

El producto escalar

Ingeniería Biomédica Curso: Matemáticas I

Fernando Chamizo <https://matematicas.uam.es/~fernando.chamizo/>

1. Definición y propiedades

El *producto escalar* en \mathbb{R}^n que conoces de cursos pasados para $n = 2$ y $n = 3$ se extiende a dimensiones superiores de la forma obvia:

$$\vec{x} \cdot \vec{y} = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n \quad \text{con } \vec{x} = (x_1, x_2, \dots, x_n)^t, \quad \vec{y} = (y_1, y_2, \dots, y_n)^t.$$

Otra notación medianamente común para el producto escalar es $\langle \vec{x}, \vec{y} \rangle$, sobre todo para generalizaciones de $\vec{x} \cdot \vec{y}$ como las que veremos después.

Esta operación tan sencilla de multiplicar las coordenadas respectivas y sumar los resultados es interesante porque está relacionada con medir longitudes y ángulos. Concretamente, la *norma (longitud)* de un vector \vec{x} y el *ángulo* α entre dos vectores \vec{x} e \vec{y} vienen dados por las fórmulas

$$(1) \quad \|\vec{x}\| = \sqrt{\vec{x} \cdot \vec{x}} \quad \text{y} \quad \cos \alpha = \frac{\vec{x} \cdot \vec{y}}{\|\vec{x}\| \|\vec{y}\|} \quad \text{con } 0 \leq \alpha \leq \pi.$$

Geométricamente, dos vectores en el plano dibujados como flechas determinan dos ángulos (en sentido antihorario) que suman 2π , el que va de \vec{x} a \vec{y} y el que va de \vec{y} a \vec{x} , por convenio se toma siempre el más pequeño y de ahí la restricción $0 \leq \alpha \leq \pi$. En matemáticas medianamente avanzadas siempre se usan radianes, así que este π son los 180° de toda la vida. Como $\cos \frac{\pi}{2} = 0$ y el coseno no se anula en ningún otro valor en el rango de α , se concluye que dos vectores no nulos son perpendiculares si y solo si su producto escalar es cero.

Por ejemplo, los vectores $\vec{x} = (3, 4)^t$ e $\vec{y} = (-1, 7)^t$ de \mathbb{R}^2 tienen producto escalar $\vec{x} \cdot \vec{y} = 3(-1) + 4 \cdot 7 = 25$, normas $\|\vec{x}\| = \sqrt{3^2 + 4^2} = 5$, $\|\vec{y}\| = \sqrt{(-1)^2 + 7^2} = \sqrt{50} = 5\sqrt{2}$ y el ángulo entre ellos es $\pi/4$ (lo que equivale a 45°) porque $\cos \alpha = 25/(5 \cdot 5\sqrt{2}) = 1/\sqrt{2}$.

En algunas situaciones es conveniente considerar maneras de medir generalizadas en espacios vectoriales sobre \mathbb{R} . Un ejemplo cotidiano, aunque avanzado para este curso, es que en las señales (funciones) reales de periodo T resulta natural medir con $\langle f, g \rangle = \frac{1}{T} \int_0^T fg$. El voltaje de nuestros enchufes en función del tiempo es de la forma $f(t) = A \operatorname{sen}(2\pi t/T)$ con $T = 50$ (oscila 50 veces por segundo). Cuando se dice que su corriente es de 230 voltios no significa

$A = 230$ sino $\sqrt{\langle f, f \rangle} = 230$, lo que conduce a $A = 230\sqrt{2}$, por tanto, hay picos de más de 300 voltios. Este producto escalar permite cuantificar mejor lo que aprovechan los motores eléctricos de esta corriente oscilante.

La existencia de productos escalares raros útiles motivó en el desarrollo del álgebra lineal una definición abstracta muy general de producto escalar, una declaración de mínimos requeridos para que algo se pueda considerar que lejanamente tiene que ver con longitudes y ángulos.

Si V es un espacio vectorial sobre \mathbb{R} , se dice que una función $\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{R}$ es un producto escalar (generalizado) si tiene las tres siguientes propiedades:

- 1) Es *lineal en el primer argumento*¹: Para todo $\vec{x}, \vec{y}, \vec{z} \in V$, $\lambda, \mu \in \mathbb{R}$ se cumple $\langle \lambda\vec{x} + \mu\vec{y}, \vec{z} \rangle = \lambda\langle \vec{x}, \vec{z} \rangle + \mu\langle \vec{y}, \vec{z} \rangle$.
- 2) Es *simétrica*: $\langle \vec{x}, \vec{y} \rangle = \langle \vec{y}, \vec{x} \rangle$ para todo $\vec{x}, \vec{y} \in V$.
- 3) Es *definida positiva*: $\langle \vec{x}, \vec{x} \rangle > 0$ para todo $\vec{x} \in V - \{\vec{0}\}$.

Los espacios vectoriales sobre \mathbb{R} en los que hemos definido un producto escalar con estas propiedades se dice que son *espacios euclídeos*.

Por ejemplo, en la línea de lo dicho antes, $\langle P, Q \rangle = \int_{-1}^1 P Q$ define un producto escalar en el espacio de polinomios $\mathbb{R}_n[x]$ convirtiéndolo en un espacio euclídeo. La linealidad se sigue de la linealidad de la integral:

$$\langle \lambda P_1 + \mu P_2, Q \rangle = \int_{-1}^1 (\lambda P_1 + \mu P_2) Q = \lambda \int_{-1}^1 P_1 Q + \mu \int_{-1}^1 P_2 Q = \lambda \langle P_1, Q \rangle + \mu \langle P_2, Q \rangle.$$

La simetría es evidente por la comutativa del producto de polinomios. Finalmente, $\langle P, P \rangle > 0$ para todo polinomio P no nulo porque $\langle P, P \rangle = \int_{-1}^1 P^2$ es el área limitada por la gráfica de P^2 sobre el intervalo $[-1, 1]$ del eje X .

En \mathbb{R}^2 se tiene que $\langle \vec{x}, \vec{y} \rangle = x_1 y_1 + 2x_2 y_2$ define un producto escalar generalizado, distinto del habitual. La comprobación de la primera propiedad es:

$$\begin{aligned} \langle \lambda\vec{x} + \mu\vec{y}, \vec{z} \rangle &= (\lambda x_1 + \mu y_1) z_1 + 2(\lambda x_2 + \mu y_2) z_2 \\ &= \lambda(x_1 z_1 + 2x_2 z_2) + \mu(y_1 z_1 + 2y_2 z_2) = \lambda\langle \vec{x}, \vec{z} \rangle + \mu\langle \vec{y}, \vec{z} \rangle. \end{aligned}$$

La segunda propiedad es bastante trivial y también la tercera, pues $\langle \vec{x}, \vec{x} \rangle = x_1^2 + 2x_2^2$ es suma de cuadrados. Con este producto escalar, el vector $\vec{x} = (1, 2)^t$ tiene longitud $\|\vec{x}\| = \sqrt{1 \cdot 1 + 2 \cdot 2} = 3$, mientras que con el producto escalar usual mediría $\sqrt{5}$.

Por otro lado, $\langle \vec{x}, \vec{y} \rangle = x_1 y_2 - x_2 y_1$ no define un producto escalar en \mathbb{R}^2 . Se cumple la primera propiedad, pero no las otras dos ya que $\langle \vec{x}, \vec{y} \rangle = -\langle \vec{y}, \vec{x} \rangle$ y $\langle \vec{x}, \vec{x} \rangle = 0$.

La política que usaremos aquí será preferir la notación $\langle \vec{x}, \vec{y} \rangle$ para los productos escalares generalizados y reservar $\vec{x} \cdot \vec{y}$ para el producto escalar usual.

¹A menudo se pide la linealidad en ambos argumentos, pero por la segunda propiedad, basta hacerlo en uno.

En muchas aplicaciones se utilizan espacios vectoriales sobre \mathbb{C} . Es muy posible que nunca te hayan hablado del producto escalar usual en \mathbb{C}^n . La fórmula es muy similar a la de \mathbb{R}^n con una pequeña salvedad. Para $\vec{x}, \vec{y} \in \mathbb{C}^n$ se define²

$$\vec{x} \cdot \vec{y} = x_1\bar{y}_1 + x_2\bar{y}_2 + \cdots + x_n\bar{y}_n \quad \text{con} \quad \vec{x} = (x_1, x_2, \dots, x_n)^t, \quad \vec{y} = (y_1, y_2, \dots, y_n)^t$$

donde la barra sobre las y_j significa el conjugado. Esto se hace para que la norma $\|\vec{x}\|$ definida como la raíz cuadrada de $\vec{x} \cdot \vec{x}$ sea un número mayor o igual que cero. Es fácil ver que esta norma es la suma de los cuadrados de los módulos de las coordenadas.

Como ejemplo, si $\vec{v}_1 = (2+i, 1+i)^t$, $\vec{v}_2 = (-1+i, 2-i)^t$ y $\vec{v}_3 = (3, i)^t$, se tiene

$$\vec{v}_1 \cdot \vec{v}_2 = 0, \quad \vec{v}_1 \cdot \vec{v}_3 = 7+2i, \quad \vec{v}_3 \cdot \vec{v}_1 = 7-2i, \quad \vec{v}_2 \cdot \vec{v}_3 = -4+i$$

y

$$\|\vec{v}_1\| = \sqrt{|2+i|^2 + |1+i|^2} = \sqrt{2^2 + 1^2 + 1^2 + 1^2} = \sqrt{7}, \quad \|\vec{v}_2\| = \sqrt{7}, \quad \|\vec{v}_3\| = \sqrt{10}.$$

También hay una definición generalizada de producto escalar para espacios vectoriales sobre \mathbb{C} igual a la de \mathbb{R} excepto que la simetría de la segunda propiedad hay que cambiarla por

$$\langle \vec{x}, \vec{y} \rangle = \overline{\langle \vec{y}, \vec{x} \rangle}.$$

Los espacios vectoriales sobre \mathbb{C} con un producto escalar en este sentido generalizado se llaman *espacios hermíticos*. Esto es, son los análogos sobre \mathbb{C} de los espacios euclídeos.

Hay tres propiedades con nombre que satisfacen todos los productos escalares generalizados, sobre \mathbb{R} o sobre \mathbb{C} . Son consecuencias de la definición, aunque sus pruebas no son inmediatas. En este curso no las demostraremos ni haremos uso de ellas, pero son demasiado importantes como para no mencionarlas:

Desigualdad de Cauchy-Schwarz: $|\langle \vec{x}, \vec{y} \rangle| \leq \|\vec{x}\| \|\vec{y}\|$.

Desigualdad triangular: $\|\vec{x} + \vec{y}\| \leq \|\vec{x}\| + \|\vec{y}\|$.

Teorema de Pitágoras: $\|\vec{x} + \vec{y}\|^2 = \|\vec{x}\|^2 + \|\vec{y}\|^2$ siempre que $\langle \vec{x}, \vec{y} \rangle = 0$.

El nombre de la última propiedad proviene de que con el producto escalar usual en \mathbb{R}^2 , la condición $\langle \vec{x}, \vec{y} \rangle = 0$ indica que \vec{x} e \vec{y} son perpendiculares. Si haces un dibujo, verás que $\vec{x} + \vec{y}$ mide lo mismo que la hipotenusa correspondiente a los catetos \vec{x} e \vec{y} .

Solo para practicar, veamos que los vectores $\vec{x}, \vec{y} \in \mathbb{C}^3$ (con el producto escalar usual) dados por $\vec{x} = (2+i, 1+i, 1)^t$, $\vec{y} = (1-3i, 1-i, 1+9i)^t$ están bajo las hipótesis del teorema de Pitágoras y cumplen su conclusión. Las cuentas, con un poco de detalle son:

$$\vec{x} \cdot \vec{y} = (2+i)(1+3i) + (1+i)(1+i) + (1-9i) = (-1+7i) + 2i + (1-9i) = 0.$$

²En física y algunas áreas de las matemáticas las x_j y las y_j están intercambiadas de modo que la linealidad que falla es la del primer argumento.

La suma es $\vec{x} + \vec{y} = (3 - 2i, 2, 2 + 9i)^t$ y se tienen las normas

$$\begin{aligned}\|\vec{x}\| &= \sqrt{|2+i|^2 + |1+i|^2 + 1^2} = \sqrt{5+2+1} = \sqrt{8}, \\ \|\vec{y}\| &= \sqrt{|1-3i|^2 + |1-i|^2 + |1+9i|^2} = \sqrt{10+2+82} = \sqrt{94}, \\ \|\vec{x} + \vec{y}\| &= \sqrt{|3-2i|^2 + 2^2 + |2+9i|^2} = \sqrt{13+4+85} = \sqrt{102}.\end{aligned}$$

En consonancia con el teorema, se tiene $8 + 94 = 102$, ¡Pitágoras nunca falla!

2. Ortogonalidad

En la jerga del álgebra lineal se utiliza el término *ortogonalidad* para indicar la generalización de la perpendicularidad en \mathbb{R}^2 o \mathbb{R}^3 . Esto es, diremos que dos vectores \vec{x} e \vec{y} son *ortogonales* si $\langle \vec{x}, \vec{y} \rangle = 0$. Si además \vec{x} e \vec{y} son *unitarios* (esto es, $\|\vec{x}\| = \|\vec{y}\| = 1$) se dice que son *ortonormales*. Por extensión, se dice que una base $\mathcal{B} = \{\vec{u}_1, \dots, \vec{u}_n\}$ de un espacio vectorial con producto escalar es *base ortogonal* si los \vec{u}_j son ortogonales dos a dos y que es *base ortonormal* si son ortonormales. Un resultado sencillo afirma que *los vectores ortogonales no nulos son siempre linealmente independientes, por tanto, para comprobar que un conjunto de vectores ortogonales es base de un espacio V basta ver que está compuesto por dim V vectores no nulos*.

Por ejemplo $\mathcal{B} = \{\vec{u}_1 = (6, 4)^t, \vec{u}_2 = (2, -3)^t\}$ es base ortogonal de \mathbb{R}^2 . Es base porque está compuesta por $2 = \dim \mathbb{R}^2$ vectores y es ortogonal porque $\vec{u}_1 \cdot \vec{u}_2 = 0$. No es ortonormal porque $\|\vec{u}_1\| = 2\sqrt{13} \neq 1$ y tampoco \vec{u}_2 es unitario.

En \mathbb{R}^3 ,

$$\mathcal{B} = \left\{ \vec{u}_1 = \frac{1}{9}(8, 4, 1)^t, \vec{u}_2 = \frac{1}{9}(4, -7, -4)^t, \vec{u}_3 = \frac{1}{9}(1, -4, 8)^t \right\},$$

La ortonormalidad se sigue comprobando $\vec{u}_1 \cdot \vec{u}_2 = \vec{u}_1 \cdot \vec{u}_3 = \vec{u}_2 \cdot \vec{u}_3 = 0$ y $\|\vec{u}_1\| = \|\vec{u}_2\| = \|\vec{u}_3\| = 1$. Una vez sabido esto, son base porque hay 3 de ellos.

Un último ejemplo con números complejos cuyos cálculos, ya sea completos o parciales, se dejan al que quiera practicar, es:

$$\mathcal{B} = \left\{ \vec{u}_1 = (4 + 3i, 4, -2 + 6i)^t, \vec{u}_2 = (4, 4 - 3i, -2 - 6i)^t, \vec{u}_3 = (-2 + 6i, -2 - 6i, 1)^t \right\},$$

que constituye una base ortogonal de \mathbb{C}^3 . Se cumple $\|\vec{u}_1\| = \|\vec{u}_2\| = \|\vec{u}_3\| = 9$, por ejemplo $\|\vec{u}_1\|^2 = |4 + 3i|^2 + 4^2 + |-2 + 6i|^2 = 25 + 16 + 40 = 81$. Por tanto al dividir los vectores de \mathcal{B} por 9 se obtiene una base ortonormal.

Ya hemos visto, que la ortogonalidad nos ahorra comprobar la independencia lineal. Otra propiedad importante es que permite calcular las coordenadas, de manera muy simple, sin resolver sistemas.

Si $\mathcal{B} = \{\vec{u}_1, \dots, \vec{u}_n\}$ es una base ortogonal de un espacio vectorial V , para cualquier vector $\vec{x} \in V$ se tiene

$$\vec{x} = x_1 \vec{u}_1 + x_2 \vec{u}_2 + \cdots + x_n \vec{u}_n \quad \text{con} \quad x_j = \frac{\langle \vec{x}, \vec{u}_j \rangle}{\|\vec{u}_j\|^2}.$$

Es decir, $\|\vec{u}_j\|^{-2} \langle \vec{x}, \vec{u}_j \rangle$ son las coordenadas de \vec{x} en la base \mathcal{B} .

Un ejemplo de que esto funciona es hallar las coordenadas de $\vec{v} = (0, 2, 1)^t$ en la base ortonormal de \mathbb{R}^3 considerada antes. En vez de calcular x_1 , x_2 y x_3 resolviendo el sistema $\vec{v} = x_1 \vec{u}_1 + x_2 \vec{u}_2 + x_3 \vec{u}_3$, en virtud del resultado anterior, tenemos directamente

$$x_1 = \vec{v} \cdot \vec{u}_1 = 1, \quad x_2 = \vec{v} \cdot \vec{u}_2 = -2, \quad x_3 = \vec{v} \cdot \vec{u}_3 = 0.$$

Es fácil comprobar $\vec{v} = \vec{u}_1 - 2\vec{u}_2$ y por tanto este cálculo es correcto.

Otro ejemplo, esta vez con una base ortogonal, es el cálculo de las coordenadas de $\vec{x} = (1, 3, 7)^t$ en $\mathcal{B} = \{\vec{u}_1, \vec{u}_2, \vec{u}_3\}$ con

$$\vec{u}_1 = (2, 2, 1)^t, \quad \vec{u}_2 = (2, -1, -2)^t, \quad \vec{u}_3 = (1, -2, 2)^t$$

que es base ortogonal de \mathbb{R}^3 . Se tiene $\|\vec{u}_j\| = 3$ y las fórmulas de las coordenadas dan

$$x_1 = \frac{15}{9} = \frac{5}{3}, \quad x_2 = -\frac{15}{9} = -\frac{5}{3}, \quad x_3 = \frac{9}{9} = 1.$$

El resultado se comprueba fácilmente con $\vec{x} = 5(\vec{u}_1 - \vec{u}_2)/3 + \vec{u}_3$.

Veamos un último ejemplo con números complejos hallando las coordenadas de $\vec{x} = (-5 + i, 6 + i)^t$ en la base ortogonal $\mathcal{B} = \{(1 + i, i)^t, (-2 + i, 3 + i)^t\}$. Las cuentas, con detalle, son:

$$\begin{cases} x_1 = \frac{(-5 + i)(1 - i) + (6 + i)(-i)}{|1 + i|^2 + |i|^2} = \frac{-3}{3} = -1, \\ x_2 = \frac{(-5 + i)(-2 - i) + (6 + i)(3 - i)}{|-2 + i|^2 + |3 + i|^2} = \frac{30}{15} = 2, \end{cases}$$

que es coherente con $\vec{x} = -(1 + i, i)^t + 2(-2 + i, 3 + i)^t$.

La moraleja es que hallar coordenadas en bases ortogonales es mucho más fácil que en el caso general: en vez de resolver sistemas de ecuaciones debemos calcular productos escalares, lo cual resulta una ventaja computacional inmensa en algunas aplicaciones. Esto motiva buscar un procedimiento para transformar una base en otra ortogonal. Existe un algoritmo famoso con tal fin llamado *proceso de Gram-Schmidt*. La entrada es una base $\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$ que no es ortogonal y la salida otra $\{\vec{u}_1, \vec{u}_2, \dots, \vec{u}_n\}$ que sí lo es. Los pasos del algoritmo son

$$(2) \quad \vec{u}_1 = \vec{v}_1 \quad \text{y} \quad \vec{u}_j = \vec{v}_j - \sum_{k=1}^{j-1} \frac{\langle \vec{v}_j, \vec{u}_k \rangle}{\|\vec{u}_k\|^2} \vec{u}_k, \quad j = 2, 3, \dots, n.$$

Si tras este proceso dividimos \vec{u}_j por $\|\vec{u}_j\|$, lo que se llama *normalización*, los vectores resultantes serán unitarios y formarán una base ortonormal.

En el caso complejo de espacios hermíticos es importante no alterar el orden del producto escalar en el numerador porque $\langle \vec{v}_j, \vec{u}_k \rangle$ en general difiere de $\langle \vec{u}_k, \vec{v}_j \rangle$, ya que es su conjugado.

Veremos más adelante una interpretación de (2) en términos de la proyección ortogonal. Si uno la entiende, no es necesario memorizar estas fórmulas, se pueden deducir sobre la marcha.

Para seguir las cuentas veamos primero un ejemplo muy sencillo en que partimos de la base de \mathbb{R}^3

$$\mathcal{B} = \{\vec{v}_1 = (2, 0, 0)^t, \vec{v}_2 = (1, 0, 3)^t, \vec{v}_3 = (3, 5, 7)^t\}.$$

Según el algoritmo (2), tomamos $\vec{u}_1 = (2, 0, 0)^t$ como paso inicial y el siguiente, correspondiente a $j = 2$, produce

$$\vec{u}_2 = \vec{v}_2 - \frac{\langle \vec{v}_2, \vec{u}_1 \rangle}{\|\vec{u}_1\|^2} \vec{u}_1 = \begin{pmatrix} 1 \\ 0 \\ 3 \end{pmatrix} - \frac{2}{4} \begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 3 \end{pmatrix}.$$

Finalmente para $j = 3$ se obtiene

$$\vec{u}_3 = \vec{v}_3 - \frac{\langle \vec{v}_3, \vec{u}_1 \rangle}{\|\vec{u}_1\|^2} \vec{u}_1 - \frac{\langle \vec{v}_3, \vec{u}_2 \rangle}{\|\vec{u}_2\|^2} \vec{u}_2 = \begin{pmatrix} 3 \\ 5 \\ 7 \end{pmatrix} - \frac{6}{4} \begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix} - \frac{21}{9} \begin{pmatrix} 0 \\ 0 \\ 3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 5 \end{pmatrix}.$$

Es evidente que $\mathcal{B} = \{\vec{u}_1, \vec{u}_2, \vec{u}_3\}$ es ortogonal. Si queremos una base ortonormal al cambiar \vec{u}_j por $\vec{u}_j/\|\vec{u}_j\|$ obtenemos una reordenación de la base canónica.

Veamos ahora el resultado del proceso de Gram-Schmidt aplicado a los vectores $\vec{v}_1 = (4, -3)^t$ y $\vec{v}_2 = (2, 1)^t$ de \mathbb{R}^2 con el producto escalar usual. La dimensión es dos y, entonces, hay dos pasos, el primero de ellos trivial:

$$\vec{u}_1 = \begin{pmatrix} 4 \\ -3 \end{pmatrix}, \quad \vec{u}_2 = \begin{pmatrix} 2 \\ 1 \end{pmatrix} - \frac{8-3}{25} \begin{pmatrix} 4 \\ -3 \end{pmatrix} = \begin{pmatrix} 6/5 \\ 8/5 \end{pmatrix}.$$

Si normalizamos, resulta la base ortonormal $\{\frac{1}{5}(4, -3)^t, \frac{1}{5}(3, 4)^t\}$.

Al hilo de los ejemplos anteriores, ortogonalizar en \mathbb{R}^n o \mathbb{C}^n es un poco superfluo porque disponemos de la base canónica que ya es ortogonal. En general, no tenemos nada similar en sus subespacios y el proceso cobra sentido ahí. Por ejemplo, consideremos

$$V = \{\vec{x} \in \mathbb{R}^3 : x_1 + 2x_2 - 5x_3 = 0\}.$$

Su dimensión es 2. Lo usual es que para hallar una base elijamos $x_2 = \lambda, x_3 = \mu$ de modo que

$$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} -2\lambda + 5\mu \\ \lambda \\ \mu \end{pmatrix} = \lambda \vec{v}_1 + \mu \vec{v}_2 \quad \text{con} \quad \vec{v}_1 = \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix}, \quad \vec{v}_2 = \begin{pmatrix} 5 \\ 0 \\ 1 \end{pmatrix}.$$

La base $\{\vec{v}_1, \vec{v}_2\}$ no es ortogonal. Para forzar a que lo sea, después de tomar $\vec{u}_1 = \vec{v}_1$ solo hay que dar un paso de Gram-Schmidt:

$$\vec{u}_2 = \vec{v}_2 - \frac{\langle \vec{v}_2, \vec{u}_1 \rangle}{\|\vec{u}_1\|^2} \vec{u}_1 = \begin{pmatrix} 5 \\ 0 \\ 1 \end{pmatrix} - \frac{-10}{5} \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}.$$

A modo de comprobación, está claro que $\vec{u}_1 \cdot \vec{u}_2 = 0$ y $\vec{u}_2 \in V$. Una base ortonormal sería

$$\left\{ \frac{1}{\sqrt{5}}(-2, 1, 0)^t, \frac{1}{\sqrt{6}}(1, 2, 1)^t \right\}.$$

Para practicar con los números complejos, hallemos ahora una base ortogonal del subespacio $\mathcal{L}(\{\vec{v}_1, \vec{v}_2\}) \subset \mathbb{C}^3$ donde

$$\vec{v}_1 = (2 + 3i, 1 + 2i, 2)^t \quad \text{y} \quad \vec{v}_2 = (8 + 7i, -3, -6 + 2i)^t.$$

Como antes, tras tomar $\vec{u}_1 = \vec{v}_1$, todo se reduce al cálculo

$$\frac{\langle \vec{v}_2, \vec{u}_1 \rangle}{\|\vec{u}_1\|^2} \vec{u}_1 = \frac{(8 + 7i)(2 - 3i) + (-3)(1 - 2i) + (-6 + 2i)2}{|2 + 3i|^2 + |1 + 2i|^2 + |2|^2} = \frac{22}{22} = 1.$$

Con ello, una base ortonormal es

$$\vec{u}_1 = \vec{v}_1 = (2 + 3i, 1 + 2i, 2)^t \quad \text{y} \quad \vec{u}_2 = \vec{v}_2 - \vec{u}_1 = (6 + 4i, -4 - 2i, -8 + 2i)^t.$$

Si queremos comprobar el resultado, debemos verificar $\vec{u}_1 \cdot \vec{u}_2$ con

$$(2 + 3i)(6 - 4i) + (1 + 2i)(-4 + 2i) + 2(-8 - 2i) = (24 + 10i) + (-8 - 6i) + (-16 - 4i) = 0.$$

Finalmente, veamos un ejemplo con un producto escalar en un espacio de polinomios.

Consideramos la base $\{1, x, x^2\}$ de $\mathbb{R}_2[x]$ (los polinomios reales de grado a lo más dos) con el producto escalar $\langle P, Q \rangle = \int_{-1}^1 PQ$. Si $\{P_1, P_2, P_3\}$ es la base que se obtiene con el proceso de Gram-Schmidt, de acuerdo con las fórmulas (2) debemos tomar $P_1 = 1$ y

$$P_2 = x - \frac{\int_{-1}^1 1x \, dx}{\int_{-1}^1 1^2 \, dx} 1 = x, \quad P_3 = x^2 - \frac{\int_{-1}^1 1x^2 \, dx}{\int_{-1}^1 1^2 \, dx} 1 - \frac{\int_{-1}^1 x^3 \, dx}{\int_{-1}^1 x^2 \, dx} x = x^2 - \frac{1}{3}.$$

Si queremos una base ortonormal, tenemos que dividir entre la norma y tras algunos cálculos se obtiene

$$\left\{ \frac{1}{\sqrt{2}}, \sqrt{\frac{3}{2}} x, \sqrt{\frac{5}{8}}(3x^2 - 1) \right\}.$$

Aunque este ejemplo parezca muy abstracto, al generalizarlo a $\mathbb{R}_n[x]$ aparece en diversas aplicaciones, por ejemplo en la forma de los orbitales atómicos.

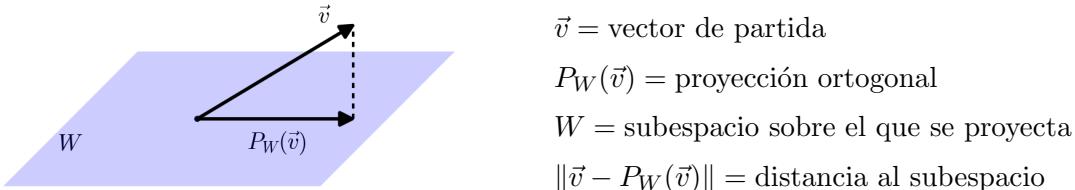
3. La proyección ortogonal

Ahora vamos a ver una de las construcciones más importantes del curso. Está basada en el siguiente resultado:

Sea V un espacio vectorial (de dimensión finita) con producto escalar y W un subespacio suyo. Para cada $\vec{v} \in V$ existe un único $\vec{w} \in W$ y un único \vec{u} ortogonal a todos los vectores de W tales que $\vec{v} = \vec{w} + \vec{u}$.

Al vector \vec{w} asociado a \vec{v} se le llama *proyección ortogonal* de \vec{v} sobre W y escribiremos $\vec{w} = P_W(\vec{v})$. Si a uno le gusta la teoría, no es difícil ver que $P_W : V \rightarrow W$ es una aplicación lineal sobreyectiva con núcleo los vectores ortogonales a todos los de W , que se suelen denotar con W^\perp . Tampoco es difícil entender, bajo el supuesto gusto por la teoría, que la existencia de la proyección ortogonal (la prueba del resultado anterior) es consecuencia de que con el proceso de Gram-Schmidt es posible extender una base ortogonal de W a una de V .

En términos visuales y geométricos, la proyección ortogonal es la sombra de un vector sobre una recta o un plano o sus generalizaciones n -dimensionales cuando los rayos de luz son perpendiculares a ellos. Hay otras proyecciones que sirven para representar sombras oblicuas, pero resultan matemáticamente menos importantes y no forman parte de este curso. La idea mental que debemos tener es algo del tipo:



Como se ha indicado, se dice que $d(\vec{v}, W) = \|\vec{v} - P_W(\vec{v})\|$ es la *distancia* del vector al subespacio. Realmente es una distancia en el sentido de que indica la longitud del camino más corto:

$$\|\vec{v} - P_W(\vec{v})\| \leq \|\vec{v} - \vec{w}\| \quad \text{para todo } \vec{w} \in W.$$

Es esto lo que hace que la proyección ortogonal sea tan importante, da la mejor aproximación de $\vec{v} \in V$ con un vector de W .

Un esquema común en las aplicaciones es que V es un espacio demasiado grande para nuestras capacidades técnicas (limitaciones de almacenamiento, funcionamiento en tiempo real, etc.) y pasamos a un espacio asequible más pequeño W . La proyección ortogonal minimiza la pérdida de precisión. Por ejemplo, en cálculo numérico un problema natural es aproximar funciones relativamente complicadas por polinomios y, una vez definido un producto escalar adecuado, la proyección ortogonal sirve para obtener buenas aproximaciones de manera eficiente. Algo similar aparece cuando uno quiere sintetizar lo más fielmente posible una señal a

partir de cierto subespacio de señales posibles que es capaz de generar el *hardware* y, con este fin, se proyecta ortogonalmente sobre dicho subespacio.

Vamos a ver dos métodos para el cálculo efectivo de la proyección ortogonal y otro más que solo funciona en situaciones muy especiales.

Método 1. Si $\mathcal{B} = \{\vec{u}_1, \dots, \vec{u}_n\}$ es una base ortogonal de W ,

$$P_W(\vec{v}) = \frac{\langle \vec{v}, \vec{u}_1 \rangle}{\|\vec{u}_1\|^2} \vec{u}_1 + \frac{\langle \vec{v}, \vec{u}_2 \rangle}{\|\vec{u}_2\|^2} \vec{u}_2 + \cdots + \frac{\langle \vec{v}, \vec{u}_n \rangle}{\|\vec{u}_n\|^2} \vec{u}_n.$$

Método 2. Si $\mathcal{B} = \{\vec{v}_1, \dots, \vec{v}_n\}$ es una base de W (no necesariamente ortogonal) entonces

$$P_W(\vec{v}) = \lambda_1 \vec{v}_1 + \lambda_2 \vec{v}_2 + \cdots + \lambda_n \vec{v}_n$$

donde $\lambda_1, \dots, \lambda_n$ vienen dados por la solución del sistema de ecuaciones lineales

$$\left\langle \vec{v} - \sum_{j=1}^n \lambda_j \vec{v}_j, \vec{v}_k \right\rangle = 0, \quad k = 1, 2, \dots, n.$$

A primera vista, puede sonar extraño que esto sea realmente un sistema. En los ejemplos siguientes quedará claro.

Veamos un primer ejemplo con el único propósito de practicar con ambos métodos, pues más adelante se introducirá una manera mucho más breve de resolverlo. Consideremos el problema de calcular la proyección ortogonal de $\vec{v} = (4, 3, -2)^t \in \mathbb{R}^3$ sobre

$$W = \{\vec{x} \in \mathbb{R}^3 : 3x_1 - x_2 - x_3 = 0\}$$

y hallar la distancia de \vec{v} a W (geométricamente, la distancia de un punto a un plano).

Una base de W se obtiene resolviendo el “sistema”, lo cual es trivial en este caso: $x_2 = \lambda$, $x_3 = \mu$, $x_1 = (\lambda + \mu)/3$. Así pues, $\vec{x} = \lambda(1/3, 1, 0)^t + \mu(1/3, 0, 1)^t$ y los vectores $(1/3, 1, 0)^t$ y $(1/3, 0, 1)^t$ forman una base. Para evitar fracciones (no es obligatorio) quitamos denominadores y escogemos en su lugar la base de W

$$\mathcal{B} = \{\vec{v}_1, \vec{v}_2\} = \{(1, 3, 0)^t, (1, 0, 3)^t\}.$$

Para el primer método debemos ortogonalizarla, con este fin, se aplica el proceso de Gram-Schmidt obteniendo $\vec{u}_1 = \vec{v}_1$ y

$$\vec{u}_2 = \vec{v}_2 - \frac{\vec{v}_2 \cdot \vec{u}_1}{\|\vec{u}_1\|^2} \vec{u}_1 = \vec{v}_2 - \frac{1}{10} \vec{u}_1 = \frac{3}{10} (3, -1, 10)^t.$$

El factor $3/10$ es irrelevante para la ortogonalidad, por tanto, se puede omitir pasando $\vec{u}_2 \mapsto \frac{10}{3}\vec{u}_2 = (3, -1, 10)^t$. Finalmente, aplicando la fórmula del primer método se llega a

$$P_W(\vec{v}) = \frac{13}{10} \begin{pmatrix} 1 \\ 3 \\ 0 \end{pmatrix} + \frac{-11}{110} \begin{pmatrix} 3 \\ -1 \\ 10 \end{pmatrix} = \begin{pmatrix} 1 \\ 4 \\ -1 \end{pmatrix}.$$

Por supuesto, el resultado debe estar en W y es fácil comprobar que es así.

Con el segundo método no necesitamos ortogonalizar la base $\mathcal{B} = \{\vec{v}_1, \vec{v}_2\}$. En la fórmula, $n = 2$ y operando (ley distributiva en el primer argumento) se obtiene

$$\begin{cases} \vec{v} \cdot \vec{v}_1 - \lambda_1 \vec{v}_1 \cdot \vec{v}_1 - \lambda_2 \vec{v}_2 \cdot \vec{v}_1 = 0 \\ \vec{v} \cdot \vec{v}_2 - \lambda_1 \vec{v}_1 \cdot \vec{v}_2 - \lambda_2 \vec{v}_2 \cdot \vec{v}_2 = 0 \end{cases}$$

Ahora está claro que esto es un sistema. Usando los datos,

$$\vec{v} \cdot \vec{v}_1 = 13, \quad \vec{v} \cdot \vec{v}_2 = -2, \quad \vec{v}_1 \cdot \vec{v}_1 = 10, \quad \vec{v}_1 \cdot \vec{v}_2 = \vec{v}_2 \cdot \vec{v}_1 = 1, \quad \vec{v}_2 \cdot \vec{v}_2 = 10$$

y el sistema es

$$\begin{cases} 10\lambda_1 + \lambda_2 = 13 \\ \lambda_1 + 10\lambda_2 = -2 \end{cases} \Rightarrow \lambda_1 = \frac{4}{3}, \quad \lambda_2 = -\frac{1}{3}.$$

Por tanto la proyección ortogonal es

$$P_W(\vec{v}) = \frac{4}{3} \begin{pmatrix} 1 \\ 3 \\ 0 \end{pmatrix} - \frac{1}{3} \begin{pmatrix} 1 \\ 4 \\ -1 \end{pmatrix} = \begin{pmatrix} 1 \\ 4 \\ -1 \end{pmatrix},$$

que coincide con el resultado anterior.

La distancia del sería:

$$d(\vec{v}, W) = \|\vec{v} - P_W(\vec{v})\| = \sqrt{3^2 + (-1)^2 + (-1)^2} = \sqrt{11}.$$

Supongamos en \mathbb{R}^n un subespacio definido por una sola ecuación

$$W = \{\vec{x} \in \mathbb{R}^n : a_1x_1 + a_2x_2 + \dots + a_nx_n = 0\} \quad \text{con } n > 2.$$

Se puede reescribir como $W = \{\vec{x} \in \mathbb{R}^n : \vec{a} \cdot \vec{x} = 0\}$, donde $\vec{a} = (a_1, \dots, a_n)^t$. Imaginemos que $\{\vec{u}_1, \dots, \vec{u}_{n-1}\}$ es una base ortogonal de W , entonces añadiendo \vec{a} lo será de \mathbb{R}^n y según lo visto en la sección anterior, para cualquier $\vec{v} \in \mathbb{R}^n$

$$\vec{v} = \frac{\vec{v} \cdot \vec{u}_1}{\|\vec{u}_1\|^2} \vec{u}_1 + \frac{\vec{v} \cdot \vec{u}_2}{\|\vec{u}_2\|^2} \vec{u}_2 + \dots + \frac{\vec{v} \cdot \vec{u}_{n-1}}{\|\vec{u}_{n-1}\|^2} \vec{u}_{n-1} + \frac{\vec{v} \cdot \vec{a}}{\|\vec{a}\|^2} \vec{a}.$$

Los primeros $n - 1$ sumando dan la proyección sobre W , según el primer método, y despejando se obtiene la fórmula sencilla:

$$P_W(\vec{v}) = \vec{v} - \frac{\vec{v} \cdot \vec{a}}{\|\vec{a}\|^2} \vec{a}.$$

No hay que olvidar que este “tercer método” solo sirve para subespacios como el indicado. Se puede extender su utilidad a \mathbb{C}^n escogiendo como \vec{a} el vector formado por los conjugados de los coeficientes. ¿Sabrías justificar por qué hay que conjugar?

Ahora podemos rehacer de manera más simple el ejemplo de proyectar $\vec{v} = (4, 3, -2)^t$ sobre $W = \{\vec{x} \in \mathbb{R}^3 : 3x_1 - x_2 - x_3 = 0\}$. Se tiene $\vec{a} = (3, -1, -1)^t$ y la fórmula da

$$P_W(\vec{v}) = \vec{v} - \frac{\vec{v} \cdot \vec{a}}{\|\vec{a}\|^2} \vec{a} = (4, 3, -1)^t - \frac{11}{11}(3, -1, -1)^t = (1, 4, -1)^t$$

que coincide con lo obtenido antes.

Consideremos $V = \mathbb{R}^4$ y un subespacio $W \subset V$ dado por

$$W = \{\vec{x} \in \mathbb{R}^4 : x_1 - 2x_2 + x_3 = x_1 - 3x_2 + x_3 + x_4 = 0\}.$$

Queremos calcular la proyección ortogonal de $\vec{v} = (4, 8, -4, 12)^t$ sobre W . Ahora no hay atajos, porque aunque los vectores $(1, -2, 1, 0)$ y $(1, -3, 1, 1)$ son ortogonales a los de W no es cierto que W sea el espacio ortogonal a un solo vector.

Para ambos métodos, calculamos una base de W resolviendo el sistema por eliminación de Gauss:

$$\begin{pmatrix} 1 & -2 & 1 & 0 \\ 1 & -3 & 1 & 1 \end{pmatrix} \xrightarrow[f_2 \leftrightarrow f_2 - f_1]{\quad} \begin{pmatrix} 1 & -2 & 1 & 0 \\ 0 & -1 & 0 & 1 \end{pmatrix}.$$

Con $x_3 = \lambda$, $x_4 = \mu$ se sigue $x_2 = \mu$ y $x_1 = 2\mu - \lambda$, por tanto

$$\vec{x} \in W \Leftrightarrow \vec{x} = \lambda \vec{v}_1 + \mu \vec{v}_2 \text{ donde } \vec{v}_1 = (-1, 0, 1, 0)^t, \vec{v}_2 = (2, 1, 0, 1)^t.$$

Se tiene entonces que $\mathcal{B} = \{\vec{v}_1, \vec{v}_2\}$ es base de W .

Para aplicar el primer método debemos hallar una base ortogonal de W . El proceso de Gram-Schmidt afirma que para ello basta reemplazar \vec{v}_2 por $\vec{v}_2 - \|\vec{v}_1\|^{-2}(\vec{v}_2 \cdot \vec{v}_1)\vec{v}_1$. De esta forma, la base ortogonal resultante es

$$\{\vec{u}_1, \vec{u}_2\} = \{(-1, 0, 1, 0)^t, (1, 1, 1, 1)^t\}.$$

Según la fórmula, la proyección ortogonal es

$$P_W(\vec{v}) = \frac{\vec{v} \cdot \vec{u}_1}{\|\vec{u}_1\|^2} \vec{u}_1 + \frac{\vec{v} \cdot \vec{u}_2}{\|\vec{u}_2\|^2} \vec{u}_2 = (9, 5, 1, 5)^t.$$

Si aplicamos en su lugar el segundo método, llegamos al sistema de dos ecuaciones con dos incógnitas:

$$\begin{cases} (\vec{v} - \lambda_1 \vec{v}_1 - \lambda_2 \vec{v}_2) \cdot \vec{v}_1 = 0, \\ (\vec{v} - \lambda_1 \vec{v}_1 - \lambda_2 \vec{v}_2) \cdot \vec{v}_2 = 0, \end{cases} \iff \begin{cases} -8 - 2\lambda_1 + 2\lambda_2 = 0, \\ 28 + 2\lambda_1 - 6\lambda_2 = 0. \end{cases}$$

La solución $\lambda_1 = 1, \lambda_2 = 5$ implica $P_W(\vec{v}) = \vec{v}_1 + 5\vec{v}_2 = (9, 5, 1, 5)^t$ que coincide con el resultado anterior.

Pasemos ahora a un ejemplo más abstracto en el espacio de polinomios $V = \mathbb{R}_2[x]$ dotado con el producto escalar dado por $\langle P, Q \rangle = \int_{-1}^1 PQ$. Queremos calcular la proyección ortogonal $P_W(8x^2)$ y la distancia $d(8x^2, W)$ donde W es el subespacio que tiene como base (no ortogonal) $\mathcal{B} = \{P_1, P_2\} = \{x+1, x^2+1\}$.

El segundo método requiere resolver el sistema:

$$\begin{cases} \langle 8x^2, P_1 \rangle - \lambda_1 \langle P_1, P_1 \rangle - \lambda_2 \langle P_1, P_2 \rangle = 0, \\ \langle 8x^2, P_2 \rangle - \lambda_1 \langle P_2, P_1 \rangle - \lambda_2 \langle P_2, P_2 \rangle = 0. \end{cases}$$

Los productos escalares de $8x^2$ con los P_j son:

$$\langle 8x^2, P_1 \rangle = 8 \int_{-1}^1 (x^3 + x^2) dx = \frac{16}{3} \quad \text{y} \quad \langle 8x^2, P_2 \rangle = 8 \int_{-1}^1 (x^4 + x^2) dx = \frac{16}{5} + \frac{16}{3} = \frac{16}{15}.$$

Mientras que los de los P_j con ellos mismos son:

$$\langle P_1, P_1 \rangle = \int_{-1}^1 (x+1)^2 dx = \frac{8}{3}, \quad \langle P_1, P_2 \rangle = \int_{-1}^1 (x+1)(x^2+1) dx = \int_{-1}^1 (x^2+1) dx = \frac{2}{3} + 2 = \frac{8}{3},$$

donde en la segunda integral se ha usado que x^n integra cero en $[-1, 1]$ si n es impar, y

$$\langle P_2, P_2 \rangle = \int_{-1}^1 (x^2 + 1)^2 dx = \int_{-1}^1 (x^4 + 2x^2 + 1) dx = \frac{2}{5} + \frac{4}{3} + 2 = \frac{56}{15}.$$

Así pues, numéricamente, el sistema original y el mismo tras ligeras simplificaciones son

$$\begin{cases} \frac{16}{3} - \frac{8}{3}\lambda_1 - \frac{8}{3}\lambda_2 = 0, \\ \frac{128}{15} - \frac{8}{3}\lambda_1 - \frac{56}{15}\lambda_2 = 0, \end{cases} \iff \begin{cases} \lambda_1 + \lambda_2 = 2, \\ 5\lambda_1 + 7\lambda_2 = 16. \end{cases}$$

La solución es $\lambda_1 = -1, \lambda_2 = 3$ y la proyección ortogonal resulta

$$P_W(8x^2) = -(x+1) + 3(x^2+1) = 3x^2 - x + 2.$$

Con esto, la distancia es

$$d(8x^2, W) = \|8x^2 - P_W(8x^2)\| = \|5x^2 + x - 2\| = \left(\int_{-1}^1 (5x^2 + x - 2)^2 dx \right)^{1/2} = \frac{4}{\sqrt{3}}.$$

Para no olvidarnos de los números complejos, terminaremos calculando la proyección ortogonal de $\vec{v} = (2+i, 1+7i)^t$ sobre la “recta compleja” dada por el subespacio

$$W = \{\vec{x} \in \mathbb{C}^2 : (1+i)x_1 - 3ix_2 = 0\}.$$

Tomando $x_2 = \lambda$ se deduce $x_1 = 3(1+i)\lambda/2$ por tanto $W = \mathcal{L}(\{\vec{u}\})$ con $\vec{u} = (\frac{3}{2} + \frac{3}{2}i, 1)^t$. Solo hay un vector y no es necesario ortogonalizar, basta aplicar la fórmula del primer método con $n = 1$. Los cálculos con todo detalle son:

$$P_W(\vec{v}) = \frac{(2+i)(\frac{3}{2} - \frac{3}{2}i) + (1+7i) \cdot 1}{|\frac{3}{2} + \frac{3}{2}i|^2 + 1^2} \begin{pmatrix} \frac{3}{2} + \frac{3}{2}i \\ 1 \end{pmatrix} = (1+i) \begin{pmatrix} \frac{3}{2} + \frac{3}{2}i \\ 1 \end{pmatrix} = \begin{pmatrix} 3i \\ 1+i \end{pmatrix}.$$

Ligeramente más sencillo desde el punto de vista computacional hubiera sido tomar $x_2 = 2\lambda$ para eliminar los denominadores. Otra posibilidad es $x_2 = (1+i)\lambda$ para no dividir por números complejos.

El atajo para subespacios definidos por una sola ecuación no es un realmente un atajo aquí, porque la fórmula del primer método no ha requerido cálculos adicionales, de ahí la limitación $n > 2$ que impusimos. Si uno se empeña en seguir este falso atajo, hay que recordar la conjugación en \vec{a} , esto es, $\vec{a} = (1-i, 3i)^t$. De este modo

$$P_W(\vec{v}) = \vec{v} - \frac{\vec{v} \cdot \vec{a}}{\|\vec{a}\|^2} \vec{a} = \begin{pmatrix} 2+i \\ 1+7i \end{pmatrix} - \frac{(2+i)(1+i) + (1+7i)(-3i)}{|1-i|^2 + |3i|^2} \begin{pmatrix} 1-i \\ 3i \end{pmatrix},$$

que, operando, da el mismo resultado que antes.

Apéndice. Hasta ahora hemos empleado varias fórmulas sin justificar. Estos párrafos opcionales muestran que su deducción es suficientemente simple como para que alguien con habilidad matemática las pueda obtener sobre la marcha, sin necesidad de memorizarlas. La conveniencia de utilizar el cerebro o bien para aprender las fórmulas o bien para entender el razonamiento depende de cada uno.

Coordenadas en una base ortogonal $\{\vec{u}_1, \dots, \vec{u}_n\}$. Por la definición de coordenadas de \vec{x} , estas son los x_j que cumplen $\vec{x} = x_1\vec{u}_1 + \dots + x_n\vec{u}_n$. Si multiplicamos escalarmente (a la derecha) por \vec{u}_j , al ser la base ortogonal, se obtiene $\langle \vec{x}, \vec{u}_j \rangle = 0 + x_j \langle \vec{u}_j, \vec{u}_j \rangle + \dots + 0$, de donde $x_j = \langle \vec{x}, \vec{u}_j \rangle / \|\vec{u}_j\|^2$.

Proyección ortogonal con $\{\vec{u}_1, \dots, \vec{u}_n\}$ base ortogonal de W (método 1). Por la definición de proyección ortogonal, $\vec{v} = P_W(\vec{v}) + \vec{u}$ con \vec{u} ortogonal a todos los vectores de W . Como $P_W(\vec{v}) \in W$, tendrá unas coordenadas en la base elegida. Entonces, es de la forma $x_1\vec{u}_1 + \dots + x_n\vec{u}_n$ y basta aplicar el apartado anterior.

Proyección ortogonal con $\{\vec{v}_1, \dots, \vec{v}_n\}$ base no ortogonal de W (método 2). Con la notación anterior, $\vec{v} = P_W(\vec{v}) + \vec{u}$ y $P_W(\vec{v}) \in W$, será de la forma $\lambda_1\vec{v}_1 + \dots + \lambda_n\vec{v}_n = \sum_{j=1}^n \lambda_j \vec{v}_j$. Como $\vec{u} = \vec{v} - P_W(\vec{v})$ es ortogonal a todos los vectores de W , en particular lo será a todos los \vec{v}_k y, anulando los productos escalares con ellos, se obtienen las ecuaciones del método.

Proceso de Gram-Schmidt. Considerando el subespacio $W_{j-1} = \mathcal{L}(\{\vec{u}_1, \dots, \vec{u}_{j-1}\})$ para $j > 1$, la definición de la proyección ortogonal asegura que $\vec{u}_j = \vec{v}_j - P_{W_{j-1}}(\vec{v}_j)$ es ortogonal a $\vec{u}_1, \dots, \vec{u}_{j-1}$ y basta aplicar la fórmula del método 1.

4. Matrices ortogonales y unitarias

Se dice que una matriz $A \in \mathcal{M}_n(\mathbb{C})$ es *unitaria* si cumple

$$\overline{A}^t A = I,$$

donde la barra indica conjugación. Si $A \in \mathcal{M}_n(\mathbb{R})$ se suele decir que es *ortogonal* (aunque estrictamente también es unitaria). Obviamente, en este caso la conjugación es superflua.

Es fácil ver que

$$A\overline{A}^t = I, \quad A^t\overline{A} = I, \quad \overline{A}A^t = I \quad \text{y} \quad A^{-1} = \overline{A}^t$$

son definiciones equivalentes a la original. Esta última caracterización muestra que obtener la inversa de una matriz unitaria no requiere ningún cálculo, algo que choca con la cantidad de operaciones que conlleva para matrices generales cuando la dimensión crece.

Una última equivalencia es que $A \in \mathcal{M}_n(\mathbb{C})$ es unitaria si y solo si sus columnas forman una base ortonormal de \mathbb{C}^n . Esto se deduce directamente de $A^t\overline{A} = I$ por la forma de multiplicar matrices y la definición del producto escalar usual en \mathbb{C}^n .

Está claro que los nombres “unitaria” y “ortogonal” aplicados a estas matrices no son muy afortunados, pero están demasiado asentados como para tomar iniciativas en su contra. Por lo que acabamos de ver, sería más lógico llamar a ambas “matrices ortonormales”.

Para terminar con las propiedades teóricas, a partir de unas matrices unitarias se pueden generar otras usando que $A, B \in \mathcal{M}_n(\mathbb{C})$ unitarias implica $AB, \overline{A}^t, \overline{B}^t$ unitarias.

Pasemos ahora a algunos ejemplos 2×2 de matrices ortogonales y unitarias:

$$\begin{pmatrix} 3/5 & 4/5 \\ -4/5 & 3/5 \end{pmatrix}, \quad \frac{1}{13} \begin{pmatrix} 5 & 12 \\ 12 & -5 \end{pmatrix}, \quad \frac{1}{2} \begin{pmatrix} \sqrt{2} + i & -i \\ i & -\sqrt{2} + i \end{pmatrix}, \quad \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}.$$

Las dos primeras son ortogonales si las consideramos en $\mathcal{M}_2(\mathbb{R})$ y unitarias si las consideramos en $\mathcal{M}_2(\mathbb{C})$. Las dos últimas son unitarias y no tiene sentido plantearse si son ortogonales porque no son matrices reales. El cálculo detallado de que la tercera matriz es unitaria sería:

$$\overline{A}^t A = \frac{1}{4} \begin{pmatrix} \sqrt{2} - i & -i \\ i & -\sqrt{2} - i \end{pmatrix} \begin{pmatrix} \sqrt{2} + i & -i \\ i & -\sqrt{2} + i \end{pmatrix} = \frac{1}{4} \begin{pmatrix} 3+1 & 0 \\ 0 & 3+1 \end{pmatrix} = I.$$

La propiedad de ser unitaria u ortogonal es tan singular que prácticamente cualquier matriz que nos inventemos al azar no lo será.

Con matrices diagonales es posible encontrar ejemplos sencillos en dimensión mayor, por ejemplo cualquier matriz diagonal A con $a_{ii} \in \{-1, 1\}$ será ortogonal y, en general, si los a_{ii} son números complejos de módulo uno, será unitaria. Dos ejemplos de dimensión 4 son

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad \text{y} \quad \begin{pmatrix} i & 0 & 0 & 0 \\ 0 & \frac{1+i}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}.$$

Por otro lado, ejemplos de matrices ortogonales no diagonales en dimensión 3 son:

$$\begin{pmatrix} 0 & -1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \quad \text{y} \quad \frac{1}{9} \begin{pmatrix} 1 & 8 & 4 \\ -4 & 4 & -7 \\ 8 & 1 & -4 \end{pmatrix}.$$

En cuanto uno tenga un mínima curiosidad matemática se preguntará a qué vienen estas definiciones. La gracia de las matrices unitarias y ortogonales es que preservan el producto escalar usual, es decir, que son las matrices que cumplen

$$A\vec{x} \cdot A\vec{y} = \vec{x} \cdot \vec{y}.$$

En particular, tomando $\vec{x} = \vec{y}$ se sigue $\|A\vec{x}\| = \|\vec{x}\|$, esto es, conservan las longitudes.

Geométricamente, para \mathbb{R}^n con $n = 2, 3$ los endomorfismos $f(\vec{x}) = A\vec{x}$ donde A es ortogonal dan todas las transformaciones que fijan el origen y preservan las distancias. Por tanto, estas matrices representan giros, simetrías o la composición de ambos.

Las matrices unitarias parecen ser objetos demasiado matemáticos, pero surgen en diferentes aplicaciones. Una de las más punteras es que el *modelo estándar* que clasifica las partículas elementales (como los *quarks*) utiliza el grupo de simetrías $SU(3) \times SU(2) \times U(1)$. Quizá hayas leído esta afirmación sin más explicaciones en libros de divulgación. Pues bien, $SU(N)$ significa las matrices unitarias de $\mathcal{M}_N(\mathbb{C})$ que tienen determinante 1 y $U(N)$ los números complejos de módulo 1, que se pueden identificar con las matrices unitarias 1×1 . En general, $U(N)$ es el nombre que recibe el conjunto de matrices unitarias de $\mathcal{M}_N(\mathbb{C})$.