Capítulo 3 - Ortogonalidad (3ra. parte) [1]

Facultad de Cs. Exactas, Ingeniería y Agrimensura
Universidad Nacional de Rosario

[1] Siguiendo *Linear Algebra and its applications*, G. Strang.

Outline

V: espacio vectorial con producto interno,
W_1, W_2: subespacios vectoriales de V:

- W_1 y W_2 son *ortogonales* si, para todo $u \in W_1$ y todo $v \in W_2$, se verifica $u \perp v$. Notamos: $W_1 \perp W_2$.
- *Condición necesaria*: $W_1 \perp W_2 \Longrightarrow W_1 \cap W_2 = \{0\}$.
- *Condición suficiente*: Sean U_1, U_2 tales que $W_1 = \langle U_1 \rangle$ y $W_2 = \langle U_2 \rangle$. Entonces, $\forall u \in U_1, v \in U_2, u \perp v \Longrightarrow W_1 \perp W_2$.
- $W_1 \perp W_2 \Longrightarrow dim(W_1) + dim(W_2) \leq dim(V)$.
- Si $W_1 \perp W_2$ y $dim(W_1) + dim(W_2) = dim(V)$ entonces $W_2 = \{u \in V : u \perp v$ para todo $v \in W_1\}$.
- $\{u \in V : u \perp v$ para todo $v \in W_1\}$ es el *complemento ortogonal de* W_1. Lo notamos $\{u \in V : u \perp v$ para todo $v \in W_1\} = W_1^{\perp}$.
- $W_1^{\perp}$ es un subespacio de V, $W_1^{\perp} \perp W_1$ y $(W_1^{\perp})^{\perp} = W_1$.

Tenemos entonces:

$$W_1 \perp W_2 \ , \ dim(W_1) + dim(W_2) = dim(V) \Longrightarrow W_2 = W_1^{\perp}.$$

Teorema Fundamental del Álgebra Lineal (Parte II):
Sea A una matriz $m \times n$. Entonces,

1. El espacio fila y el espacio nulo de A son complementos ortogonales en $\mathbb{R}^n$.
2. El espacio columna y el espacio nulo a izquierda de A son complementos ortogonales en $\mathbb{R}^m$.

Dijimos la clase pasada: *Veremos más adelante que la recíproca del lema anterior es cierta, esto es,*

$$W_1 \perp W_2 \ , \ dim(W_1) + dim(W_2) = dim(V) \Longleftrightarrow W_2 = W_1^{\perp}.$$

Por ahora podemos probarlo en el caso de que alguno de los subespacios tenga una *base ortogonal*.

Teorema: Sea V un espacio vectorial con producto interno, W_1, W_2 subespacios ortogonales de V, W_1 tiene una base ortogonal. Entonces, $W_2 = W_1^{\perp}$ si y solo si $dim(W_1) + dim(W_2) = dim(V)$

Prueba: Ya probamos que si $dim(W_1) + dim(W_2) = dim(V)$ entonces $W_2 = W_1^{\perp}$. Sólo resta probar la recíproca.

Sean W_1 y W_2 tales que $W_2 = W_1^{\perp}$. Como son ortogonales, ya sabemos que $dim(W_1) + dim(W_2) \leq dim(V)$. Más aún, sabemos que si $\mathscr{B}_i$ es una base de W_i para $i = 1,2$, entonces $\mathscr{B} = \mathscr{B}_1 \cup \mathscr{B}_2$ es un conjunto de vectores l.i.. Sea $\mathscr{B}_1 = \{w^i : i = 1, \ldots, p\}$ una base ortogonal de W_1 y $W =< \mathscr{B} >$.

Supongamos que $dim(W_1) + dim(W_2) \leq dim(V) - 1$. Entonces, existe $v \in V \setminus W$.
Construimos

$$z = v - \sum_{i=1}^{p} \frac{\langle v, w^i \rangle}{\|w^i\|^2} w^i.$$

Veremos que $z \perp w^j$ para todo $j = 1, \ldots, p$.

Prueba (continuación)
En efecto, dado $j \in \{1,\ldots,p\}$,

$$\left\langle z, w^j \right\rangle = \left\langle v - \sum_{i=1}^{p} \frac{\left\langle v, w^i \right\rangle}{\|w^i\|^2} w^i, w^j \right\rangle = \left\langle v, w^j \right\rangle - \sum_{i=1}^{p} \frac{\left\langle v, w^i \right\rangle}{\|w^i\|^2} \left\langle w^i, w^j \right\rangle.$$

Como $\mathscr{B}_1 = \{w^i : i = 1,\ldots,p\}$ es una base ortogonal de W_1, tenemos que $\left\langle w^i, w^j \right\rangle = 0$ si $i \neq j$. Por lo tanto,

$$\left\langle z, w^j \right\rangle = \left\langle v, w^j \right\rangle - \frac{\left\langle v, w^j \right\rangle}{\|w^j\|^2} \left\langle w^j, w^j \right\rangle = 0.$$

Como $z \neq 0$ (justificar) y $z \perp w^j$ para todo $j = 1,\ldots,p$ resulta que $z \perp w$ para todo $w \in W_1$. Por lo tanto $z \in W_1^{\perp} = W_2$, una contradicción (justificar).

Por lo tanto, $dim(W_1) + dim(W_2) = dim(V)$. □

Veremos que la hipótesis " W_1 *tiene una base ortogonal*" no es demasiado restrictiva.

V: espacio vectorial con producto interno.

Veremos un algoritmo que, a partir de p vectores l.i. de V, genera p vectores ortogonales no nulos en V.

Proceso de Ortogonalización de Gram-Schmidt:
ENTRADA: $\{v^i : i = 1, \ldots, p\}$: vectores l.i. de V.
SALIDA: $\{w^i : i = 1, \ldots, p\}$: vectores ortogonales no nulos de V.

PROCESO:

1. $w^1 := v^1$.
2. Para $j = 2, \ldots, p$ hacer

$$w^j = v^j - \sum_{i=1}^{j-1} \frac{\langle v^j, w^i \rangle}{\|w^i\|^2} w^i.$$

Veamos que el algoritmo es correcto, o sea, para todo i, $w^i \neq 0$ y, para todo $i \neq j$, $w^i \perp w^j$.

Debemos probar que:

- $w^i \neq 0$ para todo i,
- $w^i \perp w^j$ para todo $i \neq j$.

Observar que, como $v^i \neq 0$ para todo $i = 1, \ldots, p$ (justificar), es suficiente probar, para todo $i = 2, \ldots, p$,

$$w^i \neq 0 \qquad \text{y} \qquad w^i \perp w^j \ \ \forall j : 1 \leq j \leq i-1.$$

Probamos por inducción lo siguiente:

Para todo $i = 2, \ldots, p$,

- w^i es una combinación lineal de v^j, $1 \leq j \leq i$,
- $w^i \neq 0$,
- $w^i \perp w^j$ para todo $j : 1 \leq j \leq i-1$.

Para todo $i = 2, \ldots, p$,

1. w^i es una combinación lineal de $v^j, j = 1, \ldots, i$,
2. $w^i \neq 0$,
3. $w^i \perp w^j$ para todo $j : 1 \leq j \leq i-1$.

Prueba:

- $i = 2$:

$$w^2 = v^2 - \frac{\langle v^2, w^1 \rangle}{\|w^1\|^2} w^1 = v^2 - \frac{\langle v^2, v^1 \rangle}{\|v^1\|^2} v^1.$$

Claramente, w^2 es una combinación lineal de v^1 y v^2. Por la lineal independencia de v^1 y v^2 podemos concluir que $w^2 \neq 0$.

Falta verificar $w^2 \perp w^1$, o sea, $w^2 \perp v^1$.

$$\langle w^2, v^1 \rangle = \langle v^2, v^1 \rangle - \frac{\langle v^2, v^1 \rangle}{\|v^1\|^2} \langle v^1, v^1 \rangle = 0$$

Para todo $i = 2, \ldots, p$,

1. w^i es una combinación lineal de $v^j, j = 1, \ldots, i$,
2. $w^i \neq 0$,
3. $w^i \perp w^j$ para todo $j : 1 \leq j \leq i-1$.

Prueba: (continuación)

- Sea k tal que $2 \leq k \leq p-1$ y w^j verifica $(1), (2)$ y (3) para todo $j \leq k$. Debemos probar que w^{k+1} verifica $(1), (2)$ y (3).

  Sabemos que
  $$w^{k+1} = v^{k+1} - \sum_{i=1}^{k} \frac{\left\langle v^{k+1}, w^i \right\rangle}{\left\| w^i \right\|^2} w^i.$$

  Por hipótesis de inducción, para todo $i = 1, \ldots, k$, w^i es combinación lineal de v^t, $t = 1, \ldots, i$. Por lo tanto, w^{k+1} es combinación lineal de v^i, $i = 1, \ldots, k+1$. Otra vez, la lineal independencia de v^i, $i = 1, \ldots, k+1$ nos asegura que $w^{k+1} \neq 0$.

Prueba:

- (continuación) Sólo falta probar que $w^{k+1} \perp w^j$ para todo $j = 1, \ldots, k$.

$$\left\langle w^{k+1}, w^j \right\rangle = \left\langle v^{k+1}, w^j \right\rangle - \sum_{i=1}^{k} \frac{\left\langle v^{k+1}, w^i \right\rangle}{\left\| w^i \right\|^2} \left\langle w^i, w^j \right\rangle.$$

Por hipótesis de inducción, $\left\langle w^i, w^j \right\rangle = 0$ para todo $1 \leq j \leq k$, $i = 1, \ldots, k-1$. Por lo tanto,

$$\left\langle w^{k+1}, w^j \right\rangle = \left\langle v^{k+1}, w^j \right\rangle - \left\langle v^{k+1}, w^j \right\rangle = 0.$$

$\square$

El Proceso de Ortogonalización de Gram-Schmidt nos permite probar el siguiente importante resultado:

Teorema: Todo espacio vectorial con producto interno y de dimensión finita tiene una base ortonormal.
Prueba: Ejercicio.

Habíamos probado:

- V, espacio vectorial con producto interno, $W_1 \perp W_2$ tal que W_1 tiene una base ortogonal. Entonces,
  $W_2 = W_1^{\perp}$ si y solo si $dim(W_1) + dim(W_2) = dim(V)$

Gracias a Gram-Schmidt ahora podemos probar y dar por válido el teorema que enunciamos en la clase pasada, sin prueba:
Teorema: Sea V un espacio vectorial de dimensión finita, con producto interno y W_1, W_2 subespacios ortogonales de V. Entonces, $W_2 = W_1^{\perp}$ si y solo si $dim(W_1) + dim(W_2) = dim(V)$.

También vimos el siguiente resultado, consecuencia del teorema anterior:

Teorema (de descomposición por complementos ortogonales): V, espacio vectorial de dimensión finita con producto interno. Entonces, $V = W \oplus W^{\perp}$ para todo W subespacio de V.

Corolario: $\mathbb{R}^n = C(A^T) \oplus N(A)$ y $\mathbb{R}^m = C(A) \oplus N(A^T)$, para toda matriz A de tamaño $m \times n$.

V: espacio vectorial con producto interno
W: subespacio vectorial de V

$V = W \oplus W^{\perp}$: para todo $v \in V$, existen únicos $w \in W$ y $w^{\perp} \in W^{\perp}$ tales que $v = w + w^{\perp}$.

¿Cómo obtenemos a w y $w^{\perp}$?

Si $\mathscr{B}_1 = \{w^i : i = 1, \ldots, p\}$ y $\mathscr{B}_2 = \{z^i : i = 1, \ldots, k\}$ son bases ortonormales de W_1 y W_2, respectivamente, entonces

$$w = \sum_{i=1}^{p} \left\langle v, w^i \right\rangle w^i \quad \text{y} \quad w^{\perp} = \sum_{i=1}^{k} \left\langle v, z^i \right\rangle z^i.$$

Vimos ejemplos en $\mathbb{R}^2$ y $\mathbb{R}^3$ donde se visualizaba que $w = proy_{s/W}\, v$ y $w^{\perp} = proy_{s/W^{\perp}}\, v$.

Esto nos llevó a la definición general de proyección ortogonal de un vector sobre un subespacio.

Proyección sobre subespacios

V: espacio vectorial con producto interno
W: subespacio de V con base ortogonal $\{w^1, \ldots, w^p\}$.

El *vector proyección ortogonal de* $v \in V$ *sobre* W es

$$proy_{s/W}\, v = \sum_{i=1}^{p} \|v\| \, cos(v\hat{\ } w^i) \, \frac{w^i}{\|w^i\|} = \sum_{i=1}^{p} \frac{\langle v, w^i \rangle}{\|w^i\|^2} \, w^i.$$

Terminamos la clase anterior comentando el rol fundamental de las proyecciones en aplicaciones. Veamos un ejemplo sencillo y simplificado.

Para conocer la velocidad (constante) v con la que un satélite viaja hacia Marte, hacemos $k+1$ mediciones en tiempos $t_0 = 0 < t_1 < \ldots < t_k$ y obtenemos las distancias b_i, $i = 0, \ldots k$. Sabemos que v debe verificar,

$$t_i v = b_i - b_0, \;\; i = 1, \ldots, k.$$

Sistema con k ecuaciones y una variable, los errores de medición no nos aseguran que tenga solución.

Sabemos que v debe verificar,

$$t_i v = b_i - b_0, \;\; i = 1, \ldots, k.$$

Sean $T = (t_1, \ldots, t_k)^T$ y $b = (b_1 - b_0, \ldots, b_k - b_0)^T$. T matriz $k \times 1$, $b \in \mathbb{R}^k$ y el sistema tiene la forma $Tv = b$, $v \in \mathbb{R}$.
Si $b \in C(T)$, el sistema tiene solución (única). Caso contrario,

¿qué valor de v tomamos?

Buscamos *la mejor aproximación de b en $C(T)$*. Esto es, buscamos $\tilde{b} \in C(T)$ tal que la distancia $\left\| b - \tilde{b} \right\|$ sea mínima. Luego elegimos v tal que $Tv = \tilde{b}$.

¿Cómo conseguimos $\tilde{b}$?

Como $\tilde{b} \in C(T)$, sabemos que $\tilde{b} = \alpha T$, para algún $\alpha \in \mathbb{R}$.

Si $k=3$, $C(T)$ es una recta en $\mathbb{R}^3$ que pasa por el origen y $b \in \mathbb{R}^3$ un punto por fuera de la recta. Es claro que *el punto más cercano a b* en la recta, se encuentra en el pie de la perpendicular a la recta que pasa por b. O sea, buscamos $\tilde{b}$ tal que $(b-\tilde{b}) \perp T$. Como $\tilde{b} = \alpha T$ para algún $\alpha \in \mathbb{R}$, tenemos:

$$\left\langle b-\tilde{b}, T\right\rangle = \langle b-\alpha T, T\rangle = \langle b, T\rangle - \langle \alpha T, T\rangle = \langle b, T\rangle - \alpha \langle T, T\rangle = 0$$

y por lo tanto $\alpha = \frac{\langle b,T\rangle}{\|T\|^2}$. Así,

$$\tilde{b} = \alpha T = \frac{\langle b,T\rangle}{\|T\|^2}\, T = proy_{s/C(T)}\, b.$$

Si tuviéramos un sistema inconsistente $Ax=b$, con A matriz 3×2 y $b \in \mathbb{R}^2$ y queremos encontrar *la mejor solución* $\hat{x}$ del sistema, buscaríamos $\tilde{b}$ *la mejor aproximación de b* sobre $C(A)$.

En este caso $C(A)$ puede pensarse como un plano en $\mathbb{R}^3$ que pasa por el origen y b un vector fuera de él. Nuevamente, para encontrar *la mejor aproximación* de b en el plano, buscaríamos el pie de la perpendicular al plano que pasa por b.

Ejercicio: Sea W un subespacio de $\mathbb{R}^3$ con base ortonormal $\{w_1, w_2\}$ y $b \in \mathbb{R}^3 \setminus W$. Sea $\tilde{b} \in W$ tal que $b - \tilde{b} \perp W$. Probar que $\tilde{b} = proy_{s/W} b$.

Esta idea intuitiva respecto a la relación entre *la mejor aproximación* y la proyección son correctas en todos los espacios de dimensión finita (y para algunos de dimensión infinita).

Formalizamos definiciones:

Definición: Dado un espacio vectorial V con producto interno, W un subespacio de V y $v \in V$. Decimos que $\bar{w}$ es *una mejor aproximación de v en W* si, para todo $w \in W$,

$$\|\bar{w} - v\| \leq \|w - v\| .$$

Tenemos el siguiente resultado:

Teorema: Sea W un subespacio vectorial de V, espacio vectorial sobre $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ con producto interno. Sea $v \in V$.

1. $\bar{w} \in W$ es una mejor aproximación de v en W si y solo si $v - \bar{w}$ es ortogonal a W.
2. Si existe una mejor aproximación de v en W entonces es única.
3. Si W es de dimensión finita y $\{z^i : i = 1, \ldots, k\}$ es una base ortonormal de W, entonces el vector

$$proy_{s/W}v = \sum_{i=1}^{k} \left\langle v, z^i \right\rangle z^i$$

   es la única mejor aproximación a v en W.

Prueba:

1. Para todo $w \in W$ se verifica

$$\|w-v\|^2 = \|(w-\bar{w})+(\bar{w}-v)\|^2 =$$
$$= \langle (w-\bar{w})+(\bar{w}-v),(w-\bar{w})+(\bar{w}-v)\rangle =$$
$$= \|w-\bar{w}\|^2 + 2\mathscr{R}e(\langle w-\bar{w},\bar{w}-v\rangle) + \|\bar{w}-v\|^2 .$$

Sea $\bar{w} \in W$ tal que $(v-\bar{w})\perp W$.
Entonces,

$$\|w-v\|^2 = \|w-\bar{w}\|^2 + \|\bar{w}-v\|^2 \geq \|\bar{w}-v\|^2 .$$

Por lo tanto $\bar{w}$ una mejor aproximación de v en W.

Sea ahora $\bar{w} \in W$ tal que, para todo $w \in W$, $\|w-v\| \geq \|\bar{w}-v\|$.
Debemos probar que $(v-\bar{w})\perp W$ o, equivalentemente,
$\langle v-\bar{w},w\rangle = 0$ para todo $w \in W$.

Prueba:

1. (continuación)
   Como para todo $w \in W$.
   $$\|w - v\|^2 = \|w - \bar{w}\|^2 + 2\mathscr{R}e(\langle w - \bar{w}, \bar{w} - v\rangle) + \|\bar{w} - v\|^2$$
   y $\|w - v\| \geq \|\bar{w} - v\|$, obtenemos
   $$\|\bar{w} - w\|^2 + 2\mathscr{R}e(\langle w - \bar{w}, \bar{w} - v\rangle) \geq 0$$
   o, equivalentemente (justificar), para todo $u \in W$,
   $$\|u\|^2 + 2\mathscr{R}e(\langle u, \bar{w} - v\rangle) \geq 0$$
   Supongamos que existe $u^* \in W$ tal que $\langle u^*, \bar{w} - v\rangle \neq 0$. Sin pérdida de generalidad, podemos suponer que $\mathscr{R}e(\langle u^*, \bar{w} - v\rangle) < 0$. (Justificar). Sea
   $$\alpha = \frac{\mathscr{R}e(\langle u^*, v - \bar{w}\rangle)}{\|u^*\|^2} > 0$$
   y consideremos $\frac{\alpha}{2}u^* \in W$.

Prueba:

1. (continuación) Como $\frac{\alpha}{2}u^* \in W$, vale

$$\left\|\frac{\alpha}{2}u^*\right\|^2 + 2\mathscr{R}e(\left\langle \frac{\alpha}{2}u^*, \bar{w} - v \right\rangle) \geq 0$$

o, equivalentemente,

$$\frac{\alpha^2}{4}\|u^*\|^2 - \alpha\mathscr{R}e(\langle u^*, v - \bar{w}\rangle) \geq 0.$$

Entonces,

$$\frac{\alpha}{4}\|u^*\|^2 - \mathscr{R}e(\langle u^*, v - \bar{w}\rangle) \geq 0,$$

resultando

$$\frac{\alpha}{4} \geq \frac{\mathscr{R}e(\langle u^*, v - \bar{w}\rangle)}{\|u^*\|^2} = \alpha,$$

una contradicción. Por lo tanto, para todo $u \in W$ resuelta $\langle u, \bar{w} - v\rangle = 0$ y $(\bar{w} - v)\perp W$.

Prueba: (continuación)

(2) Sean $z, \bar{w} \in W$ mejores aproximaciones de v en W. Entonces, $\bar{w} - z \in W$.

Además, $\bar{w} - z = (\bar{w} - v) + (v - z)$. Como z y $\bar{w}$ son mejores aproximaciones de v en W, por el item (1) sabemos que $(\bar{w} - v) \in W^{\perp}$ y $(v - z) \in W^{\perp}$. Por lo tanto, $\bar{w} - z \in W^{\perp}$.

Tenemos entonces $\bar{w} - z \in W \cap W^{\perp}$ y resulta $\bar{w} - z = 0$.

Prueba: (continuación)

(3) De acuerdo a los items anteriores, sólo debemos probar que

$$\bar{w} = \sum_{i=1}^{k} \left\langle v, z^i \right\rangle z^i$$

verifica $(v - \bar{w}) \perp W$. Para ello, basta probar $\left\langle v - \bar{w}, z^j \right\rangle = 0$, para todo $j = 1, \ldots, n$.

Es fácil ver que, para todo $j = 1, \ldots, n$,

$$\left\langle \bar{w}, z^j \right\rangle = \sum_{i=1}^{k} \left\langle v, z^i \right\rangle \left\langle z^i, z^j \right\rangle = \left\langle v, z^j \right\rangle \left\langle z^j, z^j \right\rangle = \left\langle v, z^j \right\rangle.$$

Por lo tanto, para todo $j = 1, \ldots, n$,

$$\left\langle v - \bar{w}, z^j \right\rangle = \left\langle v, z^j \right\rangle - \left\langle \bar{w}, z^j \right\rangle = \left\langle v, z^j \right\rangle - \left\langle v, z^j \right\rangle = 0$$

resultando $\bar{w}$ la mejor aproximación de v en W.

Volvamos al algoritmo de $G-S$. Su paso *recursivo* era:

Para $j=2,\ldots,p$ hacer

$$w^j = v^j - \sum_{i=1}^{j-1} \frac{\left\langle v^j, w^i \right\rangle}{\|w^i\|^2} w^i.$$

Observemos que

$$\sum_{i=1}^{j-1} \frac{\left\langle v^j, w^i \right\rangle}{\|w^i\|^2} w^i$$

no es otra cosa que la proyección de v^j sobre el subespacio $W^{(j-1)}$ generado por los vectores (ortogonales) $w^1,\ldots,w^{j-1}$.
Por lo tanto,

$$w^j = v^j - proy_{s/W^{(j-1)}}\, v^j$$

es el *vector error* en la aproximación de v^j en $W^{(j-1)}$. También es la componente de v^j sobre el complemento ortogonal de $W^{(j-1)}$.

Veremos en la práctica, a través de ejemplos, cómo estos procesos se aplican en espacios vectoriales de funciones.

Vimos, por ejemplo, que las funciones potencias son base del espacio de polinomios. Considerando la integral del producto de la funciones en el intervalo $[-1, 1]$ como producto interno, las potencias no resultan ortogonales. Aplicando G-S a estas funciones, obtenemos los denominados *polinomios de Legendre*.

Matrices de Proyección

A partir de ahora, trabajamos en $\mathbb{R}^n$ con suma, producto por escalar y producto interno habituales.

En el ejemplo que vimos para aproximar la velocidad del satélite, proyectábamos el vector b correspondiente a las distancias observadas sobre el subespacio generado por el vector de tiempos de medición T. La velocidad aproximada es $v \in \mathbb{R}$ tal que $Tv = proy_{s/<T>}\, b$.

En general, la proyección de un vector $b \in \mathbb{R}^n$ sobre la recta generada por otro vector $a \in \mathbb{R}^n$ es

$$proy_{s/<a>}\, b = \frac{a^T b}{a^T a}\, a.$$

Observar que $\underbrace{(a^T b)}_{\alpha \in \mathbb{R}}\, a = a\,(a^T b) = (a\, a^T)\, b$. Si consideramos la matriz $P = \frac{1}{a^T a}\, aa^T$, resulta

$$proy_{s/<a>}\, b = Pb.$$

Así, la proyección de un vector de $\mathbb{R}^n$ sobre una recta puede ser pensada como una transformación lineal de $\mathbb{R}^n$ en $\mathbb{R}^n$.

Observaciones:

- P es simétrica. Prueba: $(a\,a^T)^T = (a^T)^T\ a^T = aa^T$
- $P^2 = P$. Prueba: (i) $Pb \in < a >$. $\quad (ii)\, d \in < a > \Longrightarrow proy_{s/<a>}d = d$.

Ejemplo: Proyección sobre la recta en $\mathbb{R}^2$ que forma un ángulo θ con el eje x. La recta es el subespacio generado por el vector $a = (cos\theta, sen\theta)$. Como $\|a\| = 1$,

$$P = a\,a^T = \begin{bmatrix} cos^2\theta & cos\theta sen\theta \\ cos\theta sen\theta & sen^2\theta \end{bmatrix}$$

Esta transformación lineal ya la habíamos visto en el Capítulo 2.

Veamos ahora las proyecciones sobre un subespacio general W de $\mathbb{R}^n$.
Si $\{a^i : i = 1, \ldots, k\}$ es una base de W, definimos A la matriz $n \times k$ cuya columna i-ésima $A^i = a^i$ de manera que $W = C(A)$. Dado $b \in \mathbb{R}^n$, buscamos $\hat{b}$, la proyección de b en $W = C(A)$. O sea, $\hat{b} = A\hat{x}$ para algún $\hat{x} \in \mathbb{R}^k$.

Sabemos que la proyección que buscamos es *la mejor aproximación de* b en W y por lo tanto, el *error* $e = b - A\hat{x}$ debe ser perpendicular al espacio $C(A)$.
O sea, $b - A\hat{x} \in N(A^T)$.
Tenemos entonces:

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

Observación: Si un sistema $Ax = b$ es inconsistente, la *mejor solución* $\hat{x}$ (que minimiza el error) es solución del *sistema de ecuaciones normales*:

$$(A^T A)\,\hat{x} = A^T b.$$

Ejercicio: Si las columnas de A son l.i., A^TA es inversible.

En ese caso,

$$\hat{x} = (A^TA)^{-1}A^Tb$$

y la proyección $\hat{b}$ de b en el espacio columna es

$$\hat{b} = A\hat{x} = A(A^TA)^{-1}A^Tb.$$

Ejemplo: Sean

$$A = \begin{bmatrix} 1 & 2 \\ 1 & 3 \\ 0 & 0 \end{bmatrix} \;,\; b = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}.$$

Aquí el subespacio $W = C(A)$ resulta ser el plano xy en $\mathbb{R}^3$ y es fácil intuir que $\hat{b} = (4,5,0)^T$.

Ejemplo(continuación):
Aquí

$$A^T A = \begin{bmatrix} 2 & 5 \\ 5 & 13 \end{bmatrix} \quad , \quad \hat{x} = (A^T A)^{-1} A^T b = \begin{bmatrix} 2 \\ 1 \end{bmatrix},$$

y finalmente,

$$\hat{b} = A\hat{x} = \begin{bmatrix} 4 \\ 5 \\ 0 \end{bmatrix}.$$

Hemos visto entonces que, dada A una matriz $n \times k$ con columnas l.i. y $b \in \mathbb{R}^n$, la proyección de b sobre el espacio $C(A)$ es $\hat{b} = A(A^T A)^{-1} A^T b$. Nuevamente, la proyección sobre $C(A)$ puede ser vista como una tranformación lineal definida por la *matriz de proyección*

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

Observación: Si P es la matriz de proyección sobre $C(A)$ y $b \in \mathbb{R}^n$, $Pb \in C(A)$ y $e = b - Pb \in N(A^T)$. Por lo tanto, P nos da la *componente de* b en el espacio $C(A)$ y e es la componente sobre el complemento ortogonal de $C(A)$, esto es, $N(A^T)$.

Veamos algunas propiedades de las matrices de proyección que las caracterizan:

Teorema: Sea A una matriz $n \times k$ con columnas l.i. y $P = A(A^TA)^{-1}A^T$. Entonces $P^2 = PP = P$ y $P^T = P$. Recíprocamente, toda matriz simétrica tal que $P^2 = P$ representa una proyección.

Prueba: Ejercicio. (Ayuda: para probar la recíproca, probar que Pb es la proyección de b sobre $C(P)$).

Observaciones:

- Si A es inversible, $P = A(A^TA)^{-1}A^T = I$. ¿Tiene sentido esto?
- Si A es inversible, sus columnas son una base de $\mathbb{R}^n$. Si la base es una base ortonormal (sabemos que existe, por G-S), ¿quién es A^TA?.
  En ese caso, $A^TA = I$.
- Si las columas de A son ortonormales, A no necesariamente cuadrada, $A^TA = I$.

Definición: Una matriz ortogonal es una matriz cuadrada con columnas ortonormales.

Ejemplos:

- Las matrices de rotación.
- Las matrices de permutación.

PROYECCIONES Y MATRICES CON COLUMNAS ORTONORMALES

Propiedades: Si Q es una matriz ortogonal $n \times n$, entonces $Q^T = Q^{-1}$. Además, para todo $x \in \mathbb{R}^n$, $\|Qx\| = \|x\|$ (preserva longitud). (Ejercicio).

Volvamos a la proyección de un vector $b \in \mathbb{R}^n$ sobre el espacio columna de una matriz A de tamaño $n \times k$. Si las columas de A son ortonormales (siempre es posible gracias a G-S), encontrar *la mejor solución de* $Ax = b$ resulta mucho más fácil.

- Las ecuaciones $A^T A\hat{x} = A^T b$ se reducen a $\hat{x} = A^T b$.
- Así, $\hat{b} = A\hat{x} = AA^T b$.
- La matriz de proyección es $P = AA^T$.

Sea A una matriz $n \times k$ cuyas columnas son l.i. y Q^i $i = 1, \ldots, k$, los vectores ortonormales obtenidos al aplicar G-S a los vectores columna de A y luego normalizarlos (dividiendo por su norma). Sea Q es la matriz $n \times k$ de columnas Q^i, $i = 1, \ldots, k$.

Se puede probar que, en general, $A = QR$, donde R es una matriz triangular superior de tamaño $k \times k$.

Esto es lo que se denomina *factorización* QR de una matriz. Contar con la factorización QR de una matriz simplifica la resolución de sistemas inconsistentes por aproximación.

Teorema: Toda matriz A de tamaño $n \times k$ y con columnas l.i. puede ser factorizada $A = QR$ con Q una matriz $n \times k$ cuyas columnas son ortonormales y R, triangular superior e inversible de tamaño $k \times k$.

Prueba: Ejercicio.
Ayuda: Probar que $Q^T A$ es triangular superior e inversible.