

## Módulo 2: Álgebra lineal

David R. Montalván Hernández  
david.montalvan@itam.mx

# Contenido

<b>1. Espacios vectoriales</b>	<b>1</b>
1.1. Espacios vectoriales . . . . .	1
1.2. Combinaciones lineales y bases . . . . .	3
1.3. Dimensión de un espacio vectorial . . . . .	5
<b>2. Matrices</b>	<b>7</b>
2.1. Definiciones básicas . . . . .	7
2.2. Tipos especiales de matrices . . . . .	8
2.3. Operaciones con matrices . . . . .	10
2.4. Sistemas de ecuaciones lineales . . . . .	12
2.5. Determinantes y matrices inversas . . . . .	14
<b>3. Producto escalar y ortogonalidad</b>	<b>17</b>
3.1. Producto escalar . . . . .	17
3.2. Ortogonalidad . . . . .	19
<b>4. Valores y vectores característicos</b>	<b>21</b>
<b>5. Descomposición por valores singulares</b>	<b>24</b>
5.1. Reducción de la dimensión de los datos . . . . .	25
<b>6. Análisis de componentes principales</b>	<b>26</b>
6.1. Formas cuadráticas . . . . .	26
6.2. Componentes principales . . . . .	27

# Capítulo 1

## Espacios vectoriales

### 1.1. Espacios vectoriales

Iniciaremos estas notas repasando algunos conceptos básicos del álgebra lineal, estos conceptos son necesarios para entender el desarrollo de los temas posteriores.

**Definición 1.1** (Campo):

Sea  $\mathbb{F}$  un subconjunto de los números complejos  $\mathbb{C}$ . Decimos que  $\mathbb{F}$  es un campo, si satisface las siguientes condiciones:

1. Si  $x, y \in \mathbb{F}$ , entonces  $x + y \in \mathbb{F}$ ,  $xy \in \mathbb{F}$  (cerradura bajo la suma y el producto).
2. Los elementos 0 y 1 pertenecen a  $\mathbb{F}$  (existencia del neutro aditivo y neutro multiplicativo).
3. Si  $x \in \mathbb{F}$ , entonces  $-x$  es también un elemento de  $\mathbb{F}$  (existencia del inverso aditivo).
4. Si  $x \in \mathbb{F}$  y  $x \neq 0$ , entonces  $x^{-1} \in \mathbb{F}$  (existencia del inverso multiplicativo).

A los elementos de un campo se les llamará **números** o **escalares**.

De acuerdo a la definición 1.1, los siguientes conjuntos son campos:

- El conjunto de los números reales  $\mathbb{R}$ .
- El conjunto de los números racionales  $\mathbb{Q}$ .

- El conjunto de los números complejos  $\mathbb{C}$ .

**Ejercicio 1.1:**

El conjunto de los números enteros,  $\mathbb{Z}$ , ¿es un campo?

**Definición 1.2** (Espacio vectorial):

Decimos que un conjunto  $\mathbb{V}$  (vectores) sobre un campo  $\mathbb{F}$  (escalares), es un espacio vectorial si:

Existen dos operaciones definidas  $(\cdot, +)$  con la propiedad de cerradura:

1.  $a \cdot \mathbf{v} \in \mathbb{V}$ , para todo  $a \in \mathbb{F}$ ,  $\mathbf{v} \in \mathbb{V}$
2.  $\mathbf{v} + \mathbf{w} \in \mathbb{V}$ , para todo  $\mathbf{v}, \mathbf{w} \in \mathbb{V}$

tales que:

- a) Para  $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{V}$ , se tiene que:

$$(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$$

- b) Existe un elemento  $\mathbf{0} \in \mathbb{V}$ , tal que  $\mathbf{0} + \mathbf{v} = \mathbf{v}$  para todo  $\mathbf{v} \in \mathbb{V}$ .

- c) Para todo  $\mathbf{v} \in \mathbb{V}$  existe un elemento  $-\mathbf{v} \in \mathbb{V}$ , tal que  $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$ .

- d) Para todo  $\mathbf{v}, \mathbf{w} \in \mathbb{V}$ , se tiene que  $\mathbf{v} + \mathbf{w} = \mathbf{w} + \mathbf{v}$ .

- e) Para todo  $a \in \mathbb{F}$  y para todo  $\mathbf{v}, \mathbf{w} \in \mathbb{V}$ , tenemos que  $a \cdot (\mathbf{v} + \mathbf{w}) = a \cdot \mathbf{v} + a \cdot \mathbf{w}$ .

- f) Para todo  $a, b \in \mathbb{F}$  y  $\mathbf{v} \in \mathbb{V}$ , se tiene que  $a \cdot (b \cdot \mathbf{v}) = (ab) \cdot \mathbf{v}$ .

- g) Para todo  $a, b \in \mathbb{F}$  y  $\mathbf{v} \in \mathbb{V}$ , tenemos  $(a + b) \cdot \mathbf{v} = a \cdot \mathbf{v} + b \cdot \mathbf{v}$ .

- h) Para todo  $\mathbf{v} \in \mathbb{V}$  y  $1 \in \mathbb{F}$ , tenemos que  $1 \cdot \mathbf{v} = \mathbf{v}$ .

**Nota 1.1:**

En la definición 1.2, para evitar cargar la notación, escribiremos  $a \cdot \mathbf{v}$  como  $a\mathbf{v}$ , mientras que para  $\mathbf{v} + (-\mathbf{w})$  utilizaremos  $\mathbf{v} - \mathbf{w}$ .

**Ejercicio 1.2:**

Sea  $\mathbb{V} = \mathbb{F}^n$ , el conjunto de  $n$ -tuplas con elementos de un campo  $\mathbb{F}$ . Sean  $\mathbf{a} = (a_1, \dots, a_n)$  y  $\mathbf{b} = (b_1, \dots, b_n)$  vectores de  $\mathbb{V}$ , con  $a_i, b_i \in \mathbb{F}$  para toda  $i$ . Definimos  $\mathbf{a} + \mathbf{b} = (a_1 + b_1, \dots, a_n + b_n)$  y para  $c \in \mathbb{F}$ ,  $c\mathbf{a} = (ca_1, \dots, ca_n)$ . Además el elemento  $\mathbf{0} \in \mathbb{V}$  se define como  $(0, \dots, 0)$ .

Demuestre que  $\mathbb{V}$  es un espacio vectorial sobre el campo  $\mathbb{F}$ , en particular  $\mathbb{R}^n$  es un espacio vectorial sobre  $\mathbb{R}$ .

**Ejercicio 1.3:**

Demuestre que si  $\mathbb{V}$  es un espacio vectorial sobre un campo  $\mathbb{F}$ , entonces para todo  $\mathbf{v} \in \mathbb{V}$  tenemos que:

$$0\mathbf{v} = \mathbf{0}$$

En donde 0 es el elemento cero de  $\mathbb{F}$  y  $\mathbf{0}$  es el elemento cero de  $\mathbb{V}$ .

Sugerencia: Sume  $\mathbf{v}$  en el lado izquierdo de la ecuación.

**Ejercicio 1.4:**

¿Es el conjunto  $\mathbb{R}^n$ , un espacio vectorial sobre el campo de los números complejos  $\mathbb{C}$ ?

**Definición 1.3** (Subespacio vectorial):

Sea  $\mathbb{V}$  un espacio vectorial y  $\mathbb{W} \subset \mathbb{V}$ . Decimos que  $\mathbb{W}$  es un subespacio vectorial si:

- a) Si  $\mathbf{v}, \mathbf{w} \in \mathbb{W}$ , entonces  $\mathbf{v} + \mathbf{w} \in \mathbb{W}$ .
- b) Si  $\mathbf{v} \in \mathbb{W}$  y  $c$  es un escalar, entonces  $c\mathbf{v} \in \mathbb{W}$
- c)  $\mathbf{0} \in \mathbb{W}$

**Ejercicio 1.5:**

Si  $\mathbb{V} = \mathbb{R}^n$  y  $\mathbb{F} = \mathbb{R}$ , demuestre que el conjunto de vectores en  $\mathbb{V}$  cuya primera coordenada es igual a cero forma un subespacio vectorial de  $\mathbb{V}$ .

## 1.2. Combinaciones lineales y bases

**Definición 1.4** (Combinación lineal de vectores):

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

$$a_1\mathbf{v}_1 + \dots + a_n\mathbf{v}_n$$

con  $a_i \in \mathbb{F}$ ,  $\mathbf{v}_i \in \mathbb{V}$  para toda  $i$ , es llamada una combinación lineal de  $\mathbf{v}_1, \dots, \mathbf{v}_n$ .

Si todo elemento  $\mathbf{v} \in \mathbb{V}$  se puede expresar como una combinación lineal de vectores  $\mathbf{v}_1, \dots, \mathbf{v}_n$  de  $\mathbb{V}$ , es decir, si existen escalares  $a_1, \dots, a_n$  tales que

$$\mathbf{v} = a_1\mathbf{v}_1 + \dots + a_n\mathbf{v}_n$$

entonces decimos que los vectores,  $\mathbf{v}_1, \dots, \mathbf{v}_n$ , **generan el espacio**  $\mathbb{V}$

Para entender el concepto de base, es necesario primero definir la independencia entre un conjunto de vectores.

**Definición 1.5** (Dependencia lineal de vectores):

Sea  $\mathbb{V}$  un espacio vectorial sobre un campo  $\mathbb{F}$  y sea  $\mathbf{v}_1, \dots, \mathbf{v}_n$  un conjunto de vectores de  $\mathbb{V}$ . Decimos que  $\mathbf{v}_1, \dots, \mathbf{v}_n$  son linealmente dependientes sobre el campo  $\mathbb{F}$  si existen escalares  $a_1, \dots, a_n$  con al menos un  $a_i$  distinto de cero, tales que

$$a_1\mathbf{v}_1 + \dots + a_n\mathbf{v}_n = \mathbf{0}$$

en donde  $\mathbf{0}$  es el vector cero de  $\mathbb{V}$

De acuerdo a la definición 1.5, en un conjunto linealmente dependiente de vectores,  $\mathbf{v}_1, \dots, \mathbf{v}_n$ , existe un vector  $\mathbf{v}_i$ , tal que

$$\mathbf{v}_i = \sum_{j \neq i} -\frac{a_j}{a_i} \mathbf{v}_j$$

De la misma forma, decimos que un conjunto de vectores,  $\mathbf{v}_1, \dots, \mathbf{v}_n$ , son **linealmente independientes**, si la igualdad

$$a_1\mathbf{v}_1 + \dots + a_n\mathbf{v}_n = \mathbf{0}$$

implica que  $a_1, \dots, a_n = 0$ ; en otras palabras, no es posible expresar algún vector  $\mathbf{v}_i$  como combinación lineal de los demás.

**Ejercicio 1.6:**

Demuestre que si  $\mathbb{V} = \mathbb{R}^n$  y  $F = \mathbb{R}$ , entonces el conjunto de vectores

$$\begin{aligned} \mathbf{e}_1 &= (1, 0, \dots, 0) \\ \mathbf{e}_2 &= (0, 1, \dots, 0) \\ &\vdots \\ \mathbf{e}_n &= (0, 0, \dots, 1) \end{aligned}$$

es un conjunto linealmente independiente.

**Ejercicio 1.7:**

Demuestre que si  $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$  es un conjunto de vectores linealmente independiente, entonces el conjunto  $\{\mathbf{v}_1, \dots, \mathbf{v}_n, \mathbf{0}\}$  es linealmente dependiente. Sugerencia: Primero demuestre para todo escalar  $c$ ,  $c\mathbf{0} = \mathbf{0}$ .

**Definición 1.6** (Base de un espacio vectorial):

Decimos que un conjunto de vectores  $\mathbf{v}_1, \dots, \mathbf{v}_n$  de un espacio vectorial  $\mathbb{V}$  sobre un campo  $\mathbb{F}$ , forma una base de  $\mathbb{V}$  si:

- $v_1, \dots, v_n$  son linealmente independientes.
- $v_1, \dots, v_n$  generan  $\mathbb{V}$ .

**Ejercicio 1.8:**

Demuestre que el conjunto de vectores del ejercicio 1.6 forma una base de  $\mathbb{R}^n$ .

### 1.3. Dimensión de un espacio vectorial

**Definición 1.7** (Dimensión de un espacio vectorial):

Sea  $\mathcal{B} = \{v_1, \dots, v_n\}$  una base para el espacio vectorial  $\mathbb{V}$ , tal que  $|\mathcal{B}| = n < \infty$ . Entonces decimos que la dimensión de  $\mathbb{V}$ , denotada como  $\dim(\mathbb{V})$ , es igual a  $n$ . En otras palabras, la dimensión de un espacio vectorial, es el número de vectores linealmente independientes, necesarios para generar dicho espacio.

Si un espacio  $\mathbb{V}$  tiene una base,  $\mathcal{B}$ , cuya cardinalidad no es finita, decimos que la dimensión de  $\mathbb{V}$  es infinita. Si  $\mathbb{V} = \{\mathbf{0}\}$ , entonces decimos que la dimensión de  $\mathbb{V}$  es 0 (no existe una base que genere este espacio).

Los siguientes resultados nos serán de utilidad.

**Teorema 1.1:**

Sea  $\mathbb{V}$  un espacio vectorial sobre un campo  $\mathbb{F}$ . Sea  $v_1, \dots, v_m$  una base de  $\mathbb{V}$  y sea  $w_1, \dots, w_n$  un conjunto de elementos de  $\mathbb{V}$ , supongamos que  $n > m$ . Entonces  $w_1, \dots, w_n$  forman un conjunto linealmente dependiente.

**Demostración:**

La idea de la demostración es la siguiente:

Se supone que  $w_1, \dots, w_n$  son linealmente independientes, por lo tanto  $w_i \neq \mathbf{0}$ , para todo  $i$  y además

$$w_1 = a_1 v_1 + \dots + a_m v_m$$

con algún  $a_i \neq 0$ . Sin pérdida de generalidad (ya que es posible volver a numerar los  $v_i$ ) supongamos que  $a_1 \neq 0$ , tenemos entonces que existen escalares  $a'_i$  tales que:

$$v_1 = a'_1 w_1 + a'_2 w_2 + \dots + a'_m w_m$$

Utilizando el mismo argumento para

$$w_2 = b_1 v_1 + \dots + b_m v_m$$

podemos expresar  $\mathbf{v}_2$  como una combinación lineal de  $\mathbf{w}_1, \mathbf{w}_2, \mathbf{v}_3, \dots, \mathbf{v}_m$ . Continuando de la misma forma podemos expresar  $\mathbf{v}_1, \dots, \mathbf{v}_m$  como combinación de  $\mathbf{w}_1, \dots, \mathbf{w}_m$ . Ya que los  $\mathbf{v}_i$  generan  $\mathbb{V}$ , se sigue que los  $\mathbf{w}_i$  también generan este espacio y por lo tanto, como  $n > m$ , se tiene que existen escalares  $c_i$  tales que:

$$\mathbf{w}_n = c_1 \mathbf{w}_1 + \dots + c_m \mathbf{w}_m$$

con lo que llegamos a una contradicción con nuestro supuesto inicial de independencia lineal entre  $\mathbf{w}_1, \dots, \mathbf{w}_n$ .  $\square$

**Ejercicio 1.9:**

Establezca las dimensiones de los siguientes subespacios vectoriales:

- $\mathbb{R}^n$  sobre el campo  $\mathbb{R}$ .
- Recta en  $\mathbb{R}^2$  que pasa por el origen.
- Plano en  $\mathbb{R}^3$  que pasa por el origen.
- $\mathbb{Q}$  sobre  $\mathbb{Z}$  (sugerencia: It's a trap!!!)



## Capítulo 2

# Matrices

### 2.1. Definiciones básicas

**Definición 2.1** (Matriz):

Sean  $\mathbb{F}$  un campo y  $n, m$  entero positivos. Una matriz de tamaño  $m \times n$  es un arreglo rectangular ( $m$  renglones y  $n$  columnas) de números en  $\mathbb{F}$

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

Si  $n = m$ , decimos que es una matriz cuadrada. Para referirnos al elemento en el renglón  $i$ , columna  $j$  utilizaremos la notación  $a_{ij}$ .

De acuerdo a la definición 2.1, un vector renglón

$$\mathbf{a} = (a_1 \quad \dots \quad a_n)$$

es una matriz de tamaño  $1 \times n$ , mientras que un vector columna

$$\mathbf{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$$

es una matriz de tamaño  $n \times 1$ .

**Definición 2.2** (Matriz transpuesta):

Sea  $\mathbf{A}$  una matriz de tamaño  $m \times n$ . La transpuesta de  $\mathbf{A}$ , es la matriz,  $\mathbf{B}$ , de tamaño  $n \times m$  tal que  $b_{ij} = a_{ji}$ , es decir, se intercambian renglones por columnas. Denotaremos a la matriz transpuesta de  $\mathbf{A}$  como  $\mathbf{A}^T$ .

**Ejercicio 2.1:**

Si

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

obtener  $\mathbf{A}^T$ .

**Definición 2.3** (Matriz simétrica):

Si una matriz  $\mathbf{A}$  es tal que  $\mathbf{A} = \mathbf{A}^T$ , se dice que  $\mathbf{A}$  es una matriz **simétrica**.

**Definición 2.4** (Diagonal de una matriz cuadrada):

Sea  $\mathbf{A}$  una matriz cuadrada de  $n \times n$ , la diagonal de  $\mathbf{A}$  son los elementos  $a_{11}, a_{22}, \dots, a_{nn}$ .

**2.2. Tipos especiales de matrices****Definición 2.5** (Matriz cero):

La matriz cero se define como

$$\mathbf{0} = \begin{pmatrix} 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix}$$

**Definición 2.6** (Matriz diagonal):

Una matriz cuadrada de  $n \times n$ , se dice que es una **matriz diagonal** si los únicos elementos distintos de cero son aquellos que pertenecen a su diagonal, es decir, tiene la siguiente forma:

$$\begin{pmatrix} a_{11} & 0 & \dots & 0 \\ 0 & a_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_{nn} \end{pmatrix}$$

En particular, si  $a_{ii} = 1$  para toda  $i$ , la matriz recibe el nombre de **matriz identidad** o **matriz unitaria**

$$\mathbf{I} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}$$

**Definición 2.7** (Matriz triangular):

Se dice que una matriz cuadrada de  $n \times n$  es **triangular superior** si los elementos que están debajo de la diagonal son igual a cero, es decir,  $a_{ij} = 0$  para  $i > j$ .

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ 0 & a_{22} & \dots & a_{2n} \\ \dots & \dots & \ddots & \dots \\ 0 & 0 & \dots & a_{nn} \end{pmatrix}$$

De manera similar, se dice que es **triangular inferior** si  $a_{ij} = 0$  para  $i < j$ .

$$\begin{pmatrix} a_{11} & 0 & \dots & 0 \\ a_{21} & a_{22} & \dots & 0 \\ \dots & \dots & \ddots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}$$

**Ejercicio 2.2:**

¿Cuál es la dimensión del espacio de matrices de  $2 \times 2$ ? Dé una base para este espacio.

**Ejercicio 2.3:**

¿Cuál es la dimensión del espacio de matrices de  $m \times n$ ? Dé una base para este espacio.

**Ejercicio 2.4:**

¿Cuál es la dimensión del espacio de matrices triangulares superiores de tamaño  $n \times n$ ?

**Ejercicio 2.5:**

¿Cuál es la dimensión del espacio de matrices simétricas de  $2 \times 2$ ? Dé una base para este espacio.

**Ejercicio 2.6:**

¿Cuál es la dimensión del espacio de matrices simétricas de  $n \times n$ ? Dé una base para este espacio.

**Ejercicio 2.7:**

¿Cuál es la dimensión del espacio formado por matrices diagonales de  $n \times n$ ? Dé una base.

## 2.3. Operaciones con matrices

**Definición 2.8** (Suma de matrices):

Sean  $\mathbf{A}$  y  $\mathbf{B}$  matrices de tamaño  $m \times n$ , definimos la suma  $\mathbf{A} + \mathbf{B}$  como la matriz  $\mathbf{C}$  tal que  $c_{ij} = a_{ij} + b_{ij}$ . En otras palabras, la suma de matrices se realiza componente por componente.

**Ejercicio 2.8:**

Si

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

y

$$\mathbf{B} = \begin{pmatrix} 5 & 1 & -1 \\ 2 & 1 & -1 \end{pmatrix}$$

Calcule  $\mathbf{A} + \mathbf{B}$ .

Claramente, para la matriz cero de la definición 2.5, se tiene que  $\mathbf{A} + \mathbf{0} = \mathbf{A}$ .

**Definición 2.9** (Producto escalar):

Sea  $\mathbf{A}$  una matriz y  $c$  un escalar, definimos la matriz  $\mathbf{B} = c\mathbf{A}$  como la matriz cuya componente  $b_{ij} = ca_{ij}$ . Es decir, multiplicamos cada elemento de  $\mathbf{A}$  por el escalar  $c$ .

**Ejercicio 2.9:**

Utilizando las definiciones 2.8 y 3.1, demuestre que el espacio de matrices de tamaño  $m \times n$  con elementos dentro un campo  $\mathbb{F}$ , es un espacio vectorial.

**Ejercicio 2.10:**

Demuestre que para cualquier matriz cuadrada  $\mathbf{A}$ , la matriz  $\mathbf{A} + \mathbf{A}^T$  es una matriz simétrica.

**Ejercicio 2.11:**

Si  $\mathbf{A}$  y  $\mathbf{B}$  son matrices de  $m \times n$ , demuestre que

$$(\mathbf{A} + \mathbf{B})^T = \mathbf{A}^T + \mathbf{B}^T.$$

**Ejercicio 2.12:**

Si  $\mathbf{A}$  es una matriz de  $m \times n$  y  $c$  un escalar, demuestre que

$$(c\mathbf{A})^T = c\mathbf{A}^T.$$

**Definición 2.10** (Producto punto):

Sean  $\mathbf{a} = (a_1, \dots, a_n)$  y  $\mathbf{b} = (b_1, \dots, b_n)$  dos vectores<sup>1</sup> con entradas sobre un campo  $\mathbb{F}$ , el **producto punto** entre  $\mathbf{a}$  y  $\mathbf{b}$  se define como

$$\langle \mathbf{a}, \mathbf{b} \rangle = \sum_{i=1}^n a_i b_i$$

**Proposición 2.1:**

El producto punto cumple lo siguiente:

- a)  $\langle \mathbf{a}, \mathbf{b} \rangle = \langle \mathbf{b}, \mathbf{a} \rangle$ .
- b)  $\langle \mathbf{a}, \mathbf{b} + \mathbf{c} \rangle = \langle \mathbf{a}, \mathbf{b} \rangle + \langle \mathbf{a}, \mathbf{c} \rangle = \langle \mathbf{b} + \mathbf{c}, \mathbf{a} \rangle$ .
- c) Si  $k$  es un escalar, entonces

$$\langle k\mathbf{a}, \mathbf{b} \rangle = k \langle \mathbf{a}, \mathbf{b} \rangle = \langle \mathbf{a}, k\mathbf{b} \rangle.$$

- d) Si los componentes de  $\mathbf{a}$  son números reales, entonces

$$\langle \mathbf{a}, \mathbf{a} \rangle = \sum_{i=1}^n a_i^2 \geq 0.$$

- e) Si  $\langle \mathbf{a}, \mathbf{b} \rangle = 0$  para todo  $\mathbf{b}$ , entonces  $\mathbf{a} = (0, \dots, 0)$ .

**Ejercicio 2.13:**

Demuestre la proposición 2.1.

**Definición 2.11** (Producto de matrices):

Sea  $\mathbf{A}$  una matriz de  $m \times n$  y  $\mathbf{B}$  una matriz de  $n \times k$ . El producto  $\mathbf{AB}$ , se define como la matriz de tamaño  $m \times k$ :

$$\begin{pmatrix} \langle \mathbf{a}_1, \mathbf{b}_1 \rangle & \langle \mathbf{a}_1, \mathbf{b}_2 \rangle & \dots & \langle \mathbf{a}_1, \mathbf{b}_k \rangle \\ \langle \mathbf{a}_2, \mathbf{b}_1 \rangle & \langle \mathbf{a}_2, \mathbf{b}_2 \rangle & \dots & \langle \mathbf{a}_2, \mathbf{b}_k \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \mathbf{a}_m, \mathbf{b}_1 \rangle & \langle \mathbf{a}_m, \mathbf{b}_2 \rangle & \dots & \langle \mathbf{a}_m, \mathbf{b}_k \rangle \end{pmatrix}$$

en donde  $\mathbf{a}_i$  representa el  $i$ -ésimo renglón de  $\mathbf{A}$  y  $\mathbf{b}_i$  representa la  $i$ -ésima columna de  $\mathbf{B}$ .

---

<sup>1</sup>Aunque aquí se expresan como vectores renglón, también pueden ser considerados como vectores columna.

**Ejercicio 2.14:**

Programa una función que realice la multiplicación de matrices.

Sugerencia: Programa primero una función que realice el producto punto de vectores.

Como se hace observa en la definición 2.11, para que el producto de dos matrices esté bien definido, el número de columnas de la matriz  $\mathbf{A}$  debe de ser igual al número de renglones de la matriz  $\mathbf{B}$ . En consecuencia, el producto de matrices **no es conmutativo**.

**Proposición 2.2:**

Algunas propiedades del producto de matrices son:

a)  $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$ .

b) Si  $k$  es un escalar

$$\mathbf{A}(k\mathbf{B}) = k(\mathbf{AB})$$

c)  $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$ .

d)  $(\mathbf{AB})^T = \mathbf{B}^T \mathbf{A}^T$  (demostrar).

e) Si  $\mathbf{A}$  es una matriz cuadrada, entonces  $\mathbf{A}^0 = \mathbf{I}$  y para enteros no negativos,  $r$  y  $s$ , tenemos que  $\mathbf{A}^{r+s} = \mathbf{A}^r \mathbf{A}^s$ .

f) El producto punto de la definición 2.10, puede escribirse como el producto de una matriz renglón con una matriz columna:

$$\langle \mathbf{a}, \mathbf{b} \rangle = (a_1, \dots, a_n) \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$$

En la literatura, es muy común representar al vector  $\mathbf{a}$  como un vector columna y al vector  $\mathbf{b}$  como un vector renglón. por lo tanto

$$\langle \mathbf{a}, \mathbf{b} \rangle = \mathbf{a}^T \mathbf{b}.$$

**2.4. Sistemas de ecuaciones lineales**

**Definición 2.12** (Sistemas de ecuaciones lineales):

Sean  $\mathbb{F}$  un campo,  $\mathbf{A}$  una matriz de  $m \times n$  y  $b_1, \dots, b_m$  escalares de  $\mathbb{F}$ .

Ecuaciones del tipo

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

son llamados sistemas de ecuaciones lineales. Los elementos de la matriz  $\mathbf{A}$  son llamados **coeficientes** y los elementos  $x_1, \dots, x_n$  **incógnitas**.

Si  $b_1 = b_2 = \dots = b_m = 0$ , entonces decimos que es un sistema **homogéneo** y el sistema homogéneo asociado al sistema anterior es:

$$\begin{aligned}a_{11}x_1 + \dots + a_{1n}x_n &= 0 \\a_{21}x_1 + \dots + a_{2n}x_n &= 0 \\&\vdots \\a_{m1}x_1 + \dots + a_{mn}x_n &= 0\end{aligned}$$

Este sistema siempre tiene una solución, conocida como **solución trivial** ¿cuál es?

Utilizando las columnas de la matriz  $\mathbf{A}$  de la definición 2.12, podemos reescribir el sistema de ecuaciones como:

$$x_1 \begin{pmatrix} a_{11} \\ \vdots \\ a_{m1} \end{pmatrix} + \dots + x_n \begin{pmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix},$$

o denotando a la  $j$ -ésima columna de  $\mathbf{A}$  como  $\mathbf{a}_j$  y al vector (columna)  $(b_1, \dots, b_m)^T$  como  $\mathbf{b}$

$$x_1 \mathbf{a}_1 + \dots + x_n \mathbf{a}_n = \mathbf{b}.$$

En el caso de un sistema homogéneo, la existencia de una solución no trivial implica una dependencia lineal entre los vectores formados por las columnas de la matriz  $\mathbf{A}$ .

**Teorema 2.1:**

Sea

$$\begin{aligned}a_{11}x_1 + \dots + a_{1n}x_n &= 0 \\a_{21}x_1 + \dots + a_{2n}x_n &= 0 \\&\vdots \\a_{m1}x_1 + \dots + a_{mn}x_n &= 0\end{aligned}$$

un sistema homogéneo de ecuaciones lineales. Si  $n > m$ , entonces existe una solución no trivial.

**Teorema 2.2:**

Sea

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

un sistema no homogéneo de ecuaciones lineales. Si  $n = m$  y las columnas de la matriz  $\mathbf{A}$  son linealmente independientes, entonces existe una solución única para el sistema.

**Ejercicio 2.15:**

Demuestre los teoremas 2.1 y 2.2.

Sugerencia: Utilice el teorema 1.1.

**Ejercicio 2.16:**

Suponga que se tiene un sistema lineal homogéneo y que  $n = m$ . Además suponga que las columnas de la matriz  $\mathbf{A}$  son linealmente independientes. Demuestre que la única solución es la solución trivial.

## 2.5. Determinantes y matrices inversas

**Definición 2.13** (Matriz inversa):

Sea  $\mathbf{A}$  una matriz cuadrada de  $n \times n$ . Una matriz,  $\mathbf{B}$ , de  $n \times n$ ; se dice que es la **inversa** de  $\mathbf{A}$  si:

$$\mathbf{AB} = \mathbf{BA} = \mathbf{I}$$

en donde  $\mathbf{I}$  es la matriz identidad de tamaño  $n \times n$  (ver definición 2.6). En este caso decimos que  $\mathbf{A}$  es **invertible** o **no singular**. Denotaremos a la inversa de  $\mathbf{A}$  como  $\mathbf{A}^{-1}$ . Si una matriz no tiene inversa, decimos que es **no invertible** o **singular**.

**Nota 2.1:**

Es importante observar que la definición 2.13 únicamente aplica para matrices cuadradas.



**Teorema 2.3:**

Sean  $\mathbf{A}$  y  $\mathbf{B}$  matrices cuadradas del mismo tamaño, entonces tenemos que:

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

**Teorema 2.4:**

Si una matriz es singular, entonces al menos uno de sus renglones (o columnas) es una combinación lineal de los otros renglones (o columnas).

**Ejercicio 2.17:**

Demuestre que si  $\mathbf{A}$  es una matriz invertible, entonces su inversa es única.

Para poder encontrar la inversa de una matriz, es necesario primero definir lo que es el determinante de una matriz.

**Definición 2.14** (Determinante - matriz  $2 \times 2$ ):

Sea

$$\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

una matriz de  $2 \times 2$ , el determinante de  $\mathbf{A}$ , se define como:

$$\det(\mathbf{A}) = |\mathbf{A}| = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc.$$

La definición de determinante para matrices de mayor tamaño se puede definir de forma inductiva.

**Definición 2.15** (Determinante - matriz  $n \times n$ ):

Sea  $\mathbf{A}$  una matriz de  $n \times n$ , el determinante de  $\mathbf{A}$  se define como:

$$\det(\mathbf{A}) = a_{1j}C_{1j} + \dots + a_{nj}C_{nj}.$$

En donde  $C_{ij} = (-1)^{i+j}M_{ij}$  con  $M_{ij}$  el determinante obtenido al eliminar el renglón  $i$  y la columna  $j$  de la matriz  $\mathbf{A}$ .

**Ejercicio 2.18:**

Utilizando las definiciones 2.14 y 2.15 calcule el determinante de la matriz

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

**Teorema 2.5:**

Para una matriz de  $n \times n$   $\mathbf{A}$ , el determinante cumple lo siguiente:

- a)  $\mathbf{A}$  es no singular si y sólo si  $|\mathbf{A}| \neq 0$
- b)  $|\mathbf{A}^T| = |\mathbf{A}|$
- c)  $|\mathbf{A}^{-1}| = |\mathbf{A}|^{-1}$
- d)  $|\mathbf{A}| = 0$  para  $\mathbf{A}$  singular.
- e)  $|c\mathbf{A}| = c^n |\mathbf{A}|$  para  $c$  un escalar.
- f)  $|\mathbf{AB}| = |\mathbf{A}||\mathbf{B}|$  con  $\mathbf{B}$  una matriz de  $n \times n$

**Teorema 2.6:**

Sea  $\mathbf{A}$  una matriz invertible, entonces

$$\mathbf{A}^{-1} = \frac{1}{|\mathbf{A}|} \mathbf{C}^T$$

en donde

$$\mathbf{C} = \begin{pmatrix} C_{11} & C_{12} & \dots & C_{1n} \\ C_{21} & C_{22} & \dots & C_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ C_{n1} & C_{n2} & \dots & C_{nn} \end{pmatrix}$$

en donde  $C_{ij} = (-1)^{i+j} M_{ij}$  con  $M_{ij}$  el determinante obtenido al eliminar el renglón  $i$  y la columna  $j$  de la matriz  $\mathbf{A}$ .

**Ejercicio 2.19:**

Demuestre que una matriz cuadrada con un renglón o una columna llena de ceros, no tiene inversa.

**Ejercicio 2.20:**

Demuestre que una matriz cuadrada y triangular (superior o inferior) cuya diagonal tiene un elemento igual a cero, no tiene inversa.

## Capítulo 3

# Producto escalar y ortogonalidad

### 3.1. Producto escalar

Como vimos, el producto punto satisface las propiedades dadas en la proposición 2.1. Este producto es un caso particular de un mapeo llamado **producto escalar**.

**Definición 3.1** (Producto escalar):

Sea  $\mathbb{V}$  un espacio vectorial sobre un campo  $\mathbb{F}$ . Un **producto escalar** (o producto interno) es un mapeo  $\langle \cdot, \cdot \rangle : \mathbb{V} \times \mathbb{V} \rightarrow \mathbb{F}$  que satisface las siguientes propiedades:

- a)  $\langle \mathbf{v}, \mathbf{v} \rangle \geq 0$ , con igualdad si y sólo si  $\mathbf{v} = \mathbf{0}$ .
- b)  $\langle \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{w}, \mathbf{v} \rangle$  para todo  $\mathbf{v}, \mathbf{w} \in \mathbb{V}$  (simetría).
- c)  $\langle \mathbf{u}, \mathbf{v} + \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{u}, \mathbf{w} \rangle$ .
- d) Si  $k \in \mathbb{F}$ , entonces:

$$\langle k\mathbf{u}, \mathbf{v} \rangle = k \langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{u}, k\mathbf{v} \rangle$$

- e) Si además se satisface que para un elemento  $\mathbf{v} \in \mathbb{V}$  fijo tal que para todo  $\mathbf{w} \in \mathbb{V}$  se tiene que  $\langle \mathbf{v}, \mathbf{w} \rangle = 0$ , entonces  $\mathbf{v} = \mathbf{0}$ . Decimos que el producto escalar es **no degenerado**.

**Ejercicio 3.1:**

Sea  $\mathbb{V}$  el espacio de funciones continuas con valores en  $\mathbb{R}$  y definidas en el intervalo  $[0, 1]$ . Demuestre que

$$\langle f, g \rangle = \int_0^1 f(t)g(t)dt$$

es un producto escalar.

**Definición 3.2** (Norma de un vector):

Sea  $\mathbb{V}$  un espacio vectorial sobre un campo  $\mathbb{F}$ . Sea  $\langle \cdot, \cdot \rangle$  un producto escalar definido en  $\mathbb{V}$ . La **norma** de un vector  $\mathbf{v} \in \mathbb{V}$  se define como:

$$\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}.$$

**Ejercicio 3.2:**

En el caso del producto punto, ¿cuál es la expresión de la norma de un vector? De acuerdo a esta expresión, dé una expresión, en términos de la norma, para la distancia euclidiana en  $\mathbb{R}^n$  entre dos vectores.

**Proposición 3.1:**

Sea  $\mathbb{V}$  un espacio vectorial sobre un campo  $\mathbb{F}$ , algunas de las propiedades de una norma son las siguientes:

- a)  $\|k\mathbf{v}\| = |k| \|\mathbf{v}\|$ , para  $k \in \mathbb{F}$ ,  $\mathbf{v} \in \mathbb{V}$ .
- b)  $\langle \mathbf{v}, \mathbf{w} \rangle^2 \leq \|\mathbf{v}\|^2 \|\mathbf{w}\|^2$  con igualdad si y sólo si  $\mathbf{v} = k\mathbf{w}$  para algún escalar  $k \in \mathbb{F}$  (Desigualdad de Cauchy-Schwarz).
- c)  $\|\mathbf{v} + \mathbf{w}\| \leq \|\mathbf{v}\| + \|\mathbf{w}\|$  (Desigualdad del triángulo).
- d)  $\|\mathbf{v} + \mathbf{w}\|^2 + \|\mathbf{v} - \mathbf{w}\|^2 = 2(\|\mathbf{v}\|^2 + \|\mathbf{w}\|^2)$  (Ley del paralelogramo).

**Ejercicio 3.3:**

Utilizando la desigualdad de Cauchy-Schwarz, demuestre la desigualdad del triángulo.

Sugerencia: exprese  $\|\mathbf{v} + \mathbf{w}\|^2$  en términos del producto escalar.

**Definición 3.3** (Vector unitario):

Un vector  $\mathbf{v}$  en un espacio vectorial  $\mathbb{F}$  sobre el cual se ha definido una norma, se dice que es un **vector unitario** (o vector dirección), si  $\|\mathbf{v}\| = 1$ .

**Ejercicio 3.4:**

Demostrar la siguiente afirmación:

Dado un vector  $\mathbf{v} \neq \mathbf{0}$ , podemos formar un vector unitario,  $\mathbf{w}$ , con la misma dirección que  $\mathbf{v}$  de la siguiente manera:

$$\mathbf{w} = \frac{\mathbf{v}}{\|\mathbf{v}\|}.$$

**3.2. Ortogonalidad**

**Definición 3.4** (Vectores ortogonales y paralelos):

Decimos que dos vectores son **ortogonales** si

$$\langle \mathbf{v}, \mathbf{w} \rangle = 0.$$

Si

$$\langle \mathbf{v}, \mathbf{w} \rangle = \|\mathbf{v}\| \|\mathbf{w}\|$$

decimos que son **paralelos**. En otras palabras, dos vectores son paralelos si uno es múltiplo de otro (ver desigualdad de Cauchy-Schwarz).

**Definición 3.5** (Conjunto ortonormales):

Sea  $\mathcal{X} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$  un conjunto de vectores distintos del vector cero. Decimos que el conjunto  $\mathcal{X}$  es un **conjunto ortogonal** si  $\langle \mathbf{v}_i, \mathbf{v}_j \rangle = 0$  para todo  $i \neq j$ . Si además tenemos que  $\|\mathbf{v}_i\| = 1$  para toda  $i$ , entonces decimos que  $\mathcal{X}$  es un **conjunto ortonormal**.

**Ejercicio 3.5:**

Sea  $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$  un conjunto de vectores ortogonales, demuestre que

$$\left\| \sum_{i=1}^n \mathbf{v}_i \right\|^2 = \sum_{i=1}^n \|\mathbf{v}_i\|^2.$$

Sugerencia: Primero analice el caso en que  $n = 2$  y argumente la generalización.

**Ejercicio 3.6:**

Demuestre que un conjunto ortogonal de vectores distintos del vector cero,  $\mathcal{X} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ , forma un conjunto linealmente independiente. El recíproco de esta proposición ¿es verdadero?

**Definición 3.6** (Bases ortogonales y ortonormales):

Llamamos **base ortogonal** a una base formada por un conjunto de vectores ortogonales, si además estos vectores son ortonormales, la base es llamada **base ortonormal**.

**Definición 3.7** (Matrices ortogonales y ortonormales):

Supongamos que tenemos una matriz,  $\mathbf{A}$ , de  $n \times n$ . Decimos que  $\mathbf{A}$  es una matriz ortogonal, si sus vectores columna forman un conjunto de vectores ortogonales. Si además estos vectores columna son ortonormales, decimos que  $\mathbf{A}$  es una matriz ortonormal.

**Nota 3.1:**

Nosotros hacemos distinción entre matrices ortogonales y ortonormales, en la literatura es común utilizar únicamente el término matriz ortogonal para referirse a ambos tipos de matrices.

**Teorema 3.1:**

Sea  $\mathbf{A}$  una matriz de  $n \times n$ . Entonces los siguientes puntos son equivalentes:

- I) Las columnas de  $\mathbf{A}$  son vectores ortonormales.
- II)  $\mathbf{A}^T \mathbf{A} = \mathbf{I} = \mathbf{A} \mathbf{A}^T$ .
- III) Los renglones de  $\mathbf{A}$  son vectores ortonormales.

Por lo tanto, para una matriz ortonormal, tenemos que  $\mathbf{A}^{-1} = \mathbf{A}^T$ .

**Ejercicio 3.7:**

Demuestre el teorema 3.1.

## Capítulo 4

# Valores y vectores característicos

Cuando multiplicamos un vector  $\mathbf{x}$  por una matriz  $\mathbf{A}$ , lo que realmente se está haciendo es una transformación lineal del vector  $\mathbf{x}$ . Esta transformación, usualmente, cambia el ángulo de este vector.

Sin embargo, para ciertos vectores  $\mathbf{x}$ , la transformación  $\mathbf{Ax}$  es igual a  $\lambda\mathbf{x}$ , para un escalar  $\lambda$ . Geométricamente, esto se interpreta como una contracción (si  $\lambda < 1$ ) o una expansión (si  $\lambda > 1$ ) del vector.

En esta sección nos restringimos a matrices sobre el campo de los números complejos  $\mathbb{C}$ .

**Definición 4.1** (Valores y vectores característicos):

Sea  $\mathbf{A}$  una matriz cuadrada de  $n \times n$ . Un **valor característico**<sup>1</sup> de la matriz  $\mathbf{A}$  es un escalar tal que satisface la ecuación (4.1)

$$\mathbf{Ax} = \lambda\mathbf{x} \tag{4.1}$$

para un vector  $\mathbf{x} \neq \mathbf{0}$ . Cualquier vector  $\mathbf{x}$  que satisface lo anterior, es llamado **vector característico**<sup>2</sup> asociado a  $\lambda$ .

**Ejercicio 4.1:**

En la definición 4.1, ¿es necesario que la matriz  $\mathbf{A}$  sea una matriz cuadrada? Argumente su respuesta.

De acuerdo a la ecuación (4.1), se tiene que  $\lambda$  es un valor característico de la matriz  $\mathbf{A}$  si y sólo si el sistema (homogéneo) de ecuaciones lineales

---

<sup>1</sup>También llamado un eigenvalue

<sup>2</sup>Eigenvector

$(\mathbf{A} - \lambda \mathbf{I})$  tiene una solución no trivial  $\mathbf{x} \neq \mathbf{0}$ . Esto quiere decir que cualquier  $\lambda$  que haga que la matriz  $\mathbf{A} - \lambda \mathbf{I}$  se vuelva una matriz singular, será un valor característico de la matriz  $\mathbf{A}$ . De manera equivalente, basta encontrar un  $\lambda$ , tal que  $|\lambda \mathbf{I} - \mathbf{A}| = 0$ .

**Definición 4.2** (Polinomio característico):

El polinomio característico de una matriz  $\mathbf{A}$  de  $n \times n$ , es el polinomio dado por

$$p_{\mathbf{A}}(\lambda) = |\lambda \mathbf{I} - \mathbf{A}|. \quad (4.2)$$

Es fácil ver entonces, que las raíces del polinomio característico son los valores característicos de la matriz  $\mathbf{A}$ .

Observemos que el polinomio característico es un polinomio de grado  $n$  y por lo tanto, de acuerdo al teorema fundamental del álgebra, tiene  $n$  raíces (ya sean complejas o reales). Así, una matriz cuadrada de  $n \times n$  tiene  $n$  valores característicos, algunos de ellos complejos, otros reales e incluso puede haber valores repetidos.

**Ejercicio 4.2:**

Demuestre que si una matriz,  $\mathbf{A}$ , de tamaño  $n \times n$ , tiene menos de  $n$  columnas linealmente independientes, entonces existe un valor característico  $\lambda$ .

**Ejercicio 4.3:**

Si  $\mathbf{U}$  es una matriz triangular superior, encuentre sus valores característicos y dé un posible vector característico asociado a cada uno de ellos.

**Teorema 4.1** (Propiedades de valores y vectores característicos):

Sea  $\mathbf{A}$  una matriz de  $n \times n$ , tenemos entonces:

- a) Si  $\lambda$  es un valor característico de la matriz  $\mathbf{A}$ , entonces también es un valor característico de  $\mathbf{A}^T$
- b) El producto de los valores característicos de  $\mathbf{A}$ , es igual a  $|\mathbf{A}|$ .
- c) La suma de los valores característicos de  $\mathbf{A}$ , es igual a la **traza** de  $\mathbf{A}$ ,  $tr(\mathbf{A})$ , es decir, igual a la suma de los elementos en la diagonal de  $\mathbf{A}$

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



- d) Si  $\{\lambda_1, \lambda_2, \dots, \lambda_k\}$  es un conjunto de valores característicos de  $\mathbf{A}$ , todos ellos distintos, entonces el conjunto con los vectores característicos  $\{\mathbf{x}_1, \dots, \mathbf{x}_k\}$  es un conjunto linealmente independiente (Demostrar).
- e) Si  $\mathbf{A}$  es una matriz **simétrica** sobre el campo  $\mathbb{R}$ , entonces todos sus valores característicos pertenecen a  $\mathbb{R}$ .
- f) Si  $\mathbf{x}_1$  y  $\mathbf{x}_2$  son dos vectores característicos, distintos, de una matriz real simétrica  $\mathbf{A}$  y  $\lambda_1, \lambda_2$  son sus respectivos valores característicos. Entonces  $\mathbf{x}_1$  y  $\mathbf{x}_2$  son ortogonales (utilizando el producto punto como producto escalar). (Demostrar)

Finalizaremos este capítulo con el siguiente resultado:

**Teorema 4.2** (Descomposición espectral para matrices simétricas):

Si  $\mathbf{A}$  es una matriz simétrica sobre el campo  $\mathbb{R}$ , entonces existe una matriz ortogonal (propriadamente ortonormal)  $\mathbf{P}$  y una matriz diagonal  $\mathbf{\Lambda}$  tales que

$$\mathbf{P}^T \mathbf{A} \mathbf{P} = \mathbf{\Lambda}.$$

La diagonal de  $\mathbf{\Lambda}$  está formada por los valores característicos de  $\mathbf{A}$  y las columnas de  $\mathbf{P}$  son los vectores característicos correspondientes.

## Capítulo 5

# Descomposición por valores singulares

La descomposición por valores singulares (SVD por sus siglas en inglés) es una de los métodos mas utilizados para factorizar matrices rectangulares. Particularmente, esta descomposición es una de las técnicas más utilizadas para reducir la dimensión en los conjuntos de datos.

Antes de introducir la descomposición necesitamos la siguiente definición.

**Definición 5.1** (Rango de una matriz):

El **rango de las columnas** de una matriz  $\mathbf{A}$ , es el número de columnas de  $\mathbf{A}$  que son linealmente independientes.

El **rango de los renglones** de una matriz  $\mathbf{A}$ , es el número de renglones de  $\mathbf{A}$  que son linealmente independientes.

De acuerdo a la definición 5.1, el rango de una matriz es una medida de redundancia, una matriz con un rango bajo (pocos renglones o columnas linealmente independientes) indica que la información contenida en ella es redundante. De acuerdo a lo anterior, una matriz no singular no tiene redundancia

**Teorema 5.1** (Descomposición por valores singulares):

Sea  $\mathbf{A}$  una matriz de  $m \times n$  sobre el campo  $\mathbb{R}$  con un rango  $r \leq \min(m, n)$ . Existen matrices ortogonales  $\mathbf{U}$  de  $m \times m$  y  $\mathbf{V}$  de  $n \times n$ , tales que

$$\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^T \quad \text{en donde } \mathbf{D} = \begin{pmatrix} \Sigma & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{pmatrix} \text{ y } \Sigma = \begin{pmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_r \end{pmatrix}$$

con  $\mathbf{D}$  una matriz de  $m \times n$ ,  $\Sigma$  es una matriz de  $r \times r$  y los  $\sigma_i$ 's son números reales tales que  $\sigma_1 \geq \sigma_2 \dots \geq \sigma_r \geq 0$  estos números son llamados **valores singulares** de la matriz  $\mathbf{A}$ . Las matrices  $\mathbf{O}$  están formadas por ceros y tienen dimensiones que hacen que el producto anterior esté bien definido.

Esta descomposición también se puede expresar de la siguiente manera:

$$\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^T = (\mathbf{U}_1 : \mathbf{U}_2) \begin{pmatrix} \Sigma & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{pmatrix} \begin{pmatrix} \mathbf{V}_1^T \\ \mathbf{V}_2^T \end{pmatrix} = \mathbf{U}_1 \Sigma \mathbf{V}_1^T.$$

en donde  $\mathbf{U}_1$  y  $\mathbf{V}_1$  son matrices de  $m \times r$  y  $n \times r$  respectivamente.

## 5.1. Reducción de la dimensión de los datos

Supongamos que tenemos una matriz de  $m \times n$ ,  $\mathbf{A}$ , con un rango  $r < \min(m, n)$ . De acuerdo al teorema 5.1, tenemos que  $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^T$ . Utilizando la ortogonalidad de  $\mathbf{U}$ , tenemos que

$$\mathbf{U}^T \mathbf{A} = \Sigma \mathbf{V}^T$$

lo que también puede expresarse como

$$\begin{pmatrix} \mathbf{U}_1^T \\ \mathbf{U}_2^T \end{pmatrix} \mathbf{A} = \begin{pmatrix} \Sigma & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{pmatrix} \begin{pmatrix} \mathbf{V}_1^T \\ \mathbf{V}_2^T \end{pmatrix} = \begin{pmatrix} \Sigma \mathbf{V}_1^T \\ \mathbf{O} \end{pmatrix}.$$

La matriz  $\mathbf{U}_1^T \mathbf{A} = \Sigma \mathbf{V}_1^T$  es una matriz de  $r \times n$  con rango  $r$ , que no presenta redundancia en la información. Por lo tanto, en lugar de utilizar la matriz  $\mathbf{A}$ , se puede reducir la dimensión de los datos y su redundancia utilizando la matriz  $\Sigma \mathbf{V}_1^T$ .

Observemos además que ya que la matriz  $\Sigma$  es una matriz diagonal, el único efecto que tiene sobre  $\mathbf{V}_1^T$  es estirar o encoger sus vectores columna (de acuerdo a si los  $\sigma_i$ 's son mayores o menores a 1) por lo tanto, también podríamos utilizar la matriz  $\mathbf{V}_1^T$  en lugar de  $\mathbf{A}$ .

## Capítulo 6

# Análisis de componentes principales

El análisis de componentes principales (PCA por sus siglas en inglés), está basado en la descomposición espectral del teorema 4.2. Este análisis, es una de las formas más utilizadas para expresar un conjunto de variables correlacionadas utilizando un conjunto de variables no correlacionadas y es, junto con la descomposición por valores singulares, una de las técnicas más comunes para reducir la dimensión en un conjunto de datos.

### 6.1. Formas cuadráticas

Antes de detallar el análisis de componentes principales, es útil tener los siguientes resultados.

**Definición 6.1** (Forma cuadrática):

Sea  $\mathbf{A}$  una matriz de  $n \times n$ , la **forma cuadrática** asociada a  $\mathbf{A}$ , es la expresión

$$\mathbf{x}^T \mathbf{A} \mathbf{x}$$

en donde  $\mathbf{x}$  es un vector (columna) de dimensión  $n$ .

Observemos que la forma cuadrática de una matriz es un escalar.

**Definición 6.2** (Matrices definidas):

Sea  $\mathbf{A}$  una matriz de  $n \times n$ , decimos que  $\mathbf{A}$  es una matriz **definida positiva** si  $\mathbf{x}^T \mathbf{A} \mathbf{x} > 0$  para todo vector  $n$ -dimensional  $\mathbf{x}$  distinto del vector cero.

Si  $\mathbf{x}^T \mathbf{A} \mathbf{x} < 0$ , para todo vector  $\mathbf{x} \neq \mathbf{0}$ , decimos que  $\mathbf{A}$  es una matriz **definida negativa**.

Finalmente, si  $\mathbf{x}^T \mathbf{A} \mathbf{x} \geq 0$  para todo vector  $\mathbf{x} \neq \mathbf{0}$ , entonces decimos que es una matriz **semidefinida positiva**. Si  $\mathbf{x}^T \mathbf{A} \mathbf{x} \leq 0$ , para todo vector  $\mathbf{x} \neq \mathbf{0}$ , decimos que  $\mathbf{A}$  es una matriz **semidefinida negativa**.

**Proposición 6.1:**

Para una matriz,  $\mathbf{A}$  de  $n \times n$ , las siguientes condiciones son equivalentes:

- a)  $\mathbf{A}$  es una matriz simétrica definida positiva.
- b)  $\mathbf{x}^T \mathbf{A} \mathbf{x} > 0$  para todo  $\mathbf{x} \neq \mathbf{0}$ .
- c) Todos los valores característicos de  $\mathbf{A}$  son positivos.

**Proposición 6.2:**

Para una matriz,  $\mathbf{A}$  de  $n \times n$  sobre el campo  $\mathbb{R}$ , las siguientes condiciones son equivalentes:

- a)  $\mathbf{A}$  es una matriz simétrica semidefinida positiva.
- b)  $\mathbf{x}^T \mathbf{A} \mathbf{x} \geq 0$  para todo  $\mathbf{x} \neq \mathbf{0}$ .
- c) Todos los valores característicos de  $\mathbf{A}$  son no negativos.

**Ejercicio 6.1:**

Demuestre la proposición 6.1.

## 6.2. Componentes principales

Supongamos que tenemos una matriz,  $\mathbf{X}$  de  $m \times n$  sobre el campo  $\mathbb{R}$ . Cada columna de  $\mathbf{X}$  (denotadas por  $\mathbf{x}_1, \dots, \mathbf{x}_n$ ) representa una variable y cada renglón una observación para cada variable. Supongamos además que cada columna tiene media igual a cero y desviación estándar igual a uno.

Los **componentes principales** de la matriz  $\mathbf{X}$ , se definen como una combinación lineal de las columnas de  $\mathbf{X}$ , en donde los coeficientes se eligen de tal forma que:

- Cada componente principal no está correlacionado con los demás.
- El primer componente principal explica la mayor variación en los datos de  $\mathbf{X}$ , el segundo explica la mayor cantidad de variación que todavía falta por explicar y así sucesivamente.

Ya que cada columna de  $\mathbf{X}$  está estandarizada (media cero y desviación estándar igual a uno) podemos formar la matriz,  $\mathbf{V}$ , de correlaciones entre las variables de  $\mathbf{X}$ .

$$\mathbf{V} = \frac{1}{m} \mathbf{X}^T \mathbf{X}.$$

Utilizando los valores propios de la matriz  $\mathbf{V}$ , ordenados de forma decreciente,  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ , podemos formar la matriz diagonal  $\mathbf{D}$ . Además podemos formar la matriz ortogonal,  $\mathbf{W}$ , cuya columna  $i$  corresponde al vector característico correspondiente a  $\lambda_i$ . La matriz de componentes principales,  $\mathbf{P}$ , es una matriz de  $m \times n$  resultado del siguiente producto:

$$\mathbf{P} = \mathbf{XW}.$$

**Ejercicio 6.2:**

Compruebe que el  $j$ -ésimo componente principal, es decir, la columna  $j$ -ésima de  $\mathbf{P}$ , se escribe como una combinación lineal de las columnas de  $\mathbf{X}$  en donde los coeficientes están dados por el  $j$ -ésimo vector propio de la matriz  $\mathbf{V}$ .

**Definición 6.3** (Variación total):

La variación total de la matriz  $\mathbf{X}$ , es la suma de los valores característicos de la matriz  $\mathbf{V}$ . Es decir

$$\lambda_1 + \dots + \lambda_n.$$

Utilizando la variación total de la definición 6.3, podemos calcular la proporción de variación que es explicada por el  $j$ -ésimo componente principal

$$\frac{\lambda_j}{\lambda_1 + \dots + \lambda_n}.$$

De esta manera, para reducir la dimensión de nuestros datos, en lugar de utilizar las  $n$  columnas de la matriz  $\mathbf{X}$  podemos utilizar las primeras  $k$  columnas de la matriz  $\mathbf{P}$ .

El número  $k$  usualmente se determina estableciendo un umbral  $r$  y tomando  $k$  como

$$k = \min \left\{ k^* : \frac{\lambda_1 + \dots + \lambda_{k^*}}{\lambda_1 + \dots + \lambda_n} > r \right\}.$$