

Capítulo 7

Ortogonalidad y mínimos cuadrados

7.1. Producto interno, longitud y ortogonalidad

Definición 7.1. Un **producto interno** en un espacio vectorial $\mathcal{V}$ es una función que asocia a cada par de vectores $\mathbf{u}, \mathbf{v}$ un número real $\langle \mathbf{u}, \mathbf{v} \rangle$, satisfaciendo los siguientes axiomas para todo $\mathbf{u}, \mathbf{v}, \mathbf{w}$ en $\mathcal{V}$ y c escalar:

1. $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$
2. $\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle$
3. $\langle c\mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{u}, c\mathbf{v} \rangle = c \langle \mathbf{u}, \mathbf{v} \rangle$
4. $\langle \mathbf{u}, \mathbf{u} \rangle \geq 0$ y $\langle \mathbf{u}, \mathbf{u} \rangle = 0$ solo cuando $\mathbf{u} = \mathbf{0}$.

Ejemplo 7.2. En $\mathbf{R}^n$, podemos escribir un producto interno $\langle \mathbf{u}, \mathbf{v} \rangle$ de dos vectores $\mathbf{u}$ y $\mathbf{v}$ como su **producto escalar** $\mathbf{u} \cdot \mathbf{v}$:

$$\mathbf{u} = (u_1, u_2, \dots, u_n), \mathbf{v} = (v_1, v_2, \dots, v_n) \Rightarrow \mathbf{u} \cdot \mathbf{v} \equiv u_1 v_1 + u_2 v_2 + \dots + u_n v_n$$

Será muy útil escribir el producto escalar en forma matricial

$$\mathbf{u} \cdot \mathbf{v} = \mathbf{u}^T \mathbf{v} = \begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = u_1 v_1 + u_2 v_2 + \dots + u_n v_n$$

No es difícil comprobar que el producto escalar satisface todos los axiomas para ser un producto interno.

Ejemplo 7.3. Definiendo, para $f, g \in C[a, b]$ (el espacio vectorial de funciones continuas en $[a, b]$)

$$\langle f, g \rangle = \int_a^b f(t)g(t) dt$$

podemos mostrar, con las propiedades de la integral, que se cumplen todos los axiomas del producto interno.

Los espacios vectoriales que disponen de un producto interno se denominan *espacios con producto interno*. El ejemplo más importante de espacio con producto interno es el caso de $\mathbf{R}^n$ con el producto escalar (ejemplo 7.2) denominado **espacio euclídeo**. Por ello, vamos a escribir **en color rojo** las definiciones o propiedades que son específicas del espacio euclídeo $\mathbf{R}^n$ (se denomina con la misma notación que el espacio vectorial)

Norma

Definición 7.4. La **longitud** (o **norma**) de un vector $\mathbf{v} \in \mathcal{V}$ es

$$\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle} = \sqrt{\mathbf{v} \cdot \mathbf{v}} = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}$$

La primera observación es que el único vector cuya norma es 0, es el vector nulo $\mathbf{0}$, como consecuencia inmediata de la definición 7.1, punto 4. La longitud de un múltiplo escalar $c\mathbf{v}$ es $|c|$ veces la longitud de $\mathbf{v}$, porque $\sqrt{\langle c\mathbf{v}, c\mathbf{v} \rangle} = \sqrt{c^2 \langle \mathbf{v}, \mathbf{v} \rangle} = |c| \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$.

Es muy útil calcular, dado un vector $\mathbf{v}$, un vector $\mathbf{u}$ de longitud unidad en la misma dirección y sentido. Se obtiene con el denominado proceso de **normalización**: $\mathbf{u} = (1/\|\mathbf{v}\|)\mathbf{v}$ (dividir el vector por su norma) Los vectores cuya longitud es 1 se denominan **vectores unitarios**.

Ejemplo 7.5. Si $\mathcal{W} = \text{Gen} \left\{ \begin{bmatrix} 2 \\ 3 \end{bmatrix} \right\}$ encuentre una base de $\mathcal{W}$ que consista en un vector unitario.

Distancia

Definición 7.6. La **distancia** entre dos vectores $\mathbf{u}, \mathbf{v}$ de un espacio vectorial se define y denota por

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

Ejercicio 7.7. Calculad la distancia entre $\mathbf{u} = (1, 2, 3)$ y $\mathbf{v} = (1, 6, 6)$.

Vectores ortogonales

La siguiente definición generaliza la noción de perpendicularidad a espacios vectoriales arbitrarios.

Definición 7.8. Dos vectores $\mathbf{u}, \mathbf{v}$ de un espacio vectorial son **ortogonales** si

$$\langle \mathbf{u}, \mathbf{v} \rangle = 0 \quad (\mathbf{u} \cdot \mathbf{v} = 0)$$

El siguiente resultado clásico se generaliza y es mucho más transparente usando la teoría vectorial.

Teorema 7.9 (Teorema de Pitágoras). Dos vectores $\mathbf{u}, \mathbf{v}$ son ortogonales si y sólo si $\|\mathbf{u} + \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$.

Demostración. $\|\mathbf{u} + \mathbf{v}\|^2 = \langle \mathbf{u} + \mathbf{v}, \mathbf{u} + \mathbf{v} \rangle = \langle \mathbf{u}, \mathbf{u} \rangle + \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{v}, \mathbf{u} \rangle + \langle \mathbf{v}, \mathbf{v} \rangle = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 + 2\langle \mathbf{u}, \mathbf{v} \rangle$. $\square$

Ángulo

En Bachillerato se suele definir el producto escalar de dos vectores en $\mathbf{R}^2$ con la propiedad demostrada en el siguiente teorema.

Teorema 7.10. Sean $\mathbf{v}, \mathbf{w} \in \mathbf{R}^2$ y sea $0 \leq \theta \leq \pi$ el ángulo entre ellos. Entonces

$$\mathbf{v} \cdot \mathbf{w} = \|\mathbf{v}\| \|\mathbf{w}\| \cos \theta \quad (7.1)$$

Demostración. La demostración se realiza calculando $\|\mathbf{w} - \mathbf{v}\|^2$ y estudiando el triángulo formado por $\mathbf{v}$ y $\mathbf{w}$:

Observando la figura 7.1 y planteando el triángulo rectángulo de la derecha tenemos que $\|\mathbf{w} - \mathbf{v}\|^2 = x^2 + \|\mathbf{w}\|^2 \sin^2 \theta = (\|\mathbf{v}\| - \|\mathbf{w}\| \cos \theta)^2 + \|\mathbf{w}\|^2 \sin^2 \theta = \|\mathbf{v}\|^2 - 2\|\mathbf{v}\| \|\mathbf{w}\| \cos \theta + \|\mathbf{w}\|^2 \cos^2 \theta + \|\mathbf{w}\|^2 \sin^2 \theta$, con lo que hemos demostrado el teorema del coseno, según el cual

$$\|\mathbf{w} - \mathbf{v}\|^2 = \|\mathbf{w}\|^2 + \|\mathbf{v}\|^2 - 2\|\mathbf{w}\| \|\mathbf{v}\| \cos \theta.$$

Igualando esta expresión al resultado al desarrollo de $\|\mathbf{w} - \mathbf{v}\|^2 = (\mathbf{w} - \mathbf{v}) \cdot (\mathbf{w} - \mathbf{v}) = \|\mathbf{w}\|^2 + \|\mathbf{v}\|^2 - 2\mathbf{v} \cdot \mathbf{w}$ se deduce que $\mathbf{v} \cdot \mathbf{w} = \|\mathbf{v}\| \|\mathbf{w}\| \cos \theta$. $\square$

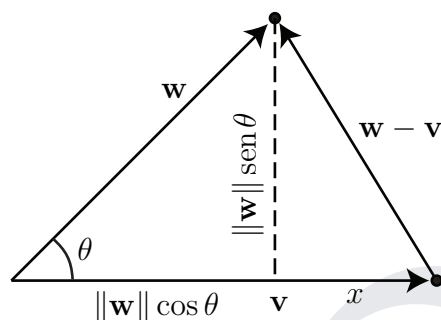


Figura 7.1: el teorema del coseno.

El mismo resultado se obtiene en $\mathbf{R}^3$, ya que dos vectores cualesquiera de $\mathbf{R}^3$ subtienden un plano, isomorfo a $\mathbf{R}^2$, y al cual el problema se puede reducir. De hecho, la fórmula (7.1) se puede utilizar para *definir* el ángulo entre dos vectores de un espacio $\mathbf{R}^n$ de cualquier dimensión.

Definición 7.11 (Ángulo). *Dados dos vectores $\mathbf{v}, \mathbf{w} \in \mathbf{R}^n$, el ángulo θ que subtienden es aquel comprendido entre 0 y π , que satisface*

$$\cos \theta = \frac{\mathbf{v} \cdot \mathbf{w}}{\|\mathbf{v}\| \|\mathbf{w}\|}, \quad 0 \leq \theta \leq \pi. \quad (7.2)$$

Esta definición no es incoherente debido a que el lado derecho de 7.2 es en valor absoluto menor o igual que 1, por la desigualdad de Cauchy-Schwarz (7.8), que demostraremos más adelante.

Ejercicio 7.12. Encontrad el ángulo entre los vectores $\mathbf{v} = (\cos \phi, \sin \phi)$ y $\mathbf{w} = -(\sin \phi, \cos \phi)$.

El complemento ortogonal

Si un vector $\mathbf{x}$ es ortogonal a todos los vectores $\mathbf{v}$ que se encuentran en un subespacio $\mathcal{W}$, se dice que $\mathbf{x}$ es **ortogonal** a $\mathcal{W}$. Dado un subespacio $\mathcal{W} \subset \mathcal{V}$ de un espacio vectorial $\mathcal{V}$, el conjunto de todos los vectores $\mathbf{x} \in \mathcal{V}$ ortogonales a $\mathcal{W}$ se denomina **complemento ortogonal** de $\mathcal{W}$, denotándose por $\mathcal{W}^\perp$:

$$\mathcal{W}^\perp = \{\mathbf{x} \in \mathcal{V} : \mathbf{x} \text{ es ortogonal a } \mathcal{W}\}$$

Ejemplo 7.13. Plano y recta ortogonales en $\mathbf{R}^3$.

Ejercicio 7.14. Si $\mathbf{v}$ está en $\mathcal{W}$ y en $\mathcal{W}^\perp$ simultáneamente, demostrad que sólo puede ser $\mathbf{v} = \mathbf{0}$.

Proposición 7.15.

1. $\mathbf{x} \in \mathcal{W}^\perp$ si y sólo si $\mathbf{x}$ es ortogonal a todos los vectores de un conjunto de generadores (cualquiera) de $\mathcal{W}$.
2. $\mathcal{W}^\perp$ es un subespacio de $\mathcal{V}$.

Teorema 7.16. Sea una matriz A de $m \times n$. 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 :

$$(\text{Fil } A)^\perp = \text{Nul } A \quad \text{y} \quad (\text{Col } A)^\perp = \text{Nul } A^T$$

Demostración. Úse la regla fila-columna para demostrar el primer resultado. El segundo se demuestra aplicando el primero a la matriz A^T . $\square$

7.2. Conjuntos ortogonales

Un conjunto de vectores $\{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ es un **conjunto ortogonal** si sus vectores son ortogonales dos a dos, es decir, $\langle \mathbf{u}_i, \mathbf{u}_j \rangle = 0$ cuando $i \neq j$.

Ejemplo 7.17. Muéstrese que $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ con

$$\mathbf{u}_1 = \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}, \quad \mathbf{u}_3 = \begin{bmatrix} -1/2 \\ -2 \\ 7/2 \end{bmatrix}$$

es ortogonal. **Solución:** $\mathbf{u}_1 \cdot \mathbf{u}_2$, $\mathbf{u}_2 \cdot \mathbf{u}_3$, $\mathbf{u}_3 \cdot \mathbf{u}_1$. **Dibujar.**

Teorema 7.18. Si $S = \{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ es un conjunto ortogonal de vectores no nulos de un espacio vectorial $\mathcal{V}$, entonces S es linealmente independiente, y por lo tanto S es una base de $\text{Gen } S$.

Demostración. Multiplíquese por $\mathbf{u}_i$ la combinación lineal $c_1 \mathbf{u}_1 + \dots + c_p \mathbf{u}_p = \mathbf{0}$.

Completar $\square$

Definición 7.19. Una **base ortogonal** de un espacio vectorial es una base que es también un conjunto ortogonal.

Teorema 7.20. Sea $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ una base ortogonal de un espacio vectorial $\mathcal{W}$. Entonces si $\mathbf{y} \in \mathcal{W}$ es un vector que escrito en función de la base es

$$\mathbf{y} = c_1 \mathbf{u}_1 + c_2 \mathbf{u}_2 + \dots + c_p \mathbf{u}_p$$

las coordenadas c_i se pueden calcular mediante una fórmula directa

$$c_i = \frac{\langle \mathbf{y}, \mathbf{u}_i \rangle}{\langle \mathbf{u}_i, \mathbf{u}_i \rangle} = \frac{\mathbf{y} \cdot \mathbf{u}_i}{\mathbf{u}_i \cdot \mathbf{u}_i} \quad (i = 1, \dots, p) \quad (7.3)$$

Demostración. Basta con multiplicar escalarmente la expresión de $\mathbf{y}$ por un vector $\mathbf{u}_i$:

$$\begin{aligned} \mathbf{u}_i \cdot \mathbf{y} &= \mathbf{u}_i \cdot (c_1 \mathbf{u}_1 + c_2 \mathbf{u}_2 + \dots + c_p \mathbf{u}_p) = c_1 \mathbf{u}_i \cdot \mathbf{u}_1 + \dots + c_i \mathbf{u}_i \cdot \mathbf{u}_i + \dots + c_p \mathbf{u}_p \cdot \mathbf{u}_i \\ &= c_1 \cdot 0 + \dots + c_i \mathbf{u}_i \cdot \mathbf{u}_i + \dots + c_p \cdot 0 = c_i \mathbf{u}_i \cdot \mathbf{u}_i \end{aligned}$$

□

En una base ortogonal, el cálculo de las coordenadas se simplifica mucho, porque hay una operación que las produce directamente, sin necesidad de resolver un sistema de ecuaciones o invertir matrices.

Ejemplo 7.21. Demuéstrese que el conjunto $\mathcal{S} = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ del ejemplo 7.17 es una base ortogonal de $\mathbf{R}^3$. Una vez hecho esto, encuentre

las coordenadas $[y]_{\mathcal{S}}$ del vector $\mathbf{y} = \begin{bmatrix} 6 \\ 1 \\ -8 \end{bmatrix}$ en la base $\mathcal{S}$.

Solución: $\mathbf{y} = \mathbf{u}_1 - 2\mathbf{u}_2 - 2\mathbf{u}_3$.

Proyección ortogonal

Proyección ortogonal sobre rectas. Sea dado un vector $\mathbf{u}$, no nulo, de $\mathbf{R}^n$. Considérese el problema de descomponer cualquier vector $\mathbf{y}$ de $\mathbf{R}^n$ en su *componente paralela* a $\mathbf{u}$ y su *componente perpendicular* a $\mathbf{u}$. Es decir, $\mathbf{y}$ será la suma de dos vectores

$$\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z} \quad (7.4)$$

donde $\mathbf{z}$ es ortogonal a $\mathbf{u}$, $\mathbf{u} \cdot \mathbf{z} = 0$, y $\hat{\mathbf{y}}$ es paralelo a $\mathbf{u}$, es decir $\hat{\mathbf{y}} = \alpha \mathbf{u}$ para algún escalar $\alpha \in \mathbf{R}$. **Hacer dibujo.** En $\mathbf{R}^2$ y quizás $\mathbf{R}^3$ nuestra intuición geométrica nos dice que esta descomposición es posible y, aparentemente, única, pero puede que en $\mathbf{R}^n$ la descomposición (7.4) no sea posible, o sea posible de muchas maneras.

La condición de ortogonalidad de $\mathbf{z}$ y $\mathbf{u}$ es

$$0 = \mathbf{z} \cdot \mathbf{u} = (\mathbf{y} - \hat{\mathbf{y}}) \cdot \mathbf{u} = (\mathbf{y} - \alpha \mathbf{u}) \cdot \mathbf{u} = \mathbf{y} \cdot \mathbf{u} - \alpha \mathbf{u} \cdot \mathbf{u}$$

que tiene solución (única) $\alpha = \frac{\mathbf{y} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}}$, lo cual implica a su vez que $\hat{\mathbf{y}} = \frac{\mathbf{y} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}} \mathbf{u}$. El vector $\hat{\mathbf{y}}$ es la componente de $\mathbf{y}$ paralela a $\mathbf{u}$, o **proyección ortogonal de $\mathbf{y}$ sobre $\mathbf{u}$** , y $\mathbf{z} = \mathbf{y} - \hat{\mathbf{y}}$ es la componente perpendicular o **componente ortogonal a $\mathbf{u}$ de $\mathbf{y}$** . En un espacio vectorial general, la proyección ortogonal de un vector $\mathbf{y}$ sobre otro $\mathbf{u}$ se define de forma evidente $\hat{\mathbf{y}} = \frac{\langle \mathbf{y}, \mathbf{u} \rangle}{\langle \mathbf{u}, \mathbf{u} \rangle} \mathbf{u}$.

Una observación importante es que la proyección ortogonal de $\mathbf{y}$ sobre $\mathbf{u}$ no depende de la longitud de $\mathbf{u}$, sólo de su dirección (ni siquiera de su sentido): proyectar $\mathbf{y}$ sobre cualquier múltiplo $c\mathbf{u}$ da lugar a $\hat{\mathbf{y}} = \frac{\mathbf{y} \cdot (c\mathbf{u})}{(c\mathbf{u}) \cdot (c\mathbf{u})} c\mathbf{u} = \frac{\mathbf{y} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}} \mathbf{u}$, lo mismo que sobre $\mathbf{u}$. Entonces, realmente la proyección ortogonal de un vector $\mathbf{y}$ está determinada sólo por la recta que genera $\mathbf{u}$, es decir, el *subespacio* L generado por $\mathbf{u}$. Por ello, se usa la siguiente notación:

$\hat{\mathbf{y}} = \text{proy}_L \mathbf{y} = \frac{\langle \mathbf{y}, \mathbf{u} \rangle}{\langle \mathbf{u}, \mathbf{u} \rangle} \mathbf{u} = \frac{\mathbf{y} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}} \mathbf{u}$	proyección ortogonal de $\mathbf{y}$ sobre la recta L generada por $\mathbf{u}$
---	---

Ejemplo 7.22. Sean $\mathbf{y} = \begin{bmatrix} 7 \\ 6 \end{bmatrix}$ y $\mathbf{u} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$. Calcúlese la proyección ortogonal de $\mathbf{y}$ sobre $\mathbf{u}$. Escribese $\mathbf{y}$ como la suma de un vector en $L = \text{Gen}\{\mathbf{u}\}$ y otro ortogonal a $\mathbf{u}$. Se suele abusar de la notación escribiendo $\text{proy}_{\mathbf{u}}$ en vez de proy_L .

Solución: La proyección ortogonal es

$$\hat{\mathbf{y}} = \text{proy}_{\mathbf{u}} \mathbf{y} = \frac{\mathbf{y} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}} \mathbf{u} = \frac{7 \cdot 4 + 6 \cdot 2}{4 \cdot 4 + 2 \cdot 2} \begin{bmatrix} 4 \\ 2 \end{bmatrix} = 2 \begin{bmatrix} 4 \\ 2 \end{bmatrix} = \begin{bmatrix} 8 \\ 4 \end{bmatrix}$$

La componente ortogonal es

$$\mathbf{z} = \mathbf{y} - \hat{\mathbf{y}} = \begin{bmatrix} 7 \\ 6 \end{bmatrix} - \begin{bmatrix} 8 \\ 4 \end{bmatrix} = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$$

Si todo es correcto, $\hat{\mathbf{y}} \cdot \mathbf{z} = 0$. Efectivamente $(8, 4) \cdot (-1, 2) = -8 + 8 = 0$.

Ejercicio 7.23. Por proyección de un vector $\mathbf{y}$ en la dirección de otro $\mathbf{u}$ se entiende muchas veces la norma de la proyección ortogonal. Demostrad que esa norma es

$$\|\text{proy}_{\mathbf{u}} \mathbf{y}\| = \frac{\mathbf{y} \cdot \mathbf{u}}{\|\mathbf{u}\|}.$$

La distancia de un vector $\mathbf{y}$ a una recta L se define como la norma de su componente ortogonal $\mathbf{z} = \mathbf{y} - \text{proy}_L \mathbf{y}$ a la recta:

$$\text{dist}(\mathbf{y}, L) = \|\mathbf{y} - \text{proy}_L \mathbf{y}\|$$

En el caso del ejemplo 7.22 la distancia de $\mathbf{y} = \begin{bmatrix} 7 \\ 6 \end{bmatrix}$ a la recta generada por $\mathbf{u} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$ es

$$\|\mathbf{y} - \text{proy}_L \mathbf{y}\| = \|\mathbf{z}\| = \sqrt{(-1)^2 + 2^2} = \sqrt{5}$$

Interpretación geométrica del teorema 7.20. En $\mathbf{R}^n$, con la fórmula de la proyección ortogonal a la vista, podemos deducir que las coordenadas c_i de un vector $\mathbf{y}$ respecto a una base ortogonal $\{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ son las proyecciones ortogonales con respecto a los ejes L_i generados por los vectores $\mathbf{u}_i$ de la base:

$$\begin{aligned} \mathbf{y} &= \frac{\mathbf{y} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1} \mathbf{u}_1 + \frac{\mathbf{y} \cdot \mathbf{u}_2}{\mathbf{u}_2 \cdot \mathbf{u}_2} \mathbf{u}_2 + \dots + \frac{\mathbf{y} \cdot \mathbf{u}_n}{\mathbf{u}_n \cdot \mathbf{u}_n} \mathbf{u}_n \\ &= \text{proy}_{L_1} \mathbf{y} + \text{proy}_{L_2} \mathbf{y} + \dots + \text{proy}_{L_n} \mathbf{y} \quad (7.5) \end{aligned}$$

En términos geométricos, estos ejes forman un *sistema de referencia ortogonal* de $\mathbf{R}^n = \text{Gen}\{\mathbf{u}_1, \dots, \mathbf{u}_n\}$, es decir, de ejes ortogonales. Un sistema de referencia asociado a una base no ortogonal es mucho más incómodo de utilizar, puesto que sus coordenadas han de averiguarse a través de la resolución de un sistema, de la matriz del cambio de base, o de otros procedimientos que siempre involucran la inversión de una matriz. Las proyecciones ortogonales son más sencillas de calcular, puesto que involucran únicamente productos escalares.

Conjuntos ortonormales

Se puede dar un paso más en la simplificación de las bases a utilizar: conseguir encontrar una **base ortonormal** del espacio considerado. Una base ortonormal es una base ortogonal $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ compuesta de *vectores unitarios*, es decir, $\|\mathbf{u}_i\| = 1$ para $i = 1, \dots, p$. Eso era precisamente la base canónica $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ de $\mathbf{R}^n$, una base ortonormal (pero no la única).

Ejemplo 7.24. Normalizando los vectores de una base ortogonal se consigue una base ortonormal. Por ejemplo, normalizando la base del ejemplo 7.17 se obtiene que

$$\mathbf{v}_1 = \begin{bmatrix} 3/\sqrt{11} \\ 1/\sqrt{11} \\ 1/\sqrt{11} \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} -1/\sqrt{6} \\ 2/\sqrt{6} \\ 1/\sqrt{6} \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} -1/\sqrt{66} \\ -4/\sqrt{66} \\ 7/\sqrt{66} \end{bmatrix}$$

es una base ortonormal.

Cuando se disponen de un conjunto ortonormal de vectores de $\mathbf{R}^m$, se puede formar una matriz cuyas columnas sean esos vectores. Este tipo de matrices es importante en la práctica, como veremos más adelante, y se pueden caracterizar mediante el siguiente resultado.

Teorema 7.25. *La matriz U de $m \times n$ tiene columnas ortonormales si y sólo si*

$$U^T U = I_n \quad (7.6)$$

Demostración. Es directa. □

Teorema 7.26. *Si U es una matriz $m \times n$ de columnas y $\mathbf{x}$ e $\mathbf{y}$ vectores de $\mathbf{R}^n$, entonces*

1. $\|U\mathbf{x}\| = \|\mathbf{x}\|$
2. $(U\mathbf{x}) \cdot (U\mathbf{y}) = \mathbf{x} \cdot \mathbf{y}$
3. $(U\mathbf{x}) \cdot (U\mathbf{y}) = 0$ si y sólo si $\mathbf{x} \cdot \mathbf{y} = 0$

Demostración. Se prueba primero 2:

$$(U\mathbf{x}) \cdot (U\mathbf{y}) = (U\mathbf{x})^T (U\mathbf{y}) = \mathbf{x}^T U^T U \mathbf{y} = \mathbf{x}^T \mathbf{y} = \mathbf{x} \cdot \mathbf{y}$$

Los casos 1 y 3 son claras consecuencias del caso 2. □

Es importante interpretar geoméricamente las propiedades anteriores. Consideremos una aplicación matricial $T(\mathbf{x}) = U\mathbf{x}$ de $\mathbf{R}^n$ a $\mathbf{R}^m$. La propiedad 2 significa que los productos escalares son respetados por la aplicación. Las propiedades 1 y 2 significan que, en consecuencia, la norma y la ortogonalidad también son preservadas por la aplicación.

Ejemplo 7.27. Sea $U = \begin{bmatrix} 1/\sqrt{2} & 2/3 \\ 1/\sqrt{2} & -2/3 \\ 0 & 1/3 \end{bmatrix}$ y $\mathbf{x} = \begin{bmatrix} \sqrt{2} \\ 3 \end{bmatrix}$. Las columnas de U son ortonormales y

$$U^T U = \begin{bmatrix} 1/\sqrt{2} & \sqrt{2} & 0 \\ 2/3 & -2/3 & 1/3 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} & 2/3 \\ 1/\sqrt{2} & -2/3 \\ 0 & 1/3 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Se tiene que $U\mathbf{x} = \begin{bmatrix} 1/\sqrt{2} & 2/3 \\ 1/\sqrt{2} & -2/3 \\ 0 & 1/3 \end{bmatrix} \begin{bmatrix} \sqrt{2} \\ 3 \end{bmatrix} = \begin{bmatrix} 3 \\ -1 \\ 11 \end{bmatrix}$ por lo que

$$\|U\mathbf{x}\| = \sqrt{9+1+1} = \sqrt{11} \quad \text{y} \quad \|\mathbf{x}\| = \sqrt{2+9} = \sqrt{11}$$

Si $T(\mathbf{x}) = U\mathbf{x}$ es una transformación lineal matricial $T: \mathbf{R}^n \rightarrow \mathbf{R}^n$, la matriz U es cuadrada. A las matrices cuadradas cuyas columnas son un conjunto ortogonal se las denomina **matrices ortogonales**. Tienen la siguiente propiedad: $U^{-1} = U^T$, ya que por (7.6) U^T es la inversa por la izquierda de U , y entonces por el teorema 2.27 k. también lo es por la derecha.

Ejemplo 7.28. La matriz $U = \begin{bmatrix} 3/\sqrt{11} & -1/\sqrt{6} & -1/\sqrt{66} \\ 1/\sqrt{11} & 2/\sqrt{6} & -4/\sqrt{66} \\ 1/\sqrt{11} & 1/\sqrt{6} & 7/\sqrt{66} \end{bmatrix}$ es ortogonal, porque es cuadrada y sus columnas son ortonormales.

Ejercicio 7.29. Demostrad que las *filas* de una matriz ortogonal forman un conjunto ortonormal.

7.3. Proyecciones ortogonales

Teorema 7.30 (La descomposición ortogonal). Sea $\mathcal{W}$ un subespacio de un espacio vectorial $\mathcal{V}$. Todo $\mathbf{y}$ de $\mathcal{V}$ puede escribirse de manera única como una suma

$$\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z}$$

con $\hat{\mathbf{y}}$ en $\mathcal{W}$ y $\mathbf{z}$ en $\mathcal{W}^\perp$. Si $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ es una base ortogonal de $\mathcal{W}$, entonces

$$\begin{aligned} \hat{\mathbf{y}} &= \frac{\langle \mathbf{y}, \mathbf{u}_1 \rangle}{\langle \mathbf{u}_1, \mathbf{u}_1 \rangle} \mathbf{u}_1 + \dots + \frac{\langle \mathbf{y}, \mathbf{u}_p \rangle}{\langle \mathbf{u}_p, \mathbf{u}_p \rangle} \mathbf{u}_p & \mathbf{z} &= \mathbf{y} - \hat{\mathbf{y}} \\ \hat{\mathbf{y}} &= \frac{\mathbf{y} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1} \mathbf{u}_1 + \dots + \frac{\mathbf{y} \cdot \mathbf{u}_p}{\mathbf{u}_p \cdot \mathbf{u}_p} \mathbf{u}_p & \mathbf{z} &= \mathbf{y} - \hat{\mathbf{y}} \end{aligned} \quad (7.7)$$

Se dice que $\text{proy}_{\mathcal{W}} \mathbf{y} = \hat{\mathbf{y}}$ es la **proyección ortogonal de $\mathbf{y}$ sobre $\mathcal{W}$** .

Demostración. Calculando

$$\mathbf{z} \cdot \hat{\mathbf{y}} = (\mathbf{y} - \hat{\mathbf{y}}) \cdot \hat{\mathbf{y}} = \frac{\mathbf{y} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1} (\mathbf{y} - \hat{\mathbf{y}}) \cdot \mathbf{u}_1 + \dots + \frac{\mathbf{y} \cdot \mathbf{u}_p}{\mathbf{u}_p \cdot \mathbf{u}_p} (\mathbf{y} - \hat{\mathbf{y}}) \cdot \mathbf{u}_p$$

Pero todos los términos del lado derecho ($i = 1, \dots, p$)

$$\begin{aligned}\langle \mathbf{y} - \hat{\mathbf{y}}, \mathbf{u}_i \rangle &= \langle \mathbf{y}, \mathbf{u}_i \rangle - \langle \hat{\mathbf{y}}, \mathbf{u}_i \rangle \\ &= \langle \mathbf{y}, \mathbf{u}_i \rangle - \left\langle \frac{\langle \mathbf{y}, \mathbf{u}_1 \rangle}{\langle \mathbf{u}_1, \mathbf{u}_1 \rangle} \mathbf{u}_1 + \dots + \frac{\langle \mathbf{y}, \mathbf{u}_p \rangle}{\langle \mathbf{u}_p, \mathbf{u}_p \rangle} \mathbf{u}_p, \mathbf{u}_i \right\rangle \\ &= \langle \mathbf{y}, \mathbf{u}_i \rangle - \frac{\langle \mathbf{y}, \mathbf{u}_i \rangle}{\langle \mathbf{u}_i, \mathbf{u}_i \rangle} \langle \mathbf{u}_i, \mathbf{u}_i \rangle = \langle \mathbf{y}, \mathbf{u}_i \rangle - \langle \mathbf{y}, \mathbf{u}_i \rangle = 0\end{aligned}$$

son cero. Por tanto, $\mathbf{z}$ es ortogonal a $\mathcal{W}$, $\mathbf{z} \in \mathcal{W}^\perp$. La descomposición es única por lo siguiente. Supongamos que existen dos maneras de descomponer $\mathbf{y}$, $\mathbf{y} = \hat{\mathbf{y}}_1 + \mathbf{z}_1$ y $\mathbf{y} = \hat{\mathbf{y}}_2 + \mathbf{z}_2$, con $\hat{\mathbf{y}}_1, \hat{\mathbf{y}}_2$ en $\mathcal{W}$ y $\mathbf{z}_1, \mathbf{z}_2$ en $\mathcal{W}^\perp$. Entonces $\hat{\mathbf{y}}_1 + \mathbf{z}_1 = \hat{\mathbf{y}}_2 + \mathbf{z}_2$, es decir

$$\hat{\mathbf{y}}_1 - \hat{\mathbf{y}}_2 = \mathbf{z}_1 - \mathbf{z}_2$$

donde el lado derecho es un vector de $\mathcal{W}$ y el izquierdo de $\mathcal{W}^\perp$, y son iguales. El único vector común a $\mathcal{W}$ y $\mathcal{W}^\perp$ es $\mathbf{0}$, así que $\hat{\mathbf{y}}_1 = \hat{\mathbf{y}}_2$ y $\mathbf{z}_1 = \mathbf{z}_2$. $\square$

Ejemplo 7.31. Sean $\mathbf{u}_1 = \begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}$ y $\mathbf{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$. El conjunto $\{\mathbf{u}_1, \mathbf{u}_2\}$ es

una base ortogonal de $\mathcal{W} = \text{Gen}\{\mathbf{u}_1, \mathbf{u}_2\}$, porque son dos vectores ortogonales y, por tanto, linealmente independientes. La descomposición ortogonal de $\mathbf{y}$ como suma de un vector de $\mathcal{W}$ más otro de $\mathcal{W}^\perp$

$$\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z}$$

se obtiene con

$$\hat{\mathbf{y}} = \frac{\mathbf{y} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1} \mathbf{u}_1 + \frac{\mathbf{y} \cdot \mathbf{u}_2}{\mathbf{u}_2 \cdot \mathbf{u}_2} \mathbf{u}_2 = \frac{9}{30} \begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix} + \frac{3}{6} \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -2/5 \\ 2 \\ 1/5 \end{bmatrix}$$

$\mathbf{y}$

$$\mathbf{z} = \mathbf{y} - \hat{\mathbf{y}} = \begin{bmatrix} 7/5 \\ 0 \\ 14/5 \end{bmatrix}$$

siendo fácil (y útil) comprobar que es ortogonal a $\mathcal{W}$

$$\mathbf{z} \cdot \mathbf{u}_1 = \mathbf{z} \cdot \mathbf{u}_2 = 0$$

La descomposición es entonces

$$\mathbf{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \hat{\mathbf{y}} + \mathbf{z} = \begin{bmatrix} -2/5 \\ 2 \\ 1/5 \end{bmatrix} + \begin{bmatrix} 7/5 \\ 0 \\ 14/5 \end{bmatrix}$$

La interpretación geométrica de la proyección ortogonal de un vector sobre un subespacio es sencilla: cada uno de los términos en $\hat{\mathbf{y}}$ es la proyección ortogonal de $\mathbf{y}$ sobre el eje de una base ortogonal de $\mathcal{W}$. Sumando todas esas proyecciones, se obtiene un vector de $\mathcal{W}$ que es la proyección ortogonal de $\mathbf{y}$ sobre $\mathcal{W}$. Es interesante hacer notar que *la proyección ortogonal* es siempre la misma, independientemente de la base ortogonal para $\mathcal{W}$ que estemos usando en la fórmula (7.7).

Dos desigualdades. Dado un vector $\mathbf{v}$ en un espacio con producto interno, y dado un subespacio $\mathcal{W}$ de dimensión finita, se puede aplicar el teorema de Pitágoras a la descomposición ortogonal

$$\|\mathbf{v}\|^2 = \|\text{proy}_{\mathcal{W}} \mathbf{v}\|^2 + \|\mathbf{v} - \text{proy}_{\mathcal{W}} \mathbf{v}\|^2$$

Esto implica que la norma de la proyección ortogonal $\|\text{proy}_{\mathcal{W}} \mathbf{v}\|$ siempre es menor o igual que la norma del vector $\|\mathbf{v}\|$. Y la siguiente importante desigualdad.

Teorema 7.32 (Desigualdad de Cauchy-Schwarz). *Para todo $\mathbf{u}, \mathbf{v} \in \mathcal{V}$*

$$|\langle \mathbf{u}, \mathbf{v} \rangle| \leq \|\mathbf{u}\| \|\mathbf{v}\| \quad (7.8)$$

Demostración. Si $\mathbf{u} = \mathbf{0}$ la igualdad se cumple. Suponiendo $\mathbf{u} \neq \mathbf{0}$, denotando $\mathcal{W} = \text{Gen}\{\mathbf{u}\}$

$$\|\text{proy}_{\mathcal{W}} \mathbf{v}\| = \left\| \frac{\langle \mathbf{v}, \mathbf{u} \rangle}{\langle \mathbf{u}, \mathbf{u} \rangle} \mathbf{u} \right\| = \frac{|\langle \mathbf{v}, \mathbf{u} \rangle|}{|\langle \mathbf{u}, \mathbf{u} \rangle|} \|\mathbf{u}\| = \frac{|\langle \mathbf{v}, \mathbf{u} \rangle|}{\|\mathbf{u}\|^2} \|\mathbf{u}\| = \frac{|\langle \mathbf{v}, \mathbf{u} \rangle|}{\|\mathbf{u}\|}$$

Como $\|\text{proy}_{\mathcal{W}} \mathbf{v}\| \leq \|\mathbf{v}\|$

$$\frac{|\langle \mathbf{v}, \mathbf{u} \rangle|}{\|\mathbf{u}\|} \leq \|\mathbf{v}\| \quad \Leftrightarrow \quad |\langle \mathbf{y}, \mathbf{v} \rangle| \leq \|\mathbf{u}\| \|\mathbf{v}\|$$

□

Teorema 7.33 (Desigualdad triangular). *Para todo $\mathbf{u}, \mathbf{v} \in \mathcal{V}$*

$$\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\| \quad (7.9)$$

Demostración.

$$\begin{aligned}\|\mathbf{u} + \mathbf{v}\|^2 &= \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 + 2\langle \mathbf{u}, \mathbf{v} \rangle \leq \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 + 2|\langle \mathbf{u}, \mathbf{v} \rangle| \\ &\leq \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 + 2\|\mathbf{u}\|\|\mathbf{v}\| = (\|\mathbf{u}\| + \|\mathbf{v}\|)^2\end{aligned}$$

□

Ejercicio 7.34. ¿Puede ser que $\|\mathbf{u} + \mathbf{v}\|^2 \geq \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$?

Propiedades de las proyecciones ortogonales. La primera observación es que si $\mathbf{y}$ está dentro de $\mathcal{W}$, entonces su proyección ortogonal es él mismo:

Si $\mathbf{y}$ está en $\mathcal{W} = \text{Gen}\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$, entonces $\text{proy}_{\mathcal{W}} \mathbf{y} = \mathbf{y}$

como se puede deducir comparando las coordenadas de $\mathbf{y}$ en la base ortogonal (7.5) con la fórmula de la proyección (7.7) (es la misma expresión) En este caso, el vector normal $\mathbf{z}$ sería $\mathbf{0}$. De hecho, esto es consecuencia del siguiente teorema de enorme utilidad práctica.

Teorema 7.35 (Teorema de mejor aproximación). Sea $\mathcal{W}$ un subespacio de un espacio vectorial $\mathcal{V}$, e $\mathbf{y}$ un vector de $\mathcal{V}$. Entonces $\hat{\mathbf{y}}$, la proyección ortogonal de $\mathbf{y}$ sobre $\mathcal{W}$, es el punto más cercano a $\mathbf{y}$ de $\mathcal{W}$:

$$\text{dist}(\mathbf{y}, \hat{\mathbf{y}}) < \text{dist}(\mathbf{y}, \mathbf{w}) \quad (\text{ó } \|\hat{\mathbf{y}} - \mathbf{y}\| < \|\mathbf{w} - \mathbf{y}\|)$$

para todo $\mathbf{w}$ en $\mathcal{W}$ distinto de $\hat{\mathbf{y}}$.

Una interpretación de este teorema es la siguiente. Supongamos que sólo disponemos de los vectores del subespacio $\mathcal{W}$ para describir los vectores del espacio total $\mathcal{V}$. Entonces, dado un vector general $\mathbf{y} \in \mathcal{V}$, el mejor vector dentro de $\mathcal{W}$ para describir $\mathbf{y}$ es su proyección ortogonal $\hat{\mathbf{y}}$, en el sentido de que es el más cercano a $\mathbf{y}$ de todos los que están en $\mathcal{W}$. Si el vector $\mathbf{y}$ ya estaba en $\mathcal{W}$, su proyección ortogonal es él mismo, $\hat{\mathbf{y}} = \mathbf{y}$. Este teorema demuestra de una manera alternativa la unicidad de la proyección ortogonal $\hat{\mathbf{y}}$, que es independiente de la base ortogonal de $\mathcal{W}$ que estemos usando. Adicionalmente, el vector $\mathbf{z} = \mathbf{y} - \hat{\mathbf{y}}$ es el *vector error*, siendo el error la distancia de $\mathbf{y}$ a $\mathcal{W}$, es decir $\|\mathbf{z}\| = \|\mathbf{y} - \hat{\mathbf{y}}\|$.

Demostración. Si $\mathbf{v} \in \mathcal{W}$ es un vector cualquiera de $\mathcal{W}$, entonces, como $\hat{\mathbf{y}} \in \mathcal{W}$, también $\hat{\mathbf{y}} - \mathbf{v} \in \mathcal{W}$. El vector $\mathbf{z} = \mathbf{y} - \hat{\mathbf{y}}$ es ortogonal a cualquier vector en $\mathcal{W}$, por el teorema 7.8, y en particular a $\hat{\mathbf{y}} - \mathbf{v}$. Por ello, el teorema de Pitágoras 7.9 afirma que

$$\|\mathbf{y} - \hat{\mathbf{y}}\|^2 + \|\hat{\mathbf{y}} - \mathbf{v}\|^2 = \|\mathbf{y} - \hat{\mathbf{y}} + \hat{\mathbf{y}} - \mathbf{v}\|^2 = \|\mathbf{y} - \mathbf{v}\|^2$$

Es decir, la distancia (al cuadrado) $\|\mathbf{y} - \mathbf{v}\|^2$ de $\mathbf{y}$ a cualquier vector $\mathbf{v} \neq \hat{\mathbf{y}}$ de $\mathcal{W}$ es mayor que la distancia (al cuadrado) $\|\mathbf{y} - \hat{\mathbf{y}}\|^2$ de $\mathbf{y}$ a $\hat{\mathbf{y}}$. □

Ejemplo 7.36. Sean $\mathbf{u}_1 = \begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}$, $\mathbf{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ y $\mathcal{W} = \text{Gen}\{\mathbf{u}_1, \mathbf{u}_2\}$ como en el ejemplo 7.31. La proyección ortogonal de $\mathbf{y}$ sobre $\mathcal{W}$ era

$$\hat{\mathbf{y}} = \frac{\mathbf{y} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1} \mathbf{u}_1 + \frac{\mathbf{y} \cdot \mathbf{u}_2}{\mathbf{u}_2 \cdot \mathbf{u}_2} \mathbf{u}_2 = \frac{9}{30} \begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix} + \frac{3}{6} \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -2/5 \\ 2 \\ 1/5 \end{bmatrix}$$

que es el punto más cercano a $\mathbf{y}$ de $\mathcal{W}$.

La distancia de un punto $\mathbf{y}$ a un subespacio se define como la distancia del punto al punto más cercano de $\mathcal{W}$. Ese punto más cercano, como hemos demostrado, es el correspondiente a la proyección ortogonal de $\mathbf{y}$ sobre $\mathcal{W}$. Esta distancia es la norma del vector $\mathbf{z} = \mathbf{y} - \hat{\mathbf{y}}$. En el caso del ejemplo 7.31, es $\sqrt{(\frac{7}{5})^2 + (\frac{14}{5})^2} = \frac{1}{5}\sqrt{49 + 196} = \frac{7}{5}\sqrt{5}$.

Ejemplo 7.37. Si $\mathbf{y} = \begin{bmatrix} -1 \\ -5 \\ 10 \end{bmatrix}$, $\mathbf{u}_1 = \begin{bmatrix} 5 \\ -2 \\ 1 \end{bmatrix}$ y $\mathbf{u}_2 = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}$, la distancia de $\mathbf{y}$ a $\mathcal{W} = \text{Gen}\{\mathbf{u}_1, \mathbf{u}_2\}$ es la longitud $\|\mathbf{y} - \hat{\mathbf{y}}\|$, donde $\hat{\mathbf{y}} = \text{proy}_{\mathcal{W}} \mathbf{y}$. La proyección $\hat{\mathbf{y}}$ se calcula con una base ortogonal de $\mathcal{W}$ que, afortunadamente, $\mathbf{u}_1$ y $\mathbf{u}_2$ forman. Entonces:

$$\begin{aligned} \hat{\mathbf{y}} = \text{proy}_{\mathcal{W}} \mathbf{y} &= \frac{\mathbf{y} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1} \mathbf{u}_1 + \frac{\mathbf{y} \cdot \mathbf{u}_2}{\mathbf{u}_2 \cdot \mathbf{u}_2} \mathbf{u}_2 = \frac{15}{30} \begin{bmatrix} 5 \\ -2 \\ 1 \end{bmatrix} - \frac{21}{6} \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} \\ &= \begin{bmatrix} 5/2 \\ -1 \\ 1/2 \end{bmatrix} - \begin{bmatrix} 7/2 \\ 7 \\ -7/2 \end{bmatrix} = \begin{bmatrix} -1 \\ -8 \\ 4 \end{bmatrix} \end{aligned}$$

y

$$\mathbf{z} = \mathbf{y} - \hat{\mathbf{y}} = \begin{bmatrix} -1 \\ -5 \\ 10 \end{bmatrix} - \begin{bmatrix} -1 \\ -8 \\ 4 \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \\ 6 \end{bmatrix}, \quad \|\mathbf{z}\| = \sqrt{3^2 + 6^2} = 3\sqrt{5}$$

que es la distancia de $\mathbf{y}$ a $\mathcal{W}$.

Cuando la base de $\mathcal{W}$ de que disponemos no sólo es ortogonal, sino ortonormal, la expresión de la proyección ortogonal se simplifica, e incluso adquiere una interesante forma matricial.

Teorema 7.38. Si $\mathcal{B} = \{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ es una base ortonormal de un subespacio $\mathcal{W}$ de un espacio vectorial $\mathcal{V}$, entonces

$$\begin{aligned}\text{proy}_{\mathcal{W}} \mathbf{y} &= \langle \mathbf{y}, \mathbf{u}_1 \rangle \mathbf{u}_1 + \dots + \langle \mathbf{y}, \mathbf{u}_p \rangle \mathbf{u}_p \\ \text{proy}_{\mathcal{W}} \mathbf{y} &= (\mathbf{y} \cdot \mathbf{u}_1) \mathbf{u}_1 + \dots + (\mathbf{y} \cdot \mathbf{u}_p) \mathbf{u}_p\end{aligned}\quad (7.10)$$

Si $\mathcal{V} = \mathbf{R}^n$, y $U = [\mathbf{u}_1 \mid \mathbf{u}_2 \mid \dots \mid \mathbf{u}_p]$ entonces

$$[\mathbf{y}]_{\mathcal{B}} = U^T \mathbf{y} \quad \text{y} \quad \text{proy}_{\mathcal{W}} \mathbf{y} = U U^T \mathbf{y} \quad \forall \mathbf{y} \in \mathbf{R}^n \quad (7.11)$$

Demostración. Las fórmulas (7.10) de la proyección son consecuencia directa de que $\langle \mathbf{u}_i, \mathbf{u}_i \rangle = 1$ en la fórmula (7.7). Las coordenadas de $\mathbf{y}$ en la base ortonormal $\mathcal{B}$ son los coeficientes $c_i = \mathbf{y} \cdot \mathbf{u}_i = \mathbf{u}_i^T \mathbf{y}$ en (7.10), con lo que se demuestra que $[\mathbf{y}]_{\mathcal{B}} = U^T \mathbf{y}$. La fórmula matricial es simplemente la regla del producto matriz-vector que expresa $\text{proy}_{\mathcal{W}} \mathbf{y}$ como combinación lineal de las columnas $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ de U . $\square$

Suele ser más ventajoso usar la fórmula (7.7) que (7.10), con una base dada ortogonal, puesto que normalizar suele dar lugar a la introducción de raíces cuadradas que complican los cálculos.

Destaquemos que hemos formado, a partir de una matriz U de $n \times p$, dos matrices, $U^T U$ de $p \times p$ y $U U^T$ de $n \times n$ que se comportan, si $\mathcal{W}$ es $\text{Col } U$, como

$$\begin{aligned}U^T U \mathbf{x} &= I_p \mathbf{x} = \mathbf{x} & \forall \mathbf{x} \in \mathbf{R}^p \\ U U^T \mathbf{y} &= \text{proy}_{\mathcal{W}} \mathbf{y} & \forall \mathbf{y} \in \mathbf{R}^n\end{aligned}$$

La matriz $U U^T$ es una *matriz de proyección* o, en la jerga habitual, un **proyector ortogonal**. Su existencia demuestra que la proyección ortogonal es una operación lineal, y en aplicaciones es interesante disponer de una matriz que implementa la operación, pudiendo ser utilizada cuando se deben proyectar muchos vectores sobre el mismo subespacio.

Ejercicio 7.39. Es fácil diagonalizar un proyector ortogonal visualizando geométricamente la proyección: ¿cuáles serán los espacios propios? ¿Y los valores propios correspondientes?

Cuando U es cuadrada (de $n \times n$) con columnas ortonormales, tenemos una matriz ortogonal, $\text{Col } U$ es todo $\mathbf{R}^n$, las columnas forman una base de $\mathbf{R}^n$ y tanto $U^T U$ como $U U^T$ son la identidad $U^T U = U U^T = I_n$ (el proyector es la identidad cuando el espacio sobre el que se proyecta es el espacio total)

7.4. El proceso de Gram-Schmidt

Las fórmulas de la proyección ortogonal dadas en las secciones anteriores dependen em modo esencial de que se disponga de una base ortogonal $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ del subespacio considerado. Pero, dado un subespacio $\mathcal{W}$, ¿y si no disponemos de una base ortogonal suya? A continuación veremos un procedimiento que construye explícitamente una base ortogonal partiendo de una base cualquiera del subespacio. Dicho sea de paso, este procedimiento también demuestra que siempre existe una base ortonormal de un subespacio de dimensión finita.

Ejemplo 7.40. Sea $\mathcal{W} = \text{Gen}\{\mathbf{x}_1, \mathbf{x}_2\}$ con $\mathbf{x}_1 = \begin{bmatrix} 3 \\ 6 \\ 0 \end{bmatrix}$ y $\mathbf{x}_2 = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$. Constrúyase

una base ortogonal de $\mathcal{W}$.

Se toma como primer vector a $\mathbf{x}_1$, $\mathbf{v}_1 = \mathbf{x}_1$. Como segundo vector, a la componente ortogonal de $\mathbf{x}_2$ respecto a $\mathbf{v}_1$: $\mathbf{v}_2 = \mathbf{x}_2 - \text{proy}_{\mathbf{v}_1} \mathbf{x}_2$:

$$\mathbf{v}_2 = \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1} \mathbf{v}_1 = \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix}$$

El conjunto $\{\mathbf{v}_1, \mathbf{v}_2\}$ es ortogonal, y ambos vectores están en $\mathcal{W}$. Es la base ortogonal deseada.

Ejemplo 7.41. Sean $\mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$, $\mathbf{x}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}$, $\mathbf{x}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$, con lo que $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$ es

linealmente dependiente, y base del subespacio que generan en $\mathbf{R}^4$. Constrúyase una base ortogonal de este espacio.

Los dos primeros pasos son los mismos del ejemplo anterior:

$$\mathbf{v}_1 = \mathbf{x}_1$$

$$\mathbf{v}_2 = \mathbf{x}_2 - \text{proy}_{\mathbf{v}_1} \mathbf{x}_2 = \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1} \mathbf{v}_1 = \begin{bmatrix} -3/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{bmatrix}$$

El último vector se puede simplificar, para facilitar los cálculos poste-

riores, eligiendo

$$\mathbf{v}'_2 = 4\mathbf{v}_2 = \begin{bmatrix} -3 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

Para encontrar el tercer vector, calculemos la componente ortogonal de $\mathbf{x}_3$ al subespacio generado por los dos vectores anteriores $\mathcal{W}_2 = \text{Gen}\{\mathbf{v}_1, \mathbf{v}'_2\}$:

$$\mathbf{v}_3 = \mathbf{x}_3 - \text{proy}_{\mathcal{W}_2} \mathbf{x}_3 = \mathbf{x}_3 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1} \mathbf{v}_1 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2} \mathbf{v}_2 = \begin{bmatrix} 0 \\ -2/3 \\ 1/3 \\ 1/3 \end{bmatrix}$$

que podemos sustituir, si queremos, por $\mathbf{v}'_3 = 3\mathbf{v}_3$. La base ortogonal es la formada por los vectores

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{v}'_2 = \begin{bmatrix} -3 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{v}'_3 = \begin{bmatrix} 0 \\ -2 \\ 1 \\ 1 \end{bmatrix}$$

Teorema 7.42. Dada una base $\{\mathbf{x}_1, \dots, \mathbf{x}_p\}$ de un subespacio $\mathcal{W}$ de $\mathcal{V}$, una base ortogonal de $\mathcal{W}$ es la formada por los vectores

$$\begin{aligned} \mathbf{v}_1 &= \mathbf{x}_1 \\ \mathbf{v}_2 &= \mathbf{x}_2 - \frac{\langle \mathbf{x}_2, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle} \mathbf{v}_1 \\ \mathbf{v}_3 &= \mathbf{x}_3 - \frac{\langle \mathbf{x}_3, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle} \mathbf{v}_1 - \frac{\langle \mathbf{x}_3, \mathbf{v}_2 \rangle}{\langle \mathbf{v}_2, \mathbf{v}_2 \rangle} \mathbf{v}_2 \\ &\vdots \\ \mathbf{v}_p &= \mathbf{x}_p - \frac{\langle \mathbf{x}_p, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle} \mathbf{v}_1 - \frac{\langle \mathbf{x}_p, \mathbf{v}_2 \rangle}{\langle \mathbf{v}_2, \mathbf{v}_2 \rangle} \mathbf{v}_2 - \dots - \frac{\langle \mathbf{x}_p, \mathbf{v}_{p-1} \rangle}{\langle \mathbf{v}_{p-1}, \mathbf{v}_{p-1} \rangle} \mathbf{v}_{p-1} \end{aligned}$$

7.5. Mínimos cuadrados

Como sabemos, hay sistemas $A\mathbf{x} = \mathbf{b}$ que son inconsistentes y no tienen solución. En la práctica, nos encontramos muchas veces con problemas de este

tipo. Basta con que la matriz A tenga más filas que columnas, para que la probabilidad de que $\mathbf{b}$ no esté en $\text{Col } A$ sea elevada y, por tanto (ver la proposición teorema 3.18) el sistema no tenga solución.

La idea es encontrar el vector $\mathbf{y} = A\mathbf{x}$ del subespacio $\text{Col } A$ que, ya que no puede ser $\mathbf{b}$, al menos sea el que más cercano se encuentre. Es decir, queremos encontrar el o los $\mathbf{x}$ tales que $\text{dist}(A\mathbf{x}, \mathbf{b}) = \|\mathbf{b} - A\mathbf{x}\|$ sea mínima, para A y $\mathbf{b}$ dados. O, lo que es lo mismo, si b_i e y_i , con $i = 1, \dots, m$ son las componentes de $\mathbf{b}$ e $\mathbf{y}$, queremos encontrar la $\mathbf{x}$ tal que, si $\mathbf{y} = A\mathbf{x}$, sea mínima $\|\mathbf{b} - \mathbf{y}\|^2 = (b_1 - y_1)^2 + \dots + (b_m - y_m)^2$, una suma de cuadrados (de ahí el nombre de mínimos cuadrados)

Definición 7.43. Sean A de $m \times n$ y $\mathbf{b}$ en $\mathbf{R}^m$. Una **solución por mínimos cuadrados** de $A\mathbf{x} = \mathbf{b}$ es un vector $\hat{\mathbf{x}}$ tal que

$$\|\mathbf{b} - A\hat{\mathbf{x}}\| \leq \|\mathbf{b} - A\mathbf{x}\|$$

para todo $\mathbf{x}$ en $\mathbf{R}^n$.

El teorema de mejor aproximación 7.35 afirma que la solución es aquella (o aquellas) $\hat{\mathbf{x}}$ tal que $\mathbf{y} = A\hat{\mathbf{x}}$ es la proyección $\hat{\mathbf{b}} = \text{proy}_{\text{Col } A} \mathbf{b}$ de $\mathbf{b}$ sobre el espacio columna de A :

$$A\hat{\mathbf{x}} = \hat{\mathbf{b}} \quad (7.12)$$

Este sistema tiene siempre solución porque $\hat{\mathbf{b}}$ está en el espacio columna de A .

El denominado *vector error* $\mathbf{b} - \hat{\mathbf{b}}$ es, por el teorema 7.8 de descomposición ortogonal, la componente de $\mathbf{b}$ ortogonal al subespacio sobre el que se proyecta, $\text{Col } A$. Por tanto, es ortogonal a todas las columnas $\mathbf{a}_j$ de A : $\mathbf{a}_j \cdot (\mathbf{b} - A\hat{\mathbf{x}}) = \mathbf{a}_j^T (\mathbf{b} - A\hat{\mathbf{x}}) = 0$. Podemos escribir esto matricialmente como

$$A^T (\mathbf{b} - A\hat{\mathbf{x}}) = \mathbf{0}$$

(Obsérvese que podíamos haber utilizado el teorema 7.16 para deducir este resultado: el ortogonal a $\text{Col } A$ es $\text{Nul } A^T$) Desarrollando

$$A^T \mathbf{b} - A^T A\hat{\mathbf{x}} = \mathbf{0} \quad \Leftrightarrow \quad A^T A\hat{\mathbf{x}} = A^T \mathbf{b}$$

Es decir, la ecuación fundamental de los cálculos de mínimos cuadrados es $A^T A\mathbf{x} = A^T \mathbf{b}$, un sistema lineal denominado **ecuaciones normales para $A\mathbf{x} = \mathbf{b}$** .

Teorema 7.44. El conjunto de soluciones por mínimos cuadrados de $A\mathbf{x} = \mathbf{b}$ es el conjunto solución de las ecuaciones normales $A^T A\hat{\mathbf{x}} = A^T \mathbf{b}$.

Demostración. Hemos demostrado que si $\hat{\mathbf{x}}$ es una solución por mínimos cuadrados, entonces es solución de las ecuaciones normales. Viceversa, si $\hat{\mathbf{x}}$ es solución de las ecuaciones normales, $A\hat{\mathbf{x}} - \mathbf{b}$ es ortogonal a todas las columnas de A . El vector $\mathbf{z} = \mathbf{b} - A\hat{\mathbf{x}}$ es la componente ortogonal de $\mathbf{b}$ a $\text{Col } A$, y la componente paralela es $\mathbf{b} - \mathbf{z} = A\hat{\mathbf{x}}$, única por el teorema 7.8 de descomposición ortogonal. Entonces ha de ser igual a la proyección de $\mathbf{b}$ sobre $\text{Col } A$, es decir $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$, y $\hat{\mathbf{x}}$ es una solución por mínimos cuadrados. $\square$

El error cometido por la o las soluciones de mínimos cuadrados es $\|\mathbf{b} - \hat{\mathbf{b}}\|$.

Ejemplo 7.45.

$$A = \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}$$

Entonces

$$A^T A \mathbf{x} = A^T \mathbf{b} \Leftrightarrow \begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 19 \\ 11 \end{bmatrix}$$

Como $A^T A$ es invertible

$$\hat{\mathbf{x}} = (A^T A)^{-1} A^T \mathbf{b} = \frac{1}{84} \begin{bmatrix} 5 & -1 \\ -1 & 17 \end{bmatrix} \begin{bmatrix} 19 \\ 11 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

El error de mínimos cuadrados $\|\mathbf{b} - A\hat{\mathbf{x}}\|$, como

$$\mathbf{b} = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix} \quad \text{y} \quad A\hat{\mathbf{x}} = \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \\ 3 \end{bmatrix} \Rightarrow \mathbf{b} - A\hat{\mathbf{x}} = \begin{bmatrix} -2 \\ -4 \\ 8 \end{bmatrix}$$

$$\text{es } \|\mathbf{b} - A\hat{\mathbf{x}}\| = \sqrt{(-2)^2 + (-4)^2 + 8^2} = \sqrt{84}$$

En el caso anterior $A^T A$ era invertible. No siempre es así.

Ejemplo 7.46.

$$A = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} -3 \\ -1 \\ 0 \\ 2 \\ 5 \\ 1 \end{bmatrix} \Rightarrow \hat{\mathbf{x}} = \begin{bmatrix} 3 \\ -5 \\ -2 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

Hay un criterio para determinar cuándo la solución al problema de mínimos cuadrados es única.

Teorema 7.47. *La matriz $A^T A$ es invertible si y sólo si las columnas de A son linealmente independientes. En ese caso, la solución $\hat{\mathbf{x}}$ por mínimos cuadrados de $A\mathbf{x} = \mathbf{b}$ es única y es*

$$\hat{\mathbf{x}} = (A^T A)^{-1} A^T \mathbf{b}$$

Demostración. Las columnas de A son linealmente independientes si y solo si el espacio nulo $\text{Nul } A \neq \emptyset$ no es trivial. La matriz $A^T A$ es cuadrada, y es invertible si y solo si su espacio nulo $\text{Nul}(A^T A) \neq \emptyset$ no es trivial. Veamos que $\text{Nul}(A^T A) = \text{Nul } A$, por lo que el teorema quedará demostrado. Si $\mathbf{0} = A^T A\mathbf{x}$, entonces $\mathbf{0} = \mathbf{x}^T A^T A\mathbf{x} = \|A\mathbf{x}\|^2$, luego $\text{Nul}(A^T A) \subset \text{Nul } A$. Y evidentemente $\text{Nul } A \subset \text{Nul}(A^T A)$, luego son dos conjuntos iguales. $\square$

La pseudoinversa de Moore-Penrose y el proyector ortogonal

Cuando tenemos un sistema $A\mathbf{x} = \mathbf{b}$ inconsistente se puede encontrar la solución por mínimos cuadrados resolviendo las ecuaciones normales $A^T A\hat{\mathbf{x}} = A^T \mathbf{b}$. En el caso en que las columnas de A son linealmente independientes, $A^T A$ es invertible, y podemos encontrar la solución de las ecuaciones normales invirtiendo esa matriz:

$$A\hat{\mathbf{x}} = \mathbf{b} \Rightarrow \hat{\mathbf{x}} = (A^T A)^{-1} A^T \mathbf{b} \quad \text{si } \exists (A^T A)^{-1}$$

A la matriz $A^+ = (A^T A)^{-1} A^T$ que da la solución de mínimos cuadrados se la denomina *pseudoinversa* de A

$$A\hat{\mathbf{x}} = \mathbf{b} \Rightarrow \hat{\mathbf{x}} = A^+ \mathbf{b}$$

La pseudoinversa A^+ es una *inversa por la izquierda* de A

$$A^+ A = (A^T A)^{-1} A^T A = I$$

Si A es cuadrada e invertible, la pseudoinversa coincide con la inversa

$$(A^T A)^{-1} A^T = A^{-1} (A^T)^{-1} A^T = A^{-1}$$

Se puede obtener una fórmula para la proyección $\hat{\mathbf{b}} = \text{proy}_{\text{Col } A} \mathbf{b}$ en el caso de que A tenga columnas independientes. En efecto, según (7.12) $\hat{\mathbf{b}} = A\hat{\mathbf{x}}$. Por tanto

$$\text{proy}_{\text{Col } A} \mathbf{b} = A(A^T A)^{-1} A^T \mathbf{b}$$

La matriz $A(A^T A)^{-1} A^T$ es el *proyector ortogonal* sobre $\text{Col } A$, cuando las columnas de A forman una base de $\text{Col } A$ (son linealmente independientes). Es interesante observar que esta fórmula generaliza (7.11) aplicable en el caso de que la base de $\mathcal{W} = \text{Col } A$ sea ortonormal. Efectivamente, si este es el caso entonces $A^T A = I$, y el proyector se convierte en

$$A(A^T A)^{-1} A^T = AA^T$$

7.6. Regresión lineal

Un caso típico es de la medición de un sistema que sabemos reacciona de forma lineal a un estímulo. Por ejemplo, la determinación de una resistencia R mediante mediciones de las intensidades que la atraviesan $i_1, i_2, \dots, i_n$ para distintos valores del voltaje aplicado $v_1, v_2, \dots, v_n$. Estas mediciones conducen a un sistema de ecuaciones $i_1 R = v_1, i_1 R = v_2, \dots, i_n R = v_n$ con una sola

incógnita, de matriz ampliada $\begin{bmatrix} i_1 & v_1 \\ \vdots & \vdots \\ i_n & v_n \end{bmatrix}$. El sistema difícilmente será compatible, pero nosotros necesitamos determinar de algún modo la resistencia.

$$X\beta = \mathbf{y} \Leftrightarrow \begin{bmatrix} i_1 \\ \vdots \\ i_n \end{bmatrix} R = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}$$

Lo podemos hacer por mínimos cuadrados:

$$\begin{aligned} X^T X &= \begin{bmatrix} i_1 & \cdots & i_n \end{bmatrix} \begin{bmatrix} i_1 \\ \vdots \\ i_n \end{bmatrix} = i_1^2 + \cdots + i_n^2 \\ X^T \mathbf{y} &= \begin{bmatrix} i_1 & \cdots & i_n \end{bmatrix} \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} = i_1 v_1 + \cdots + i_n v_n \\ X^T X \beta &= X^T \mathbf{y} \Leftrightarrow (i_1^2 + \cdots + i_n^2) R = i_1 v_1 + \cdots + i_n v_n \end{aligned}$$

y la solución es

$$R = \frac{i_1 v_1 + \cdots + i_n v_n}{i_1^2 + \cdots + i_n^2} \neq \frac{1}{n} \left(\frac{v_1}{i_1} + \cdots + \frac{v_n}{i_n} \right)$$

En general, un problema de regresión lineal de ajuste de la ley $y = \beta_0 + \beta_1 x$ se puede plantear matricialmente como

$$X\beta = \mathbf{y}, \quad \text{con} \quad X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

y esto es un problema de mínimos cuadrados. El error cuadrático que se minimiza es

$$\|\mathbf{y} - X\beta\|^2 = (y_1 - \beta_0 - \beta_1 x_1)^2 + \cdots + (y_n - \beta_0 - \beta_1 x_n)^2.$$

Las ecuaciones normales son $X^T X\beta = X^T \mathbf{y}$ y, si los x_i no son todos iguales, hay solución única:

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

$$\begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum (x_i)^2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix},$$

$$\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \frac{1}{n \sum (x_i)^2 - (\sum x_i)^2} \begin{bmatrix} \sum (x_i)^2 & -\sum x_i \\ -\sum x_i & n \end{bmatrix} \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix}$$

con lo que la recta apropiada es

$$y = \left\{ \frac{n \sum x_i y_i - (\sum x_i)(\sum y_i)}{n \sum (x_i)^2 - (\sum x_i)^2} \right\} x + \left\{ \frac{(\sum (x_i)^2)(\sum y_i) - (\sum x_i)(\sum x_i y_i)}{n \sum (x_i)^2 - (\sum x_i)^2} \right\}.$$

Por ejemplo, un modelo de un sistema que se comporta aproximadamente de un modo lineal consiste en suponer que si $\mathbf{x}$ es un vector que codifica las entradas del sistema, la salida o respuesta es, según este modelo lineal, un vector $\mathbf{y} = A\mathbf{x}$ donde A es una matriz característica del sistema. En la práctica desconocemos esta matriz, y la tenemos que determinar realizando mediciones de diversas salidas $\mathbf{y}$ que son respuestas a correspondientes entradas $\mathbf{x}$. Introduciendo suficientes entradas $\mathbf{x}_1, \dots, \mathbf{x}_p$ obtenemos, con un proceso de medición sobre el sistema, salidas $\mathbf{y}_1, \dots, \mathbf{y}_p$ correspondientes y con ellas podemos plantear un sistema de ecuaciones

$$\mathbf{y}_1 = A\mathbf{x}_1, \mathbf{y}_2 = A\mathbf{x}_2, \dots, \mathbf{y}_p = A\mathbf{x}_p \quad (7.13)$$

Si $\mathbf{x}$ tiene n componentes, y A es de $m \times n$, las relaciones anteriores son $p \times n$ ecuaciones sobre los coeficientes a_{ij} de A . Si hacemos $p = n$ mediciones, podemos escribir esas ecuaciones en forma matricial

$$\begin{bmatrix} \mathbf{y}_1 & \mathbf{y}_2 & \cdots & \mathbf{y}_n \end{bmatrix} = A \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_n \end{bmatrix} \Leftrightarrow Y = AX$$

y resolver para la matriz $A = YX^{-1}$. Este procedimiento puede presentar muchos inconvenientes: hay errores en la determinación de las entradas $\mathbf{x}_i$ y las salidas $\mathbf{y}_i$. No sería muy lógico elegir dos entradas $\mathbf{x}_i$ y $\mathbf{x}_j$ iguales ($\mathbf{x}_i = \mathbf{x}_j$ para dos mediciones distintas, pero si lo hiciéramos es muy posible que las salidas correspondientes $\mathbf{y}_i, \mathbf{y}_j$ fueran (ligeramente) distintas, con lo cual el sistema sería inconsistente (ver (7.13)). Lo más habitual es realizar muchas más mediciones de las necesarias, $p > n$, con lo cual la matriz X

7.7. Respuestas a los ejercicios

7.14 Como todos los vectores de $\mathcal{W}^\perp$ son ortogonales a todos los vectores de $\mathcal{W}$, $\mathbf{v}$ tiene que ser ortogonal a sí mismo: $\langle \mathbf{v}, \mathbf{v} \rangle = 0$. Eso implica, por la propiedad correspondiente del producto interno, que $\mathbf{v} = \mathbf{0}$.

7.29 Como $U^T = U^{-1}$ no sólo será $U^T U = I$, sino también $U U^T = I$.

7.8. Resumen

Definición (Producto interno). Para todo $\mathbf{u}, \mathbf{v}, \mathbf{w}$ en $\mathcal{V}$ y c escalar:

1. $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$
2. $\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle$
3. $\langle c\mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{u}, c\mathbf{v} \rangle = c\langle \mathbf{v}, \mathbf{u} \rangle$
4. $\langle \mathbf{u}, \mathbf{u} \rangle \geq 0$ y $\langle \mathbf{u}, \mathbf{u} \rangle = 0$ solo cuando $\mathbf{u} = \mathbf{0}$.

Definición.

1. La **longitud o norma**

$$\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle} = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}$$

2. La **distancia** entre $\mathbf{u}$ y $\mathbf{v}$

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

3. Vectores **ortogonales** $\mathbf{u} \perp \mathbf{v}$

$$\langle \mathbf{u}, \mathbf{v} \rangle = 0 \quad (\mathbf{u} \cdot \mathbf{v} = 0)$$

4. **Ángulo** entre dos vectores

$$\cos \theta = \frac{\mathbf{v} \cdot \mathbf{w}}{\|\mathbf{v}\| \|\mathbf{w}\|}, \quad 0 \leq \theta \leq \pi.$$

Teorema (Pitágoras). $\mathbf{u} \perp \mathbf{v}$ si y sólo si $\|\mathbf{u} + \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$.

Definición (Complemento ortogonal). Si $\mathcal{W} \subset \mathcal{V}$

$$\mathcal{W}^\perp = \{\mathbf{x} \in \mathcal{V} : \mathbf{x} \text{ es ortogonal a } \mathcal{W}\}$$

Proposición.

1. $\mathbf{x} \in \mathcal{W}^\perp$ si y sólo si $\mathbf{x}$ es ortogonal a todos los vectores de un conjunto de generadores (cualquiera) de $\mathcal{W}$.
2. $\mathcal{W}^\perp$ es un subespacio de $\mathcal{V}$.

Teorema. A de $m \times n$.

$$(\text{Fil } A)^\perp = \text{Nul } A \quad \text{y} \quad (\text{Col } A)^\perp = \text{Nul } A^T$$

Teorema. $S = \{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ ortogonal, Para todo $\mathbf{y} \in \mathcal{V}$
 $\mathbf{u}_i \neq \mathbf{0} \forall i \Rightarrow S$ linealmente independiente.

Teorema. Sea $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ una base ortogonal de $\mathcal{W}$. Entonces $\forall \mathbf{y} \in \mathcal{W}$

$$\mathbf{y} = \frac{\mathbf{y} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1} \mathbf{u}_1 + \cdots + \frac{\mathbf{y} \cdot \mathbf{u}_p}{\mathbf{u}_p \cdot \mathbf{u}_p} \mathbf{u}_p$$

$$\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z}$$

con $\hat{\mathbf{y}}$ en $\mathcal{W}$ y $\mathbf{z}$ en $\mathcal{W}^\perp$. Si $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ es una base ortogonal de $\mathcal{W}$, entonces

$$\hat{\mathbf{y}} = \frac{\mathbf{y} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1} \mathbf{u}_1 + \cdots + \frac{\mathbf{y} \cdot \mathbf{u}_p}{\mathbf{u}_p \cdot \mathbf{u}_p} \mathbf{u}_p$$

Teorema (La descomposición ortogonal). Sea $\mathcal{W} \subset \mathcal{V}$ un subespacio de $\mathcal{V}$. Se dice que $\text{proy}_{\mathcal{W}} \mathbf{y} = \hat{\mathbf{y}}$ es la **proyección ortogonal de $\mathbf{y}$ sobre $\mathcal{W}$** .