Geometría analítica

3.1 Normas

Cuando pensamos en vectores geométricos, es decir, segmentos de línea dirigidos que comienzan en el origen, intuitivamente la longitud de un vector es la distancia desde el "extremo" de ese segmento de línea dirigido hasta el origen. A continuación, discutiremos la noción de la longitud de los vectores usando el concepto de una norma.

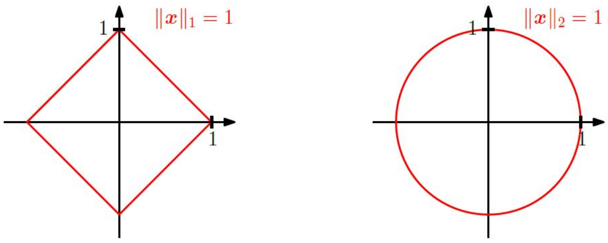


Figura 1: Para diferentes normas, las líneas rojas indican el conjunto de vectores con norma 1. Izquierda: norma Manhattan; Derecha: distancia Euclidiana.

Definición 3.1 (Norma). Una norma en un espacio vectorial V es una función

$$\|\cdot\| : V \to \mathbb{R}, \tag{3.1}$$

$$x \mapsto \|x\|, \tag{3.2}$$

que asigna a cada vector x su longitud $\|x\| \in \mathbb{R}$, de forma que para todo $\lambda \in \mathbb{R}$ y $x, y \in V$, se cumplen las siguientes propiedades:

- **Homogeneidad absoluta:** $\|\lambda x\| = |\lambda|\|x\|$
- **Desigualdad triangular:** $\|x + y\| \leq \|x\| + \|y\|$
- **Definida positiva:** $\|x\| \geq 0$ y $\|x\| = 0 \iff x = 0$

En términos geométricos, la desigualdad triangular establece que, para cualquier triángulo, la suma de las longitudes de dos de sus lados siempre debe ser mayor o igual que la longitud del lado restante; ver la Figura 3.2 como ilustración.

La Definición 3.1 se expresa en términos de un espacio vectorial general V (Sección 2.4), pero en este libro solo consideraremos un espacio vectorial de dimensión finita $\mathbb{R}^n$.

Recuerda que para un vector $x \in \mathbb{R}^n$, denotamos los elementos del vector usando un subíndice, es decir, x_i es el i-ésimo elemento del vector x.

Ejemplo 3.1 (Norma de Manhattan)
La norma de Manhattan en $\mathbb{R}^n$ se define para $x \in \mathbb{R}^n$ como

$$\|x\|_1 := \sum_{i=1}^{n} |x_i|, \tag{3.3}$$

donde $|\cdot|$ representa el valor absoluto. El panel izquierdo de la Figura 3.3 muestra todos los vectores $x \in \mathbb{R}^2$ con $\|x\|_1 = 1$. La norma de Manhattan también se conoce como norma ℓ_1.

Ejemplo 3.2 (Norma Euclidiana)
La norma Euclidiana de $x \in \mathbb{R}^n$ se define como

$$\|x\|_2 := \sqrt{\sum_{i=1}^{n} x_i^2} = \sqrt{x^\top x} \tag{3.4}$$

y calcula la distancia Euclidiana de x al origen. El panel derecho de la Figura 3.3 muestra todos los vectores $x \in \mathbb{R}^2$ con $\|x\|_2 = 1$. La norma Euclidiana también se conoce como norma ℓ_2.

3.2 Productos Internos

Los productos internos permiten la introducción de conceptos geométricos intuitivos, como la longitud de un vector y el ángulo o la distancia entre dos vectores. Un propósito principal de los productos internos es determinar si dos vectores son ortogonales entre sí.

3.2.1 Producto Punto

Es posible que ya estemos familiarizados con un tipo particular de producto interno, el producto escalar o producto punto en $\mathbb{R}^n$, que se define como

$$x^\top y = \sum_{i=1}^{n} x_i y_i. \tag{3.5}$$

Nos referiremos a este producto interno particular como el producto punto en este libro. Sin embargo, los productos internos son conceptos más generales con propiedades específicas, que ahora introduciremos.

3.2.2 Productos Internos Generales

Recordemos la aplicación lineal de la Sección 2.7, donde podemos reorganizar dicha aplicación respecto a la suma y a la multiplicación por un escalar. Una aplicación bilineal Ω es una aplicación con dos argumentos, y es lineal en cada argumento. Es decir, cuando trabajamos en un espacio vectorial V, se cumple que para todos $x, y, z \in V$ y $\lambda, \psi \in \mathbb{R}$:

$$\Omega(\lambda x + \psi y, z) = \lambda\Omega(x, z) + \psi\Omega(y, z) \tag{3.6}$$

$$\Omega(x, \lambda y + \psi z) = \lambda\Omega(x, y) + \psi\Omega(x, z) \tag{3.7}$$

Aquí, la ecuación (3.6) afirma que Ω es lineal en el primer argumento, y la ecuación (3.7) afirma que Ω es lineal en el segundo argumento (ver también (2.87)). **Definición 3.2.** Sea V un espacio vectorial y $\Omega : V \times V \to \mathbb{R}$ una aplicación bilineal que toma dos vectores y los mapea en un número real. Entonces:

- Ω se llama simétrica si $\Omega(x, y) = \Omega(y, x)$ para todo $x, y \in V$, es decir, el orden de los argumentos no importa.
- Ω se llama positiva definida si
$$\forall x \in V \setminus \{0\} : \Omega(x, x) > 0, \quad \Omega(0, 0) = 0. \tag{3.8}$$

Definición 3.3. Sea V un espacio vectorial y $\Omega : V \times V \to \mathbb{R}$ una aplicación bilineal que toma dos vectores y los mapea en un número real. Entonces:

- Una aplicación bilineal, simétrica y positiva definida $\Omega : V \times V \to \mathbb{R}$ se llama un producto interno en V. Típicamente escribimos $\langle x, y\rangle$ en lugar de $\Omega(x, y)$.
- El par $(V, \langle\cdot, \cdot\rangle)$ se llama un espacio con producto interno (o espacio vectorial real con producto interno). Si usamos el producto punto definido en (3.5), llamamos a $(V, \langle\cdot, \cdot\rangle)$ un espacio euclidiano.

Nos referiremos a estos espacios como espacios con producto interno en este libro.

Ejemplo 3.3 (Producto interno que no es el producto punto)
Consideremos $V = \mathbb{R}^2$. Si definimos

$$\langle x, y\rangle := x_1y_1 - (x_1y_2 + x_2y_1) + 2x_2y_2 \tag{3.9}$$

entonces $\langle \cdot, \cdot\rangle$ es un producto interno, pero distinto del producto punto. La demostración se dejará como un ejercicio.

Demostración. Verificaremos que la expresión

$$\langle x, y\rangle := x_1y_1 - (x_1y_2 + x_2y_1) + 2x_2y_2$$

define un producto interno en $\mathbb{R}^2$ cumpliendo las tres propiedades:

- **Linealidad en el primer argumento.** Sea $\alpha \in \mathbb{R}$ y $x, x', y \in \mathbb{R}^2$. Se tiene:

$$\langle \alpha x + x', y\rangle = \alpha\langle x, y\rangle + \langle x', y\rangle$$

comprobado mediante expansión directa.
- **Simetría.** Se verifica que:

$$\langle x, y\rangle = x_1y_1 - x_1y_2 - x_2y_1 + 2x_2y_2 = \langle y, x\rangle$$

por la conmutatividad del producto real.
- **Positividad definida.** Evaluando $\langle x, x\rangle$:

$$\langle x, x\rangle = x_1^2 - 2x_1x_2 + 2x_2^2 = (x_1 - x_2)^2 + x_2^2 \geq 0$$

y es cero si y solo si $x_1 = x_2 = 0$, es decir, $x = 0$.

Por lo tanto, $\langle \cdot, \cdot\rangle$ define un producto interno válido, aunque distinto del producto punto.

3.2.3 Matrices Simétricas y Definidas Positivas

Las matrices simétricas y definidas positivas desempeñan un papel importante en el aprendizaje automático, y se definen a través del producto interno. En la Sección 4.3, volveremos a tratar las matrices simétricas y definidas positivas en el contexto de descomposiciones matriciales. La idea de matrices semidefinidas positivas simétricas es clave en la definición de núcleos (ver Sección 12.4).

Consideremos un espacio vectorial de dimensión n con un producto interno $\langle \cdot, \cdot\rangle : V \times V \to \mathbb{R}$ (ver Definición 3.3) y una base ordenada $B = (b_1, \ldots, b_n)$ de V. Recordando la Sección 2.6.1, cualquier vector $x, y \in V$ puede escribirse como combinaciones lineales de los vectores base, de modo que $x = \sum_{i=1}^n \psi_i b_i$ y $y = \sum_{j=1}^n \lambda_j b_j$ para escalares $\psi_i, \lambda_j \in \mathbb{R}$. Debido a la bilinealidad del producto interno, se tiene que para todo $x, y \in V$:

$$\langle x, y\rangle = \left\langle \sum_{i=1}^n \psi_i b_i, \sum_{j=1}^n \lambda_j b_j \right\rangle = \sum_{i=1}^n \sum_{j=1}^n \psi_i \langle b_i, b_j\rangle \lambda_j = \hat{x}^\top A \hat{y}, \tag{3.10}$$

donde $A_{ij} := \langle b_i, b_j\rangle$ y $\hat{x}, \hat{y}$ son las coordenadas de x e y con respecto a la base B. Esto implica que el producto interno $\langle \cdot, \cdot\rangle$ está determinado por la matriz A. La simetría del producto interno también implica que A es simétrica. Además, la positividad definida del producto interno implica que:

$$\forall x \in V \setminus \{0\} : x^\top A x > 0. \tag{3.11}$$

Definición 3.4 (Matriz Simétrica y Definida Positiva). Una matriz simétrica $A \in \mathbb{R}^{n\times n}$ que satisface la condición (3.11) se llama simétrica y definida positiva, o simplemente definida positiva. Si sólo se cumple $x^\top A x \geq 0$ en (3.11), entonces A se denomina simétrica y semidefinida positiva.

Ejemplo 3.4 (Matrices Simétricas y Definidas Positivas)

Consideremos las matrices

$$A_1 = \begin{bmatrix} 9 & 6 \\ 6 & 5 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 9 & 6 \\ 6 & 3 \end{bmatrix}. \tag{3.12}$$

A_1 es definida positiva porque es simétrica y

$$x^\top A_1 x = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 9 & 6 \\ 6 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{3.13a}$$

$$= 9x_1^2 + 12x_1x_2 + 5x_2^2 = (3x_1 + 2x_2)^2 + x_2^2 > 0 \tag{3.13b}$$

para todo $x \in V \setminus \{0\}$. En contraste, A_2 es simétrica pero no definida positiva, porque

$$x^\top A_2 x = 9x_1^2 + 12x_1x_2 + 3x_2^2 = (3x_1 + 2x_2)^2 - x_2^2$$

puede ser menor que cero. Por ejemplo, para $x = \begin{bmatrix} 2 \\ -3 \end{bmatrix}$.

Si $\mathbf{A} \in \mathbb{R}^{n\times n}$ es simétrica y definida positiva, entonces

$$\langle x, y \rangle = \hat{x}^\top \mathbf{A} \hat{y} \tag{3.14}$$

define un producto interno con respecto a una base ordenada B, donde $\hat{x}$ y $\hat{y}$ son las representaciones coordenadas de $x, y \in V$ con respecto a B.

Teorema 3.5. *Para un espacio vectorial V de dimensión finita sobre los reales y una base ordenada B de V, se cumple que $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ es un producto interno si y sólo si existe una matriz $\mathbf{A} \in \mathbb{R}^{n\times n}$ simétrica y definida positiva tal que*

$$\langle x, y \rangle = \hat{x}^\top \mathbf{A} \hat{y}. \tag{3.15}$$

Las siguientes propiedades se cumplen si $\mathbf{A} \in \mathbb{R}^{n\times n}$ es simétrica y definida positiva:

- El núcleo (espacio nulo) de $\mathbf{A}$ contiene únicamente al vector 0 porque $x^\top \mathbf{A} x > 0$ para todo $x \neq 0$. Esto implica que $\mathbf{A}x \neq 0$ si $x \neq 0$.
- Los elementos diagonales a_{ii} de $\mathbf{A}$ son positivos porque $a_{ii} = e_i^\top \mathbf{A} e_i > 0$, donde e_i es el i-ésimo vector de la base estándar en $\mathbb{R}^n$.

3.3 Longitudes y Distancias

En la Sección 3.1, ya discutimos normas que podemos usar para calcular la longitud de un vector. Los productos internos y las normas están estrechamente relacionados en el sentido de que todo producto interno induce una norma

$$\|x\| := \sqrt{\langle x, x \rangle} \tag{3.16}$$

de forma natural, lo cual nos permite calcular longitudes de vectores usando el producto interno. Sin embargo, no toda norma es inducida por un producto interno. La norma Manhattan (3.3) es un ejemplo de norma sin un producto interno correspondiente. A continuación, nos enfocaremos en normas que sí son inducidas por productos internos e introduciremos conceptos geométricos como longitudes, distancias y ángulos.

Observación (Desigualdad de Cauchy-Schwarz). Para un espacio con producto interno $(V, \langle \cdot, \cdot \rangle)$, la norma inducida $\| \cdot \|$ satisface la *desigualdad de Cauchy-Schwarz*:

$$|\langle x, y \rangle| \leq \|x\| \|y\|. \tag{3.17}$$

Ejemplo 3.5 (Longitudes de vectores usando productos internos)

En geometría, a menudo nos interesa la longitud de los vectores. Podemos usar ahora un producto interno para calcularla usando la ecuación (3.16). Tomemos $x = [1, 1]^\top \in \mathbb{R}^2$. Si usamos el producto punto como producto interno, con (3.16) obtenemos

$$\|x\| = \sqrt{x^\top x} = \sqrt{1^2 + 1^2} = \sqrt{2} \tag{3.18}$$

como la longitud de x. Ahora elijamos un producto interno diferente:

$$\langle x, y \rangle := x^\top \begin{bmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{bmatrix} y = x_1 y_1 - \frac{1}{2}(x_1 y_2 + x_2 y_1) + x_2 y_2. \tag{3.19}$$

Si calculamos la norma de un vector, entonces este producto interno devuelve valores más pequeños que el producto punto si x_1 y x_2 tienen el mismo signo (y $x_1 x_2 > 0$); de lo contrario, devuelve valores mayores que el producto punto. Con este producto interno obtenemos

$$\langle x, x \rangle = x_1^2 - x_1 x_2 + x_2^2 = 1 - 1 + 1 = 1 \quad \Longrightarrow \quad \|x\| = \sqrt{1} = 1, \tag{3.20}$$

de modo que x es "más corto" con este producto interno que con el producto punto.

Definición 3.6 (Distancia y métrica). Consideremos un espacio con producto interno $(V, \langle \cdot, \cdot \rangle)$. Entonces,

$$d(x, y) := \|x - y\| = \sqrt{\langle x - y, x - y \rangle} \tag{3.21}$$

se llama la *distancia* entre x y y para $x, y \in V$. Si usamos el producto punto como producto interno, entonces la distancia se llama *distancia euclidiana*. La aplicación

$$d : V \times V \to \mathbb{R} \tag{3.22}$$
$$(x, y) \mapsto d(x, y) \tag{3.23}$$

se llama una *métrica*.

Observación. De manera similar a la longitud de un vector, la distancia entre vectores no requiere un producto interno: una norma es suficiente. Si la norma proviene de un producto interno, la distancia puede variar dependiendo de la elección del producto interno. ♢

Una métrica d satisface lo siguiente:

1. d es *positiva definida*, es decir, $d(x, y) \geq 0$ para todo $x, y \in V$ y $d(x, y) = 0 \iff x = y$.
2. d es *simétrica*, es decir, $d(x, y) = d(y, x)$ para todo $x, y \in V$.
3. *Desigualdad triangular:* $d(x, z) \leq d(x, y) + d(y, z)$ para todo $x, y, z \in V$.

Observación. A primera vista, las listas de propiedades de productos internos y métricas parecen muy similares. Sin embargo, al comparar la Definición 3.3 con la Definición 3.6, observamos que $\langle x, y \rangle$ y $d(x, y)$ se comportan en direcciones opuestas. Vectores x y y muy similares dan como resultado un valor grande para el producto interno y un valor pequeño para la métrica.

3.4 Ángulos y Ortogonalidad

Además de permitir la definición de la longitud de vectores y de la distancia entre dos vectores, los productos internos también capturan la geometría del espacio vectorial al definir el ángulo ω entre dos vectores. Usamos la desigualdad de Cauchy-Schwarz (3.17) para definir ángulos ω en espacios con producto interno entre dos vectores x, y, y esta noción coincide con nuestra intuición en $\mathbb{R}^2$ y $\mathbb{R}^3$. Supongamos que $x \neq 0$, $y \neq 0$. Entonces

$$-1 \leq \frac{\langle x, y \rangle}{\|x\| \, \|y\|} \leq 1. \tag{3.24}$$

Por lo tanto, existe un único $\omega \in [0, \pi]$, ilustrado en la Figura 2, tal que

$$\cos \omega = \frac{\langle x, y \rangle}{\|x\| \, \|y\|}. \tag{3.25}$$

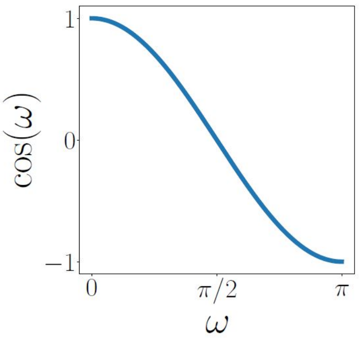


Figura 2: Cuando se restringe a $[0, \pi]$, entonces $f(\omega) = \cos(\omega)$ devuelve un valor único en el intervalo $[-1, 1]$.

El número ω es el *ángulo* entre los vectores x y y. Intuitivamente, el ángulo entre dos vectores nos indica qué tan similares son sus orientaciones. Por ejemplo, usando el producto punto, el ángulo entre x y $y = 4x$, es decir, y es una versión escalada de x, es 0: sus orientaciones son iguales.

Ejemplo 3.6 (Ángulo entre vectores)

Calculemos el ángulo entre $x = [1, 1]^\top \in \mathbb{R}^2$ y $y = [1, 2]^\top \in \mathbb{R}^2$; ver la Figura 3.5, donde usamos el producto punto como producto interno. Entonces obtenemos

$$\cos\omega = \frac{\langle x, y\rangle}{\sqrt{\langle x, x\rangle\langle y, y\rangle}} = \frac{x^\top y}{\sqrt{x^\top x\, y^\top y}} = \frac{3}{\sqrt{10}}, \tag{3.26}$$

y el ángulo entre los dos vectores es $\arccos\left(\frac{3}{\sqrt{10}}\right) \approx 0{,}32$ radianes, lo que corresponde a aproximadamente 18°.

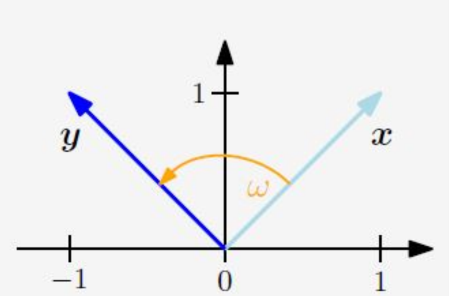
Una característica clave del producto interno es que también nos permite caracterizar vectores que son ortogonales.

Definición 3.7 (Ortogonalidad). Dos vectores x y y son *ortogonales* si y solo si $\langle x, y\rangle = 0$, y escribimos $x \perp y$. Si además $\|x\| = 1 = \|y\|$, es decir, los vectores son unitarios, entonces x y y son *ortonormales*.

Una implicancia de esta definición es que el **vector cero** es ortogonal a todo vector en el espacio vectorial.

Observación. La ortogonalidad es la generalización del concepto de perpendicularidad a formas bilineales que no necesariamente son el producto punto. En nuestro contexto, geométricamente, podemos pensar en vectores ortogonales como aquellos que forman un ángulo recto con respecto a un producto interno específico. ◊

Ejemplo 3.7 (Vectores ortogonales)



Consideremos dos vectores $x = [1, 1]^\top$, $y = [-1, 1]^\top \in \mathbb{R}^2$; ver la Figura 3.6. Nos interesa determinar el ángulo ω entre ellos usando dos productos internos distintos. Usando el producto punto como producto interno, obtenemos un ángulo ω entre x y y de 90°, tal que $x \perp y$. Sin embargo, si elegimos el producto interno

$$\langle x, y\rangle = x^\top \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} y, \tag{3.27}$$

obtenemos que el ángulo ω entre x y y está dado por

$$\cos\omega = \frac{\langle x, y\rangle}{\|x\|\,\|y\|} = \frac{1}{3} \implies \omega \approx 1{,}91\,\text{rad} \approx 109{,}5°, \tag{3.28}$$

y x y y no son ortogonales. Por lo tanto, vectores que son ortogonales con respecto a un producto interno no necesariamente lo son con respecto a un producto interno diferente.

Definición 3.8 (Matriz ortogonal). Una matriz cuadrada $A \in \mathbb{R}^{n \times n}$ es una *matriz ortogonal* si y sólo si sus columnas son ortonormales, de modo que

$$AA^\top = I = A^\top A, \tag{3.29}$$

lo que implica que

$$A^{-1} = A^\top, \tag{3.30}$$

es decir, la inversa se obtiene simplemente trasponiendo la matriz.

Las transformaciones mediante matrices ortogonales son especiales porque la longitud de un vector x no cambia al transformarse usando una matriz ortogonal A. Para el producto punto, obtenemos

$$\|Ax\|^2 = (Ax)^\top (Ax) = x^\top A^\top A x = x^\top I x = x^\top x = \|x\|^2. \tag{3.31}$$

Además, el ángulo entre dos vectores x, y, medido por su producto interno, también permanece inalterado cuando ambos se transforman usando una matriz ortogonal A. Suponiendo el producto punto como producto interno, el ángulo entre las imágenes Ax y Ay está dado por

$$\cos\omega = \frac{(Ax)^\top (Ay)}{\|Ax\|\|Ay\|} = \frac{x^\top A^\top A y}{\sqrt{x^\top A^\top A x y^\top A^\top A y}} = \frac{x^\top y}{\|x\|\|y\|}, \tag{3.32}$$

lo que da exactamente el mismo ángulo entre x y y. Esto significa que las matrices ortogonales A con $A^\top = A^{-1}$ preservan tanto los ángulos como las distancias. Resulta que las matrices ortogonales definen transformaciones que son rotaciones (con posibilidad de reflejos). En la Sección 3.9, discutiremos más detalles sobre rotaciones.

3.5 Base ortonormal

En las Secciones 3.3 y 3.4, usamos productos internos para calcular la longitud de los vectores y el ángulo entre ellos. A continuación, discutiremos el caso especial en el que los vectores base son ortogonales entre sí y cuya longitud es 1. Llamaremos a esta base una base ortonormal.

Presentamos esto de manera más formal:

Definición 3.9 (Base ortonormal). Consideremos un espacio vectorial de dimensión n llamado V y una base $\{b_1, \ldots, b_n\}$ de V. Si

$$\langle b_i, b_j \rangle = 0 \quad \text{para } i \neq j \tag{3.33}$$

$$\langle b_i, b_i \rangle = 1 \tag{3.34}$$

para todo $i, j = 1, \ldots, n$, entonces la base se denomina *base ortonormal (ONB)*. Si solo se cumple (3.33), la base se llama *base ortogonal*. Nótese que (3.34) implica que cada vector base tiene norma 1.

Recordando la Sección 2.6.1, podemos usar la eliminación de Gauss para hallar una base para un espacio vectorial generado por un conjunto de vectores. Supongamos que se nos da un conjunto $\{\tilde{b}_1, \ldots, \tilde{b}_n\}$ de vectores base no ortogonales y no normalizados. Concatenamos estos vectores en una matriz $\tilde{B} = [\tilde{b}_1, \ldots, \tilde{b}_n]$ y aplicamos eliminación de Gauss a la matriz aumentada (ver Sección 2.3.2) $\left[\tilde{B}\,\tilde{B}^\top \,|\tilde{B}\right]$ para obtener una base ortonormal. Este método constructivo e iterativo para construir una base ortonormal $\{b_1, \ldots, b_n\}$ se llama el *proceso de Gram–Schmidt* (Strang, 2003).

Ejemplo 3.8 (Base ortonormal)

La base canónica/estándar para un espacio vectorial euclidiano $\mathbb{R}^n$ es una base ortonormal, donde el producto interno es el producto punto de vectores.
En $\mathbb{R}^2$, los vectores

$$b_1 = \frac{1}{\sqrt{2}}\begin{bmatrix}1\\1\end{bmatrix}, \quad b_2 = \frac{1}{\sqrt{2}}\begin{bmatrix}1\\-1\end{bmatrix} \tag{3.35}$$

forman una base ortonormal ya que $b_1^\top b_2 = 0$ y $\|b_1\| = 1 = \|b_2\|$.

Explotaremos el concepto de una base ortonormal en el Capítulo 12 y el Capítulo 10 cuando discutamos máquinas de vectores de soporte y análisis de componentes principales.

3.6 Complemento ortogonal

Habiendo definido ortogonalidad, ahora estudiaremos espacios vectoriales que son ortogonales entre sí. Esto jugará un rol importante en el Capítulo 10, cuando discutamos reducción de dimensionalidad lineal desde una perspectiva geométrica.

Considera un espacio vectorial de dimensión D, V, y un subespacio de dimensión M, $U \subseteq V$. Entonces su complemento ortogonal $U^{\perp}$ es un subespacio de dimensión $(D - M)$ de V y contiene todos los vectores en V que son ortogonales a todo vector en U. Además, $U \cap U^{\perp} = \{0\}$, de modo que cualquier vector $x \in V$ puede ser descompuesto de manera única como

$$x = \sum_{m=1}^{M} \lambda_m b_m + \sum_{j=1}^{D-M} \psi_j b_j^{\perp}, \quad \lambda_m, \psi_j \in \mathbb{R}, \tag{3.36}$$

donde $(b_1, \ldots, b_M)$ es una base de U y $(b_1^{\perp}, \ldots, b_{D-M}^{\perp})$ es una base de $U^{\perp}$.

Por lo tanto, el complemento ortogonal también puede usarse para describir un plano U (subespacio bidimensional) en un espacio vectorial tridimensional. Más específicamente, el vector w con $\|w\| = 1$, que es ortogonal al plano U, es el vector base de $U^{\perp}$. La Figura 3.7 ilustra esta configuración. Todos los vectores que son ortogonales a w deben (por construcción) estar en el plano U. El vector w se llama el vector normal de U.

En general, los complementos ortogonales pueden usarse para describir hiperplanos en espacios vectoriales y afines de dimensión n.

3.7 Producto Interno de Funciones

Hasta ahora, hemos estudiado propiedades de productos internos para calcular longitudes, ángulos y distancias. Nos enfocamos en productos internos de vectores de dimensión finita. A continuación, veremos un ejemplo de productos internos de un tipo diferente de vectores: productos internos de funciones.

Los productos internos que discutimos hasta ahora fueron definidos para vectores con un número finito de entradas. Podemos pensar en un vector $x \in \mathbb{R}^n$ como una función con n valores funcionales. El concepto de producto interno puede generalizarse a vectores con un número infinito de entradas (infinitamente numerables) y también a funciones de valores continuos (infinitamente no numerables). Entonces, la suma sobre componentes individuales de vectores (ver Ecuación (3.5), por ejemplo) se convierte en una integral.

Un producto interno de dos funciones $u : \mathbb{R} \to \mathbb{R}$ y $v : \mathbb{R} \to \mathbb{R}$ puede definirse como la integral definida

$$\langle u, v \rangle := \int_a^b u(x)v(x)\, dx \tag{3.37}$$

para límites inferiores y superiores $a, b < \infty$, respectivamente. Como con nuestro producto interno usual, podemos definir normas y ortogonalidad observando el producto interno. Si la Ecuación (3.37) evalúa a 0, entonces las funciones u y v son ortogonales.

Para hacer este producto interno matemáticamente preciso, necesitamos ocuparnos de medidas y la definición de integrales, lo que lleva a la definición de un espacio de Hilbert. Además, a diferencia de productos internos en vectores de dimensión finita, los productos internos de funciones pueden diverger (tener valor infinito). Todo esto requiere adentrarse en detalles más intrincados del análisis real y funcional, lo cual no cubrimos en este libro.

Ejemplo 3.9 (Producto Interno de Funciones)

Si elegimos $u = \sin(x)$ y $v = \cos(x)$, el integrando $f(x) = u(x)v(x)$ de la Ecuación (3.37) se muestra en la Figura 3.8. Vemos que esta función es impar, es decir, $f(-x) = -f(x)$. Por lo tanto, la integral con límites $a = -\pi$, $b = \pi$ de este producto interno se evalúa como 0. Por lo tanto, sin y cos son funciones ortogonales.

Observación. También se cumple que la colección de funciones

$$\{1, \cos(x), \cos(2x), \cos(3x), \ldots\} \tag{3.38}$$

es ortogonal si integramos desde $-\pi$ hasta π, es decir, cualquier par de funciones es ortogonal entre sí. La colección de funciones en (3.38) abarca un gran subespacio de funciones que son pares y periódicas en $[-\pi, \pi]$, y proyectar funciones sobre este subespacio es la idea fundamental detrás de las series de Fourier.

3.8 Proyecciones Ortogonales

Las proyecciones son una clase importante de transformaciones lineales (además de rotaciones y reflexiones) y desempeñan un papel importante en gráficos, teoría de la codificación, estadística y aprendizaje automático. En aprendizaje automático, a menudo trabajamos con datos de alta dimensión. Los datos de alta dimensión suelen ser difíciles de analizar o visualizar. Sin embargo, los datos de alta dimensión suelen tener la propiedad de que solo unas pocas dimensiones contienen la mayor parte de la información, y la mayoría de las otras dimensiones no son esenciales para describir las propiedades clave de los datos. Cuando comprimimos o visualizamos datos de alta dimensión, perdemos información. Para minimizar esta pérdida por compresión, idealmente identificamos las dimensiones más informativas de los datos.

Como se discutió en el Capítulo 1, los datos pueden representarse como vectores, y en este capítulo discutiremos algunas herramientas fundamentales para la compresión de datos. Más específicamente, podemos proyectar los datos originales de alta dimensión sobre un espacio de características de menor dimensión y trabajar en este espacio de menor dimensión para aprender más sobre el conjunto de datos y extraer patrones relevantes.

Por ejemplo,Los algoritmos de aprendizaje, como el análisis de componentes principales (PCA) de Pearson (1901) y Hotelling (1933), así como las redes neuronales profundas, explotan intensamente la idea de reducción de dimensionalidad.

En lo que sigue, nos enfocaremos en las proyecciones ortogonales, que usaremos en el Capítulo 10 para reducción lineal de dimensionalidad y en el Capítulo 12 para clasificación. Incluso la regresión lineal, que discutiremos en el Capítulo 9, puede interpretarse usando proyecciones ortogonales.

Para un subespacio de menor dimensión dado, las proyecciones ortogonales de datos de alta dimensión retienen la mayor cantidad de información posible y minimizan la diferencia/error entre los datos originales y su proyección correspondiente. Una ilustración de dicha proyección ortogonal se muestra en la Figura 3.9. Antes de detallar cómo obtener estas proyecciones, definamos formalmente qué es una proyección.

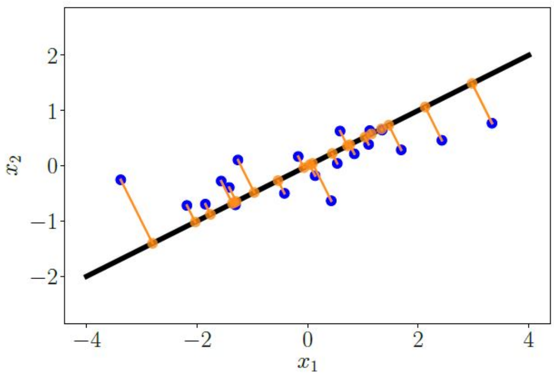


Figura 3: Proyección ortogonal (puntos naranjas) de un conjunto de datos bidimensional (puntos azules) sobre un subespacio unidimensional (línea recta).

Definición 3.10 (**Proyección**). Sea V un espacio vectorial y $U \subseteq V$ un subespacio de V. Una aplicación lineal $\pi : V \to U$ se llama una **proyección** si

$$\pi^2 = \pi \circ \pi = \pi.$$

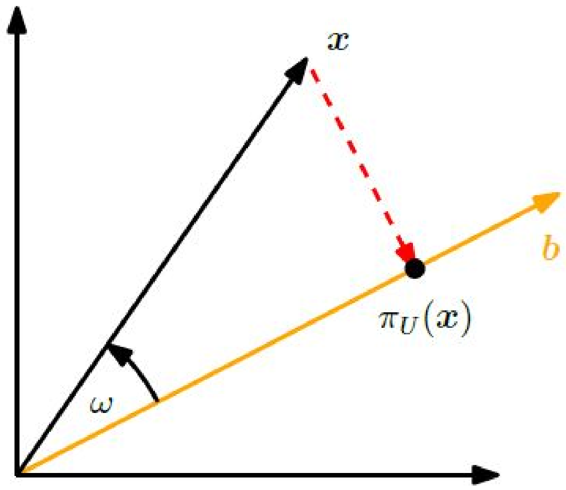
Dado que las aplicaciones lineales pueden expresarse mediante matrices de transformación (ver Sección 2.7), la definición anterior se aplica igualmente a un tipo especial de matrices de transformación, las **matrices de proyección** P_π, que cumplen la propiedad

$$P_\pi^2 = P_\pi.$$

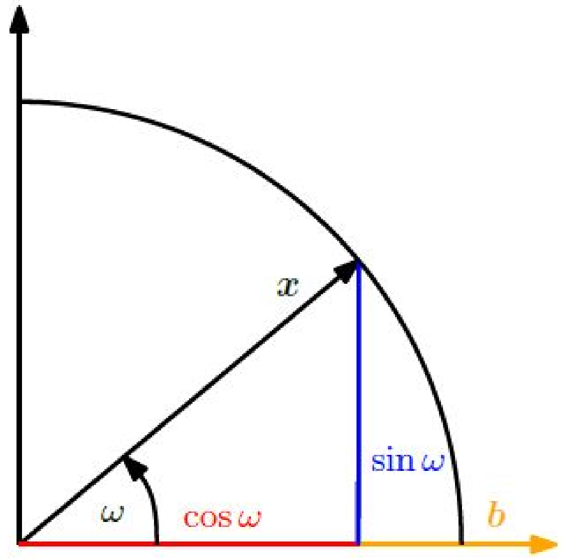
En lo que sigue, derivaremos **proyecciones ortogonales** de vectores en el espacio con producto interno $(\mathbb{R}^n, \langle \cdot, \cdot \rangle)$ sobre subespacios. Comenzaremos con subespacios unidimensionales, que también se llaman **líneas**. A menos que se indique lo contrario, asumimos que el producto punto es el producto interno:

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^\top \mathbf{y}.$$

Figura 3.10. Visualización geométrica de la proyección ortogonal sobre un subespacio unidimensional.



(a) Proyección de $\mathbf{x} \in \mathbb{R}^2$ sobre un subespacio U con vector base $\mathbf{b}$.



(b) Proyección de un vector bidimensional $\mathbf{x}$ con $\|\mathbf{x}\| = 1$ sobre un subespacio unidimensional generado por $\mathbf{b}$.

3.8.1 *Proyección sobre subespacios unidimensionales (líneas)*

Supongamos que se nos da una línea (subespacio unidimensional) que pasa por el origen con vector base $\mathbf{b} \in \mathbb{R}^n$. La línea es un subespacio unidimensional $U \subseteq \mathbb{R}^n$ generado por $\mathbf{b}$. Cuando proyectamos $\mathbf{x} \in \mathbb{R}^n$ sobre U, buscamos el vector $\pi_U(\mathbf{x}) \in U$ que esté más cerca de $\mathbf{x}$. Usando argumentos geométricos, caractericemos algunas propiedades de la proyección $\pi_U(\mathbf{x})$ (la Figura 3.10(a) sirve como ilustración):

- La proyección $\pi_U(\mathbf{x})$ está más cerca de $\mathbf{x}$, donde "más cerca" implica que la distancia $\|\mathbf{x} - \pi_U(\mathbf{x})\|$ es mínima. De ello se deduce que el segmento $\pi_U(\mathbf{x}) - \mathbf{x}$ es ortogonal a U, y por lo tanto, el vector base $\mathbf{b}$ de U. La condición de ortogonalidad da:
$$\langle \pi_U(\mathbf{x}) - \mathbf{x}, \mathbf{b} \rangle = 0,$$
ya que los ángulos entre vectores se definen mediante el producto interno.
- La proyección $\pi_U(\mathbf{x})$ de $\mathbf{x}$ sobre U debe ser un elemento de U y, por lo tanto, un múltiplo del vector base $\mathbf{b}$ que genera U. Por lo tanto, $\pi_U(\mathbf{x}) = \lambda\mathbf{b}$ para algún $\lambda \in \mathbb{R}$.

En los siguientes tres pasos, determinamos la coordenada λ, la proyección $\pi_U(\mathbf{x}) \in U$, y la matriz de proyección P_π que mapea cualquier $\mathbf{x} \in \mathbb{R}^n$ sobre U:

1. Encontrar la coordenada λ. Partimos de la condición de ortogonalidad, que establece que el vector proyectado $\pi_U(\mathbf{x})$ debe satisfacer:

$$\langle \mathbf{x} - \pi_U(\mathbf{x}), \mathbf{b} \rangle = 0.$$

Como $\pi_U(\mathbf{x}) = \lambda\mathbf{b}$, esta expresión se convierte en:

$$\langle \mathbf{x} - \lambda\mathbf{b}, \mathbf{b} \rangle = 0. \tag{3.39}$$

Aplicando la bilinealidad del producto interno, obtenemos:

$$\langle \mathbf{x}, \mathbf{b} \rangle - \lambda\langle \mathbf{b}, \mathbf{b} \rangle = 0,$$

lo cual implica que:

$$\lambda = \frac{\langle \mathbf{x}, \mathbf{b} \rangle}{\langle \mathbf{b}, \mathbf{b} \rangle} = \frac{\langle \mathbf{b}, \mathbf{x} \rangle}{\|\mathbf{b}\|^2}. \tag{3.40}$$

En este paso hemos utilizado que el producto interno es simétrico. Si tomamos $\langle \cdot, \cdot \rangle$ como el producto punto, entonces:

$$\lambda = \frac{\mathbf{b}^\top \mathbf{x}}{\mathbf{b}^\top \mathbf{b}} = \frac{\mathbf{b}^\top \mathbf{x}}{\|\mathbf{b}\|^2}. \tag{3.41}$$

Si además se cumple que $\|\mathbf{b}\| = 1$, entonces la coordenada λ de la proyección está dada simplemente por:

$$\lambda = \mathbf{b}^\top \mathbf{x}.$$

2. Encontrar el punto de proyección $\pi_U(\mathbf{x}) \in U$. Dado que $\pi_U(\mathbf{x}) = \lambda \mathbf{b}$, obtenemos inmediatamente usando (3.40) que:

$$\pi_U(\mathbf{x}) = \lambda \mathbf{b} = \frac{\langle \mathbf{x}, \mathbf{b} \rangle}{\|\mathbf{b}\|^2} \mathbf{b} = \frac{\mathbf{b}^\top \mathbf{x}}{\|\mathbf{b}\|^2} \mathbf{b}, \tag{3.42}$$

donde la última igualdad es válida únicamente si usamos el producto punto. También podemos calcular la longitud de $\pi_U(\mathbf{x})$ mediante la Definición 3.1:

$$\|\pi_U(\mathbf{x})\| = \|\lambda \mathbf{b}\| = |\lambda| \cdot \|\mathbf{b}\|. \tag{3.43}$$

Por lo tanto, nuestra proyección tiene longitud $|\lambda|$ por la longitud de $\mathbf{b}$. Esto también refuerza la intuición de que λ es la coordenada de $\pi_U(\mathbf{x})$ con respecto al vector base $\mathbf{b}$ que genera nuestro subespacio unidimensional U.

Si usamos el producto punto como producto interno, obtenemos:

$$\|\pi_U(\mathbf{x})\| \overset{(3.42)}{=} \left| \frac{\mathbf{b}^\top \mathbf{x}}{\|\mathbf{b}\|^2} \right| \cdot \|\mathbf{b}\| \overset{(3.25)}{=} |\cos \omega| \cdot \|\mathbf{x}\| \cdot \|\mathbf{b}\| \cdot \frac{1}{\|\mathbf{b}\|} = |\cos \omega| \cdot \|\mathbf{x}\|. \tag{3.44}$$

Aquí, ω es el ángulo entre $\mathbf{x}$ y $\mathbf{b}$. Esta ecuación debería resultarte familiar de la trigonometría: si $\|\mathbf{x}\| = 1$, entonces $\mathbf{x}$ yace sobre el círculo unitario. Se deduce que la proyección de $\mathbf{x}$ sobre el eje horizontal generado por $\mathbf{b}$ es exactamente $\cos \omega$, y la longitud del vector proyectado $\pi_U(\mathbf{x}) = |\cos \omega|$. Una ilustración se muestra en la Figura 3.10(b).

3. Encontrar la matriz de proyección P_π. Sabemos que una proyección es una transformación lineal (ver Definición 3.10). Por lo tanto, existe una matriz de proyección P_π tal que

$$\pi_U(\mathbf{x}) = P_\pi \mathbf{x}.$$

Usando el producto punto como producto interno:

$$\pi_U(\mathbf{x}) = \lambda \mathbf{b} = \mathbf{b} \lambda = \mathbf{b} \frac{\mathbf{b}^\top \mathbf{x}}{\|\mathbf{b}\|^2} = \frac{\mathbf{b}\mathbf{b}^\top}{\|\mathbf{b}\|^2} \mathbf{x}, \tag{3.45}$$

vemos inmediatamente que:

$$P_\pi = \frac{\mathbf{b}\mathbf{b}^\top}{\|\mathbf{b}\|^2}. \tag{3.46}$$

Nótese que $\mathbf{b}\mathbf{b}^\top$ (y por consiguiente, P_π) es una matriz simétrica (de rango 1), y que $\|\mathbf{b}\|^2 = \langle \mathbf{b}, \mathbf{b} \rangle$ es un escalar.

La matriz de proyección P_π proyecta cualquier vector $\mathbf{x} \in \mathbb{R}^n$ sobre la recta que pasa por el origen y tiene dirección $\mathbf{b}$ (equivalentemente, el subespacio U generado por $\mathbf{b}$).

Observación. La proyección $\pi_U(\mathbf{x}) \in \mathbb{R}^n$ sigue siendo un vector de dimensión n y no un escalar. Sin embargo, ya no necesitamos n coordenadas para representar la proyección, sino únicamente una, si queremos expresarla respecto al vector base $\mathbf{b}$ que genera el subespacio U: λ.

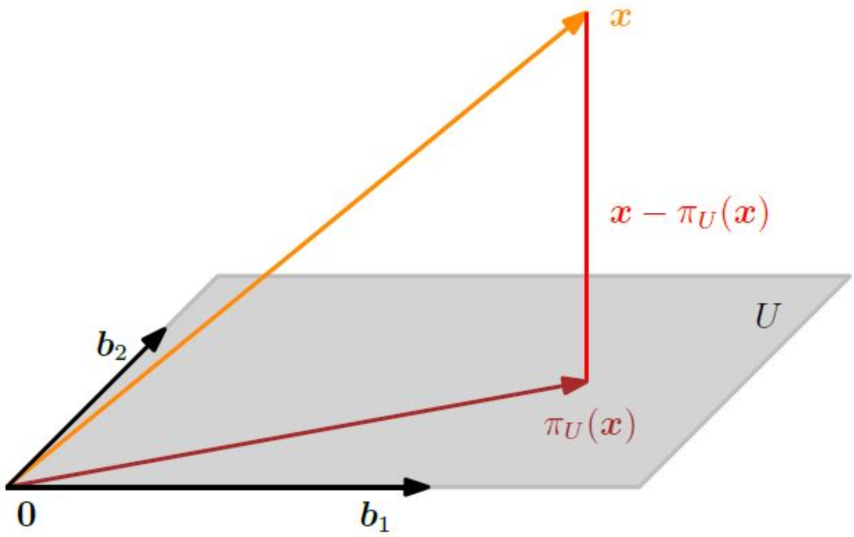


Figura 3.11
Proyección sobre un subespacio bidimensional U con base $\mathbf{b}_1$, $\mathbf{b}_2$. La proyección $\pi_U(\mathbf{x})$ de $\mathbf{x} \in \mathbb{R}^3$ sobre U puede expresarse como una combinación lineal de $\mathbf{b}_1$, $\mathbf{b}_2$, y el vector de desplazamiento $\mathbf{x} - \pi_U(\mathbf{x})$ es ortogonal tanto a $\mathbf{b}_1$ como a $\mathbf{b}_2$.

Ejemplo 3.10 (Proyección sobre una recta)

Calcular la matriz de proyección P_π sobre la recta que pasa por el origen y está generada por $\mathbf{b} = \begin{bmatrix} 1 & 2 & 2 \end{bmatrix}^\top$. $\mathbf{b}$ representa una dirección y también una base del subespacio unidimensional (la recta que pasa por el origen).

Usando la ecuación (3.46), obtenemos:

$$P_\pi = \frac{\mathbf{b}\mathbf{b}^\top}{\mathbf{b}^\top\mathbf{b}} = \frac{1}{9}\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}\begin{bmatrix} 1 & 2 & 2 \end{bmatrix} = \frac{1}{9}\begin{bmatrix} 1 & 2 & 2 \\ 2 & 4 & 4 \\ 2 & 4 & 4 \end{bmatrix}. \tag{3.47}$$

Ahora elijamos un vector particular $\mathbf{x}$ y veamos si pertenece al subespacio generado por $\mathbf{b}$. Para $\mathbf{x} = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}^\top$, la proyección es:

$$\pi_U(\mathbf{x}) = P_\pi \mathbf{x} = \frac{1}{9}\begin{bmatrix} 1 & 2 & 2 \\ 2 & 4 & 4 \\ 2 & 4 & 4 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \frac{1}{9}\begin{bmatrix} 5 \\ 10 \\ 10 \end{bmatrix} \in \operatorname{span}\left(\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}\right). \tag{3.48}$$

Observa que aplicar nuevamente P_π a $\pi_U(\mathbf{x})$ no cambia el resultado, es decir, $P_\pi \pi_U(\mathbf{x}) = \pi_U(\mathbf{x})$. Esto es coherente con la Definición 3.10, que establece que toda matriz de proyección P_π satisface:

$$P_\pi^2 \mathbf{x} = P_\pi \mathbf{x}, \quad \text{para todo } \mathbf{x}.$$

3.8.2 Proyección sobre subespacios generales

A continuación, analizamos las proyecciones ortogonales de vectores $\mathbf{x} \in \mathbb{R}^n$ sobre subespacios de menor dimensión $U \subseteq \mathbb{R}^n$ con $\dim(U) = m \geq 1$. Una ilustración se presenta en la Figura 3.11.

Supongamos que $(\mathbf{b}_1, \dots, \mathbf{b}_m)$ es una base ordenada de U. Cualquier proyección $\pi_U(\mathbf{x})$ sobre U es necesariamente un elemento de U. Por lo tanto, puede representarse como combinación lineal de los vectores base:

$$\pi_U(\mathbf{x}) = \sum_{i=1}^{m} \lambda_i \mathbf{b}_i.$$

Al igual que en el caso unidimensional, seguimos un procedimiento de tres pasos para hallar la proyección $\pi_U(\mathbf{x})$ y la matriz de proyección P_π.

1. Hallar las coordenadas $\lambda_1, \dots, \lambda_m$ de la proyección (segun la base de U), de modo que la combinación lineal:

$$\pi_U(\mathbf{x}) = \sum_{i=1}^{m} \lambda_i \mathbf{b}_i = B\boldsymbol{\lambda}, \quad B = \begin{bmatrix} \mathbf{b}_1 & \dots & \mathbf{b}_m \end{bmatrix} \in \mathbb{R}^{n \times m}, \quad \boldsymbol{\lambda} = \begin{bmatrix} \lambda_1 & \dots & \lambda_m \end{bmatrix}^\top \in \mathbb{R}^m. \tag{3.49}$$

La proyección es el punto más cercano a $\mathbf{x} \in \mathbb{R}^n$, lo que implica que el vector de conexión $\mathbf{x} - \pi_U(\mathbf{x})$ es ortogonal a todos los vectores base $\mathbf{b}_i$. Esto da lugar a m condiciones simultáneas (asumiendo el producto punto):

$$\langle \mathbf{b}_1, \mathbf{x} - \pi_U(\mathbf{x}) \rangle = \mathbf{b}_1^\top (\mathbf{x} - \pi_U(\mathbf{x})) = 0 \tag{3.51}$$

$$\vdots$$

$$\langle \mathbf{b}_m, \mathbf{x} - \pi_U(\mathbf{x}) \rangle = \mathbf{b}_m^\top (\mathbf{x} - \pi_U(\mathbf{x})) = 0. \tag{3.52}$$

Sabiendo que $\pi_U(\mathbf{x}) = B\boldsymbol{\lambda}$, reescribimos:

$$\mathbf{b}_1^\top (\mathbf{x} - B\boldsymbol{\lambda}) = 0 \quad \vdots \quad \mathbf{b}_m^\top (\mathbf{x} - B\boldsymbol{\lambda}) = 0. \tag{3.53}$$

Así obtenemos un sistema de ecuaciones lineales homogéneo:

$$\begin{bmatrix} \mathbf{b}_1^\top \\ \vdots \\ \mathbf{b}_m^\top \end{bmatrix} (\mathbf{x} - B\boldsymbol{\lambda}) = 0 \iff B^\top (\mathbf{x} - B\boldsymbol{\lambda}) = 0 \iff B^\top B\boldsymbol{\lambda} = B^\top \mathbf{x}. \tag{3.55}$$

Esta última ecuación se conoce como la ecuación normal. Dado que los vectores $\mathbf{b}_1, \ldots, \mathbf{b}_m$ forman una base de U y son linealmente independientes, $B^\top B \in \mathbb{R}^{m \times m}$ es regular (invertible), por lo que podemos resolver:

$$\boldsymbol{\lambda} = (B^\top B)^{-1} B^\top \mathbf{x}. \tag{3.57}$$

La matriz $(B^\top B)^{-1} B^\top$ se conoce como la *pseudo-inversa* de B, y puede calcularse incluso cuando B no es cuadrada. Solo se requiere que $B^\top B$ sea definida positiva (lo cual ocurre si B tiene rango completo). En aplicaciones prácticas (como regresión lineal), se suele añadir un *término de regularización* εI para garantizar estabilidad numérica y positividad definida. Este ridge"puede derivarse rigurosamente usando inferencia bayesiana (ver Capítulo 9).

2. Hallar la proyección $\pi_U(\mathbf{x}) \in U$ Ya establecimos que $\pi_U(\mathbf{x}) = B\boldsymbol{\lambda}$. Por lo tanto, usando (3.57):

$$\pi_U(\mathbf{x}) = B(B^\top B)^{-1} B^\top \mathbf{x}. \tag{3.58}$$

3. Hallar la matriz de proyección P_π A partir de (3.58), vemos que la matriz que satisface $P_\pi \mathbf{x} = \pi_U(\mathbf{x})$ es:

$$P_\pi = B(B^\top B)^{-1} B^\top. \tag{3.59}$$

Observación. La solución para proyectar sobre subespacios generales incluye el caso unidimensional como caso particular. Si $\dim(U) = 1$, entonces $B^\top B \in \mathbb{R}$ es un escalar y podemos reescribir (3.59) como:

$$P_\pi = \frac{BB^\top}{B^\top B},$$

que es exactamente la fórmula de la proyección en (3.46).

Ejemplo 3.11 (Proyección sobre un subespacio bidimensional)

Sea un subespacio $U = \operatorname{span}\left(\begin{bmatrix}1\\1\\1\end{bmatrix}, \begin{bmatrix}0\\1\\2\end{bmatrix}\right) \subseteq \mathbb{R}^3$ y $\mathbf{x} = \begin{bmatrix}6\\0\\0\end{bmatrix} \in \mathbb{R}^3$.

Queremos encontrar las coordenadas $\boldsymbol{\lambda}$ de $\mathbf{x}$ en términos de la base del subespacio U, el punto proyectado $\pi_U(\mathbf{x})$ y la matriz de proyección P_π.

Primero, notamos que el conjunto generador de U es una base (por independencia lineal), y escribimos los vectores base de U en una matriz:

$$B = \begin{bmatrix}1 & 0\\1 & 1\\1 & 2\end{bmatrix}.$$

Segundo, calculamos las matrices $B^\top B$ y $B^\top \mathbf{x}$:

$$B^\top B = \begin{bmatrix}1 & 1 & 1\\0 & 1 & 2\end{bmatrix}\begin{bmatrix}1 & 0\\1 & 1\\1 & 2\end{bmatrix} = \begin{bmatrix}3 & 3\\3 & 5\end{bmatrix}, \quad B^\top \mathbf{x} = \begin{bmatrix}1 & 1 & 1\\0 & 1 & 2\end{bmatrix}\begin{bmatrix}6\\0\\0\end{bmatrix} = \begin{bmatrix}6\\0\end{bmatrix}. \tag{3.60}$$

Tercero, resolvemos la ecuación normal $B^\top B\boldsymbol{\lambda} = B^\top \mathbf{x}$ para obtener $\boldsymbol{\lambda}$:

$$\begin{bmatrix}3 & 3\\3 & 5\end{bmatrix}\begin{bmatrix}\lambda_1\\\lambda_2\end{bmatrix} = \begin{bmatrix}6\\0\end{bmatrix} \iff \boldsymbol{\lambda} = \begin{bmatrix}5\\-3\end{bmatrix}. \tag{3.61}$$

Cuarto, la proyección $\pi_U(\mathbf{x})$ de $\mathbf{x}$ sobre U, es decir, el espacio columna de B, se puede calcular como:

$$\pi_U(\mathbf{x}) = B\boldsymbol{\lambda} = \begin{bmatrix}1 & 0\\1 & 1\\1 & 2\end{bmatrix}\begin{bmatrix}5\\-3\end{bmatrix} = \begin{bmatrix}5\\2\\-1\end{bmatrix}. \tag{3.62}$$

El *error de proyección* correspondiente es la norma del vector diferencia entre el original y su proyección sobre U:

$$\|\mathbf{x} - \pi_U(\mathbf{x})\| = \left\|\begin{bmatrix}1\\-2\\1\end{bmatrix}\right\| = \sqrt{6}. \tag{3.63}$$

Quinto, la matriz de proyección (para cualquier $\mathbf{x} \in \mathbb{R}^3$) está dada por:

$$P_\pi = B(B^\top B)^{-1}B^\top = \frac{1}{6}\begin{bmatrix}5 & 2 & -1\\2 & 2 & 2\\-1 & 2 & 5\end{bmatrix}. \tag{3.64}$$

Para verificar los resultados, podemos (a) comprobar que el vector de desplazamiento $\pi_U(\mathbf{x}) - \mathbf{x}$ es ortogonal a todos los vectores base de U, y (b) verificar que $P_\pi^2 = P_\pi$ (ver Definición 3.10).

3.8.3 *Ortogonalización de Gram-Schmidt*

Las proyecciones están en el núcleo del método de Gram-Schmidt, que nos permite transformar constructivamente cualquier base $(\mathbf{b}_1, \ldots, \mathbf{b}_n)$ de un espacio vectorial n-dimensional V en una base ortogonal/ortonormal $(\mathbf{u}_1, \ldots, \mathbf{u}_n)$ de V. Tal base siempre existe (Liesen y Mehrmann, 2015) y se cumple que:

$$\operatorname{span}[\mathbf{b}_1, \ldots, \mathbf{b}_n] = \operatorname{span}[\mathbf{u}_1, \ldots, \mathbf{u}_n].$$

El método de *ortogonalización de Gram-Schmidt* construye iterativamente una base ortogonal $(\mathbf{u}_1, \ldots, \mathbf{u}_n)$ a partir de cualquier base $(\mathbf{b}_1, \ldots, \mathbf{b}_n)$ de V de la siguiente forma:

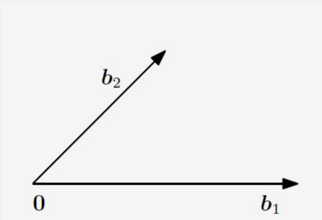
$$\mathbf{u}_1 := \mathbf{b}_1 \tag{3.67}$$

$$\mathbf{u}_k := \mathbf{b}_k - \pi_{\operatorname{span}\{\mathbf{u}_1, \ldots, \mathbf{u}_{k-1}\}}(\mathbf{b}_k), \quad k = 2, \ldots, n. \tag{3.68}$$

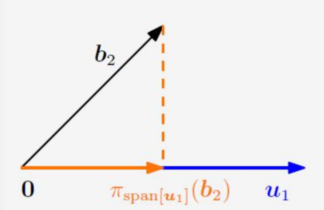
En la ecuación (3.68), el vector base $\mathbf{b}_k$ se proyecta sobre el subespacio generado por los primeros $k-1$ vectores ortogonales $\mathbf{u}_1, \ldots, \mathbf{u}_{k-1}$ (ver Sección 3.8.2). Esta proyección luego se resta de $\mathbf{b}_k$, y el resultado es un vector $\mathbf{u}_k$ que es ortogonal al subespacio $(k-1)$-dimensional generado por $\mathbf{u}_1, \ldots, \mathbf{u}_{k-1}$.

Repitiendo este procedimiento para todos los vectores base $\mathbf{b}_1, \ldots, \mathbf{b}_n$, se obtiene una base ortogonal $(\mathbf{u}_1, \ldots, \mathbf{u}_n)$ de V. Si normalizamos los $\mathbf{u}_k$, obtenemos una base ortonormal (ONB) donde $\|\mathbf{u}_k\| = 1$ para $k = 1, \ldots, n$.

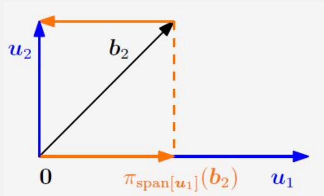
Ejemplo 3.12 (Ortogonalización de Gram-Schmidt)



(a) Vectores base originales no ortogonales $\mathbf{b}_1, \mathbf{b}_2$.



(b) Primer nuevo vector base $\mathbf{u}_1 = \mathbf{b}_1$ y proyección de $\mathbf{b}_2$ sobre el subespacio generado por $\mathbf{u}_1$.



(c) Vectores base ortogonales $\mathbf{u}_1$ y $\mathbf{u}_2 = \mathbf{b}_2 - \pi_{\mathrm{span}\{\mathbf{u}_1\}}(\mathbf{b}_2)$.

Consideremos una base $(\mathbf{b}_1, \mathbf{b}_2)$ de $\mathbb{R}^2$, donde

$$\mathbf{b}_1 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}, \quad \mathbf{b}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}; \tag{3.69}$$

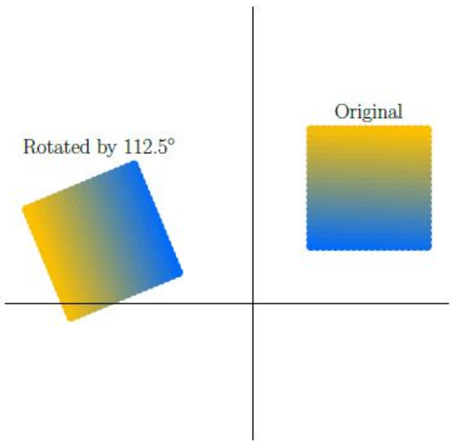
ver también la Figura 3.12(a). Usando el método de Gram-Schmidt, construimos una base ortogonal $(\mathbf{u}_1, \mathbf{u}_2)$ de $\mathbb{R}^2$ como sigue (asumiendo el producto punto como producto interno):

$$\mathbf{u}_1 := \mathbf{b}_1 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}, \tag{3.70}$$

$$\mathbf{u}_2 := \mathbf{b}_2 - \pi_{\mathrm{span}\{\mathbf{u}_1\}}(\mathbf{b}_2) \overset{(3.45)}{=} \mathbf{b}_2 - \frac{\mathbf{u}_1 \mathbf{u}_1^\top}{\|\mathbf{u}_1\|^2} \mathbf{b}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \tag{3.71}$$

3.9 Rotaciones

Figura 3.14



Una rotación gira objetos en un plano respecto al origen. Si el ángulo de rotación es positivo, rotamos en sentido antihorario.

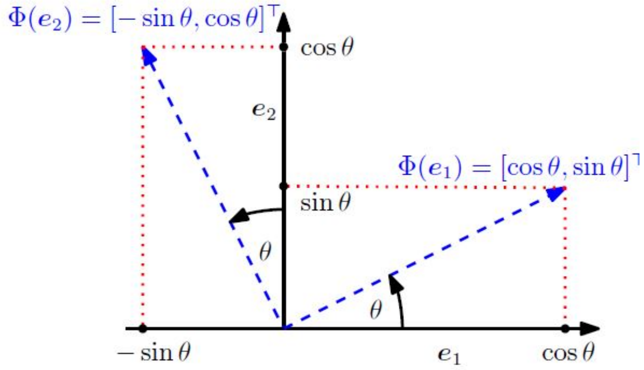
La preservación de la longitud y del ángulo, como se discutió en la Sección 3.4, son las dos características de las aplicaciones lineales con matrices de transformación ortogonales. En lo que sigue, analizaremos más de cerca matrices de transformación ortogonales específicas, que describen rotaciones.

Una rotación es una aplicación lineal (más específicamente, un automorfismo de un espacio vectorial euclídeo) que rota un plano en un ángulo θ alrededor del origen, es decir, el origen es un punto fijo. Para un ángulo positivo $\theta > 0$, por convención, rotamos en sentido antihorario. Un ejemplo se muestra en la Figura 3.14, donde la matriz de transformación es:

$$\mathbf{R} = \begin{bmatrix} -0{,}38 & -0{,}92 \\ 0{,}92 & -0{,}38 \end{bmatrix}. \tag{3.74}$$

3.9.1 Rotaciones en $\mathbb{R}^2$

Figura 3.16



Rotación de la base canónica en $\mathbb{R}^2$ por un ángulo θ.

Consideremos la base canónica

$$\left\{ \mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\} \subset \mathbb{R}^2,$$

la cual define el sistema de coordenadas estándar en $\mathbb{R}^2$. Nuestro objetivo es rotar este sistema de coordenadas por un ángulo θ, como se ilustra en la Figura 3.16. Observa que los vectores rotados siguen siendo linealmente independientes y, por lo tanto, conforman una base de $\mathbb{R}^2$. Esto implica que la rotación realiza un cambio de base.

Las rotaciones Φ son aplicaciones lineales, por lo que podemos expresarlas mediante una *matriz de rotación* $\mathbf{R}(\theta)$. La trigonometría (ver Figura 3.16) nos permite determinar las coordenadas de los ejes rotados (la imagen de Φ) con respecto a la base canónica de $\mathbb{R}^2$. Obtenemos:

$$\Phi(\mathbf{e}_1) = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}, \quad \Phi(\mathbf{e}_2) = \begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix}. \tag{3.75}$$

Por lo tanto, la matriz de rotación que realiza el cambio de base a las coordenadas rotadas $\mathbf{R}(\theta)$ está dada por:

$$\mathbf{R}(\theta) = [\Phi(\mathbf{e}_1) \quad \Phi(\mathbf{e}_2)] = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}. \tag{3.76}$$

Ejercicios

3.1 Demuestra que $\langle \cdot, \cdot \rangle$ definido para todo $x = [x_1, x_2]^\top \in \mathbb{R}^2$ y $y = [y_1, y_2]^\top \in \mathbb{R}^2$ por

$$\langle x, y \rangle := x_1 y_1 - (x_1 y_2 + x_2 y_1) + 2(x_2 y_2)$$

es un producto interno.

3.2 Considera $\mathbb{R}^2$ con $\langle \cdot, \cdot \rangle$ definido para todo x y y en $\mathbb{R}^2$ como

$$\langle x, y \rangle := x^\top \underbrace{\begin{bmatrix} 2 & 0 \\ 1 & 2 \end{bmatrix}}_{=:A} y.$$

¿Es $\langle \cdot, \cdot \rangle$ un producto interno?

3.3 Calcula la distancia entre

$$x = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad y = \begin{bmatrix} -1 \\ -1 \\ 0 \end{bmatrix}$$

usando:

a) $\langle x, y \rangle := x^\top y$

b) $\langle x, y \rangle := x^\top A y, \quad$ donde $A := \begin{bmatrix} 2 & 1 & 0 \\ 1 & 3 & -1 \\ 0 & -1 & 2 \end{bmatrix}$

3.4 Calcula el ángulo entre

$$x = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad y = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$$

usando:

a) $\langle x, y \rangle := x^\top y$

b) $\langle x, y \rangle := x^\top B y, \quad$ donde $B := \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}$

3.5 Considera el espacio vectorial euclidiano $\mathbb{R}^5$ con el producto punto. Un subespacio $U \subseteq \mathbb{R}^5$ y un vector $x \in \mathbb{R}^5$ están dados por

$$U = \operatorname{span}\left[\begin{bmatrix} 0 \\ -1 \\ 2 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ -3 \\ 1 \\ -1 \\ 2 \end{bmatrix}, \begin{bmatrix} -3 \\ 4 \\ 2 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ -3 \\ 5 \\ 0 \\ 7 \end{bmatrix}\right], \quad x = \begin{bmatrix} -1 \\ -9 \\ -1 \\ 4 \\ 1 \end{bmatrix}.$$

a) Determina la proyección ortogonal $\pi_U(x)$ de x sobre U.

b) Determina la distancia $d(x, U)$.