Espacios euclídeos y unitarios

Fernando Chamizo

9 de febrero de 2019

1. Formas bilineales y sesquilineales

En lo sucesivo V será un espacio vectorial sobre un cuerpo $\mathbb{K}$ con $\mathbb{K}=\mathbb{R}$ o $\mathbb{K}=\mathbb{C}$. Como es habitual, cuando no se diga lo contrario las coordenadas se suponen respecto a la base canónica y a veces se identifica un vector con sus coordenadas.

Una *forma bilineal* es una aplicación $\varphi : V\times V\longrightarrow \mathbb{K}$ que es lineal en ambos argumentos. Es decir,

$$(1)\qquad \begin{cases}\varphi(\lambda\vec{u}+\mu\vec{v},\vec{w})=\lambda\varphi(\vec{u},\vec{w})+\mu\varphi(\vec{v},\vec{w})\\ \varphi(\vec{u},\lambda\vec{v}+\mu\vec{w})=\lambda\varphi(\vec{u},\vec{v})+\mu\varphi(\vec{u},\vec{w})\end{cases}\qquad \text{para todo } \vec{u},\vec{v},\vec{w}\in V,\quad \lambda,\mu\in\mathbb{K}.$$

<u>Observación</u>. Para nuestros propósitos posteriores el caso más relevante es $\mathbb{K}=\mathbb{R}$ aunque nada impide considerar $\mathbb{C}$ o incluso otros cuerpos (conjuntos donde se puede sumar, restar, multiplicar y dividir salvo por cero).

Ejemplo. Para $V=\mathbb{R}^2$, si $\vec{x}=(x_1,x_2)^t$, $\vec{y}=(y_1,y_2)^t$ y definimos $\varphi(\vec{x},\vec{y})$ como el determinante $\det(\vec{x},\vec{y})=x_1y_2-x_2y_1$ entonces es una forma bilineal por las propiedades de los determinantes o un cálculo directo.

Si $\dim V=n<\infty$ entonces, fijada una base, es posible representar cada forma bilineal con una matriz. Si $\vec{x}=\sum_i x_i\vec{e}_i$, $\vec{y}=\sum_j y_j\vec{e}_j$ con $\mathcal{B}=\{\vec{e}_1,\dots,\vec{e}_n\}$ una base de V, para $\varphi : V\times V\longrightarrow\mathbb{K}$ forma bilineal se tiene

$$(2)\qquad \varphi(\vec{x},\vec{y})=\sum_i x_i\varphi\big(\vec{e}_i,\sum_j y_j\vec{e}_j\big)=\sum_i\sum_j x_iy_j\varphi\big(\vec{e}_i,\vec{e}_j\big).$$

Definiendo la matriz cuadrada $A=(a_{ij})_{i,j=1}^n$ con $a_{ij}=\varphi\big(\vec{e}_i,\vec{e}_j\big)$, se puede escribir

$$(3)\qquad \varphi(\vec{x},\vec{y})=(x_1,\dots,x_n)A\begin{pmatrix}y_1\\ \vdots\\ y_n\end{pmatrix}=\vec{x}^tA\vec{y}$$

donde en la última igualdad se han identificado los vectores y sus coordenadas en la base $\mathcal{B}$. Se dice que A construida de esta manera es la *matriz de φ en la base $\mathcal{B}$*. Evidentemente dada una matriz $n\times n$ la fórmula $\vec{x}^t A\vec{y}$ define una forma bilineal, así que fijada una base las matrices $n\times n$ y las aplicaciones bilineales están en correspondencia uno a uno. Simplemente el coeficiente de $x_i y_j$ es a_{ij}.

Ejemplo. Para φ como en el ejemplo anterior y $\mathcal{B} = \{\vec{e}_1, \vec{e}_2\}$ la base canónica de $\mathbb{R}^2$, se tiene $\varphi(\vec{e}_1, \vec{e}_1) = \varphi(\vec{e}_2, \vec{e}_2) = 0$ y $\varphi(\vec{e}_1, \vec{e}_2) = -\varphi(\vec{e}_2, \vec{e}_1) = 1$, por tanto la matriz de φ (en la base canónica) es

$$(4)\qquad A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \qquad \text{que concuerda con} \quad (x_1, x_2)\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = x_1y_2 - x_2y_1.$$

Supongamos que cambiamos de una base $\mathcal{B}$ a una base $\mathcal{B}'$ de modo que la relación entre las antiguas coordenadas en términos de las nuevas sea:

$$(5)\qquad \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = C\begin{pmatrix} x'_1 \\ \vdots \\ x'_n \end{pmatrix} \qquad \text{(columnas de } C = \text{coordenadas de los vectores de } \mathcal{B}' \text{ en la base } \mathcal{B}).$$

Entonces sustituyendo en (3) se obtiene

$$(6)\qquad \varphi(\vec{x}, \vec{y}) = \left(C\begin{pmatrix} x'_1 \\ \vdots \\ x'_n \end{pmatrix}\right)^t AC\begin{pmatrix} y'_1 \\ \vdots \\ y'_n \end{pmatrix} = (x'_1, \dots, x'_n)C^tAC\begin{pmatrix} y'_1 \\ \vdots \\ y'_n \end{pmatrix}$$

y se deduce que la matriz de φ en la base $\mathcal{B}'$ es C^tAC. Esta es la fórmula de cambio de base para matrices de aplicaciones bilineales.

Ejemplo. Consideremos la base $\mathcal{B}' = \{\vec{u}_1, \vec{u}_2\}$ donde $\vec{u}_1 = (1, 2)^t$ y $\vec{u}_2 = (-1, 1)^t$ (coordenadas con respecto a la base canónica $\mathcal{B}$). Podemos calcular la matriz de la forma bilineal φ de los ejemplos anteriores en la base $\mathcal{B}'$ usando la fórmula:

$$(7)\qquad C^tAC = \begin{pmatrix} 1 & 2 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}\begin{pmatrix} 1 & -1 \\ 2 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 3 \\ -3 & 0 \end{pmatrix}.$$

En realidad es más rápido usar un cálculo directo:

$$(8)\qquad \varphi(\vec{u}_1, \vec{u}_1) = \varphi(\vec{u}_2, \vec{u}_2) = 0, \qquad \varphi(\vec{u}_1, \vec{u}_2) = -\varphi(\vec{u}_2, \vec{u}_1) = 3.$$

Se dice que una forma bilineal es *simétrica* si $\varphi(\vec{x}, \vec{y}) = \varphi(\vec{y}, \vec{x})$ y que es *antisimétrica* si $\varphi(\vec{x}, \vec{y}) = -\varphi(\vec{y}, \vec{x})$. Es fácil ver que (en el caso de dimensión finita) lo primero ocurre si y solo sí la matriz es simétrica $A = A^t$ y lo segundo si es antisimétrica $A = -A^t$.

En ciertas situaciones uno necesita cierta “positividad” al trabajar con $\mathbb{C}$ y eso sugiere definir una variante de las formas bilineales que no es exactamente lineal en una de las variables.

Se dice que una aplicación $\varphi : V \times V \longrightarrow \mathbb{C}$ (con V espacio vectorial sobre $\mathbb{C}$) es una *forma sesquilineal* si es lineal en la primera variable y antilineal en la segunda. Es decir,

$$(9)\qquad \begin{cases} \varphi(\lambda\vec{u}+\mu\vec{v},\vec{w}) = \lambda\varphi(\vec{u},\vec{w})+\mu\varphi(\vec{v},\vec{w}) \\ \varphi(\vec{u},\lambda\vec{v}+\mu\vec{w}) = \bar{\lambda}\varphi(\vec{u},\vec{v})+\bar{\mu}\varphi(\vec{u},\vec{w}) \end{cases} \qquad \text{para todo } \vec{u},\vec{v},\vec{w}\in V, \quad \lambda,\mu\in\mathbb{C};$$

donde $\bar{\lambda}$ y $\bar{\mu}$ son los conjugados de λ y μ.

Observación. En física casi siempre se consideran formas sobre $\mathbb{C}$ que son lineales en la segunda variable y antilineales en la primera. Es solo una cuestión de convenio. Aquí se sigue el habitual en álgebra lineal aun reconociendo que el empleado en física y otras áreas de las matemáticas tiene ventajas.

Todo lo visto anteriormente para las formas bilineales se cumple con las modificaciones obvias para las sesquilineales.

En el caso de dimensión finita, la matriz se define igualmente como $A = (a_{ij})_{i,j=1}^n$ con $a_{ij} = \varphi\big(\vec{e}_i,\vec{e}_j\big)$ y en este caso se tiene

$$(10)\qquad \varphi(\vec{x},\vec{y}) = (x_1,\dots,x_n)A\begin{pmatrix}\bar{y}_1\\ \vdots\\ \bar{y}_n\end{pmatrix} = \vec{x}^tA\bar{\vec{y}}$$

con lo que a_{ij} es el coeficiente de $x_i\bar{y}_j$. El cambio de base se reduce de nuevo a sustituir (5) en esta fórmula para obtener que la matriz de la forma sesquilineal φ en la base $\mathcal{B}'$ es $C^tA\overline{C}$.

Las simetrías de interés son algo distintas que las del caso bilineal porque las dos variables se comportan de manera intrínsecamente diferente. Se dice que una forma sesquilineal es *hermítica* si $\varphi(\vec{x},\vec{y}) = \overline{\varphi(\vec{y},\vec{x})}$ y que es *antihermítica* si $\varphi(\vec{x},\vec{y}) = -\overline{\varphi(\vec{y},\vec{x})}$. Es fácil ver que (en el caso de dimensión finita) lo primero ocurre si y solo si la matriz cumple $A = \overline{A}^t$ y lo segundo si y solo si $A = -\overline{A}^t$. Correspondientemente se habla de matrices *hermíticas* y *antihermíticas*. En física es habitual la notación $A^\dagger$ en lugar de $\overline{A}^t$.

Ejemplo. Para $\vec{x} = (x_1,x_2)^t\in\mathbb{C}^2$, $\vec{y} = (y_1,y_2)^t\in\mathbb{C}^2$ definamos

$$(11)\qquad \varphi(\vec{x},\vec{y}) = 2x_1\bar{y}_1 + ix_1\bar{y}_2 - ix_2\bar{y}_1 + x_2\bar{y}_2$$

que en forma matricial es

$$(12)\qquad \varphi(\vec{x},\vec{y}) = (x_1,x_2)A\begin{pmatrix}\bar{y}_1\\ \bar{y}_2\end{pmatrix} \qquad \text{con} \quad A = \begin{pmatrix}2 & i\\ -i & 1\end{pmatrix}.$$

Entonces es una forma sesquilineal de matriz A (en la base canónica). Se tiene $A = \overline{A}^t$, por tanto es hermítica. Esto también se puede comprobar con un cálculo directo ya que $\varphi(\vec{y},\vec{x}) = 2y_1\bar{x}_1 + iy_1\bar{x}_2 - iy_2\bar{x}_1 + y_2\bar{x}_2$ implica $\varphi(\vec{x},\vec{y}) = \overline{\varphi(\vec{y},\vec{x})}$.

2. Productos escalares

En $\mathbb{R}^2$ y $\mathbb{R}^3$ el producto escalar de toda la vida nos sirve para medir la longitud de un vector con $\|\vec{a}\| = \sqrt{\vec{a}\cdot\vec{a}}$ o el ángulo entre dos vectores no nulos con $\cos\alpha = \vec{a}\cdot\vec{b}/(\|\vec{a}\|\|\vec{b}\|)$. En diversas situaciones en física y matemáticas conviene considerar otras formas de medir longitudes y ángulos (por ejemplo la relatividad general postula que la gravitación se explica como una variación de un "producto escalar" natural) y extender estas nociones a otros espacios. En este sentido es interesante definir un análogo del producto escalar para espacios vectoriales sobre $\mathbb{C}$ (en física cuántica sirve para medir probabilidades).

En un espacio vectorial V sobre $\mathbb{R}$ se dice que $\varphi : V \times V \longrightarrow \mathbb{R}$ es un *producto escalar* si cumple las propiedades:

1. φ es una forma bilineal.
2. φ es simétrica.
3. $\varphi(\vec{x},\vec{x}) > 0$ para todo $\vec{x} \neq \vec{0}$ (es *definida positiva*).

Una vez que tenemos un producto escalar, se dice que V es un *espacio euclídeo*.

Observación. A menudo se escribe $\langle\vec{x},\vec{y}\rangle$ en vez de $\varphi(\vec{x},\vec{y})$ para indicar un producto escalar. En física se usa mucho la llamada notación *bra-ket* que en este caso sería $\langle\vec{x}|\vec{y}\rangle$ (aunque rara vez se escribirían las flechas de los vectores).

Ejemplo. Comprobemos que $\langle\vec{x},\vec{y}\rangle$ dado por la forma bilineal de matriz (en la base canónica)

$$A = \begin{pmatrix} 1 & -1 & -1 \\ -1 & 2 & 2 \\ -1 & 2 & 6 \end{pmatrix} \tag{13}$$

es un producto escalar. Por definición cumple la primera propiedad, la segunda se sigue de $A = A^t$ y la tercera es más difícil. Si escribimos $\vec{x} = (x,y,z)^t$ lo que tenemos que probar es que el polinomio de tres variables

$$\langle\vec{x},\vec{x}\rangle = x^2 + 2y^2 + 6z^2 - 2xy - 2xz + 4yz \tag{14}$$

es mayor que cero para $x, y, z \in \mathbb{R}$ excepto cuando $x = y = z = 0$. Una manera de verlo es agruparlo como $(x-y-z)^2 + (y+z)^2 + (2z)^2$.

Más adelante en el curso veremos un algoritmo para escribir como suma de cuadrados que nos permitirá decidir siempre la tercera propiedad a partir de la matriz. En realidad hay un método simple basado en determinantes que funciona bien en dimensiones bajas y que asumiremos por ahora sin demostración:

Se dice que una matriz simétrica $A \in \mathcal{M}_n(\mathbb{R})$ es *definida positiva* si es la matriz de una forma bilineal que es un producto escalar. El *criterio de Sylvester* afirma que esto ocurre si y solo si

$$
\Delta_k = \begin{vmatrix} a_{11} & \dots & a_{1k} \\ \vdots & & \vdots \\ a_{k1} & \dots & a_{kk} \end{vmatrix} > 0 \qquad \text{para todo } 1 \le k \le n. \tag{15}
$$

Ejemplo. Al aplicar este criterio a la matriz de (13), se deduce que $\langle \vec{x}, \vec{y} \rangle$ es un producto escalar de

$$
\Delta_1 = \begin{vmatrix} 1 \end{vmatrix} = 1 > 0, \qquad \Delta_2 = \begin{vmatrix} 1 & -1 \\ -1 & 2 \end{vmatrix} = 1 > 0 \quad \text{y} \quad \Delta_3 = \begin{vmatrix} 1 & -1 & -1 \\ -1 & 2 & 2 \\ -1 & 2 & 6 \end{vmatrix} = 4 > 0. \tag{16}
$$

Ejemplo. El tensor de Minkowski en dos dimensiones es la forma bilineal en $\mathbb{R}^2$ dada por $\varphi(\vec{x}, \vec{y}) = x_1 y_1 - x_2 y_2$ obviamente no es definida positiva (si uno se empeña en usar el criterio, $\Delta_2 = -1 < 0$). Sin embargo, al escribir $x_1 = ct$, $x_2 = x$ con c, t y x la velocidad de la luz, el incremento de tiempo y el incremento de espacio, respectivamente; de acuerdo con la relatividad especial solo los eventos con $\varphi(\vec{x}, \vec{x}) > 0$ pueden ser alcanzados por una partícula material desde el origen. Es decir, a efectos de la realidad física se comporta como un producto escalar aunque estrictamente no lo sea. A veces se dice que es *pseudo producto escalar*.

Una vez que tenemos un producto escalar en un espacio vectorial sobre $\mathbb{R}$, podemos definir la *norma* (*longitud*) y el *ángulo* entre dos vectores no nulos mediante fórmulas como las de $\mathbb{R}^2$ y $\mathbb{R}^3$:

$$
\|\vec{x}\| = \sqrt{\langle \vec{x}, \vec{x} \rangle} \qquad \text{y} \qquad \cos\alpha = \frac{\langle \vec{x}, \vec{y} \rangle}{\|\vec{x}\| \|\vec{y}\|} \quad \text{con } 0 \le \alpha \le \pi. \tag{17}
$$

Es obvio que la longitud es positiva excepto por $\|\vec{0}\| = 0$, concordando con la idea de longitud, pero no está en absoluto claro que el ángulo α esté bien definido, es decir que el valor asignado coseno sea un número en $[-1, 1]$. Este es el contenido de una de las desigualdades más empleadas en matemáticas.

Desigualdad de Cauchy-Schwarz. En cualquier espacio euclídeo V se cumple

$$
|\langle \vec{x}, \vec{y} \rangle| \le \|\vec{x}\| \|\vec{y}\| \qquad \text{para todo} \quad \vec{x}, \vec{y} \in V. \tag{18}
$$

Ademas la igualdad se da solo si uno de los vectores es múltiplo de otro (lo que incluye el caso en que alguno sea nulo).

Es instructivo ver la prueba habitual que es breve e ingeniosa. Es muy fácil comprobar que si uno es múltiplo de otro se tiene igualdad. Tenemos que demostrar que en el resto de los casos hay desigualdad estricta. Para cualquier $\lambda \in \mathbb{R}$ se tiene

$$
\langle \lambda\vec{x} - \vec{y}, \lambda\vec{x} - \vec{y} \rangle = \|\lambda\vec{x} - \vec{y}\|^2 > 0. \tag{19}
$$

Utilizando que el producto escalar es bilineal, lo anterior se escribe como

$$\lambda^2\|\vec{x}\|^2 - 2\lambda\langle\vec{x},\vec{y}\rangle + \|\vec{y}\|^2 > 0. \tag{20}$$

La parte derecha es un polinomio cuadrático en λ y estamos diciendo que no tiene raíces reales (porque es positivo) así que el discriminante de la ecuación cuadrática correspondiente es negativo y este es $b^2 - 4ac = 4\langle\vec{x},\vec{y}\rangle^2 - 4\|\vec{x}\|^2\|\vec{y}\|^2$ de donde se deduce (18).

Ejemplo. Las funciones continuas $f : [0,1] \longrightarrow \mathbb{R}$ con el producto escalar $\langle f, g\rangle = \int_0^1 fg$ son un espacio euclídeo y entonces la desigualdad de Cauchy-Schwarz implica

$$\Big(\int_0^1 fg\Big)^2 \le \int_0^1 f^2 \int_0^1 g^2. \tag{21}$$

Una de las consecuencias más importantes de la desigualdad de Cauchy-Schwarz, y comparable en importancia, es la llamada *desigualdad triangular*

$$\|\vec{x}+\vec{y}\| \le \|\vec{x}\| + \|\vec{y}\| \qquad \text{para todo} \quad \vec{x},\vec{y} \in V. \tag{22}$$

La prueba se reduce a elevar al cuadrado, de modo que el lado derecho pasa a ser $\|\vec{x}\|^2 + \|\vec{y}\|^2 + 2\|\vec{x}\|\|\vec{y}\|$ mientras que el lado izquierdo pasa a ser $\langle\vec{x}+\vec{y},\vec{x}+\vec{y}\rangle = \langle\vec{x},\vec{x}\rangle + \langle\vec{y},\vec{y}\rangle + 2\langle\vec{x},\vec{y}\rangle$, utilizando que es bilineal, y (18) da el resultado.

En espacio vectoriales sobre $\mathbb{C}$ un producto escalar que tome valores complejos impide una definición adecuada de ángulo pero como veremos más adelante, lo más importante respecto a los ángulos es salvar el concepto de perpendicularidad que corresponde al producto escalar nulo. Es posible lograr este objetivo preservando una definición con sentido de longitud si se consideran formas sesquilineales en lugar de bilineales. Concretamente:

En un espacio vectorial V sobre $\mathbb{C}$ se dice que $\varphi : V \times V \longrightarrow \mathbb{C}$ es un *producto hermítico* (o también un *producto escalar*) si cumple las propiedades:

1. φ es una forma sesquilineal.
2. φ es hermítica.
3. $\varphi(\vec{x},\vec{x}) > 0$ para todo $\vec{x} \neq \vec{0}$ (es *definida positiva*).

Cuando asociamos a V una φ con estas características decimos que se ha convertido en un *espacio hermítico* (o en un *espacio unitario*).

Observación. También en este caso se aplican los convenios de notación antes indicados. Además el criterio de Sylvester también es válido en este contexto (bajo la hipótesis de que la matriz de φ es hermítica) así como la desigualdad de Cauchy-Schwarz, donde en este caso $|\langle\vec{x},\vec{y}\rangle|$ significa el módulo como número complejo. Eligiendo un número complejo u con $|u| = 1$

siempre se puede conseguir $\langle u\vec{x},\vec{y}\rangle = |\langle \vec{x},\vec{y}\rangle|$ y por ello se puede aprovechar la demostración del caso real cambiando $\vec{x}$ por $u\vec{x}$.

Ejemplo. En $\mathbb{C}^n$ es obvio que $\langle \vec{x},\vec{y}\rangle = x_1\bar{y}_1 + x_2\bar{y}_2 + \dots x_n\bar{y}_n$ es un producto hermítico (escalar). Es el usual en $\mathbb{C}^n$.

Ejemplo. En $\mathbb{C}^2$ consideramos las formas $\varphi_1(\vec{x},\vec{y}) = 3x_1\bar{y}_1 + (1+i)x_1\bar{y}_2 + (1+i)x_2\bar{y}_1 + x_2\bar{y}_2$ y $\varphi_2(\vec{x},\vec{y}) = 3x_1\bar{y}_1 + (1+i)x_1\bar{y}_2 + (1-i)x_2\bar{y}_1 + x_2\bar{y}_2$. Sus matrices son respectivamente

$$(23) \qquad A_1 = \begin{pmatrix} 3 & 1+i \\ 1+i & 1 \end{pmatrix} \qquad \text{y} \qquad A_2 = \begin{pmatrix} 3 & 1+i \\ 1-i & 1 \end{pmatrix}.$$

La forma sesquilineal φ_1 no es hermítica porque $A_1 \neq \overline{A}_1^t$ mientras que φ_2 sí lo es porque $A_2 = \overline{A}_2^t$. Por otro lado, $3 > 0$ y $|A_2| = 1 > 0$ implican que φ_2 es definida positiva. Por tanto φ_1 no define un producto hermítico en $\mathbb{C}^2$ y φ_2 sí.

Observación. En un contexto matemático más amplio se habla de *norma* en un espacio vectorial para referirse a culquier manera de medir longitudes con las propiedades 1) $\|\vec{x}\| > 0$ si $\vec{x} \neq 0$, 2) $\|\lambda\vec{x}\| = |\lambda|\|\vec{x}\|$ y 3) $\|\vec{x}+\vec{y}\| \le \|\vec{x}\| + \|\vec{y}\|$. Ya hemos visto que en los espacios euclídeos y hermíticos siempre $\sqrt{\langle \vec{x},\vec{x}\rangle}$ es una norma pero en este sentido amplio es posible definir normas que no provienen de un producto escalar. Un ejemplo es $\|\vec{x}\| = \text{máx}(|x_1|,|x_2|)$ que satisface las propiedades en $\mathbb{R}^2$ y $\mathbb{C}^2$.

3. Ortogonalidad

Se dice que dos vectores son *ortogonales* si su producto escalar (o hermítico en el caso de espacios unitarios) es nulo.

Observación. La ortogonalidad es importante porque a través de las proyecciones ortogonales del próximo capítulo está relacionada a menudo con hallar distancias mínimas o, en general, con optimizar, por lo cual participa en muchas aplicaciones.

Con este concepto se generaliza uno de los teoremas más antiguos de las matemáticas:

Teorema de Pitágoras. En cualquier espacio euclídeo o hermítico V

$$(24) \qquad \|\vec{x}+\vec{y}\|^2 = \|\vec{x}\|^2 + \|\vec{y}\|^2 \qquad \text{para cualesquiera } \vec{x},\vec{y} \in V \text{ ortogonales.}$$

La prueba se reduce a la igualdad $\langle \vec{x}+\vec{y},\vec{x}+\vec{y}\rangle = \langle \vec{x},\vec{x}\rangle + \langle \vec{y},\vec{y}\rangle + \langle \vec{x},\vec{y}\rangle + \langle \vec{y},\vec{x}\rangle$.

Como es lógico, llamamos *base ortogonal* de un espacio vectorial euclídeo o hermítico a la compuesta por vectores ortogonales dos a dos. Dada una bases arbitraria de un espacio vectorial de dimensión finita, hallar las coordenadas de un vector requiere resolver un sistema lineal de ecuaciones pero cuando tenemos una base ortogonal $\mathcal{B}=\{\vec{u}_1,\dots,\vec{u}_n\}$ es mucho más sencillo ya que calculando el producto escalar con $\vec{u}_j$

$$\vec{x}=x_1\vec{u}_1+x_2\vec{u}_2+\dots+x_n\vec{u}_n \quad\Rightarrow\quad \langle\vec{x},\vec{u}_j\rangle=x_j\|\vec{u}_j\|^2 \quad\Rightarrow\quad x_j=\frac{\langle\vec{x},\vec{u}_j\rangle}{\|\vec{u}_j\|^2}. \tag{25}$$

Si los elementos de la base $\mathcal{B}$ no solo son ortogonales sino también son *vectores unitarios*, es decir, tienen norma 1, se dice que $\mathcal{B}$ es una *base ortonormal* y entonces $x_j=\langle\vec{x},\vec{u}_j\rangle$, así que las coordenadas coinciden con los productos escalares con elementos de la base.

El mismo argumento empleado en (25) muestra que los conjuntos de vectores ortogonales no nulos son siempre linealmente independientes.

Debido a las ventajas de las bases ortogonales y ortonormales, es conveniente tener un algoritmo que transforme una base cualquiera de un espacio de dimensión finita en una de ellas. Dada una base ortogonal es fácil convertirla en una ortonormal sin más que dividir cada vector por su norma para convertirlo en unitario, lo que se llama habitualmente *normalización*, así que todo el problema es obtener una base ortogonal.

El método más natural para resolver este problema es el llamado *proceso de Gram-Schmidt*. Dada una base arbitraria $\{\vec{u}_1,\vec{u}_2,\dots,\vec{u}_n\}$ produce una base ortogonal $\{\vec{v}_1,\vec{v}_2,\dots,\vec{v}_n\}$ por medio de las fórmulas

$$\vec{v}_1=\vec{u}_1 \qquad \text{y} \qquad \vec{v}_j=\vec{u}_j-\sum_{k=1}^{j-1}\frac{\langle\vec{u}_j,\vec{v}_k\rangle}{\|\vec{v}_k\|^2}\,\vec{v}_k, \qquad j=2,3,\dots,n. \tag{26}$$

Geométricamente este algoritmo corresponde a restar a cada $\vec{u}_j$ la proyección ortogonal (que estudiaremos más adelante) sobre el subespacio generado por $\{\vec{v}_1,\dots,\vec{v}_{j-1}\}$.

Una consecuencia de este algoritmo es que todo espacio vectorial de dimensión finita euclídeo o hermítico tiene bases ortonormales.

Ejemplo. Por ejemplo, la base $\{1,x,x^2\}$ de $\mathbb{R}_2[x]$ (polinomios reales de grado a lo más 2) no es ortonormal con el producto escalar $\langle P,Q\rangle=\int_{-1}^1 PQ$. Con el proceso de Gram-Schmidt conseguimos una base ortogonal $\{P_1,P_2,P_3\}$ tomando $P_1=1$ y

$$P_2=x-\frac{\int_{-1}^1 1x\,dx}{\int_{-1}^1 1^2\,dx}1=x, \quad P_3=x^2-\frac{\int_{-1}^1 1x^2\,dx}{\int_{-1}^1 1^2\,dx}1-\frac{\int_{-1}^1 x^3\,dx}{\int_{-1}^1 x^2\,dx}x=x^2-\frac{1}{3}. \tag{27}$$

Si queremos una base ortonormal, dividimos por la norma obteniéndose

$$\Big\{\frac{1}{\sqrt{2}},\sqrt{\frac{3}{2}}\,x,\sqrt{\frac{5}{8}}\big(3x^2-1\big)\Big\}. \tag{28}$$

Aunque este ejemplo parezca demasiado abstracto, los polinomios obtenidos al aplicar el proceso de Gram-Schmidt a $\{1,x,\dots,x^n\}$ con $\langle P,Q\rangle=\int_{-1}^1 PQ$ aparecen en varios temas de física, por ejemplo sirven para dar el llamado desarrollo multipolar del potencial gravitatorio y son necesarios para describir la forma de los orbitales del átomo de hidrógeno.

El proceso de Gram-Schmidt prueba que existen bases ortonormales de un espacio euclídeo o hermítico de dimensión finita. En una de tales bases $\mathcal{B} = \{\vec{u}_1, \dots \vec{u}_n\}$ la matriz de la forma correspondiente al producto escalar es la identidad. Esto implica que, en el caso de dimensión finita, con un cambio de base cualquier producto escalar se transforma en el usual. Esto no significa que la definición general de producto escalar sea superflua pues las coordenadas en una base pueden tener un significado físico o matemático especial que convenga preservar. Por dar una analogía, las elipses se reducen a circunferencias con cambios sencillos de variable pero eso no significa que las elipses sean irrelevantes o que sus propiedades se deduzcan fácilmente de aquellas de las circunferencias.

<u>Observación</u>. Por la fórmula del cambio de base, el signo del determinante de la matriz de una forma bilineal en $\mathbb{R}$ o sesquilineal, no varía (porque $|C||C^t| > 0$, $|C||\overline{C}^t| > 0$). Con ello, el criterio de Sylvester se puede deducir del proceso de Gram-Schmidt. Aquí va una demostración esquemática: Sea $\{\vec{e}_1, \dots, \vec{e}_n\}$ la base canónica de $\mathbb{K}^n$ y V_k el subespacio generado por $\mathcal{B}_k = \{\vec{e}_1, \dots, \vec{e}_k\}$. La forma $\varphi : \mathbb{K}^n \times \mathbb{K}^n \longrightarrow \mathbb{K}$ de matriz A tiene matriz A_k con $\Delta_k = |A_k|$ cuando la restringimos a V_k. Si φ es producto escalar, cambiando a una base ortonormal la transformaríamos en la identidad I y $|I| = 1 > 0$, así que $\Delta_k > 0$. Recíprocamente, si no es producto escalar, no podríamos completar el proceso de Gram-Schmidt en $\mathcal{B}_n$ (alguna norma saldría no positiva). Digamos que nos paramos en $\mathcal{B}'_k = \{\vec{u}_1, \dots, \vec{u}_{k-1}, \vec{u}\}$ base de V_k con $\{\vec{u}_1, \dots, \vec{u}_{k-1}\}$ ortonormal y $\varphi(\vec{u}, \vec{u}_j) = 0$ pero $\varphi(\vec{u}, \vec{u}) \le 0$. En V_k con la base $\mathcal{B}_k$, φ tiene una matriz con determinante Δ_k y en la base $\mathcal{B}'_k$ el determinante sería $\varphi(\vec{u}, \vec{u})$, así pues $\Delta_k \le 0$.

Dado un espacio vectorial euclídeo o hermítico V y un subespacio $W \subset V$ se define el *complemento ortogonal* de W como

$$W^\perp = \big\{\vec{v} \in V \,:\, \langle \vec{v}, \vec{w}\rangle = 0 \quad \text{para todo } \vec{w} \in W\big\}. \tag{29}$$

Esto es, $W^\perp$ está formado por todos los vectores que son ortogonales a los vectores de W. Es fácil ver que es un subespacio vectorial de V.

<u>Observación</u>. La ortogonalidad entre los vectores de W y de $W^\perp$ indica a veces en física una especie de irrelevancia o separación. Por ejemplo, en el principio de d'Alambert las diferencias $\vec{F} - m\vec{a}$ que son ortogonales a los desplazamientos virtuales no tienen influencia en el comportamiento mecánico o en física cuántica entre los estados de sistemas que están en W y en $W^\perp$ hay una probabilidad de transición nula.

En el caso de dimensión finita se cumple que $V = W \oplus W^\perp$ cualquiera que sea el subespacio W. Recordemos que el símbolo $\oplus$ indica la *suma directa*. Esto significa que $W \cap W^\perp = \{\vec{0}\}$ (lo que se deduce de que el producto escalar es definido positivo) y que cualquier $\vec{v} \in V$ se puede escribir como $\vec{x} + \vec{y}$ con $\vec{x} \in W$ y $\vec{y} \in W^\perp$. Para probar esta última afirmación, consideremos una base ortonormal de V $\mathcal{B} = \{\vec{u}_1, \dots, \vec{u}_m, \vec{u}_{m+1}, \dots, \vec{u}_n\}$ con $\vec{u}_j \in W$ si $1 \le j \le m = \dim W$, lo cual se consigue con el proceso de Gram-Schmidt tomando los m primeros vectores en W. Automáticamente $\vec{u}_j \in W^\perp$ y por ser una base ortonormal

$$\vec{v} = \big(\langle \vec{v}, \vec{u}_1\rangle \vec{u}_1 + \cdots + \langle \vec{v}, \vec{u}_m\rangle \vec{u}_m\big) + \big(\langle \vec{v}, \vec{u}_{m+1}\rangle \vec{u}_{m+1} + \cdots + \langle \vec{v}, \vec{u}_n\rangle \vec{u}_n\big). \tag{30}$$

Entonces $\vec{x}$ es el primer paréntesis y $\vec{y}$ el segundo.

También en dimensión finita se cumple $(W^\perp)^\perp = W$. Aunque esta propiedad parezca obvia, el mundo de los espacios vectoriales de dimensión infinita es tan raro que allí deja de ser cierto en general.

Ejemplo. En $\mathbb{C}^4$ con el producto escalar usual consideramos el subespacio $W = \{\vec{x} \in \mathbb{C}^4 : x_2 + x_3 + (1+i)x_4 = 0\}$, vamos a hallar bases ortogonales de W y $W^\perp$. En primer lugar nos percatamos de que los vectores de W son justamente los $\vec{x} \in \mathbb{C}^4$ que cumple $\langle \vec{x}, \vec{y}\rangle = 0$ con $\vec{y} = (0, 1, 1, 1-i)^t$. Esto prueba inmediatamente que $\vec{y}$ genera $W^\perp$ y así $\{\vec{y}\}$ es ya una base ortonormal (trivial) de $W^\perp$. De aquí $\dim W = 3$ y $\vec{u}_1 = (1,0,0,0)^t$, $\vec{u}_2 = (0,1,-1,0)^t$, $\vec{u}_3 = (0,-1-i,0,1)^t$ conforman una base de W porque son linealmente independientes. El proceso de Gram-Schmidt es bastante trivial para ellos y da lugar a la base ortogonal $\big\{\vec{u_1}, \vec{u}_2, \vec{v}_3 + (1+i)\vec{v}_2/2\big\}$. El último vector es $\big(0, -\frac{1}{2}(1+i), -\frac{1}{2}(1+i), 1\big)^t$.

Ejemplo. [Opcional] Sea V el espacio vectorial sobre $\mathbb{R}$ de todas las sucesiones reales $\{x_n\}_{n=1}^\infty$ con soporte finito (las que son nulas a partir de cierto momento) con el producto escalar $\sum_{n=1}^\infty x_n y_n$ y sea W el subespacio de las sucesiones de V con $\sum_{n=1}^\infty x_n = 0$. Si $\{y_n\}_{n=1}^\infty \in W^\perp$ entonces como $W^\perp \subset V$ existe N tal que $y_n = 0$ para $n \geq N$ y tomando la sucesión con $x_n = 1$ si $n \leq N$ y $x_n = -1$ si $N < n \leq 2N$ se tiene $\{x_n\}_{n=1}^\infty \in W$ y por tanto $0 = \sum_{n=1}^\infty x_n y_n = \sum_{n=1}^N y_n$ lo que implica $\{y_n\}_{n=1}^\infty \in W$ por tanto $\{y_n\}_{n=1}^\infty$ es la sucesión nula, entonces $W^\perp = \{0\}$ y $(W^\perp)^\perp = V \neq W$. Como hemos visto, esto no puede ocurrir en espacios de dimensión finita.