Índice

Capítulo 7

Espacios con producto interno

7.1. Producto interno

Definición. Un **producto interno** sobre un espacio vectorial real (o complejo) $\mathcal{V}$ es una operación que asigna a cada par ordenado de vectores $x, y \in \mathcal{V}$ un escalar real (o complejo) $\langle x|y\rangle$ tal que se cumplen las siguientes propiedades:

$$(i) \quad \langle x|x\rangle \in \mathbb{R}, \langle x|x\rangle \geq 0 \quad \wedge \quad \langle x|x\rangle = 0 \iff x = 0_{\mathcal{V}}. \tag{7.1}$$

$$(ii) \quad \langle x|\alpha y\rangle = \alpha \langle x|y\rangle, \forall \alpha \in \mathbb{R} \,(\text{o } \mathbb{C}). \tag{7.2}$$

$$(iii) \quad \langle x|y+z\rangle = \langle x|y\rangle + \langle x|z\rangle \tag{7.3}$$

$$(iv) \quad \langle x|y\rangle = \overline{\langle y|x\rangle}. \tag{7.4}$$

∗ Si $\mathcal{V}$ es un espacio vectorial real, la propiedad (iv) se reduce a

$$\langle x|y\rangle = \langle y|x\rangle .$$

∗ Un espacio vectorial real o complejo, equipado con un producto interno, se denomina **Espacio con producto interno**.

Ejemplos

1) $\mathbb{R}^n$ equipado con el producto escalar estándar $\langle x|y\rangle = x^T y$.

2) $\mathbb{C}^n$ equipado con el producto escalar estándar $\langle x|y\rangle = x^* y$.

3) $\mathbb{R}^{m\times n}$ equipado con $\langle A|B\rangle = \text{Traza}(A^T B)$.

4) $\mathbb{C}^{m\times n}$ equipado con $\langle A|B\rangle = \text{Traza}(A^* B)$.

5) $\mathcal{V} = P_2(x)$, el espacio de los polinomios de grado menor o igual a dos con coeficientes reales, equipado con el producto interno

$$\langle P(x)|Q(x)\rangle = \sum_{i=0}^{2} p_i q_i, \tag{7.5}$$

donde $P(x) = p_0 + p_1 x + p_2 x^2$ y $Q(x) = q_0 + q_1 x + q_2 x^2$.

6) $\mathcal{V} \in C[a,b]$, el espacio de todas las funciones reales y continuas en $[a,b]$. Si $f(x)$, $g(x) \in C[a,b]$, y definimos

$$\langle f(x)|g(x)\rangle = \int_a^b f(x)g(x)\,dx,$$

luego $\mathcal{V}$ es un espacio con producto interno. Desarrollaremos este ejemplo en clase.

* * *

7.2. Norma

Definición. Una **norma** sobre un espacio vectorial real o complejo $\mathcal{V}$ es una función que asocia a cada vector $v \in \mathcal{V}$ un número real $||v||$, llamado norma de v, de modo tal que se satisfacen las siguientes propiedades:

$$(i) \quad ||v|| \geq 0 \quad \wedge \quad ||v|| = 0 \iff v = 0_{\mathcal{V}} \tag{7.6}$$

$$(ii) \quad ||\alpha v|| = |\alpha| \, ||v||, \quad \forall \alpha \in \mathbb{R} \text{ (o } \mathbb{C}) \tag{7.7}$$

$$(iii) \quad ||u + v|| \leq ||u|| + ||v|| \tag{7.8}$$

∗ Un espacio vectorial real o complejo, equipado con una norma, se denomina **espacio vectorial normado**, o **espacio normado.**

∗ De su definición se desprende que la norma nos permite asignar "tamaños" a vectores de un espacio vectorial. También nos permite cuantificar la diferencia (distancia) entre dos vectores. Veremos en clase un ejemplo de cuándo resulta particularmente útil esto último (en relación a los algoritmos numéricos).

Una norma conocida

Sea $x = (x_1, x_2) \in \mathbb{R}^2$. ¿Es $||x|| = \sqrt{x_1^2 + x_2^2}$ una norma? Vamos a comprobarlo.

(i) $\sqrt{x_1^2 + x_2^2} \geq 0 \wedge \sqrt{x_1^2 + x_2^2} = 0 \iff x = 0$ ✓

(ii) Ahora debemos considerar αx. Luego,

$$\alpha x = \alpha(x_1, x_2) = (\alpha x_1, \alpha x_2) \quad \Rightarrow \quad ||\alpha x|| = \sqrt{\alpha^2 x_1^2 + \alpha^2 x_2^2} = |\alpha|\sqrt{x_1^2 + x_2^2} = |\alpha| \, ||x|| \quad \checkmark \tag{7.9}$$

(iii) Sean $x = (x_1, x_2)$, $y = (y_1, y_2)$. Como $x + y = (x_1 + y_1, x_2 + y_2)$, debemos probar que

$$||x + y|| \leq ||x|| + ||y||, \tag{7.10}$$

es decir, que

$$\sqrt{(x_1 + y_1)^2 + (x_2 + y_2)^2} \leq \sqrt{x_1^2 + x_2^2} + \sqrt{y_1^2 + y_2^2}, \tag{7.11}$$

para lo cual nos ayudamos con el siguiente esquema
Visualmente, es claro que (iii) se cumple.

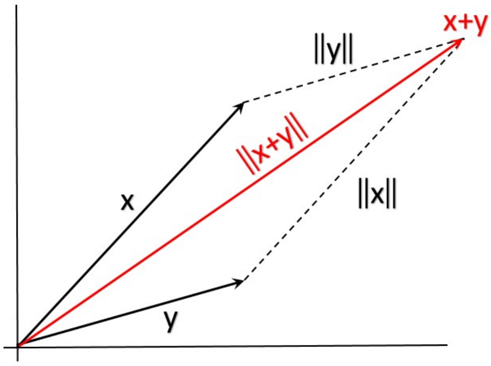

Figura 7.1: Desigualdad triangular en $\mathbb{R}^2$

En espacios más generales, donde no tenemos el beneficio de la visualización de los vectores, es necesario tener una herramienta que nos permita determinar si la desigualdad triangular se cumple. Veremos para ello, en breve, la desigualdad CBS[1].

Otras normas

1) **Norma Euclídea**

$$||x|| = \left(\sum_{i=1}^{n} x_i^2\right)^{\frac{1}{2}} = \sqrt{x^T x}, \qquad \text{si } x \in \mathbb{R}^n, \tag{7.12}$$

$$||x|| = \left(\sum_{i=1}^{n} |x_i|^2\right)^{\frac{1}{2}} = \sqrt{x^* x}, \qquad \text{si } x \in \mathbb{C}^n. \tag{7.13}$$

2) **Norma 1**

$$||x||_1 = \sum_{i=1}^{n} |x_i|, \qquad \text{si } x \in \mathbb{C}^n. \tag{7.14}$$

3) **Norma ∞**

$$||x||_\infty = \max\{|x_1|, |x_2|, ..., |x_n|\}, \qquad \text{si } x \in \mathbb{C}^n. \tag{7.15}$$

4) **Norma p**

$$||x||_p = \left(\sum_{i=1}^{p} |x_i|^p\right)^{\frac{1}{p}}, \qquad \text{si } x \in \mathbb{C}^n. \tag{7.16}$$

Observen que para $p = 2$, tenemos la norma euclídea. Las normas $p = 1, 2, \infty$ son las más frecuentemente utilizadas.

* * *

[1]CBS proviene de las iniciales de Cauchy, Bunyakovskii y Schwarz, autores de distintas versiones de este teorema.

Resulta ilustrativo comparar los tamaños de las distintas normas de un dado vector; definamos para ello la **esfera unidad**:

$$S_p = \{x/||x||_p = 1\}, \tag{7.17}$$

es decir, la esfera unidad de un espacio vectorial, respecto de una norma dada, es el conjunto de vectores cuya norma es igual a 1.

Considerando el caso particular $\mathcal{V} = \mathbb{R}^2$, y $p = 1, 2, \infty$, podemos ver en la figura 7.2 esquemas de las correspondientes esferas unidad. De aquí es fácil ver que $S_1 \subset S_2 \subset S_\infty$, y por lo tanto, $||x||_\infty \leq ||x||_2 \leq ||x||_1$.

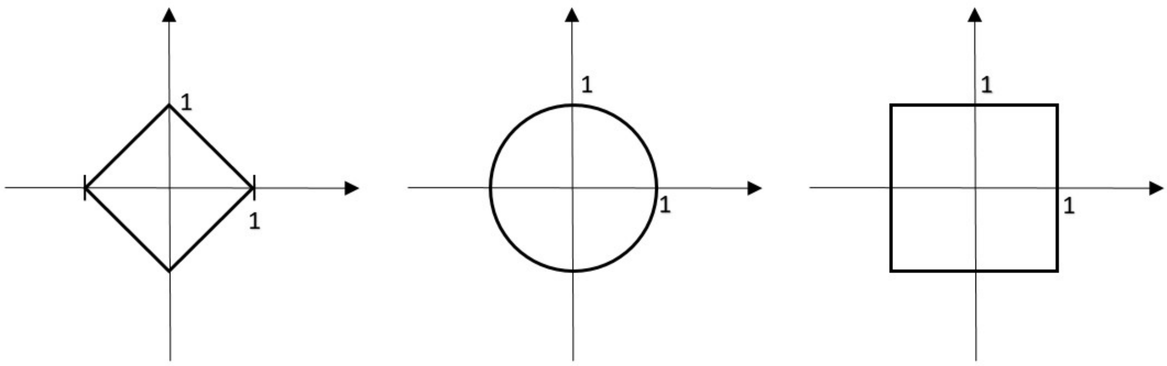

Figura 7.2: Izquierda: S_1, centro: S_2, derecha: S_∞

Consideremos el caso particular del punto x=(1,1). Calculando las distintas normas para este punto encontramos que

$$||x||_\infty = 1, \qquad ||x||_2 = \sqrt{2}, \qquad ||x||_1 = 2, \tag{7.18}$$

que satisface la desigualdad anterior.

$*$ Las normas que acabamos de ver están definidas en términos de las coordenadas de los vectores. Es posible definir otras más generales.

7.2.1. Normas de Matrices

Si bien es posible definir normas matriciales en analogía a lo hecho en espacios $\mathbb{C}^n$, o $\mathbb{R}^n$, es conveniente, desde el punto de vista de la utilidad, definir una norma más restrictiva:

Definición. La **norma de una matriz** A, denotada por $||A||$, es una función sobre el espacio de las matrices de cualquier orden (en $\mathbb{R}$ o $\mathbb{C}$) en los reales tal que satisface las siguientes propiedades:

$$(i) \quad ||A|| \geq 0 \quad \wedge \quad ||A|| = 0 \iff A = 0 \tag{7.19}$$

$$(ii) \quad ||\alpha A|| = |\alpha|\,||A||, \quad \forall \alpha \in \mathbb{R} \text{ (ó } \mathbb{C}) \tag{7.20}$$

$$(iii) \quad ||A + B|| \leq ||A|| + ||B|| \qquad \text{(si } A \text{ y } B \text{ son compatibles para la suma)} \tag{7.21}$$

$$(iv) \quad ||A\,B|| \leq ||A||\,||B|| \qquad \text{(si } A \text{ y } B \text{ son compatibles para el producto)} \tag{7.22}$$

Teorema (S/D). Una norma definida sobre vectores en $\mathbb{C}^p$, $p = m, n$ **induce** una norma en $\mathbb{C}^{m\times n}$ haciendo

$$||A|| = \max_{||x||=1} ||A\,x||, \quad A \in \mathbb{C}^{m\times n}, x \in \mathbb{C}^{n\times 1}.$$

$*$ En palabras, la norma inducida $||A||$ representa el máximo que un vector de la esfera unidad puede ser extendido por la matriz A.

$*$ Se dice que una norma matricial es compatible con una norma de vector si para todas las matrices $\mathbb{C}^{m\times n}$ vale que

$$||A\,x|| \leq ||A||\,||x||. \tag{7.23}$$

Otras normas matriciales

1) **Norma de Frobenius**

$$||A||_F = \left(\sum_{i,j} |a_{ij}|^2\right)^{\frac{1}{2}}$$

2) **Norma 2**

$$||A||_2 = \max_{||x||_2=1} ||A\,x||_2$$

3) **Norma 1**

$$||A||_1 = \max_{||x||_1=1} ||A\,x||_1$$

Vamos a demostrar que $||A||_1 = \max_j \sum_i |a_{ij}|$, es decir, la norma 1 matricial, es el máximo de la suma de los valores absolutos de los elementos de cada columna.

Sea x un vector con $||x||_1 = 1$. Luego,

$$||Ax||_1 = \sum_i \left|\sum_j a_{ij}\,x_j\right| \leq \sum_i \sum_j |a_{ij}|\,|x_j| = \sum_j \left(|x_j| \sum_i |a_{ij}|\right) \leq$$

$$\leq \left(\sum_j |x_j|\right)\left(\max_j \sum_i |a_{ij}|\right) = \max_j \sum_i |a_{ij}|$$

Hemos podido acotar $||Ax||_1$, pero aún no sabemos si el valor máximo es posible de alcanzar. Consideremos que A_{*k} es la columna con el mayor valor de la suma de los valores absolutos de sus elementos. Ahora tomemos el vector $x = e_k$, que sabemos que tiene $||e_k||_1 = 1$. Reemplazando en la expresión de arriba

$$||Ax||_1 = ||Ae_k||_1 = \sum_i |a_{ik}| = \max_j \sum_i |a_{ij}|, \tag{7.24}$$

y queda demostrado que

$$\boxed{||A||_1 = \max_j \sum_i |a_{ij}|.}$$

4) **Norma ∞**

$$||A||_\infty = \max_{||x||_\infty=1} ||A\,x||_\infty$$

Análogamente al caso de la norma 1, puede mostrarse (hacerlo como tarea) que

$$\boxed{||A||_\infty = \max_i \sum_j |a_{ij}|.} \tag{7.25}$$

$* * *$

Volvamos a trabajar nuevamente con espacios abstractos; dejando de lado por un rato ejemplos particulares.

Teorema (Desigualdad CBS). Si $\mathcal{V}$ es un espacio vectorial con producto interno, entonces se cumple que

$$|\langle x|y\rangle| \leq \sqrt{\langle x|x\rangle}\sqrt{\langle y|y\rangle}, \qquad \forall x, y \in \mathcal{V}. \tag{7.26}$$

La igualdad vale sí y sólo sí $y = \alpha x$, para $\alpha = \langle x|y\rangle / \langle x|x\rangle$.

Demostración

Sean $x \neq 0$ y $\alpha = \langle x|y\rangle / \langle x|x\rangle$, entonces

$$0 \leq \langle \alpha\, x - y|\alpha\, x - y\rangle = \overline{\alpha}\,\langle x|\alpha\, x - y\rangle - \langle y|\alpha\, x - y\rangle\,. \tag{7.27}$$

Pero

$$\langle x|\alpha\, x - y\rangle = \alpha\,\langle x|x\rangle - \langle x|y\rangle = \frac{\langle x|y\rangle}{\cancel{\langle x|x\rangle}}\cancel{\langle x|x\rangle} - \langle x|y\rangle = 0, \tag{7.28}$$

por nuestra elección de α. Luego

$$0 \leq \langle \alpha\, x - y|\alpha\, x - y\rangle = -\,\langle y|\alpha\, x - y\rangle = \langle y|y\rangle - \alpha\,\langle y|x\rangle = \langle y|y\rangle - \frac{\langle x|y\rangle}{\langle x|x\rangle}\,\langle y|x\rangle\,. \tag{7.29}$$

Como $\langle y|x\rangle = \overline{\langle x|y\rangle}$, tenemos que $\langle x|y\rangle\ \langle y|x\rangle = |\langle x|y\rangle|^2$. Luego,

$$\langle y|y\rangle - \frac{|\langle x|y\rangle|^2}{\langle x|x\rangle} \geq 0, \tag{7.30}$$

o

$$\frac{\langle x|x\rangle\,\langle y|y\rangle - |\langle x|y\rangle|^2}{\langle x|x\rangle} \geq 0, \tag{7.31}$$

y, por lo tanto,

$$\boxed{|\langle x|y\rangle| \leq \sqrt{\langle x|x\rangle}\sqrt{\langle y|y\rangle}.} \tag{7.32}$$

Consideremos finalmente el caso particular de la igualdad

$$|\langle x|y\rangle| = \sqrt{\langle x|x\rangle}\sqrt{\langle y|y\rangle}. \tag{7.33}$$

De la definición de α, tendremos ahora que

$$|\alpha| = \frac{|\langle x|y\rangle|}{\langle x|x\rangle} = \frac{\sqrt{\langle x|x\rangle}\sqrt{\langle y|y\rangle}}{\langle x|x\rangle} = \frac{\sqrt{\langle y|y\rangle}}{\sqrt{\langle x|x\rangle}}, \tag{7.34}$$

donde hemos usado (7.33). Teniendo en cuenta que $|\alpha| = \sqrt{\alpha\overline{\alpha}}$ podemos escribir la expresión anterior como

$$\sqrt{\langle y|y\rangle} = \sqrt{\alpha\overline{\alpha}}\sqrt{\langle x|x\rangle} = \sqrt{\langle \alpha x|\alpha x\rangle}, \tag{7.35}$$

de donde se desprende que, en este caso particular, se cumple que

$$\boxed{y = \alpha x.} \tag{7.36}$$

□

Teorema: Si $\mathcal{V}$ es un espacio vectorial con producto interno $\langle\cdot|\cdot\rangle$ y $x \in \mathcal{V}$, entonces

$$||x|| = \sqrt{\langle x|x\rangle} \tag{7.37}$$

define una norma sobre $\mathcal{V}$.

Demostración

Debemos probar que se verifican las tres propiedades requeridas.

- (i) Que cumple $||x|| \geq 0$ y $||x|| = 0 \Leftrightarrow x = 0$ se desprende de la definición de producto interno. ✓
- (ii) $||\alpha\, x|| = \sqrt{\langle \alpha\, x | \alpha\, x \rangle} = \sqrt{\overline{\alpha}\alpha\, \langle x|x\rangle} = \sqrt{|\alpha|^2\, \langle x|x\rangle} = |\alpha|\, \sqrt{\langle x|x\rangle} = |\alpha|\, ||x||$. ✓
- (iii) Resta probar la desigualdad triangular.

$$\begin{aligned} ||x+y||^2 &= \langle x+y|x+y\rangle \\ &= \langle x|x\rangle + \langle x|y\rangle + \langle y|x\rangle + \langle y|y\rangle \\ &= ||x||^2 + ||y||^2 + 2\,\mathrm{Re}(\langle x|y\rangle) \\ &\leq ||x||^2 + ||y||^2 + 2\,|\,\langle x|y\rangle\,| \\ &\leq ||x||^2 + ||y||^2 + 2\,||x||\,||y|| \qquad \text{(por desigualdad CBS)} \\ &= (||x|| + ||y||)^2 \end{aligned} \tag{7.38}$$

En consecuencia,

$$||x+y|| \leq ||x|| + ||y|| \quad ✓ \tag{7.39}$$

■

De acuerdo a este teorema, siempre es posible definir una norma en un espacio vectorial con producto interno. Decimos, en este caso, que **el producto interno induce una norma**.

Cabe preguntarse si vale la recíproca, es decir, si una norma puede inducir un producto interno. El siguiente teorema nos dice que, en general, esto NO es posible.

Teorema (S/D) - Identidad del Paralelogramo

Para una dada norma $||\cdot||$ sobre un espacio vectorial $\mathcal{V}$, existe un producto interno sobre $\mathcal{V}$ tal que $\langle\cdot|\cdot\rangle = ||\cdot||^2$ sí y sólo sí vale la identidad del paralelogramo, es decir, si

$$||x+y||^2 + ||x-y||^2 = 2\,(||x||^2 + ||y||^2), \qquad \forall x, y \in \mathcal{V}. \tag{7.40}$$

Ejemplo

Sea $||\cdot||_p$ y sea $\mathcal{V} = \mathbb{R}^n$. Tomemos $x = e_1$ e $y = e_2$.

$$\begin{aligned} ||x+y||_p^2 + ||x-y||_p^2 &= ||e_1+e_2||_p^2 + ||e_1-e_2||_p^2 \\ &= 2^{2/p} + 2^{2/p} \\ &= 2\cdot 2^{2/p} = 2^{(p+2)/p} \end{aligned} \tag{7.41}$$

Por otro lado,

$$2\,(||x||_p^2 + ||y||_p^2) = 2\,(1^{1/p} + 1^{1/p}) = 4 \tag{7.42}$$

Se ve que para $p \neq 2$, $2^{(p+2)/p} \neq 4$, y no vale la identidad del paralelogramo. Por lo tanto, las normas con $p \neq 2$ no inducen un producto interno sobre el espacio vectorial $\mathcal{V}$.

* * *

7.3. Ortogonalidad

Definición. Sean $x, y \in \mathcal{V}$, espacio vectorial con producto interno, y sea $||\cdot|| = \sqrt{\langle\cdot|\cdot\rangle}$ la norma inducida por el producto interno, entonces

- (i) decimos que x e y **son vectores ortogonales** si $\langle x|y\rangle = 0$,
- (ii) llamamos **distancia entre** los **vectores** x e y al número real $d(x, y)$ tal que

$$d(x, y) = ||x - y||. \tag{7.43}$$

∗ Observen que ortogonalidad es una propiedad que sólo puede ser definida en espacios con producto interno.

Ejemplo

Sea $\mathcal{V} \in C[0, 1]$ y sean $f(x) = x$ y $g(x) = 3x - 2$, dos miembros de $\mathcal{V}$. Calcular la distancia y el producto interno.

$$\begin{aligned} d(f(x), g(x)) = ||f(x) - g(x)|| &= (\langle f(x) - g(x)|f(x) - g(x)\rangle)^{\frac{1}{2}} \\ &= \left(\int_0^1 (f(x) - g(x))^2 dx\right)^{\frac{1}{2}} \\ &= \left(\int_0^1 (-2x + 2)^2 dx\right)^{\frac{1}{2}} \end{aligned}$$

$$\langle f(x)|g(x)\rangle = \int_0^1 x(3x - 2)dx = 0 \quad \Rightarrow \quad f(x) \perp g(x).$$

∗ ∗ ∗

Teorema. Sea $\mathcal{V}$ un **espacio vectorial real sobre el cuerpo de los reales** con producto interno $\langle\cdot|\cdot\rangle$ y norma inducida $||\cdot|| = \sqrt{\langle\cdot|\cdot\rangle}$, entonces se cumple que

$$||x + y||^2 = ||x||^2 + ||y||^2 \iff x \perp y, \qquad x, y \in \mathcal{V}. \tag{7.44}$$

Demostración

Utilizando que la norma es inducida y operando trivialmente se encuentra que

$$||x + y||^2 = \langle x + y|x + y\rangle = \langle x|x\rangle + \langle y|y\rangle + \langle x|y\rangle + \langle y|x\rangle = ||x||^2 + ||y||^2 + 2\langle x|y\rangle. \tag{7.45}$$

Por lo tanto,

$$||x + y||^2 = ||x||^2 + ||y||^2 \quad \iff \quad x \perp y. \tag{7.46}$$

■

∗ En $\mathbb{R}^2$ y $\mathbb{R}^3$, éste es el Teorema de Pitágoras.

∗ El teorema no vale en espacios vectoriales complejos, donde las cuentas anteriores cambian levemente:

$$\begin{aligned} ||x + y||^2 = \langle x + y|x + y\rangle &= \langle x|x\rangle + \langle y|y\rangle + \langle x|y\rangle + \langle y|x\rangle \\ &= ||x||^2 + ||y||^2 + \langle x|y\rangle + \overline{\langle x|y\rangle} \\ &= ||x||^2 + ||y||^2 + 2\,\mathrm{Re}\,(\langle x|y\rangle). \end{aligned}$$

Ahora basta con que $\mathrm{Re}(\langle x|y\rangle) = 0$ para que valga $||x + y||^2 = ||x||^2 + ||y||^2$, y no es necesario que se dé que $x \perp y$.

* * *

Definición. Un conjunto de vectores no nulos $S = \{s_1, s_2, \ldots, s_k\} \in \mathcal{V}$, espacio vectorial con producto interno, es un **conjunto ortogonal** si

$$\langle s_i | s_j \rangle = 0, \quad \text{si} \quad i \neq j, \quad \forall i, j = 1, \ldots, k.$$

$*$ Si, además, se cumple que $||s_i||$=1, $\forall\, i = 1, \ldots, k$, el conjunto S se denomina **ortonormal**. Por supuesto, $||(s_i/||s_i||)|| = 1$.

$*$ Los conjuntos ortogonales tienen características que los hacen interesantes, por ejemplo,

Teorema. Si $S = \{s_1, s_2, \ldots, s_k\}$ es un conjunto ortogonal $\Rightarrow$ es linealmente independiente.

Demostración

Consideremos la combinación lineal

$$\alpha_1\, s_1 + \alpha_2\, s_2 + \cdots + \alpha_k\, s_k = 0_{\mathcal{V}}, \tag{7.47}$$

y tomemos el producto interno con el vector s_i

$$\langle s_i | \alpha_1\, s_1 + \alpha_2\, s_2 + \cdots + \alpha_k\, s_k \rangle = \langle s_i | 0_{\mathcal{V}} \rangle = 0. \tag{7.48}$$

Utilizando las propiedades que debe cumplir un producto interno, tenemos que

$$\alpha_1\, \langle s_i | s_1 \rangle + \ldots + \alpha_i\, \langle s_i | s_i \rangle + \ldots + \alpha_k\, \langle s_i | s_k \rangle = 0, \tag{7.49}$$

y por ser S un conjunto ortogonal, la expresión anterior se reduce a

$$\alpha_i\, \langle s_i | s_i \rangle = 0, \tag{7.50}$$

donde sabemos que $\langle s_i | s_i \rangle \neq 0$. Por lo tanto, concluimos que

$$\alpha_i = 0. \tag{7.51}$$

Este proceso es válido para $i = 1, \ldots, k$, por lo que se demuestra que el conjunto S es linealmente independiente.

■

$*$ *Corolario:* Un conjunto ortogonal (ortonormal) de n vectores de un espacio vectorial n-dimensional es una base.

$*$ Cuando se desea encontrar las coordenadas de un vector respecto de una base, en general debe resolverse un sistema lineal de ecuaciones, con tantas incógnitas como dimensiones tenga el espacio vectorial.

En el caso en que la base sea ortonormal, las coordenadas pueden ser fácilmente obtenidas, como haremos a continuación. Consideremos un conjunto $S = \{s_1, s_2, \ldots, s_n\}$, base ortonormal del espacio $\mathcal{V}$, y sea $u \in \mathcal{V}$. Entonces podemos escribir

$$u = \sum_{j=1}^{n} \alpha_j s_j. \tag{7.52}$$

Las coordenadas de u pueden hallarse haciendo el producto interno

$$\langle s_i\, | u \rangle = \left\langle s_i \middle| \sum_{j=1}^{n} \alpha_j s_j \right\rangle = \alpha_i, \tag{7.53}$$

para $i = 1, ..., n$. Así, obtenemos

$$\boxed{u = \langle s_1|u\rangle\, s_1 + \langle s_2|u\rangle\, s_2 + \cdots + \langle s_n|u\rangle\, s_n.} \tag{7.54}$$

A esta expresión se la denomina desarrollo de Fourier de u y los $\langle s_i|u\rangle$, $i = 1, \ldots, n$, son los coeficientes de Fourier del desarrollo. Por ser S base, los α_i son únicos.

$*$ Si la base es ortogonal, pero no ortonormal, entonces

$$\alpha_i = \frac{\langle s_i|u\rangle}{||s_i||^2}, \qquad i = 1, \ldots, n. \tag{7.55}$$

¿Pueden demostrarlo?

$* * *$

$*$ Dada una base cualquiera de un espacio vectorial n-dimensional, es válido preguntarse si es posible encontrar a partir de ella otra base pero que sea ortonormal. La respuesta es sí y el método para obtenerla se denomina **procedimiento de Gram-Schmidt**, que estudiaremos un poco más adelante.

$* * *$

7.4. Subespacios Complementarios - Complemento ortogonal

Definición. Dados dos subespacios $\mathcal{X}$ e $\mathcal{Y}$ de un espacio vectorial $\mathcal{V}$, se llama **subespacio suma** $\mathcal{X} + \mathcal{Y}$ al subespacio

$$\mathcal{X} + \mathcal{Y} = \{x + y / x \in \mathcal{X}, y \in \mathcal{Y}\}.$$

Definición. Dos **subespacios** $\mathcal{X}$ e $\mathcal{Y}$ de un espacio vectorial $\mathcal{V}$, se dicen **complementarios** cuando se cumple que

$$\mathcal{V} = \mathcal{X} + \mathcal{Y} \quad \wedge \quad \mathcal{X} \cap \mathcal{Y} = 0_{\mathcal{V}}.$$

En este caso $\mathcal{V}$ se llama **suma directa** de $\mathcal{X}$ e $\mathcal{Y}$, y se denota por $\mathcal{V} = \mathcal{X} \oplus \mathcal{Y}$.

Teorema (S/D). Para un espacio vectorial $\mathcal{V}$ con subespacios $\mathcal{X}$ e $\mathcal{Y}$ y bases $B_{\mathcal{X}}$ y $B_{\mathcal{Y}}$, respectivamente, son equivalentes las siguientes afirmaciones:

- (i) $\mathcal{V} = \mathcal{X} \oplus \mathcal{Y}$
- (ii) para cada $v \in \mathcal{V}$, existen $x \in \mathcal{X}$ e $y \in \mathcal{Y}$ **únicos** tales que

$$v = x + y \tag{7.56}$$

- (iii) $B_{\mathcal{X}} \cap B_{\mathcal{Y}} = \emptyset$ (vacío) y $B_{\mathcal{X}} \cup B_{\mathcal{Y}}$ es una base de $\mathcal{V}$.

7.5. Proyecciones

Definición. Sea $\mathcal{V} = \mathcal{X} \oplus \mathcal{Y}$, de modo que para cada $v \in \mathcal{V}$ existen vectores únicos $x \in \mathcal{X}$ e $y \in \mathcal{Y}$ tales que $v = x + y$.

- (i) El vector x se llama proyección de v sobre $\mathcal{X}$ a lo largo de $\mathcal{Y}$.
- (ii) El vector y se llama proyección de v sobre $\mathcal{Y}$ a lo largo de $\mathcal{X}$.

Teorema (de Proyectores). Sean $\mathcal{X}$ e $\mathcal{Y}$ subespacios complementarios de un espacio vectorial $\mathcal{V}$ de modo tal que cada $v \in \mathcal{V}$ puede escribirse de forma única como $v = x + y$, con $x \in \mathcal{X}$ e $y \in \mathcal{Y}$. El único operador lineal $\mathcal{P}$ definido por $\mathcal{P}(v) = x$ se llama **proyector sobre $\mathcal{X}$ a lo largo de $\mathcal{Y}$**, y tiene las siguientes propiedades:

(I) $\mathcal{P}^2 = \mathcal{P}$

(II) id$-\mathcal{P}$ es el proyector complementario sobre $\mathcal{Y}$ a lo largo de $\mathcal{X}$

(III) $\mathcal{R}(\mathcal{P}) = \mathcal{N}(\mathrm{id} - \mathcal{P}) = \mathcal{X} \quad \wedge \quad \mathcal{R}(\mathrm{id} - \mathcal{P}) = \mathcal{N}(\mathcal{P}) = \mathcal{Y}$

(IV) Si $\mathcal{V} = \mathbb{R}^n$ (o $\mathbb{C}^n$) $\Rightarrow$ la matriz que representa a $\mathcal{P}$ en la base estándar (S) está dada por el siguiente producto matricial

$$[\mathcal{P}]_{\mathrm{SS}} = \begin{pmatrix} \mathrm{X} & 0_{n\times(n-r)} \end{pmatrix} \begin{pmatrix} \mathrm{X} & \mathrm{Y} \end{pmatrix}^{-1} = \begin{pmatrix} \mathrm{X} & \mathrm{Y} \end{pmatrix} \begin{pmatrix} \mathrm{I}_{r\times r} & 0_{r\times(n-r)} \\ 0_{(n-r)\times r} & 0_{(n-r)\times(n-r)} \end{pmatrix} \begin{pmatrix} \mathrm{X} & \mathrm{Y} \end{pmatrix}^{-1} \tag{7.57}$$

donde las columnas de $\mathrm{X}^{n\times r}$ e $\mathrm{Y}^{n\times(n-r)}$ son los vectores que forman las respectivas bases $B_{\mathcal{X}}$ y $B_{\mathcal{Y}}$ de $\mathcal{X}$ e $\mathcal{Y}$.

Demostración

Antes de comenzar conviene verificar la unicidad del operador lineal definido por $\mathcal{P}(v) = x$ que menciona el enunciado. Consideremos para ello que existen dos operadores lineales $\mathcal{P}_1$ y $\mathcal{P}_2$ tales que

$$\mathcal{P}_1(v) = x, \qquad\qquad \mathcal{P}_2(v) = x. \tag{7.58}$$

Como esto debe ocurrir para todo $v \in \mathcal{V}$, esto implica que $\mathcal{P}_1(v) = \mathcal{P}_2(v)$, es decir, que $\mathcal{P}_1 = \mathcal{P}_2$.

¿Es $\mathcal{P}$ un homomorfismo? Sabemos que es una función, faltaría verificar si preserva estructura. Sean $v_1 = x_1 + y_1$ y $v_2 = x_2 + y_2$ dos vectores de $\mathcal{V}$, con $x_1, x_2 \in \mathcal{X}$ e $y_1, y_2 \in \mathcal{Y}$. Entonces

$$\mathcal{P}(v_1 + v_2) = x_1 + x_2 = \mathcal{P}(v_1) + \mathcal{P}(v_2), \tag{7.59}$$

$$\mathcal{P}(\alpha v_1) = \alpha\, x_1 = \alpha\, \mathcal{P}(v_1), \tag{7.60}$$

y vemos que se preserva la estructura.

Ahora sí vamos a demostrar los distintos ítems del teorema.

(I)

$$\mathcal{P}^2(v) = \mathcal{P}(\mathcal{P}(v)) = \mathcal{P}(x) = x = \mathcal{P}(v), \quad \forall\, v \quad \Rightarrow \quad \boxed{\mathcal{P}^2 = \mathcal{P}} \tag{7.61}$$

(II)

$$v = x + y = \mathcal{P}(v) + y \quad \Rightarrow \quad y = v - \mathcal{P}(v) = \mathrm{id}(v) - \mathcal{P}(v) = (\mathrm{id} - \mathcal{P})(v). \tag{7.62}$$

Luego, id $-\,\mathcal{P}$ es el proyector complementario sobre $\mathcal{Y}$ a lo largo de $\mathcal{X}$.

(III) Salen de las definiciones de imagen $\mathcal{R}$ y núcleo $\mathcal{N}$.

(IV) *(iv)* Sean $B_{\mathcal{X}} = \{x_1, \ldots, x_r\}$ y y $B_{\mathcal{Y}} = \{y_1, \ldots, y_{n-r}\}$ bases de $\mathcal{X}$ e $\mathcal{Y}$ respectivamente. Entonces sabemos que $B = B_{\mathcal{X}} \cup B_{\mathcal{Y}}$ es base de $\mathcal{V}$. La representación matricial de $\mathcal{P}$ con respecto a la base B será

$$[\mathcal{P}]_{BB} = \begin{pmatrix} | & & | & | & & | \\ [\mathcal{P}(x_1)]_B & \ldots & [\mathcal{P}(x_r)]_B & [\mathcal{P}(y_1)]_B & \ldots & [\mathcal{P}(y_{n-r})]_B \\ | & & | & | & & | \end{pmatrix}$$

$$= \begin{pmatrix} | & & | & | & & | \\ [x_1]_B & \dots & [x_r]_B & [0]_B & \dots & [0]_B \\ | & & | & | & & | \end{pmatrix}$$

$$= \begin{pmatrix} e_1 & \dots & e_r & 0_{n\times 1} & \dots & 0_{n\times 1} \end{pmatrix}$$

$$= \begin{pmatrix} I_{r\times r} & 0_{r\times(n-r)} \\ 0_{(n-r)\times r} & 0_{(n-r)\times(n-r)} \end{pmatrix}$$

Sea ahora S la base canónica de $\mathbb{R}^n$ (o $\mathbb{C}^n$). Para cambiar de base hacemos

$$[\mathcal{P}]_{\mathrm{SS}} = C_{BS}\,[\mathcal{P}]_{BB}\,C_{BS}^{-1}, \tag{7.63}$$

y como

$$C_{BS} = \begin{pmatrix} | & & | & | & & | \\ x_1 & \dots & x_r & y_1 & \dots & y_{n-r} \\ | & & | & | & & | \end{pmatrix} = \begin{pmatrix} \mathrm{X} & \mathrm{Y} \end{pmatrix}, \tag{7.64}$$

obtenemos finalmente

$$[\mathcal{P}]_{\mathrm{SS}} = \begin{pmatrix} \mathrm{X} & \mathrm{Y} \end{pmatrix} \underbrace{\begin{pmatrix} \mathrm{I}_{r\times r} & 0_{r\times(n-r)} \\ 0_{(n-r)\times r} & 0_{(n-r)\times(n-r)} \end{pmatrix}}_{Q} \begin{pmatrix} \mathrm{X} & \mathrm{Y} \end{pmatrix}^{-1} = \boxed{\begin{pmatrix} \mathrm{X} & 0 \end{pmatrix}\begin{pmatrix} \mathrm{X} & \mathrm{Y} \end{pmatrix}^{-1}.} \tag{7.65}$$

Notemos que $[\mathcal{P}]_{\mathrm{SS}}$ es semejante a Q.

■

Ejercicio

Sean $\mathcal{A}$ y $\mathcal{B}$ dos subespacios de $\mathbb{R}^3$, expandidos por

$$B_{\mathcal{A}} = \left\{ \begin{pmatrix} 1 \\ -1 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ -2 \end{pmatrix} \right\}, \qquad B_{\mathcal{B}} = \left\{ \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \right\}.$$

(a) Determinar si $\mathcal{A}$ y $\mathcal{B}$ son complementarios. (b) Hallar el proyector sobre $\mathcal{A}$ a lo largo de $\mathcal{B}$. (c) Dado $v = (-2\ 1\ 3)^T$, hallar su proyección sobre $\mathcal{A}$ a lo largo de $\mathcal{B}$, y la proyección sobre $\mathcal{B}$ a lo largo de $\mathcal{A}$.

Resolución

Para determinar si los subespacios son complementarios vamos a comenzar probando que $B_{\mathcal{A}} \cup B_{\mathcal{B}}$ forman un conjunto linealmente independiente.

$$c_1 \begin{pmatrix} 1 \\ -1 \\ -1 \end{pmatrix} + c_2 \begin{pmatrix} 0 \\ 1 \\ -2 \end{pmatrix} + c_3 \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \tag{7.66}$$

que podemos escribir en forma matricial como

$$\begin{pmatrix} 1 & 0 & 1 \\ -1 & 1 & -1 \\ -1 & -2 & 0 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}. \tag{7.67}$$

Dado que el determinante de la matriz es distinto de cero, la solución es única e igual a la trivial. Esto significa que el conjunto $B_{\mathcal{A}} \cup B_{\mathcal{B}}$ es linealmente independiente. Esto nos dice dos cosas:

1. $B_{\mathcal{A}}$ es base de $\mathcal{A}$,
2. $\mathcal{A}$ y $\mathcal{B}$ son complementarios, ya que al ser $B_{\mathcal{A}} \cup B_{\mathcal{B}}$ linealmente independiente, el único vector en común que tienen (los subespacios) es el nulo.

En consecuencia,

$$\mathbb{R}^3 = \mathcal{A} \oplus \mathcal{B}. \tag{7.68}$$

Para encontrar el proyector sobre $\mathcal{A}$ a lo largo de $\mathcal{B}$ hacemos:

$$\mathcal{P} = \begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ -1 & -2 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 1 \\ -1 & 1 & -1 \\ -1 & -2 & 0 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ -1 & -2 & 0 \end{pmatrix} \begin{pmatrix} -2 & -2 & -1 \\ 1 & 1 & 0 \\ 3 & 2 & 1 \end{pmatrix} = \begin{pmatrix} -2 & -2 & -1 \\ 3 & 3 & 1 \\ 0 & 0 & 1 \end{pmatrix} \tag{7.69}$$

Calculamos la proyección del vector v sobre el subespacio $\mathcal{A}$ a lo largo de $\mathcal{B}$:

$$\mathcal{P}v = \begin{pmatrix} -2 & -2 & -1 \\ 3 & 3 & 1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} -2 \\ 1 \\ 3 \end{pmatrix} = \begin{pmatrix} -1 \\ 0 \\ 3 \end{pmatrix} \tag{7.70}$$

Para verificar que efectivamente $\mathcal{P}v \in \mathcal{A}$ basta con ver que $B_{\mathcal{A}} \cup \{\begin{pmatrix} -1 & 0 & 3 \end{pmatrix}^T\}$ es linealmente dependiente.

Verifiquemos que $\mathcal{P}$ es idempotente.

$$\mathcal{P}^2 = \begin{pmatrix} -2 & -2 & -1 \\ 3 & 3 & 1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} -2 & -2 & -1 \\ 3 & 3 & 1 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} -2 & -2 & -1 \\ 3 & 3 & 1 \\ 0 & 0 & 1 \end{pmatrix} \tag{7.71}$$

Resta hallar la proyección de v sobre $\mathcal{B}$ a lo largo de $\mathcal{A}$. Esto se puede hacer sencillamente calculando

$$\mathcal{P}' = I - \mathcal{P}. \tag{7.72}$$

Lo dejamos como ejercicio.

* * *

Vimos que un proyector es idempotente. Cabe preguntase si un operador lineal idempotente siempre es un proyector. La respuesta es sí, y se demuestra en el siguiente teorema.

Teorema. Un operador lineal $\mathcal{P}$ sobre un espacio vectorial $\mathcal{V}$ es un proyector sí y sólo sí $\mathcal{P}^2 = \mathcal{P}$.

Demostración

$\Rightarrow$) (ya fue demostrado en el teorema anterior)

$\Leftarrow$) Necesitamos probar que la idempotencia de $\mathcal{P}$ implica que $\mathcal{R}(\mathcal{P})$ y $\mathcal{N}(\mathcal{P})$ son subespacios complementarios.

Dado que para todo $v \in \mathcal{V}$,

$$v = v - \mathcal{P}(v) + \mathcal{P}(v) = (\mathrm{id} - \mathcal{P})(v) + \mathcal{P}(v), \tag{7.73}$$

donde $\mathcal{P}(v) \in \mathcal{R}(\mathcal{P})$ e $(\mathrm{id} - \mathcal{P})(v) \in \mathcal{N}(\mathcal{P})$[2], podemos decir que

$$\mathcal{V} = \mathcal{R}(\mathcal{P}) + \mathcal{N}(\mathcal{P}). \tag{7.74}$$

Además, debemos probar que $\mathcal{R}(\mathcal{P}) \cap \mathcal{N}(\mathcal{P}) = \{0_{\mathcal{V}}\}$. Supongamos que $x \in \mathcal{R}(\mathcal{P}) \cap \mathcal{N}(\mathcal{P})$. Entonces x deberá verificar simultáneamente

$$x = \mathcal{P}(v) \qquad \text{y} \qquad \mathcal{P}(x) = 0_{\mathcal{V}}. \tag{7.75}$$

Usando estas expresiones y la idempotencia, podemos escribir

$$x = \mathcal{P}(v) = \mathcal{P}\left(\mathcal{P}(v)\right) = \mathcal{P}(x) = 0_{\mathcal{V}}, \tag{7.76}$$

y vemos que el único valor posible de x es el $0_{\mathcal{V}}$.

Tenemos entonces que $\mathcal{R}(\mathcal{P})$ y $\mathcal{N}(\mathcal{P})$ son subespacios complementarios. Por lo tanto, todo $v \in \mathcal{V}$ puede escribirse unívocamente como $v = x + y$, con $x \in \mathcal{R}(\mathcal{P})$ e $y \in \mathcal{N}(\mathcal{P})$ y $\mathcal{P}(v) = x \therefore \mathcal{P}$ es proyector.

[2] $\mathcal{P}((\mathrm{id} - \mathcal{P})(v)) = \mathcal{P}(\mathrm{id}(v)) - \mathcal{P}(\mathcal{P}(v)) = \mathcal{P}(v) - \mathcal{P}^2(v) = \mathcal{P}(v) - \mathcal{P}(v) = 0_{\mathcal{V}}$

■

* * *

Definición. Se denomina **complemento ortogonal** $\mathcal{M}^\perp$ de un conjunto $\mathcal{M} \subset \mathcal{V}$, espacio vectorial con producto interno, al conjunto de todos los vectores de $\mathcal{V}$ ortogonales a todos los de $\mathcal{M}$. Esto es,

$$\mathcal{M}^\perp = \{x \in \mathcal{V} / \langle m|x\rangle = 0, \ \forall m \in \mathcal{M}\}. \tag{7.77}$$

Ejemplo 1

$\mathcal{M} = \{x\}$, $x \in \mathbb{R}^2$ (un único vector de $\mathbb{R}^2$). Luego, $\mathcal{M}^\perp$ es la recta que pasa por el origen y es $\perp$ a x.

Ejemplo 2

Si $\mathcal{M}$ es un plano que pasa por el origen en $\mathbb{R}^3$, entonces $\mathcal{M}^\perp$ es la recta que pasa por el origen y que es perpendicular al plano.

$*$ En la definición de complemento ortogonal, el conjunto $\mathcal{M}$ no necesariamente es un subespacio de $\mathcal{V}$, pero se puede probar que $\mathcal{M}^\perp$ sí lo es. Para hacerlo, consideremos a los vectores $z \in \mathcal{M}$ y $x, y \in \mathcal{M}^\perp$.

Veamos si $x + y \in \mathcal{M}^\perp$.

$$\langle z|x+y\rangle = \langle z|x\rangle + \langle z|y\rangle = 0 \quad \Rightarrow \quad x + y \in \mathcal{M}^\perp. \tag{7.78}$$

Por otro lado

$$\langle z|\alpha x\rangle = \alpha \langle z|x\rangle = 0 \quad \Rightarrow \quad \alpha x \in \mathcal{M}^\perp. \tag{7.79}$$

En consecuencia, $\mathcal{M}^\perp$ es subespacio de $\mathcal{V}$.

7.6. Subespacios complementarios ortogonales

Teorema. Si $\mathcal{M}$ es un subespacio de un espacio vectorial $\mathcal{V}$ de dimensión n con producto interno, entonces

$$\mathcal{V} = \mathcal{M} \oplus \mathcal{M}^\perp. \tag{7.80}$$

Más aún, si $\mathcal{A}$ es un subespacio tal que $\mathcal{V} = \mathcal{M} \oplus \mathcal{A} \ \wedge \ \mathcal{A} \perp \mathcal{M} \Rightarrow \mathcal{A} = \mathcal{M}^\perp$.

Demostración

Que $\mathcal{M} \cap \mathcal{M}^\perp = 0_\mathcal{V}$ es fácilmente demostrable, ya que si $x \in \mathcal{M}$ y $x \in \mathcal{M}^\perp \Rightarrow \langle x|x\rangle = 0 \therefore x = 0_\mathcal{V}$.

Ahora debemos probar que $\mathcal{V} = \mathcal{M} + \mathcal{M}^\perp$. Sean $B_\mathcal{M}$ y $B_{\mathcal{M}^\perp}$ bases ortonormales de $\mathcal{M}$ y $\mathcal{M}^\perp$, respectivamente. $B_\mathcal{S} = B_\mathcal{M} \cup B_{\mathcal{M}^\perp}$ es una base de un subespacio $S \subseteq \mathcal{V}$. Podemos extender $B_\mathcal{S}$ hasta obtener una base ortogonal $B_\mathcal{V}$ de $\mathcal{V}$ (el procedimiento de Gram-Schmidt, que veremos a continuación, nos garantiza que esto es posible).

Sea $\mathcal{E}$ el conjunto de vectores que agregamos, de modo que $B_\mathcal{V} = B_\mathcal{M} \cup B_{\mathcal{M}^\perp} \cup \mathcal{E}$.

Como ésta es una base ortonormal, $\mathcal{E} \perp B_\mathcal{M}$. Luego, $\mathcal{E} \perp \mathcal{M} \Rightarrow \mathcal{E} \subseteq \mathcal{M}^\perp \Rightarrow \mathcal{E} \in \mathrm{span}(B_{\mathcal{M}^\perp})$.

Pero esto no es posible, ya que $B_\mathcal{V}$ es un conjunto linealmente independiente, luego $\mathcal{E} = \{\}$ y $\mathcal{V} = \mathcal{M} + \mathcal{M}^\perp$. Por lo que demostramos previamente, resulta entonces que

$$\mathcal{V} = \mathcal{M} \oplus \mathcal{M}^\perp. \tag{7.81}$$

Resta probar que el complemento ortogonal es único:

$$\mathcal{A} \perp \mathcal{M} \Rightarrow \mathcal{A} \subseteq \mathcal{M}^{\perp}, \quad \text{pero} \quad \mathcal{M} \oplus \mathcal{M}^{\perp} = \mathcal{V} = \mathcal{M} \oplus \mathcal{A} \Rightarrow \mathcal{A} = \mathcal{M}^{\perp} \tag{7.82}$$

■

∗ Corolario: $\dim(\mathcal{M}^{\perp}) = n - \dim(\mathcal{M})$. Vale también (lo damos sin prueba) que $(\mathcal{M}^{\perp})^{\perp} = \mathcal{M}$.

7.6.1. Subespacios fundamentales de una matriz

Existen 4 subespacios fundamentales de una matriz $A \in \mathbb{R}^{m \times n}$:

1. espacio fila ($\mathrm{Fil}(A)$),
2. espacio columna ($\mathrm{Col}(A)$),
3. núcleo de A ($\mathcal{N}(A)$),
4. núcleo de A^T ($\mathcal{N}(A^T)$).

Estos 4 subespacios están relacionados entre sí, como se demuestra en el siguiente teorema.

Teorema. Sea $A \in \mathbb{R}^{m \times n}$. El complemento ortogonal del espacio fila de A es el espacio nulo de A, y el complemento ortogonal del espacio columna de A es el espacio nulo de A^T. Es decir,

$$(\mathrm{Fil}(A))^{\perp} = \mathcal{N}(A) \qquad \wedge \qquad (\mathrm{Col}(A))^{\perp} = \mathcal{N}(A^T). \tag{7.83}$$

Demostración

Sea $x \in \mathbb{R}^n$. Por definición,

$$x \in (\mathrm{Fil}(A))^{\perp} \iff x \text{ es ortogonal a todas las filas de } A. \tag{7.84}$$

Pero si sucede esto, entonces $Ax = 0$. Por lo tanto,

$$x \in \mathcal{N}(A) \qquad \Rightarrow \qquad \boxed{(\mathrm{Fil}(A))^{\perp} = \mathcal{N}(A).} \tag{7.85}$$

Podemos aplicar el mismo razonamiento a la matriz A^T y obtendremos que

$$\left(\mathrm{Fil}(A^T)\right)^{\perp} = \mathcal{N}(A^T). \tag{7.86}$$

Teniendo en cuenta que

$$\mathrm{Fil}(A^T) = \mathrm{Col}(A), \tag{7.87}$$

podemos concluir que

$$\boxed{(\mathrm{Col}(A))^{\perp} = \mathcal{N}(A^T).} \tag{7.88}$$

■

∗ Corolario 1: Usando que $\left(\mathcal{M}^{\perp}\right)^{\perp} = \mathcal{M}$ para cualquier subespacio $\mathcal{M}$, podemos afirmar que

$$\mathrm{Fil}(A) = \mathcal{N}(A)^{\perp} \qquad \wedge \qquad \mathrm{Col}(A) = \mathcal{N}(A^T)^{\perp}. \tag{7.89}$$

∗ Corolario 2: Si $A \in \mathbb{R}^{m \times n}$,

$$\mathbb{R}^m = \mathrm{Col}(A) \oplus (\mathrm{Col}(A))^{\perp} = \mathrm{Col}(A) \oplus \mathcal{N}(A^T)$$

$$\mathbb{R}^n = \mathcal{N}(A) \oplus (\mathcal{N}(A))^{\perp} = \mathcal{N}(A) \oplus \mathrm{Fil}(A)$$

∗ ∗ ∗

Definición (Proyección ortogonal). Para $v \in \mathcal{V} = \mathcal{M} \oplus \mathcal{M}^\perp$, sea $v = m + n$, con $m \in \mathcal{M}$ y $n \in \mathcal{M}^\perp$. Luego

- (i) El vector m es la proyección ortogonal de v sobre $\mathcal{M}$.
- (ii) El proyector $\mathcal{P}_\mathcal{M}$ sobre $\mathcal{M}$ a lo largo de $\mathcal{M}^\perp$ es llamado proyector ortogonal sobre $\mathcal{M}$.

$*$ Como caso particular del teorema de proyectores, vale que $\mathcal{P}_\mathcal{M}$ es el único operador lineal tal que $\mathcal{P}_\mathcal{M}(v) = m$.

7.6.2. Construcción de proyectores ortogonales

Sea $\mathcal{M}$ un subespacio r-dimensional de $\mathbb{R}^n$, y sean las columnas de las matrices $M^{n\times r}$ y $N^{n\times(n-r)}$ bases de $\mathcal{M}$ y $\mathcal{M}^\perp$ respectivamente. Luego, los proyectores ortogonales sobre $\mathcal{M}$ y $\mathcal{M}^\perp$ son

$$\mathcal{P}_\mathcal{M} = M\left(M^TM\right)^{-1}M^T \qquad \wedge \qquad \mathcal{P}_{\mathcal{M}^\perp} = N\left(N^TN\right)^{-1}N^T. \tag{7.90}$$

Si además las columnas de M y N son ortonormales, tenemos que

(I) $\mathcal{P}_\mathcal{M} = MM^T$ y $\mathcal{P}_{\mathcal{M}^\perp} = NN^T$

(II) $\mathcal{P}_\mathcal{M} = U\begin{pmatrix} I_{r\times r} & 0 \\ 0 & 0\end{pmatrix}U^T$, con $U = (M\ N)$.

(III) $\mathcal{P}_{\mathcal{M}^\perp} = I_{n\times n} - \mathcal{P}_\mathcal{M}$

$*$ Si $\mathcal{M}$ es un subespacio de $\mathbb{C}^n$, en vez de usar transposición, debemos usar la operación $*$, es decir, transposición y conjugación.

$*$ Los proyectores ortogonales tienen ciertas características que no tienen los demás proyectores. Si $\mathbf{P} \in \mathbb{R}^{n\times n}$ es un proyector, las siguientes afirmaciones son equivalentes:

(I) $\mathbf{P}$ es un **proyector ortogonal**

(II) $\mathcal{R}(\mathbf{P}) \perp \mathcal{N}(\mathbf{P})$

(III) $\mathbf{P}^T = \mathbf{P}$ (es decir, $\mathbf{P}$ es proyector ortogonal $\iff \mathbf{P}^2 = \mathbf{P} = \mathbf{P}^T$)

(IV) $||\mathbf{P}||_2 = 1$ para la norma 2 matricial

7.6.3. Cuadrados mínimos a la Álgebra Lineal

Con todo lo aprendido hasta acá, podemos presentar al método de Cuadrados Mínimos con un enfoque distinto.

Definición. Sea $\mathcal{W}$ un subespacio de un espacio vectorial normado $\mathcal{V}$, y sea $v \in \mathcal{V}$. Llamamos la **mejor aproximación a v en $\mathcal{W}$** al vector $\overline{v}$ tal que

$$||v - \overline{v}|| < ||v - w||, \qquad \forall w \in \mathcal{W}, w \neq \overline{v}. \tag{7.91}$$

Teorema de la mejor aproximación. Si $\mathcal{W}$ es un subespacio de un espacio vectorial con producto interno $\mathcal{V}$ y si $v \in \mathcal{V}$, entonces la proyección ortogonal de v sobre $\mathcal{W}$, $\mathcal{P}_\mathcal{W}(v)$, es la mejor aproximación a v en $\mathcal{W}$.

Demostración

Sea $w \in \mathcal{W}$ con $w \neq \mathcal{P}_{\mathcal{W}}(v)$. Entonces

$$\mathcal{P}_{\mathcal{W}}(v) - w \in \mathcal{W}, \tag{7.92}$$

y, además,

$$v - \mathcal{P}_{\mathcal{W}}(v) \perp \mathcal{P}_{\mathcal{W}}(v) - w, \tag{7.93}$$

es decir, $v - \mathcal{P}_{\mathcal{W}}(v)$ "vive" en $\mathcal{W}^{\perp}$. Usando el teorema de Pitágoras (ojo, lo siguiente es válido sólo para espacios reales sobre los reales) podemos escribir

$$||v - \mathcal{P}_{\mathcal{W}}(v)||^2 + ||\mathcal{P}_{\mathcal{W}}(v) - w||^2 = ||v - \cancel{\mathcal{P}_{\mathcal{W}}(v)} + \cancel{\mathcal{P}_{\mathcal{W}}(v)} - w||^2 = ||v - w||^2. \tag{7.94}$$

Como

$$||\mathcal{P}_{\mathcal{W}}(v) - w||^2 > 0, \tag{7.95}$$

ya que, por hipótesis, $w \neq \mathcal{P}_{\mathcal{W}}(v)$, concluimos que

$$\boxed{||v - \mathcal{P}_{\mathcal{W}}(v)||^2 < ||v - w||^2, \qquad \forall w \in \mathcal{W}/w \neq \mathcal{P}_{\mathcal{W}}(v).} \tag{7.96}$$

En consecuencia, $\mathcal{P}_{\mathcal{W}}(v)$ es la mejor aproximación a v en $\mathcal{W}$.

■

Podemos aplicar lo anterior al método de Cuadrados Mínimos. Recordemos que el método busca encontrar un vector x que haga mínimo a

$$E_2(x) = \varepsilon^T(x)\varepsilon(x) = (Ax - b)^T(Ax - b) = ||Ax - b||_2^2, \tag{7.97}$$

es decir, estamos buscando el vector x tal que Ax sea tan próximo como sea posible al vector b. Naturalmente, el vector

$$Ax \in \mathcal{R}(A), \qquad \text{(espacio columna de } A) \tag{7.98}$$

y el teorema de la mejor aproximación nos dice que el vector de $\mathcal{R}(A)$ más cercano a b es

$$\mathcal{P}_{\mathcal{R}(A)}(b), \tag{7.99}$$

es decir, la proyección ortogonal de b sobre $\mathcal{R}(A)$. Por lo tanto, debemos encontrar un vector x tal que

$$Ax = \mathcal{P}_{\mathcal{R}(A)}(b). \tag{7.100}$$

Aplicando la proyección ortogonal sobre ambos miembros y usando la idempotencia de los proyectores

$$\mathcal{P}_{\mathcal{R}(A)}(Ax) = \mathcal{P}_{\mathcal{R}(A)}(b) \quad \Rightarrow \quad \mathcal{P}_{\mathcal{R}(A)}(Ax - b) = 0. \tag{7.101}$$

Por lo tanto, la solución que buscamos debe cumplir con que

$$Ax - b \in \mathcal{N}(\mathcal{P}_{\mathcal{R}(A)}). \tag{7.102}$$

Se puede demostrar (ver después de este teorema) que

$$\mathcal{N}(\mathcal{P}_{\mathcal{R}(A)}) = \mathcal{R}(A)^{\perp} = \mathcal{N}(A^T). \tag{7.103}$$

Por lo tanto,

$$A^T(Ax - b) = 0 \quad \Rightarrow \quad \boxed{A^TAx = A^Tb.} \tag{7.104}$$

En consecuencia, el vector x que es tal que Ax es el vector más próximo a b es aquél que es solución de las ecuaciones normales asociadas al sistema $Ax = b$.

Teorema. El complemento ortogonal del espacio columna de una matriz $A \in \mathbb{R}^{m\times n}$ es igual al núcleo de A^T. Esto es,

$$\mathcal{R}(A)^{\perp} = \mathcal{N}(A^T). \tag{7.105}$$

Demostración

Sea $A \in \mathbb{R}^{m\times n}$ y sea $x \in \mathbb{R}^n$, por lo que Ax está bien definida y, por definición, $Ax \in \mathcal{R}(A)$. Sea $y \in \mathcal{R}(A)^\perp$, entonces

$$\langle Ax|y\rangle = 0 \quad \Rightarrow \quad x^T A^T y = 0 \quad \Rightarrow \quad \langle x|A^T y\rangle = 0, \tag{7.106}$$

de lo que resulta que $A^T y$ es perpendicular a todo vector de $\mathbb{R}^n$. El único vector que cumple esto es el vector nulo, es decir, tiene que verificarse que

$$A^T y = 0. \tag{7.107}$$

Por definición, esto quiere decir que $y \in \mathcal{N}(A^T)$. En síntesis,

$$y \in \mathcal{R}(A)^\perp \quad \Rightarrow \quad y \in \mathcal{N}(A^T), \tag{7.108}$$

y concluimos que

$$\boxed{\mathcal{R}(A)^\perp = \mathcal{N}(A^T).} \tag{7.109}$$

■

7.7. Ortogonalización de Gram-Schmidt

Habiendo estudiado proyecciones ortogonales, estamos en condiciones de presentar el método de Gram-Schmidt, utilizado para obtener conjuntos ortogonales u ortonormales, a partir de conjuntos linealmente independientes. Comencemos con un ejemplo en $\mathbb{R}^2$ que nos permitirá visualizar la idea de su funcionamiento.

Ejemplo

Sea $B=\{b_1, b_2\}=\left\{\begin{pmatrix}2\\1\end{pmatrix}, \begin{pmatrix}1\\1\end{pmatrix}\right\}$ una base de $\mathbb{R}^2$. Es fácil ver que B es un conjunto linealmente independiente y que ambos vectores no son ortogonales entre sí

$$\langle b_1|b_2\rangle = b_1^T b_2 = \begin{pmatrix}2 & 1\end{pmatrix}\begin{pmatrix}1\\1\end{pmatrix} = 5 \neq 0, \tag{7.110}$$

por lo que B no es una base ortogonal. Sea ahora

$$\mathcal{M} = \operatorname{span}\left\{\begin{pmatrix}2\\1\end{pmatrix}\right\}, \tag{7.111}$$

el subespacio de $\mathbb{R}^2$ generado por el primer elemento de la base B. Construyamos el proyector ortogonal sobre él:

$$\mathcal{P}_\mathcal{M} = b_1 \left(b_1^T b_1\right)^{-1} b_1^T = \begin{pmatrix}2\\1\end{pmatrix}\left[(2\ 1)\begin{pmatrix}2\\1\end{pmatrix}\right]^{-1}(2\ 1) = \frac{1}{5}\begin{pmatrix}2\\1\end{pmatrix}(2\ 1) = \frac{1}{5}\begin{pmatrix}4 & 2\\2 & 1\end{pmatrix}. \tag{7.112}$$

La proyección ortogonal de b_2 sobre $\mathcal{M}$ es

$$\mathcal{P}_\mathcal{M}(b_2) = \frac{1}{5}\begin{pmatrix}4 & 2\\2 & 1\end{pmatrix}\begin{pmatrix}1\\1\end{pmatrix} = \frac{1}{5}\begin{pmatrix}6\\3\end{pmatrix} = \frac{3}{5}\begin{pmatrix}2\\1\end{pmatrix}. \tag{7.113}$$

Noten que el vector está en $\mathcal{M}$.

El proyector sobre el espacio complementario puede obtenerse haciendo

$$(\mathrm{id} - \mathcal{P}_\mathcal{M})(b_2) = b_2 - \mathcal{P}_\mathcal{M}(b_2) = \begin{pmatrix}1\\1\end{pmatrix} - \frac{3}{5}\begin{pmatrix}2\\1\end{pmatrix} = \begin{pmatrix}1-6/5\\1-3/5\end{pmatrix} = \begin{pmatrix}-1/5\\2/5\end{pmatrix} = \frac{1}{5}\begin{pmatrix}-1\\2\end{pmatrix} \tag{7.114}$$

Verifiquemos los vectores obtenidos son ortogonales:

$$\langle b_1|b_2 - \mathcal{P}_\mathcal{M}(b_2)\rangle = b_1^T(b_2 - \mathcal{P}_\mathcal{M}(b_2)) = (2\ 1)\frac{1}{5}\begin{pmatrix}-1\\2\end{pmatrix} = 0. \tag{7.115}$$

En consecuencia,

$$b_2 - \mathcal{P}_\mathcal{M}(b_2) \in \mathcal{M}^\perp. \tag{7.116}$$

Así, el conjunto $\{b_1, b_2 - \mathcal{P}_\mathcal{M}(b_2)\}$, con $\mathcal{P}_\mathcal{M}$ proyector ortogonal sobre el espacio generado por b_1, es un conjunto ortogonal. Podemos entonces presentar el caso general.

7.7.1. Proceso de ortogonalización de Gram-Schmidt

Si $B = \{b_1, b_2, \dots, b_n\}$ es una base de un espacio vectorial con producto interno de dimensión n, luego, la secuencia de vectores definida por

$$u_1 = \frac{b_1}{||b_1||}, \qquad u_k = \frac{b_k - \sum_{i=1}^{k-1} \langle u_i | b_k \rangle \, u_i}{||b_k - \sum_{i=1}^{k} \langle u_i | b_k \rangle \, u_i||}, \quad k = 2, 3, \dots, n \tag{7.117}$$

forma una base ortonormal.

∗ La demostración de la validez del proceso, que no mostraremos, se hace por inducción completa.

En el caso particular que $S = \{s_1, s_2, \dots, s_n\}$ sea una base de un subespacio $n-$dimensional de $\mathbb{C}^m$, la secuencia de Gram-Schmidt se encuentra como

$$u_k = \frac{(I_{m\times m} - U_k U_k^*)\, s_k}{||\,(I_{m\times m} - U_k U_k^*)\, s_k||}, \quad k = 1, 2, \dots, n, \tag{7.118}$$

en donde $U_1 {=} 0^{m\times 1}$ es un vector nulo de m elementos y $U_k^{m\times(k-1)} {=} (u_1\ u_2\ \dots\ u_{k-1})$ para $k > 1$ es la matriz cuyas columnas son los vectores ortonormales obtenidos hasta el paso previo.

∗ Desde el punto de vista numérico el proceso de ortogonalización de Gram-Schmidt presenta inconvenientes, ya que produce vectores que no son ortogonales. Esto se debe a que, tal como está dado, amplifica los errores de redondeo. Existen variaciones, equivalentes desde el punto de vista teórico, que producen resultados significativamente mejores.

7.7.2. Factorización QR

Consideremos una matriz $A^{m\times n}$ cuyas columnas forman un conjunto linealmente independiente. Escribimos $A = (a_1\ a_2\ \dots\ a_n)$, donde $a_j = (a_{1j}\ a_{2j}\ \dots\ a_{mj})^T$, $j = 1, 2, \dots, n$ son las columnas.

Si aplicamos Gram-Schmidt a los vectores columna, obtenemos un conjunto $\{q_1\ q_2\ \dots\ q_n\}$ de vectores ortonormales que, naturalmente, serán también base del espacio columna de A. De las fórmulas (7.117) podemos escribir

$$q_1 = \frac{a_1}{\mu_1}, \qquad q_k = \frac{a_k - \sum_{i=1}^{k-1} \langle q_i | a_k \rangle \, q_i}{\mu_k}, \quad k = 2, 3, \dots, n. \tag{7.119}$$

donde

$$\mu_1 = ||a_1||, \qquad \text{y} \qquad \mu_k = ||a_k - \sum_{i=1}^{k-1} \langle q_i | a_k \rangle \, q_i||. \tag{7.120}$$

De (7.119) podemos escribir

$$a_1 = \mu_1 q_1, \qquad a_k = \mu_k q_k + \langle q_1 | a_k \rangle \, q_1 + \langle q_2 | a_k \rangle \, q_2 + \dots + \langle q_{k-1} | a_k \rangle \, q_{k-1},$$

que a su vez puede incorporarse en A, para reescribirla como

$$(a_1\ a_2\ \dots\ a_n) = (q_1\ q_2\ \dots\ q_n) \begin{pmatrix} \mu_1 & \langle q_1|a_2\rangle & \langle q_1|a_3\rangle & \dots & \langle q_1|a_n\rangle \\ 0 & \mu_2 & \langle q_2|a_3\rangle & \dots & \langle q_2|a_n\rangle \\ 0 & 0 & \mu_3 & \dots & \langle q_3|a_n\rangle \\ & & \vdots & \ddots & \\ 0 & 0 & 0 & \dots & \mu_n \end{pmatrix}.$$

De acuerdo con esta igualdad matricial, toda matriz $A^{m\times n}$ con n columnas linealmente independientes puede factorizarse como

$$A^{m\times n} = Q^{m\times n} R^{n\times n}, \tag{7.121}$$

donde Q es una matriz con sus columnas ortogonales y R es una matriz triangular superior con elementos diagonales estrictamente positivos.

∗ La fórmula (7.121) se denomina *Factorización QR* de la matriz A. Cada matriz factorizable QR tiene una única descomposición posible.

$*$ Si la matriz $A^{n\times n}$ tiene sus columnas linealmente independientes, la matriz Q resultante de su factorización es *ortonormal* (observen que los vectores q_k tienen norma uno). Tales matrices tienen una serie de propiedades interesantes, que damos sin demostrar:

- Q tiene sus columnas ortonormales
- Q tiene sus filas ortonormales
- $Q^{-1} = Q^T$
- $||Qx||_2 = ||x||_2$, $\forall x \in \mathbb{R}^n$.
- Si $Q \in \mathbb{C}^{n\times n}$, se la denomina *unitaria*, y valen las propiedades anteriores, salvo que $Q^{-1} = Q^*$.

$*$ Ejemplos de matrices ortogonales son las reflexiones y las rotaciones alrededor de los ejes coordenados.