Producto escalar

Vectores ortogonales y proyecciones

La definición matemática de *producto escalar* es bastante amplia porque recoge toda expresión bilineal que sirva razonablemente para medir ángulos y distancias, aunque sea de manera muy distinta a la habitual. En este curso, sin embargo, nos fijamos sólo en el producto escalar que se ha utilizado en cursos anteriores, extendida a $\mathbb{R}^n$:

$$\vec{x} = (x_1, x_2, \dots, x_n), \quad \vec{y} = (y_1, y_2, \dots, y_n) \qquad \Longrightarrow \qquad \vec{x} \cdot \vec{y} = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n.$$

Al multiplicar escalarmente un vector por sí mismo obtenemos su *longitud* (*norma*, *módulo*) al cuadrado, y el ángulo $0 \le \alpha \le \pi$ entre dos vectores no nulos $\vec{x}$, $\vec{y}$ viene determinado por una conocida fórmula para $\cos\alpha$:

$$|\vec{x}| = \sqrt{\vec{x} \cdot \vec{x}} \qquad \text{y} \qquad \cos\alpha = \frac{\vec{x} \cdot \vec{y}}{|\vec{x}|\,|\vec{y}|}.$$

Ejemplo. Los vectores $\vec{x} = (2, -2)$ y $\vec{y} = (3, 0)$ tienen longitudes $2\sqrt{2}$ y 3 respectivamente y su ángulo cumple $\cos\alpha = 1/\sqrt{2}$, por tanto es $\pi/4$ (esto son 45°), como se ve claro en un dibujo.

El caso más importante en cuanto a ángulo entre vectores es el de perpendicularidad que corresponde a $\vec{x} \cdot \vec{y} = 0$. En este caso se dice que los vectores son *ortogonales*. Cuando tenemos más vectores, se dice que son ortogonales si lo son dos a dos. Y se dice que son *ortonormales* si además son *unitarios*, es decir, de longitud uno.

Ejemplo. Los vectores $\vec{a} = (1, -8, -4)$, $\vec{b} = (-4, -4, 7)$, $\vec{c} = (-8, 1, -4)$ son ortogonales porque $\vec{a} \cdot \vec{b} = \vec{a} \cdot \vec{c} = \vec{b} \cdot \vec{c} = 0$ pero no son ortonormales porque $|\vec{a}| = 9$. De hecho todos tienen longitud 9 y entonces $\vec{a}/9$, $\vec{b}/9$ y $\vec{c}/9$, sí son ortonormales.

Dado un vector $\vec{v}$ muchas veces surge de manera natural el problema de encontrar un vector en un subespacio V que esté lo más cerca posible de $\vec{v}$. A este vector se le llama *proyección ortogonal* de $\vec{v}$ sobre V y lo denotaremos mediante $\text{Proy}_V(\vec{v})$. Si $\{\vec{x}_1, \vec{x}_2, \dots, \vec{x}_n\}$ es una base ortogonal de V, viene dado por la fórmula:

$$\text{(1)} \qquad \text{Proy}_V(\vec{v}) = \frac{\vec{v} \cdot \vec{x}_1}{|\vec{x}_1|^2}\,\vec{x}_1 + \frac{\vec{v} \cdot \vec{x}_2}{|\vec{x}_2|^2}\,\vec{x}_2 + \cdots + \frac{\vec{v} \cdot \vec{x}_n}{|\vec{x}_n|^2}\,\vec{x}_n.$$

La proyección ortogonal aparece en muchas aplicaciones en la forma de la mejor aproximación, como se ha presentado aquí, sin embargo, como el nombre sugiere, tiene una interpretación geométrica: $\text{Proy}_V(\vec{v})$ es el vector al que llegamos cuando vamos del extremo de $\vec{v}$ y nos dirigimos en perpendicular hacia V. Esta interpretación geométrica nos indica que siempre se cumple que $\vec{v} - \text{Proy}_V(\vec{v})$ y $\text{Proy}_V(\vec{v})$ son ortogonales y por tanto en los problemas de proyecciones ortogonales es aconsejable comprobar al final

$$\big(\vec{v} - \text{Proy}_V(\vec{v})\big) \cdot \text{Proy}_V(\vec{v}) = 0$$

para asegurarse de que la solución es correcta.

Ejemplo. Consideremos el subespacio de $\mathbb{R}^4$ que tiene por base $\{\vec{u}_1, \vec{u}_2, \vec{u}_3\}$ con

$$\vec{u}_1 = (1,0,0,0), \qquad \vec{u}_2 = (0,2,0,-1), \qquad \vec{u}_3 = (0,1,5,2).$$

Es una base ortogonal porque $\vec{u}_1 \cdot \vec{u}_2 = \vec{u}_1 \cdot \vec{u}_3 = \vec{u}_2 \cdot \vec{u}_3 = 0$. Veamos cuál es la proyección de $\vec{v} = (1,1,1,1)$ sobre el subespacio generado por ellos. Basta aplicar la fórmula

$$\begin{aligned} \text{Proy}_V(\vec{v}) &= \frac{\vec{v}\cdot\vec{u}_1}{|\vec{u}_1|^2}\vec{u}_1 + \frac{\vec{v}\cdot\vec{u}_2}{|\vec{u}_2|^2}\vec{u}_2 + \frac{\vec{v}\cdot\vec{u}_3}{|\vec{u}_3|^2}\vec{u}_3 \\ &= \frac{1}{1}(1,0,0,0) + \frac{1}{5}(0,2,0,-1) + \frac{8}{30}(0,1,5,2). \end{aligned}$$

El resultado es entonces $(1, 2/3, 4/3, 1/3)$. A modo de comprobación se puede notar que este vector es ortogonal a $\vec{v} - \text{Proy}_V(\vec{v}) = (0, 1/3, -1/3, 2/3)$.

Es importante hacer hincapié en que la fórmula (1) sólo es válida si $\{\vec{x}_1, \vec{x}_2, \dots, \vec{x}_n\}$ es una base ortogonal del subespacio V. Entonces surge el problema de transformar una base que no es ortogonal en otra que sí lo es. El *proceso de ortogonalización de Gram-Schmidt* es un algoritmo con este propósito.

Digamos que $\{\vec{b}_1, \vec{b}_2, \dots, \vec{b}_n\}$ es una base no ortogonal de un subespacio vectorial V. El proceso de Gram-Schmidt produce una base ortogonal $\{\vec{x}_1, \vec{x}_2, \dots, \vec{x}_n\}$ de V a través de las fórmulas

$$\vec{x}_1 = \vec{b}_1, \qquad \vec{x}_i = \vec{b}_i - \sum_{j=1}^{i-1} \frac{\vec{b}_i \cdot \vec{x}_j}{\|\vec{x}_j\|^2}\,\vec{x}_j, \quad i = 2,3,\dots,n.$$

De una forma más fácil de recordar, se puede interpretar el algoritmo teniendo en cuenta (1) y diciendo que a cada $\vec{b}_i$ se le resta su proyección ortogonal sobre $V_{i-1} = \mathcal{L}\big(\{\vec{x}_1, \dots, \vec{x}_{i-1}\}\big)$. Esto es:

$$\vec{x}_1 = \vec{b}_1, \qquad \vec{x}_i = \vec{b}_i - \text{Proy}_{V_{i-1}}(\vec{b}_i), \quad i = 2,3,\dots,n.$$

Ejemplo. Apliquemos el proceso de Gram-Schmidt para ortogonalizar los vectores

$$\vec{u}_1 = (0,1,-1), \qquad \vec{u}_2 = (1,-2,0), \qquad \vec{u}_3 = (4,-1,-1).$$

El primer vector se deja inalterado: $\vec{x}_1 = \vec{u}_1 = (0,1,-1)$. Al segundo se le resta su proyección sobre $\vec{x}_1$

$$\vec{x}_2 = \vec{u}_2 - \frac{\vec{u}_2\cdot\vec{x}_1}{|\vec{x}_1|^2}\vec{x}_1 = (1,-2,0) - \frac{-2}{2}(0,1,-1) = (1,-1,-1).$$

Y, finalmente, al tercero, se le resta su proyección sobre el subespacio que generan $\vec{x}_1$ y $\vec{x}_2$

$$\vec{x}_3 = \vec{u}_3 - \frac{\vec{u}_3 \cdot \vec{x}_1}{|\vec{x}_1|^2}\vec{x}_1 - \frac{\vec{u}_3 \cdot \vec{x}_2}{|\vec{x}_2|^2}\vec{x}_2 = (4, -1, -1) - \frac{0}{2}(0, 1, -1) - \frac{6}{3}(1, -1, -1) = (2, 1, 1).$$

Es muy fácil comprobar que realmente $(0, 1, -1)$, $(1, -1, -1)$ y $(2, 1, 1)$ son ortogonales.

En resumen, cuando queramos hallar $\text{Proy}_V(\vec{v})$ tenemos que hallar primero una base de V, después ortogonalizarla y finalmente aplicar la fórmula (1).

Ejemplo. Para calcular la proyección ortogonal del vector $\vec{v} = (-2, -2, 3)$ sobre el subespacio $V = \{(x, y, z) \in \mathbb{R}^3 \ : \ x - y + 2z = 0\}$, hallamos primero una base de V. Esto pasa por resolver la ecuación $x - y + 2z = 0$. Su solución se escribe como

$$(x, y, z) = \lambda(1, 1, 0) + \mu(-2, 0, 1)$$

y $\vec{b}_1 = (1, 1, 0)$, $\vec{b}_2 = (-2, 0, 1)$, forman una base. Al ortogonalizarla con el proceso de Gram-Schmidt se obtiene:

$$\vec{x}_1 = (1, 1, 0), \qquad \vec{x}_2 = (-2, 0, 1) - \frac{\vec{b}_2 \cdot \vec{x}_1}{|\vec{x}_1|^2}\vec{x}_1 = (-1, 1, 1).$$

Finalmente aplicamos la fórmula (1) con la base $\{\vec{x}_1, \vec{x}_2\}$,

$$\text{Proy}_V(\vec{v}) = \frac{\vec{v} \cdot \vec{x}_1}{|\vec{x}_1|^2}\,\vec{x}_1 + \frac{\vec{v} \cdot \vec{x}_2}{|\vec{x}_2|^2}\,\vec{x}_2 = -\frac{4}{2}(1, 1, 0) + \frac{3}{3}(-1, 1, 1) = (-3, -1, 1).$$

La comprobación es ver que éste vector es ortogonal a $\vec{v} - \text{Proy}_V(\vec{v}) = (1, -1, 2)$.

Matrices ortogonales y simétricas

Entre las aplicaciones lineales hay algunas destacadas que preservan el producto escalar y por tanto ángulos y distancias. Corresponden a giros y simetrías y su matriz se dice que es una *matriz ortogonal*.

Una definición más sencilla y operativa es que una matriz A es ortogonal si verifica

$$AA^t = I,$$

donde A^t es lo que se llama la *matriz traspuesta* de A que es la propia A intercambiando filas y columnas (haciendo una simetría por la diagonal). Esto significa que $A^{-1} = A^t$, así que para las matrices ortogonales invertir es tan fácil como trasponer.

Otra definición equivalente, aunque menos útil, es que una matriz es ortogonal si sus columnas forman una base ortonormal.

Ejemplo. La matriz

$$A = \begin{pmatrix} 3/5 & -4/5 \\ 4/5 & 3/5 \end{pmatrix} \qquad \text{verifica} \qquad \begin{pmatrix} 3/5 & -4/5 \\ 4/5 & 3/5 \end{pmatrix} \begin{pmatrix} 3/5 & 4/5 \\ -4/5 & 3/5 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

por tanto es ortogonal. También es fácil ver que las columnas son vectores ortonormales.

Un importante teorema afirma que una matriz simétrica diagonaliza sobre $\mathbb{R}$ en una base ortonormal. Esto significa que todos los autovalores reales y que los autovectores se pueden escoger de forma que sean ortonormales. En ese caso, la matriz P que aparece al diagonalizar, es ortogonal.

Ejemplo. Vamos a hallar una base ortonormal en la que diagonalice la matriz simétrica

$$A = \begin{pmatrix} -2 & 3 \\ 3 & 6 \end{pmatrix}.$$

Resolviendo $|A - \lambda I| = 0$ se obtienen los autovalores $\lambda_1 = 7$ y $\lambda_2 = -3$. Para hallar los autovectores hay que resolver, respectivamente,

$$\begin{pmatrix} -2-7 & 3 \\ 3 & 6-7 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \qquad \text{y} \qquad \begin{pmatrix} -2+3 & 3 \\ 3 & 6+3 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Posibles soluciones son $(1,3)$ y $(3,-1)$. Como esperábamos, son ortogonales. Para tener una base ortonormal de autovectores (la base ortonormal en la que diagonaliza A) basta dividir entre la norma. Entonces, ésta es $\big\{(1/\sqrt{10}, 3/\sqrt{10}), (3/\sqrt{10}, -1/\sqrt{10})\big\}$. Al ponerlos en columnas obtenemos una matriz P que es ortogonal y que sirve para diagonalizar:

$$A = P \begin{pmatrix} 7 & 0 \\ 0 & -3 \end{pmatrix} P^{-1} \qquad \text{con} \qquad P = \frac{1}{\sqrt{10}} \begin{pmatrix} 1 & 3 \\ 3 & -1 \end{pmatrix}.$$

Gracias al teorema antes mencionado, siempre los autovectores que correspondan a autovalores distintos serán ortogonales y en general basta dividir entre la norma para obtener la base ortonormal. La única excepción es que si hay autovalores múltiples habrá varios autovectores independientes correspondientes a cada uno de ellos y hay que preocuparse de escogerlos ortogonales.

Ejemplo. La matriz

$$A = \begin{pmatrix} 3 & 2 & \sqrt{2} \\ 2 & 3 & \sqrt{2} \\ \sqrt{2} & \sqrt{2} & 2 \end{pmatrix}$$

es simétrica y por tanto debe haber una base ortonormal de autovectores en la que diagonaliza. Tras unos cálculos se obtiene $|A - \lambda I| = (\lambda - 1)^2(6 - \lambda)$, entonces los valores propios son $\lambda_1 = 1$ (doble) y $\lambda_2 = 6$.

Al resolver $(A - I)\vec{x} = \vec{0}$ por Gauss de la manera habitual, se sigue que los autovectores correspondientes a $\lambda_1 = 1$ son combinaciones lineales de $\vec{v}_1 = (-1, 1, 0)$ y de

$\vec{v}_2 = (-1/\sqrt{2}, 0, 1)$. Por otro lado, resolviendo $(A - 6I)\vec{x} = \vec{0}$, los autovectores correspondientes a $\lambda_2 = 6$ son proporcionales a $\vec{v}_3 = (\sqrt{2}, \sqrt{2}, 1)$. Los vectores $\vec{v}_1$ y $\vec{v}_2$ son ortogonales a $\vec{v}_3$ pero no entre sí. Al ortogonalizar los dos primeros con el proceso de Gram-Schmidt hay reemplazar el segundo por

$$\vec{v}_2' = (-1/\sqrt{2}, 0, 1) - \frac{1}{2\sqrt{2}}(-1, 1, 0) = \big(- \frac{1}{2\sqrt{2}}, -\frac{1}{2\sqrt{2}}, 1\big).$$

Finalmente dividimos por la norma para hacer que estos vectores ortogonales pasen a ser ortonormales. Esto es, la base ortonormal de autovectores es $\{\vec{u}_1, \vec{u}_2, \vec{u}_3\}$ con $\vec{u}_1 = \vec{v}_1/|\vec{v}_1|$, $\vec{u}_2 = \vec{v}_2'/|\vec{v}_2'|$, $\vec{u}_3 = \vec{v}_3/|\vec{v}_3|$. Estos vectores escritos en columnas dan lugar a la matriz ortogonal

$$P = \begin{pmatrix} -1/\sqrt{2} & -1/\sqrt{10} & \sqrt{2/5} \\ 1/\sqrt{2} & -1/\sqrt{10} & \sqrt{2/5} \\ 0 & \sqrt{4/5} & 1/\sqrt{5} \end{pmatrix}.$$

La cual verifica

$$A = P \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 6 \end{pmatrix} P^{-1} = P \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 6 \end{pmatrix} P^t.$$

Referencias. Hay muchos libros de álgebra lineal y casi todos tienen contenidos parecidos. Uno con muchos ejemplos y buenas explicaciones es [HVZ12]. Una faceta del álgebra lineal, en la que desafortunadamente no incidimos en este curso, es la cantidad de aplicaciones que tiene. Éstas aplicaciones están en gran medida sustentadas por la posibilidad de programar eficientemente muchos cálculos de álgebra lineal. Un libro que cubre las aplicaciones y los cálculos numéricos es [Str80]. Por otro lado, [Gol86] satisfará a los que tengan interés en la interpretación geométrica y física del álgebra lineal, aunque quizá no sea fácil de encontrar. Por último, para los estudiantes muy avanzados, [Lax97] es un libro escrito por un matemático de primera línea que constituye una excepción a la uniformidad de temas de los libros de álgebra lineal.

Referencias

[Gol86] L. I. Golovina. *Algebra Lineal y Algunas de sus Aplicaciones.* Mir, 1986.

[HVZ12] E. Hernández, M.J. Vázquez, and M.A. Zurro. *Álgebra lineal y Geometría.* Pearson/Addison Wesley, tercera edition, 2012.

[Lax97] P. D. Lax. *Linear algebra.* Pure and Applied Mathematics (New York). John Wiley & Sons, Inc., New York, 1997. A Wiley-Interscience Publication.

[Str80] G. Strang. *Linear algebra and its applications.* Academic Press, New York-London, second edition, 1980.