Capítulo 4

El producto escalar

4.1. Definición y propiedades

Todos conocemos el *producto escalar* en $\mathbb{R}^n$ que hemos usado en cursos pasados con $n = 2$ y $n = 3$. Se extiende a $n > 3$ de la forma obvia:

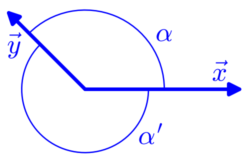
$$\vec{x} \cdot \vec{y} = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n \qquad \text{con} \quad \vec{x} = (x_1, x_2, \ldots, x_n)^t, \quad \vec{y} = (y_1, y_2, \ldots, y_n)^t.$$

Pensando los vectores como matrices columna, se tiene $\vec{x} \cdot \vec{y} = \vec{x}^t \vec{y}$. Otra notación medianamente común para el producto escalares $\langle \vec{x}, \vec{y} \rangle$, sobre todo para generalizaciones 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

$$(4.1) \qquad \|\vec{x}\| = \sqrt{\langle \vec{x}, \vec{x} \rangle} \qquad \text{y} \qquad \cos\alpha = \frac{\langle \vec{x}, \vec{y} \rangle}{\|\vec{x}\| \|\vec{y}\|} \quad \text{con } 0 \le \alpha \le \pi.$$

Geométricamente dos vectores en el plano dibujados como flechas determinan dos ángulos 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 $0 \le \alpha \le \pi$.



$$\alpha' = 2\pi - \alpha, \qquad \cos\alpha = \cos\alpha'.$$

En matemáticas medianamente avanzadas siempre se usan radianes, así que π son los 180° que la gente de a pie conoce (si le pides al pizzero 45° de *pizza* se reirá de ti pero seguramente te entenderá. Prueba sin embargo a pedirle $\pi/4$).

En cuanto uno tenga un poco de curiosidad científica se plantea por qué las fórmulas (4.1) están de acuerdo con nuestro concepto intuitivo de longitudes y ángulos al menos en el caso $n = 2$. Consideremos $\vec{x} = (x_1, x_2)^t$, $\vec{y} = (y_1, y_2)^t$. La primera fórmula es solo una manifestación del teorema de Pitágoras. Si dibujamos la flecha que representa $\vec{x}$, esta será la hipotenusa de un triángulo rectángulo con catetos de longitudes x_1 y x_2, salvo el signo:

$$\vec{x} = \sqrt{x_1^2 + x_2^2} = \sqrt{\vec{x} \cdot \vec{x}}.$$

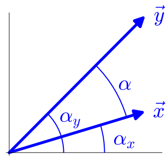
Para la segunda fórmula pensemos que los ángulos que van del eje OX a los vectores $\vec{x}$ e $\vec{y}$ son α_x y α_y, respectivamente. En ecuaciones, esto significa, empleando trigonometría elemental

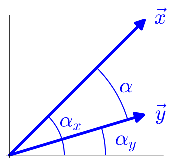
$$x_1 = \|\vec{x}\| \cos \alpha_x, \qquad x_2 = \|\vec{x}\| \operatorname{sen} \alpha_x, \qquad y_1 = \|\vec{y}\| \cos \alpha_y, \qquad y_2 = \|\vec{y}\| \operatorname{sen} \alpha_y.$$

De aquí

$$\frac{\vec{x} \cdot \vec{y}}{\|\vec{x}\|\|\vec{y}\|} = \cos \alpha_x \cos \alpha_y + \operatorname{sen} \alpha_x \operatorname{sen} \alpha_y = \cos(\alpha_x - \alpha_y)$$

donde se ha usado la fórmula de adición del coseno. La observación final es que el ángulo α entre $\vec{x}$ e $\vec{y}$ es $\pm(\alpha_x - \alpha_y)$ donde el signo no afecta al valor del coseno y depende de si tenemos que ir de $\vec{x}$ a $\vec{y}$ o viceversa.

$\alpha = -(\alpha_x - \alpha_y)$ 

$\alpha = \alpha_x - \alpha_y$ 

Veamos ahora algunos indicios de por qué es conveniente a veces considerar generalizaciones del producto escalar a espacios bien diferentes de $\mathbb{R}^n$ o definidos por fórmulas distintas de la habitual.

Consideremos una señal que dura un segundo representada por una función $f : [0,1] \longrightarrow \mathbb{R}$. En la práctica, sobre todo en nuestro mundo digital, la *muestreamos*, lo que significa que la sustituimos por el vector $\big(f(1/N), f(2/N), \dots, f(N/N)\big)^t \in \mathbb{R}^N$ correspondiente a medir la señal cada $1/N$ segundos con N grande (en señales de audio es habitual $N = 44100$). El producto escalar correspondiente a dos señales f y g sería $\sum_{k=1}^N f(k/N)g(k/N)$. Para que esto no resulte enorme se suele normalizar dividiendo entre N y en el límite $N \to \infty$, cuando el muestreo es infinitamente fino, se obtiene $\int_0^1 fg$. Por esta razón para señales que duran entre los tiempos $t = a$ y $t = b$ es natural definir los productos escalares

$$\langle f, g\rangle = \int_a^b fg \qquad \text{o} \qquad \langle f, g\rangle = \frac{1}{b-a}\int_a^b fg$$

que solo difieren en decidir si se normaliza también con la duración de la señal.

Por supuesto que uno puede definir lo que le venga en gana, otra cosa es que sirva para algo. Resulta que estos productos escalares sirven y mucho por varias razones, entre otras que en física e ingeniería $\langle f, f\rangle$ está relacionado con la energía. Por dar un ejemplo simple y un poco tonto (este no es un curso de teoría de señales), seguro que sabes que el voltaje de los enchufes de nuestra casa es actualmente[1] de 230 voltios y seguramente también sepas que es una onda sinusoidodal que oscila una vez cada $T = 1/50$ segundos, matemáticamente $f(t) = A\operatorname{sen}(2\pi t/T)$. Con esta información estarás tentado a creer que lo que sale de tu enchufe tiene $A = 230$. No es así, cuando se habla del voltaje de los enchufes lo que se da es $\|f\| = \sqrt{\langle f, f\rangle}$ con el último producto escalar. Es decir,

$$230 = \Big(\frac{1}{T}\int_0^T A^2 \operatorname{sen}^2(2\pi t/T)\, dt\Big)^{1/2} = \Big(\frac{1}{T}\cdot\frac{A^2T}{2}\Big)^{1/2} = \frac{1}{\sqrt{2}}A.$$

En resumidas cuentas, lo que sale del enchufe es $230\sqrt{2}\operatorname{sen}(2\pi t/T)$ que tiene picos de tensión de más de 300 voltios. ¿Por qué se habla entonces de 230? Porque esta es la tensión que se aprovecha para crear energía, la que correspondería a una corriente continua con igual rendimiento.

Incluso si no nos salimos de nuestro confortable $\mathbb{R}^2$ o $\mathbb{R}^3$ hay razones para considerar productos escalares diferentes del habitual. Supongamos que alguien viniera con la siguiente fórmula bajo el brazo:

$$\langle\vec{x}, \vec{y}\rangle = 2x_1y_1 + x_1y_2 + x_2y_1 + x_2y_2 \qquad \text{para} \quad \vec{x} = (x_1, x_2)^t,\ \vec{y} = (y_1, y_2)^t \in \mathbb{R}^2.$$

Pensaríamos que es muy raro porque sus longitudes y ángulos a través de (4.1) serían bien diferentes de las nuestras, por ejemplo $(1, 0)^t$ mediría $\sqrt{2}$ y el ángulo entre $(1, 0)^t$ y $(0, 1)^t$ sería $\alpha = \pi/4$. Sin embargo es posible mostrar que podría desarrollar con ello una geometría perfectamente coherente, aunque distinta de la habitual. La razón para esta inesperada coherencia, y lo que justifica que esta nueva fórmula para el producto escalar no sea tan loca, es que si empleamos la base $\{\vec{u}_1 = (1, 1)^t, \vec{u}_2 = (0, 1)^t\}$ entonces lo que llamamos los vectores de coordenadas $(x_1, x_2)^t$ e $(y_1, y_2)^t$ son $x_1\vec{u}_1 + x_2\vec{u}_2$ e $y_1\vec{u}_1 + y_2\vec{u}_2$ cuyo producto escalar usual es

$$\begin{aligned}(x_1\vec{u}_1 + x_2\vec{u}_2)\cdot(y_1\vec{u}_1 + y_2\vec{u}_2) &= x_1y_1\vec{u}_1\cdot\vec{u}_1 + x_1y_2\vec{u}_1\cdot\vec{u}_2 + x_2y_1\vec{u}_2\cdot\vec{u}_1 + x_2y_2\vec{u}_2\cdot\vec{u}_2,\\ &= 2x_1y_1 + x_1y_2 + x_2y_1 + x_2y_2.\end{aligned}$$

Es decir, que $\langle\vec{x}, \vec{y}\rangle$ con su extraña fórmula no es más que el producto escalar de toda la vida cuando usamos cierta base distinta de la canónica, un sistema de referencia deformado.

La existencia de productos escalares raros motivó en el desarrollo del álgebra lineal dar un 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.

[1] Incluso si eres más talludito que tus compañeros quizá no te percataste de que en 2003 se cambió oficialmente de 220 voltios a 230 para unificar con Europa. Si eres tan viejo como yo, sí recordarás los 125 voltios de la infancia y que en el cambio a los 220 el gobierno distribuyó transformadores en los hogares para que los electrodomésticos funcionaran.

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

1) Es lineal en cada argumento: Para todo $\vec{x},\vec{y},\vec{z}\in V$, $\lambda,\mu\in\mathbb{R}$

$$\langle\lambda\vec{x}+\mu\vec{y},\vec{z}\rangle = \lambda\langle\vec{x},\vec{z}\rangle+\mu\langle\vec{y},\vec{z}\rangle \qquad \text{y} \qquad \langle\vec{x},\lambda\vec{y}+\mu\vec{z}\rangle = \lambda\langle\vec{x},\vec{y}\rangle+\mu\langle\vec{x},\vec{z}\rangle.$$

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*.

La última propiedad exigida asegura que las longitudes definidas con la primera fórmula de (4.1) tienen sentido, no dan números imaginarios. Mostrar que los ángulos también lo tienen requiere mostrar que en la segunda fórmula el resultado está en $[-1,1]$. Este resultado no es de ningún modo inmediato y tiene nombre propio (normalmente dos y a veces tres).

Proposición 4.1.1 (Desigualdad de Cauchy-Schwarz)**.** *En cualquier espacio euclídeo V se cumple*

$$|\langle\vec{x},\vec{y}\rangle| \le \|\vec{x}\|\|\vec{y}\| \qquad \textit{para todo} \quad \vec{x},\vec{y}\in V.$$

Aquí las barras que rodean $\langle\vec{x},\vec{y}\rangle$ significan valor absoluto, es decir, que pongas siempre el signo positivo.

Demostración. Si $\vec{y}$ es un múltiplo de $\vec{x}$ entonces el resultado es evidente, incluso con igualdad. En otro caso, por la última propiedad $\langle\lambda\vec{x}-\vec{y},\lambda\vec{x}-\vec{y}\rangle > 0$. Usando la linealidad (la primera propiedad) esto es lo mismo que

$$0 < \lambda^2\langle\vec{x},\vec{x}\rangle - \lambda\langle\vec{x},\vec{y}\rangle - \lambda\langle\vec{y},\vec{x}\rangle + \langle\vec{y},\vec{y}\rangle = \lambda^2\|\vec{x}\|^2 - 2\lambda\langle\vec{x},\vec{y}\rangle + \|\vec{y}\|^2$$

donde se ha usado la definición de norma y la simetría en la igualdad. La última expresión es un polinomio cuadrático en λ que no tiene raíces reales (es positivo) por tanto su discriminante $b^2-4ac = 4\langle\vec{x},\vec{y}\rangle^2 - 4\|\vec{x}\|^2\|\vec{y}\|^2$ es negativo y esto equivale al resultado deseado. □

Estrechamente relacionada con esta desigualdad hay otra bien conocida acerca de longitudes.

Proposición 4.1.2 (Desigualdad triangular)**.** *En un espacio euclídeo V se cumple*

$$\|\vec{x}+\vec{y}\| \le \|\vec{x}\|+\|\vec{y}\| \qquad \textit{para todo} \quad \vec{x},\vec{y}\in V.$$

Demostración. Elevando al cuadrado y recordando la definición de la norma en (4.1), hay que probar

$$\langle\vec{x}+\vec{y},\vec{x}+\vec{y}\rangle \le \langle\vec{x},\vec{x}\rangle + 2\|\vec{x}\|\|\vec{y}\| + \langle\vec{y},\vec{y}\rangle.$$

Cuando expandimos el primer miembro por la linealidad y cancelamos términos iguales, esto equivale a $\langle\vec{x},\vec{y}\rangle+\langle\vec{y},\vec{x}\rangle \le 2\|\vec{x}\|\|\vec{y}\|$ que es consecuencia de la desigualdad de Cauchy-Schwarz. □

Del último resultado general que destacaremos es posible que te sorprenda el nombre, ciertamente el aludido se extrañará si en su metempsicosis ve esto.

Proposición 4.1.3 (Teorema de Pitágoras)**.** *En cualquier espacio euclídeo V se verifica*

$$\|\vec{x}+\vec{y}\|^2 = \|\vec{x}\|^2 + \|\vec{y}\|^2 \qquad \textit{para cualesquiera } \vec{x},\vec{y}\in V \textit{ con } \langle\vec{x},\vec{y}\rangle = 0.$$

El nombre proviene de que, según la segunda fórmula de (4.1), $\langle\vec{x},\vec{y}\rangle = 0$ corresponde a $\alpha = \pi/2$. Con ello, al poner $\vec{y}$ al final de $\vec{x}$ forman los catetos de un triángulo rectángulo cuya hipotenusa mide $\|\vec{x}+\vec{y}\|$.

Demostración. Utilizando, como antes, la linealidad del producto escalar

$$\langle\vec{x}+\vec{y},\vec{x}+\vec{y}\rangle = \langle\vec{x},\vec{x}\rangle + \langle\vec{x},\vec{y}\rangle + \langle\vec{y},\vec{x}\rangle + \langle\vec{y},\vec{y}\rangle$$

y la hipótesis y la simetría aseguran que los dos términos intermedios son nulos. □

Después de toda esta sección centrada en $\mathbb{R}$ cabe preguntarse qué ocurre en $\mathbb{C}$. Es muy posible que nunca te hayan definido en $\mathbb{C}^n$ ni siquiera el producto escalar "usual". La definición es muy similar a la de $\mathbb{R}^n$ con una pequeña salvedad. Para $\vec{x},\vec{y}\in\mathbb{C}^n$ se define[2]

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

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. 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

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

También hay una definición generalizada de producto escalar para espacios vectoriales sobre $\mathbb{C}$ igual a la de $\mathbb{R}$ excepto que la linealidad en el segundo argumento hay que modificarla a

$$\langle\vec{x},\lambda\vec{y}+\mu\vec{z}\rangle = \overline{\lambda}\langle\vec{x},\vec{y}\rangle + \overline{\mu}\langle\vec{x},\vec{z}\rangle$$

y la simetría a

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

Con ello la desigualdad de Cauchy-Schwarz, la triangular y el teorema de Pitágoras se siguen cumpliendo en este contexto, entendiendo que en la Proposición 4.1.1, $|\langle\vec{x},\vec{y}\rangle|$ significa el módulo de $\langle\vec{x},\vec{y}\rangle$. 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.

[2]Aquí pongo, con bastante desconfianza, la definición habitual en álgebra lineal que es la que aparece en fuentes bibliográficas principales del curso como [22] y [13], pero en física y en otras partes de las matemáticas las x_j y las y_j están intercambiadas de modo que la linealidad que falla es la del primer argumento.

Exprimiendo el silicio [opcional]. En `matlab/octave` el apóstrofo indica la operación † que coincide con la trasposición en el caso real, así que si `v` y `w` son vectores columna `v'*w` da el producto escalar de `v` por `w` en el caso real mientras que si queremos que esto funcione también en el caso complejo hay que cambiar el orden[3] escribiendo `w'*v` o conjugar el resultado anterior, lo cual se consigue mediante `conj`. Para comprobar los ejemplos anteriores en $\mathbb{C}^2$ podríamos escribir

```
% Definición de los vectores:
v1 = [2+i; 1+i];
v2 = [-1+i; 2-i];
v3 = [3; i];

% Productos escalares:
conj(v1'*v2)
conj(v1'*v3)
conj(v3'*v1)
conj(v2'*v3)
```

y si esto de conjugar nos parece enrevesado, sería equivalente

```
% Productos escalares:
v2'*v1
v3'*v1
v1'*v3
v3'*v2
```

Hay un comando `norm` que funciona correctamente con números complejos. Por ejemplo, `norm(v1)` mostraría una aproximación de $\sqrt{7}$.

En `sagemath` existe una instrucción `dot_product` que se limita a calcular $\sum x_k y_k$ por tanto funciona en $\mathbb{R}^n$. Para que sirva también en $\mathbb{C}^n$ hay que conjugar el segundo argumento con `conjugate`. Así el análogo del código anterior sería:

```
# Definición de los vectores:
v1 = vector(CC, [2+i, 1+i])
v2 = vector(CC, [-1+i, 2-i])
v3 = vector(CC, [3, i])

# Productos escalares:
print( v1.dot_product(v2.conjugate()) )
print( v1.dot_product(v3.conjugate()) )
print( v3.dot_product(v1.conjugate()) )
print( v2.dot_product(v3.conjugate()) )
```

Hay también un comando `norm` como en `matlab/octave`. El código anterior produce resultados con decimales. La razón para ello es que estamos trabajando en $\mathbb{C}$, indicado mediante `CC`. Una forma de conseguir resultados exactos forzando el cálculo simbólico es omitir `CC` (y la coma) o sustituirlo por `SR`. Otra forma más artificiosa es cambiar al cuerpo $K = \{a + bi \,:\, a, b \in \mathbb{Q}\}$ reemplazando las cuatro primeras líneas por

```
# Definición del nuevo cuerpo (redefine i)
K.<i> = QuadraticField(-1)
# Definición de los vectores:
v1 = vector(K, [2+i, 1+i])
v2 = vector(K, [-1+i, 2-i])
v3 = vector(K, [3, i])
```

[3] Una razón más para que se cambie el convenio en álgebra lineal poniendo el conjugado en el primer argumento.

4.2. Ortogonalidad

En el teorema de Pitágoras (Proposición 4.1.3) apareció la condición $\langle \vec{x}, \vec{y} \rangle = 0$ que en $\mathbb{R}^2$ o en $\mathbb{R}^3$ no es otra cosa que la perpendicularidad. Dentro de la jerga del álgebra lineal se prefiere el término *ortogonalidad*. Es decir, se dice 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* se dice que son *ortonormales*. Por extensión, diremos 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.

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 linealmente independientes de $\mathbb{R}^2$ 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 $\mathbb{R}^3$,

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

es base ortonormal. Son tres vectores linealmente independientes de $\mathbb{R}^3$ y 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 $\|\vec{u}_1\| = \|\vec{u}_2\| = \|\vec{u}_3\| = 1$.

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

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

que es 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.

En los ejemplos anteriores la comprobación de la independencia lineal es en realidad redundante porque se deduce de la ortogonalidad. A pesar de que este es un resultado muy fácil es suficientemente notable para destacarlo y completarlo con un cálculo de las coordenadas en una base. Es conveniente dar un vistazo a la demostración para convencerse de que no hay nada que memorizar.

Proposición 4.2.1. *Si $\vec{u}_1, \dots, \vec{u}_n$ son vectores ortogonales no nulos en un espacio vectorial sobre $\mathbb{R}$ o $\mathbb{C}$ con un producto escalar, entonces son linealmente independientes. Además si V es el subespacio generado por ellos, para cada $\vec{x} \in V$ se tiene*

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

Es decir, $\|\vec{u}_j\|^{-2}\langle \vec{x}, \vec{u}_j \rangle$ son las coordenadas de $\vec{x}$ en la base $\mathcal{B} = \{\vec{u}_1, \dots, \vec{u}_n\}$.

Demostración. Al calcular el producto escalar de $\vec{x} = x_1\vec{u}_1 + x_2\vec{u}_2 + \cdots + x_n\vec{u}_n$ con $\vec{u}_j$ se obtiene $\langle \vec{x}, \vec{u}_j\rangle = x_j\langle \vec{u}_j, \vec{u}_j\rangle$. Si $\vec{x} = \vec{0}$ se deduce $x_j = 0$ y con ello la independencia lineal y para $\vec{x}$ general la fórmula indicada. □

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 hallar x_1, x_2 y x_3 resolviendo $\vec{v} = x_1\vec{u}_1 + x_2\vec{u}_2 + x_3\vec{u}_3$, en virtud del resultado anterior tenemos directamente

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

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

La moraleja es que hallar coordenadas en bases ortogonales, como la canónica, es mucho más fácil que en el caso general: en vez de resolver sistemas de ecuaciones debemos calcular productos escalares. Esta ventaja motiva encontrar 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

$$\vec{u}_1 = \vec{v}_1 \qquad \text{y} \qquad \vec{u}_j = \vec{v}_j - \sum_{k=1}^{j-1} \frac{\langle \vec{v}_j, \vec{u}_k\rangle}{\|\vec{u}_k\|^2}\,\vec{u}_k, \qquad j = 2,3,\dots,n. \tag{4.2}$$

En la próxima sección veremos la interpretación de ello en términos de la proyección ortogonal. Comprendiendo esta interpretación no es necesario recordar estas fórmulas (al menos yo no lo hago).

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.

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 (4.2) 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

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

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.

Al hilo del ejemplo anterior, 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. Por ejemplo, consideremos

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

Su dimensión es 2 por el Corolario 2.2.3. Lo usual es que para hallar una base elijamos $x_2 = \lambda$, $x_3 = \mu$ de modo que

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

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 más 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

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

Finalmente veamos un ejemplo con un producto escalar en un espacio de polinomios. Aunque parezca muy abstracto, al generalizarlo a $\mathbb{R}_n[x]$ aparece en diversas aplicaciones, por ejemplo en la forma de los orbitales atómicos.

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 (4.2) 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

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

El cabo suelto teórico que queda es mostrar que el proceso de Gram-Schmidt realmente cumple su cometido.

Proposición 4.2.2. *Al aplicar el algoritmo* (4.2) *a una base* $\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$ *de un espacio vectorial* V*, con producto escalar, se obtiene siempre una base ortogonal* $\{\vec{u}_1, \vec{u}_2, \dots, \vec{u}_n\}$ *de* V*. En particular todos los espacios vectoriales de dimensión finita y con producto escalar admiten bases ortogonales.*

Demostración. Los vectores $\vec{u}_1$ y $\vec{u}_2 = \vec{v}_2 - \|\vec{u}_1\|^{-2}\langle \vec{v}_2, \vec{u}_1\rangle \vec{u}_1$ son ortogonales como muestra el cálculo de $\langle \vec{u}_2, \vec{u}_1\rangle$. En general, procedemos por inducción finita, si damos por supuesto que $\vec{u}_1, \dots, \vec{u}_{j-1}$ son ortogonales entre sí entonces $\vec{u}_j$ es ortogonal a cualquiera de ellos, digamos a $\vec{u}_\ell$, ya que

$$\langle \vec{u}_j, \vec{u}_\ell\rangle = \langle \vec{v}_j, \vec{u}_\ell\rangle - \sum_{k=1}^{j-1} \frac{\langle \vec{v}_j, \vec{u}_k\rangle}{\|\vec{u}_k\|^2} \langle \vec{u}_k, \vec{u}_\ell\rangle = \langle \vec{v}_j, \vec{u}_\ell\rangle - \frac{\langle \vec{v}_j, \vec{u}_\ell\rangle}{\|\vec{u}_\ell\|^2} \langle \vec{u}_\ell, \vec{u}_\ell\rangle = 0.$$

Por otro lado, $\vec{u}_1$ está en el subespacio generado por $\vec{v}_1$ ya que ambos vectores son idénticos y en general si $\vec{u}_1, \dots, \vec{u}_{j-1} \in \mathcal{L}\big(\{\vec{v}_1, \dots, \vec{v}_{j-1}\}\big)$ entonces necesariamente $\vec{u}_j \in \mathcal{L}\big(\{\vec{v}_1, \dots, \vec{v}_j\}\big)$, gracias a la fórmula que define el algoritmo. Además no es nulo porque $\vec{v}_j \notin \mathcal{L}\big(\{\vec{v}_1, \dots, \vec{v}_{j-1}\}\big)$ (por ser base). Con ello $\{\vec{u}_1, \vec{u}_2, \dots, \vec{u}_n\}$ genera un subespacio de V y como $n = \dim V$ y son vectores ortogonales no nulos debe ser base de V. □

Exprimiendo el silicio [opcional]**.** Aparentemente en `matlab/octave` no hay una instrucción directa que lleve a cabo el proceso de Gram-Schmidt aunque uno puede encontrar en la red muchas implementaciones con pequeños programas. Si nuestro propósito es ortonormalizar sin importar el método, el comando `orth` aplicado a una matriz ortonormaliza el subespacio generado por sus columnas. Es decir que sirve para hallar bases ortonormales en subespacios presentados como $W = \mathcal{L}\big(\{\vec{v}_1, \dots, \vec{v}_n\}\big)$ incluso si $\{\vec{v}_1, \dots, \vec{v}_n\}$ no es una base. Un ejemplo relacionado con un subespacio que hemos considerado antes es:

```
% Vectores de partida
v1 = [-2;1;0];
v2 = [5;0;1];
v3 = v1 + v2;
v4 = 2*v1 - 7*v2;

% Las columnas de A generan W
A = [v1,v2,v3,v4];

% Las columnas del resultado son
% base ortonormal de W
orth(A)
```

Por supuesto la tercera y la cuarta columnas de `A` son superfluas pues `v1` y `v2` ya generan el subespacio. La salida que se obtiene es:

0.982980	−0.020408
−0.048520	0.929684
0.177188	0.367792

que difiere de los vectores que habíamos obtenido con el proceso de Gram-Schmidt. El comando `orth` también funciona con números complejos.

En `sagemath` sí existe `gram_schmidt` que aplicado sobre una matriz ortogonaliza sus filas con el proceso de Gram-Schmidt. Entonces con el código

```
# Vectores
v1 = [-2,1,0]
v2 = [5,0,1]
```

```
A = matrix(QQ, [ v1, v2 ])

G, M = A.gram_schmidt()

# El resultado son las filas de G
print(G)

# La matriz M cumple A = M*G
```

obtenemos una matriz cuyas filas son $(-2,1,0)$ y $(1,2,1)$, el mismo resultado que habíamos conseguido nosotros. La matriz `M` almacena los coeficientes que aparecen en el proceso y, como se indica, cumple `A = M*G`. Si queremos una base ortonormal basta indicar `orthonormal=True` en el argumento de `gram_schmidt`. Si hacemos esto en el código anterior obtendremos un error porque no podemos trabajar ya de forma simbólica en $\mathbb{Q}$, representado por `QQ`, debemos reemplazar `QQ` por `RDF` que muestra cálculos con decimales. El comando `gram_schmidt` también funciona con números complejos, en este caso hay que usar `CDF` para los cálculos inexactos.

4.3. La proyección ortogonal

En vez de plantearse si dos vectores son ortogonales, cabe preguntarse si dos subespacios vectoriales lo son. En algunas aplicaciones esto indica que los subespacios no tienen absolutamente nada que ver[4]. En relación con esto, en un espacio vectorial V con producto escalar, dado un subespacio $W \subset V$ se define el *complemento ortogonal* de W como

$$W^{\perp} = \{\vec{v} \in V \,:\, \langle \vec{v}, \vec{w} \rangle = 0 \quad \text{para todo } \vec{w} \in W\}. \tag{4.3}$$

Esto es, $W^{\perp}$ está constituido por todos los vectores que son ortogonales a los vectores de W. Es muy fácil comprobar que es un subespacio vectorial de V. Un subespacio y su complemento ortogonal determinan una descomposición del espacio total

Proposición 4.3.1. *Sea V un espacio vectorial de dimensión finita con producto escalar y sea W un subespacio suyo. Entonces* $\dim W^{\perp} = \dim V - \dim W$ *y para cada $\vec{v} \in V$ existe un único $\vec{w} \in W$ y un único $\vec{u} \in W^{\perp}$ tales que $\vec{v} = \vec{w} + \vec{u}$.*

La situación descrita en la proposición se resume en la jerga del álgebra lineal diciendo que V es la *suma directa*[5] de W y $W^{\perp}$ y se escribe $V = W \oplus W^{\perp}$.

Al vector $\vec{w}$ asociado a $\vec{v}$ se le llama *proyección ortogonal* de $\vec{v}$ en W. En otros términos, la proyección ortogonal sobre W es la función $P_W : V \longrightarrow W$ tal que $P_W(\vec{v}) = \vec{w}$. No es difícil ver que P_W es una aplicación lineal. En términos ópticos y geométricos, la proyección ortogonal da 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 son matemáticamente menos importantes y no forman parte de este curso.

[4] Por ejemplo, en física cuántica significa que la probabilidad de transición es nula entre los estados correspondientes a cada uno de ellos

[5] En general se dice que V es suma directa de dos subespacios W_1 y W_2, y se escribe $V = W_1 \oplus W_2$, cuando $V = W_1 + W_2$ (cada vector de V es suma de vectores de los subespacios) y $W_1 \cap W_2 = \{\vec{0}\}$.

La demostración de la Proposición 4.3.1 da un método para calcular la proyección ortogonal aunque también es conveniente tener en cuenta la primera de las siguientes propiedades básicas:

Proposición 4.3.2. *Con la notación anterior se cumple*

$$1)\ \vec{v} = P_W(\vec{v}) + P_{W^\perp}(\vec{v}) \qquad y \qquad 2)\ \|\vec{v} - P_W(\vec{v})\| \le \|\vec{v} - \vec{w}\| \quad \text{para todo } \vec{w} \in W.$$

Demostración de la Proposición 4.3.1. Si tenemos una base $\{\vec{u}_1, \dots, \vec{u}_m\}$ de W siempre la podemos completar a una base $\mathcal{B} = \{\vec{u}_1, \dots, \vec{u}_m, \vec{u}_{m+1}, \dots, \vec{u}_n\}$ de V, simplemente añadiendo sucesivamente vectores linealmente independientes hasta que esto no sea posible. Aplicando el proceso de Gram-Schmidt, que deja los primeros vectores invariantes, podemos suponer que $\mathcal{B}$ es ortonormal. Automáticamente $\vec{u}_j \in W^\perp$ para $j > m$ y por ser una base ortonormal, recordando la Proposición 4.2.1, se tiene

$$\vec{v} = \big(\langle \vec{v}, \vec{u}_1\rangle \vec{u}_1 + \dots + \langle \vec{v}, \vec{u}_m\rangle \vec{u}_m\big) + \big(\langle \vec{v}, \vec{u}_{m+1}\rangle \vec{u}_{m+1} + \dots + \langle \vec{v}, \vec{u}_n\rangle \vec{u}_n\big).$$

El primer paréntesis es $\vec{w}$ y el segundo $\vec{u}$. Con nuestra notación $n = \dim V$, $m = \dim W$ y, considerando el caso $\vec{v} \in W^\perp$, deducimos que todos los vectores de $W^\perp$ son combinación lineal de $\vec{u}_{m+1}, \dots, \vec{u}_n$, que son $n-m$ vectores linealmente independientes, de donde $\dim W^\perp = \dim V - \dim W$.

La unicidad de $\vec{w}$ y $\vec{u}$ se sigue porque si $\vec{w}_1 + \vec{u}_1 = \vec{w}_2 + \vec{u}_2$ entonces $\vec{w}_1 - \vec{w}_2 = \vec{u}_2 - \vec{u}_1$ por tanto $\vec{w}_1 - \vec{w}_2$ estaría a la vez en W y en $W^\perp$ y el único vector ortogonal a sí mismo es el $\vec{0}$. □

Demostración de la Proposición 4.3.2. Por la simetría del producto escalar, los vectores de W están en el ortogonal de $W^\perp$. Eso significa que en la igualdad $\vec{v} = \vec{u} + \vec{w}$ se tiene $\vec{u} \in W^\perp$ y $\vec{w} \in (W^\perp)^\perp$, esto es, $P_{W^\perp}(\vec{v}) = \vec{u}$ y se sigue la primera propiedad.

Para la segunda, nótese que $\vec{x} = \vec{v} - P_W(\vec{v}) \in W^\perp$ e $\vec{y} = P_W(\vec{v}) - \vec{w} \in W$. El teorema de Pitágoras (Proposición 4.1.3) asegura $\|\vec{v} - \vec{w}\|^2 = \|\vec{x} + \vec{y}\|^2 \ge \|\vec{x}\|^2$. □

La primera propiedad es útil para los cálculos porque a veces es más fácil calcular la proyección ortogonal sobre $W^\perp$ que sobre W.

La segunda propiedad es útil en diversas aplicaciones porque dice que la proyección ortogonal es la mejor aproximación si no queremos salirnos de un subespacio. Por ejemplo en cálculo numérico es relevante aproximar funciones relativamente complicadas por elementos de subespacios de polinomios. En diferentes ámbitos del procesamiento de señales uno quiere sintetizar lo más fielmente posible una señal a partir de cierto subespacio de señales posibles que es capaz de generar el hardware. En cierta manera, la compresión que se produce en los ficheros JPEG se debe también a que la mayor parte de las imágenes admiten una aproximación muy buena por la proyección ortogonal en un espacio de dimensión comparativamente pequeña. Genéricamente se llama método de mínimos cuadrados a la aproximación de datos por medio de aplicaciones lineales utilizando la proyección ortogonal y la famosa *recta de regresión* es solo el ejemplo más sencillo de ello.

La *distancia* entre dos subconjuntos de $\mathbb{R}^n$ (o de $\mathbb{C}^n$) se define como la mínima distancia entre los puntos que las integran. Identificando las flechas que representan los vectores con los puntos de sus extremos, lo que está diciendo la segunda propiedad es que la distancia del punto que corresponde a $\vec{v}$ a los puntos del subespacio W es justamente $\|\vec{v} - P_W(\vec{v})\|$. De esta forma, se consiguen generalizar a dimensiones superiores las fórmulas de distancia de punto a recta y punto a plano estudiadas en cursos anteriores.

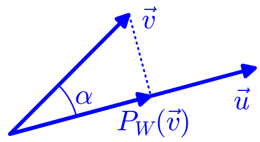
Si leemos la demostración de la Proposición 4.3.1 veremos que calcular $P_W(\vec{v})$ cuando tenemos una base ortonormal $\mathcal{B} = \{\vec{u}_1, \dots, \vec{u}_n\}$ es tan sencillo como calcular productos escalares. Si $\mathcal{B}$ es ortogonal, el análogo cambiando $\vec{u}_j$ por $\vec{u}_j/\|\vec{u}_j\|$ es

$$P_W(\vec{v}) = \frac{\langle \vec{v}, \vec{u}_1 \rangle}{\|\vec{u}_1\|^2}\vec{u}_1 + \frac{\langle \vec{v}, \vec{u}_2 \rangle}{\|\vec{u}_2\|^2}\vec{u}_2 + \cdots + \frac{\langle \vec{v}, \vec{u}_n \rangle}{\|\vec{u}_n\|^2}\vec{u}_n. \tag{4.4}$$

Esto no es otra cosa que hallar las coordenadas de $\vec{w}$ en $\mathcal{B}$ usando la Proposición 4.2.1, lo que se reducía a tomar productos escalares con $\vec{u}_j$. Por tanto es una fórmula que se puede improvisar y además explica por qué el proceso de Gram-Schmidt (4.2) funciona: en cada paso estamos calculando $\vec{u}_j = \vec{v}_j - P_{W_j}(\vec{v}_j)$ donde $W_j = \mathcal{L}(\{\vec{u}_1, \dots, \vec{u}_{j-1}\})$ y, como $\vec{u}_j \in W_j^\perp$, necesariamente es ortogonal a $\vec{u}_1, \dots, \vec{u}_{j-1}$.

Un caso particular es la proyección ortogonal sobre el espacio unidimensional $W = \mathcal{L}(\{\vec{u}\})$. Al tener un solo vector no hay que ortogonalizar nada. La fórmula resultante está de acuerdo con lo que se sigue en el plano usando simple trigonometría recordando (4.1).

$$P_W(\vec{v}) = \frac{\langle \vec{v}, \vec{u} \rangle}{\|\vec{u}\|^2}\vec{u}$$



$$P_W(\vec{v}) = \|\vec{v}\| \cos\alpha \frac{\vec{u}}{\|\vec{u}\|}.$$

Para ilustrar el uso de la primera propiedad de la Proposición 4.3.2 consideremos el problema de calcular la proyección ortogonal de $\vec{v} = (4, 3, -2)^t \in \mathbb{R}^3$ sobre

$$W = \{\vec{x} \in \mathbb{R}^3 \ : \ 3x - y - z = 0\}.$$

La manera "larga" de hacerlo sería buscar una base de W, que tendría dos vectores, y seguir el procedimiento de la prueba de la Proposición 4.3.1 o (4.4). Al ser $W^\perp$ un espacio unidimensional porque $\dim W^\perp = \dim V - \dim W = 3 - 2$, es más ventajoso calcular $P_{W^\perp}(\vec{v})$. Simplemente hay que fijarse en que $W = \{\vec{x} \in \mathbb{R}^3 \ : \ \vec{x} \cdot \vec{u}\}$ con $\vec{u} = (3, -1, -1)^t$. Así pues $\{\vec{u}\}$ es una base de $W^\perp$ y se tiene

$$P_{W^\perp}(\vec{v}) = \frac{\vec{v} \cdot \vec{u}}{\|\vec{u}\|}\vec{u} = (3, -1, -1)^t \quad \Rightarrow \quad P_W(\vec{v}) = \vec{v} - (3, -1, -1)^t = (1, 4, -1)^t$$

donde se ha usado la primera propiedad de la Proposición 4.3.2.

La distancia del punto $Q = (4, 3, -2)$ al plano determinado por W sería:

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

Si recuerdas la fórmula de la distancia de un punto a un plano, podrás corroborar este resultado.

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.

Si deseamos utilizar (4.4), lo que debemos hacer es calcular una base ortogonal de W. En primer lugar calculamos una base de W resolviendo el sistema por eliminación de Gauss:

$$\begin{pmatrix} 1 & -2 & 1 & 0 \\ 1 & -3 & 1 & 1 \end{pmatrix} \underset{f_2 \mapsto f_2 - f_1}{\longrightarrow} \begin{pmatrix} 1 & -2 & 1 & 0 \\ 0 & -1 & 0 & 1 \end{pmatrix}.$$

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

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

Se tiene entonces que $\{\vec{v}_1, \vec{v}_2\}$ es base de W. El proceso de Gram-Schmidt afirma que reemplazando $\vec{v}_2$ por $\vec{v}_2 - \|\vec{v}_1\|^{-2}(\vec{v}_2 \cdot \vec{v}_1)\vec{v}_1$, se convierte en una base ortogonal, concretamente en

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

Aplicando (4.4), la proyección es

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

Una alternativa que no requiere hallar una base ortogonal es usar que $\vec{v} - P_W(\vec{v})$ debe pertenecer a $W^\perp$ por tanto su producto escalar con $\vec{v}_1$ y $\vec{v}_2$ es nulo. Escribiendo $P_W(\vec{v}) = \lambda_1\vec{v}_1 + \lambda_2\vec{v}_2$, lo cual es lícito porque $P_W(\vec{v}) \in W$, llegamos a un sistema de dos ecuaciones con dos incógnitas:

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

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.

Supongamos que el ejemplo anterior nos pidieran una base de $W^\perp$. Esto es inmediato si reescribimos W como

$$W = \{\vec{x} \in \mathbb{R}^4 \ : \ \vec{u}_1 \cdot \vec{x} = \vec{u}_2 \cdot \vec{x} = 0\} \qquad \text{con} \quad \vec{u}_1 = (1, -2, 1, 0)^t, \ \vec{u}_2 = (1, -3, 1, 1)^t.$$

Esto dice que los vectores de W son justamente los que tienen a $\vec{u}_1$ y a $\vec{u}_2$ como ortogonales, por tanto estos generan $W^\perp$ y conforman una base suya ya que son linealmente independientes.

Veamos ahora un ejemplo más abstracto en nuestro espacio favorito de polinomios $V = \mathbb{R}_2[x]$ en el que definiremos el producto escalar extraño dado por

$$\langle P, Q\rangle = P(-1)Q(-1) + P(0)Q(0) + P(1)Q(1).$$

Siguiendo los requerimientos de un examen anterior, vamos a calcular la proyección ortogonal de x^2 sobre el subespacio

$$W = \mathcal{L}\big(\{P_1,\ P_2\}\big) \qquad \text{con} \quad P_1 = (x+1)^2, \quad P_2 = x^2+1.$$

Si $Q = \lambda_1 P_1 + \lambda_2 P_2 = P_W(x^2)$ es la proyección ortogonal buscada, entonces $x^2 - Q \in W^\perp$ o lo que es lo mismo $\langle P_1, x^2 - Q\rangle = 0$ y $\langle P_2, x^2 - Q\rangle = 0$. Unos cálculos sencillos muestran $\langle P_1, x^2\rangle = \langle P_2, x^2\rangle = 4$ con lo que las ecuaciones anteriores son:

$$\begin{cases} 4 - \lambda_1\langle P_1, P_1\rangle - \lambda_2\langle P_1, P_2\rangle = 0 \\ 4 - \lambda_1\langle P_2, P_1\rangle - \lambda_2\langle P_2, P_2\rangle = 0 \end{cases} \quad \Rightarrow \quad \begin{cases} 4 = 17\lambda_1 + 9\lambda_2 \\ 4 = 9\lambda_1 + 9\lambda_2 \end{cases}$$

donde se ha sustituido $\langle P_1, P_1\rangle = 17$ y $\langle P_1, P_2\rangle = \langle P_2, P_2\rangle = 9$. Resolviendo el sistema, $\lambda_1 = 0$ y $\lambda_2 = 4/9$ por tanto la proyección ortogonal es $Q = 4(x^2+1)/9$. Este sería el polinomio que mejor aproxima a x^2 dentro de nuestro subespacio.

En la línea de lo visto anteriormente, otra forma de llegar a la solución es ortogonalizar $\{P_1, P_2\}$. Aplicando el proceso de Gram-Schmidt y quitando denominadores, se obtiene la base ortogonal $\{Q_1, Q_2\}$ con $Q_1 = (x+1)^2$ y $Q_2 = 4x^2 - 9x + 4$ y con ello la proyección ortogonal es:

$$\frac{\langle x^2, Q_1\rangle}{\|Q_1\|^2}Q_1 + \frac{\langle x^2, Q_2\rangle}{\|Q_2\|^2}Q_2 = \frac{4Q_1}{0^2+1^2+4^2} + \frac{16Q_2}{17^2+4^2+(-1)^2} = \frac{4}{9}(x^2+1).$$

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

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

Tomando $x_2 = \lambda$ se deduce $x_1 = 3(1+i)\lambda/2$ por tanto $W = \mathcal{L}\big(\{\vec{u}\}\big)$ con $\vec{u} = \big(\frac{3}{2} + \frac{3}{2}i, 1\big)^t$. Basta entonces usar (4.4) para $n = 1$. Los cálculos con mucho detalle son:

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

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.

Es posible automatizar el cálculo de proyecciones ortogonales con una fórmula aunque esto no es tan conveniente en la práctica si no somos una máquina, porque las cuentas se vuelven opacas y porque no siempre estamos bajo las hipótesis de aplicarla directamente.

Proposición 4.3.3. *Dado un subespacio $W \subset K^n$ con $K = \mathbb{R}$ o $\mathbb{C}$, sea S una matriz cuyas columnas forman una base de W. Entonces*

$$P_W(\vec{v}) = S(S^\dagger S)^{-1} S^\dagger \vec{v} \qquad \textit{para todo} \quad \vec{v} \in K^n.$$

Si alguien está tentado a escribir $(S^\dagger S)^{-1} = S^{-1}(S^\dagger)^{-1}$ y simplificar, que se contenga porque S en general no es cuadrada y no tiene sentido invertirla. De hecho solo es cuadrada cuando $W = K^n$ en cuyo caso es evidente de partida que $P_W(\vec{v}) = \vec{v}$.

Demostración. Sea $\{\vec{v}_1, \dots, \vec{v}_n\}$ la base de W representada por las columnas de S. La proyección ortogonal es de la forma $\sum_j \lambda_j \vec{v}_j$ para ciertos λ_j que es nuestro objetivo calcular. En forma matricial tenemos

$$P_W(\vec{v}) = S\vec{\lambda} \qquad \text{con} \quad \vec{\lambda} = (\lambda_1, \dots, \lambda_n)^t.$$

Lo que caracteriza los λ_j es que $\vec{v} - P_W(\vec{v})$ debe ser ortogonal a todos los $\vec{v}_i$, es decir,

$$0 = \vec{v} \cdot \vec{v}_i - \sum_j \lambda_j \vec{v}_j \cdot \vec{v}_i = \vec{v}_i^\dagger \vec{v} - \sum_j \vec{v}_i^\dagger \vec{v}_j \lambda_j$$

donde en la última expresión hay que entender que los vectores se consideran como matrices columna. Por la forma en que se multiplican las matrices, $\vec{v}_i^\dagger \vec{v}$ es la i-ésima coordenada del vector $S^\dagger \vec{v}$ y $\vec{v}_i^\dagger \vec{v}_j$ es el elemento ij de $S^\dagger S$. Se concluye $S^\dagger \vec{v} = S^\dagger S \vec{\lambda}$ y de aquí $\vec{\lambda} = (S^\dagger S)^{-1} S^\dagger \vec{v}$. Necesariamente $S^\dagger S \in \mathcal{M}_n(K)$ es invertible porque si no tuviera rango máximo $\vec{v} = \vec{0}$, $\vec{\lambda} \neq \vec{0}$ sería una posibilidad absurda. $\square$

Una ventaja de la Proposición 4.3.3 es que produce explícitamente la matriz del endomorfismo proyección ortogonal. Veamos un ejemplo que podamos comprobar con cálculos anteriores.

Recordemos que habíamos hallado $P_W(\vec{v})$ para cierto $\vec{v}$ y un subespacio $W \subset \mathbb{R}^4$ con una base compuesta por $\vec{v}_1 = (-1, 0, 1, 0)^t$ y $\vec{v}_2 = (2, 1, 0, 1)^t$. Nos preguntamos cuál es la matriz de $P_W : \mathbb{R}^4 \longrightarrow \mathbb{R}^4$. A partir de las coordenadas de los vectores de la base

$$S = \begin{pmatrix} -1 & 2 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \Rightarrow \quad S^\dagger S = \begin{pmatrix} 2 & -2 \\ -2 & 6 \end{pmatrix} \quad \Rightarrow \quad (S^\dagger S)^{-1} = \frac{1}{4} \begin{pmatrix} 3 & 1 \\ 1 & 1 \end{pmatrix}.$$

Por tanto la matriz del endomorfismo que da la proyección ortogonal es:

$$\frac{1}{4} \begin{pmatrix} -1 & 2 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 3 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} -1 & 0 & 1 & 0 \\ 2 & 1 & 0 & 1 \end{pmatrix} = \frac{1}{4} \begin{pmatrix} 3 & 1 & -1 & 1 \\ 1 & 1 & 1 & 1 \\ -1 & 1 & 3 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}.$$

Si la aplicamos a $\vec{v} = (4, 8, -4, 12)^t$, obtendremos $P_W(\vec{v}) = (9, 5, 1, 5)^t$, como antes.

Exprimiendo el silicio [opcional]. El siguiente código implementa en `sagemath` la función `proj_Wv` que calcula la proyección ortogonal sobre un subespacio a partir de una base suya utilizando la fórmula anterior. Su uso está ilustrado con dos de los ejemplos anteriores.

```
def proj_Wv(S,v):
        G = S.conjugate_transpose()*S
        M = S*G.inverse()*S.conjugate_transpose()
        return M*v

# S = Base de W en columna
# v = vector a proyectar

S = matrix(4,2,[-1,2,0,1,1,0,0,1])
v = vector([4,8,-4,12])
print(proj_Wv(S,v))
print('----------------')

S = matrix(2,1,[3/2+3/2*I, 1])
v = vector([2+I, 1+7*I])
print(proj_Wv(S,v))
```

La salida es `(9, 5, 1, 5)` y `(3*I, I + 1)`, en consonancia con los resultados obtenidos.

Ya puestos a la solución automática de los ejemplos, la siguiente variante `proj_Wv2` (que llama a `proj_Wv`) permite introducir las ecuaciones del subespacio en vez de una base suya.

```
def proj_Wv2(A, v):
        W = A.right_kernel()
        S = matrix([item for item in W.basis()]).transpose()
        return proj_Wv(S,v)

# A = matrix con las ecuaciones de W
# v = vector a proyectar

A = matrix(QQ,2,4,[1,-2,1,0, 1,-3,1,1])
v = vector([4,8,-4,12])
print(proj_Wv2(A,v))
print('----------------')

A = matrix(1,2,[1+I, -3*I])
v = vector([2+I, 1+7*I])
print(proj_Wv2(A,v))
print('----------------')
```

4.4. Matrices ortogonales y unitarias

Una pregunta recurrente en matemáticas es el tipo de funciones que preservan ciertas cantidades o estructuras. En este sentido, a pesar de que no ha sido exactamente nuestro enfoque, las aplicaciones lineales son las funciones que preservan la estructura de espacio vectorial.

En esta línea e incluso dentro de la geometría básica, cabe preguntarse qué funciones preservan las distancias en el plano o el espacio. Claramente los giros lo hacen, la distancia de Barcelona a Madrid no cambia a lo largo del día con la rotación de la Tierra. Un prurito matemático lleva a plantearse una demostración que determine cuáles son exactamente. Aquí restringiremos nuestro análisis a las aplicaciones lineales, a cambio no nos quedaremos en el plano o el espacio sino que estudiaremos la situación en $\mathbb{R}^n$ y en $\mathbb{C}^n$ para cualquier dimensión.

Proposición 4.4.1. *Sea $f : K^n \longrightarrow K^n$ con $K = \mathbb{R}, \mathbb{C}$ una aplicación lineal que preserva las longitudes, es decir, tal que $\|f(\vec{x})\| = \|\vec{x}\|$ para todo $\vec{x} \in K^n$, entonces f también preserva el producto escalar, es decir $f(\vec{x}) \cdot f(\vec{y}) = \vec{x} \cdot \vec{y}$ para todo $\vec{x} \in K^n$.*

Demostración. Por hipótesis y por la linealidad de f

$$\|\vec{x} + \vec{y}\|^2 = \|f(\vec{x} + \vec{y})\|^2 = \|f(\vec{x}) + f(\vec{y})\|^2.$$

Utilizando que $\|\vec{z}\|^2 = \vec{z} \cdot \vec{z}$ al desarrollar se sigue

$$\|\vec{x}\|^2 + \|\vec{y}\|^2 + 2\vec{x} \cdot \vec{y} = \|f(\vec{x})\|^2 + \|f(\vec{y})\|^2 + 2f(\vec{x}) \cdot f(\vec{y}).$$

Los dos primeros sumandos de ambos miembros se cancelan porque f preserva las distancias y se obtiene el resultado buscado. □

En términos geométricos se puede interpretar este resultado a la luz de (4.1) diciendo que preservar longitudes implica preservar ángulos, algo que no es tan intuitivo.

Ya habíamos visto (Proposición 2.3.1) que en K^n todas las aplicaciones lineales venían dadas por multiplicar por una matriz. En nuestro caso, tratamos con endomorfismos y esta matriz es necesariamente cuadrada. El siguiente resultado resuelve totalmente el problema que nos habíamos planteado.

Proposición 4.4.2. *Un endomorfismo $f : K^n \longrightarrow K^n$ con $K = \mathbb{R}$ o $\mathbb{C}$ dado por $f(\vec{x}) = A\vec{x}$ preserva el producto escalar usual en K^n si y solo si*

$$1)\ A^tA = I \quad \textit{cuando } K = \mathbb{R}, \qquad 2)\ A^\dagger A = I \quad \textit{cuando } K = \mathbb{C}.$$

Multiplicando a la derecha por A^{-1} y a la derecha por A, estas condiciones equivalen a $AA^t = I$ y $AA^\dagger = I$. Ambas condiciones en realidad se pueden resumir en $A^\dagger A = I$ (o $AA^\dagger = I$) porque $A^\dagger = A^t$ cuando $K = \mathbb{R}$.

Demostración. Basta considerar el caso $K = \mathbb{C}$. Pensando los vectores como matrices columna, el producto escalar usual es $\vec{x} \cdot \vec{y} = \vec{x}^t\overline{\vec{y}}$ y se tiene $f(\vec{x}) \cdot f(\vec{y}) = \vec{x} \cdot \vec{y}$ si y solo si $\vec{x}^tA^t\overline{A}\,\overline{\vec{y}} = \vec{x}^tI\overline{\vec{y}}$. Al ser $\vec{x}$ e $\vec{y}$ arbitrarios esto equivale a $A^t\overline{A} = I$ y conjugando a $A^\dagger A = I$. □

Las matrices $A \in \mathcal{M}_n(\mathbb{R})$ que cumplen lo primero se llaman *matrices ortogonales* y las matrices $A \in \mathcal{M}_n(\mathbb{C})$ que cumplen lo segundo se llaman *matrices unitarias*. Se mire como se mire, los nombres no son muy afortunados pero están demasiado asentados como para tomar iniciativas en su contra. Quizá sería más adecuado llamar a ambas "matrices ortonormales" gracias al siguiente resultado que es una mera observación si uno tiene claro el procedimiento para multiplicar matrices.

Proposición 4.4.3. *Una matriz $A \in \mathcal{M}_n(K)$ es ortogonal (cuando $K = \mathbb{R}$) o unitaria (cuando $K = \mathbb{C}$) si y solo si sus columnas forman una base ortonormal de K^n.*

Demostración. De nuevo, basta considerar el caso $K = \mathbb{C}$, y las condiciones para ser unitaria son $\sum_k b_{ik}a_{kj} = \delta_{ij}$ con δ_{ij} los elementos de la matriz identidad y $b_{ik} = \overline{a}_{ki}$. La igualdad con el sumatorio significa que los vectores $(a_{ki})_{k=1}^n$ y $(a_{kj})_{k=1}^n$ son ortogonales para $i \neq j$ y con $i = j$ tenemos un vector unitario. Estos vectores son las columnas i-ésima y j-ésima de A. $\square$

Antes de seguir, aquí van algunos ejemplos 2×2 de matrices ortogonales y unitarias:

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

Las dos primeras son ortogonales si las consideramos en $A \in \mathcal{M}_n(\mathbb{R})$ y unitarias si las consideramos en $A \in \mathcal{M}_n(\mathbb{C})$. 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:

$$A^\dagger A = \frac{1}{4}\begin{pmatrix} \sqrt{2}-i & -i \\ i & -\sqrt{2}-i \end{pmatrix}\begin{pmatrix} \sqrt{2}+i & -i \\ i & -\sqrt{2}+i \end{pmatrix} = \frac{1}{4}\begin{pmatrix} 3+1 & 0 \\ 0 & 3+1 \end{pmatrix} = I_2.$$

Gracias a la Proposición 4.4.3, con cualquiera de las bases ortonormales de ejemplos anteriores obtendremos matrices ortogonales y unitarias. Recuperando dos de ellos, de las siguientes matrices la primera es ortogonal ($K = \mathbb{R}$) y unitaria ($K = \mathbb{C}$) y la segunda es unitaria.

$$\frac{1}{9}\begin{pmatrix} 8 & 4 & 1 \\ 4 & -7 & -4 \\ 1 & -4 & 8 \end{pmatrix}, \qquad \frac{1}{9}\begin{pmatrix} 4+3i & 4 & -2+6i \\ 4 & 4-3i & -2-6i \\ -2+6i & -2-6i & 1 \end{pmatrix}$$

Ambas son además simétricas pero eso es casualidad[6], rompemos tal simetría intercambiando dos filas sin violar la propiedad de ser ortogonal o unitaria.

No es difícil parametrizar todas las matrices 2×2 ortogonales y unitarias. Veámoslo en el caso ortogonal con la excusa de entender sus significado geométrico.

Proposición 4.4.4. *Todas las matrices $A \in \mathcal{M}_2(\mathbb{R})$ ortogonales son de la forma*

$$A = \begin{pmatrix} \cos\alpha & -\operatorname{sen}\alpha \\ \operatorname{sen}\alpha & \cos\alpha \end{pmatrix} \qquad \text{o de la forma} \quad A = \begin{pmatrix} \cos\alpha & \operatorname{sen}\alpha \\ \operatorname{sen}\alpha & -\cos\alpha \end{pmatrix}.$$

En el primer caso la aplicación lineal $f(\vec{x}) = A\vec{x}$ corresponde a un giro de ángulo α alrededor del origen y en el segundo a una simetría por la recta que forma un ángulo $\alpha/2$ con el eje X.

[6] En realidad no tanto: hay un truco para inventarse rápidamente bases ortonormales con números exactos y siempre da matrices simétricas. Este truco consiste en tomar un vector $\vec{u}$ con $\|\vec{u}\|^2 = 2$ y considerar las columnas de la llamada *matriz de Householder* $I - \vec{u}\vec{u}^t$.

Demostración. Por la Proposición 4.4.3, las columnas de A son vectores unitarios así que existen α y β tales que $a_{11} = \cos\alpha$, $a_{21} = \operatorname{sen}\alpha$ y $a_{12} = \cos\beta$, $a_{22} = \operatorname{sen}\beta$. La ortogonalidad de ambas columnas implica

$$0 = \cos\alpha\cos\beta + \operatorname{sen}\alpha\operatorname{sen}\beta = \cos(\alpha - \beta).$$

Entonces salvo múltiplos enteros de 2π se tiene $\alpha - \beta = -\pi/2$ o $\alpha - \beta = \pi/2$. Despejando β se obtienen ambos tipos de matrices.

Para estudiar el significado geométrico de f analizamos su efecto sobre una base. Para el primer tipo, notamos que $f(\vec{e}_1)$ y $f(\vec{e}_2)$ son los vectores de la base canónica $\{\vec{e}_1, \vec{e}_2\}$ girados un ángulo α. Por otro lado, usando la base ortonormal $\{\vec{u}_1, \vec{u}_2\}$ con $\vec{u}_1 = \big(\cos(\alpha/2), \operatorname{sen}(\alpha/2)\big)^t$, que es vector director de la recta del enunciado, y $\vec{u}_2 = \big(\operatorname{sen}(\alpha/2), -\cos(\alpha/2)\big)^t$, que es vector normal, se tiene $f(\vec{u}_1) = \vec{u}_1$ y $f(\vec{u}_2) = -\vec{u}_2$. Así pues actúa como la simetría mencionada sobre esta base. □

La segunda parte de la proposición anterior es un primer paso para resolver el problema planteado al inicio de la sección de determinar todas las funciones que preservan las distancias en el plano y en el espacio. Estas funciones se llaman *movimientos*. En el plano están las traslaciones, los giros por un punto (no necesariamente el origen), las simetrías y las simetrías deslizantes (aplicar una simetría y después trasladar la imagen en la dirección de su eje). En el espacio es un poco más complicado [13] pero en definitiva todo lo que hay son giros, simetrías, traslaciones y combinaciones de ellos.

Exprimiendo el silicio [opcional]. Gracias a la Proposición 4.4.3 las matrices ortogonales y unitarias están estrechamente ligadas a las bases ortonormales. El siguiente código muestra cómo generar matrices ortogonales y unitarias con `matlab/octave` ortonormalizando las columnas de matrices aleatorias reales o complejas mediante `orth`.

```
N = 3
% Matriz aleatoria NxN
A = rand(N);

% Matriz ortogonal
M = orth(A)

% Matriz aleatoria compleja
A = A + i*rand(N);

% Matriz unitaria
U = orth(A)
```

Se cumple que `M*M'` y `U*U'` son aproximaciones de la identidad.

Las cinco primeras líneas del siguiente código `sagemath` crean una matriz ortogonal partiendo de un vector arbitrario $\vec{v}$, usando el truco de las *matrices de Householder* que consiste en que si $\|\vec{u}\|^2 = 2$ entonces $I - \vec{u}\vec{u}^t$ es ortogonal y simétrica. La línea 4 toma $\vec{u} = \sqrt{2}\vec{v}/\|\vec{v}\|$ de una manera enrevesada para forzar el cálculo simbólico. Con el vector de partida indicado, se obtiene uno de los ejemplos que habíamos dado.

```
1 # Vector arbitrario
2 v = matrix([1,-4,-1])
```

```
3  N = v.ncols()
4  v *= sqrt( 2/(v*v.transpose())[0][0] )
5  H = identity_matrix(N)-v.transpose()*v
6
7  # Matriz ortogonal (de Householder) correspondiente
8  print(H)
9  print('-------------------')
10
11 # Cambios arbitrarios de filas
12 for k in range(N):
13         H.swap_rows(randint(0,N-1),randint(0,N-1))
14 print(H)
15 print('-------------------')
16
17 # ortogonal*unitaria*ortogonal = unitaria
18 D = diagonal_matrix( [I^randint(0,3) for _ in range(N)])
19 U = H*D*H
20 print(U)
```

El resto del código permuta algunas filas para romper la simetría (líneas 12–13) y genera una matriz unitaria usando que el producto de matrices unitarias es unitaria y que una ortogonal es unitaria. Concretamente se genera una matriz diagonal unitaria aleatoria muy sencilla y se multiplica por delante y por detrás por la matriz ortogonal antes hallada (líneas 18–19).

Si efectuamos `H*H.transpose()` y `U*U.conjugate_transpose()` obtendremos matrices identidad.