

Ortogonalidad

7.1. Producto escalar y ortogonalidad

Lección 9,
18 abr 2023

Longitud, norma o módulo de vectores y distancias entre puntos

Generalizando la fórmula pitagórica de la longitud de un vector de \mathbf{R}^2 o de \mathbf{R}^3 , definimos la *norma, longitud o módulo* de un vector $\mathbf{v} = (v_1, \dots, v_n)$ en \mathbf{R}^n mediante la fórmula

$$\|\mathbf{v}\| = \sqrt{v_1^2 + \dots + v_n^2}.$$

La “función norma” definida en \mathbf{R}^n tiene todas las propiedades algebraicas del módulo de un vector del plano o del espacio, a saber:

Propiedades de la norma:

1. $\|\mathbf{v}\| \geq 0$, siendo $\|\mathbf{v}\| = 0$ solamente en el caso $\mathbf{v} = \mathbf{0}$.
2. $\|c\mathbf{v}\| = |c|\|\mathbf{v}\|$.
3. $\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$ (*desigualdad triangular*).

Longitudes de vectores y distancias

La posibilidad de calcular longitudes de vectores de un espacio vectorial es equivalente a la de calcular distancias entre pares de puntos de ese espacio vectorial ya que, por un lado, la longitud de un vector es la distancia desde el punto representado por ese vector al origen de coordenadas (toda longitud de un vector es una distancia) y, por otro lado, la distancia entre los puntos representados por dos vectores \mathbf{u}, \mathbf{v} es igual a la longitud del vector diferencia $\mathbf{u} - \mathbf{v}$ (toda distancia es la longitud de un vector). Así, por el mero hecho de poder calcular la norma de cada vector, podemos definir la distancia entre dos vectores cualesquiera:

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

Vectores unitarios. Un vector de longitud (o norma, o módulo) igual a 1 se llama un *vector unitario*. Dado que la norma de un vector es un número no negativo, las siguientes ecuaciones son expresiones equivalentes de que \mathbf{u} es un vector unitario:

$$\|\mathbf{u}\| = 1, \quad \|\mathbf{u}\|^2 = 1.$$

El vector
normalizado de
otro.

La idea más importante que se debe tener en mente en relación con los vectores unitarios es el hecho de que todo vector no nulo $\mathbf{v} \neq \mathbf{0}$ se puede reescalar para convertirlo en un vector unitario. Esto se conoce como *normalizar* el vector. Si \mathbf{v} es un vector no nulo, el vector unitario $\mathbf{u} = \frac{1}{\|\mathbf{v}\|} \mathbf{v}$ es el normalizado de \mathbf{v} .

7.1.1 Ejercicio de tarea. Demuestra, usando únicamente las propiedades de la norma, que para todo vector no nulo \mathbf{v} , el vector $\mathbf{u} = \frac{1}{\|\mathbf{v}\|} \mathbf{v}$ es un vector unitario.

$$\text{Solución: } \|\mathbf{u}\| = \left\| \frac{\|\mathbf{u}\|}{1} \mathbf{u} \right\| = \left| \frac{\|\mathbf{u}\|}{1} \right| \|\mathbf{u}\| = \frac{\|\mathbf{u}\|}{1} \|\mathbf{u}\| = 1.$$

Producto escalar

A continuación vamos a ver que la posibilidad de calcular longitudes de vectores de un espacio vectorial real se puede obtener como consecuencia de poder realizar una operación algebraica con vectores llamada *producto escalar*, la cual es conocida por los estudiantes de Física para los vectores del plano y del espacio.

Empezamos observando que el cuadrado de la norma de un vector de \mathbf{R}^n ,

$$\|\mathbf{v}\|^2 = v_1^2 + \cdots + v_n^2$$

se puede obtener como resultado de una operación que ya conocemos, a saber: la multiplicación de matrices. Efectivamente, $\|\mathbf{v}\|^2$ es igual al producto de la matriz $1 \times n$ obtenida al hallar la traspuesta de \mathbf{v} (considerado como un vector columna) por el propio vector \mathbf{v} (como matriz $n \times 1$):

$$\mathbf{v}^T \mathbf{v} = (v_1 \cdots v_n) \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = v_1^2 + \cdots + v_n^2 = \|\mathbf{v}\|^2.$$

producto escalar

Esta propiedad nos lleva definir la operación de *producto escalar* de dos vectores de \mathbf{R}^n mediante

$$\mathbf{v} \cdot \mathbf{u} = \mathbf{u}^T \mathbf{v} = (u_1 \cdots u_n) \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = u_1 v_1 + \cdots + u_n v_n \quad (7.1)$$

La norma euclidiana de los vectores se expresa en términos del producto escalar mediante

$$\|\mathbf{v}\| = \sqrt{\mathbf{v} \cdot \mathbf{v}}. \quad (7.2)$$

Propiedades del producto escalar en \mathbf{R}^n

La definición (7.1) del producto escalar en \mathbf{R}^n nos permite demostrar muy fácilmente las siguientes propiedades:

1. $(c\mathbf{u}) \cdot \mathbf{v} = c(\mathbf{u} \cdot \mathbf{v})$.
2. $(\mathbf{u} + \mathbf{w}) \cdot \mathbf{v} = \mathbf{u} \cdot \mathbf{v} + \mathbf{w} \cdot \mathbf{v}$.
3. $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$ (conmutatividad o simetría).
4. $\mathbf{u} \cdot \mathbf{u} \geq 0$ siendo $\mathbf{u} \cdot \mathbf{u} = 0$ solamente en el caso $\mathbf{u} = \mathbf{0}$ (definida positiva).

En vista de la tercera propiedad, las dos primeras son equivalentes a (cada una de) las siguientes:

$$1' \quad (cu) \cdot v = c(u \cdot v) \text{ y } (u + w) \cdot v = u \cdot v + w \cdot v.$$

(“linealidad en la primera variable”: la aplicación $(\cdot) \cdot v$ definida por $x \mapsto x \cdot v$ es lineal).

$$2' \quad u \cdot (cv) = c(u \cdot v) \text{ y } u \cdot (v + w) = u \cdot v + u \cdot w.$$

(“linealidad en la segunda variable”: la aplicación $u \cdot (\cdot)$ definida por $x \mapsto u \cdot x$ es lineal).

Al expresar de esta forma las propiedades características del producto escalar en \mathbf{R}^n , vemos que se puede decir que el producto escalar en \mathbf{R}^n es una función real de dos variables vectoriales que es *bilineal, simétrica y definida positiva*.

7.1.2 Ejercicio de tarea. Demuestra que para cualquier matriz A , el producto $A^T A$ es una matriz cuadrada simétrica.

$$\text{Solución: } (\mathbf{v}_i^T \mathbf{v}_j)^T = \mathbf{v}_j^T (\mathbf{v}_i)^T = \mathbf{v}_j^T \mathbf{v}_i^T$$

Consecuencia de la propiedad “definida positiva”: Para cualquier matriz A , la matriz cuadrada simétrica $A^T A$ tiene el mismo espacio nulo que A .

Por un lado, es evidente que los vectores del espacio nulo de A son también vectores del espacio nulo de $A^T A$ ya que si $Ax = \mathbf{0}$ entonces también $A^T Ax = \mathbf{0}$.

Supongamos ahora que x es un vector del espacio nulo de $A^T A$, es decir, $A^T Ax = \mathbf{0}$. Entonces $x^T A^T Ax = 0$. Pero esto significa que el vector Ax tiene norma cero porque el miembro de la izquierda en la última expresión es igual al cuadrado de la norma de Ax :

$$x^T A^T Ax = (Ax)^T Ax = \|Ax\|^2.$$

Así pues, Ax tiene norma cero y por la propiedad de la norma de ser definida positiva, el propio Ax es el vector cero, lo que significa que x pertenece al espacio nulo de A .

Corolario 1: Para cualquier matriz A , la matriz cuadrada simétrica $A^T A$ tiene el mismo rango que A .

Esto se debe a que el rango es igual al número de columnas pivote, que a su vez es igual al número de columnas menos el número de columnas sin pivote. Como el número de columnas sin pivote es la dimensión del espacio nulo y $A^T A$ tiene el mismo número de columnas que A y el mismo espacio nulo, se deduce que los rangos son iguales.

Corolario 2: Si A es una matriz $m \times n$, las n columnas de A son vectores linealmente independientes en \mathbf{R}^m si y sólo si las n columnas de $A^T A$ son independientes (es decir, forman una base de \mathbf{R}^n).

Este corolario es una consecuencia clara del anterior, pero también se puede deducir directamente porque decir que las columnas de una matriz son linealmente independientes es lo mismo que decir que el espacio nulo de esa matriz es el espacio cero (que sólo contiene al vector nulo). Otra forma de expresar este corolario es la siguiente:

Una matriz A tiene una inversa por la izquierda si y sólo si la matriz $A^T A$ es inversible.

Perpendicularidad

La posibilidad de calcular la longitud de los vectores nos permite definir el concepto de “perpendicularidad” y por lo tanto de “ángulo recto”. Además, la posibilidad de calcular el producto escalar de dos vectores cualesquiera nos permite definir también el *ángulo* que forman dos vectores no nulos cualesquiera. En lo que sigue supondremos implícitamente que los vectores pertenecen a \mathbf{R}^n .

Vectores perpendiculares

Definición de perpendicularidad u ortogonalidad.

Definición: Diremos que dos vectores \mathbf{u}, \mathbf{v} son *perpendiculares* u *ortogonales* si las dos diagonales, $\mathbf{u} + \mathbf{v}$ y $\mathbf{u} - \mathbf{v}$, del paralelogramo que forman tienen la misma longitud, y escribiremos

$$\mathbf{u} \perp \mathbf{v} \quad \text{si y sólo si} \quad \|\mathbf{u} + \mathbf{v}\| = \|\mathbf{u} - \mathbf{v}\|.$$

Dado que la norma o longitud de vectores se puede expresar en términos del producto escalar, normas al cuadrado

$$\|\mathbf{u} + \mathbf{v}\|^2 = (\mathbf{u} + \mathbf{v}) \cdot (\mathbf{u} + \mathbf{v}) = \mathbf{u} \cdot \mathbf{u} + \mathbf{v} \cdot \mathbf{v} + 2\mathbf{u} \cdot \mathbf{v} \quad (7.3)$$

$$\|\mathbf{u} - \mathbf{v}\|^2 = (\mathbf{u} - \mathbf{v}) \cdot (\mathbf{u} - \mathbf{v}) = \mathbf{u} \cdot \mathbf{u} + \mathbf{v} \cdot \mathbf{v} - 2\mathbf{u} \cdot \mathbf{v} \quad (7.4)$$

Por tanto podemos escribir:

$$\mathbf{u} \perp \mathbf{v} \quad \text{si y sólo si} \quad \mathbf{u} \cdot \mathbf{v} = 0.$$

Así, usando el producto escalar, es muy sencillo saber si dos vectores de \mathbf{R}^n son perpendiculares: si su producto escalar es cero lo son, en caso contrario no lo son.

7.1.3 Ejercicio de tarea. ¿Cuál de las propiedades del producto escalar implica que el único vector ortogonal a sí mismo es el vector cero?

Solución: La propiedad de "definida positiva".

7.1.4 Ejercicio de tarea. Demuestra que si \mathbf{x} e \mathbf{y} son vectores no nulos tales que $\mathbf{x} = \lambda \mathbf{y}$ entonces \mathbf{x} e \mathbf{y} no pueden ser ortogonales.

Solución: $\mathbf{x} \cdot \mathbf{y} = \lambda \mathbf{y} \cdot \mathbf{y} = \lambda \|\mathbf{y}\|^2 \neq 0$

7.1.5 Ejercicio de tarea. Demuestra que si $\mathbf{x} = (x_1, \dots, x_n)$ e $\mathbf{y} = (y_1, \dots, y_n)$ son vectores no nulos de \mathbf{R}^n "del mismo cuadrante" (en el sentido de que para cada i el signo de x_i es igual al de y_i) entonces \mathbf{x} e \mathbf{y} no pueden ser ortogonales.

Solución: $\mathbf{x} \cdot \mathbf{y} = \sum x_i y_i$ es una suma de sumandos no negativos y no todos nulos, luego no puede ser cero.

7.1.6 Ejercicio de tarea. Demuestra que si $\mathbf{x} = (x_1, \dots, x_n)$ e $\mathbf{y} = (y_1, \dots, y_n)$ son vectores no nulos de \mathbf{R}^n "de cuadrantes opuestos" (en el sentido de que para cada i el signo de x_i es opuesto al de y_i) entonces \mathbf{x} e \mathbf{y} no pueden ser ortogonales.

Solución: $\mathbf{x} \cdot \mathbf{y} = \sum x_i y_i$ es una suma de sumandos negativos o nulos (pero no todos nulos), luego no puede ser cero.

El ejemplo fundamental de vectores perpendiculares es el los vectores de la base canónica de \mathbf{R}^n , que son las columnas $\mathbf{e}_1, \dots, \mathbf{e}_n$ de la matriz identidad $n \times n$.

Componentes longitudinal y transversal de un vector respecto a una dirección dada

Proyección ortogonal de un vector sobre la dirección determinada por otro

proyección ortogonal

Se llama *proyección ortogonal* de un vector \mathbf{v} sobre (la recta definida por) un vector no nulo \mathbf{u} , y se denota por $\text{proy}_{\mathbf{u}}(\mathbf{v})$, a aquél múltiplo de \mathbf{u} que restado de \mathbf{v} da como resultado un vector perpendicular a \mathbf{u} . Es decir: un múltiplo de \mathbf{u} , $c\mathbf{u}$, es la proyección ortogonal de \mathbf{v} sobre \mathbf{u} si el vector diferencia $\mathbf{v} - c\mathbf{u}$ es perpendicular a \mathbf{u} , es decir, si

$$(\mathbf{v} - c\mathbf{u}) \cdot \mathbf{u} = 0.$$

De esta ecuación se deduce:

$$c = \frac{\mathbf{v} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}} \quad \text{y por tanto:} \quad \text{proy}_{\mathbf{u}}(\mathbf{v}) = \left(\frac{\mathbf{v} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}} \right) \mathbf{u}. \quad (7.5)$$

Componentes longitudinal y transversal

La proyección ortogonal nos permite calcular la descomposición de un vector dado en suma de dos vectores que sean uno de ellos paralelo y el otro perpendicular a una dirección dada. Sea \mathbf{v} un vector dado y supongamos dada una dirección (o subespacio vectorial unidimensional) mediante un vector \mathbf{u} . Las componentes *longitudinal* (paralela) y *transversal* (perpendicular) de \mathbf{v} respecto a la dirección de \mathbf{u} son dos vectores denotados respectivamente \mathbf{v}_{\parallel} y \mathbf{v}_{\perp} tales que

$$\mathbf{v}_{\parallel} + \mathbf{v}_{\perp} = \mathbf{v}, \quad \mathbf{v}_{\parallel} = c\mathbf{u} \quad \text{y} \quad \mathbf{v}_{\perp} \cdot \mathbf{u} = 0$$

En consecuencia, la componente longitudinal es la proyección ortogonal y la componente transversal es la diferencia entre el vector dado y su proyección ortogonal:

$$\mathbf{v}_{\parallel} = \text{proy}_{\mathbf{u}}(\mathbf{v}), \quad \mathbf{v}_{\perp} = \mathbf{v} - \text{proy}_{\mathbf{u}}(\mathbf{v}).$$

Ángulo entre dos vectores y la fórmula del coseno para el producto escalar

Sabiendo calcular la proyección ortogonal de un vector sobre otro, es muy sencillo medir el ángulo entre dos vectores no nulos \mathbf{u} , \mathbf{v} . Para ello basta hallar la proyección ortogonal de uno de ellos sobre el otro: $\text{proy}_{\mathbf{u}}(\mathbf{v})$ y observar que se obtiene un triángulo rectángulo en el que \mathbf{v} es la “hipotenusa” y las componentes longitudinal y transversal de \mathbf{v} son los “catetos”. Suponiendo $\mathbf{v} \cdot \mathbf{u} > 0$, el ángulo entre \mathbf{u} y \mathbf{v} es el que tiene a $\mathbf{v}_{\parallel} = \text{proy}_{\mathbf{u}}(\mathbf{v})$ como “cateto adyacente”, por lo que la definición “cateto adyacente partido por la hipotenusa” nos dice que su coseno es:

$$\cos \alpha = \frac{\|\text{proy}_{\mathbf{u}}(\mathbf{v})\|}{\|\mathbf{v}\|} = \frac{1}{\|\mathbf{v}\|} \left\| \left(\frac{\mathbf{v} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}} \right) \mathbf{u} \right\| = \frac{\mathbf{v} \cdot \mathbf{u}}{\|\mathbf{v}\| \|\mathbf{u}\|^2} \|\mathbf{u}\| = \frac{\mathbf{v} \cdot \mathbf{u}}{\|\mathbf{v}\| \|\mathbf{u}\|}.$$

Esto concuerda con la conocida fórmula de los libros de Física para el producto escalar:

$$\mathbf{v} \cdot \mathbf{u} = \|\mathbf{v}\| \|\mathbf{u}\| \cos \alpha,$$

la cual es válida para cualesquiera vectores y no sólo para los que cumplen $\mathbf{v} \cdot \mathbf{u} > 0$.

Complemento ortogonal

DEFINICIÓN 7.1.1

Vector ortogonal a un subconjunto de \mathbb{R}^n

Dado un subconjunto S de vectores de \mathbb{R}^n , se dice que un vector $\mathbf{v} \in \mathbb{R}^n$ es ortogonal a S si es ortogonal a todos y cada uno de los vectores de S .

Si \mathbf{v} es un vector de \mathbb{R}^n , entonces evidentemente *todo* vector de \mathbb{R}^n que sea ortogonal a \mathbf{v} es también ortogonal a todos los múltiplos de \mathbf{v} . Esto implica que cada vector ortogonal a \mathbf{v} es ortogonal al subespacio generado por \mathbf{v} . Esta propiedad sigue siendo cierta si en lugar de un vector \mathbf{v} nos dan dos o más vectores $\mathbf{v}_1, \dots, \mathbf{v}_k$:

PROPOSICIÓN 7.1.1

Si $S = \{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ un subconjunto de vectores de \mathbb{R}^n y un vector $\mathbf{w} \in \mathbb{R}^n$ es ortogonal a S , entonces \mathbf{w} es ortogonal al subespacio $\text{Gen}\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ generado por S .

Demostración:

Hay que demostrar que si $\mathbf{v}_1 \cdot \mathbf{w} = \dots = \mathbf{v}_k \cdot \mathbf{w} = 0$, entonces \mathbf{w} es ortogonal a toda combinación lineal de los \mathbf{v}_i . Esto es inmediato:

$$(c_1 \mathbf{v}_1 + \dots + c_k \mathbf{v}_k) \cdot \mathbf{w} = c_1 \mathbf{v}_1 \cdot \mathbf{w} + \dots + c_k \mathbf{v}_k \cdot \mathbf{w} = c_1 \times 0 + \dots + c_k \times 0 = 0.$$

Ejemplo: Si \mathbf{w} es una solución de un sistema homogéneo $A\mathbf{x} = \mathbf{0}$ entonces \mathbf{w} es ortogonal a cada fila de A . En consecuencia es ortogonal al espacio fila de A .

PROPOSICIÓN 7.1.2

Si W es un subespacio vectorial de \mathbf{R}^n , el conjunto de todos los vectores $\mathbf{w} \in \mathbf{R}^n$ que son ortogonales a W es un subespacio vectorial de \mathbf{R}^n .

La demostración de esta última proposición se deja como ejercicio (ver el ejercicio 29).

DEFINICIÓN 7.1.2

Complemento ortogonal de un subespacio

Se llama complemento ortogonal de un subespacio vectorial W de \mathbf{R}^n , y se denota W^\perp , al subespacio de \mathbf{R}^n formado por todos los vectores de \mathbf{R}^n que son ortogonales a W .

Ejemplos:

- (1) El complemento ortogonal del subespacio cero de \mathbf{R}^n es todo \mathbf{R}^n : $\{\mathbf{0}\}^\perp = \mathbf{R}^n$.
- (2) El complemento ortogonal de \mathbf{R}^n es el subespacio cero: $(\mathbf{R}^n)^\perp = \{\mathbf{0}\}$.
- (3) El complemento ortogonal de un plano por el origen de \mathbf{R}^3 es la recta que pasa por el origen y es perpendicular a ese plano.
- (4) Si W es un subespacio vectorial de \mathbf{R}^n y $U = W^\perp$, entonces $U^\perp = W$. Dicho de otra forma:

$$(W^\perp)^\perp = W.$$

TEOREMA 7.1.1

Si A es una matriz $m \times n$, el complemento ortogonal del espacio fila de A es el espacio nulo de A :

$$(\text{Fil } A)^\perp = \text{Nul } A.$$

Demostración:

Un vector cualquiera del espacio nulo es una solución del sistema homogéneo $A\mathbf{x} = \mathbf{0}$, por tanto satisface cada ecuación, lo que significa que es ortogonal a cada fila de A . Recíprocamente, si un vector es ortogonal a todas las filas de A , entonces cumple cada ecuación del sistema homogéneo $A\mathbf{x} = \mathbf{0}$ y por tanto es una solución del sistema y es un vector del espacio nulo.

Poniendo A^T en lugar de A en el teorema anterior y teniendo en cuenta que el espacio fila de A^T es lo mismo que el espacio columna de A , se obtiene:

COROLARIO 7.1.1

Si A es una matriz $m \times n$, el complemento ortogonal del espacio columna de A es el espacio nulo de A^T :

$$(\text{Col } A)^\perp = \text{Nul } A^T.$$

Justificación del nombre «complemento» ortogonal: Si W es un subespacio de \mathbf{R}^n y A es una matriz cuyas columnas son una base de W (de forma que $W = \text{Col } A$), entonces $W^\perp = \text{Nul } A^T$, lo que significa que $\dim W + \dim W^\perp = n$. Por otra parte, el único vector que pertenece a W y a W^\perp es el cero (porque es el único vector ortogonal a sí mismo); en consecuencia $W \cap W^\perp = \{\mathbf{0}\}$.

Enlaces a los ejercicios de tarea de esta sección

Usa los siguientes enlaces para visualizar cada uno de los ejercicios de tarea que aparecen en esta sección:

Enlaces: [Ejercicio 1](#), [Ejercicio 2](#), [Ejercicio 3](#), [Ejercicio 4](#), [Ejercicio 5](#), [Ejercicio 6](#).

Ejercicios de la sección 7.1 Producto escalar y ortogonalidad

En los ejercicios 1 a 8, calcula las cantidades indicadas usando los siguientes vectores:

$$\mathbf{u} = \begin{pmatrix} -1 \\ 2 \end{pmatrix}, \mathbf{v} = \begin{pmatrix} 4 \\ 6 \end{pmatrix}, \mathbf{w} = \begin{pmatrix} 3 \\ -1 \\ -5 \end{pmatrix}, \mathbf{x} = \begin{pmatrix} 6 \\ -2 \\ 3 \end{pmatrix}$$

1. $\mathbf{u} \cdot \mathbf{u}$, $\mathbf{v} \cdot \mathbf{u}$ y $\frac{\mathbf{v} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}}$.

2. $\frac{1}{\mathbf{u} \cdot \mathbf{u}} \mathbf{u}$.

3. $\mathbf{w} \cdot \mathbf{w}$, $\mathbf{x} \cdot \mathbf{w}$ y $\frac{\mathbf{x} \cdot \mathbf{w}}{\mathbf{w} \cdot \mathbf{w}}$.

4. $\frac{1}{\mathbf{w} \cdot \mathbf{w}} \mathbf{w}$.

5. $\left(\frac{\mathbf{x} \cdot \mathbf{w}}{\mathbf{x} \cdot \mathbf{x}} \right) \mathbf{x}$.

6. $\left(\frac{\mathbf{u} \cdot \mathbf{v}}{\mathbf{v} \cdot \mathbf{v}} \right) \mathbf{v}$.

7. $\|\mathbf{w}\|$.

8. $\|\mathbf{x}\|$.

En los ejercicios 9 a 12, halla un vector unitario en la dirección del vector dado:

9. $\begin{pmatrix} -30 \\ 40 \end{pmatrix}$.

10. $\begin{pmatrix} -6 \\ 4 \\ -3 \end{pmatrix}$.

11. $\begin{pmatrix} 7/4 \\ 1/2 \\ 1 \end{pmatrix}$.

12. $\begin{pmatrix} 8/3 \\ 2 \end{pmatrix}$.

13. Calcula la distancia entre los vectores

$$\mathbf{x} = \begin{pmatrix} 10 \\ 3 \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} -1 \\ -5 \end{pmatrix}.$$

14. Calcula la distancia entre los vectores

$$\mathbf{u} = \begin{pmatrix} 0 \\ -5 \\ 2 \end{pmatrix}, \quad \mathbf{z} = \begin{pmatrix} -4 \\ -1 \\ 8 \end{pmatrix}.$$

En los ejercicios 15 a 18 averigua si los dos vectores dados son ortogonales:

15. $\mathbf{a} = \begin{pmatrix} 8 \\ -5 \end{pmatrix}, \mathbf{b} = \begin{pmatrix} -2 \\ -3 \end{pmatrix}$.

16. $\mathbf{u} = \begin{pmatrix} 12 \\ 3 \\ -5 \end{pmatrix}, \mathbf{v} = \begin{pmatrix} 2 \\ -3 \\ 3 \end{pmatrix}$.

17. $\mathbf{a} = \begin{pmatrix} 3 \\ 2 \\ -5 \\ 0 \end{pmatrix}, \mathbf{b} = \begin{pmatrix} -4 \\ 1 \\ -2 \\ 6 \end{pmatrix}$.

18. $\mathbf{u} = \begin{pmatrix} -3 \\ 7 \\ 4 \\ 0 \end{pmatrix}, \mathbf{v} = \begin{pmatrix} 1 \\ -8 \\ 15 \\ -7 \end{pmatrix}$.

En los ejercicios 19 y 20 todos los vectores son de \mathbf{R}^n . Indica para cada afirmación si es verdadera o falsa, justificando tus respuestas.

19.

(a) $\mathbf{v} \cdot \mathbf{v} = \|\mathbf{v}\|^2$.

(b) Para cualquier escalar c , se cumple $\mathbf{u} \cdot (c\mathbf{v}) = c(\mathbf{u} \cdot \mathbf{v})$.

(c) Si la distancia de \mathbf{u} a \mathbf{v} es igual a la distancia de \mathbf{u} a $-\mathbf{v}$ entonces \mathbf{u} y \mathbf{v} son ortogonales.

(d) Para cualquier matriz cuadrada A , los vectores de $\text{Col } A$ son ortogonales a los de $\text{Nul } A$.

(e) Si los vectores $\mathbf{v}_1, \dots, \mathbf{v}_p$ generan un subespacio W y si \mathbf{x} es ortogonal a cada \mathbf{v}_j para $j = 1, \dots, p$ entonces \mathbf{x} pertenece a W^\perp .

20.

- (a) $\mathbf{u} \cdot \mathbf{v} - \mathbf{v} \cdot \mathbf{u} = 0$.
- (b) Para cualquier escalar c , se cumple $\|c\mathbf{v}\| = |c|\|\mathbf{v}\|$.
- (c) Si \mathbf{x} es ortogonal a cada vector de un subespacio W , entonces \mathbf{x} pertenece a W^\perp .
- (d) Si $\|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 = \|\mathbf{u} + \mathbf{v}\|^2$ entonces \mathbf{u} y \mathbf{v} son ortogonales.
- (e) Para cualquier matriz $m \times n$ A , los vectores del espacio nulo de A son ortogonales a los vectores del espacio fila de A .

21. Sea $\mathbf{u} = (u_1, u_2, u_3)$. Explica por qué $\mathbf{u} \cdot \mathbf{u} \geq 0$. ¿En qué caso se cumpliría $\mathbf{u} \cdot \mathbf{u} = 0$?

22. Sean $\mathbf{u} = \begin{pmatrix} 2 \\ -5 \\ -1 \end{pmatrix}$ y $\mathbf{v} = \begin{pmatrix} -7 \\ -4 \\ 6 \end{pmatrix}$. Calcula $\|\mathbf{u}\|^2$, $\|\mathbf{v}\|^2$, $\|\mathbf{u} + \mathbf{v}\|^2$ y $\mathbf{u} \cdot \mathbf{v}$. Explica los resultados.

23. Demuestra la *ley del paralelogramo* para vectores \mathbf{u} y \mathbf{v} de \mathbb{R}^n :

$$\|\mathbf{u} + \mathbf{v}\|^2 + \|\mathbf{u} - \mathbf{v}\|^2 = 2\|\mathbf{u}\|^2 + 2\|\mathbf{v}\|^2.$$

24. Describe geoméricamente el conjunto H de los vectores $\begin{pmatrix} x \\ y \end{pmatrix}$ que son perpendiculares a un vector dado $\mathbf{v} = \begin{pmatrix} a \\ b \end{pmatrix}$ en \mathbb{R}^2 . Considera separadamente los casos $\mathbf{v} = \mathbf{0}$ y $\mathbf{v} \neq \mathbf{0}$.

25. Sea $\mathbf{u} = \begin{pmatrix} 5 \\ -6 \\ 7 \end{pmatrix}$ y sea W el conjunto de todos los vectores \mathbf{x} de \mathbb{R}^3 tales que $\mathbf{u} \cdot \mathbf{x} = 0$. Describe W geoméricamente. ¿Cuál es la matriz de la que W es el espacio nulo?

26. Supongamos que \mathbf{y} es un vector ortogonal a \mathbf{u} y a \mathbf{v} . Demuestra que \mathbf{y} es ortogonal a $\mathbf{u} + \mathbf{v}$.

27. Supongamos que \mathbf{y} es un vector ortogonal a \mathbf{u} y a \mathbf{v} . Demuestra que \mathbf{y} es ortogonal a todo vector \mathbf{w} de $\text{Gen}\{\mathbf{u}, \mathbf{v}\}$.

28. Supongamos que \mathbf{x} es un vector ortogonal a cada uno de los vectores $\mathbf{v}_1, \dots, \mathbf{v}_p$. Demuestra que \mathbf{x} es ortogonal a todo vector \mathbf{w} de $W = \text{Gen}\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$.

29. Sea W un subespacio de \mathbb{R}^n y sea W^\perp su *complemento ortogonal* (conjunto de todos los vectores ortogonales a W). Sigue los siguientes pasos para demostrar que W^\perp es un subespacio de \mathbb{R}^n :

- (a) Sea $\mathbf{z} \in W^\perp$ y sea \mathbf{u} un vector cualquiera de W . Entonces $\mathbf{z} \cdot \mathbf{u} = 0$. Si c es un escalar, demuestra que $c\mathbf{z}$ es ortogonal a \mathbf{u} . (Puesto que \mathbf{u} es arbitrario, esto demuestra que $c\mathbf{z}$ pertenece a W^\perp .)
- (b) Sean $\mathbf{z}_1, \mathbf{z}_2 \in W^\perp$ y sea \mathbf{u} un vector cualquiera de W . Demuestra que $\mathbf{z}_1 + \mathbf{z}_2$ es ortogonal a \mathbf{u} . (¿Qué se deduce de esto acerca de $\mathbf{z}_1 + \mathbf{z}_2$?)
- (c) Completa la demostración de que W^\perp es un subespacio de \mathbb{R}^n .

30. Demuestra que si \mathbf{x} pertenece a W y a W^\perp , entonces $\mathbf{x} = \mathbf{0}$.

7.2. Proyección ortogonal sobre un subespacio

Conjuntos ortogonales, vectores unitarios y conjuntos ortonormales

Se llama *conjunto ortogonal* de vectores a todo conjunto de vectores $\mathbf{u}_1, \dots, \mathbf{u}_k$ tales que cada uno es perpendicular (u ortogonal) a todos los demás, es decir, tales que $\mathbf{u}_i \cdot \mathbf{u}_j = 0$ siempre que $i \neq j$. Esto es lo mismo que decir que la matriz $A = [\mathbf{u}_1, \dots, \mathbf{u}_k]$ cuyas columnas son esos vectores tiene la propiedad de que la matriz cuadrada simétrica $A^T A$ es una *matriz diagonal*, pues los elementos del producto $A^T A$ son los productos escalares $\mathbf{u}_i \cdot \mathbf{u}_j$ de los vectores $\mathbf{u}_1, \dots, \mathbf{u}_k$.

7.2.1 Ejercicio de tarea. Demuestra que para cualquier matriz A , el elemento (i, j) de la matriz $A^T A$ es igual al producto escalar $\mathbf{u}_i \cdot \mathbf{u}_j$.

Solución: El elemento (i, j) de $A^T A$ es $\mathbf{u}_i^T \mathbf{u}_j = \mathbf{u}_i \cdot \mathbf{u}_j = \mathbf{u}_j \cdot \mathbf{u}_i$.

Si las columnas de A forman un conjunto ortogonal y son todas diferentes de $\mathbf{0}$ entonces $A^T A$ es una matriz diagonal cuyos elementos de la diagonal son todos diferentes de 0; por tanto es inversible. Como sabemos que las columnas de A son independientes si y sólo si $A^T A$ es inversible, tenemos:

Propiedad: Todo conjunto ortogonal de vectores no nulos, $\mathbf{u}_1, \dots, \mathbf{u}_k$, es un conjunto libre.

Conjuntos ortonormales. Un conjunto ortogonal de vectores, los cuales, además, son *unitarios* se llama un conjunto *ortonormal*. A partir de un conjunto ortogonal cualquiera de vectores no nulos se puede obtener un conjunto ortonormal sin más que reescalar cada vector apropiadamente.

Las columnas de una matriz U forman un conjunto ortonormal si y sólo si $U^T U = I$.

Matrices Ortogonales. Se llama *matriz ortogonal* a toda *matriz cuadrada* cuyas columnas son vectores unitarios ortogonales dos a dos, es decir una matriz cuadrada cuyas columnas forman un *conjunto ortonormal*.

¡Cuidado! Una matriz ortogonal **no es** aquella cuyas columnas forman un conjunto ortogonal. Ni siquiera aquella cuyas columnas forman un conjunto ortonormal. Es requisito de una matriz ortogonal que sea una matriz *cuadrada*.

Caracterización: Si A es una matriz ortogonal, $A^T A$ es una matriz identidad y por tanto A^T es la inversa de A . Recíprocamente: Si A es una matriz cuadrada tal que $A^T A = I$ entonces las columnas de A forman un conjunto ortonormal y por tanto A es una matriz ortogonal. En consecuencia:

Una matriz A es una matriz ortogonal si y sólo si es cuadrada y su traspuesta es su inversa.

Y también:

Las columnas de una matriz ortogonal $n \times n$ forman una base ortonormal de \mathbf{R}^n . Y recíprocamente: Si A tiene n filas y sus columnas forman una base ortonormal de \mathbf{R}^n , entonces A es una matriz ortogonal.

Proyección ortogonal sobre un subespacio

Al igual que ocurre con la descomposición de un vector obtenida a partir de su proyección ortogonal sobre un subespacio unidimensional, dado un vector \mathbf{v} y un subespacio W de \mathbf{R}^n , existe una única descomposición de \mathbf{v} en suma de dos vectores:

$$\mathbf{v} = \mathbf{v}_\perp + \mathbf{v}_\parallel$$

tales que el primero sea ortogonal a (cada vector de) W mientras que el segundo pertenece a W . La componente \mathbf{v}_\perp , ortogonal a W , se llama *la componente transversal* (u ortogonal) de \mathbf{v} respecto a W , mientras que la otra componente, \mathbf{v}_\parallel , contenida en el subespacio W , se llama *la componente longitudinal* (o paralela) de \mathbf{v} respecto a W y también la *proyección ortogonal* de \mathbf{v} sobre W .

Dada una base $\{\mathbf{u}_1, \dots, \mathbf{u}_r\}$ de W , es posible plantear un sistema de ecuaciones para calcular las coordenadas c_i de la proyección ortogonal de un vector cualquiera \mathbf{v} sobre W :

$$\mathbf{v}_\parallel = \text{proy}_W(\mathbf{v}) = c_1 \mathbf{u}_1 + \dots + c_r \mathbf{u}_r.$$

La condición que necesitamos aplicar para plantear dicho sistema de ecuaciones es que la diferencia $\mathbf{v} - \mathbf{v}_\parallel$ es perpendicular a cada vector de la base:

$$(\mathbf{v} - \mathbf{v}_\parallel) \cdot \mathbf{u}_i = 0.$$

Así pues, en el sistema de ecuaciones satisfecho por los coeficientes c_i , la ecuación j es:

$$(\mathbf{u}_1 \cdot \mathbf{u}_j) c_1 + \dots + (\mathbf{u}_r \cdot \mathbf{u}_j) c_r = \mathbf{v} \cdot \mathbf{u}_j \quad (7.6)$$

Para hallar las coordenadas c_1, \dots, c_r de la proyección ortogonal es, pues, necesario resolver un sistema de ecuaciones cuya matriz de coeficientes es la matriz simétrica:

$$\begin{pmatrix} \mathbf{u}_1 \cdot \mathbf{u}_1 & \dots & \mathbf{u}_1 \cdot \mathbf{u}_r \\ \vdots & \ddots & \vdots \\ \mathbf{u}_r \cdot \mathbf{u}_1 & \dots & \mathbf{u}_r \cdot \mathbf{u}_r \end{pmatrix}. \quad (7.7)$$

Cálculo de la proyección ortogonal conocida una base ortogonal

base ortogonal

Existe un caso en el que la matriz de coeficientes (7.7) es una matriz diagonal y por tanto la solución del sistema es inmediata. Nos referimos al caso en que todos los productos $\mathbf{u}_i \cdot \mathbf{u}_j$ fuera de la diagonal son cero; es decir, el caso en el que la base $\{\mathbf{u}_1, \dots, \mathbf{u}_r\}$ de W es una *base ortogonal*:

$$\mathbf{u}_i \cdot \mathbf{u}_j = 0 \quad \text{para } i \neq j.$$

En este caso, la ecuación (7.6) se reduce a:

$$(\mathbf{u}_j \cdot \mathbf{u}_j)c_j = \mathbf{v} \cdot \mathbf{u}_j$$

y por tanto las coordenadas de la proyección ortogonal son:

$$c_1 = \frac{\mathbf{v} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1}, \dots, c_r = \frac{\mathbf{v} \cdot \mathbf{u}_r}{\mathbf{u}_r \cdot \mathbf{u}_r}.$$

La fórmula de la proyección ortogonal queda entonces:

$$\mathbf{v}_{\parallel} = \text{proy}_W(\mathbf{v}) = \left(\frac{\mathbf{v} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1} \right) \mathbf{u}_1 + \dots + \left(\frac{\mathbf{v} \cdot \mathbf{u}_r}{\mathbf{u}_r \cdot \mathbf{u}_r} \right) \mathbf{u}_r.$$

Ejemplo: Hallar las componentes longitudinal y transversal del vector \mathbf{v} respecto al subespacio $W = \text{Gen}\{\mathbf{u}_1, \mathbf{u}_2\}$ de \mathbf{R}^3 donde:

$$\mathbf{v} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \quad \mathbf{u}_1 = \begin{pmatrix} 2 \\ 5 \\ -1 \end{pmatrix}, \quad \mathbf{u}_2 = \begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix}.$$

Solución: Empezamos comprobando que la base dada es ortogonal: calculamos el producto escalar $\mathbf{u}_1 \cdot \mathbf{u}_2 = -4 + 5 - 1 = 0$. Entonces hallamos la componente longitudinal que es la proyección ortogonal de \mathbf{v} sobre W :

$$\mathbf{v}_{\parallel} = \text{proy}_W(\mathbf{v}) = \left(\frac{\mathbf{v} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1} \right) \mathbf{u}_1 + \left(\frac{\mathbf{v} \cdot \mathbf{u}_2}{\mathbf{u}_2 \cdot \mathbf{u}_2} \right) \mathbf{u}_2.$$

Calculando los productos escalares: $\mathbf{v} \cdot \mathbf{u}_1 = 2 + 10 - 3 = 9$, $\mathbf{u}_1 \cdot \mathbf{u}_1 = 4 + 25 + 1 = 30$, $\mathbf{v} \cdot \mathbf{u}_2 = -2 + 2 + 3 = 3$, $\mathbf{u}_2 \cdot \mathbf{u}_2 = 4 + 1 + 1 = 6$, tenemos:

$$\mathbf{v}_{\parallel} = \frac{9}{30} \begin{pmatrix} 2 \\ 5 \\ -1 \end{pmatrix} + \frac{3}{6} \begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} -2/5 \\ 2 \\ 1/5 \end{pmatrix}, \quad \mathbf{v}_{\perp} = \mathbf{v} - \mathbf{v}_{\parallel} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} - \begin{pmatrix} -2/5 \\ 2 \\ 1/5 \end{pmatrix} = \begin{pmatrix} 7/5 \\ 0 \\ 14/5 \end{pmatrix} = \frac{7}{5} \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}.$$

Propiedades de la proyección ortogonal

Proyección de un vector contenido en W

Si \mathbf{v} pertenece al subespacio W entonces (y sólo entonces) $\text{proy}_W(\mathbf{v}) = \mathbf{v}$.

Propiedad de mínimo de la proyección ortogonal

La proyección ortogonal de un vector \mathbf{v} sobre un subespacio W , $\text{proy}_W(\mathbf{v})$, es el vector de W que menos dista de \mathbf{v} :

$$\|\mathbf{v} - \text{proy}_W(\mathbf{v})\| = \min_{\mathbf{x} \in W} \|\mathbf{v} - \mathbf{x}\|. \quad (7.8)$$

Propiedad de
mínimo de la
proyección
ortogonal.

En otras palabras: Para cualquier vector \mathbf{x} de W distinto de la proyección ortogonal $\text{proy}_W(\mathbf{v})$, la distancia de \mathbf{x} a \mathbf{v} es mayor que la distancia de $\text{proy}_W(\mathbf{v})$ a \mathbf{v} :

$$\|\mathbf{v} - \text{proy}_W(\mathbf{v})\| < \|\mathbf{v} - \mathbf{x}\|. \quad (7.9)$$

Esta propiedad es una simple consecuencia del teorema de Pitágoras. Para convencerse de ello basta observar que la componente transversal $\mathbf{v}_\perp = \mathbf{v} - \text{proy}_W(\mathbf{v})$ es, por definición, perpendicular a todo vector de W , y es, en particular, perpendicular al vector diferencia $\text{proy}_W(\mathbf{v}) - \mathbf{x}$ (que —por hipótesis— es distinto de cero). Por ser perpendiculares, estos dos vectores son los catetos de un triángulo rectángulo cuya hipotenusa es su suma:

$$(\mathbf{v} - \text{proy}_W(\mathbf{v})) + (\text{proy}_W(\mathbf{v}) - \mathbf{x}) = \mathbf{v} - \mathbf{x}.$$

Entonces, el teorema de Pitágoras dice:

$$\|\mathbf{v} - \text{proy}_W(\mathbf{v})\|^2 + \|\text{proy}_W(\mathbf{v}) - \mathbf{x}\|^2 = \|\mathbf{v} - \mathbf{x}\|^2$$

y en consecuencia, la hipótesis $\|\mathbf{x} - \text{proy}_W(\mathbf{v})\| > 0$ implica (7.9).

Como la distancia de un punto a un conjunto se define como el mínimo de las distancias del punto a los puntos del conjunto, la propiedad de mínimo de la proyección ortogonal es equivalente a decir que:

La distancia de un vector a un subespacio es igual a la norma o longitud de su componente transversal.

$$\text{dist}(\mathbf{v}, W) = \|\mathbf{v}_\perp\|.$$

Enlaces a los ejercicios de tarea de esta sección

Usa los siguientes enlaces para visualizar cada uno de los ejercicios de tarea que aparecen en esta sección:

Enlaces: [Ejercicio 1](#).

Ejercicios de la sección 7.2 Proyección ortogonal

En los ejercicios 1 a 4, demuestra que $\{\mathbf{u}_1, \mathbf{u}_2\}$ o $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ es una base ortogonal respectivamente de \mathbf{R}^2 o \mathbf{R}^3 y expresa \mathbf{x} como combinación lineal de las \mathbf{u}_i :

1. $\mathbf{u}_1 = \begin{pmatrix} 2 \\ -3 \end{pmatrix}, \mathbf{u}_2 = \begin{pmatrix} 6 \\ 4 \end{pmatrix}, \quad \mathbf{x} = \begin{pmatrix} 9 \\ -7 \end{pmatrix}.$

2. $\mathbf{u}_1 = \begin{pmatrix} 3 \\ 1 \end{pmatrix}, \mathbf{u}_2 = \begin{pmatrix} -2 \\ 6 \end{pmatrix}, \quad \mathbf{x} = \begin{pmatrix} -6 \\ 3 \end{pmatrix}.$

3. $\mathbf{u}_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \mathbf{u}_2 = \begin{pmatrix} -1 \\ 4 \\ 1 \end{pmatrix}, \mathbf{u}_3 = \begin{pmatrix} 2 \\ 1 \\ -2 \end{pmatrix}, \quad \mathbf{x} = \begin{pmatrix} 8 \\ -4 \\ -3 \end{pmatrix}.$

4. $\mathbf{u}_1 = \begin{pmatrix} 3 \\ -3 \\ 0 \end{pmatrix}, \mathbf{u}_2 = \begin{pmatrix} 2 \\ 2 \\ -1 \end{pmatrix}, \mathbf{u}_3 = \begin{pmatrix} 1 \\ 1 \\ 4 \end{pmatrix}, \quad \mathbf{x} = \begin{pmatrix} 5 \\ -3 \\ 1 \end{pmatrix}.$

5. Calcula la proyección ortogonal del vector $\begin{pmatrix} 1 \\ 7 \end{pmatrix}$ sobre la recta cuya dirección es la del vector $\begin{pmatrix} -4 \\ 7 \end{pmatrix}$.

6. Calcula la proyección ortogonal de $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$ sobre la recta cuya dirección es la del vector $\begin{pmatrix} -1 \\ 3 \end{pmatrix}$.

7. Sean $\mathbf{y} = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$ y $\mathbf{u} = \begin{pmatrix} 4 \\ -7 \end{pmatrix}$. Escribe \mathbf{y} como suma de dos vectores, uno en $\text{Gen}\{\mathbf{u}\}$ y otro ortogonal a \mathbf{u} .

8. Sean $\mathbf{y} = \begin{pmatrix} 2 \\ 6 \end{pmatrix}$ y $\mathbf{u} = \begin{pmatrix} 7 \\ 1 \end{pmatrix}$. Escribe \mathbf{y} como suma de dos vectores, uno en $\text{Gen}\{\mathbf{u}\}$ y otro ortogonal a \mathbf{u} .

9. Sean $\mathbf{y} = \begin{pmatrix} -3 \\ 9 \end{pmatrix}$ y $\mathbf{u} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$. Calcula la distancia de \mathbf{y} a la recta que pasa por el origen y por \mathbf{u} .

10. Sean $\mathbf{y} = \begin{pmatrix} 3 \\ 1 \end{pmatrix}$ y $\mathbf{u} = \begin{pmatrix} 8 \\ 6 \end{pmatrix}$. Calcula la distancia de \mathbf{y} a la recta que pasa por el origen y por \mathbf{u} .

En cada uno de los ejercicios 11 a 16 se da un conjunto de vectores. Averigua en qué casos es un conjunto ortogonal y en aquellos que lo sea, normaliza los vectores para obtener un conjunto ortonormal.

11. $\begin{pmatrix} 1/3 \\ 1/3 \\ 1/3 \end{pmatrix}, \begin{pmatrix} -1/2 \\ 0 \\ 1/2 \end{pmatrix}$. 12. $\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \\ 0 \end{pmatrix}$.

13. $\begin{pmatrix} -2/3 \\ 1/3 \\ 2/3 \end{pmatrix}, \begin{pmatrix} 1/3 \\ 2/3 \\ 0 \end{pmatrix}$. 14. $\begin{pmatrix} -0'6 \\ 0'8 \end{pmatrix}, \begin{pmatrix} 0'8 \\ 0'6 \end{pmatrix}$.

15. $\begin{pmatrix} 1/\sqrt{10} \\ 3/\sqrt{20} \\ 3/\sqrt{20} \end{pmatrix}, \begin{pmatrix} 3/\sqrt{10} \\ -1/\sqrt{20} \\ -1/\sqrt{20} \end{pmatrix}, \begin{pmatrix} 0 \\ -1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$.

16. $\begin{pmatrix} 1/\sqrt{18} \\ 4/\sqrt{18} \\ 1/\sqrt{18} \end{pmatrix}, \begin{pmatrix} 1/\sqrt{2} \\ 0 \\ -1/\sqrt{2} \end{pmatrix}, \begin{pmatrix} -2/3 \\ 1/3 \\ -2/3 \end{pmatrix}$.

17. Demuestra que para toda matriz $m \times n$, U , cuyas n columnas forman un conjunto ortonormal, si \mathbf{x} e \mathbf{y} son vectores de \mathbf{R}^n entonces la aplicación lineal $\mathbf{x} \mapsto U\mathbf{x}$:

- (a) Conserva el producto escalar: $(U\mathbf{x}) \cdot (U\mathbf{y}) = \mathbf{x} \cdot \mathbf{y}$.
- (b) Conserva las normas: $\|U\mathbf{x}\| = \|\mathbf{x}\|$.
- (c) Conserva la perpendicularidad: $(U\mathbf{x}) \cdot (U\mathbf{y}) = 0$ si y sólo si $\mathbf{x} \cdot \mathbf{y} = 0$.

En los ejercicios 18 y 19 todos los vectores son de \mathbf{R}^n . Indica para cada afirmación si es verdadera o falsa, justificando tus respuestas.

18.

- (a) No todo conjunto ortogonal de \mathbf{R}^n es libre.
- (b) Si \mathbf{y} es una combinación lineal de vectores no nulos de un conjunto ortogonal, entonces los coeficientes en la combinación lineal se pueden calcular sin necesidad de realizar operaciones elementales de fila en una matriz.
- (c) Si los vectores de un conjunto de vectores ortogonales no nulos se normalizan entonces puede ocurrir que el conjunto resultante ya no sea ortogonal.
- (d) Una matriz con columnas ortogonales es una matriz ortogonal.
- (e) Si L es una recta por el origen e \mathbf{y}_{\parallel} es la proyección ortogonal de \mathbf{y} sobre L , entonces $\|\mathbf{y}_{\parallel}\|$ es la distancia de \mathbf{y} a L .

19.

- (a) No todo conjunto de vectores de \mathbf{R}^n linealmente independientes es ortogonal.
- (b) Si un conjunto $S = \{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ tiene la propiedad de que $\mathbf{u}_i \cdot \mathbf{u}_j = 0$ siempre que $i \neq j$, entonces S es un conjunto ortonormal.
- (c) Si las columnas de una matriz $m \times n$, A , son ortonormales, entonces la aplicación lineal $\mathbf{x} \mapsto A\mathbf{x}$ conserva las normas.
- (d) La proyección ortogonal de \mathbf{y} sobre \mathbf{v} es la misma que la proyección ortogonal de \mathbf{y} sobre $c\mathbf{v}$ para cualquier $c \neq 0$.
- (e) Toda matriz ortogonal es inversible.

20. Sea W un subespacio de \mathbf{R}^n generado por n vectores no nulos ortogonales. Demuestra $W = \mathbf{R}^n$.

21. Sea U una matriz cuadrada con columnas ortonormales. Explica por qué U es inversible.

22. Sea U una matriz $n \times n$ con columnas ortonormales. Demuestra que las filas de U forman una base ortonormal de \mathbf{R}^n .

23. Sean U y V matrices ortogonales. Demuestra que UV es también una matriz ortogonal (o sea, que su traspuesta es su inversa).

24. Dado un vector no nulo \mathbf{u} de \mathbf{R}^n , sea $L = \text{Gen}\{\mathbf{u}\}$. Demuestra que la aplicación $\mathbf{x} \mapsto \text{proy}_L \mathbf{x}$ es una aplicación lineal.

25. Dado un vector no nulo \mathbf{u} de \mathbf{R}^n , sea $L = \text{Gen}\{\mathbf{u}\}$. Para cada $\mathbf{y} \in \mathbf{R}^n$, la reflexión de \mathbf{y} en L es el vector $\text{refl}_L \mathbf{y}$ definido por $\text{refl}_L \mathbf{y} = 2(\text{proy}_L \mathbf{y}) - \mathbf{y}$. Demuestra que la aplicación $\mathbf{x} \mapsto \text{refl}_L \mathbf{x}$ es una aplicación lineal.

En los ejercicios 26 y 27 se da una base ortogonal $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4\}$ de \mathbf{R}^4 .

26. $\mathbf{u}_1 = \begin{pmatrix} 0 \\ 1 \\ -4 \\ -1 \end{pmatrix}, \mathbf{u}_2 = \begin{pmatrix} 3 \\ 5 \\ 1 \\ 1 \end{pmatrix}, \mathbf{u}_3 = \begin{pmatrix} 1 \\ 0 \\ 1 \\ -4 \end{pmatrix}, \mathbf{u}_4 = \begin{pmatrix} 5 \\ -3 \\ -1 \\ 1 \end{pmatrix}$. Escribe el vector $\mathbf{x} = \begin{pmatrix} 10 \\ -8 \\ 2 \\ 0 \end{pmatrix}$ como suma de dos vectores, uno en $\text{Gen}\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ y otro en $\text{Gen}\{\mathbf{u}_4\}$.

27. $\mathbf{u}_1 = \begin{pmatrix} 1 \\ 2 \\ 1 \\ 1 \end{pmatrix}, \mathbf{u}_2 = \begin{pmatrix} -2 \\ 1 \\ -1 \\ 1 \end{pmatrix}, \mathbf{u}_3 = \begin{pmatrix} 1 \\ 1 \\ -2 \\ -1 \end{pmatrix}, \mathbf{u}_4 = \begin{pmatrix} -1 \\ 1 \\ 1 \\ -2 \end{pmatrix}$. Escribe el vector $\mathbf{v} = \begin{pmatrix} 4 \\ 5 \\ -3 \\ 3 \end{pmatrix}$ como suma de dos vectores, uno en $\text{Gen}\{\mathbf{u}_1\}$ y otro en $\text{Gen}\{\mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4\}$.

7. Ortogonalidad

En los ejercicios 28 a 31 comprueba que $\{\mathbf{u}_1, \mathbf{u}_2\}$ es un conjunto ortogonal y halla la proyección ortogonal de \mathbf{y} sobre $\text{Gen}\{\mathbf{u}_1, \mathbf{u}_2\}$.

$$28. \mathbf{y} = \begin{pmatrix} -1 \\ 4 \\ 3 \end{pmatrix}, \mathbf{u}_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \mathbf{u}_2 = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}.$$

$$29. \mathbf{y} = \begin{pmatrix} 6 \\ 3 \\ -2 \end{pmatrix}, \mathbf{u}_1 = \begin{pmatrix} 3 \\ 4 \\ 0 \end{pmatrix}, \mathbf{u}_2 = \begin{pmatrix} -4 \\ 3 \\ 0 \end{pmatrix}.$$

$$30. \mathbf{y} = \begin{pmatrix} -1 \\ 2 \\ 6 \end{pmatrix}, \mathbf{u}_1 = \begin{pmatrix} 3 \\ -1 \\ 2 \end{pmatrix}, \mathbf{u}_2 = \begin{pmatrix} 1 \\ -1 \\ -2 \end{pmatrix}.$$

$$31. \mathbf{y} = \begin{pmatrix} 6 \\ 4 \\ 1 \end{pmatrix}, \mathbf{u}_1 = \begin{pmatrix} -4 \\ -1 \\ 1 \end{pmatrix}, \mathbf{u}_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}.$$

En los ejercicios 32 a 39, sea W el espacio generado por los vectores \mathbf{u}_1 y \mathbf{u}_2 . Escribe \mathbf{y} como suma de un vector de W y un vector ortogonal a W .

$$32. \mathbf{y} = \begin{pmatrix} 1 \\ 3 \\ 5 \end{pmatrix}, \mathbf{u}_1 = \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix}, \mathbf{u}_2 = \begin{pmatrix} 5 \\ 1 \\ 4 \end{pmatrix}.$$

$$33. \mathbf{y} = \begin{pmatrix} -1 \\ 4 \\ 3 \end{pmatrix}, \mathbf{u}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \mathbf{u}_2 = \begin{pmatrix} -1 \\ 3 \\ -2 \end{pmatrix}.$$

$$34. \mathbf{y} = \begin{pmatrix} 4 \\ 3 \\ 3 \\ -1 \end{pmatrix}, \mathbf{u}_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \\ -1 \end{pmatrix}, \mathbf{u}_2 = \begin{pmatrix} -1 \\ 3 \\ 1 \\ -2 \end{pmatrix}, \mathbf{u}_3 = \begin{pmatrix} -1 \\ 0 \\ 1 \\ 1 \end{pmatrix}.$$

$$35. \mathbf{y} = \begin{pmatrix} 3 \\ 4 \\ 5 \\ 6 \end{pmatrix}, \mathbf{u}_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \\ -1 \end{pmatrix}, \mathbf{u}_2 = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 1 \end{pmatrix}, \mathbf{u}_3 = \begin{pmatrix} 0 \\ -1 \\ 1 \\ -1 \end{pmatrix}.$$

$$36. \mathbf{y} = \begin{pmatrix} 3 \\ 1 \\ 5 \\ 1 \end{pmatrix}, \mathbf{u}_1 = \begin{pmatrix} 3 \\ 1 \\ -1 \\ 1 \end{pmatrix}, \mathbf{u}_2 = \begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \end{pmatrix}.$$

$$37. \mathbf{y} = \begin{pmatrix} 3 \\ -1 \\ 1 \\ 13 \end{pmatrix}, \mathbf{u}_1 = \begin{pmatrix} 1 \\ -2 \\ -1 \\ 2 \end{pmatrix}, \mathbf{u}_2 = \begin{pmatrix} -4 \\ 1 \\ 0 \\ 3 \end{pmatrix}.$$

$$38. \mathbf{y} = \begin{pmatrix} 3 \\ -7 \\ 2 \\ 3 \end{pmatrix}, \mathbf{u}_1 = \begin{pmatrix} 2 \\ -1 \\ -3 \\ 1 \end{pmatrix}, \mathbf{u}_2 = \begin{pmatrix} 1 \\ 1 \\ 0 \\ -1 \end{pmatrix}.$$

7.2. Proyección ortogonal

$$39. \mathbf{y} = \begin{pmatrix} 2 \\ 4 \\ 0 \\ -1 \end{pmatrix}, \mathbf{u}_1 = \begin{pmatrix} 2 \\ 0 \\ -1 \\ -3 \end{pmatrix}, \mathbf{u}_2 = \begin{pmatrix} 5 \\ -2 \\ 4 \\ 2 \end{pmatrix}.$$

40. Sean $\mathbf{y} = \begin{pmatrix} 5 \\ 9 \\ -5 \end{pmatrix}$, $\mathbf{u}_1 = \begin{pmatrix} 1 \\ 3 \\ -2 \end{pmatrix}$, $\mathbf{u}_2 = \begin{pmatrix} 5 \\ 1 \\ 4 \end{pmatrix}$. Halla la distancia desde \mathbf{y} hasta el plano de \mathbf{R}^3 generado por \mathbf{u}_1 y \mathbf{u}_2 .

41. Sean \mathbf{y} , \mathbf{u}_1 y \mathbf{u}_2 como en el ejercicio 32. Halla la distancia desde \mathbf{y} hasta el subespacio de \mathbf{R}^4 generado por \mathbf{u}_1 y \mathbf{u}_2 .

42. Sean $\mathbf{y} = \begin{pmatrix} 4 \\ 8 \\ 1 \end{pmatrix}$, $\mathbf{u}_1 = \begin{pmatrix} 2/3 \\ 1/3 \\ 2/3 \end{pmatrix}$, $\mathbf{u}_2 = \begin{pmatrix} -2/3 \\ 2/3 \\ 1/3 \end{pmatrix}$ y $W = \text{Gen}\{\mathbf{u}_1, \mathbf{u}_2\}$. Sea $U = [\mathbf{u}_1 \ \mathbf{u}_2]$ la matriz cuyas columnas son \mathbf{u}_1 y \mathbf{u}_2 .

- Calcula $U^T U$ y $U U^T$.
- Calcula $\text{proy}_W \mathbf{y}$ y $(U U^T) \mathbf{y}$.

43. Sean $\mathbf{y} = \begin{pmatrix} 7 \\ 9 \end{pmatrix}$, $\mathbf{u}_1 = \begin{pmatrix} 1/\sqrt{10} \\ -3/\sqrt{10} \end{pmatrix}$ y $W = \text{Gen}\{\mathbf{u}_1\}$.

- Si U es la matriz cuya única columna es \mathbf{u}_1 , calcula $U^T U$ y $U U^T$.
- Calcula $\text{proy}_W \mathbf{y}$ y $(U U^T) \mathbf{y}$.

44. Sean $\mathbf{u}_1 = \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix}$, $\mathbf{u}_2 = \begin{pmatrix} 5 \\ -1 \\ 2 \end{pmatrix}$ y $\mathbf{u}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$. Observa que \mathbf{u}_1 y \mathbf{u}_2 son mutuamente ortogonales pero que \mathbf{u}_3 no es ortogonal ni a \mathbf{u}_1 ni a \mathbf{u}_2 . Se puede demostrar que \mathbf{u}_3 no pertenece al subespacio generado por \mathbf{u}_1 y \mathbf{u}_2 . Usa este hecho para hallar un vector no nulo \mathbf{v} de \mathbf{R}^3 que sea ortogonal a \mathbf{u}_1 y a \mathbf{u}_2 .

45. Sean $\mathbf{u}_1, \mathbf{u}_2$ los del ejercicio 44 y sea $\mathbf{u}_4 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$. Se puede demostrar que \mathbf{u}_4 no pertenece al subespacio generado por \mathbf{u}_1 y \mathbf{u}_2 . Usa este hecho (sin necesidad de demostrarlo) para hallar un vector no nulo \mathbf{v} de \mathbf{R}^3 que sea ortogonal a \mathbf{u}_1 y a \mathbf{u}_2 .

En los ejercicios 46 y 47 todos los vectores y subespacios están en \mathbf{R}^n . Indica para cada afirmación si es verdadera o falsa, justificando tus respuestas.

- Si \mathbf{z} es ortogonal a \mathbf{u}_1 y a \mathbf{u}_2 y $W = \text{Gen}\{\mathbf{u}_1, \mathbf{u}_2\}$, entonces \mathbf{z} pertenece a W^\perp .
 - Para cada \mathbf{y} y cada subespacio W , el vector $\mathbf{y} - \text{proy}_W \mathbf{y}$ es ortogonal a W .
 - La proyección ortogonal de \mathbf{y} sobre un subespacio W de \mathbf{R}^n puede depender en algunos casos de la base ortogonal de W usada para calcularla.

- (d) Si \mathbf{y} pertenece a un subespacio W , entonces la proyección ortogonal de \mathbf{y} sobre W es el propio \mathbf{y} .
- (e) Si las columnas de una matriz $n \times p$, U , son ortonormales entonces $UU^T \mathbf{y}$ es la proyección ortogonal de \mathbf{y} sobre el espacio columna de U .
- 47.
- (a) Si W es un subespacio de \mathbf{R}^n y \mathbf{v} pertenece tanto a W como a W^\perp , entonces \mathbf{v} es el vector cero.
- (b) Si $\mathbf{y} = \mathbf{z}_1 + \mathbf{z}_2$, donde \mathbf{z}_1 está en un subespacio W y \mathbf{z}_2 está en W^\perp entonces \mathbf{z}_1 es la proyección ortogonal de \mathbf{y} sobre W .
- (c) La aproximación óptima de \mathbf{y} por vectores de un subespacio W es el vector $\mathbf{y} - \text{proy}_W \mathbf{y}$.
- (d) Si las columnas de una matriz $n \times p$, U , son ortonormales entonces $UU^T \mathbf{x} = \mathbf{x}$ para todo $\mathbf{x} \in \mathbf{R}^n$.
48. Si A es una matriz $m \times n$, entonces todo vector $\mathbf{x} \in \mathbf{R}^n$ se puede poner como suma $\mathbf{x} = \mathbf{p} + \mathbf{u}$ donde $\mathbf{p} \in \text{Fil } A$ y $\mathbf{u} \in \text{Nul } A$.
49. Sea W el subespacio de \mathbf{R}^n generado por un conjunto ortogonal $\{\mathbf{w}_1, \dots, \mathbf{w}_p\}$ y sea $\{\mathbf{v}_1, \dots, \mathbf{v}_q\}$ una base ortogonal de W^\perp . ¿Por qué es $\{\mathbf{w}_1, \dots, \mathbf{w}_p, \mathbf{v}_1, \dots, \mathbf{v}_q\}$ un conjunto ortogonal? ¿Por qué genera \mathbf{R}^n ? Demuestra que $\dim W + \dim W^\perp = n$.

Lección 10,
2 may 2023

7.3. Ortogonalización de Gram-Schmidt y factorización QR

Ortogonalización de Gram-Schmidt

El método de ortogonalización de Gram-Schmidt nos permite transformar una base cualquiera $\{\mathbf{v}_1, \dots, \mathbf{v}_r\}$ de un subespacio W de \mathbf{R}^n en una base ortogonal del mismo. Para comprender la idea fundamental de este método basta comprender el caso especial en el que tenemos solamente dos vectores linealmente independientes, \mathbf{v}_1 y \mathbf{v}_2 . En este caso, para obtener un sistema ortogonal a partir de estos vectores basta sustituir \mathbf{v}_2 por su componente transversal respecto a \mathbf{v}_1 , de manera que pasamos de la base $\{\mathbf{v}_1, \mathbf{v}_2\}$ a la base ortogonal $\{\mathbf{u}_1, \mathbf{u}_2\}$ definiendo:

$$\begin{aligned}\mathbf{u}_1 &= \mathbf{v}_1, \\ \mathbf{u}_2 &= \mathbf{v}_2 - \text{proy}_{\mathbf{u}_1}(\mathbf{v}_2) = \mathbf{v}_2 - \left(\frac{\mathbf{v}_2 \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1} \right) \mathbf{u}_1.\end{aligned}$$

Esta idea puede repetirse el número de veces que sea necesario para convertir cualquier conjunto libre de vectores en un conjunto ortogonal. En general, un sistema de r vectores independientes $\{\mathbf{v}_1, \dots, \mathbf{v}_r\}$ se transforma en un sistema ortogonal en r pasos en cada uno de los cuales (excepto en el primero) se sustituye un vector \mathbf{v}_{i+1} por su componente transversal respecto al subespacio $W_i = \text{Gen}\{\mathbf{v}_1, \dots, \mathbf{v}_i\} = \text{Gen}\{\mathbf{u}_1, \dots, \mathbf{u}_i\}$ generado por los vectores anteriores:

$$\mathbf{u}_{i+1} = \mathbf{v}_{i+1} - \text{proy}_{W_i}(\mathbf{v}_{i+1}) = \mathbf{v}_{i+1} - \left(\frac{\mathbf{v}_{i+1} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1} \right) \mathbf{u}_1 - \dots - \left(\frac{\mathbf{v}_{i+1} \cdot \mathbf{u}_i}{\mathbf{u}_i \cdot \mathbf{u}_i} \right) \mathbf{u}_i.$$

Por ejemplo, si partimos de tres vectores independientes $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$, éstos quedarían transformados en un conjunto ortogonal en tres pasos:

$$\begin{aligned}\mathbf{u}_1 &= \mathbf{v}_1, \\ \mathbf{u}_2 &= \mathbf{v}_2 - \text{proy}_{\mathbf{u}_1}(\mathbf{v}_2) = \mathbf{v}_2 - \left(\frac{\mathbf{v}_2 \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1} \right) \mathbf{u}_1, \\ \mathbf{u}_3 &= \mathbf{v}_3 - \text{proy}_{\{\mathbf{u}_1, \mathbf{u}_2\}}(\mathbf{v}_3) = \mathbf{v}_3 - \left(\frac{\mathbf{v}_3 \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1} \right) \mathbf{u}_1 - \left(\frac{\mathbf{v}_3 \cdot \mathbf{u}_2}{\mathbf{u}_2 \cdot \mathbf{u}_2} \right) \mathbf{u}_2.\end{aligned}$$

Cuando se realizan cálculos a mano conviene tener presente que los vectores \mathbf{u}_i que se van obteniendo paso a paso pueden ser a su vez sustituidos por un múltiplo escalar cualquiera, $c\mathbf{u}_i$, con el fin de eliminar denominadores de las fracciones que pudieran haberse obtenido y de esta forma facilitar los cálculos del paso siguiente. Veamos esto con un ejemplo.

Ejemplo:

Aplicar el proceso de Gram-Schmidt al conjunto libre de vectores de \mathbf{R}^4 siguiente:

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \quad \mathbf{v}_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \quad \mathbf{v}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \end{pmatrix}.$$

Solución: Tomando $\mathbf{u}_1 = \mathbf{v}_1$ calculamos

$$\mathbf{u}_2 = \mathbf{v}_2 - \left(\frac{\mathbf{v}_2 \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1} \right) \mathbf{u}_1 = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 1 \end{pmatrix} - \frac{3}{4} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} -3/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{pmatrix}; \quad \text{y multiplicando el resultado por 4:} \quad \xrightarrow{\times 4} \begin{pmatrix} -3 \\ 1 \\ 1 \\ 1 \end{pmatrix}.$$

Finalmente:

$$\mathbf{u}_3 = \mathbf{v}_3 - \left(\frac{\mathbf{v}_3 \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1} \right) \mathbf{u}_1 - \left(\frac{\mathbf{v}_3 \cdot \mathbf{u}_2}{\mathbf{u}_2 \cdot \mathbf{u}_2} \right) \mathbf{u}_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \end{pmatrix} - \frac{2}{4} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} - \frac{2}{12} \begin{pmatrix} -3 \\ 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ -2/3 \\ 1/3 \\ 1/3 \end{pmatrix} \xrightarrow{\times 3} \begin{pmatrix} 0 \\ -2 \\ 1 \\ 1 \end{pmatrix},$$

con lo cual hemos obtenido una base ortogonal del espacio generado por los vectores dados, $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$, que es la siguiente:

$$\mathbf{u}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \quad \mathbf{u}_2 = \begin{pmatrix} -3 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \quad \mathbf{u}_3 = \begin{pmatrix} 0 \\ -2 \\ 1 \\ 1 \end{pmatrix}.$$

Ortonormalización

En algunos casos, puede ser necesario obtener una base *ortonormal*, es decir, una base en la que no solo sean los vectores ortogonales dos a dos, sino que además, cada vector de la base sea un vector unitario. Una vez conseguida una base ortogonal, no hay más que “normalizar” cada uno de los vectores hallados para convertirla en una base ortonormal haciendo una operación de reescalado con cada vector.

Por ejemplo, si después de obtener la base del ejemplo anterior deseamos convertirla en un sistema ortonormal, calculamos la norma de cada vector:

$$\begin{aligned} \|\mathbf{u}_1\| &= \sqrt{(1)^2 + (1)^2 + (1)^2 + (1)^2} = 2, \\ \|\mathbf{u}_2\| &= \sqrt{(-3)^2 + (1)^2 + (1)^2 + (1)^2} = \sqrt{12}, \\ \|\mathbf{u}_3\| &= \sqrt{(0)^2 + (-2)^2 + (1)^2 + (1)^2} = \sqrt{6}, \end{aligned}$$

y dividiendo cada vector por su norma obtenemos la base ortonormal:

$$\left\{ \begin{pmatrix} 1/2 \\ 1/2 \\ 1/2 \\ 1/2 \end{pmatrix}, \begin{pmatrix} -3/\sqrt{12} \\ 1/\sqrt{12} \\ 1/\sqrt{12} \\ 1/\sqrt{12} \end{pmatrix}, \begin{pmatrix} 0 \\ -2/\sqrt{6} \\ 1/\sqrt{6} \\ 1/\sqrt{6} \end{pmatrix} \right\}.$$

Factorización QR

Sea A una matriz real $m \times n$ cuyas n columnas, $\mathbf{a}_1, \dots, \mathbf{a}_n$, son vectores linealmente independientes de \mathbf{R}^m . Si aplicamos el proceso de Gram-Schmidt al sistema formado por las columnas de A , obtendremos una base ortogonal

$$\{\mathbf{a}_1, \dots, \mathbf{a}_n\} \xrightarrow{\text{G-S}} \{\mathbf{u}_1, \dots, \mathbf{u}_n\} = \mathcal{U}$$

del espacio columna de A . Pero además, en el paso k del proceso hemos obtenido una base que sólo difiere de la formada por las columnas de A en los k primeros vectores, es decir una base de la forma

$$\{\mathbf{u}_1, \dots, \mathbf{u}_k, \mathbf{a}_{k+1}, \dots, \mathbf{a}_n\}$$

donde los vectores $\mathbf{u}_1, \dots, \mathbf{u}_k$ generan el mismo subespacio de \mathbf{R}^m que las k primeras columnas de A , $\mathbf{a}_1, \dots, \mathbf{a}_k$. En consecuencia cada \mathbf{a}_k pertenece al espacio generado por los primeros k vectores de la nueva base \mathcal{U} y por tanto respecto a esa base sus coordenadas $k+1$ a n -ésima son cero:

$$\mathbf{a}_k = r_{1k}\mathbf{u}_1 + \dots + r_{kk}\mathbf{u}_k + 0\mathbf{u}_{k+1} + \dots + 0\mathbf{u}_n = [\mathbf{u}_1 \dots \mathbf{u}_n] \begin{pmatrix} r_{1k} \\ \vdots \\ r_{kk} \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Además, la coordenada r_{kk} es distinta de cero porque \mathbf{a}_k es linealmente independiente las columnas $\mathbf{a}_1, \dots, \mathbf{a}_{k-1}$ de A y por tanto también de los vectores $\mathbf{u}_1, \dots, \mathbf{u}_{k-1}$. Por tanto esos vectores de coordenadas forman una matriz triangular superior inversible:

$$R = \begin{pmatrix} r_{11} & r_{12} & \dots & r_{1n} \\ 0 & r_{22} & \dots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & r_{nn} \end{pmatrix}.$$

Por otra parte, podemos hacer que el signo de los r_{kk} sea positivo sin más que ajustar el signo del correspondiente vector \mathbf{u}_k en el proceso de ortogonalización de Gram-Schmidt. Si hacemos eso, la matriz R será una matriz *triangular superior con elementos diagonales positivos* y junto con la matriz $Q = [\mathbf{u}_1 \dots \mathbf{u}_n]$ proporciona una factorización de A :

$$A = QR. \quad (7.10)$$

Si, además, los vectores \mathbf{u}_i obtenidos en el proceso de Gram-Schmidt se habían normalizado de forma que el conjunto obtenido, $\{\mathbf{u}_1, \dots, \mathbf{u}_n\}$, era un conjunto ortonormal, llegamos a esta conclusión:

El proceso de Gram-Schmidt aplicado a un conjunto libre $\{\mathbf{a}_1, \dots, \mathbf{a}_n\}$ permite obtener una factorización $A = QR$ de la matriz $A = [\mathbf{a}_1, \dots, \mathbf{a}_n]$ donde Q es una matriz de columnas ortonormales y R es una matriz triangular superior con todos los elementos diagonales positivos.

Esta factorización de la matriz A se conoce como la factorización QR.

La matriz Q de la factorización QR es una matriz cuyas columnas $\mathbf{u}_1 \dots \mathbf{u}_n$ forman un sistema ortonormal. Debido a que todos los productos escalares $\mathbf{u}_i \cdot \mathbf{u}_j = \mathbf{u}_i^T \mathbf{u}_j$ son iguales a cero si $i \neq j$ e iguales a 1 si $i = j$, tenemos

$$Q^T Q = \begin{bmatrix} \mathbf{u}_1^T \\ \vdots \\ \mathbf{u}_n^T \end{bmatrix} [\mathbf{u}_1 \dots \mathbf{u}_n] = \begin{pmatrix} \mathbf{u}_1 \cdot \mathbf{u}_1 & \dots & \mathbf{u}_1 \cdot \mathbf{u}_n \\ \vdots & \ddots & \vdots \\ \mathbf{u}_n \cdot \mathbf{u}_1 & \dots & \mathbf{u}_n \cdot \mathbf{u}_n \end{pmatrix} = \begin{pmatrix} 1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 1 \end{pmatrix} = I_n.$$

Esta es la propiedad característica de las matrices con columnas ortonormales. Es decir: *Una matriz tiene columnas ortonormales si y sólo si su traspuesta es su inversa por la izquierda.*

Debido a la propiedad característica de las matrices con columnas ortonormales, la matriz R de la factorización QR de A se puede calcular de forma sencilla a partir de Q y de A ya que de (7.10), multiplicando por la izquierda por Q^T obtenemos:

$$R = Q^T A.$$

Si, además, la matriz Q (y por tanto A) es cuadrada, entonces es inversible y su inversa es su traspuesta. Tal matriz se llama una *matriz ortogonal*. Es decir: *Una matriz es una matriz ortogonal si y sólo si es inversible y su traspuesta es su inversa.*

La factorización QR de una matriz cuadrada A , si existe, es la única factorización de A como producto de una matriz ortogonal por una matriz inversible triangular superior con todos los elementos de la diagonal positivos.

Ejemplo:

En la sección anterior aplicamos el método de Gram-Schmidt para ortonormalizar las columnas de la matriz

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

y obtuvimos:

$$Q = \begin{pmatrix} 1/2 & -3/\sqrt{12} & 0 \\ 1/2 & 1/\sqrt{12} & -2/\sqrt{6} \\ 1/2 & 1/\sqrt{12} & 1/\sqrt{6} \\ 1/2 & 1/\sqrt{12} & 1/\sqrt{6} \end{pmatrix}$$

En consecuencia, podemos calcular R así:

$$R = Q^T A = \begin{pmatrix} 1/2 & 1/2 & 1/2 & 1/2 \\ -3/\sqrt{12} & 1/\sqrt{12} & 1/\sqrt{12} & 1/\sqrt{12} \\ 0 & -2/\sqrt{6} & 1/\sqrt{6} & 1/\sqrt{6} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 3/2 & 1 \\ 0 & 3/\sqrt{12} & 2/\sqrt{12} \\ 0 & 0 & 2/\sqrt{6} \end{pmatrix}.$$

Como todos los elementos de la diagonal son positivos, hemos encontrado la factorización QR. Si alguno de los elementos de la diagonal hubiese resultado negativo, habría que cambiar de signo a todos los elementos de su fila y hacer lo mismo con la correspondiente columna de Q .

Ejercicios de la sección 7.3 Gram-Schmidt y Factorización QR

En los ejercicios 1 a 6, se da una base de un subespacio W . Usa el proceso de Gram-Schmidt para obtener una base ortogonal de W .

1. $\begin{pmatrix} 3 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 8 \\ 5 \\ -6 \end{pmatrix}.$

2. $\begin{pmatrix} 0 \\ 4 \\ 2 \end{pmatrix}, \begin{pmatrix} 5 \\ 6 \\ -7 \end{pmatrix}.$

3. $\begin{pmatrix} 2 \\ -5 \\ 1 \end{pmatrix}, \begin{pmatrix} 4 \\ -1 \\ 2 \end{pmatrix}.$

4. $\begin{pmatrix} 3 \\ -4 \\ 5 \end{pmatrix}, \begin{pmatrix} -3 \\ 14 \\ -7 \end{pmatrix}.$

5. $\begin{pmatrix} 1 \\ -4 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 7 \\ -7 \\ -4 \\ 1 \end{pmatrix}.$

6. $\begin{pmatrix} 3 \\ -1 \\ 2 \\ -1 \end{pmatrix}, \begin{pmatrix} -5 \\ 9 \\ -9 \\ 3 \end{pmatrix}.$

7. Halla una base ortonormal del subespacio generado por los vectores del ejercicio 3.

8. Halla una base ortonormal del subespacio generado por los vectores del ejercicio 4.

En los ejercicios 9 a 12 halla una base ortogonal del espacio columna de la matriz dada:

$$9. \begin{pmatrix} 3 & -5 & 1 \\ 1 & 1 & 1 \\ -1 & 5 & -2 \\ 3 & -7 & 8 \end{pmatrix}.$$

$$10. \begin{pmatrix} -1 & 6 & 6 \\ 3 & -8 & 3 \\ 1 & -2 & 6 \\ 1 & -4 & -3 \end{pmatrix}.$$

$$11. \begin{pmatrix} 1 & 2 & 5 \\ -1 & 1 & -4 \\ -1 & 4 & -3 \\ 1 & -4 & 7 \\ 1 & 2 & 1 \end{pmatrix}.$$

$$12. \begin{pmatrix} 1 & 3 & 5 \\ -1 & -3 & 1 \\ 0 & 2 & 3 \\ 1 & 5 & 2 \\ 1 & 5 & 8 \end{pmatrix}.$$

En los ejercicios 13 y 14 las columnas de Q fueron halladas aplicando el proceso de Gram-Schmidt a las columnas de A . Halla una matriz triangular superior R tal que $A = QR$.

$$13. A = \begin{pmatrix} -2 & 3 \\ 5 & 7 \\ 2 & -2 \\ 4 & 6 \end{pmatrix}, Q = \begin{pmatrix} -2/7 & 5/7 \\ 5/7 & 2/7 \\ 2/7 & -4/7 \\ 4/7 & 2/7 \end{pmatrix}.$$

$$14. A = \begin{pmatrix} 5 & 9 \\ 1 & 7 \\ -3 & -5 \\ 1 & 5 \end{pmatrix}, Q = \begin{pmatrix} 5/6 & -1/6 \\ 1/6 & 5/6 \\ -3/6 & 1/6 \\ 1/6 & 3/6 \end{pmatrix}.$$

15. Halla una factorización QR de la matriz del ejercicio 11.

16. Halla una factorización QR de la matriz del ejercicio 12.

En los ejercicios 17 y 18 todos los vectores y subespacios están en \mathbf{R}^n . Indica para cada afirmación si es verdadera o falsa, justificando tus respuestas.

17.

(a) Si $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ es una base ortogonal de W , entonces al multiplicar \mathbf{v}_3 por un escalar c se obtiene una nueva base ortogonal $\{\mathbf{v}_1, \mathbf{v}_2, c\mathbf{v}_3\}$.

(b) El proceso de Gram-Schmidt produce, a partir de un conjunto libre $S = \{\mathbf{x}_1, \dots, \mathbf{x}_r\}$, un conjunto ortogonal $U = \{\mathbf{v}_1, \dots, \mathbf{v}_r\}$ con la propiedad de que para cada k de 1 a r los k primeros vectores de U generan el mismo subespacio que los k primeros vectores de S .

(c) Si $A = QR$ donde Q tiene columnas ortonormales, entonces $R = Q^T A$.

18.

(a) Si $W = \text{Gen}\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$ donde $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ son linealmente independientes y si $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ es un conjunto ortogonal en W entonces $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ es una base de W .

(b) Si \mathbf{x} no pertenece a un subespacio W , entonces $\mathbf{x} - \text{proy}_W \mathbf{x}$ es distinto de cero.

(c) En una factorización QR, $A = QR$ (donde las columnas de A son linealmente independientes), las columnas de Q forman una base ortonormal del espacio columna de A .

7.4. Mínimos cuadrados

La propiedad de mínimo de la proyección ortogonal

Ya hemos visto en la sección 7.2 (ver página 10) que si H es un subespacio de \mathbf{R}^n y \mathbf{b} es un vector de \mathbf{R}^n entonces la proyección ortogonal de \mathbf{b} sobre H , \mathbf{b}_{\parallel} , tiene la siguiente propiedad de mínimo:

La proyección ortogonal \mathbf{b}_{\parallel} de \mathbf{b} sobre H es el vector de H que menos dista de \mathbf{b} .

Esto se debe a que para cualquier vector $\mathbf{x} \in H$, la distancia de \mathbf{x} a \mathbf{b} es la hipotenusa de un triángulo rectángulo del que la distancia de \mathbf{b}_{\parallel} a \mathbf{b} es un cateto y por tanto

$$\|\mathbf{x} - \mathbf{b}\| \geq \|\mathbf{b}_{\parallel} - \mathbf{b}\|.$$

El problema de mínimos cuadrados asociado con un sistema de ecuaciones lineales

Supongamos que nos dan un sistema de ecuaciones lineales, $A\mathbf{x} = \mathbf{b}$, que podría ser compatible o no.

problema de
mínimos
cuadrados

Se llama *problema de mínimos cuadrados* asociado a dicho sistema al problema de hallar un vector \mathbf{x} tal que la distancia de $A\mathbf{x}$ a \mathbf{b} (es decir, la norma $\|A\mathbf{x} - \mathbf{b}\|$) sea lo menor posible.

Por la propiedad de mínimo de la proyección ortogonal, ese mínimo es la distancia de \mathbf{b} a su proyección ortogonal sobre el espacio columna de A . En consecuencia las posibles soluciones del problema de mínimos cuadrados son las soluciones del sistema

$$A\mathbf{x} = \mathbf{b}_{\parallel}, \quad (\text{donde } \mathbf{b}_{\parallel} = \text{proy}_{\text{Col } A}(\mathbf{b})) \quad (7.11)$$

(que necesariamente será siempre compatible). Si $\hat{\mathbf{x}}$ es una solución de dicho problema de mínimos cuadrados, la cantidad

$$\epsilon = \|A\hat{\mathbf{x}} - \mathbf{b}\|$$

se llama el “error” en dicha solución, mientras que la cantidad $\epsilon^2 = \|A\hat{\mathbf{x}} - \mathbf{b}\|^2$ es el “error cuadrático” y el vector $A\hat{\mathbf{x}} - \mathbf{b}$ es el *vector de error*. Error, error cuadrático y vector de error.

Según lo dicho, para resolver un problema de mínimos cuadrados habría que plantear el sistema (7.11) y para ello habría que calcular la proyección ortogonal de \mathbf{b} sobre el espacio columna de A , lo cual puede ser complicado si las columnas de A no son ortogonales. Sin embargo vamos a hacer algunas observaciones que nos permitirán encontrar la misma solución de una forma más sencilla.

Solución del problema de mínimos cuadrados. Ecuaciones normales de un sistema

Recordemos que la componente transversal de \mathbf{b} respecto al subespacio $\text{Col } A$, $\mathbf{b}_{\perp} = \mathbf{b} - \mathbf{b}_{\parallel}$, tiene la propiedad de ser ortogonal a cada vector de $\text{Col } A$ y en particular a cada columna de A , así que si los vectores $\mathbf{a}_1, \dots, \mathbf{a}_n$ son las columnas de A , entonces, para cada $i = 1, \dots, n$,

$$0 = \mathbf{a}_i \cdot \mathbf{b}_{\perp} = \mathbf{a}_i \cdot (\mathbf{b} - \mathbf{b}_{\parallel}) \quad \text{de donde, } \mathbf{a}_i^T (\mathbf{b} - \mathbf{b}_{\parallel}) = 0,$$

Como los \mathbf{a}_i^T son las filas de la matriz traspuesta A^T , estas ecuaciones para los distintos valores de i se resumen en:

$$A^T(\mathbf{b} - \mathbf{b}_{\parallel}) = \mathbf{0}, \quad \text{o} \quad A^T \mathbf{b}_{\parallel} = A^T \mathbf{b}.$$

Por tanto si un vector \mathbf{x} es solución de (7.11) entonces también verifica $A^T A \mathbf{x} = A^T \mathbf{b}_{\parallel} = A^T \mathbf{b}$, es decir:

$$A^T A \mathbf{x} = A^T \mathbf{b}.$$

Recíprocamente, si \mathbf{x} cumple esta ecuación entonces $A^T(\mathbf{b} - A\mathbf{x}) = \mathbf{0}$, es decir, $\mathbf{b} - A\mathbf{x}$ es ortogonal a todas las columnas de A y por tanto a todos los vectores de $\text{Col } A$. Esto quiere decir que $A\mathbf{x}$ es un vector del espacio columna de A que restado de \mathbf{b} resulta en un vector ortogonal a $\text{Col } A$. Por el teorema de la proyección ortogonal, esto significa que $A\mathbf{x}$ es igual a $\text{proy}_{\text{Col } A}(\mathbf{b})$.

En conclusión, los sistemas

$$A\mathbf{x} = \mathbf{b}_{\parallel} \quad \text{y} \quad A^T A \mathbf{x} = A^T \mathbf{b}$$

tienen exactamente las mismas soluciones. El sistema de ecuaciones $A^T A \mathbf{x} = A^T \mathbf{b}$ (que tiene tantas ecuaciones como incógnitas) se conoce como *las ecuaciones normales del sistema $A\mathbf{x} = \mathbf{b}$* . Ecuaciones normales

La solución del problema de mínimos cuadrados asociado con un sistema de ecuaciones lineales $A\mathbf{x} = \mathbf{b}$ es el conjunto solución de sus ecuaciones normales.

Ejemplo: Resolver el problema de mínimos cuadrados asociado con el sistema $A\mathbf{x} = \mathbf{b}$ donde

$$A = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} -3 \\ -1 \\ 0 \\ 2 \\ 5 \\ 1 \end{pmatrix}.$$

Hallar también el error asociado con la solución por mínimos cuadrados.

Solución: Calculamos

$$A^T A = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 6 & 2 & 2 & 2 \\ 2 & 2 & 0 & 0 \\ 2 & 0 & 2 & 0 \\ 2 & 0 & 0 & 2 \end{pmatrix}$$

$$A^T \mathbf{b} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} -3 \\ -1 \\ 0 \\ 2 \\ 5 \\ 1 \end{pmatrix} = \begin{pmatrix} 4 \\ -4 \\ 2 \\ 6 \end{pmatrix}$$

Con lo cual ahora nos queda resolver el sistema $A^T A \mathbf{x} = A^T \mathbf{b}$, cuya matriz ampliada es:

$$\begin{pmatrix} 6 & 2 & 2 & 2 & 4 \\ 2 & 2 & 0 & 0 & -4 \\ 2 & 0 & 2 & 0 & 2 \\ 2 & 0 & 0 & 2 & 6 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & 0 & 1 & 3 \\ 0 & 1 & 0 & -1 & -5 \\ 0 & 0 & 1 & -1 & -2 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

(Observación: Se podría también multiplicar A^T por la matriz ampliada del sistema original y luego reducir.)

La solución general es:

$$\mathbf{x} = \begin{pmatrix} 3 \\ -5 \\ -2 \\ 0 \end{pmatrix} + x_4 \begin{pmatrix} -1 \\ 1 \\ 1 \\ 1 \end{pmatrix}.$$

El error en cualquiera de estas soluciones es el mismo e igual la norma $\epsilon = \|\mathbf{Ax} - \mathbf{b}\|$. Para calcularlo podemos coger cualquier solución particular, por ejemplo la que corresponde al valor cero de todos los parámetros. Entonces el error es la norma de

$$\mathbf{Ax}_0 - \mathbf{b} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 3 \\ -5 \\ -2 \\ 0 \end{pmatrix} - \begin{pmatrix} -3 \\ -1 \\ 0 \\ 2 \\ 5 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \\ -2 \\ 2 \end{pmatrix},$$

luego $\epsilon = \|\mathbf{Ax}_0 - \mathbf{b}\| = \sqrt{12}$.

Caso de coeficientes con columnas ortogonales

Muchos problemas de regresión lineal que necesitamos resolver por mínimos cuadrados dan lugar a una matriz de coeficientes cuyas columnas son ortogonales dos a dos. En esos casos, la solución del problema de mínimos cuadrados del sistema $\mathbf{Ax} = \mathbf{b}$ se puede obtener de forma sencilla calculando previamente la proyección ortogonal $\mathbf{b}_{\parallel} = \text{proy}_{\text{Col } A}(\mathbf{b})$ y luego resolviendo el sistema

$$\mathbf{Ax} = \mathbf{b}_{\parallel}.$$

Por ejemplo: Sean

$$A = \begin{pmatrix} -6 & 1 \\ -2 & 1 \\ 1 & 1 \\ 7 & 1 \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} -1 \\ 2 \\ 1 \\ 6 \end{pmatrix}$$

Se comprueba fácilmente que las columnas de A son ortogonales: $\mathbf{a}_1 \cdot \mathbf{a}_2 = -6 - 2 + 1 + 7 = 0$. En consecuencia, la proyección ortogonal de \mathbf{b} sobre $\text{Col } A$ es fácil de calcular:

$$\mathbf{b}_{\parallel} = \text{proy}_{\text{Col } A}(\mathbf{b}) = \left(\frac{\mathbf{b} \cdot \mathbf{a}_1}{\mathbf{a}_1 \cdot \mathbf{a}_1} \right) \mathbf{a}_1 + \left(\frac{\mathbf{b} \cdot \mathbf{a}_2}{\mathbf{a}_2 \cdot \mathbf{a}_2} \right) \mathbf{a}_2 = \frac{45}{90} \mathbf{a}_1 + \frac{8}{4} \mathbf{a}_2 = \begin{pmatrix} -3 \\ -1 \\ 1/2 \\ 7/2 \end{pmatrix} + \begin{pmatrix} 2 \\ 2 \\ 2 \\ 2 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \\ 5/2 \\ 11/2 \end{pmatrix}$$

y sólo tendríamos que resolver el sistema $A\mathbf{x} = \mathbf{b}_{\parallel}$. Pero ni siquiera necesitamos resolverlo porque ya conocemos los coeficientes de la combinación lineal de las columnas de A que es igual a \mathbf{b}_{\parallel} . Estos coeficientes son $\frac{45}{90} = \frac{1}{2}$ y $\frac{8}{4} = 2$, por tanto la solución por mínimos cuadrados es:

$$\mathbf{x} = \begin{pmatrix} 1/2 \\ 2 \end{pmatrix}$$

y el error asociado es la norma de

$$A\mathbf{x} - \mathbf{b} = \mathbf{b}_{\parallel} - \mathbf{b} = \begin{pmatrix} -1 \\ 1 \\ 5/2 \\ 11/2 \end{pmatrix} - \begin{pmatrix} -1 \\ 2 \\ 1 \\ 6 \end{pmatrix} = \begin{pmatrix} 0 \\ -1 \\ 3/2 \\ -1/2 \end{pmatrix},$$

o sea, $\epsilon = \|\mathbf{b}_{\parallel} - \mathbf{b}\| = \sqrt{1 + \left(\frac{3}{2}\right)^2 + \left(\frac{1}{2}\right)^2} = \frac{1}{2}\sqrt{14}$.

TEOREMA 7.4.1

Si las columnas $\mathbf{a}_1, \dots, \mathbf{a}_n$ de la matriz A son no nulas y ortogonales dos a dos entonces el problema de mínimos cuadrados asociado al sistema $A\mathbf{x} = \mathbf{b}$ tiene solución única, que está dada por los coeficientes que expresan de la proyección ortogonal de \mathbf{b} sobre $\text{Col } A$ como combinación lineal de las columnas de A :

$$\left(\frac{\mathbf{b} \cdot \mathbf{a}_1}{\mathbf{a}_1 \cdot \mathbf{a}_1}, \dots, \frac{\mathbf{b} \cdot \mathbf{a}_n}{\mathbf{a}_n \cdot \mathbf{a}_n} \right).$$

Unicidad de la solución del problema de mínimos cuadrados

Si A es una matriz $m \times n$, toda solución de $A\mathbf{x} = \mathbf{0}$ es, evidentemente solución de $A^T A\mathbf{x} = \mathbf{0}$, pero, recíprocamente, también toda solución de $A^T A\mathbf{x} = \mathbf{0}$ es solución de la anterior porque $A^T A\mathbf{x} = \mathbf{0}$ implica $\mathbf{x}^T A^T A\mathbf{x} = 0$, es decir $(A\mathbf{x}) \cdot (A\mathbf{x}) = 0$, que implica $A\mathbf{x} = \mathbf{0}$. Esto significa que:

*Las matrices A y $A^T A$ tienen el mismo espacio nulo (y por tanto, como tienen el mismo número de columnas, tienen también el mismo rango. Ver los ejercicios 17 a 20 y ver la sección **Consecuencia de la propiedad "definida positiva"** y sus corolarios (pág. 3)).*

Si las columnas de A son linealmente independientes, entonces el espacio nulo de A es cero y por tanto, según lo anterior, también el espacio nulo de $A^T A$ es cero, lo que significa que las columnas de $A^T A$ son linealmente independientes y (por ser una matriz cuadrada) $A^T A$ es inversible. Recíprocamente, si $A^T A$ es inversible, sus columnas, $A^T \mathbf{a}_1, \dots, A^T \mathbf{a}_n$ son independientes y esto

implica que las de A , $\mathbf{a}_1, \dots, \mathbf{a}_n$, también lo son (de lo contrario, las de $A^T A$ tendrían la misma relación de dependencia lineal). En consecuencia:

LEMA 7.4.1

Las columnas de A son linealmente independientes si y sólo si la matriz $A^T A$ es invertible.

TEOREMA 7.4.2

Si las columnas de la matriz de coeficientes de un sistema $A\mathbf{x} = \mathbf{b}$ son linealmente independientes, entonces el problema de mínimos cuadrados asociado tiene solución única, independientemente de cuál sea el vector de términos independientes \mathbf{b} , y esa solución es:

$$\mathbf{x} = (A^T A)^{-1} A^T \mathbf{b}. \quad (7.12)$$

Aplicación de los mínimos cuadrados al cálculo de la proyección ortogonal

Dada una base $\{\mathbf{u}_1, \dots, \mathbf{u}_r\}$ de un subespacio W de \mathbf{R}^n , hemos visto cómo plantear un sistema de ecuaciones para calcular las coordenadas c_i de la proyección ortogonal de un vector cualquiera \mathbf{v} sobre W :

$$\text{proy}_W(\mathbf{v}) = c_1 \mathbf{u}_1 + \dots + c_r \mathbf{u}_r.$$

En el sistema de ecuaciones satisfecho por las coordenadas c_i , la ecuación j es:

$$(\mathbf{u}_1 \cdot \mathbf{u}_j) c_1 + \dots + (\mathbf{u}_r \cdot \mathbf{u}_j) c_r = \mathbf{v} \cdot \mathbf{u}_j. \quad (7.13)$$

En términos de la matriz $U = [\mathbf{u}_1 \ \dots \ \mathbf{u}_r]$ cuyas columnas son los vectores de la base de W , la matriz de coeficientes de este sistema es la matriz simétrica $U^T U$:

$$U^T U = \begin{pmatrix} \mathbf{u}_1 \cdot \mathbf{u}_1 & \dots & \mathbf{u}_1 \cdot \mathbf{u}_r \\ \vdots & \ddots & \vdots \\ \mathbf{u}_r \cdot \mathbf{u}_1 & \dots & \mathbf{u}_r \cdot \mathbf{u}_r \end{pmatrix}$$

y las ecuaciones (7.13) significan que el vector \mathbf{c} de coordenadas verifica:

$$U^T U \mathbf{c} = U^T \mathbf{v}.$$

Por otra parte, el producto $U\mathbf{c}$ es justamente la combinación lineal $c_1 \mathbf{u}_1 + \dots + c_r \mathbf{u}_r$, luego la proyección ortogonal de \mathbf{v} sobre W es:

$$\text{proy}_W(\mathbf{v}) = c_1 \mathbf{u}_1 + \dots + c_r \mathbf{u}_r = U\mathbf{c},$$

en consecuencia tenemos lo siguiente:

PROPOSICIÓN 7.4.3

Si $\{\mathbf{u}_1, \dots, \mathbf{u}_r\}$ es una base de un subespacio W de \mathbf{R}^n , para hallar la proyección ortogonal de \mathbf{v} sobre W basta resolver el problema de mínimos cuadrados $U\mathbf{x} = \mathbf{v}$ cuya matriz de coeficientes es $U = [\mathbf{u}_1 \ \dots \ \mathbf{u}_r]$. Elegida una solución \mathbf{x} , la proyección ortogonal es: $\text{proy}_W(\mathbf{v}) = U\mathbf{x}$.

Teniendo en cuenta que los vectores $\{\mathbf{u}_1, \dots, \mathbf{u}_r\}$ son linealmente independientes (por ser base de W) y por tanto, según (7.12), $\mathbf{x} = (U^T U)^{-1} U^T \mathbf{v}$, vemos que

$$\text{proy}_W \mathbf{v} = U(U^T U)^{-1} U^T \mathbf{v}$$

y tenemos:

Si la columnas de U son linealmente independientes, la matriz

$$P_U = U(U^T U)^{-1} U^T$$

es la matriz de proyección sobre el espacio $W = \text{Col } U$.

En el caso particular de que la base $\{\mathbf{u}_1, \dots, \mathbf{u}_r\}$ del subespacio W sea una base ortonormal, entonces $U^T U = I_r$ y

$$P_U = U U^T, \quad \text{y} \quad \text{proy}_{\text{Col } U} \mathbf{v} = U U^T \mathbf{v}.$$

Ajuste polinómico mediante mínimos cuadrados

El problema de ajuste polinómico consiste en hallar un polinomio de grado fijado de antemano que “mejor se ajuste” a unos datos dados $(x_1, y_1), \dots, (x_k, y_k)$. Un ejemplo de esto es hallar la recta (polinomio de grado 1) que pase por dos puntos con diferente abscisas, $(x_1, y_1), (x_2, y_2)$, $x_1 \neq x_2$ (si las abscisas fuesen iguales la recta que pasa por los puntos no sería la gráfica de ninguna función). En este ejemplo la solución es muy sencilla, ya que el problema es el de hallar una recta de la forma $y = ax + b$ que pase por los dos puntos dados, es decir, que cumpla

$$\begin{aligned} ax_1 + b &= y_1 \\ ax_2 + b &= y_2 \end{aligned}$$

es decir,

$$\begin{pmatrix} x_1 & 1 \\ x_2 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}, \quad \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} x_1 & 1 \\ x_2 & 1 \end{pmatrix}^{-1} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \frac{1}{x_1 - x_2} \begin{pmatrix} 1 & -1 \\ -x_2 & x_1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$$

cuya solución es

$$a = \frac{y_1 - y_2}{x_1 - x_2}, \quad b = \frac{x_1 y_2 - x_2 y_1}{x_1 - x_2} \quad (7.14)$$

siempre que $x_1 \neq x_2$. Sin embargo, si los datos son tres puntos en lugar de dos, ya no existirá en general ninguna recta $y = ax + b$ que pase por los tres puntos porque éstos pueden no estar alineados. En ese caso, nuestro sistema será de la forma:

$$\begin{aligned} ax_1 + b &= y_1 \\ ax_2 + b &= y_2 \\ ax_3 + b &= y_3 \end{aligned}$$

y lo que todavía podemos hacer es hallar la solución de mínimos cuadrados resolviendo el sistema cuya matriz ampliada es el producto

$$\begin{pmatrix} x_1 & x_2 & x_3 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 & 1 & y_1 \\ x_2 & 1 & y_2 \\ x_3 & 1 & y_3 \end{pmatrix} = \begin{pmatrix} \sum_i x_i^2 & \sum_i x_i & \sum_i x_i y_i \\ \sum_i x_i & 3 & \sum_i y_i \end{pmatrix}.$$

Así se llega a las fórmulas de la *recta de regresión* que se estudian en Estadística.

En el problema general de ajuste polinómico se buscan los coeficientes a_0, a_1, \dots, a_n del polinomio $p(x) = a_0 + a_1x + \dots + a_nx^n$ de grado n que mejor se ajusta a unos datos dados $(x_1, y_1), \dots, (x_k, y_k)$. En este contexto “mejor ajuste” ocurre cuando el “error cuadrático”

$$\epsilon^2 = (p(x_1) - y_1)^2 + \dots + (p(x_k) - y_k)^2$$

tiene el menor valor posible. Claramente, este valor es igual a cero precisamente cuando el polinomio “pasa por todos los puntos”. Esta suma de cuadrados es el cuadrado de la norma eucídea del vector diferencia $\mathbf{p} - \mathbf{y}$, donde \mathbf{p} es el vector

$$\mathbf{p} = \begin{pmatrix} p(x_1) \\ \vdots \\ p(x_k) \end{pmatrix} = \begin{pmatrix} 1 & x_1 & x_1^2 & \dots & x_1^n \\ 1 & x_2 & x_2^2 & \dots & x_2^n \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_k & x_k^2 & \dots & x_k^n \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$$

Por tanto nuestro problema es minimizar $\|\mathbf{Ax} - \mathbf{y}\|$ con

$$A = \begin{pmatrix} 1 & x_1 & x_1^2 & \dots & x_1^n \\ 1 & x_2 & x_2^2 & \dots & x_2^n \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_k & x_k^2 & \dots & x_k^n \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{pmatrix}.$$

Otras variantes de este problema general se obtienen al imponer restricciones adicionales a los coeficientes de $p(x)$. Por ejemplo, en el siguiente ejercicio se busca un polinomio de segundo grado con el término independiente igual a cero, por tanto la parábola está determinada por dos coeficientes:

Ejercicio: Hallar la curva de ecuación $y = ax^2 + bx$ que mejor se ajusta, en el sentido de los mínimos cuadrados, a los siguientes puntos:

$$(1, 1), \quad (-1, 1), \quad (2, 2), \quad (-2, 2).$$

Al resolver este ejercicio conviene observar que al plantear el sistema $\mathbf{Ax} = \mathbf{b}$, las dos columnas de la matriz de coeficientes son perpendiculares y por tanto el método de los mínimos cuadrados se puede simplificar según lo dicho en la sección “Caso de coeficientes con columnas ortogonales” (página 20).

Solución:

El sistema a resolver es:

$$\begin{aligned} a \cdot 1^2 + b \cdot 1 &= 1 \\ a \cdot (-1)^2 + b \cdot (-1) &= 1 \\ a \cdot 2^2 + b \cdot 2 &= 2 \\ a \cdot (-2)^2 + b \cdot (-2) &= 2 \end{aligned}$$

De aquí, las ecuaciones se pueden expresar como

$$\begin{aligned} a + b &= 1 \\ a - b &= 1 \\ 4a + 2b &= 2 \\ 4a - 2b &= 2 \end{aligned} \quad \text{o bien:, en la forma matricial } \mathbf{Ax} = \mathbf{b}, \quad \begin{pmatrix} 1 & 1 \\ 1 & -1 \\ 4 & 2 \\ 4 & -2 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 2 \\ 2 \end{pmatrix}$$

Atención: Al plantear un problema de mínimos cuadrados hay que tener cuidado de *no* realizar operaciones elementales “simplificativas” en las ecuaciones del sistema original, pues con ello se podría estar cambiando la solución. Simplificaciones de ese tipo sólo se pueden realizar de forma segura en la *ecuaciones normales* del sistema. En este ejemplo, si simplificásemos las dos últimas ecuaciones dividiéndolas entre 2 llegaríamos al resultado $y = \frac{6}{10}x^2$, que no es la solución correcta (ver más abajo). La solución de mínimos cuadrados de este sistema se puede hallar resolviendo el sistema $A\mathbf{x} = \mathbf{b}_{\parallel}$ donde \mathbf{b}_{\parallel} es la proyección ortogonal de \mathbf{b} sobre el espacio columna de A . En este caso, como las dos columnas $\mathbf{a}_1, \mathbf{a}_2$ de A son ortogonales, el vector \mathbf{b}_{\parallel} se halla fácilmente:

$$\mathbf{b}_{\parallel} = \left(\frac{\mathbf{b} \cdot \mathbf{a}_1}{\mathbf{a}_1 \cdot \mathbf{a}_1} \right) \mathbf{a}_1 + \left(\frac{\mathbf{b} \cdot \mathbf{a}_2}{\mathbf{a}_2 \cdot \mathbf{a}_2} \right) \mathbf{a}_2 = \frac{18}{34} \begin{pmatrix} 1 \\ 1 \\ 4 \\ 4 \end{pmatrix} + \frac{0}{10} \begin{pmatrix} 1 \\ -1 \\ 2 \\ -2 \end{pmatrix}.$$

La solución de $A\mathbf{x} = \mathbf{b}_{\parallel}$ nos la dan evidentemente los coeficientes en esta combinación lineal:

$$\mathbf{x} = \begin{pmatrix} \frac{9}{17} \\ 0 \end{pmatrix}, \text{ es decir, } a = \frac{9}{17}, b = 0,$$

por tanto la curva pedida es

$$y = \frac{9}{17}x^2.$$

Ejercicios de la sección 7.4 Mínimos cuadrados

En los ejercicios 1 a 4, halla la solución de mínimos cuadrados del sistema $A\mathbf{x} = \mathbf{b}$ mediante la resolución de las ecuaciones normales.

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

$$2. A = \begin{pmatrix} 2 & 1 \\ -2 & 0 \\ 2 & 3 \end{pmatrix}, \mathbf{b} = \begin{pmatrix} -5 \\ 8 \\ 1 \end{pmatrix}$$

$$3. A = \begin{pmatrix} 1 & -2 \\ -1 & 2 \\ 0 & 3 \\ 2 & 5 \end{pmatrix}, \mathbf{b} = \begin{pmatrix} 3 \\ 1 \\ -4 \\ 2 \end{pmatrix}$$

$$4. A = \begin{pmatrix} 1 & 3 \\ 1 & -1 \\ 1 & 1 \end{pmatrix}, \mathbf{b} = \begin{pmatrix} 5 \\ 1 \\ 0 \end{pmatrix}$$

En los ejercicios 5 y 6, describe todas las soluciones de mínimos cuadrados de la ecuación $A\mathbf{x} = \mathbf{b}$.

$$5. A = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix}, \mathbf{b} = \begin{pmatrix} 1 \\ 3 \\ 8 \\ 2 \end{pmatrix}$$

$$6. A = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix}, \mathbf{b} = \begin{pmatrix} 7 \\ 2 \\ 3 \\ 6 \\ 5 \\ 4 \end{pmatrix}$$

7. Calcula el error asociado con la solución de mínimos cuadrados del ejercicio 3.

8. Calcula el error asociado con la solución de mínimos cuadrados del ejercicio 4.

En los ejercicios 9 a 12, halla la proyección ortogonal de \mathbf{b} sobre $\text{Col } A$ y la solución de mínimos cuadrados de $A\mathbf{x} = \mathbf{b}$.

$$9. A = \begin{pmatrix} 1 & 5 \\ 3 & 1 \\ -2 & 4 \end{pmatrix}, \mathbf{b} = \begin{pmatrix} 4 \\ -2 \\ -3 \end{pmatrix}$$

$$10. A = \begin{pmatrix} 1 & 2 \\ -1 & 4 \\ 1 & 2 \end{pmatrix}, \mathbf{b} = \begin{pmatrix} 3 \\ -1 \\ 5 \end{pmatrix}$$

$$11. A = \begin{pmatrix} 4 & 0 & 1 \\ 1 & -5 & 1 \\ 6 & 1 & 0 \\ 1 & -1 & -5 \end{pmatrix}, \mathbf{b} = \begin{pmatrix} 9 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

$$12. A = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & 1 \\ -1 & 1 & -1 \end{pmatrix}, \mathbf{b} = \begin{pmatrix} 2 \\ 5 \\ 6 \\ 6 \end{pmatrix}$$

En los ejercicios 13 y 14, usa la factorización QR de A dada para hallar la solución de mínimos cuadrados de $A\mathbf{x} = \mathbf{b}$.

$$13. A = \begin{pmatrix} 2 & 3 \\ 2 & 4 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} \frac{2}{3} & -\frac{1}{3} \\ \frac{2}{3} & \frac{2}{3} \\ \frac{1}{3} & -\frac{2}{3} \end{pmatrix} \begin{pmatrix} 3 & 5 \\ 0 & 1 \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} 7 \\ 1 \end{pmatrix}.$$

$$14. A = \begin{pmatrix} 1 & -1 \\ 1 & 4 \\ 1 & -1 \\ 1 & 4 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} 2 & 3 \\ 0 & 5 \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} -1 \\ 6 \\ 3 \\ 7 \end{pmatrix}.$$

En los ejercicios 15 y 16, A es una matriz $m \times n$ y $\mathbf{b} \in \mathbb{R}^n$. Indica para cada enunciado si es verdadero o falso. Justifica tus respuestas.

15.

- El problema de mínimos cuadrados asociado a un sistema $A\mathbf{x} = \mathbf{b}$ consiste en hallar un \mathbf{x} que haga $A\mathbf{x}$ lo más cerca posible de \mathbf{b} .
- Una solución del problema de mínimos cuadrados de $A\mathbf{x} = \mathbf{b}$ es un vector $\hat{\mathbf{x}}$ que satisface $A\hat{\mathbf{x}} = \mathbf{b}_{\parallel}$ donde \mathbf{b}_{\parallel} es la proyección ortogonal de \mathbf{b} sobre $\text{Col } A$.
- Una solución del problema de mínimos cuadrados de $A\mathbf{x} = \mathbf{b}$ es un vector $\hat{\mathbf{u}}$ tal que $\|\mathbf{b} - A\hat{\mathbf{u}}\| \leq \|\mathbf{b} - A\mathbf{u}\|$ para todo $\mathbf{u} \in \mathbb{R}^n$.
- Toda solución de $A^T A\mathbf{x} = A^T \mathbf{b}$ es una solución de mínimos cuadrados de $A\mathbf{x} = \mathbf{b}$.
- Si las columnas de A son linealmente independientes, entonces el problema de mínimos cuadrados de la ecuación $A\mathbf{x} = \mathbf{b}$ tiene exactamente una solución.

16.

- Si \mathbf{b} pertenece al espacio columna de A , entonces toda solución de $A\mathbf{x} = \mathbf{b}$ es una solución de mínimos cuadrados.
- La solución del problema de mínimos cuadrados de $A\mathbf{x} = \mathbf{b}$ es el vector del espacio columna de A más cercano a \mathbf{b} .
- Una solución de mínimos cuadrados de $A\mathbf{x} = \mathbf{b}$ es una lista de coeficientes que, cuando se usan para formar una combinación lineal de las columnas de A , el resultado es la proyección ortogonal de \mathbf{b} sobre $\text{Col } A$.
- Si $\hat{\mathbf{x}}$ es una solución del problema de mínimos cuadrados de $A\mathbf{x} = \mathbf{b}$, entonces $\hat{\mathbf{x}} = (A^T A)^{-1} A^T \mathbf{b}$.

- Las ecuaciones normales siempre proporcionan un método seguro para calcular soluciones de problemas de mínimos cuadrados.
- Si A tiene una factorización QR, $A = QR$, entonces la mejor forma de hallar la solución del problema de mínimos cuadrados de $A\mathbf{x} = \mathbf{b}$ es calcular $\hat{\mathbf{x}} = R^{-1} Q^T \mathbf{b}$.

17. Sea A una matriz $m \times n$. Usa los pasos que se indican a continuación para demostrar que un vector $\mathbf{x} \in \mathbb{R}^n$ satisface $A\mathbf{x} = \mathbf{0}$ si y sólo si $A^T A\mathbf{x} = \mathbf{0}$, es decir, que $\text{Nul } A = \text{Nul } A^T A$.

- Demuestra que si $A\mathbf{x} = \mathbf{0}$, entonces $A^T A\mathbf{x} = \mathbf{0}$.
- Suponiendo $A^T A\mathbf{x} = \mathbf{0}$, explica por qué $\mathbf{x}^T A^T A\mathbf{x} = 0$ y usa esto para demostrar que $A\mathbf{x} = \mathbf{0}$.

18. Sea A una matriz $m \times n$ tal que $A^T A$ es invertible. Demuestra que las columnas de A son linealmente independientes. (Cuidado: no se puede suponer que A sea invertible ni que sea siquiera cuadrada.)

19. Sea A una matriz $m \times n$ cuyas columnas son linealmente independientes.

- Usa el ejercicio 17 para demostrar que $A^T A$ es una matriz invertible.
- Explica por qué A no puede tener menos filas que columnas.
- Halla el rango de A .

20. Usa el ejercicio 17 para demostrar que para cualquier matriz $m \times n$, A , $\text{Rang}(A^T A) = \text{Rang}(A)$.

Pista: ¿Cuántas columnas tiene $A^T A$? ¿Qué relación tiene esto con el rango de $A^T A$?

21. Sea A una matriz $m \times n$ cuyas columnas son linealmente independientes y sea $\mathbf{b} \in \mathbb{R}^m$. Usa las ecuaciones normales del sistema $A\mathbf{x} = \mathbf{b}$ para obtener una fórmula que nos dé la proyección ortogonal de \mathbf{b} sobre $\text{Col } A$.

Pista: Halla primero una fórmula para la solución, $\hat{\mathbf{x}}$, del problema de mínimos cuadrados $A\mathbf{x} = \mathbf{b}$.

En los ejercicios 22 a 25 halla la ecuación $y = ax + b$ de la recta de regresión que mejor se ajuste a los puntos dados.

22. $(0, 1), (1, 1), (2, 2), (3, 2)$.

23. $(1, 0), (2, 1), (4, 2), (5, 3)$.

24. $(-1, 0), (0, 1), (1, 2), (2, 4)$.

25. $(2, 3), (3, 2), (5, 1), (6, 0)$.

26. Cierta experimento produjo los datos $(1, 1'8), (2, 2'7), (3, 3'4), (4, 3'8)$, y $(5, 3'9)$ halla la función de la forma $y = ax + \beta x^2$ que mejor se ajusta a esos datos.

27. Cierta experimento produjo los datos $(1, 7'9), (2, 5'4)$ y $(3, -0'9)$ halla la función de la forma $y = A \cos x + B \sin x$ que mejor se ajusta a esos datos.