Capítulo 7. Espacios vectoriales euclídeos

Introducción

La herramienta esencial para poder saltar del Álgebra Lineal a la Geometría es el comcepto de *producto escalar*, que permite hablar de longitudes, distancias y ángulos. No debemos pensar solamente en el producto escalar clásico, conocido desde el bachillerato, sino que debemos considerar unos "productos escalares" más generales (por ejemplo, midiendo las longitudes de manera diferente según en qué dirección).

Cuando en un espacio vectorial tenemos un producto escalar, decimos que estamos en un *espacio vectorial euclídeo*. Este capítulo se dedicará a su estudio.

Definición.- Sea V un espacio vectorial (de dimensión finita para nosotros, aunque esa condición no es fundamental). Un producto escalar en V no es más que una forma bilineal simétrica en V que sea definida positiva. Es costumbre usar la notación $u \cdot v$ para denotar el producto escalar de los vectores u y v; también es frecuente la notación $< u, v >$.

El par (V, f) formado por un espacio vectorial V y un producto escalar en él, f, es un *espacio vectorial euclídeo*. Es habitual mencionar sólo a V, dando por sobreentendido el producto escalar. Ese abuso no es peligroso si no hay duda acerca de cuál es la forma bilineal a la que nos referimos; en otro caso, habrá que especificarlo.

Ejemplos y ejercicios.

$< (x, y), (x', y') >= x \cdot x' + y \cdot y'$ es el producto escalar usual en $\mathbf{R}^2$.

$< (x, y), (x', y') >= 3x \cdot x' + 2y \cdot y'$ es un producto escalar diferente.

$f((x, y), (x', y')) = x \cdot x' + 2x \cdot y' + 2y \cdot x' + y \cdot y'$ es una forma bilineal simétrica en $\mathbf{R}^2$, pero no es un producto escalar (piense qué le falta).

$< (x, y, z), (x', y', z') >= x \cdot x' + y \cdot y' + z \cdot z'$ es el producto escalar usual en $\mathbf{R}^3$.

$< f, g >= \int_0^1 f(x) \cdot g(x)dx$ es un producto escalar en el espacio vectorial $C^0([0, 1], \mathbf{R})$.

Definición.- En el espacio vectorial euclídeo V, llamamos *norma* de un vector u (también *módulo* de u) a la raíz cuadrada de $< u, u >$: $\|u\| = \sqrt{< u, u >}$.

Ha de subrayarse que el radicando no es negativo, por ser $<, >$ definida positiva, y que la norma de u es 0 si u el vector nulo (y sólo en ese caso). Un vector cuya norma sea igual a 1 se llama *vector unitario*.

Ejercicio: Calcule la norma de las funciones $\varphi(x) = \cos(x)$, $\psi(x) = \sin(x)$, $\zeta(x) = \cos(nx)$, $\eta(x) = \sin(nx)$, en el espacio $C^0([-\pi, \pi], \mathbf{R})$ con el producto escalar dado por la integral.

La norma permite hablar de longitud, y por tanto de distancia (la distancia de u a v será la norma de la diferencia $u - v$). Por otra parte, en un espacio vectorial euclídeo también podemos medir ángulos: recordemos que el producto escalar (usual) de dos vectores es igual al producto de sus módulos por el coseno del ángulo que forman los vectores. Extender esto al caso de productos escalares generales requiere un paso previo:

Proposición (desigualdad de Schwarz).- En un eve V se tiene $|< u, v >| \leq \|u\| \cdot \|v\|$, cualesquiera que sean los vectores u y v.

Demostración: Para todo $x \in \mathbf{R}$ se tiene $< u + x \cdot v, u + x \cdot v > \geq 0$, lo que equivale (por la bilinealidad) a $\|u\|^2 + 2x \cdot < u, v > +x^2 \cdot \|v\|^2 \geq 0$.

Lo que tenemos ahí es un polinomio de segundo grado que siempre es positivo (o nulo, como poco), lo que nos indica que su discriminante es negativo o nulo, es decir: $4 \cdot < u, v >^2 -4 \cdot \|u\|^2 \cdot \|v\|^2 \leq 0$, , sea $< u, v >^2 \leq \|u\|^2 \cdot \|v\|^2$, de donde se deduce la desigualdad de Schwarz extrayendo las raíces cuadradas.

Puede modificarse la demostración dándole a x el valor $\frac{-< u,v >}{\|v\|^2}$, Complete los detalles como ejercicio.

En consecuencia, el cociente $\frac{|< u,v >|}{\|u\| \cdot \|v\|}$ está siempre entre -1 y 1, lo que permite verlo como el coseno de un ángulo, y definir así el ángulo entre u y v como arccos $\frac{|< u,v >|}{\|u\| \cdot \|v\|}$.

Ejercicio: Discuta cuándo la desigualdad de Schwarz se convierte en una igualdad, e interprete el resultado geométricamente.

Ejemplos:

La desigualdad de Schwarz se traduce de diversas maneras en distintos espacios. Así, $|x_1 \cdot y_1 + x_2 \cdot y_2| \leq \sqrt{x_1^2 + x_2^2} \cdot \sqrt{y_1^2 + y_2^2}$, en el plano $\mathbf{R}^2$ con el producto escalar habitual.

En el espacio $C^0([0, 1], \mathbf{R})$ considerado más arriba, se tiene $| \int_0^1 (f \cdot g)| \leq \sqrt{\int_0^1 f^2} \cdot \sqrt{\int_0^1 g^2}$.

El ángulo que forman los enlaces en el átomo de metano (que tiene la forma de un tetraedro regular) se puede calcular observando que tomando un sistema de referencia con origen en el centro del tetraedro y ejes adecuadamente dispuestos, un vértice está en el punto $(0, 0, 3)$ y los otros tres en puntos $(a, b, -1)$. Así, el producto escalar es igual a -3, en tanto que el módulo de cada vector es 3, por lo que el coseno del ángulo que forman es igual a $-1/3$, y ese ángulo resulta ser de 109^{o} aproximadamente.

Ejercicio: Calcule el ángulo que forman dos caras contiguas de un octaedro regular. También las de un tetraedro regular.

Una consecuencia sencilla de la desigualdad de Schwarz es la llamada *desigualdad triangular*: $\|u+v\| \leq \|u\|+\|v\|$. Se demuestra observando que $\|u+v\|^2 =< u+v, u+v >= \|u\|^2+2 < u, v > +\|v\|^2 \leq \|u\|^2 + 2| < u, v > | + \|v\|^2$ y que, por la desigualdad de Schwarz, la última expresión es menor o igual que $(\|u\| + \|v\|)^2$; extrayendo las raíces cuadradas se llega a la desigualdad buscada.

Ejercicio: Demuestre que $\|u\| - \|v\| \leq \|u - v\|$ para cualesquiera u y v. ¿Qué condición han de cumplir para que se dé la igualdad?

Ortogonalidad

Los conceptos relativos a formas bilineales simétricas siguen vigentes, claro está, en el caso del producto escalar, pero en algunos casos adquieren nuevos nombres. Así, la matriz que representa al producto escalar en una base, a veces recibe el nombre de *matriz de Gram*. Y no decimos que dos vectores son *conjugados*, sino que son *ortogonales* o *perpendiculares*. Igualmente hablamos del *subespacio ortogonal* en vez del subespacio conjugado. El cambio de denominación es pertinente porque la ortogonalidad goza de algunas propiedades que no tiene la mera conjugación. Veamos una:

Proposición.- Si unos vectores son ortogonales dos a dos y ninguno es el vector nulo (lo que llamamos una *familia ortogonal*), entonces son linealmente independientes.

Demostración: Sea $\lambda_1 \cdot v_1 + \ldots + \lambda_k \cdot v_k = 0$, entonces, multiplicando escalarmente por v_1 resulta $\lambda_1 \cdot < v_1, v_1 > + \ldots + \lambda_k \cdot < v_k, v_1 > = < 0, v_1 > = 0$, que se reduce a $\lambda_1 \cdot \|v_1\|^2 = 0$ por la ortogonalidad de v_1 con los demás vectores. Como $v_1 \neq 0$, su norma no se anula, y por tanto será $\lambda_1 = 0$. Del mismo modo, los demás coeficientes son también nulos, y los vectores son linealmente independientes.

Nótese que el resultado análogo para formas bilineales simétricas no es cierto en general, porque puede haber vectores no nulos en el núcleo.

Consecuencia.- En un eve de dimensión n, una famila ortogonal de n vectores es una base. Naturalmente, una base así se llamará *base ortogonal* del espacio en cuestión. Si los vectores de la base son todos unitarios, la llamamos *base ortonormal*. Pasar de una base ortogonal a una ortonormal es tan sencillo como dividir cada vector por su norma.

La existencia de bases ortogonales y ortonormales está garantizada por la posibilidad de diagonalizar formas cuadráticas. De hecho, el procedimiento que vimos para diagonalizar una forma cuadrática mediante oe nos sirve perfectamente para hallar bases ortogonales en un eve. Otro método muy célebre para conseguirlo es el conocido como *procedimiento de ortogonalización de Gram-Schmidt*, que se describe a continuación:

Se empieza con una base cualquiera del espacio, $\{v_1, \ldots, v_n\}$ y se fabrica otra base $\{u_1, \ldots, u_n\}$ paso a paso, comenzando por elegir $u_1 = v_1$. El segundo vector será $u_2 = v_2 + \alpha u_1$, donde el escalar α se escoge de manera que u_2 sea ortogonal a u_1, lo que nos da $\alpha = - < v_2, u_1 > / \|u_1\|^2$. El tercer vector es $u_3 = v_3 + \alpha u_1 + \beta u_2$, siendo los escalares α y β los adecuados para que u_3 sea ortogonal a u_1 y u_2, esto es $\alpha = - < v_3, u_1 > / \|u_1\|^2$, $\beta = - < v_3, u_2 > / \|u_2\|^2$. Así se prosigue hasta completar la base $\{u_1, \ldots, u_n\}$, que resulta ortogonal por cómo se construye.

Observación: Merece destacarse que aunque los vectores $u_2, u_3, \ldots, u_n$ no coinciden con $v_2, v_3, \ldots, v_n$, los subespacios $L(u_1)$, $L(u_1, u_2)$, $L(u_1, u_2, u_3)$ etc., coinciden con $L(v_1)$, $L(v_1, v_2)$, $L(v_1, v_2, v_3)$ etc.

Ejercicio.- Encuentre una base ortogonal de $\mathbf{R}^3$ (con el producto escalar usual) por este procedimiento partiendo de la base $\{(0, 1, 1), (1, 0, 1), (1, 1, 0)\}$.

Ejercicio.- Encuentre por este procedimiento una base ortogonal de $\mathbf{R}^3$ (con el producto escalar dado por la forma cuadrática $q(x, y, z) = 2x^2 + y^2 + 4z^2 + 2xy - 2yz$) partiendo de la base canónica.

Ejercicio.- Encuentre una base ortogonal del espacio de polinomios $\mathbf{R}_4[x]$ (con el producto escalar dado por la integral entre -1 y 1) por este procedimiento partiendo de la base $\{1, x, x^2, x^3, x^4\}$.

Por otra parte, es fácil ver que cualquier familia ortogonal se puede ampliar a una base ortogonal de todo el espacio: sencillamente, amplíese de cualquier manera a una base de V y luego aplíquese el método de Gram-Schmidt (que dejará invariados los primeros vectores puesto que ya eran ortogonales) para conseguir la base ortogonal.

Una ventaja evidente de las bases ortogonales es que las coordenadas de un vector se calculan con mucha facilidad, y para calcular una coordenada concreta no hay que tener en cuenta todos los vectores de la base, sino sólo el vector involucrado. En efecto, se tiene:

Proposición.- Si $\{v_1, \ldots, v_n\}$ es una base ortogonal del eve V, y $v = \lambda_1 v_1 + \ldots + \lambda_n v_n$ es un vector cualquiera de V, entonces la k-ésima coordenada, λ_k, es igual al producto escalar $< v, v_k >$ dividido por el cuadrado de la norma de v_k. En particular, si la base es ortonormal, cada coordenada es igual al producto escalar del vector v por el correspondiente vector de la base.

Demostración: Es inmediata, basta con realizar el producto escalar de $v = \lambda_1 v_1 + \ldots + \lambda_n v_n$ por v_k; a la izquierda nos queda $< v, v_k >$, y a la derecha $\lambda_k < v_k, v_k >$, debido a la ortogonalidad de los vectores de la base.

Ejercicio.- Calcule las coordenadas del vector $(3, 5, -4)$ en el espacio vectorial $\mathbf{R}^3$ respecto de la base $\{(1, 1, 1), (1, 0, -1), (1, -2, 1)\}$ (que es ortogonal) y respecto de la base $\{(0, 1, 1), (1, 0, 1), (1, 1, 0)\}$ (que no lo es). Observe la diferencia.

Observación: La importancia de la ortogonalidad va mucho más allá de las ventajas para calcular las coordenadas de los vectores. Aunque no podemos entrar en ello, quiero comentar que los sistemas de ecuaciones referidos a bases ortogonales son mucho menos sensibles a perturbaciones en los datos que los referidos a bases cualesquiera. Además, los polinomios ortogonales permiten calcular numéricamente integrales con gran eficacia y rapidez; pero no debemos desviarnos de nuestro temario.

En cuanto al subespacio ortogonal a un subespacio dado (o a una familia de vectores dada), la situación es particularmente grata, puesto que se tiene:

Proposición.- Si S es un subespacio vectorial de V, entonces el subespacio S^0 formado por los vectores ortogonales a todos los de S es suplementario de S (le llamaremos el *suplemento ortogonal de* S).

Demostración: En pocas palabras, se toma una base ortogonal de S y se amplía hasta una base ortogonal de V. Los vectores añadidos engendran un subespacio de V que es evidentemente suplementario de S. Comprobar que ése es precisamente el subespacio S^0 es mera rutina.

Proyección ortogonal sobre un subespacio.- La estructura geométrica que aparece al disponer de un producto escalar hace que en un eve podamos hablar de proyectar sobre un subespacio en una dirección privilegiada, que es la perpendicular. Recordemos que hace varios capítulos se habló de la proyección sobre un subespacio S_1 paralelamente a otro S_2 siempre que los dos subespacios sean suplementarios. Ahora, la presencia del suplemento ortogonal nos permite hablar de proyectar ortogonalmente sobre S, lo que significa hacerlo paralelamente a S^0. Con detalle, cada vector $v \in V$ se puede descomponer (de una sola manera) como una suma $v = v_1 + v_2$ donde $v_1 \in S$ y $v_2 \in S^0$; asociando a v el vector v_1 estamos definiendo un endomorfismo en V: $p(v) = v_1$. Otra manera de decir lo mismo es que $p(v) \in S, p(v) - v \in S^0$. El núcleo de ese endomorfismo p es el subespacio S^0 y la imagen es S. Un ejercicio sencillo consiste en comprobar esas afirmaciones, así como que es un endomorfismo idempotente ($p^2 = p$), que sus únicos autovalores son 0 y 1, y que es diagonalizable. Hágalo.

Pariente cercana de la proyección ortogonal es la simetría respecto de S, que asocia a cada vector de V el que sería su "imagen especular" en ese subespacio, es decir, el único vector, w, tal que $w + v = 2p(v)$. Es evidente que esta simetría viene definida por $\sigma(v) = 2p(v) - v$. Queda como tarea encontrar núcleo e imagen de σ, así como sus autovalores y autoespacios.

Ejemplo.- Calcule la proyección ortogonal del vector $v = (1, 1, 2)$ sobre el plano π de ecuación $x + 2y - z = 0$ (en $\mathbf{R}^3$ con el producto escalar usual). Calcule también su simétrico respecto del mismo plano. Haga otro tanto con un vector cualquiera (x, y, z).

A) Enfoque elemental.- El vector $n = (1, 2, -1)$ es perpendicular al plano π. Proyectamos el vector v sobre n: $v' = k \cdot n$ con $k = \frac{v \cdot n}{n \cdot n} = \frac{1}{6}$.

La proyección de v sobre el plano π es igual a $p(v) = v - v' = (1, 1, 2) - (1/6, 1/3, -1/6) = (5/6, 2/3, 13/6)$.

El simétrico de v respecto de π es $\sigma(v) = v - 2v' = (1, 1, 2) - (1/3, 2/3, -1/3) = (2/3, 1/3, 7/3)$.

Del mismo modo, si $v = (x, y, z)$, resultará que $k = \frac{v \cdot n}{n \cdot n} = \frac{x+2y-z}{6}$, por lo que su proyección y su simétrico serán:

$p(v) = v - v' = (x, y, z) - \frac{x+2y-z}{6}(1, 2, -1) = \frac{1}{6}(5x - 2y + z, -2x + 2y + 2z, x + 2y + 5z)$,

$\sigma(v) = v - 2v' = (x, y, z) - \frac{x+2y-z}{3}(1, 2, -1) = \frac{1}{3}(2x - 2y + z, -2x - y + 2z, x + 2y + 2z)$.

B) Segundo enfoque elemental.- Elegimos una base de π, y la escogemos ortogonal: $\{u_1 = (1, 0, 1), u_2 = (-1, 1, 1)\}$. Al añadirle n, tenemos una base ortogonal de $\mathbf{R}^3$: $\mathrm{B} = \{u_1, u_2, n\}$.

Un vector, v, se descompone como $v = x^1 \cdot u_1 + x^2 \cdot u_2 + x^3 \cdot n$, donde las coordenadas se calculan fácilmente (debido a la ortogonalidad de la base): $x^1 = \frac{u_1 \cdot v}{u_1 \cdot u_1}, x^2 = \frac{u_2 \cdot v}{u_2 \cdot u_2}, x^3 = \frac{n \cdot v}{n \cdot n}$.

La proyección de v sobre π es $p(v) = x^1 \cdot u_1 + x^2 \cdot u_2$, es decir (en el caso general, $v = (x, y, z)$):

$p(v) = \frac{x+z}{2} \cdot (1, 0, 1) + \frac{-x+y+z}{3} \cdot (-1, 1, 1) = \frac{1}{6}(5x - 2y + z, -2x + 2y + 2z, x + 2y + 5z)$.

El simétrico se puede deducir de la fórmula $v + \sigma(v) = 2 \cdot p(v)$, de modo que

$\sigma(v) = 2 \cdot p(v) - v = \frac{1}{3}(5x - 2y + z, -2x + 2y + 2z, x + 2y + 5z) - (x, y, z) = \frac{1}{3}(2x - 2y + z, -2x - y + 2z, x + 2y + 2z)$.

C) Enfoque astuto.- Elegimos una base de $\mathbf{R}^3$ adaptada al problema, es decir formada por vectores tomados en el subespacio π y en su ortogonal: $\mathrm{B} = \{u_1, u_2, n\}$.

Así, $p(u_1) = u_1, p(u_2) = u_2, p(n) = 0$, y la matriz de p en esa base es $M = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$.

Si $P = \begin{pmatrix} 1 & -1 & 1 \\ 0 & 1 & 2 \\ 1 & 1 & -1 \end{pmatrix}$ es la matriz de paso de esa base B a la base canónica, entonces la matriz de p en la base canónica es

$$A = P \cdot M \cdot P^{-1} = P \cdot M \cdot \begin{pmatrix} 1/2 & 0 & 1/2 \\ -1/3 & 1/3 & 1/3 \\ 1/6 & 1/3 & -1/6 \end{pmatrix} = \begin{pmatrix} 5/6 & -1/3 & 1/6 \\ -1/3 & 1/3 & 1/3 \\ 1/6 & 1/3 & 5/6 \end{pmatrix}$$

Por tanto, podemos calcular $p(v)$ multiplicando: $A \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} (5x-2y+z)/6 \\ (-x+y+z)/3 \\ (x+2y+5z)/6 \end{pmatrix}$.

En cuanto a la simetría, procedemos del mismo modo:

$\sigma(u_1) = u_1, \sigma(u_2) = u_2, \sigma(n) = -n$, y la matriz de σ en la base B es $N = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$, por lo que en la base canónica es $Z = P \cdot M \cdot P^{-1} = \begin{pmatrix} 2/3 & -2/3 & 1/3 \\ -2/3 & -1/3 & 2/3 \\ 1/3 & 2/3 & 2/3 \end{pmatrix}$.

Se sigue que $\sigma(v)$ se obtiene mediante el producto $Z \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} (2x-2y+z)/3 \\ (-2x-y+2z)/3 \\ (x+2y+2z)/3 \end{pmatrix}$.

Nota.- Obsérvese que la matriz inversa P^{-1} se parece mucho a la traspuesta P^T. Eso es debido a que es un cambio de base entre dos bases ortogonales, B y C. Si B hubiese sido ortonormal, como C, no estaríamos hablando de que P^{-1} y P^T se parecen, sino de que son idénticas. Pruebe a dividir cada vector de la base B por su módulo y verá lo que sucede (piense también si vale la pena el precio pagado: unas raíces cuadradas que vamos arrastrando).

Nota.- Los cálculos de invertir matrices y multiplicarlas pueden resultar tediosos y hay riesgo de errores. La elección de bases ortogonales facilita el cálculo de la inversa (es casi la traspuesta). De todos modos, yo me he servido de Octave UPM.

La proyección ortogonal resuelve un problema de optimización: dado un vector, v, y un subespacio, U, de todos los vectores de U la mejor aproximación a v (en el sentido de que la norma de la diferencia sea la menor) la da la proyección ortogonal de v sobre U.

Ejercicio.- Calcule la proyección ortogonal del vector $(1,1,2)$ sobre la recta $2x = y = 2z$ (en $\mathbf{R}^3$ con el producto escalar usual). Calcule también su simétrico respecto de la misma recta. Haga otro tanto con un vector cualquiera (x, y, z).

El ejemplo anterior, lo mismo que el ejercicio que le sigue, puede dar la impresión de que el enfoque elemental es suficiente. Sin duda, en tres dimensiones las proyeccines son muy sencillas, de cualquier manera que se planteen, pues uno de los subespacios (aquel sobre el que se proyecta o su ortogonal) será una recta. Pero si nos situamos en un espacio de dimensión superior, de forma que S y su ortogonal tengan dimensión mayor que 1, acaso no sea inmediato cómo descomponer un vector para proyectarlo sobre S, y el tercer planteamiento (elegir bases adecuadas al problema) se impone.

Ejercicio.- Si V es un eve de dimensión finita y S es un subespacio vectorial de V, la proyección ortogonal sobre S es un endomorfismo $p : V \to V$. Compruebe que $p^2 = p$ y que los únicos autovalores de p son 0 y 1. Discuta si V admite una base de autovectores de p o no. Traduzca los resultados anteriores a la matriz de p en una base de V.

Ejercicio.- Los resultados del ejercicio anterior son igualmente válidos si p no es la proyección ortogonal, sino la proyección paralela a algún otro subespacio suplementario de S. Compruébelo. Estudie el caso concreto en que V es el espacio $\mathbf{R}^3$ con el producto escalar usual, S es el plano de ecuación $2x + y + z = 0$, y los suplementarios que se consideran son el complemento ortogonal de S y el eje z.

Ejercicio.- Volviendo a la proyección ortogonal, demuestre que $< u, p(v) >=< p(u), v >$ cualesquiera que sean los vectores u y v. Muestre con un contraejemplo que eso no sucede si p designa una proyección paralela a un subespacio que no es el complemento ortogonal.

Ejercicio (Traducción del ejercicio anterior a términos matriciales).- Si A es la matriz de la proyección ortogonal sobre S (respecto de cierta base B) y G es la matriz de Gram del producto escalar en esa misma base, pruebe que $G \cdot A = A^T \cdot G$ (indicación: el producto escalar $< u, p(v) >$ se expresa como $u_B^T \cdot G \cdot A \cdot v_B$, mientras que $< p(u), v >$ se expresa como $(A \cdot u_B)^T \cdot G \cdot v_B$). Concluya que si la base B es ortonormal, entonces la matriz A es simétrica.

Ejercicio (Simetrías) .- Si V es un eve de dimensión finita y S es un subespacio vectorial de V, la simetría respecto de S es otro endomorfismo $\sigma : V \to V$. Compruebe que $\sigma^2 = 1$ y que los únicos autovalores de σ son 1 y -1. Discuta si V admite una base de autovectores de σ o no. Traduzca los resultados anteriores a la matriz de σ en una base de V.

Ejercicio (Traducción del ejercicio anterior a términos matriciales).- Si Z es la matriz de la simetría respecto de S (en cierta base B) y G es la matriz de Gram del producto escalar en esa misma base, pruebe que $G \cdot Z = Z^T \cdot G$ (indicación: es inmediato, expresando σ en función de p)). Concluya que si la base B es ortonormal, entonces la matriz Z es simétrica.

Una aplicación importante: ajuste por mínimos cuadrados.- A menudo, para calcular algunas magnitudes se realizan mediciones con las que se pueden evaluar esas magnitudes, frecuentemente por medio de unas ecuaciones. Los errores y la imprecisión en las medidas suelen conducir a sistemas de ecuaciones incompatibles que, sin embargo, interesa resolver. Veamos un ejemplo muy simple:

Al realizar dos experimentos para calcular una cantidad, x, llegamos a las ecuaciones $2x = 9$ y $3x = 12$. Claramente, es imposible que ambas ecuaciones se satisfagan simultáneamente, pero no es prudente descartarlas sin más: podemos pensar que la cantidad x que pretendemos conocer estará entre los valores 4 y $9/2$. Es natural preguntarse cuál sería el valor más adecuado, si hemos de decantarnos por alguno. Una idea sensata consiste en escoger x de modo que haga mínimo el error promedio entre las 2 ecuaciones; la forma más frecuente de hacerlo es tomar la suma de los cuadrados $(2x-9)^2 + (3x-12)^2$ y elegir x para que sea mínima esa suma. Un sencillo cálculo de derivadas nos conduce a que el mínimo se alcanza para $x = 54/13$. Si realizamos otro experimento más, obtenemos una tercera ecuación, digamos $5x = 21$. Al tratar ahora de ajustar lo mejor posible el resultado, procuramos minimizar la suma $(2x-9)^2 + (3x-12)^2 + (5x-21)^2$, lo que nos da la nueva "solución" $x = 159/38$.

Geométricamente, podemos pensar que el primer sistema no tiene solución porque el vector $(9, 12)$ no es múltiplo del $(2, 3)$, y las ecuaciones piden que $(2, 3)x = (9, 12)$, lo mejor que podemos conseguir es elegir x de manera que $(2, 3)x$ se aproxime tanto como sea posible a $(9, 12)$, lo que supone hacer que $(2, 3)x$ sea la proyección ortogonal del vector $(9, 12)$ sobre el subespacio generado por $(2, 3)$; esa proyección no es otra que $(108/13, 162/13)$, o sea $(2, 3)\cdot 54/13$. En el segundo sistema, sucede otro tanto con el vector $(9, 12, 21)$ y la recta generada por $(2, 3, 5)$.

Podemos encontrarnos con más ecuaciones e incógnitas, esto es, con un sistema $Ax = b$ que sea incompatible, a menudo por tener demasiadas ecuaciones. La incompatibilidad significa que el vector b no pertenece al espacio generado por las columnas de A. Si buscamos la aproximación mejor, pensaremos en un vector de ese espacio, es decir As, lo más parecido posible a b, lo que significa que la diferencia $As - b$ será ortogonal al espacio columna de A; eso se expresa algebraicamente por medio de $A^T \cdot (As - b) = 0$, es decir $A^T As = A^T b$. La solución de este sistema, s (que existirá y será única si las columnas de A son linealmente independientes) es la que tomamos como solución aproximada del sistema incompatible $Ax = b$.

Ejemplo.- El sistema

$$\begin{cases} x + y = 1, \\ x + 2y = 3, \\ x + 3y = 4 \end{cases}$$

es incompatible, porque el vector $b^T = (1, 3, 4)$ no es combinación lineal de $(1, 1, 1)$ y $(1, 2, 3)$. Su proyección sobre el plano que estos generan se calcula fácilmente, basta con imponer que $As - b$ sea ortogonal a ese plano, lo que es tanto como $A^T As = A^T b$. Como la matriz A es

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix}$$

resulta que

$$A^T \cdot A = \begin{pmatrix} 3 & 6 \\ 6 & 14 \end{pmatrix}$$

y el sistema para hallar la solución aproximada es

$$\begin{cases} 3x + 6y = 8, \\ 6x + 14y = 19 \end{cases}$$

cuya solución es $x = -1/3, y = 3/2$ que es la solución exacta del sistema

$$\begin{cases} x + y = 7/6, \\ x + 2y = 8/3, \\ x + 3y = 25/6 \end{cases}$$

como corresponde al hecho de que $(7/6, 8/3, 25/6)$ es la proyección ortogonal del vector $(1, 3, 4)$ sobre el plano generado por $1, 1, 1)$ y $(1, 2, 3)$.

El libro de Strang "Álgebra Lineal y sus aplicaciones" trata este asunto (el de los mínimos cuadrados y las proyecciones ortogonales) en su capítulo tercero, cuya lectura recomiendo (al menos, la primera parte, si es que uno no quiere embarcarse en un viaje que le muestre la descomposición QR, series de Fourier, pseudoinversas y descomposición en valores singulares; apasionante periplo que rebasa nuestros objetivos). Ahí pueden encontrar fórmulas explícitas y explicaciones detalladas que amplíen lo que acabo de exponer aquí.

En la asignatura de Estadística, se topa uno con el ajuste por mínimos cuadrados enseguida, al encajar datos por medio de la recta de regresión. Bueno, en realidad, uno se encuentra el ajuste por mínimos cuadrados por todas partes; cito el caso anterior por mencionar algo próximo en sus estudios.

En un ámbito totalmente diferente, la compresión de archivos, imágenes o música (tal como sucede al almacenarla en CD o MP4) está basada también en una proyección ortogonal sobre un subespacio: si una melodía viene representada por una función, la podemos recuperar aproximadamente almacenando los coeficientes de su proyección ortogonal sobre un subespacio generado por ciertas funciones básicas (senos y cosenos, en el caso más sencillo; "wavelets" si nos ponemos más al día), lo que supone un notable ahorro de espacio y puede ser suficiente para nuestro oído.

Aplicaciones ortogonales.- Las proyecciones y las simetrías nos han metido en el terreno de las aplicaciones lineales en un eve. Es natural pensar que no todos los homomorfismos en una eve tienen el mismo interés, y que su comportamiento con el producto escalar será relevante. Así, un endomorfismo del espacio vectorial euclídeo V que "respete" el producto escalar recibirá una atención especial. Tales endomorfismos se llaman *aplicaciones ortogonales*, y les vamos a dedicar un poco de tiempo.

Definición.- Una *aplicación ortogonal* en el espacio vectorial euclídeo V, es un endomorfismo de V que conserve el producto escalar, esto es $< f(u), f(v) >=< u, v >$.

Observación.- Una aplicación ortogonal conservará la norma (es decir, $\|f(u)\| = \|u\|$) y la ortogonalidad, por lo que transformará una base ortonormal de V en otra base ortonormal. De hecho, esa propiedad caracteriza a las aplicaciones ortogonales. Lo escribimos:

Proposición.- Si f es un endomorfismo del eve V que transforma una base ortonormal en otra base ortonormal, entonces f es una aplicación ortogonal, y a la inversa.

Demostración: Es evidente.

Otra observación.- Una aplicación lineal que conserve la norma será una aplicación ortogonal, puesto que el producto escalar se expresa con ayuda de la norma: $< u, v >= \frac{1}{2} \cdot (\|u + v\|^2 - \|u\|^2 - \|v\|^2)$.

Otra más.- Una aplicación que conserve el producto escalar es necesariamente lineal, por lo que no es preciso exigirlo en la definición de aplicación ortogonal. Para comprobarlo, calculemos el producto escalar $< f(\lambda u) - \lambda f(u), f(\lambda u) - \lambda f(u) >$ y resulta ser igual a $< f(\lambda u), f(\lambda u) > -2\lambda < f(\lambda u), f(u) > +\lambda^2 < f(u), f(u) >$, que por conservar f el producto escalar es igual a $< \lambda u, \lambda u > -2\lambda < \lambda u, u > +\lambda^2 < u, u >$, o sea, a 0. Como el producto escalar es definido positivo, se sigue que $f(\lambda u) - \lambda f(u) = 0$. De manera análoga se prueba que $f(u + v) - f(u) - f(v) = 0$, y f es lineal.

Si B es una base de V, y la matriz de f en la base B es A, entonces la condición de ortogonalidad se traduce por $A^T \cdot G \cdot A = G$, siendo G la matriz de Gram en esa base, puesto que $u_B^T G v_B$ es el producto escalar de u por v y el de sus imágenes es $(Au_B)^T G(Av_B)$, que es igual a $u_B^T A^T G A v_B$. Si la base es ortonormal, la matriz de Gram es la matriz unidad, de modo que la condición se simplifica: $A^T \cdot A = I$, así que la matriz A tiene a su traspuesta por inversa.

Definición.- Una matriz cuya inversa es su traspuesta se llama *matriz ortogonal.*

Así pues, las aplicaciones ortogonales vienen representadas en bases ortonormales por matrices ortogonales. El estudio de las matrices ortogonales es un capítulo interesante del Álgebra que no vamos a abordar.

Ejercicio.- Compruebe que una matriz es ortogonal cuando sus vectores columnas son unitarios y ortogonales dos a dos. Discuta si se puede decir lo mismo de las filas.

Proposición.- El determinante de una matriz ortogonal es 1 o -1.

Demostración: Puesto que el determinante de una matriz es igual al de su traspuesta, que el determinante del producto de dos matrices es igual al producto de sus determinantes, y que el determinante de la matriz unidad es 1, se deduce de $A^T \cdot A = I$ que $|A|^2 = 1$, luego $|A| = \pm 1$.

Definición.- Una transformación ortogonal cuyo determinante (el de su matriz en una base ortonormal) es igual a 1 se llama una *aplicación ortogonal directa*. Una *aplicación ortogonal inversa* es aquella cuyo determinante es igual a -1.

Ejemplo.- La fórmula $f(x, y) = (-y, x)$ define una aplicación ortogonal en $\mathbf{R}^2$. Lo mismo podemos decir de $g(x, y) = (y, x)$. Aquélla es directa, mientras que ésta es inversa. Geométricamente, f representa un giro de 90° y g una simetría. Sus matrices en la base canónica son:

$$F = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

$$G = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

Los autovalores de G son 1 y -1, con autovectores ortogonales, en tanto que f no tiene autovalores reales (sí los tiene imaginarios, y no son otros que $\imath$ y $-\imath$, que son conjugados y de módulo 1).

Nada en ese ejemplo es casual. Todos los autovalores de una aplicación ortogonal han de tener módulo 1, puesto que si fuese otro el módulo alargaría o acortaría los vectores en esa proporción; por tanto, si son reales sólo pueden ser 1 y -1. Por otra parte, los autovectores asociados a autovalores distintos han de ser ortogonales, pues si $f(u) = \lambda u$ y $g(v) = \mu v$, entonces λ y μ serán 1 y -1, ya que son distintos, por lo cual $< u, v >=< f(u), f(v) >=< \lambda u, \mu v >= \lambda\mu < u, v >= - < u, v >$ de donde se sigue que $< u, v >= 0$.

Ejercicio.- Demuestre que una transformación ortogonal es un isomorfismo, y que su inversa también es una transformación ortogonal.

Ejercicio.- Demuestre que la composición de dos transformaciones ortogonales es una transformación ortogonal.

Transformaciones ortogonales en $\mathbf{R}$, $\mathbf{R}^2$ y $\mathbf{R}^3$.

Las únicas transforamciones ortogonales en $\mathbf{R}$ son $f(x) = x$ y $f(x) = -x$: la identidad (que es directa) y la simetría respecto del origen (que es inversa).

En $\mathbf{R}^2$, se nos ocurren varias transformaciones ortogonales: giros alrededor del origen y simetrías respecto de una recta. Veamos que no hay otras.

Si f es una aplicación ortogonal inversa en $\mathbf{R}^2$, entonces sus dos autovalores han de ser reales (si fueran complejos, serían conjugados y su producto sería positivo, con lo que f no sería inversa) y al ser su producto igual a -1 sólo pueden ser 1 y -1. En la base ortonormal formada por autovectores correspondientes, la matriz de f es

$$F = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

por lo que, en esa base, f se expresa como $f(x, y) = (x, -y)$, revelando que se trata de una simetría respecto del primer autovector. Así pues, las únicas transformaciones ortogonales inversas en $\mathbf{R}^2$ son las simetrías.

Si f es directa, puede tener el autovalor doble 1 (en cuyo caso f es la identidad), el autovalor doble -1 (y f será la opuesta de la identidad, o sea, el giro de 180º) o no tener autovalores reales. En este último caso, si la matriz de f en la base canónica es

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

entonces se han de dar las siguientes igualdades (para que A sea una matriz ortogonal y su determinante sea 1):

$a^2 + c^2 = 1$, $b^2 + d^2 = 1$, $ab + cd = 0$, $ad - bc = 1$.

De las dos primeras, deducimos que habrá unos ángulos, digamos φ, ψ tales que $a = \cos\varphi, c = \sin\varphi$, $d = \cos\psi, b = \sin\psi$; la última nos dice que $\cos(\varphi + \psi) = 1$, luego los ángulos φ, ψ son opuestos. Así pues, la matriz A es

$$A = \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix}$$

que representa un giro de amplitud φ alrededor del origen.

De lo anterior se deduce inmediatamente que la composición de dos simetrías en el plano es un giro, y la composición de un giro y una simetría es otra simetría, puesto que al componer dos transformaciones ortogonales inversas obtendremos una transformación ortogonal directa y al componer una aplicación ortogonal directa con una inversa resulta una inversa. El hecho es casi obvio intuitivamente, pero la clasificación que acabamos de hacer no deja lugar a la menor duda.

Un estudio análogo de la situación en $\mathbf{R}^3$ nos lleva a concluir que siempre habrá una recta invariante, y que las transformaciones directas son los giros (con esa recta invariante como eje) y las transformaciones inversas son las simetrías respecto de un plano o la composición de una simetría con un giro alrededor de la recta invariante (que será perpendicular al plano). Nótese que una simetría respecto de una recta es un giro de 180º alrededor de ella, y que una simetría respecto del origen es un giro seguido de una simetría respecto a un plano.

Como la composición de dos transformaciones directas es otra transformación directa, podemos deducir que la composición de dos giros en $\mathbf{R}^3$ es otro giro en el espacio, cuyo eje y amplitud pueden ser más o menos difíciles de intuir, pero resultan sencillos de calcular: basta con hallar los autovectores asociados al autovalor 1 y calcular el efecto que tiene la composición en un vector perpendicular al eje de giro. Veamos un ejemplo sencillo:

Las matrices

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}$$

$$B = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

describen sendos giros de amplitudes 90º y 270º alrededor de los ejes OX y OZ respectivamente. Su composición está descrita por la matriz producto

$$AB = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & -1 \\ -1 & 0 & 0 \end{pmatrix}$$

que es una matriz ortogonal directa. ¿Qué giro representa? Es claro que un autovector asociado al autovalor 1 es el $(1, 1, -1)$, lo que hace ver que el eje de giro es la recta $x = y = -z$. El vector $(0, 1, 1)$ es perpendicular a esa recta y se transforma en en $(1, -1, 0)$. El ángulo que forman ambos vectores se calcula mediante el producto escalar y resulta ser de 120º, algo que se ha calculado sin apenas esfuerzo pero que acaso resultara poco intuitivo. He aquí un ejemplo de la potencia del cálculo matricial.

Como ejercicio, calcule el resultado de realizar ambos giros en el otro orden.

Base recíproca. Coordenadas covariantes.

Recordemos que si $B = \{v_1, \ldots, v_n\}$ es una base ortonormal de un eve V, entonces las coordenadas de los vectores en B se expresan de manera muy sencilla: si $u = \sum x^j \cdot v_j$, entonces $x^j =< u, u_j >$, lo que constituye una notable ventaja.

Si la base B no es ortonormal, no podemos calcular las coordenadas con esa facilidad, pero existe otra base, $B^* = \{v^1, \ldots, v^n\}$ que hace esa tarea: si $u = \sum x^j \cdot v_j$, entonces $x^j =< u, u^j >$.

De momento es dudoso que sea así (lo justificaremos enseguida), pero es fácil comprobar que la condición de que las coordenadas se calculen por ese sencillo procedimiento equivale a que los productos escalares entre los vectores de las bases B y B^* sean $< v^j, v_k >= \delta_j^k$ (hágalo, si le pica la curiosidad). Recuerdo que el signo δ_j^k, conocido como delta de Kronecker, vale 1 cuando $j = k$ y 0 cuando $j \neq k$.

Visto así, es fácil demostrar que la base B^* existe y además una base con esas condiciones es única. La razón es que un vector ortogonal a $v_2, \ldots, v_n$ estará en el espacio ortogonal a ellos, que es de dimensión 1, lo que deja un grado de libertad para el vector v^1; al imponer que su producto escalar por v_1 sea igual a 1, queda determinado el vector. Del mismo modo están determinados los vectores $v^2, \ldots, v^n$. También es fácil comprobar que los vectores $v^1, \ldots, v^n$ son linealmente independientes, por lo que forman una base de V, que recibe el nombre de *base recíproca* de la base dada. Lo destacamos:

Definición.- Dada una base $B = \{v_1, \ldots, v_n\}$ en un eve, llamamos *base recíproca* de B a la base $B^* = \{v^1, \ldots, v^n\}$ determinada por las condiciones $< v^j, v_k >= \delta_j^k$.

Observación.- La base recíproca de una base ortonormal es ella misma. Recíprocamente, si la base recíproca de B es B, entonces es que B es ortonormal.

Otra observación.- Si B^* es la base recíproca de B, entonces ésta es la base recíproca de aquélla. Podemos escribir $B^{**} = B$. La razón es clara: los vectores de B cumplen exactamente la condición que se exige a los de B^{**}, que los productos escalares con los de B^* sean las deltas de Kronecker.

Ejemplo.- En $\mathbf{R}^2$, la base recíproca de $\{(2,1),(3,2)\}$ es $\{(2,-3),(-1,2)\}$, porque $(2,1) \cdot (2,-3) = 1, (2,1) \cdot (-1,2) = 0, (3,2) \cdot (2,-3) = 0, (3,2) \cdot (-1,2) = 1$. Y la base recíproca de $\{(2,-3),(-1,2)\}$ es $\{(2,1),(3,2)\}$, por la misma razón.

Ejercicio.- Calcule la base recíproca de $\{(0,1,1),(1,0,1),(1,1,0)\}$ (en $\mathbf{R}^3$, con el producto escalar usual).

Hablar de una base, B, y otra más, B^*, es hablar de la matriz de paso: ¿cómo se expresa el cambio de base matricialmente?, o sea, ¿cómo se relacionan las coordenadas de un vector en la primera base con las del mismo vector en la segunda? Aquí está la respuesta:

Proposición.- Si G es la matriz de Gram del producto escalar en la base B y B^* es la base recíproca, entonces la matriz de paso $P = (B : B^*)$ que nos da los vectores de la base recíproca en función de los de la primitiva es la matriz inversa de G: $P = G^{-1}$.

Demostración: Hay varias maneras de demostrarlo. Una muy sencilla es la siguiente:

La matriz de paso de B a B^* será la inversa de la matriz P, y tendrá en la k-ésima columna las coordenadas de v_k en la base B^*, es decir, a_{jk} =j-ésima coordenada de v_k respecto de B^*. Dicho de otra manera, $v_k = \sum a_{jk} \cdot v^j$. Pero las coordenadas de un vector en la base B^* se calculan multiplicando escalarmente por los vectores de su base recíproca, que es B, por lo que $a_{jk} =< v_k, v_j >$, y ese es justamente el elemento que ocupa el lugar jk en la matriz de Gram, G. Por tanto, $P^{-1} = (B^* : B) = G$ y $P = G^{-1}$.

Ejemplo.- La matriz de paso de la base $B = \{(2,1),(3,2)\}$ a su recíproca $\{(2,-3),(-1,2)\}$ es (calcúlese):

$$P = \begin{pmatrix} 5 & 8 \\ 8 & 13 \end{pmatrix}$$

que es precisamente la matriz de Gram del producto escalar en la base B.

Nota.- Así pues, la matriz G^{-1} nos permite calcular la base recíproca a partie de la base B. De esta manera, resulta muy fácil encontrar la base recíproca de una base dada.

Ejemplo.- Si queremos hallar la base recíproca de $B = \{v_1 = (1,1), v_2 = (2,1)\}$, empezamos calculando la matriz de Gram y su inversa en esa base:

$$G = \begin{pmatrix} 2 & 3 \\ 3 & 5 \end{pmatrix}$$

$$P = G^{-1} = \begin{pmatrix} 5 & -3 \\ -3 & 2 \end{pmatrix}$$

Las columnas de esta matriz, P, nos proporcionan las coordenadas de los vectores de la base recíproca B^* en la base B; así, el primer vector será $v^1 = 5v_1 - 3v_2 = (-1,2)$ y el segundo, $v^2 = -3v_1 + 2v_2 = (1,-1)$.

Ejercicio.- Calcule la base recíproca de $\{(-1,1,1),(1,-1,1),(1,1,-1)\}$ (en $\mathbf{R}^3$, con el producto escalar usual). Úsela para calcular las coordenadas de un vector arbitrario en la base dada. (Solución: las coordenadas del vector (x,y,z) en la base recíproca son $(y+z)/2, (x+z)/2$ y $(x+y)/2$).

Ejercicio.- Si G es la matriz de Gram del producto escalar en la base B, entonces la matriz de Gram del producto escalar en la base recíproca, B^*, es la matriz G^{-1}.

Coordenadas covariantes y contravariantes de un vector en una base.- Dada una base B de un eve, las coordenadas de un vector v respecto de esa base se llaman a veces *coordenadas contravariantes* del vector, y se da el nombre de *coordenadas covariantes* a las coordenadas del vector en la base recíproca.

Notación: Suelen usarse subíndices para las coordenadas covariantes y superíndices para las contravariantes. Del mismo modo, los vectores de la base inicial se suelen denotar con subíndices y los de la recíproca con superíndices. Así, escribimos $B = \{v_1, \dots, v_n\}$, $B^* = \{v^1, \dots, v^n\}$, $v = x_1 v^1 + \dots + x_n v^n = x^1 v_1 + \dots + x^n v_n$, siendo x^k las coordenadas contravariantes de v en las base B y x_k las covariantes.

Ni que decir tiene que las coordenadas covariantes de un vector en una base son las contravariantes en la base recíproca y viceversa.

Ejemplo.- En la base $B = \{(2,1),(3,2)\}$, el vector $v = (2,-1)$ tiene las coordenadas contravariantes 7 y -4 (porque $7(2,1) - 4(3,2) = (2,-1)$) y las coordenadas covariantes 3 y 4 (puesto que $3(2,-3) + 4(-1,2) = (2,-1)$).

Observación.- Las coordenadas covariantes de un vector v respecto de una base B son extraordinariamente fáciles de calcular: basta con multiplicar escalarmente v por los vectores de B (puesto que esas coordenadas covariantes no son más que las coordenadas contravariantes de v respecto de B^*, que se obtienen multiplicando escalarmente por los vectores de la base recíproca de B^*, o sea, de B). Por otra parte, el cambio de las coordenadas contravariantes a las covariantes lo da por la matriz de Gram (pues esa matriz representa el cambio de la base B a su recíproca).

Ejemplo.- Las coordenadas covariantes del vector $v = (2,-1)$ en la base $B = \{(2,1),(3,2)\}$, son $3 = (2,-1)\cdot(2,1)$ y $4 = (2,-1)\cdot(3,2)$.

La matriz de Gram del producto escalar usual en la base $B = \{(2,1),(3,2)\}$, es

$$G = \begin{pmatrix} 5 & 8 \\ 8 & 13 \end{pmatrix}$$

Si multiplicamos por el vector columna $(7,-4)^T$ de las coordenadas contravariantes del vector $(2,-1)$, nos da el vector columna $(3,4)$, que son sus coordenadas covariantes.

Dedicaré un último párrafo a explicar el curioso nombre que se da a esas coordenadas: unas se llaman contravariantes porque cambian "al revés" que los vectores de la base B, mientras que las otras lo hacen "igual" que ellos, y por eso se llaman covariantes. Un ejemplo ayudará:

Si $C = \{e_1, e_2\}$ es la base canónica de $\mathbf{R}^2$, $B = \{v_1 = (2,1), v_2 = (3,2)\}$, entonces sus bases recíprocas son $C^* = \{e^1 = e_1, e^2 = e_2\} = C$ (por ser C ortonormal) y $B^* = \{v^1 = (2,-3), v^2 = (-1,2)\}$. Sea $v = (a,b)$ un vector cualquiera de $\mathbf{R}^2$; sus coordenadas covariantes y contravariantes en C son $x^1 = x_1 = a, x^2 = x_2 = b$, mientras que en B las coordenadas contravariantes serán $y^1 = 2a - 3b, y^2 = -a + 2b$ y las covariantes $y_1 = 2a + b, y_2 = 3a + 2b$ (compruébelo). La matriz de paso

$$P = (C : B) = \begin{pmatrix} 2 & 3 \\ 1 & 2 \end{pmatrix}$$

nos permite expresar cómodamente la transformación de unas coordenadas en otras y de unos vectores en otros. Así, el cambio de los vectores se puede escribir $(v_1, v_2) = (e_1, e_2) \cdot P$, mientras que el cambio de las coordenadas contravariantes se expresa como $P \cdot v_B = v_C$, o sea, $P \cdot (y^1, y^2)^T = (x^1, x^2)^T$, o si se prefiere $(y^1, y^2) = (x^1, x^2) \cdot P^{-1T}$; donde se ve que el cambio es distinto del que efectúan los vectores. Por el contrario, las coordenadas covariantes responden a la fórmula $(y_1, y_2) = (x_1, x_2) \cdot P$, análoga a la del cambio de los vectores de la base; de ahí su nombre. Si nos fijamos en cómo cambian los vectores de las bases recíprocas, observamos que siguen la pauta de las coordenadas contravariantes: $P \cdot (v^1, v^2)^T = (e^1, e^2)^T$.

La idea de emplear subíndices en unos casos y superíndices en otros sirve para recordar el tipo de objetos que manejamos (si son covariantes o contravariantes) y para que al efectuar los productos y las combinaciones lineales se multipliquen factores subindicados por otros superindicados, favoreciendo la memoria y un útil convenio para abreviar la notación (eliminando el signo $\sum$ de suma).

Comentario final: diagonalización de formas cuadráticas en espacios vectoriales euclídeos.- Un teorema que no demostraremos garantiza que dada una forma cuadrática en un espacio vectorial euclídeo se puede encontrar una base de vectores conjugados para la forma cuadrática que además es base ortonormal del eve. En términos matriciales, dada una matriz simétrica, A, existe una matriz ortogonal, P, tal que $P^T \cdot A \cdot P = P^{-1} \cdot A \cdot P = D$ es una matriz diagonal. Decimos que A se diagonaliza simultáneamente por congruencia y por semejanza. Puesto que la diagonal de la matriz D está formada por los autovalores de A, podemos estudiar una forma cuadrática (en particular, su rango y su signatura) por medio de los autovalores de su matriz en alguna base. De todos modos, suele ser más sencillo estudiarla por el procedimiento elemental que se expuso en el capítulo anterior.