-r$ filas y $n-r$ columnas de ceros de la matriz S existirán o no en función de los valores de m y n y del número de valores singulares no nulos.
Los valores $\sigma_1, \sigma_2, \ldots, \sigma_r$ de la diagonal principal de S son los valores singulares de A.
Las columnas de la matriz U se denominan **vectores singulares por la izquierda** y las columnas de V se denominan **vectores singulares por la derecha** de A.

Obviamente A es equivalente a la matriz S y tienen, por tanto, el mismo rango. Por esta razón

$$r = rg(A)\,.$$

Las matrices U y V no son únicas, pero una forma habitual de obtener tales matrices es la siguiente:

- Matriz V:

  $$V = (v_1, \cdots, v_n)\ ,$$

  con $v_1, \ldots, v_n$ los vectores propios de $K = A^t A$ normalizados y en el orden dado por los valores singulares de A.

- Matriz U:

  $$U = (u_1, \cdots, u_m)\ ,$$

  con las columnas $u_1, \cdots, u_r$ calculadas mediante

  $$\boxed{u_i = \frac{1}{\sigma_i} A\, v_i}$$

  y con $u_{r+1}, \cdots, u_m$, si existen, elegidos de manera que U sea ortogonal (usando Gram-Schmidt, si es necesario).

Ejemplo

Vamos a calcular una descomposición en valores singulares de la matriz 2×3:

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

Calculamos la matriz $A^t A$:

$$A^t A = \begin{pmatrix} 13 & 12 & 2 \\ 12 & 13 & -2 \\ 2 & -2 & 8 \end{pmatrix}$$

y sus valores propios:

$$\sigma(A^t A) = \{0, 9, 25\},$$

por lo que los valores singulares de A son $\sigma_1 = 5$, $\sigma_2 = 3$ y $\sigma_3 = 0$. Los vectores propios de $A^t A$ son:

$$v_1' = (1,1,0)^t \ , \quad v_2' = (1,-1,4)^t \ , \quad v_3' = (2,-2,-1)^t \ ,$$

que normalizados resultan:

$$v_1 = \frac{1}{\sqrt{2}}(1,1,0)^t \, , \quad v_2 = \frac{1}{3\sqrt{2}}(1,-1,4)^t \, , \quad v_3 = \frac{1}{3}(2,-2,-1)^t \, .$$

Por tanto, tenemos que

$$S = \begin{pmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \end{pmatrix}$$

$$V = \frac{1}{3\sqrt{2}} \begin{pmatrix} 3 & 1 & 2\sqrt{2} \\ 3 & -1 & -2\sqrt{2} \\ 0 & 4 & -\sqrt{2} \end{pmatrix} .$$

Vamos a calcular la matriz U, de dimensión 2×2:

$$\begin{aligned} u_1 &= \frac{1}{5} A\, v_1 = \frac{1}{\sqrt{2}}\,(1,1)^t\,, \\ u_2 &= \frac{1}{3} A\, v_2 = \frac{1}{\sqrt{2}}\,(1,-1)^t\,. \end{aligned}$$

En resumen:

$$U = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

Se puede comprobar que efectivamente $A = U\, S\, V^t$.

Las matrices obtenidas en la SVD verifican la siguiente propiedad:

Teorema

Sea $A = U\, S\, V^t$ una descomposición en valores singulares de $A \in \mathbb{R}^{m \times n}$ con $rg(A) = r$. Entonces:

i) $(u_1, \ldots, u_r)$ es una base (ortonormal) de $\mathcal{C}(A)$.

ii) $(u_{r+1}, \ldots, u_m)$ es una base (ortonormal) de $N(A^t)$.

iii) $(v_1, \ldots, v_r)$ es una base (ortonormal) de $\mathcal{C}(A^t)$.

iv) $(v_{r+1}, \ldots, v_n)$ es una base (ortonormal) de $N(A)$.

A efectos prácticos, en ocasiones es interesante tener en cuenta el siguiente resultado:

Sea la matriz $A \in \mathbb{R}^{m \times n}$; entonces:

i) El número de valores singulares no nulos de A es $r = rg(A) \leqslant \min(m, n)$.

ii) Las matrices $A^t\, A$ y $A\, A^t$ tienen los MISMOS valores propios NO NULOS.

En consecuencia, si la matriz A, de dimensión $m \times n$, con $m < n$, tiene descomposición

SVD dada por $A = U\,S\,V^t$, resulta que $A^t = V\,S^t\,U^t$ es la descomposición SVD de A^t. Como $A^t\,A$ es de dimensión $n \times n$ y $A\,A^t$ es de dimensión $m \times m$, los cálculos son más sencillos si trabajamos con esta segunda matriz. Todo lo anterior puede esquematizarse como se indica en la figura 14.1.

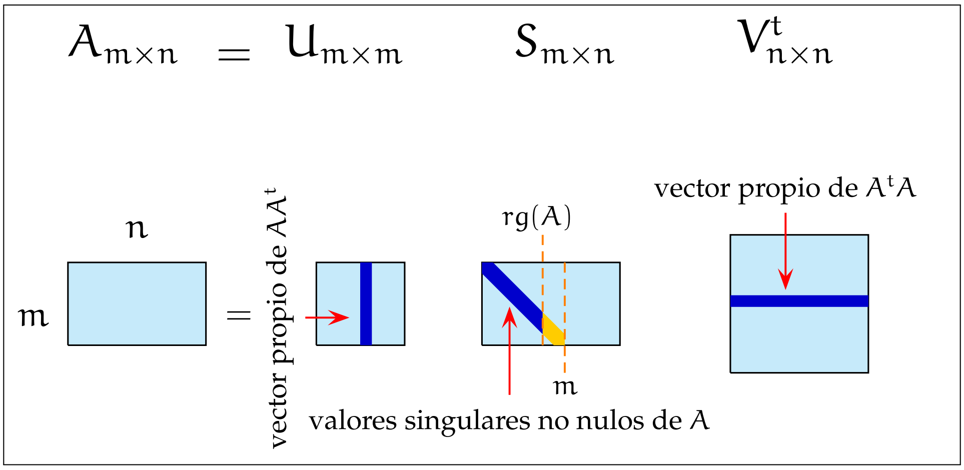


Figura 14.1: Descomposición SVD de una matriz $m \times n$.

Ejemplo

Consideremos de nuevo la matriz

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

Vamos a obtener la descomposición SVD de A^t. Calculamos

$$A\,A^t = \begin{pmatrix} 17 & 8 \\ 8 & 17 \end{pmatrix},$$

cuya ecuación característica es de grado 2; sus valores propios son $\sigma(A\,A^t) = \{9, 25\}$. Por tanto, los valores singulares de A^t son $\sigma_1 = 5, \sigma_2 = 3$ y los vectores propios de $A\,A^t$, normalizados, son

$$v_1 = \frac{1}{\sqrt{2}}(1,1)^t, \qquad v_2 = \frac{1}{\sqrt{2}}(1,-1)^t.$$

Ahora calculamos los vectores

$$
\begin{aligned}
u_1 &= \frac{1}{\sigma_1} A^t v_1 = \frac{1}{\sqrt{2}} (1,1,0)^t \,, \\
u_2 &= \frac{1}{\sigma_2} A^t v_2 = \frac{1}{3\sqrt{2}} (1,-1,4)^t \,.
\end{aligned}
$$

Un vector ortonormal a u_1 y u_2 sería

$$u_3 = \frac{1}{3} (2,-2,-1)^t \,,$$

que podemos obtener buscando un vector linealmente independiente de u_1 y u_2 y utilizando Gram-Schmidt.

Obsérvese que la descomposición SVD de la matriz A^t obtenida con estos cálculos coincide con la traspuesta de la SVD de A que obtuvimos anteriormente.

14.3. Pseudo-inversa de matrices arbitrarias

Finalmente, damos el siguiente resultado que permite obtener la pseudo-inversa de una matriz arbitraria, incluidas las que no son de rango completo.

Teorema

Sea la descomposición SVD de la matriz $A \in \mathbb{R}^{m\times n}$ de la forma $A = U\,S\,V^t$ con $U \in \mathbb{R}^{m\times m}$ y $V \in \mathbb{R}^{n\times n}$ ortogonales y $\boxed{S \in \mathbb{R}^{m\times n}}$ dada por:

$$
S = \begin{pmatrix}
\sigma_1 & 0 & \cdots & 0 & \vline & 0 \\
0 & \sigma_2 & \cdots & 0 & \vline & 0 \\
\vdots & \vdots & & \vdots & \vline & \vdots \\
0 & 0 & \cdots & \sigma_r & \vline & 0 \\
\hline
0 & 0 & \cdots & 0 & \vline & \underbrace{0}_{n-r \text{ columnas}}
\end{pmatrix}
\left.\vphantom{0}\right\} m-r \text{ filas}
$$

siendo $\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_r > 0$ y siendo $r \leqslant \text{mín}(m, n)$ el número de valores singulares no nulos (e igual al rango de A). La pseudo-inversa de A se calcula mediante la expresión:

$$A^+ = V\, S^+\, U^t$$

donde la matriz $\boxed{S^+ \in \mathbb{R}^{n\times m}}$ se calcula con

$$S^+ = \left(\begin{array}{cccc|c} 1/\sigma_1 & 0 & \cdots & 0 & 0 \\ 0 & 1/\sigma_2 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 1/\sigma_r & 0 \\ \hline 0 & 0 & \cdots & 0 & \underbrace{0}_{m-r \text{ columnas}} \end{array}\right) \begin{array}{l} \\ \\ \\ \\ \}\, n-r \text{ filas} \end{array}.$$

Ejemplo

Dada la matriz

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

cuya descomposición SVD es

$$A = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \end{pmatrix} \frac{1}{3\sqrt{2}} \begin{pmatrix} 3 & 1 & 2\sqrt{2} \\ 3 & -1 & -2\sqrt{2} \\ 0 & 4 & -\sqrt{2} \end{pmatrix}^t,$$

la pseudo-inversa de A se escribe como

$$A^+ = V\,S^+\,U^t = \frac{1}{3\sqrt{2}} \begin{pmatrix} 3 & 1 & 2\sqrt{2} \\ 3 & -1 & -2\sqrt{2} \\ 0 & 4 & -\sqrt{2} \end{pmatrix} \begin{pmatrix} \frac{1}{5} & 0 \\ 0 & \frac{1}{3} \\ 0 & 0 \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

$$= \frac{1}{45} \begin{pmatrix} 7 & 2 \\ 2 & 7 \\ 10 & -10 \end{pmatrix}.$$

Se puede comprobar que A^+ verifica las cuatro condiciones de Penrose.

14.4. Pseudo-inversa y mínimos cuadrados

Recordemos que el sistema de ecuaciones lineales $A\,x = b$ tiene al menos una solución cuando b pertenece al espacio columna de A, pero si $b \notin \mathcal{C}(A)$, entonces el sistema no tiene solución. Sin embargo, si fuera deseable encontrar algún x_0 que sea lo más cercano posible a una solución, calculamos la solución de mínimos cuadrados dada por

$$A^t\,A\,x_0 = A^t\,b\,.$$

En el caso de que la matriz $A^t\,A$ fuese invertible (o en otras palabras, si las columnas de A fueran linealmente independientes), el problema de mínimos cuadrados tiene solución única y ésta viene expresada por

$$x_0 = (A^t\,A)^{-1}\,A^t\,b\,,$$

que en este caso podemos escribir mediante

$$\boxed{x_0 = A^+\,b.}$$

Por otro lado, si la matriz $A^t A$ no fuese invertible, el problema de mínimos cuadrados tiene infinitas soluciones. En esta situación es posible demostrar que, si tenemos la SVD $A = U S V^t$, entonces

$$\boxed{x_0 = A^+ b = V S^+ U^t b}$$

proporciona *una* solución de mínimos cuadrados de todas las posibles, con la propiedad de tener norma euclídea (es decir, asociada al producto escalar usual) mínima.

Ejemplo

Consideremos el sistema $A x = b$, con

$$A = \begin{pmatrix} 6 & 2 & -4 \\ 3 & 1 & -2 \end{pmatrix}, \quad b = \begin{pmatrix} 35 \\ 70 \end{pmatrix}.$$

La matriz

$$A^t A = \begin{pmatrix} 45 & 15 & -30 \\ 15 & 5 & -10 \\ -30 & -10 & 20 \end{pmatrix}$$

no es invertible, por lo que existe un número infinito de soluciones s de mínimos cuadrados, de la forma

$$S = \left\{ s = (s_1, s_2, s_3)^t : s_1 = \frac{28}{3} + \frac{2}{3} s_3 - \frac{1}{3} s_2 \right\}.$$

Todas ellas verifican que la $\boxed{\text{diferencia}}$ $r = A s - b = (21, -42)^t$ tiene la mínima norma posible,

$$\|r\| = \|A s - b\| = \sqrt{(21)^2 + (-42)^2} = \sqrt{2205}.$$

La $\boxed{\text{solución}}$ proporcionada por la pseudo-inversa es

$$x_0 = (6, 2, -4)^t, \quad \text{con } \|x_0\| = \sqrt{56}.$$

Si consideramos cualquier otra solución de mínimos cuadrados, por ejemplo

$$x_1 = (10, 0, 1)^t \,,$$

ésta tendrá una norma mayor, en este caso $\|x_1\| = \sqrt{101}$.