Ortogonalidad

1 *Producto interior. Longitud y ortogonalidad*

Definición 1 Sean $\mathbf{u}$ y $\mathbf{v}$ vectores de $\mathbb{R}^n$. Se define el *producto escalar (o producto interior)* de $\mathbf{u}$ y $\mathbf{v}$ como

$$\mathbf{u}\cdot\mathbf{v} = \mathbf{u}^T\mathbf{v} = (u_1, u_2, \ldots, u_n)\begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} = u_1v_1 + u_2v_2 + \ldots u_nv_n.$$

Ejemplo 1 Calcular el producto escalar de $\mathbf{u} = \begin{pmatrix} 2 \\ -5 \\ 1 \end{pmatrix}$ y $\mathbf{v} = \begin{pmatrix} 3 \\ 2 \\ -3 \end{pmatrix}$.

Dados $\mathbf{u}$, $\mathbf{v}$ y $\mathbf{w}$ vectores de $\mathbb{R}^n$ y λ un escalar, el producto escalar verifica las siguientes propiedades:

1. $\mathbf{u}\cdot\mathbf{v} = \mathbf{v}\cdot\mathbf{u}$.
2. $(\mathbf{u}+\mathbf{v})\cdot\mathbf{w} = \mathbf{u}\cdot\mathbf{w} + \mathbf{v}\cdot\mathbf{w}$.
3. $(\lambda\mathbf{u})\cdot\mathbf{v} = \lambda(\mathbf{u}\cdot\mathbf{v}) = \mathbf{u}\cdot(\lambda\mathbf{v})$.
4. $\mathbf{u}\cdot\mathbf{u} \geq 0$ y $\mathbf{u}\cdot\mathbf{u} = 0$ si y sólo si $\mathbf{u} = 0$.

Definición 2 Dado un vector $\mathbf{u} \in \mathbb{R}^n$. Definimos su *norma (o longitud)* como el escalar no negativo

$$\|\mathbf{u}\| = \sqrt{\mathbf{u}^T\mathbf{u}} = \sqrt{\mathbf{u}\cdot\mathbf{u}} = \sqrt{u_1^2 + u_2^2 + \cdots + u_n^2}.$$

Ejercicio 2 Demostrar que para cualquier escalar c se tiene $\|c\mathbf{u}\| = |c|\,\|\mathbf{u}\|$.

Un vector cuya longitud es 1 se llama vector unitario. Para obtener un vector unitario u a partir de otro dado, v, basta dividir el vector $\mathbf{v}$ por su norma. Ambos vectores tienen la misma dirección, pero distinta longitud. Al proceso mediante el cual se obtiene $\mathbf{u}$ a partir de $\mathbf{v}$ se le conoce con el nombre de normalización.

Ejercicio 3 Sea $\mathbf{v} = \begin{pmatrix} 1 \\ -2 \\ 2 \\ 0 \end{pmatrix}$. Encontrar un vector unitario $\mathbf{u}$ en la misma dirección que $\mathbf{v}$.

Ejercicio 4 Sea W el subespacio generado por $\mathbf{x} = (\frac{2}{3}, 1)^T$. Encontrar un vector unitario $\mathbf{z}$ que sea una base de W.

Definición 3 Sean $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$. Se define la *distancia* entre los vectores $\mathbf{u}$ y $\mathbf{v}$ como la norma del vector $\mathbf{u}-\mathbf{v}$,

$$\operatorname{dist}(\mathbf{u},\mathbf{v}) = \|\mathbf{u}-\mathbf{v}\| = \|\mathbf{v}-\mathbf{u}\| = \operatorname{dist}(\mathbf{v},\mathbf{u}).$$

Ejercicio 5 Calcular la distancia entre $\mathbf{u} = \begin{pmatrix} 7 \\ 2 \end{pmatrix}$ y $\mathbf{v} = \begin{pmatrix} 3 \\ 2 \end{pmatrix}$.

Definición 4 Sean $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$. Diremos que $\mathbf{u}$ y $\mathbf{v}$ son *ortogonales* si $\mathbf{u}\cdot\mathbf{v} = 0$.

Obsérvese que el vector $\mathbf{0}$ es ortogonal a cualquier vector $\mathbf{u}$, porque $\mathbf{0} \cdot \mathbf{u} = 0$, siempre.

Definición 5 Dado un vector $\mathbf{v}$ y un subespacio W diremos que $\mathbf{v}$ es *ortogonal* a W si $\mathbf{v} \cdot \mathbf{w} = 0$ para todo vector $\mathbf{w} \in W$.

Dos subespacios vectoriales V y W de un espacio vectorial se dicen *ortogonales* si $\mathbf{v} \cdot \mathbf{w} = 0$ para todo vector $\mathbf{v} \in V$ y $\mathbf{w} \in W$. En este caso, diremos que V es el complemento ortogonal de W y lo denotamos por $V = W^{\perp}$ (y W es también el complemento ortogonal de V). Dicho de otra forma: el complemento ortogonal de W es el conjunto de todos los vectores $\mathbf{v}$ que son perpendiculares a W,

$$W^{\perp} = \{\mathbf{v} \in \mathbb{R}^n \ : \mathbf{v} \cdot \mathbf{w} = 0, \text{ para todo } \mathbf{w} \in W\}.$$

Observación. Obsérvese que para comprobar si un vector $\mathbf{v}$ pertenece al complemento ortogonal de un espacio dado W, basta con comprobar si $\mathbf{v}$ es ortogonal a un conjunto que genere W.

Ejemplo 6 Hay cuatro subespacios importantes $N(A), C(A), F(A), N(A^T)$. ¿Cuáles son ortogonales?

El espacio $F(A)$ es por definición el generado por las filas de A, $F(A) = \text{Gen}\{\mathbf{f_1}, \mathbf{f_2}, \dots, \mathbf{f_n}\}$ (visto como vectores columna). Para demostrar que los subespacios son ortogonales, basta con ver que dado $\mathbf{x} \in N(A)$, entonces $\mathbf{x}$ es ortogonal a cada una de las filas. Pero como $\mathbf{x} \in N(A)$, por definición

$$A\mathbf{x} = \begin{pmatrix} -\mathbf{f_1}^T- \\ -\mathbf{f_2}^T- \\ \vdots \\ -\mathbf{f_n}^T- \end{pmatrix} \mathbf{x} = 0, \qquad \text{¡Cuidado con la notación!}$$

y por tanto

$$\mathbf{f_1}^T\mathbf{x} = 0, \ \dots, \ \mathbf{f_n}^T\mathbf{x} = 0.$$

Análogamente, para demostrar que $C(A) = \text{Gen}\{\mathbf{a_1}, \mathbf{a_2}, \dots, \mathbf{a_m}\}$ es el complemento ortogonal de $N(A^T)$ comprobamos que un vector $\mathbf{y} \in N(A^T)$ arbitrario es ortogonal a todas las columnas de A: Si $\mathbf{y} \in N(A^T)$, entonces

$$A^T\mathbf{y} = \begin{pmatrix} -\mathbf{a_1}^T- \\ -\mathbf{a_2}^T- \\ \vdots \\ -\mathbf{a_m}^T- \end{pmatrix} \mathbf{y} = 0, \qquad \text{¡Cuidado con la notación!}$$

y por tanto

$$\mathbf{a_1}^T\mathbf{y} = 0, \ \dots, \ \mathbf{a_m}^T\mathbf{y} = 0.$$

Importante. Sea A una matriz $m \times n$. El complemento ortogonal del espacio fila de A es el espacio nulo de A y el complemento ortogonal del espacio columna de A es el espacio nulo de A^T:

$$(F(A))^{\perp} = N(A), \qquad (C(A))^{\perp} = N(A^T).$$

2 *Proyecciones ortogonales*

¿Cuál es la sombra que proyecta el vector $\mathbf{b} = (2, 3, 4)^T$ sobre el plano xy? ¿Y sobre el eje z? ¿Cómo podemos calcular la proyección de un vector sobre un subespacio arbitrario?

Idea. Dado $\mathbf{v} \in \mathbb{R}^n$ y W un subespacio de $\mathbb{R}^n$, queremos descomponer $\mathbf{v}$ como suma de dos vectores

$$\mathbf{v} = \mathbf{p_W} + \mathbf{e},$$

tal que $\mathbf{p_W} \in W$ y $\mathbf{e} \in W^{\perp}$.

Ejemplo 7 Proyectar un vector $\mathbf{v}$ sobre un subespacio vectorial de dimensión 1 (una recta):

Sea $\mathbf{a}$ el vector director de la recta. Buscamos λ y $\mathbf{p_W}$ tal que

$$\mathbf{p_W} = \lambda \mathbf{a}$$

Sabemos que si $\mathbf{v} = \mathbf{p_W} + \mathbf{e}$, entonces

- $\mathbf{a} \cdot \mathbf{e} = 0$
- $\mathbf{e} = \mathbf{v} - \mathbf{p_W} = \mathbf{v} - \lambda \mathbf{a}$

Juntando estas dos expresiones se obtiene:

$$0 = \mathbf{a} \cdot \mathbf{e} = \mathbf{a} \cdot (\mathbf{v} - \lambda \mathbf{a}) = \mathbf{a} \cdot \mathbf{v} - \lambda \mathbf{a} \cdot \mathbf{a}$$

y por tanto $\boxed{\lambda = \dfrac{\mathbf{a}^T \mathbf{v}}{\mathbf{a}^T \mathbf{a}} \text{ y } \mathbf{p_W} = \dfrac{\mathbf{a}^T \mathbf{v}}{\mathbf{a}^T \mathbf{a}} \mathbf{a}}$

Ejercicio 8 Calcular la proyección del vector $\mathbf{v} = (1, 1, 1)^T$ sobre la recta cuyo vector director es $\mathbf{a} = (1, 2, 2)^T$.

Si W es un subespacio vectorial generado por los vectores $\{\mathbf{a_1}, \dots, \mathbf{a_p}\}$ linealmente independientes para proyectar $\mathbf{v}$ sobre W usamos la misma idea que en dimensión 1. Buscamos

$$\mathbf{p_W} = [\mathbf{a_1}| \dots |\mathbf{a_p}]_{p \times n} \mathbf{x} = A\mathbf{x}.$$

Utilizando que $A^T \cdot \mathbf{e} = 0$ concluimos $\mathbf{x}$ tiene que ser solución del sistema $\boxed{A^T \mathbf{v} = A^T A\mathbf{x}}$ y entonces la proyección es $\boxed{\mathbf{p_W} = A(A^T A)^{-1} A^T \mathbf{v}}$.

Ejercicio 9 Dados el vector $\mathbf{v} = (6, 0, 0)^T$ y el subespacio $W = \{(1, 1, 1)^T, (0, 1, 2)^T\}$. Calcular la proyección de $\mathbf{v}$ sobre W.

Importante.

- Teorema de la descomposición ortogonal
  Sea W un subespacio de $\mathbb{R}^n$. Todo vector $\mathbf{v} \in \mathbb{R}^n$ se escribe de manera única como suma de dos vectores
  $$\mathbf{v} = \mathbf{p_W} + \mathbf{e},$$
  tales que $\mathbf{p_W} \in W$ y $\mathbf{e} \in W^\perp$.
- Distancia mínima
  $\mathbf{p_W}$ es el elemento de W que más cerca está de $\mathbf{v}$:
  $$\|\mathbf{p_W} - \mathbf{v}\| \leq \|\mathbf{p_W} - \mathbf{b}\|,$$
  para todo $\mathbf{b} \in W$.

Ejercicio 10 Encontrar la distancia del vector $\mathbf{v} = (-1, -5, 10)^T$ al subespacio W generado por $(5, -2, 1)^T, (1, 2, -1)^T$.

3 *Problemas de mínimos cuadrados*

Idea. En muchos problemas prácticos $A\mathbf{x} = \mathbf{b}$ no tiene solución. Hay demasiadas ecuaciones. Sin embargo necesitamos encontrar una "solución", de forma que $A\mathbf{x}$ se parezca lo más posible a $\mathbf{b}$.

Sea $\mathbf{e} = \mathbf{b} - A\mathbf{x}$; como el sistema no tiene solución $\mathbf{e} \neq \mathbf{0}$, y buscamos $\mathbf{x}$ tal que $\mathbf{e}$ sea lo más pequeño posible.

Definición 6 Dada A una matriz $m \times n$ y $\mathbf{b} \in \mathbb{R}^n$. Una solución de *mínimos cuadrados* de $A\mathbf{x} = \mathbf{b}$ es $\mathbf{y} \in \mathbb{R}^n$ tal que

$$\|\mathbf{b} - A\mathbf{y}\| \leq \|\mathbf{b} - A\mathbf{x}\|, \qquad \text{para todo } \mathbf{x} \in \mathbb{R}^n.$$

¿Cómo calculamos la solución de mínimos cuadrados?

> **Idea.** A es rectangular y $A\mathbf{x} = \mathbf{b}$ no tiene solución. Multiplicar ambos lados de la ecuación por A^T,
> $$A^T A\mathbf{x} = A^T\mathbf{b}.$$
> Es un nuevo sistema, que sí tiene solución. La solución de este sistema, es la solución de mínimos cuadrados del problema original.

Ejemplo 11 Dados los puntos $(0,6),(1,0)$ y $(2,0)$ encontrar la recta que esté más cerca de los tres puntos.

Buscamos $y = mx + n$ con incógnitas m y n. El sistema que tenemos que resolver se escribe en forma matricial como

$$\left(\begin{array}{cc|c} 1 & 0 & 6 \\ 1 & 1 & 0 \\ 1 & 2 & 0 \end{array}\right)$$

Es fácil ver que este sistema es incompatible.

Buscamos una solución de mínimos cuadrados. Para ello calculamos

$$A^T A = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 3 & 3 \\ 3 & 5 \end{pmatrix}, \qquad A^T\mathbf{b} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{pmatrix}\begin{pmatrix} 6 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 6 \\ 0 \end{pmatrix}$$

y resolvemos el sistema

$$\begin{pmatrix} 3 & 3 \\ 3 & 5 \end{pmatrix}\begin{pmatrix} n \\ m \end{pmatrix}\begin{pmatrix} 6 \\ 0 \end{pmatrix}.$$

La solución de mínimos cuadrados es $m = -3$ y $n = 5$, luego la recta que mejor aproxima a los tres puntos es $y = -3x + 5$.

4 *El proceso de ortogonalización de Gram-Schmidt*

Un problema de mínimos cuadrados es más fácil de resolver si $A^T A$ es diagonal.

Definición 7 Los vectores $\{\mathbf{q_1}, \ldots \mathbf{q_n}\}$ son

- *ortogonales* si $\mathbf{q_i} \cdot \mathbf{q_j} = 0$, para $i \neq j$.
- *ortonormales* si son ortogonales y $\mathbf{q_i} \cdot \mathbf{q_i} = 1$ para todo i.

Definición 8 Una matriz Q con columnas ortonormales entre sí se llama *ortogonal*

> **Propiedad.** Si Q es cuadrada y ortogonal
> $$Q^T = Q^{-1}$$

Dada una base $\{\mathbf{a_1}, \ldots, \mathbf{a_n}\}$ de un subespacio, queremos encontrar otra $\{\mathbf{q_1}, \ldots, \mathbf{q_n}\}$ base de ese mismo espacio cuyos elementos sean ortogonales (o incluso ortonormales). La ventaja que tiene el trabajar con bases ortonormales es que, si bien el espacio generado por ambas bases el mismo, las cuentas se simplifican si utilizamos $Q = [\mathbf{q_1}|\ldots|\mathbf{q_n}]$ (Recordar que $Q^T Q = I$).

> **Proceso de Gram-Schmidt.** Dados n vectores linealmente independientes $\{\mathbf{a_1}, \ldots, \mathbf{a_n}\}$ de un subespacio vectorial, construimos una base ortogonal de vectores $\{\mathbf{w_1}, \ldots, \mathbf{w_n}\}$ a partir de la base original. La base ortonormal $\{\mathbf{q_1}, \ldots, \mathbf{q_n}\}$ se obtiene dividiendo cada $\mathbf{w_i}$ por su norma.

Paso i: Elegimos $\boxed{\mathbf{w_1} = \mathbf{a_1}.}$

Paso ii: El vector $\mathbf{w_2}$ tiene que ser ortogonal a $\mathbf{w_1}$. Por el teorema de de descomposición ortogonal sabemos que

$$\mathbf{a_2} = \mathbf{p_W} + \mathbf{e},$$

donde $\mathbf{p_W}$ es la proyección de $\mathbf{a_2}$ sobre el subespacio W y $\mathbf{e} \in W^{\perp}$. Sea en este caso $W = \text{Gen}\{\mathbf{w_1}\}$. Entonces $\mathbf{e}$ y $\mathbf{w_1}$ son ortogonales. Por tanto tomamos $\mathbf{e} = \mathbf{w_2}$ y así

$$\boxed{\mathbf{w_2} = \mathbf{a_2} - \mathbf{p_W} = \mathbf{a_2} - \frac{\mathbf{w_1}^T\mathbf{a_2}}{\mathbf{w_1}^T\mathbf{w_1}}\mathbf{w_1}.}$$

Paso iii: Usamos la misma idea que en el paso anterior. El vector $\mathbf{w_3}$ tiene que ser ortogonal a $\mathbf{w_1}$ y $\mathbf{w_2}$. De nuevo, por el teorema de de descomposición ortogonal sabemos que

$$\mathbf{a_3} = \mathbf{p_W} + \mathbf{e},$$

donde $\mathbf{p_W}$ es la proyección de $\mathbf{a_2}$ sobre el subespacio W y $\mathbf{e} \in W^{\perp}$. Definamos $W = \text{Gen}\{\mathbf{w_1}, \mathbf{w_2}\}$. Entonces $\mathbf{e}$ es ortogonal a $\mathbf{w_1}$ y $\mathbf{w_2}$, (que a su vez son ortogonales entre sí). Tomamos $\mathbf{e} = \mathbf{w_3}$ y así, como estamos proyectando sobre un espacio generado por un conjunto ortogonal $\{\mathbf{w_1}, \mathbf{w_2}\}$ tenemos

$$\boxed{\mathbf{w_3} = \mathbf{a_3} - \mathbf{p_W} = \mathbf{a_3} - \frac{\mathbf{w_1}^T\mathbf{a_3}}{\mathbf{w_1}^T\mathbf{w_1}}\mathbf{w_1} - \frac{\mathbf{w_2}^T\mathbf{a_3}}{\mathbf{w_2}^T\mathbf{w_2}}\mathbf{w_2}.}$$

$$\vdots$$

Paso n: Para calcular $\mathbf{w_n}$ proyectamos $\mathbf{a_n}$ sobre el espacio generado por $\{\mathbf{w_1}, \dots \mathbf{w_{n-1}}\}$, que por la construcción que hemos hecho, son ortogonales entre sí. Entonces

$$\boxed{\mathbf{w_n} = \mathbf{a_n} - \mathbf{p_W} = \mathbf{a_n} - \frac{\mathbf{w_1}^T\mathbf{a_n}}{\mathbf{w_1}^T\mathbf{w_1}}\mathbf{w_1} - \frac{\mathbf{w_2}^T\mathbf{a_n}}{\mathbf{w_2}^T\mathbf{w_2}}\mathbf{w_2} - \cdots - \frac{\mathbf{w_{n-1}}^T\mathbf{a_n}}{\mathbf{w_{n-1}}^T\mathbf{w_{n-1}}}\mathbf{w_{n-1}}.}$$

De esta forma obtenemos una base ortogonal $\{\mathbf{w_1}, \dots, \mathbf{w_n}\}$ y tal que $\text{Gen}\{\mathbf{a_1}, \dots, \mathbf{a_n}\} = \text{Gen}\{\mathbf{w_1}, \dots, \mathbf{w_n}\}$. Para conseguir una base ortonormal basta con dividir cada uno de los vectores de la base ortogonal por su norma; es decir

$$\boxed{\{\mathbf{q_1}, \dots, \mathbf{q_n}\} = \left\{\frac{\mathbf{w_1}}{\|\mathbf{w_1}\|}, \dots, \frac{\mathbf{w_n}}{\|\mathbf{w_n}\|}\right\}.}$$

5 *Diagonalización de matrices simétricas*

Recordatorio.

- A es simétrica si $A^T = A$.
- A es diagonalizable si existen matrices D diagonal y P invertible tal que

$$A = PDP^{-1}$$

- Si Q es cuadrada y ortogonal $Q^{-1} = Q^T$

Ejemplo 12 La matriz

$$A = \begin{pmatrix} 6 & -2 & -1 \\ -2 & 6 & -1 \\ -1 & -1 & 4 \end{pmatrix}$$

tiene por valores propios $\lambda = 8, 6, 3$ y por vectores propios

$$\mathbf{v_1} = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \quad \mathbf{v_2} = \begin{pmatrix} -1 \\ -1 \\ 2 \end{pmatrix}, \quad \mathbf{v_3} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

que son ortogonales entre sí. Si $\mathbf{v_1}, \mathbf{v_2}, \mathbf{v_3}$ son vectores propios

$$\mathbf{w_1} = \frac{\mathbf{v_1}}{\|\mathbf{v_1}\|}, \quad \mathbf{w_2} = \frac{\mathbf{v_2}}{\|\mathbf{v_2}\|}, \quad \mathbf{w_3} = \frac{\mathbf{v_3}}{\|\mathbf{v_3}\|},$$

también lo son y A se puede diagonalizar como

$$A = [\mathbf{w_1}|\mathbf{w_2}|\mathbf{w_3}]D[\mathbf{w_1}|\mathbf{w_2}|\mathbf{w_3}]^{-1} = [\mathbf{w_1}|\mathbf{w_2}|\mathbf{w_3}]D[\mathbf{w_1}|\mathbf{w_2}|\mathbf{w_3}]^T.$$

¿Cuándo se puede hacer este tipo de diagonalización (*diagonalización ortogonal*)?

Importante. Dada A simétrica. Entonces

- A es siempre diagonalizable
- Vectores propios de espacios propios diferentes son siempre ortogonales
- Los vectores propios se pueden escoger todos ortogonales entre sí (con el proceso de Gram-Schmidt, cuando sea necesario)

Es más:

- A es diagonalizable ortogonalmente $\Longleftrightarrow$ A es simétrica.

Ejercicio 13 Diagonalizar ortogonalmente la siguiente matriz

$$A = \begin{pmatrix} 5 & -4 & -2 \\ -4 & 5 & 2 \\ -2 & 2 & 2 \end{pmatrix}.$$