14. Forma Canónica de Jordan

Surge ahora la pregunta sobre cuál es la forma más simple en que pueden escribirse los operadores (o matrices) no diagonalizables. El siguiente teorema nos da la respuesta:
Teorema: Sea $F : V \to V$ un operador lineal en un espacio vectorial V de dimensión finita n sobre $\mathbb{C}$. Entonces existe una base $e = (e_{11}, e_{12}, \ldots, e_{1d_1}, \ldots, e_{k1}, e_{k2}, \ldots, e_{kd_k})$, con $\sum_{i=1}^{k} d_i = n$, en la que

$$\begin{array}{rcl} F(e_{i1}) & = & \lambda_i e_{i1} \\ F(e_{ij}) & = & \lambda_i e_{ij} + e_{i,j-1}, \;\; j = 2, \ldots, d_i \end{array} \;,\;\; i = 1, \ldots, k$$

o sea, $F(e_{11}) = \lambda_1 e_{11}$, $F(e_{12}) = \lambda_1 e_{12} + e_{11}$, $\ldots$, $F(e_{1d_1}) = \lambda_1 e_{1d_1} + e_{1,d_1-1}$ y similar para $i \geq 1$. El caso diagonalizable corresponde a $d_i = 1 \;\forall\; k$, en cuyo caso $k = n$. Los parámetros λ_i no son necesariamente distintos y son los autovalores de F, como demostraremos a continuación.
La matriz $[F]_e \equiv [F]_e^e$ en esta base toma entonces la forma de bloques

$$[F]_e = A = \begin{pmatrix} A_1 & 0 & \ldots & 0 \\ 0 & A_2 & \ldots & 0 \\ & & \ldots & \\ 0 & 0 & \ldots & A_k \end{pmatrix}$$

donde A_i son matrices de $d_i \times d_i$ de la forma

$$A_i = \begin{pmatrix} \lambda_i & 1 & 0 & \ldots & 0 \\ 0 & \lambda_i & 1 & \ldots & 0 \\ & & & \ldots & \\ 0 & 0 & 0 & \ldots & 1 \\ 0 & 0 & 0 & \ldots & \lambda_i \end{pmatrix} = \lambda_i I_{d_i} + J_{d_i}\,, \quad J_{d_i} = \begin{pmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ & & & \ldots & \\ 0 & 0 & 0 & \ldots & 1 \\ 0 & 0 & 0 & \ldots & 0 \end{pmatrix}$$

con I_{d_i} la matriz identidad de $d_i \times d_i$.
Cada subespacio $S_i = \overline{(e_{i1}, e_{i2}, \ldots, e_{id_i})}$ es claramente invariante por F, ya que $F(e_{ij}) \in S(\lambda_i)$.
Es claro también que los escalares λ_i, $i = 1, \ldots, k$, son los autovalores de F, pues

$$P(\lambda) = \mathrm{Det}[F - \lambda I] = |A_1 - \lambda I_{d_1}| \ldots |A_k - \lambda I_{d_k}| = (\lambda_1 - \lambda)^{d_1} \ldots (\lambda_k - \lambda)^{d_k}$$

posee como únicas raíces a $\lambda_1, \ldots, \lambda_k$.

Cada submatriz A_i posee un único autovalor λ_i de multiplicidad d_i ($|A_i - \lambda I_{d_i}| = (\lambda_i - \lambda)^{d_i}$), pero el espacio propio correspondiente es *de dimensión 1*: dim $N[A_i - \lambda_i I_{d_i}] =$ dim $N[J_{d_i}] = 1$ pues Rango$(J_{d_i}) = d_i - 1$. Por lo tanto, la submatriz A_i no es diagonalizable si $d_i > 1$. Cada subespacio S_i contiene entonces *un único subespacio propio* de dimensión 1, que es el generado por e_{i1}, y el número total de autovectores LI de F es $k \leq n$ (uno por cada S_i).

Notemos que $(F - \lambda_i I)e_{ij} = e_{i,j-1}$ para $j > 1$, con $(F - \lambda_i I)e_{i1} = 0$, por lo que aplicando m veces el op. $(F - \lambda_i I)$ sobre e_{ij} resulta

$$(F - \lambda_i I)^m e_{ij} = \left\{ \begin{array}{cc} e_{i,j-m} & m < j \\ 0 & m \geq j \end{array} \right.$$

Los operadores no diagonalizables en espacios finitos se caracterizan pues por la existencia de vectores e_{ij} no nulos tales que $(F - \lambda_i I)^j e_{ij} = 0$ pero $(F - \lambda_i I)e_{ij} \neq 0$ si $j > 1$. Si F es diagonalizable tales vectores no existen. Notemos que conociendo e_{id_i}, los restantes vectores e_{ij} pueden obtenerse como

$$e_{ij} = (F - \lambda_i I)^{d_i - j} e_{id_i} = (F - \lambda_i I)e_{i,j+1}, \;\; j = 1, \ldots, d_i - 1$$

La ecuación previa implica también que $(F - \lambda_i I)^{d_i}(e_{ij}) = 0, \quad j = 1, \ldots, d_i$. Por lo tanto, la matriz $J_{d_i} = A_i - \lambda_i I_{d_i}$ es *nilpotente*:

$$(J_{d_i})^{d_i} = 0$$

donde 0 denota la mariz nula de $d_i \times d_i$. Por ejemplo,

$$J_2 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \;\; J_2^2 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \;\; J_3 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \;\; J_3^2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \;\; J_3^3 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

La evaluación de potencias del operador puede entonces realizarse sin mayor dificultad, ya que

$$A^m = \begin{pmatrix} A_1^m & 0 & \dots & 0 \\ 0 & A_2^m & \dots & 0 \\ & & \dots & \\ 0 & 0 & \dots & A_k^m \end{pmatrix}$$

donde, teniendo en cuenta que I_{d_i} conmuta con J_{d_i},

$$A_i^m = (\lambda_i I_{d_i} + J_{d_i})^m = \lambda_i^m I_{d_i} + m\lambda_i^{m-1} J_{d_i} + \frac{m(m-1)}{2!}\lambda_i^{m-2} J_{d_i}^2 + \dots + J_{d_i}^m$$

para $i = 1, \dots, k$. Esta expansión contiene a lo sumo d_i términos ya que $(J_{d_i})^r = 0$ si $r \geq d_i$.
En general, si $p(t)$ es un polinomio de grado m,

$$p(t) = p(\lambda_i)1 + p'(\lambda_i)(t-\lambda_i) + \dots + \frac{p^{(m)}(\lambda_i)}{m!}(t-\lambda_i)^m$$

se obtiene

$$p(A_i) = p(\lambda_i) I_{d_i} + p'(\lambda_i) J_{d_i} + \dots + \frac{p^{(d_i-1)}(\lambda_i)}{(d_i-1)!} J_{d_i}^{d_i-1}$$

Además, la forma de Jordan es muy conveniente para la evaluación de exponenciales:

$$\exp[A_i t] = \exp[\lambda_i I_{d_i} t + J_{d_i} t] = \exp[\lambda_i I_{d_i} t]\exp[J_{d_i} t] = e^{\lambda_i t}\left(I_{d_i} + J_{d_i} t + \dots + J_{d_i}^{d_i-1}\frac{t^{d_i-1}}{(d_i-1)!}\right)$$

Por lo tanto, $B(t) = \exp[A_i t]$ será una matriz triangular con elementos $B_{kj} = e^{\lambda_i t} t^{j-k}/(j-k)!$ si $k \leq j$ y $B_{kj} = 0$ si $k > j$.

Para obtener la representación de Jordan se puede, una vez obtenidos los k autovalores λ_i y autovectores e_{i1}, $i = 1, \dots, k$, resolver las ecuaciones inhomogéneas $F(e_{ij}) = \lambda_i e_{ij} + e_{i,j-1} \quad j = 2, \dots, d_i$, es decir, trabajando en forma matricial en la base e,

$$AX_{i1} = \lambda_i X_{i1}, \;\; AX_{ij} = \lambda_i X_{ij} + X_{i,j-1}, \quad j = 2, \dots, d_i, \;\; i = 1, \dots, k$$

que no poseen solución única. Otra forma más eficiente es partir de e_{id_i}, es decir, encontrar un vector X_{id_i} que satisfaga

$$(A - \lambda_i I)^{d_i} X_{id_i} = 0, \;\; (A - \lambda_i I)^{d_i-1} X_{id_i} \neq 0$$

Los vectores restantes del bloque pueden obtenerse como

$$X_{ij} = (A - \lambda_i I)^{d_i - j} X_{id_i}, \;\; j = 1, \dots, d_i - 1$$

Ejemplo:

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{pmatrix}$$

Tenemos $|A - \lambda I_3| = (1-\lambda)(2-\lambda)^2$, por lo que las raíces son $\lambda_1 = 1$, con multiplicidad 1, y $\lambda_2 = 2$, con multiplicidad 2. Como

$$A - 2I_3 = \begin{pmatrix} -1 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

posee rango 2, $N[A - 2I_3]$ posee dimensión 1, por lo que A no es diagonalizable.
Para $\lambda_2 = 2$ el sistema homogéneo $(A - \lambda I_3)X = 0$ posee la solución general $x = y$, $z = 0$, de modo que el autovector es de la forma $x(1,1,0)^t$. Eligiendo $X_{11} = (1,1,0)^t$, el vector X_{12} puede obtenerse resolviendo

$$\begin{pmatrix} -1 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$$

que da como resultado $z=1$, $x=y$. Podemos elegir entonces $X_{12}=(0,0,1)^t$. Finalmente, $A-1I_3=\begin{pmatrix}0&1&1\\0&1&1\\0&0&1\end{pmatrix}$, por lo que $(A-I_3)X_{31}=0$ conduce a $X_{31}=x(1,0,0)^t$. Obtenemos entonces

$$S=\begin{pmatrix}1&0&1\\1&0&0\\0&1&0\end{pmatrix},\quad \text{con } S^{-1}=\begin{pmatrix}0&1&0\\0&0&1\\1&-1&0\end{pmatrix}$$

y finalmente la forma de Jordan

$$A'=S^{-1}AS=\begin{pmatrix}2&1&0\\0&2&0\\0&0&1\end{pmatrix}$$

Puede comenzarse también determinando X_{12} a partir de las condiciones $(A-2I_3)X_{12}\neq 0$, $(A-2I_3)^2X_{12}=0$, y obtener luego X_{11} como $(A-2I_3)X_{12}$ (hecho así en clase).

Se obtiene también

$$\exp[A't]=\begin{pmatrix}\exp[t\begin{pmatrix}2&1\\0&2\end{pmatrix}]&0\\0&\exp(t)\end{pmatrix}=\begin{pmatrix}e^{2t}&te^{2t}&0\\0&e^{2t}&0\\0&0&e^t\end{pmatrix}$$

ya que $\exp[t\begin{pmatrix}2&1\\0&2\end{pmatrix}]=\exp[t(2I_2+J_2)]=\exp[2tI_2]\exp[tJ_2]=e^{2t}(I_2+tJ_2)=e^{2t}\begin{pmatrix}1&t\\0&1\end{pmatrix}$, pues $J_2^2=0$.

Resolución general de sistemas de ecuaciones lineales de primer orden. La solución del sistema

$$\frac{dX}{dt}=AX$$

con A constante pero no diagonalizable, puede obtenerse a partir de la forma canónica de Jordan. Tenemos, para $X(0)=X_0$ y $A=SA'S^{-1}$,

$$X=\exp[At]X_0=S\exp[A't]C=\sum_{i=1}^{k}e^{\lambda_i t}\sum_{m=1}^{d_i}c_{im}\sum_{j=0}^{m-1}V_{i,m-j}\frac{t^j}{j!}$$

donde $V_{i,m-j}$ denota las columnas de S [$S=(V_{11},V_{12},\ldots,V_{1,d_1},\ldots,V_{k,1},\ldots,V_{k,d_k})$, de forma que $S\exp[A't]=(e^{\lambda_1 t}V_{11},e^{\lambda_1 t}(V_{12}+tV_{11}),e^{\lambda_1 t}(V_{13}+tV_{12}+\frac{t^2}{2!}V_{11}),\ldots)$] y $C=S^{-1}X_0=(c_{11},c_{12},\ldots,c_{1d_1},\ldots)^t$ un vector de constantes determinadas por el vector inicial $X_0=X(0)$. Por ejemplo,

$$X=e^{\lambda t}[c_1V_1+c_2(V_2+V_1t)+c_3(V_3+V_2t+V_1t^2/2!)+\ldots]$$

en el caso de un solo bloque, con $V_j=(A-\lambda I)^{d-j}V_d$ y d la dimensión del bloque.

15. Polinomios Anuladores y Teorema de Cayley-Hamilton

Sea $F:V\to V$ un operador lineal en un espacio vectorial V de dimensión finita n.
Si $p(\lambda)=a_0+a_1\lambda+\ldots+a_m\lambda^m$ es un polinomio de grado m, podemos asociar a $p(\lambda)$ el operador lineal

$$p(F)=a_0I+a_1F+\ldots+a_mF^m$$

La matriz que representa al operador $p(F)$ en una base e es

$$\begin{aligned}[p(F)]_e &= a_0[I]_e+a_1[F]_e+\ldots+a_m[F^m]_e=a_0I_n+a_1[F]_e+\ldots+a_m[F]_e^m\\ &= p([F]_e)\end{aligned}$$

donde hemos asociado $F^0=I$ y $[F]_e^0=I_n$. Además, si escribimos $p(\lambda)$ en términos de sus m raíces λ_i

$$p(\lambda)=a_m(\lambda-\lambda_1)(\lambda-\lambda_2)\ldots(\lambda-\lambda_m)$$

entonces
$$p(F) = a_m(F-\lambda_1 I)(F-\lambda_2 I)\ldots(F-\lambda_m I)$$
ya que las potencias F^k de F conmutan todas entre si $\forall\ j\geq 0$.

Un polinomio p se dice que es *anulador* de F si $p(F)=0$ (o sea, si $p(F)$ es el operador nulo). Dado que la dimension del espacio de operadores lineales $H=\{F: V\to V,\ F \text{ lineal}\}$ en un espacio vectorial V de dimensión finita n es n^2, es claro que el conjunto $(I=F^0, F, F^2, \ldots, F^{n^2})$ es LD (pues son $n+1$). y que por lo tanto, existen siempre n^2+1 constantes $c_0,\ldots,c_n$ no todas nulas tales que $c_0 I + c_1 F + \ldots c_n F^{n^2} = 0$. Esto muestra en forma básica que siempre existe un polinomio anulador de F.

No obstante, el siguiente teorema (denominado teorema de Cayley-Hamilton) muestra que el mismo *polinomio característico asociado a F, que es de grado n, es siempre un polinomio anulador de F.*

Teorema: Si $p(\lambda)=\mathrm{Det}[F-\lambda I]$ es el polinomio característico de $F \Rightarrow p(F)=0$.
Como el polinomio característico es de grado n ($p(\lambda)=a_0+a_1\lambda+\ldots+a_n\lambda^n$ con $a_n=(-1)^n\neq 0$), el teorema implica que es siempre posible expresar F^n en términos de potencias de grado $\leq n-1$: Como $p(F)=0\Rightarrow$
$$F^n = -(a_0 + a_1 F + \ldots a_{n-1}F^{n-1})/a_n$$
Por lo tanto, *cualquier* potencia F^k con $k\geq n$ puede expresarse en términos de potencias de grado $\leq n-1$.

Demostraremos el teorema a partir de la forma canónica de Jordan. No obstante, en el caso en que F es diagonalizable, el teorema es obvio, ya que en tal caso existe una base e en la que $[F]_e$ es diagonal,
$$[F]_e = \begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ & & \ldots & \\ 0 & 0 & \ldots & \lambda_n \end{pmatrix}$$
donde λ_i, $i=1,\ldots,n$ son los autovalores de F, y
$$[p(F)]_e = p([F])_e = \begin{pmatrix} p(\lambda_1) & 0 & \ldots & 0 \\ 0 & p(\lambda_2) & \ldots & 0 \\ & & \ldots & \\ 0 & 0 & \ldots & p(\lambda_n) \end{pmatrix} \qquad (15.1)$$
para cualquier polinomio p. Pero si p es el polinomio característico, $p(\lambda_i)=0$ para $i=1,\ldots,n$ y por lo tanto $[p(F)]_e = p([F]_e)=0$ (matriz nula). Esto implica a su vez $p(F)=0$ (operador nulo).

En el caso general, utilizando la base en la que $[F]_e$ tiene la forma canónica de Jordan, tenemos, para cualquier polinomio p,
$$[p(F)]_e = p([F]_e) = \begin{pmatrix} p(A_1) & 0 & \ldots & 0 \\ 0 & p(A_2) & \ldots & 0 \\ & & \ldots & \\ 0 & 0 & \ldots & p(A_k) \end{pmatrix}$$
donde A_i, $i=1,\ldots,k$ son matrices de $d_i\times d_i$ que satisfacen $(A_i-\lambda_i I_{d_i})^{d_i}=0$ (matriz nula).
Recordemos ahora que el polinomio característico de F está dado por
$$p(\lambda) = |[F]_e - \lambda I_n| = |A_1-\lambda I_{d_1}|\ldots|A_k-\lambda I_{d_k}| = (\lambda_1-\lambda)^{d_1}\ldots(\lambda_k-\lambda)^{d_k}$$
Por lo tanto,
$$p(A_i) = (\lambda_1 I_{d_i} - A_i)^{d_1}\ldots(\lambda_i I_{d_i}-A_i)^{d_i}\ldots(\lambda_k I_{d_i}-A_i)^{d_k} = 0$$
pues $(\lambda_i I_{d_i}-A_i)^{d_i}=0$. Esto implica $[p(F)]_e=0$ y entonces $p(F)=0$. Se cumple pues, para cualquier base e', $p([F]_{e'}) = [p(F)]_{e'} = [0]_{e'} = 0$.

El teorema vale por consiguiente para matrices A generales de $n\times n$. Si $p(\lambda)=|A-\lambda I_n|$ es el polinomio característico asociado a A (de grado n) $\Rightarrow p(A)=0$ (la matriz nula de $n\times n$).
Para matrices A diagonalizables el resultado es evidente, ya que en tal caso $A=SA'S^{-1}$, con A' diagonal, y por lo tanto $p(A)=p(SA'S^{-1})=Sp(A')S^{-1}$, pero $p(A')$ tiene la forma 15.1 y es por lo tanto la matriz nula.

Escribiendo

$$p(F) = c_n F^n + c_{n-1}F^{n-1} + \ldots + c_1 F + c_0 I$$

en el caso del polinomio característico tenemos $c_n = (-1)^n \neq 0$ y $c_0 = \mathrm{Det}[F]$. Por lo tanto, como $p(F) = 0$, podemos escribir

$$F^n = -(c_{n-1}F^{n-1} + \ldots c_1 F + c_0 I)/c_n$$

de modo que F^n (y por lo tanto cualquier potencia F^k, con $k \geq n$ natural) puede escribirse en términos de los operadores $F^{n-1}, \ldots, F, I$. Más aún, si F es invertible, $c_0 \neq 0$ y multiplicando la expresión anterior por F^{-1} se obtiene

$$F^{-1} = -(c_n F^{n-1} + c_{n-1}F^{n-2} + \ldots c_1 I)/c_0$$

de modo que también F^{-1} (y por tanto cualquier potencia F^{-k}, $k > 0$ natural) puede escribirse en términos de $F^{n-1}, \ldots, F, I$.

Cabe destacar que el polinomio característico no es necesariamente el polinomio anulador de grado mínimo. Sí lo es en el caso de autovalores todos distintos o, en general, en el caso de bloques de Jordan con autovalores todos distintos.
Si F es diagonalizable, el polinomio anulador de grado mínimo es simplemente $P_m(\lambda) = \prod_i (\lambda - \lambda_i)$, donde la productoria es sobre *autovalores distintos*.
En el caso general, el polinomio anulador de grado mínimo será $P_m(\lambda) = \prod_i (\lambda - \lambda_i)^{d_i}$, donde la productoria es nuevamente sobre autovalores distintos y d_i es la dimensión del mayor bloque de Jordan asociado a λ_i. $P_m(\lambda)$ es pues de grado $\leq n$.

Ejemplo 1: Sea

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

El polinomio característico es

$$p(\lambda) = |A - \lambda I_2| = \begin{vmatrix} -\lambda & 1 \\ 1 & -\lambda \end{vmatrix} = \lambda^2 - 1$$

Se cumple entonces

$$p(A) = A^2 - I_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} - \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

Esto muestra simplemente que $A^2 = I_2$ y que por lo tanto, $A^k = I_2$ si k es par y $A^k = A$ si k impar.

Ejemplo 2:

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{pmatrix}$$

El polinomio característico es

$$p(\lambda) = |A - \lambda I_3| = (1 - \lambda)(2 - \lambda)^2 = -\lambda^3 + 5\lambda^2 - 8\lambda + 4$$

El teorema implica entonces que

$$-A^3 + 5A^2 - 8A + 4I_3 = 0$$

donde $A^2 = A.A$, $A^3 = A.A.A$ (producto matricial), como es fácil verificar. Por lo tanto, $A^3 = 5A^2 - 8A + 4I_3$ y $A^{-1} = (A^2 - 5A + 8I)/4$. Cualquier potencia A^k con k entero puede expresarse en términos de I_3, A y A^2.

Ej. 3: Matriz A de $n \times n$ de rango 1, con $n \geq 2$. Dado que A y $A' = S^{-1}AS$ poseen el mismo rango y traza, tenemos, si A' es la forma canónica de Jordan de A, $r(A') = 1$ y $\mathrm{Tr}\, A' = \mathrm{Tr}\, A$, por lo que A' tiene sólo una fila no nula. Si $\mathrm{Tr}\, A \neq 0$, la única posibilidad es que A sea diagonalizable y tenga un único autovalor no nulo igual a $\mathrm{Tr}\, A$, siendo los restantes nulos, mientras que si $\mathrm{Tr}\, A = 0$, la única posibilidad para A' es un bloque de Jordan de dimensión 2 de la forma $\left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)$, siendo todos los autovalores nulos y A no diagonalizable. Toda matriz de rango 1 es necesariamente de la forma $A = bc^t$, con b y c vectores columna no nulos (de $n \times 1$), como el lector puede fácilmente demostrar, con $\mathrm{Tr}\, A = c^t b = \mathbf{c} \cdot \mathbf{b}$. Se deja como problema hallar los autovectores asociados y el polinomio minimal en ambos casos.

16. Demostración

Daremos aquí un resumen de la demostración de la forma canónica de Jordan. En primer lugar, sabemos que todo operador lineal $F : V \to V$, con V de dimensión finita n, posee un polinomio anulador $P(x)$, tal que $P(F) = 0$ (o sea, $P(F)(v) = 0 \ \forall\ v \in V$). Existirá entonces un polinomio anulador de grado mínimo $P_m(x) = a_0x + a_1x + \ldots + a_mx^m$ (polinomio minimal), tal que $P_m(F) = a_0F + a_1F + \ldots + a_mF^m = 0$.
1) λ es raíz de $P_m(F)$ si y sólo si λ es autovalor de F. Esto indica que las raíces del polinomio minimal y el polinomio característico son las mismas. Sólo la multiplicidad puede ser diferente.
Dem.: Si λ es autovalor de $F \Rightarrow \exists\ v \neq 0$ tal que $F(v) = \lambda v$, y en tal caso $P_m(F)(v) = P_m(\lambda)v = 0$, por lo que $P_m(\lambda) = 0$, es decir, λ es raíz de $P_m(x)$.
Si λ es raíz de $P_m(x) \Rightarrow P_m(x) = Q_{m-1}(x)(x - \lambda)$. En tal caso $P_m(F)(v) = Q_{m-1}(F)(F - \lambda I)(v) = 0$ $\forall\ v \in V$, por lo que necesariamente $\exists\ v \neq 0$ tal que $(F - \lambda I)(v) = 0$, es decir, λ es autovalor de F y v autovector asociado (si tal vector no existiese tendríamos $Q_{m-1}(F)(v) = 0 \ \forall\ v \in V$ y el polinomio minimal sería $Q_{m-1}(F)$, de grado $m - 1 < m$, en contradicción con la hipótesis).
2) Si $P_m(x) = Q_1(x)Q_2(x)$, con $Q_1(x)$ y $Q_2(x)$ polinomios sin raíces comunes y $P_m(F) = Q_1(F)Q_2(F) = 0$ $\Rightarrow V = N(Q_1(F)) \oplus N(Q_2(F))$, donde $N(Q_i(F))$ $(i = 1, 2)$ denota el núcleo de $Q_i(F)$. Los subespacios $N(Q_i(F))$ son además invariantes por F.
Dem.: Al no tener raíces comunes, existen polinomios $A_1(x)$, $A_2(x)$ t.q. $1 = A_1(x)Q_1(x) + A_2(x)Q_2(x)$, o sea,

$$I = A_1(F)Q_1(F) + A_2(F)Q_2(F)$$

Por lo tanto, $\forall\ v \in V$, $v = A_1(F)Q_1(F)(v) + A_2(F)Q_2(F)(v) = v_1 + v_2$, con $v_i = A_i(F)Q_i(F)$. Pero $v_1 \in N(Q_2(F))$ pues $Q_2(F)A_1(F)Q_1(F)(v) = A_1(F)Q_1(F)Q_2(F)(v) = 0$, y análogamente, $v_2 \in N(Q_1(F))$. Esto muestra que $V = N(Q_2(F)) + N(Q_1(F))$.
Además, si $v \in N(Q_1(F))$ y $v \in N(Q_2(F)) \Rightarrow v = A_1(F)Q_1(F)(v) + Q_2(F)Q_2(F)(v) = 0$. Esto muestra que $V = N(Q_1(F)) \oplus N(Q_2(F))$.
Finalmente, si $v \in N(Q_1(F)) \Rightarrow v = A_2(F)Q_2(F)(v)$ y en tal caso $Q_1(F)F(v) = A_2(F)Q_1(F)Q_2(F)F(v) = 0$, por lo que $F(v) \in N(Q_1(F))$. Análogamente, si $v \in N(Q_2(F)) \Rightarrow F(v) \in N(Q_2(F))$, por lo que ambos núcleos son invariantes por F.
3) Generalizando, si

$$P_m(x) = (x - \lambda_1)^{d_1} \ldots (x - \lambda_k)^{d_k}$$

con $\lambda_i \neq \lambda_j$ si $i \neq j$ (las raíces distintas de $P_m(x)$) y $P_m(F) = 0 \Rightarrow V = V_1 \oplus \ldots \oplus V_k$, con $V_i = N(F - \lambda_i I)^{d_i}$. El espacio completo V puede pues subdividirse en k subespacios invariantes, núcleos de $\tilde{F}_i^{d_i}$, donde $\tilde{F}_i = (F - \lambda_i I)$. Podemos pues construir una base de V formada por las bases de V_i.
4) Para construir una base de V_i, notemos que debe existir un vector $v \neq 0$ tal que $\tilde{F}_i^{d_i}(v) = 0$ pero $\tilde{F}_i^{d_i-1}(v) \neq 0$ (de lo contrario $P_m(F)$ no sería el polinomio minimal). En tal caso, los d_i vectores no nulos

$$e_{ij} = \tilde{F}_i^{d_i-j}(v), \quad j = 1, \ldots, d_i$$

(o sea $e_{id_i} = v$, $e_{ij} = \tilde{F}_i(e_{i,j+1})$, $j = 1, \ldots, d_i - 1$) son LI. Dem.: Si

$$c_1\tilde{F}_i^{d_i-1}(v) + c_2\tilde{F}_i^{d_i-2}(v) + \ldots + c_{d_i-1}\tilde{F}_i(v) + c_{d_i}v = 0$$

aplicando $\tilde{F}_i^{d_i-1}$ al segundo miembro obtenemos $c_{d_i}\tilde{F}_i^{d_i-1}(v) = 0$, por lo que $c_{d_i} = 0$. Luego, aplicando sucesivamente $\tilde{F}_i^j$, con $j = d_i - 1, \ldots, 0$, vemos que $c_j = 0$ para $j = 1, \ldots, d_i$. Notemos además que $\tilde{F}_i(e_{i1}) = \tilde{F}_i^{d_i}(v) = 0$, o sea, $(F - \lambda_i I)e_{i1} = 0$, por lo que e_{i1} es autovector de F con autovalor λ_i. Tenemos pues, en el subespacio S_i generado por los d_i vectores $B_i = (e_{i1}, \ldots, e_{id_i})$,

$$[\tilde{F}_i]_{B_i} = \begin{pmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ & & & \ldots & \\ 0 & 0 & \ldots & 0 & 1 \\ 0 & 0 & \ldots & 0 & 0 \end{pmatrix} = J_{d_i}, \quad \text{es decir,} \quad [F_i]_{B_i} = [\tilde{F}_i]_{B_i} + \lambda_I I_{d_i} = \begin{pmatrix} \lambda_i & 1 & 0 & \ldots & 0 \\ 0 & \lambda_i & 1 & \ldots & 0 \\ & & & \ldots & \\ 0 & 0 & \ldots & \lambda_i & 1 \\ 0 & 0 & \ldots & 0 & \lambda_i \end{pmatrix}$$

Se obtiene así un bloque de Jordan de dimensión d_i (grado del término correspondiente $(x - \lambda_i)^{d_i}$ del polinomio minimal).

Puede existir otro vector $v \in N(\tilde{F}_i^{d_i})$ tal que $\tilde{F}^{d_i-1}(v) \neq 0$ pero $\tilde{F}_i^{d_i}(v) = 0$ *y que no pertenezca al espacio generado por los vectores de* B_i. Este vector generaría otro bloque de Jordan de la misma dimensión con el mismo autovalor λ_i. En general, pueden surgir vectores $v \in N(\tilde{F}_i^{d_i})$ que no pertenezcan a los subespacios generados por el conjunto de vectores anteriores y que satisfagan $\tilde{F}_i^{r-1}(v) = 0$ pero $\tilde{F}_i^{r}(v) = 0$, con $r \leq d_i$, que generarán otros bloques de Jordan de dimensión $r \leq d_i$ con el mismo autovalor. La dimension total de $N(F - \lambda_i I)^{d_i}$ será así la multiplicidad $m_i \geq d_i$ de λ_i en el polinomio característico.
Si $d_i = 1$ los bloques son de dimensión 1 y los vectores correspondientes *autovectores* de F con autovalor λ_i. Este es el caso donde F es *diagonalizable* en el subespacio asociado a λ_i, es decir, donde la dimensión de $N(F - \lambda_i I)$ coincide con la multiplicidad de λ_i como raíz del polinomio característico.
Por lo tanto, si F es diagonalizable, el polinomio minimal es $P_m(x) = (x - \lambda_1) \ldots (x - \lambda_k)$.

17. Formas lineales, bilineales y cuadráticas

17.1 Formas lineales

Estudiaremos ahora funciones escalares lineales de argumento vectorial. Sea V un espacio vectorial sobre un cuerpo K. Una forma lineal es una función $F : V \to K$ que satisface las condiciones

$$F(\alpha v) = \alpha F(v) \;\; \forall v \in V,\; \alpha \in K \tag{1}$$
$$F(v_1 + v_2) = F(v_1) + F(v_2) \;\; \forall v_1, v_2 \in V \tag{2}$$

Una forma lineal puede ser considerada como un caso particular de transformación lineal si se considera el cuerpo K como un espacio vectorial de dimensión 1 sobre el mismo K. Nótese que se satisface $F(0) = 0$.
Ejemplos (se dejan las comprobaciones para el lector):
1) Si $K = \mathbb{R}$ y $V = \mathbb{R}^2$, $F(x, y) = x + y$ es claramente una forma lineal, mientras que $G(x, y) = x + y^2$ y $H(x, y) = 1 + x$ no son formas lineales.
2) Si $V = \mathbb{R}^{n\times n}$, la traza de una matriz $A \in V$, $\mathrm{Tr}[A] = \sum_{i=1}^n A_{ii}$, es una forma lineal.
3) Si $K = \mathbb{R}$ y $V = C_{[a,b]}$ (espacio de funciones continuas $f : [a, b] \to \mathbb{R}$),

$$T(f) = \int_a^b f(x)dx$$

es una forma lineal, y también lo es (para $\rho \in C_{[a,b]}$)

$$T_\rho(f) = \int_a^b f(x)\rho(x)dx\,.$$

4) En el mismo espacio anterior, y para $a < 0 < b$, $T(f) = f(0)$ es también una forma lineal. Nótese sin embargo que en este caso no existe $\rho(x)$ continua tal que $T(f) = \int_a^b f(x)\rho(x)dx$.
5) Si $V = \mathbb{R}^n$ y w es un vector fijo de $\mathbb{R}^n$,

$$F_w(v) = w \cdot v$$

(producto escalar usual) es una forma lineal. Por ej. el primer caso de 1), $F(x, y) = x + y$, puede ser escrito como $F(x, y) = (1, 1) \cdot (x, y)$. *Toda* forma lineal en $\mathbb{R}^n$ puede ser escrita de esta manera en términos de un único vector $w \in V$, como se verá a continuación.

Si $\dim V = n$ y F no es la forma lineal nula $\Rightarrow \dim I(F) = 1$, por lo que $\dim N(F) = n - 1$. Ejemplo: Hallar el núcleo de la forma lineal del ejemplo 2.

En un espacio vectorial V de dimensión finita n, la forma lineal queda completamente determinada por los valores que asigna a los elementos de una base: Si $B = (b_1, \ldots, b_n)$ es una base de V y $v = \sum_{i=1}^n \alpha_i b_i \Rightarrow$

$$F(v) = F(\sum_{i=1}^n \alpha_i b_i) = \sum_{i=1}^n \alpha_i F(b_i) = [F]_B[v]_B$$

donde

$$[F]_B = (F(b_1), \ldots, F(b_n))$$

es la matriz fila que representa a F en la base B y

$$[v]_B = \begin{pmatrix} \alpha_1 \\ \ldots \\ \alpha_n \end{pmatrix}$$

la matriz columna de coordenadas de v en dicha base. El producto $[F]_B[v]_B$ puede entonces visualizarse como el producto escalar usual de los vectores $[F]_B$ y $([v]_B)^t$ de K^n. En $V = \mathbb{R}^n$ toda forma lineal puede pues ser escrita en la forma $F(v) = w \cdot v$, con $w = [F]_e = (\beta_1, \ldots, \beta_n)$, siendo e la base canónica y $\beta_i = F(e_i) = w \cdot e_i$.
Frente a un cambio de base,

$$b'_i = \sum_{j=1}^n S_{ji} b_j, \quad i = 1, \ldots, n$$

con S una matriz no singular ($|S| \neq 0$) se obtiene $F(b'_i) = \sum_{j=1}^n S_{ji} F(b_j)$ y por lo tanto

$$[F]_{B'} = [F]_B S$$

con lo cual, dado que $[v]_B = S[v]_{B'}$, $F(v) = [F]_B[v]_B = [F]_B S[v]_{B'} = [F]_{B'}[v]_{B'}$.

Si $F: V \to K$ y $G: V \to K$ son dos formas lineales sobre V, la combinación lineal $\alpha F + \beta G$, definida por $(\alpha F + \beta G)(v) = \alpha F(v) + \beta G(v)$, es también una forma lineal $\forall\ \alpha, \beta \in K$, como es muy fácil comprobar. El conjunto de todas las formas lineales $F: \mathbb{V} \to K$ es pues un espacio vectorial denominado *espacio dual* V^*. Si V es de dimensión finita $\Rightarrow$

$$\dim V^* = \dim V$$

ya que existe un isomorfismo entre V^* y K^n (definido por $\mathcal{G}_B(F) = [F]_B \in K^n$, con $n = \dim V$) y por lo tanto entre V^* y V. Si $B = (b_1, \ldots, b_n)$ es una base ordenada de V, la base asociada de V^* es la base dual $B^* = \{F_1, \ldots, F_n\}$, donde $F_i: V \to K$ está definido por $F(\sum_i \alpha_i b_i) = \alpha_i$, es decir,

$$F_i(b_j) = \delta_{ij}\,.$$

F_i es pues representado por el vector fila $[F_i]_B = (0, \ldots, 1_i, \ldots, 0) = e_i$.

17.2 Formas bilineales

Una función escalar de dos variables vectoriales, $A: V \times V \to K$, es una forma bilineal si satisface

$$A(\alpha v, w) = \alpha A(v, w), \quad A(v_1 + v_2, w) = A(v_1, w) + A(v_2, w)\ \ \forall v, v_1, v_2, w \in V,\ \alpha \in K$$

$$A(v, \alpha w) = \alpha A(v, w), \quad A(v, w_1 + w_2) = A(v, w_1) + A(v, w_2)\ \ \forall w, w_1, w_2, v \in V,\ \alpha \in K$$

A es entonces una forma bilineal si es lineal con respecto a sus dos argumentos vectoriales.
Ejemplos (se dejan las comprobaciones como ejercicio):
1) Si $V = \mathbb{R}^2$ y $K = \mathbb{R}$, con $v = (x, y)$, $w = (z, t)$, las siguientes funciones son formas bilineales:

$$A(v, w) = v \cdot w = xz + yt \quad \text{(producto escalar)}$$

$$B(v, w) = xt - yz \ \ \text{(determinante de } \left(\begin{smallmatrix} x & y \\ z & t \end{smallmatrix}\right))$$

2) Si $V = C_{[a,b]}$ y $K = \mathbb{R}$, las siguientes funciones son formas bilineales:

$$A(f, g) = \int_a^b f(x)g(x)dx\,, \quad B(f, g) = \int_a^b f(x)g(x)\rho(x)dx\,, \quad C(f, g) = \int_a^b \int_a^b f(x)K(x, x')g(x')dxdx'$$

donde $\rho(x)$ y $K(x, x')$ son continuas.
3) En el mismo espacio anterior, para $a < 0 < b$, también son formas bilineales $T(f, g) = f(0)g(0)$ y $H(f, g) = f(0)g(c)$, con $c \in [a, b]$ fijo. Estas formas no pueden ser escritas en la forma integral del ejemplo anterior para ρ y K continuas.
4) En $V = \mathbb{R}^{n\times 1}$,

$$A(v, w) = v^t A w$$

donde $v^t = (\alpha_1, \ldots, \alpha_n)$, $w = \begin{pmatrix} \beta_1 \\ \ldots \\ \beta_n \end{pmatrix}$ y A es una matriz real de $n \times n$, es una forma bilineal. *Toda* forma bilineal en $\mathbb{R}^{n\times 1}$ (y por lo tanto $\mathbb{R}^n$) puede escribirse de esta manera, como se verá a continuación. Por ejemplo, las formas del ejemplo 1) pueden ser escritas como $A(v, w) = (x, y)\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)\left(\begin{smallmatrix} z \\ t \end{smallmatrix}\right)$, $B(v, w) = (x, y)\left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)\left(\begin{smallmatrix} z \\ t \end{smallmatrix}\right)$.

Representación matricial. En un espacio vectorial V de dimensión finita n, la forma bilineal queda completamente determinada por los valores que asigna a pares ordenados de elementos de una base $B = (b_1, \ldots, b_n)$ de V. Si $v = \sum_{i=1}^n \alpha_i b_i$, $w = \sum_{j=1}^n \beta_j b_j \Rightarrow$

$$A(v, w) = A(\sum_{i=1}^n \alpha_i b_i, \sum_{j=1}^n \beta_j b_j) = \sum_{i=1}^n \alpha_i A(b_i, \sum_{j=1}^n \beta_j ebj) = \sum_{i=1}^n \sum_{j=1}^n \alpha_i \beta_j A(b_i, b_j)$$

La igualdad anterior puede escribirse en forma compacta matricial como

$$A(v, w) = [v]_B^t [A]_B [w]_B$$

donde $[v]_B^t = (\alpha_1, \ldots, \alpha_n)$, $[w]_B = \begin{pmatrix} \beta_1 \\ \ldots \\ \beta_n \end{pmatrix}$ y

$$[A]_B = \begin{pmatrix} A(b_1, b_1) & \ldots & A(b_1, b_n) \\ & \ldots & \\ A(b_n, b_1) & \ldots & A(b_n, b_n) \end{pmatrix}$$

es la matriz de $n \times n$ que representa a la forma bilineal A en dicha base $[([A]_B)_{ij} = A(b_i, b_j)]$.
Por ej., si $V = \mathbb{R}^2$, $K = \mathbb{R}$ y e es la base canónica, obtenemos, para los casos del ejemplo 1) y $v = (x, y) = xe_1 + ye_2$, $w = (z, t) = ze_1 + te_2$,

$$A(v, w) = xz + yt = (x, y)[A]_e(\begin{smallmatrix} z \\ t \end{smallmatrix}), \;\; [A]_e = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$B(v, w) = xt - yz = (x, y)[B]_e(\begin{smallmatrix} z \\ t \end{smallmatrix}), \;\; [B]_e = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

Por otro lado, la matriz $[C]_e = (\begin{smallmatrix} 1 & 2 \\ 3 & 4 \end{smallmatrix})$ determina la forma bilineal

$$C(v, w) = \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} z \\ t \end{pmatrix} = xz + 2xt + 3yz + 4yt$$

Una forma bilineal es *simétrica* si

$$A(v, w) = A(w, v) \; \forall v, w \in V$$

y *antisimétrica*

$$A(v, w) = -A(w, v) \; \forall v, w \in V$$

Toda forma bilineal puede escribirse como suma de una forma bilineal simétrica y otra antisimétrica:

$$A(v, w) = \frac{A(v, w) + A(w, v)}{2} + \frac{A(v, w) - A(w, v)}{2} = A_s(v, w) + A_a(v, w)$$

El conjunto de formas bilineales de $V \times V$ sobre K forma un espacio vectorial W (con las operaciones usuales de suma y multiplicación por escalar) y la anterior descomposición corresponde a la suma directa $W = W_s \oplus W_a$, con W_s, W_a los subespacios de formas bilineales simétricas y antisimétricas sobre K.
En espacios V de dimensión finita, las correspondientes matrices en cualquier base dada son simétricas y antisimétricas respectivamente:

$$A(v, w) = A(w, v) \Rightarrow [A]_B^t = [A]_B, \;\; \text{pues } A(b_i, b_j) = A(b_j, b_i)$$

$$A(v, w) = -A(w, v) \Rightarrow [A]_B^t = -[A]_B, \;\; \text{pues } A(b_i, b_j) = -A(b_j, b_i)$$

En los ejemplos anteriores, el primero (producto escalar) es una forma bilineal simétrica mientras que el segundo (determinante) es una forma antisimétrica.

Dada una forma bilineal A arbitraria, notemos que $A(v, 0) = A(0, w) = 0 \; \forall \; v, w \in V$, como el lector podrá fácilmente demostrar. Si además existe $w \neq 0$ tal que $A(v, w) = 0 \; \forall \; v \in V$, la forma bilineal se dice que es **singular**. En caso contrario se dice **no singular**.
En un espacio V de dimensión finita, *A es singular si y sólo si la matriz que la representa en una base cualquiera, $[A]_B$, es singular.*
Dem.: Si $[A]_B$ es singular, existe un vector columna $[w]_B$ no nulo tal que $[A]_B[w]_B = 0$ y por lo tanto, $A(v, w) = [v]_B^t[A]_B[w]_B = [v]_B^t 0 = 0 \; \forall \; v \in V$.
Por otro lado, si existe $w \neq 0$ tal que $A(v, w) = 0 \; \forall v \in V$, y B es una base cualquiera de $V \Rightarrow [v]_B^t[A]_B[w]_B = 0 \; \forall$ vector $[v]_B^t \in K^{n \times 1}$, lo que implica $[A]_B[w]_B = 0$. Como $[w]_B \neq 0$, la matriz $[A]_B$ es entonces singular.
En espacios de dimensión finita, si $\exists \; w \; / \; A(v, w) = 0 \; \forall \; v \in V \Rightarrow \exists \; u \in V \; / \; A(u, v) = 0 \; \forall \; v \in V$, pues si $[A]_B$ es singular $\Rightarrow [A]_B^t$ es también singular $(|[A]_B^t| = |[A]_B| = 0)$.
Notemos también que si A es no singular y $A(v, w_1) = A(v, w_2) \; \forall \; v \in V \Rightarrow w_1 = w_2$, ya que en tal caso $A(v, w_1 - w_2) = 0 \; \forall \; v \in V$ y por lo tanto $w_1 - w_2 = 0$.

17.3 Cambio de base en formas bilineales

Consideremos una forma bilineal A. Frente a un cambio de base

$$b'_i = \sum_{j=1}^{n} S_{ji} b_j, \quad i = 1, \ldots, n$$

se tiene

$$A(b'_i, b'_k) = A(\sum_{j=1}^{n} S_{ji} b_j, \sum_{l=1}^{n} S_{lk} b_l) = \sum_{j=1}^{n}\sum_{l=1}^{n} S_{ji} A(b_j, b_l) S_{lk} = (S^t [A]_B S)_{ik}$$

Se obtiene entonces la ley de transformación

$$[A]_{B'} = S^t [A]_B S$$

donde $([A]_{B'})_{ij} = A(b'_i, b'_j)$ para $i, j = 1, \ldots, n$. De esta forma,

$$A(v,w) = [v]^t_B [A]_B [w]_B = (S[v]_{B'})^t [A]_B (S[w]_{B'}) = [v]^t_{B'} S^t [A]_B S [w]_{B'} = [v]^t_{B'} [A]_{B'} [w]_{B'}$$

Nótese la diferencia con la ley de transformación de matrices que representan operadores lineales $F : V \to V$ en una base, para las que $[F]_{B'} = S^{-1}[F]_B S$. Notemos también que ($|\ldots|$ denota el determinante)

$$|[A]_{B'}| = |S^t [A]_B S| = |S^t||[A]_B||S| = |S|^2 |[A]_B|$$

por lo que el signo del determinante no depende de la base (pues $|S| \neq 0$). Si A es singular, $|[A]_B| = 0$ y entonces $|[A]_{B'}| = 0$ en cualquier base.
Otra consecuencia es que como S es no singular ($|S| \neq 0$), el rango de $[A]_B$ (dimensión del espacio fila o columna de $[A]_B$) es también independiente de la base.
Podemos también corroborar que el carácter simétrico o antisimétrico es independiente de la base elegida:

$$[A]^t_{B'} = (S^t [A]_B S)^t = S^t [A]^t_B S$$

por lo que $[A]^t_{B'} = \pm [A]_{B'}$ si $[A]^t_B = \pm [A]_B$.

Ejemplo: Para el caso del producto escalar usual en $\mathbb{R}^n$, $[A]_e = I_n$ en la base canónica e y por lo tanto $[A]_{e'} = S^t [A]_e S = S^t S$ en una base arbitraria e', tal como se adelantó en el apunte 4 sobre cambio de base.
Ejemplo: Para el caso del determinante en $\mathbb{R}^2$, $[A]_e = \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$ en la base canónica y por lo tanto, en un base e' determinada por una matriz $S = [I]^{e'}_e = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ no singular ($|S| \neq 0$),

$$[A]_{e'} = S^t [A]_e S = \left(\begin{smallmatrix} a & c \\ b & d \end{smallmatrix}\right)\left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) = (ad - bc)\left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right) = |S|[A]_e$$

$[A]_{e'}$ es pues proporcional a $[A]_e$. Este resultado es obvio pues $[A]_{e'}$ debe ser antisimétrica y toda matriz antisimétrica de 2×2 debe ser proporcional a $[A]_e = \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$.
Ejemplo: Si $[A]_B = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$ y $b'_1 = (b_1 + b_2)$, $b'_2 = b_2 - b_1$, $S = \left(\begin{smallmatrix} 1 & -1 \\ 1 & 1 \end{smallmatrix}\right)$ y por lo tanto

$$[A]_{B'} = S^t [A]_B S = \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix}$$

Así, si $v = xb_1 + yb_2 = x'b'_1 + y'b'_2$, $w = zb_1 + tb_2 = z'b'_1 + t'b'_2$,

$$A(v,w) = (x,y)[A]_B\left(\begin{smallmatrix} z \\ t \end{smallmatrix}\right) = (x', y')[A]_{B'}\left(\begin{smallmatrix} z' \\ t' \end{smallmatrix}\right)$$

o sea,

$$A(v,w) = xt + yz = 2(x'z' - y't')$$

lo que está de acuerdo con $\left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right) = S\left(\begin{smallmatrix} x' \\ y' \end{smallmatrix}\right) = \left(\begin{smallmatrix} x'-y' \\ x'+y' \end{smallmatrix}\right)$, $\left(\begin{smallmatrix} z \\ t \end{smallmatrix}\right) = S\left(\begin{smallmatrix} z' \\ t' \end{smallmatrix}\right) = \left(\begin{smallmatrix} z'-t' \\ z'+t' \end{smallmatrix}\right)$

18 Formas cuadráticas

Si A es una forma bilineal de $V \times V$ en K, la función $\tilde{A}: V \to K$ dada por

$$\tilde{A}(v) = A(v, v)$$

se denomina *forma cuadrática*. Notemos que satisface $\tilde{A}(\alpha v) = \alpha^2 \tilde{A}(v)\ \forall\ \alpha \in K,\ v \in V$: $A(\alpha v, \alpha v) = \alpha A(v, \alpha v) = \alpha^2 A(v, v)$.
Es importante notar que la forma cuadrática queda completamente determinada por la parte simétrica de la forma bilineal, ya que $A_a(v, v) = [A(v, v) - A(v, v)]/2 = 0$ y por lo tanto

$$A(v, v) = A_s(v, v)$$

Asimismo, la parte simétrica de una forma bilineal queda completamente determinada por la forma cuadrática respectiva, ya que

$$A_s(v + w, v + w) = A_s(v, v) + A_s(w, w) + 2A_s(v, w)$$

y por lo tanto

$$A_s(v, w) = [A_s(v + w, v + w) - A_s(v, v) - A_s(w, w)]/2$$

En un espacio vectorial V de dimensión finita n, podemos entonces escribir, para A *simétrica*,

$$\begin{aligned} A(v, v) &= [v]_B^t [A]_B [v]_B \\ &= \sum_{i,j=1}^{n} \alpha_i A(b_i, b_j)\alpha_j = \sum_{i=1}^{n} A(b_i, b_i)\alpha_i^2 + 2\sum_{i<j} A(b_i, b_j)\alpha_i\alpha_j \end{aligned} \tag{3}$$

Ejemplo: Si $V = \mathbb{R}^2$ y $K = \mathbb{R}$, la longitud al cuadrado de un vector $v = (x, y)$,

$$|v|^2 = x^2 + y^2$$

es una forma cuadrática y puede escribirse como

$$|v|^2 = (x, y)\binom{x}{y} = (x, y)[A]_e\binom{x}{y} = A(v, v)$$

con $[A]_e = I_2$, $A(v, w) = v \cdot w$ y e la base canónica.
También es una forma cuadrática

$$\tilde{B}(v) = 3x^2 + 5y^2 + 2xy = (x, y)[B]_e\binom{x}{y}, \quad [B]_e = \begin{pmatrix} 3 & 1 \\ 1 & 5 \end{pmatrix}$$

Toda forma cuadrática en $V = \mathbb{R}^n$ o $V = \mathbb{R}^{n\times 1}$ puede escribirse como

$$\tilde{A}(v) = v^t A v = (\alpha_1, \ldots, \alpha_n) A \begin{pmatrix} \alpha_1 \\ \ldots \\ \alpha_n \end{pmatrix} = \sum_{i=1}^{n} a_{ii}\alpha_i^2 + 2\sum_{i<j} a_{ij}\alpha_i\alpha_j$$

con $a_{ij} = A_{ij} = a_{ji}$ los elementos de la matriz real *simétrica* A de $n \times n$ ($A^t = A$).
Ejemplo: Si $V = C_{[a,b]}$ y $K = \mathbb{R}$,

$$||f||^2 \equiv \int_a^b [f(x)]^2 dx = A(f, f)$$

es una forma cuadrática. También lo es $\tilde{C}(f) = \int_a^b \int_a^b K(x, x') f(x) f(x') dx dx'$.

18.1 Forma canónica de una forma cuadrática

Teorema: Sea V un espacio vectorial de dimensión finita n sobre un cuerpo K y sea $A: V \times V \to K$ una forma bilineal *simétrica*. Entonces existe una base B' en la que

$$A(b'_i, b'_j) = \begin{cases} 0 & i \neq j \\ a'_i & i = j \end{cases}$$

es decir, $A(b'_i, b'_j) = a'_i\delta_{ij}$. Esto implica, partiendo de una base arbitraria B, que existe una matriz de cambio de base S tal que

$$[A]_{B'} = S^t[A]_B S = \begin{pmatrix} a'_1 & 0 & \dots & 0 \\ 0 & a'_2 & \dots & 0 \\ & & \dots & \\ 0 & 0 & \dots & a'_n \end{pmatrix}$$

o sea, $([A]_{B'})_{ij} = a'_i\delta_{ij}$. En dicha base la forma bilineal toma entonces la forma diagonal o canónica

$$A(v, w) = \sum_{i=1}^{n} a'_i \alpha'_i \beta'_i$$

donde $v = \sum_{i=1}^n \alpha'_i b'_i$, $w = \sum_{i=1}^n \beta'_i b'_i$, y la correspondiente forma cuadrática toma la forma canónica

$$\tilde{A}(v) = A(v, v) = \sum_{i=1}^{n} a'_i {\alpha'_i}^2$$

Antes de proceder a la demostración, cabe destacar que ni los coeficientes a'_i, ni los vectores b'_i, son únicos. Por ejemplo, en la base B'' definida por $b''_i = \gamma_i b'_i$, $i = 1, \dots, n$, tenemos $A[b''_i, b''_j] = \gamma_i^2 a'_i \delta_{ij}$, y por lo tanto A toma también la forma canónica, con $a'_i \to a''_i = \gamma_i^2 a'_i$.

Notemos también que si la forma bilineal no es simétrica, no es posible encontrar una base en la que $[A]_{B'}$ sea diagonal: Si existiese, $[A]_{B'}$ sería simétrica y por lo tanto $[A]_{B''} = S^t[A]_{B'}S$ sería también simétrica en cualquier base B'' (y la forma bilineal sería entonces simétrica).

Demostración: En el caso de que $K = \mathbb{R}$, la demostración es inmediata si recordamos que toda matriz real simétrica A es siempre diagonalizable, que todos sus autovalores son reales y que sus autovectores pueden siempre elegirse ortogonales y de longitud 1 (véase apunte de autovalores).
Por lo tanto, existirá una matriz de cambio de base S formada por autovectores normalizados de $[A]_B$, con $|S| \neq 0$ y $S^{-1} = S^t$, tal que $S^{-1}[A]_B S = S^t[A]_B S$ es *diagonal*. Si B' es dicha base de autovectores, tendremos entonces

$$[A]_{B'} = S^t[A]_B S = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ & & \dots & \\ 0 & 0 & \dots & \lambda_n \end{pmatrix}$$

con $a'_i = \lambda_i$ los autovalores de $[A]_B$.

No obstante, cabe destacar que diagonalizar $[A]_B$ no es el único procedimiento para llevar una forma cuadrática a una forma diagonal. Esto puede también lograrse utilizando la conocida y simple técnica de *completar cuadrados*, en la cual se basa la demostración del teorema para un cuerpo arbitrario K, que damos a continuación. En tales casos, los coeficientes diagonales a'_i *no son necesariamente iguales a los autovalores de* A.

Notemos primero que si encontramos una transformación lineal de coordenadas

$$\begin{pmatrix} \alpha_1 \\ \dots \\ \alpha_n \end{pmatrix} = S \begin{pmatrix} \alpha'_1 \\ \dots \\ \alpha'_n \end{pmatrix}$$

(o sea, $[v]_B = S[v]_{B'}$) con S una matriz de $n \times n$ no singular ($|S| \neq 0$), tal que

$$A(v, v) = \sum_{i,j=1}^{n} A(b_i, b_j)\alpha_i\alpha_j = \sum_{i,j,k,l} S_{ik} A(b_i, b_j) S_{jl} \alpha'_k \alpha'_l = \sum_{i=1}^{n} a'_i {\alpha'_i}^2$$

hemos entonces encontrado una base canónica para la forma bilineal, dada por

$$b'_i = \sum_{j=1}^{n} S_{ji} b_j \quad i = 1, \dots, n$$

ya que en tal caso $[v]_B = S[v]_{B'}$ y $([A]_{B'})_{ij} = (S^t[A]_BS)_{ij} = a'_i\delta_{ij}$. El problema se reduce pues al de encontrar variables α'_i relacionadas linealmente con las α_i por una transformación no singular, en las que la forma cuadrática sea diagonal.

Procederemos ahora por inducción sobre la dimensión n de V. Para $n = 1$, toda forma cuadrática tiene trivialmente la forma canónica en cualquier base: Si $v \in V \to v = \alpha b_1$ y $\tilde{A}(v) = a'_1\alpha^2$, con $a'_1 = A(b_1, b_1)$. Para $n > 1$, supongamos que hemos demostrado que toda forma cuadrática en un espacio de dimensión $n-1$ puede escribirse en la forma canónica. Entonces,

$$A(v,v) = a_{nn}\alpha_n^2 + 2(a_{n1}\alpha_n\alpha_1 + \ldots + a_{n,n-1}\alpha_n\alpha_{n-1}) + g(\alpha_1, \ldots, \alpha_{n-1})$$

donde $v = \sum_{i=1}^n \alpha_i b_i$, $a_{ij} = A(b_i, b_j)$ y g representa una forma cuadrática de dimensión $n-1$. Si $a_{nn} \neq 0$ podemos escribir

$$A(v,v) = a_{nn}(\alpha_n^2 + 2\alpha_n\sum_{j=1}^{n-1}\alpha_j a_{nj}/a_{nn}) + g(\alpha_1, \ldots, \alpha_{n-1}) = a_{nn}(\alpha_n + \sum_{j=1}^{n-1}\alpha_j a_{nj}/a_{nn})^2 + h(\alpha_1, \ldots, \alpha_{n-1})$$

donde $h = g - a_{nn}(\sum_{j=1}^{n-1}\alpha_j a_{nj}/a_{nn})^2$. Por lo tanto

$$A(v,v) = a_{nn}{\alpha'_n}^2 + h(\alpha_1, \ldots, \alpha_{n-1}), \;\; \alpha'_n = \alpha_n + \sum_{j=1}^{n-1}\alpha_j a_{nj}/a_{nn}$$

Y como h representa una forma cuadrática de dimensión $n-1$, podemos escribirla en forma canónica como $h = \sum_{i=1}^{n-1} a'_i{\alpha'_i}^2$, donde α'_i son combinaciones lineales de los α_j, $j = 1, \ldots, n-1$. Finalmente obtenemos la forma canónica

$$A(v,v) = a'_n{\alpha'_n}^2 + \sum_{i=1}^{n-1} a'_i{\alpha'_i}^2$$

donde $a'_n = a_{nn}$ y la matriz de transformación $T = S^{-1}$ es de la forma

$$T = \begin{pmatrix} T_{n-1} & 0 \\ t & 1 \end{pmatrix}$$

con T_{n-1} una matriz no singular de $(n-1)\times(n-1)$ y t el vector de $n-1$ componentes determinado por α'_n ($t_i = a_{ni}/a_{nn}$). T es por consiguiente no-singular y define una base B' determinada por $S = T^{-1}$ en la que A tiene la forma canónica.
Si $a_{nn} = 0$ pero $a_{ii} \neq 0$ para algún $i < n$, podemos proceder en forma similar realizando la correspondiente permutación $i \leftrightarrow n$. Finalmente, si todos los a_{ii} son nulos pero existe algún elemento $a_{in} \neq 0$ con $i \neq n$ (pues de lo contrario tendríamos una forma de dimensión $n-1$), podemos efectuar primero el cambio de variables $\alpha_n = \tilde{\alpha}_n + \tilde{\alpha}_i$, $\alpha_i = \tilde{\alpha}_n - \tilde{\alpha}_i$, con lo cual $2a_{ij}\alpha_i\alpha_j = 2a_{ij}(\tilde{\alpha}_n^2 - \tilde{\alpha}_i^2)$ y podemos entonces proceder como en los casos anteriores.

Ejemplo: para $V = \mathbb{R}^2$, $K = \mathbb{R}$ y $v = (x,y) = xe_1 + ye_2$, consideremos

$$A(v,v) = (x,y)[A]_e\binom{x}{y} = x^2 + y^2 + 4xy$$

que coresponde a

$$[A]_e = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$$

Si optamos por el método (muy simple) de completar cuadrados, tenemos

$$x^2 + y^2 + 4xy = x^2 + (y+2x)^2 - 4x^2 = -3x^2 + (y+2x)^2$$

por lo que podemos escribir

$$A(v,v) = -3x'^2 + y'^2, \;\; \text{con} \;\; y' = (2x+y), \; x' = x$$

Esto corresponde a la transformación

$$\binom{x'}{y'} = T\binom{x}{y}, \;\; T = \left(\begin{smallmatrix} 1 & 0 \\ 2 & 1 \end{smallmatrix}\right)$$

Por lo tanto

$$S=T^{-1}=\left(\begin{smallmatrix}1&0\\-2&1\end{smallmatrix}\right)$$

y la base en la que A toma la forma canónica queda entonces determinada por las columnas de S:

$$e'_1=e_1-2e_2\ \ e'_2=e_2$$

Se verifica entonces

$$[A]_{e'}=S^t[A]_eS=\left(\begin{smallmatrix}1&-2\\0&1\end{smallmatrix}\right)\left(\begin{smallmatrix}1&2\\2&1\end{smallmatrix}\right)\left(\begin{smallmatrix}1&0\\-2&1\end{smallmatrix}\right)=\left(\begin{smallmatrix}-3&0\\0&1\end{smallmatrix}\right)$$

es decir, $A(e'_1,e'_1)=-3$, $A(e'_2,e'_2)=1$, $A(e'_1,e'_2)=0$, como es posible corroborar directamente.

Podemos también optar por el método basado en la diagonalización de A. Tenemos $|[A]_e-\lambda I_2|=(1-\lambda)^2-4=0$, de donde $\lambda=1\pm 2$, o sea, $\lambda_1=3$, $\lambda_2=-1$.
Las componentes de los autovectores correspondientes normalizados son $[e''_1]_e=\frac{1}{\sqrt{2}}\left(\begin{smallmatrix}1\\1\end{smallmatrix}\right)$, $[e''_2]_e=\frac{1}{\sqrt{2}}\left(\begin{smallmatrix}-1\\1\end{smallmatrix}\right)$, o sea, $e''_1=(e_1+e_2)/\sqrt{2}$, $e''_2=(-e_1+e_2)/\sqrt{2}$, y la correspondiente matriz de cambio de base es

$$S=\frac{1}{\sqrt{2}}\left(\begin{smallmatrix}1&-1\\1&1\end{smallmatrix}\right),\ \ S^{-1}=S^t=\frac{1}{\sqrt{2}}\left(\begin{smallmatrix}1&1\\-1&1\end{smallmatrix}\right)$$

Se verifica entonces

$$[A]_{e''}=S^t[A]_eS=\left(\begin{smallmatrix}3&0\\0&-1\end{smallmatrix}\right)$$

Es muy importante que los autovectores esten normalizados para que $S^{-1}=S^t$. Finalmente, se obtiene,

$$A(v,v)=(x'',y'')^t[A]_{e''}\left(\begin{smallmatrix}x''\\y''\end{smallmatrix}\right)=3x''^2-y''^2$$

donde $\left(\begin{smallmatrix}x''\\y''\end{smallmatrix}\right)=[v]_{e''}=S^{-1}[v]_e=S^{-1}\left(\begin{smallmatrix}x\\y\end{smallmatrix}\right)=\frac{1}{\sqrt{2}}\left(\begin{smallmatrix}x+y\\x-y\end{smallmatrix}\right)$, o sea, $x''=(x+y)/\sqrt{2}$, $y''=(x-y)/\sqrt{2}$.
Notemos que tanto los coeficientes diagonales como las bases obtenidas con los dos procedimientos anteriores son *distintos*. La diagonalización puede llevar más tiempo pero posee la ventaja que automáticamente proporciona una base *ortogonal* en la que la forma cuadrática tiene la forma canónica, lo cual es muy importante en diversas aplicaciones físicas.
Notemos también que el número de coeficientes positivos y negativos en la forma canónica obtenidos en ambos procedimientos *es el mismo*. Esta conclusión es general y se demostrará en el siguiente teorema, de gran importancia.

18.2 Teorema de inercia de formas cuadráticas:
Sea $A(v,v):V\times V\to\mathbb{R}$ una forma cuadrática sobre $\mathbb{R}$. El número de coeficientes a'_i positivos, negativos y nulos en cualquier forma canónica de A es el mismo.
Dem.: Consideremos dos formas canónicas distintas, tal que

$$A(v,v)=\sum_{i=1}^{n}a_i\alpha_i^2=\sum_{i=1}^{n}a'_i\alpha'^2_i$$

con $v=\sum_{i=1}^n\alpha_ie_i=\sum_{i=1}^n\alpha'_ie'_i$, $A(e_i,e_j)=a_i\delta_{ij}$, $A(e'_i,e'_j)=a'_i\delta_{ij}$, y

$$\begin{pmatrix}\alpha_1\\ \dots\\ \alpha_n\end{pmatrix}=S\begin{pmatrix}\alpha'_1\\ \dots\\ \alpha'_n\end{pmatrix}$$

o sea, $\alpha_i=\sum_{j=1}^nS_{ij}\alpha'_j$ para $i=1,\dots,n$, con $|S|\neq 0$.
Supongamos ahora que $a_i\begin{cases}>0 & i=1,\dots,k\\ <0 & i=k+1,\dots,m\\ 0 & i=m+1,\dots,n\end{cases}$, $a'_i\begin{cases}>0 & i=1,\dots,p\\ <0 & i=p+1,\dots,q\\ 0 & i=q+1,\dots,n\end{cases}$. Por consiguiente,

$$A(v,v)=\sum_{i=1}^{k}|a_i|\alpha_i^2-\sum_{i=k+1}^{m}|a_i|\alpha_i^2=\sum_{i=1}^{p}|a'_i|\alpha'^2_i-\sum_{i=p+1}^{q}|a'_i|\alpha'^2_i$$

Veremos ahora que si se supone $k < p$ se llega a un absurdo. Si $k < p$, podemos elegir $v \in V$, $v \neq 0$, tal que las primeras k componentes de v en la base e sean nulas ($\alpha_i = 0$ si $i \leq k$), y tal que sus últimas $n - p$ componentes en la base e' sean también nulas ($\alpha'_i = 0$ si $i > p$). En efecto, esto conduce al sistema de k ecuaciones homogéneas $0 = \sum_{j=1}^{p} S_{ij}\alpha'_j$ para $i = 1, \ldots, k$, con $p > k$ incógnitas α'_j, $j = 1, \ldots, p$, el cual posee entonces infinitas soluciones (y por lo tanto, soluciones no nulas). Para tal vector, tendríamos

$$A(v,v) = -\sum_{i=k+1}^{m} |a_i|\alpha_i^2 = \sum_{i=1}^{p} |a'_i|\alpha_i'^2$$

pero el segundo miembre es menor o igual a 0 y el tercero mayor que 0, lo que es imposible. Por lo tanto, no puede ser $k < p$. De la misma manera se prueba que no puede ser $p < k$. Por lo tanto, la única posibilidad es $k = p$, es decir, que el número de coeficientes positivos es el mismo.
De la misma forma (se dejan los detalles para el lector) se prueba que $m - k = q - p$ (el número de coeficientes negativos es el mismo).
Finalmente, los dos resultados anteriores implican $n - m = n - q$, es decir, que el número de coeficientes nulos es el mismo.

El número k (número de coeficientes positivos de la forma canónica) se denomina índice de inercia positivo y $m - k$ (número de coeficientes negativos) índice de inercia negativo.
El rango de una forma bilineal simétrica coincide con el rango de la matriz $[A]_e$ y es por lo tanto m (es decir, el número de coeficientes no nulos).
Si A es no singular $\Rightarrow m = n$ (el número de coeficientes nulos es 0).

Ejemplo: Consideremos, para $V = \mathbb{R}^2$, $\mathbb{R} = K$,

$$A(v,v) = x^2 + y^2 + 2xy$$

Completando cuadrados llegamos fácilmente a

$$A(v,v) = (x+y)^2 = 1x'^2 + 0y'^2$$

con $x' = (x + y)$, $y' = y$. Es decir, existe un coeficiente positivo ($a_1 = 1$) y uno nulo ($a_2 = 0$).
Si en cambio optamos por diagonalizar la matriz correspondiente ($[A]_e = \left(\begin{smallmatrix} 1 & 1 \\ 1 & 1 \end{smallmatrix}\right)$), obtenemos $|A - \lambda I_2| = (1-\lambda)^2 - 1 = 0$ y por lo tanto $\lambda = 1 \pm 1$, o sea, $\lambda_1 = 2$, $\lambda_2 = 0$. Obtenemos entonces un autovalor positivo y uno nulo.

Ejemplo: Consideremos, para $V = \mathbb{R}^3$,

$$A(v,v) = x^2 + y^2 + z^2 + 2xy + 2xz$$

Completando cuadrados,

$$A(v,v) = (x+y+z)^2 - (y+z)^2 + y^2 + z^2 = (x+y+z)^2 - 2yz = x'^2 + 2y'^2 - 2z'^2$$

donde $x' = x + y + z$, $z' = (z + y)/2$, $y' = (y - z)/2$ (se reemplazó $y = z' + y'$, $z = z' - y'$).
Esto implica que $[A]_e = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}$ tendrá dos autovalores positivos y uno negativo. En efecto, $|A - \lambda I_3| = (1-\lambda)((1-\lambda)^2 - 2) = 0$ conduce a $\lambda_1 = 1 > 0$, $\lambda_2 = 1 + \sqrt{2} > 0$, $\lambda_3 = 1 - \sqrt{2} < 0$.
Obtendremos en la correspondiente base de autovectores normalizados la forma canónica

$$A(v,v) = x''^2 + (1+\sqrt{2})y''^2 + (1-\sqrt{2})z''^2$$

18.3 Formas cuadráticas positivas y aplicaciones

Una forma cuadrática sobre $K = \mathbb{R}$ se denomina definida positiva (o estríctamente positiva) si

$$A(v,v) > 0 \quad \forall\, v \neq 0$$

Es fácil ver que A **es definida positiva si y sólo si los coeficientes diagonales** a_i **de la forma canónica son todos positivos:** $a_i > 0$ para $i = 1, \ldots, n$ (es decir, $k = n$). En efecto, en tal caso

$$A(v,v) = \sum_{i=1}^{n} a_i \alpha_i^2 > 0 \;\; \forall \;\; v \neq 0$$

donde ahora hemos escrito $v = \sum_{i=1}^{n} \alpha_i b_i$, con $B = (b_1, \ldots, b_n)$ una base donde A toma la forma canónica ($A(b_i, b_j) = a_i \delta_{ij}$). Por otro lado, si $A(v,v) > 0 \; \forall \; v \neq 0$, entonces $a_i = A(b_i, b_i) > 0$.

Para una forma cuadrática definida positiva, *podemos siempre elegir una base en la que* $a_i = 1$ para $i = 1, \ldots, n$: En efecto, si $A(b_i, b_j) = a_i \delta_{ij}$, con $a_i > 0$, podemos definir la base de elementos $e_i = b_i/\sqrt{a_i}$ en la que $A(e_i, e_j) = A(b_i, b_j)/\sqrt{a_i a_j} = (a_i/\sqrt{a_i^2})\delta_{ij} = 1\delta_{ij}$.
Notemos también que el determinante de la matriz que representa una forma cuadrática positiva es positivo en cualquier base. En la base B en la que A toma la forma canónica,

$$|[A]_B| = a_1 a_2 \ldots a_n > 0$$

y en cualquier otra base B' de V,

$$|[A]_{B'}| = \begin{vmatrix} A(b'_1, b'_1) & \ldots & A(b'_1, b'_n) \\ & \ldots & \\ A(b'_n, b'_1) & \ldots & A(b'_n, b'_n) \end{vmatrix} = |S^t [A]_B S| = |S|^2 |[A]_B| > 0$$

Además notemos que A sigue siendo positiva en cualquier subespacio de V (pues $A(v,v) > 0 \; \forall \; v \neq 0$), por lo que el determinante de cualquier menor de $[A]_{B'}$ (obtenido al suprimir un número dado de columnas y las respectivas filas de $[A]_{B'}$) es también siempre positivo. Por ejemplo, si consideramos el subespacio generado por los primeros $m \leq n$ elementos de la base B', tendremos

$$|[A]_m| > 0$$

donde $[A]_m$ es la matriz de $m \times m$ de elementos $A(b'_i, b'_j)$, $i \leq m$, $j \leq m$, que representa a A en la base $(b'_1, \ldots, b'_m)$ del subespacio anterior.

Más aún, A **es definida positiva si y sólo si todos los determinantes principales en una base arbitraria** B' **de** V **son positivos**, es decir, si $|[A]_m| > 0$ para $m = 1, \ldots, n$.
Dem.: Por inducción: Para $n = 1$ es obviamente válido. Asumiendo ahora que es válido para $n-1$, entonces existe una base canónica $(e_1, \ldots, e_{n-1})$ del subespacio generado por los primeros $n-1$ vectores de la base original B', en la que $A(e_i, e_j) = \delta_{ij}$. Definiendo ahora

$$e_n = b'_n - \sum_{i=1}^{n-1} \alpha_i e_i$$

con $\alpha_i = A(e_i, b'_n)$, obtenemos $A(e_i, e_n) = A(e_i, b'_n) - \alpha_i = 0$ para $i = 1, \ldots, n-1$. Se obtiene así una base canónica $e = (e_1, \ldots, e_{n-1}, e_n)$ de V en la que $A(e_i, e_j) = \delta_{ij} A(e_i, e_i)$, con $A(e_i, e_i) = 1$ si $i \leq n-1$ y entonces $A(e_n, e_n) = |[A]_e| > 0$ (pues $[A]_e = S^{\rm tr}[A]_{B'} S$ y $|[A]_e| = |S|^2 |[A]_{B'}| > 0$). La forma cuadrática es pues definida positiva.

Aplicaciones:
1) Clasificación de puntos críticos:
Consideremos un campo escalar $G : \mathbb{R}^n \to \mathbb{R}$ derivable a segundo orden orden en un entorno de un punto crítico $\vec{r}_0$ donde $\frac{\partial G}{\partial x_i}|_{\vec{r}=\vec{r}_0} = 0$, $i = 1, \ldots, n$. El polinomio de Taylor de segundo orden de $\Delta G(\vec{r}) = G(\vec{r}) - G(\vec{r}_0)$ alrededor de $\vec{r}_0$ es una forma cuadrática en $\Delta\vec{r} = \vec{r} - \vec{r}_0 = (\Delta x_1, \ldots, \Delta x_n)$:

$$\Delta G = \frac{1}{2} \sum_{i,j=1}^{n} \frac{\partial^2 G}{\partial x_i \partial x_j}|_{\vec{r}=\vec{r}_0} \Delta x_i \Delta x_j + R_3 = \frac{1}{2}(\Delta\vec{r}) H (\Delta\vec{r})^t + R_3$$

donde H es una matriz simétrica de $n \times n$, denominada matriz *Hessiana*, de elementos

$$H_{ij} = \frac{\partial^2 G}{\partial x_i \partial x_j}|_{\vec{r}=\vec{r}_0}$$

y R_3 es el resto ($\lim_{\vec{r}\to\vec{r}_0} R_3/|\vec{r}-\vec{r}_0|^2 = 0$). Llevando la forma cuadrática anterior a una forma canónica (ya sea completando cuadrados o diagonalizando la matriz H), obtenemos

$$\Delta G = \frac{1}{2}\sum_{i=1}^{n} a'_i (\Delta x'_i)^2 + R_3$$

Si $a'_i > 0$ para $i = 1, \ldots, n$, $\Delta G > 0$ para $|\Delta\vec{r}|$ suf. pequeño y el punto crítico es un mínimo local o relativo.
Si $a'_i < 0$ para $i = 1, \ldots, n$, $\Delta G < 0$ para $|\Delta\vec{r}|$ suf. pequeño y el punto crítico es un máximo local o relativo.
Y si existen a'_i positivos y negativos, se trata de un punto silla ("saddle point").
Finalmente, si algunos a'_i son nulos y $a'_i \geq 0$ para $i = 1, \ldots, n$ (o $a'_i \leq 0$ para $i = 1, \ldots, n$) el presente criterio no decide y es necesario un desarrollo a orden más alto (que puede también no ser concluyente) o bien un análisis alternativo.
Por lo tanto, podemos clasificar el punto crítico en forma inmediata conociendo los autovalores de la matriz H (de $n \times n$), o bien simplemente completando cuadrados y observando los signos de los coeficientes diagonales a_i. El último método es en general más sencillo (pues no requiere determinar raíces de ninguna ecuación) pero el primero tiene la ventaja de determinar a la vez (mediante los autovectores de H) n direcciones *ortogonales* en las que la forma cuadrática tiene la forma canónica (y por lo tanto conocer las direcciones ortogonales en las que ΔG es positivo ($a'_i > 0$) o negativo ($a_i < 0$)). (Ver práctica para más detalles).

Notemos también que si definimos $f_{\vec{r}_0} : \mathbb{R} \to \mathbb{R}$ como

$$f_{\vec{r}_0}(t) = G(\vec{r}_0 + t\Delta\vec{r})$$

entonces

$$f''_{\vec{r}_0}(0) = \sum_{i,j} \frac{\partial^2 G}{\partial x_i \partial x_j}|_{\vec{r}=\vec{r}_0} \Delta x_i \Delta x_j = (\Delta\vec{r}) H (\Delta\vec{r})^t$$

lo cual es una forma cuadrática en $\Delta\vec{r}$ definida por la matriz simétrica H. Si H es definida positiva $\Rightarrow f_{\vec{r}_0}(t)$ es cóncava hacia arriba en $t = 0$ para *cualquier dirección* $\Delta\vec{r}$, mientras que si es definida negativa, $f_{\vec{r}_0}(t)$ será cóncava hacia abajo para cualquier dirección $\Delta\vec{r}$. En el caso general, la concavidad dependerá de la dirección de $\Delta\vec{r}$.

2) Clasificación de curvas de nivel de formas cuadráticas. Consideremos la ecuación

$$\sum_{i,j=1}^{n} x_i a_{ij} x_j = C$$

que puede reescribirse como

$$\vec{r} A (\vec{r})^t = C$$

con $\vec{r} = (x_1, \ldots, x_n)$ y A la matriz (real) de elementos a_{ij}, que puede suponerse siempre simétrica ($a_{ij} = a_{ji}$). Llevándola a una forma canónica obtenemos la ecuación equivalente

$$\sum_{i=1}^{n} a'_i {x'_i}^2 = C$$

con los x'_i relacionados linealmente con los x_i. Si todos los a'_i son positivos (y $C > 0$) la ecuación anterior determina un *elipsoide*, mientras que si los a'_i tienen signos distintos la ec. determina un hiperboloide. Si la forma canónica se obtiene diagonalizando la matriz A, los autovectores pueden elegirse normalizados y ortogonales, en cuyo caso las variables x'_i serán las coordenadas a lo largo de ejes ortogonales en los que la forma cuadrática toma la forma canónica (ejes principales). (véase práctica para más detalles).

Ejemplo: Consideremos

$$G(x, y) = x^2 + y^2 + 2\alpha xy$$

$(0,0)$ es un pto. crítico de G y la matriz H de derivadas segundas es $H=2\binom{1\ \alpha}{\alpha\ 1}$. Sus autovalores son

$$\lambda_{\pm}=2(1\pm\alpha)$$

(obtenidos de la ec. $|H-\lambda I_2|=(2-\lambda)^2-4\alpha^2=0$). Por lo tanto, Si $|\alpha|<1$ ambos autovalores son positivos y $(0,0)$ es un mínimo de G (en este caso mínimo absoluto). En cambio, si $|\alpha|>1$, un autovalor es positivo y el otro negativo (por ej., si $\alpha>1$, $\lambda_+>0$, $\lambda_-<0$), por lo que $(0,0)$ es en este caso un punto silla. Las componentes de los autovectores normalizados (y por su puesto ortogonales) de H son $[v_{\pm}]_e=\binom{\pm1}{1}/\sqrt{2}$, por lo que $S=\binom{1\ -1}{1\ \ 1}/\sqrt{2}$ y podemos escribir

$$G(x,y)=(1+\alpha)x'^2+(1-\alpha)y'^2$$

con $\binom{x'}{y'}=S^{-1}\binom{x}{y}=\binom{x+y}{x-y}/\sqrt{2}$, como puede verificarse directamente.
Si G representa la energía potencial de un sistema físico dependiente de dos coordenadas x,y en las cercanías de un punto estacionario, vemos pues que el sistema será estable sólo si $|\alpha|<1$. Si $\alpha>0$, la estabilidad del sistema en la dirección de e_2' disminuye al aumentar α, tornándose inestable para $\alpha>1$.

Cabe destacar, no obstante, que la misma conclusión puede obtenerse simplemente completando cuadrados, lo cual conduce a

$$G(x,y)=(x+\alpha y)^2+y^2(1-\alpha^2)$$

Vemos pues que el coeficiente de y^2 es positivo si $|\alpha|<1$ y negativo si $|\alpha|>1$, mientras que el primero es siempre positivo.

Si consideramos ahora la ec.

$$x^2+y^2+2\alpha xy=C$$

el mismo análisis conduce a que para $C>0$, la ec. anterior representa una elipse si $|\alpha|<1$, con ejes principales inclinados 45 grados respecto de los originales (y radios de longitud $1/\sqrt{1\pm\alpha}$ para $C=1$), mientras que si $|\alpha|>1$ la ec. representa una hipérbola.
Ejemplo: Consideremos

$$G(x,y,z)=x^2+y^2+z^2+2xy+2xz$$

que ya fue analizado. $(0,0,0)$ es claramente un punto crítico. Completando cuadrados, se obtiene

$$G(x,y,z)=(x+y+z)^2-(y+z)^2+y^2+z^2=(x+y+z)^2-2yz=x'^2+2y'^2-2z'^2$$

donde $x'=x+y+z$, $z'=(z+y)/2$, $y'=(y-z)/2$, lo que implica que $(0,0,0)$ es un punto silla. El mismo resultado se obtiene de los autovalores de la matriz $H=2\begin{pmatrix}1&1&1\\1&1&0\\1&0&1\end{pmatrix}$, que son $\lambda_1=2>0$, $\lambda_2=2+2\sqrt{2}>0$, $\lambda_3=2-2\sqrt{2}<0$.
La ecuación

$$x^2+y^2+z^2+2xy+2xz=C$$

corresponde, por lo tanto, a un hiperboloide (de una hoja para $C>0$).

Ejemplo: Consideremos la función $T:\mathbb{R}^n\to\mathbb{R}$ dada por $\quad (X=(x_1,\ldots,x_n)^t)$

$$T(X)=\sum_{i,j}x_iA_{ij}x_j+2\sum_i r_ix_i$$

con $A_{ij}=A_{ji}$. Asumiendo que la matriz de coeficientes $A\in\mathbb{R}^{n\times n}$ es invertible, podemos reescribir T como

$$T(X)=X^tAX+(R^tX+X^tR)=Y^TAY-R^tA^{-1}R$$

donde $X^t=(x_1,\ldots,x_n)$, $R^t=(r_1,\ldots,r_n)$ y $Y=X+C$, con $C=A^{-1}R$. Es decir, $T(X)$ es una forma cuadrática en $Y=X+A^{-1}R$ (o sea, $y_i=x_i+\sum_j A^{-1}_{ij}r_j$) más una constante $R^tA^{-1}R$.

Si A es singular, podemos econtrar C tal que $AC=R$ sólo si $R\in EC(A)$ (espacio columna de A). En tal caso $T=Y^tAY-C^tR$ sigue siendo una forma cuadrática en $Y=X+C$, a menos de una constante $-C^tR$.

19. Espacios Euclídeos

Un espacio vectorial V sobre el cuerpo de los reales $\mathbb{R}$ es *Euclídeo* si está equipado con una operación denominada *producto escalar* y denotada por (v, w), que asigna a todo par de vectores un escalar real que satisface

$$(v, w) = (w, v) \;\; \forall\, v, w \in V$$

$$(v, w_1 + w_2) = (v, w_1) + (v, w_2), \;\; \forall\, v, w_1, w_2 \in V$$

$$(\alpha v, w) = \alpha(v, w) \;\; \forall\, v, w \in V, \; \alpha \in \mathbb{R}$$

$$(v, v) > 0 \; \forall v \neq 0, \;\; (0, 0) = 0$$

El producto escalar en un espacio euclídeo es pues *una forma bilineal simétrica* de $V \times V$ sobre $\mathbb{R}$ tal que la correspondiente forma cuadrática es *definida positiva*. Cualquier forma bilineal de este tipo es apta para definir un producto escalar. Notemos que $(0, v) = (v, 0) = 0 \; \forall \; v \in V$.

En un espacio de dimensión finita generado por una base $B = (b_1, \ldots, b_n)$, se obtiene, eligiendo para el producto escalar una forma bilineal simétrica G asociada a una forma cuadrática definida positiva,

$$(v, w) = G(v, w) = [v]_B^t [G]_B [w]_B = \sum_{i,j=1}^{n} \alpha_i g_{ij} \beta_j, \qquad g_{ij} = ([G]_B)_{ij} = (b_i, b_j) = g_{ji}$$

donde $v = \sum_i \alpha_i b_i$, $w = \sum_i \beta_i b_i$ y $[v]_B^t = (\alpha_1, \ldots, \alpha_n)$, $[w]_B^t = (\beta_1, \ldots, \beta_n)$. Recordemos que para este tipo de formas bilineales es siempre posible elegir una base B donde $[G]_B$ es diagonal, es decir, $(b_i, b_j) = g_i \delta_{ij}$, en cuyo caso el producto escalar toma la forma

$$(v, w) = \sum_{i,j=1}^{n} \alpha_i g_i \beta_i, \qquad g_i = (b_i, b_i) > 0$$

Definiendo ahora $e_i = b_i/\sqrt{g_i}$, podemos obtener así una *base canónica* $e = (e_1, \ldots, e_n)$ en la que $(e_i, e_j) = \delta_{ij}$ y por lo tanto $[G]_e = I_n$ (matriz identidad). El producto escalar en esta base adopta entonces la forma usual

$$(v, w) = [v]_e^t [w]_e = \sum_{i=1}^{n} \alpha_i \beta_i$$

A una base de este tipo la denominaremos *base canónica* o *base ortonormal* del espacio euclídeo.
Ejemplo 1: Si $V = \mathbb{R}^n$ y $v = (x_1, \ldots, x_n)$, $v' = (x'_1, \ldots, x'_n)$, el producto escalar usual, dado por

$$(v, v') = v \cdot v' = \sum_{i=1}^{n} x_i x'_i$$

satisface las 4 condiciones requeridas. Los vectores de la base canónica $e_1 = (1, 0, \ldots, 0) \ldots e_n = (0, \ldots, 0, 1)$ satisfacen $(e_i, e_j) = \delta_{ij}$ y forman pues una base ortonormal para este producto escalar.
Ejemplo 2: Si V es el espacio $C_{[a,b]}$ de funciones reales continuas $f : [a, b] \to \mathbb{R}$ (de dimensión infinita), podemos equiparlo con el producto escalar definido por

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

que satisface todas las condiciones requeridas (probar como ejercicio).
Ejemplo 3: Si $V = \mathbb{R}^{m \times n}$ es el espacio de matrices reales de $m \times n$, podemos definir el producto escalar de dos matrices $A, B \in V$ como

$$(A, B) = \mathrm{Tr}\,[A^t B] = \sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij} B_{ij} = (B, A)$$

que satisface también todas las condiciones requeridas (probar como ejercicio).

19.1 Norma de un vector

La **norma** (o longitud) de un vector $v \in V$ se define como

$$||v|| = \sqrt{(v,v)}$$

y satisface $||v|| \geq 0 \ \forall \ v \in V$, con $||v|| = 0$ si y sólo si $v = 0$. Por ejemplo, utilizando los productos escalares anteriores, en $V = \mathbb{R}^n$ se obtiene

$$||v|| = \sqrt{\sum_{i=1}^{n} x_i^2}$$

mientras que en $V = C_{[a,b]}$,

$$||f|| = \sqrt{\int_a^b f^2(x)dx}$$

y en $V = \mathbb{R}^{m\times n}$,

$$||A|| = \sqrt{\mathrm{Tr}\,[A^t A]} = \sqrt{\sum_{i,j} A_{ij}^2}$$

Todo vector en un espacio euclídeo posee pues una norma, que es positiva si $v \neq 0$ y 0 si $v = 0$. Notemos que $\forall \alpha \in \mathbb{R}$ se cumple

$$||\alpha v|| = \sqrt{(\alpha v, \alpha v)} = \sqrt{\alpha^2 (v,v)} = |\alpha| ||v||$$

de modo que la norma de αv es $|\alpha|$ veces la longitud de v.
Un vector de norma 1 se denomina vector unitario. Todo vector v no nulo puede ser *normalizado*, es decir, convertido en vector unitario mediante la multiplicación por un escalar: Si $||\alpha v|| = |\alpha| ||v|| = 1 \Rightarrow$ basta con elegir α tal que $|\alpha| = 1/||v||$, o sea, $\alpha = \pm 1/||v||$. El vector normalizado con el mismo sentido de v es pues

$$v_n = v/||v||$$

Un conjunto C de V se dice que es **acotado** si existe $m \in \mathbb{R}$ tal que $||v|| < m \ \forall \ v \in C$. El conjunto $\{v, \ ||v|| \leq 1\}$ se l lama bola unidad, mientras que el conjunto $\{v, \ ||v|| = 1\}$ esfera unidad. Estos conjuntos no son subespacios (como el lector podrá fácilmente mostrar).

19.2 Desigualdad de Cauchy-Schwarz y ángulo entre vectores

Dados dos vectores v, w de un espacio euclídeo V, se cumple siempre la **desigualdad de Cauchy-Schwartz**

$$|(v,w)| \leq ||v||\,||w|| \qquad (19.1)$$

donde la igualdad rige si y sólo si v y w son LD (Linealmente Dependientes).
Demostración: Si v, w son LD $\Rightarrow$ $v = \alpha\omega$ (o $w = \gamma v$) en cuyo caso $|(v,w)| = |\alpha||(w,w)| = |\alpha|\,||w||^2 = ||v||\,||w||$. Esto incluye en particular el caso en que v o w es nulo ($\alpha = 0$ o $\gamma = 0$).
Si $v \neq 0$ y $w \neq 0$, se obtiene, para los correspondientes vectores normalizados $v_n = v/||v||$, $w_n = w/||w||$,

$$0 \leq ||v_n \pm w_n||^2 = (v_n \pm w_n, v_n \pm w_n) = (v_n, v_n) + (w_n, w_n) \pm 2(v_n, w_n) = 2(1 \pm (v_n, w_n))$$

lo que implica $-1 \leq (v_n, w_n) \leq 1$, o sea,

$$|(v_n, w_n)| \leq 1$$

de donde $|(v,w)| \leq ||v||\,||w||$, como se quería demostrar. Vemos también que la igualdad ($|(v_n, w_n)| = 1$) implica $||v_n \pm w_n||^2 = 0$ y por lo tanto $v_n \pm w_n = 0$, en cuyo caso v y w son LD.

Por ejemplo, en $\mathbb{R}^n$, la desigualdad de Cauchy-Schwarz implica, para $v = \sum_{i=1}^n x_i e_i$, $w = \sum_{i=1}^n y_i e_i$,

$$|\sum_{i=1}^{n} x_i x'_i| \leq \sqrt{\sum_{i=1}^{n} x_i^2}\sqrt{\sum_{i=1}^{n} x'^2_i}$$

y en $C_{[a,b]}$,

$$|\int_a^b f(x)g(x)dx| \leq \sqrt{\int_a^b f^2(x)dx}\sqrt{\int_a^b g^2(x)dx}$$

mientras que en $V = \mathbb{R}^{m\times n}$,

$$\mathrm{Tr}[A^tB] \leq \sqrt{\mathrm{Tr}[A^tA]}\sqrt{\mathrm{Tr}[B^tB]}$$

El **ángulo** θ entre dos vectores v, w no nulos se define como

$$\cos\theta = \frac{(v,w)}{||v||\,||w||} = (v_n, w_n)$$

donde $v_n = v/||v||$, $w_n = w/||w||$ son los vectores normalizados. La desigualdad de Cauchy-Schwartz asegura que $-1 \leq (v_n, w_n) \leq 1$, por lo que el ángulo θ está correctamente definido. Notemos que si $v = \alpha w$ entonces entonces $\theta = 0$ $(\alpha > 0)$ o π $(\alpha < 0)$.
Ejercicio: Determinar el ángulo entre los vectores $(1, 1, \ldots, 1)$ y $(-1, 1, \ldots, 1)$ pertenecientes a $\mathbb{R}^n$.

19.3 Desigualdad triangular y distancia entre vectores

La desigualdad de Cauchy-Schwarz permite demostrar en forma inmediata la desigualdad triangular

$$|||v|| - ||w||| \leq ||v+w|| \leq ||v|| + ||w||$$

ya que $||v+w||^2 = (v+w, v+w) = (v,v) + (w,w) + 2(v,w)$, y por lo tanto,

$$||v+w||^2 \leq ||v||^2 + 2||v||\,||w|| + ||w||^2 = (||v|| + ||w||)^2$$

$$||v+w||^2 \geq |v||^2 - 2||v||\,||w|| + ||w||^2 = (||v|| - ||w||)^2$$

Notemos que se cumple también (dado que $||-w|| = ||w||$) $|||v|| - ||w||| \leq ||v-w|| \leq ||v|| + ||w||$.
La **distancia** entre dos vectores $d(v, w)$ se define como

$$d(v,w) = ||v-w||$$

y satisface las propiedades

$$d(v,w) \geq 0, \text{ con } d(v,w) = 0 \text{ sii } v = w$$

$$d(v,w) = d(w,v)$$

$$d(v,w) \leq d(v,u) + d(u,w)$$

donde la última es consecuencia de la desigualdad triangular: $||v-w|| = ||v-u-(w-u)|| \leq ||v-u|| + ||w-u||$.

19.4 Ortogonalidad y bases ortonormales

Dos vectores $v, w \in V$ se dicen **ortogonales** si $(v, w) = 0$. En tal caso, si $v \neq 0$, $w \neq 0$, $\cos\theta = 0$ y por lo tanto, $\theta = \pi/2$.

Un conjunto de m vectores v_i son **ortogonales** si $(v_i, v_j) = 0$ $\forall$ $i \neq j$, es decir, si son mutuamente ortogonales de a pares. Y se dicen que son **ortonormales** si además tienen norma no nula e igual a 1: $(v_i, v_i) = 1$, $i = 1, \ldots, n$. Una base ortonormal es una base compuesta por vectores ortonormales. La base canónica en la que $(e_i, e_j) = \delta_{ij}$ es pues una base *ortonormal*.

Independencia lineal de vectores ortogonales: Si $v_1, v_2, \ldots, v_m$ son mutuamente ortogonales ($(v_i, v_j) = 0$ si $i \neq j$) y no nulos ($||v_i||^2 = (v_i, v_i) > 0$) $\Rightarrow$ *son linealmente independientes.*
Demostración: Si

$$\alpha_1 v_1 + \alpha_2 v_2 + \ldots + \alpha_n v_n = 0$$

multiplicando escalarmente por v_i, con $1 \leq i \leq m$, y teniendo en cuenta que $(v_i, v_j) = 0$ si $i \neq j$, se obtiene

$$(v_i, \alpha_1 v_1 + \ldots + \alpha_n v_n) = (v_i, \alpha_i v_i) = \alpha(v_i, v_i) = (v_i, 0) = 0$$

lo que implica $\alpha_i = 0$ pues $(v_i, v_i) = ||v_i||^2 > 0$. Esto muestra que son LI. La prop. recíproca no es, obviamente, válida.

Por lo tanto, si dim $V=n \Rightarrow$ cualquier conjunto de n vectores ortogonales no nulos forma una base de V.
Generalización del teorema de Pitágoras: Si v_1, v_2 son ortogonales ($(v_1,v_2)=0$) $\Rightarrow$

$$||v_1+v_2||^2=(v_1+v_2,v_1+v_2)=(v_1,v_1)+(v_2,v_2)+2(v_1,v_2)=||v_1||^2+||v_2||^2$$

Y si $(v_1,v_2,\ldots,v_m)$ son mutuamente ortogonales ($(v_i,v_j)=0$ si $i\neq j$), $\Rightarrow$

$$||\sum_{i=1}^{m}v_i||^2=(\sum_{i=1}^{m}v_i,\sum_{j=1}^{m}v_j)=\sum_{i,j=1}^{m}(v_i,v_j)=\sum_{i=1}^{m}(v_i,v_i)=\sum_{i=1}^{m}||v_i||^2$$

Expansión en una base ortonormal: Si escribimos, para un vector $v\in V$,

$$v=\sum_{i=1}^{n}x_ie_i$$

donde $(e_1,\ldots,e_n)$ es una base ortonormal de V, entonces

$$x_i=(e_i,v),\quad i=1,\ldots,n$$

ya que $(e_i,v)=(e_i,\sum_{j=1}^n x_je_j)=\sum_{j=1}^n x_j(e_i,e_j)=x_i$ por ortonormalidad de los e_i. Las coordenadas x_i de v en la base canónica se obtienen pues simplemente efectuando el producto escalar (e_i,v), no siendo necesario resolver explícitamente un sistema de ecuaciones lineales para su obtención. Además, por la generalización del teorema de Pitágoras anterior,

$$||v||^2=\sum_{i=1}^{n}||x_ie_i||^2=\sum_{i=1}^{n}x_i^2$$

Los ángulos que forma v con e_i están determinados por

$$\cos(\theta_i)=\frac{(e_i,v)}{||e_i||\,||v||}=x_i/||v||$$

(ángulos directores) y satisfacen

$$\sum_{i=1}^{n}\cos^2(\theta_i)=\sum_{i=1}^{n}x_i^2/||v||^2=1$$

Se cumple entonces $x_i=||v||\cos\theta_i$ para $i=1,\ldots,n$.

Si una base e' es ortogonal pero no necesariamente ortonormal, entonces

$$x_i=(e'_i,v)/||e'_i||^2$$

con $||v||^2=\sum_{i=1}^n||x_ie'_i||^2=\sum_{i=1}^n x_i^2||e'_i||^2$ y $\cos(\theta_i)=\frac{(e'_i,v)}{||e'_i||\,||v||}=x_i||e'_i||/||v||$. Se sigue cumpliendo que $\sum_{i=1}^n\cos^2(\theta_i)=1$.

Notemos también que si $F:V\to W$ es una transformación lineal entre espacios euclídeos V y W de dimensiones n y m respectivamente, y e, $\tilde{e}$ son bases ortonormales de V y W, entonces los elementos F_{ij} de la matriz $[F]^e_{\tilde{e}}\in\mathbb{R}^{m\times n}$ que representa a F en estas bases están dados por el producto escalar

$$F_{ij}=(\tilde{e}_i,F(e_j))$$

dado que por definición, $F(e_j)=\sum_{i=1}^m F_{ij}\tilde{e}_i$.

Relación entre bases ortonormales.
Si e es una base ortonormal de V y e' es otra base de V, tenemos

$$e'_j=\sum_{i=1}^{n}S_{ij}e_i,\quad j=1,\ldots,n$$

con $S_{ij}=(e_i,e'_j)$ y

$$(e'_j,e'_k)=\sum_{i=1}^{n}S_{ij}S_{ik}=(S^tS)_{jk}$$

Vemos que e' será una base ortonormal ($(e'_j, e'_k) = \delta_{jk}$) *si y sólo si* la matriz de cambio de base S satisface

$$S^t S = I_n$$

o sea, $S^{-1} = S^t$. Las matrices reales que satisfacen esta relación se denominan *ortonormales* (o a veces ortogonales). Dado que $(S^t S)_{ij}$ es el producto escalar de la columna i por la columna j de S, las columnas de estas matr ices son *ortonormales* ($(S^t S)_{ij} = \delta_{ij}$) formando entonces una base ortonormal de $\mathbb{R}^n$. Como la ec. anterior implica asimismo $SS^t = I_n$, las filas de S *son también ortonormales* y forman asimismo una base ortonormal de $\mathbb{R}^n$ (se prueba de la misma manera).
Notemos además que $|S| \equiv \text{Det}S = \pm 1$, pues $|S^t S| = |S|^2 = 1$.
Resumiendo, la base e' será ortonormal sii la matriz de cambio de base S es una matriz ortonormal.

Para un vector arbitrario $v = \sum_{i=1}^n x_i e_i = \sum_{i=1}^n x'_i e'_i$, tenemos entonces

$$x'_i = (e'_i, v) = \sum_{j=1}^n x_j (e'_i, e_j) = \sum_{j=1}^n S^t_{ij} x_j$$

es decir,

$$[v]_{e'} = S^t [v]_e$$

lo que esta de acuerdo con la relación general $[v]_{e'} = S^{-1}[v]_e$.

19.5 Teorema de ortogonalización de Gram-Schmidt

Las propiedades anteriores muestran claramente la ventaja de trabajar con bases y conjuntos ortonormales. Daremos ahora un método general para construir bases ortogonales de espacios y subespacios.
Sean $v_1, \ldots, v_m$ m vectores LI $\in V$, que generan un subespacio $S \subset V$ de dimensión $m \leq n = \dim V$. Entonces existen m vectores *ortogonales* no nulos $w_1, \ldots, w_m$ que general el mismo espacio (y que son, por lo tanto, combinaciones lineales de los $v_1, \ldots, v_m$).
La demostración es directamente constructiva. Comencemos con $w_1 = v_1$. Definimos luego

$$w_2 = v_2 - \alpha w_1$$

y exigimos que $0 = (w_1, w_2) = (w_1, v_2) - \alpha(w_1, w_1)$. Por lo tanto $\alpha = (w_1, v_2)/||w_1||^2$ y

$$w_2 = v_2 - \frac{(w_1, v_2)}{||w_1||^2} w_1$$

Análogamente, definimos

$$w_3 = v_3 - \alpha_2 w_2 - \alpha_1 w_1$$

Las condiciones $0 = (w_2, w_3) = (w_2, v_3) - \alpha_2 ||w_2||^2$, $0 = (w_1, w_3) = (w_1, v_3) - \alpha_1 ||w_1||^2$ (donde hemos utilizado la ortogonalidad $(w_1, w_2) = (w_2, w_1) = 0$) implican $\alpha_i = (w_i, v_i)/||w_i||^2$ para $i = 1, 2$, y por tanto

$$w_3 = v_3 - \frac{(w_2, v_3)}{||w_2||^2} w_2 - \frac{(w_1, v_3)}{||w_1||^2} w_1$$

En general, definiendo para $i = 2, \ldots, m$,

$$w_i = v_i - \sum_{j=1}^{i-1} \alpha_j w_j,$$

las $i - 1$ condiciones $(w_j, w_i) = 0$ para $j = 1, \ldots, i-1$ implican $\alpha_j = \frac{(w_j, v_i)}{||w_j||^2}$, teniendo en cuenta la ortogonalidad $(w_j, w_k) = 0$ si $j < k < i$.

Por lo tanto,

$$w_1 = v_1, \quad w_i = v_i - \sum_{j=1}^{i-1} \frac{(w_j, v_i)}{||w_j||^2} w_j, \quad i = 2, \ldots, m$$

Los m vectores w_i así definidos son *no nulos*: si $w_i = 0 \Rightarrow v_i = \sum_{j=1}^{i-1} \alpha_j w_j$, lo que implicaría, dado que los w_j son combinaciones lineales de los v_j, que los vectores originales son LD, contradiciendo la hipótesis.

Los m vectores w_i así construidos son entonces mutuamente ortogonales por construcción ($(w_i, w_j) = 0$ si $i \neq j$) y no nulos, por lo que son LI, conformando entonces una base de S. Si $m = n$, se obtiene así un método para construir una base *ortogonal* del espacio completo V. Notemos que

$$||w_i||^2 = (w_i, w_i) = (w_i, v_i) = ||v_i||^2 - \sum_{j=1}^{i-1} (w_j, v_i)^2 / ||\omega_j||^2$$

Para obtener un conjunto ortonormal, se puede normalizar al final del procedimiento ($w_i \to w_i' = w_i/||w_i||$) o en cada paso. En este último caso, el método se resume en

$$w_1' = v_1/||v_1||, \quad w_i' = [v_i - \sum_{j=1}^{i-1} (w_j', v_i) w_j'] / [||v_i||^2 - \sum_{j=1}^{i=1} (w_j', v_i)^2]^{1/2}, \;\; i = 2, \ldots, m$$

Ejemplo: Sean $v_1 = (1,1,1)$, $v_2 = (1,1,-1)$ vectores de $\mathbb{R}^3$, no ortogonales ($(v_1, v_2) = 1$, con $(v_1, v_1) = (v_2, v_2) = 3$). Aplicando el método de Gram-Schmidt, se obtiene

$$w_1 = (1,1,1), \quad w_2 = (1,1,-1) - \frac{1}{3}(1,1,1) = (2,2,-4)/3$$

que son claramente ortogonales.
Para formar una base ortogonal de $\mathbb{R}^3$ que contenga a w_1 y w_2, podemos considerar un vector cualquiera v_3 tal que (w_1, w_2, v_3) sean LI. Por ejemplo, $v_3 = (1,0,0)$. Se obtiene entonces el resultado esperado

$$w_3 = v_3 - \frac{(w_1, v_3)}{||w_1||^2} w_1 - \frac{(w_2, v_3)}{||w_2||^2} w_2 = (1,0,0) - \frac{1}{3}(1,1,1) - \frac{1}{3}\frac{2/3}{24/9}(2,2,-4) = \frac{1}{2}(1,-1,0)$$

Ejemplo: Sean $p_1(t) = 1$, $p_2(t) = t$, $p_3 = t^2$ vectores de P_2 (polinomios de grado ≤ 2). Determinar una base ortogonal de P_2 para el producto escalar $(p,q) = \int_{-1}^{1} p(t) q(t) dt$.
Aplicando el método anterior, obtenemos, notando que $(p_1, p_1) = 2$, $(p_2, p_2) = 2/3$, $(p_3, p_3) = 2/5$, $(p_1, p_2) = 0 = (p_2, p_3)$, $(p_1, p_3) = 2/3$,

$$w_1(t) = 1, \;\; w_2(t) = t, \;\; w_3 = t^2 - \frac{2/3}{2} = t^2 - 1/3$$

Si exigimos que $w_i(1) = 1$ y extendemos $P_2 \to P_\infty$ se obtienen de esta manera los polinomios de Legendre: $P_1(t) = 1$, $P_2(t) = t$, $P_3(t) = (3t^2 - 1)/2$, etc.
De la misma manera, para productos escalares del tipo $(p,q) = \int_a^b p(t) q(t) \rho(t) dt$, donde $\rho(t) > 0$ para $t \in (a,b)$, se obtienen otras familias de polinomios ortogonales.

19.6 Proyección ortogonal

Sea w un vector no nulo $\in V$ y sea $v \in V$. Podemos descomponer v como una suma de un vector v_w paralelo a w y un vector $v - v_w$ *ortogonal* a w:

$$v = v_w + (v - v_w)$$

donde exigimos $(w, v - v_w) = 0$. Esta condición determina v_w. Escribiendo $v_w = \alpha w$, obtenemos

$$(w, v - \alpha w) = (w, v) - \alpha (w, w) = 0$$

por lo que $\alpha = (w, v)/||w||^2$ y

$$v_w = \frac{(w, v)}{||w||^2} w$$

El vector v_w es la *proyección ortogonal* de v sobre w y su significado geométrico es muy claro (recordar dibujo): Si trazamos la perpendicular desde el extremo de v a la recta generada por w, obtenemos un triángulo rectángulo formado por v, v_w y $v - v_w$, siendo $v_w \perp v - v_w$. Así, $v_w = 0$ si $v \perp w$, y $v_w = v$ si $v || w$.

El vector v_w puede también interpretarse como el vector paralelo a w cuya distancia a v es **mínima**. En efecto, si $u_w = \alpha w$,

$$d^2(v, u_w) = ||v - u_w||^2 = ||v - v_w + (v_w - u_w)||^2 = ||v - v_w||^2 + ||v_w - u_w||^2 + 2(v - v_w, v_w - u_w)$$

Pero el último término es nulo pues $v - v_w$ es $\perp$ a w y por tanto a $v_w - u_w$, por lo que

$$||v - u_w||^2 = ||v - v_w||^2 + ||v_w - u_w||^2 \geq ||v - v_w||^2$$

La distancia mínima se obtiene pues para $u_w = v_w$.

Operador de Proyección: El operador de proyección sobre w queda definido por

$$P_w(v) = v_w$$

y es un operador lineal que satisface $P_w^2 = P_w$. En una base canónica de V, $(w, v) = [w]_e^t[v]_e$, $||w||^2 = [w]_e^t[w]_e$ y entonces

$$[v_w]_e = \frac{[w]_e^t[v]_e}{[w]_e^t[w]_e}[w]_e = \frac{[w]_e[w]_e^t}{[w]_e^t[w]_e}[v]_e$$

La matriz $[P]_e$ que representa a P en una base canónica ($[v_w]_e = [P]_e[v]_e$) está entonces dada por

$$[P]_e = \frac{[w]_e[w]_e^t}{[w]_e^t[w]_e}$$

Ejemplo: Proyectar el vector $v = (1, 1, 1)$ sobre $w = (1, 1, -1)$.
Tenemos, como $(v, w) = 1$ y $||w||^2 = 3$,

$$v_w = \frac{1}{3}(1, 1, -1)$$

El operador de proyección correspondiente queda definido, para $v = (x, y, z) = xe_1 + ye_2 + ze_3$ (base canónica), por

$$v_w = P_w(v) = \frac{(w, v)}{||w||^2}w = \frac{x + y - z}{3}(1, 1, -1)$$

y la correspondiente matriz es entonces

$$[P_w]_e = \frac{1}{3}\begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}(1, 1, -1) = \frac{1}{3}\begin{pmatrix} 1 & 1 & -1 \\ 1 & 1 & -1 \\ -1 & -1 & 1 \end{pmatrix}$$

de forma que $[v_w]_e = [P_w]_e[v]_e$.
Gram-Schmidt en términos de proyectores:
El proceso de ortogonalización de Gram-Schmidt puede ahora escribirse como

$$w_1 = v_1, \quad w_i = v_i - \sum_{j=1}^{i-1} P_{w_j}(v_i), \quad i = 2, \ldots, m$$

El significado es muy claro: w_i se construye a partir de v_i quitándole a este último las proyecciones sobre cada uno de los vectores anteriores w_j, $j < i$. De esta forma w_i sólo conserva la parte de v_i ortogonal al espacio generado por los w_j.

La expansión de un vector en una base ortonormal puede entonces verse también como la suma de proyecciones ortogonales: Tenemos, para $v \in V$ y $(e_1, \ldots, e_n)$ una base ortonormal,

$$v = \sum_{i=1}^{n} x_i e_i = \sum_{i=1}^{n} P_{e_i}(v)$$

ya que $x_i e_i = (e_i, v)e_i = P_{e_i}(v)$.

19.7 Subespacios ortogonales

El conjunto de vectores ortogonales a un cierto vector v es un subespacio de V: Si $(v, w_1) = 0$, $(v, w_2) = 0$ $\Rightarrow (v, w_1 + w_2) = (v, w_1) + (v, w_2) = 0$ y $(v, \alpha w_1) = \alpha(v, w_1) = 0$. Además es no vacío pues $(v, 0) = 0$.
El conjunto de vectores ortogonales a todos los vectores de un cierto subespacio $S \subset V$ es también un subespacio (se prueba de la misma forma), denominado *complemento ortogonal* de S o $S_\perp$.

Mostraremos a continuación que $V = S \oplus S_\perp$.

Demostración: Sea $v \in V$ y v_s un vector $\in S$. Mostraremos que es siempre posible escribir

$$v = v_s + (v - v_s)$$

con $v_s \in S$ y $v - v_s \in S_\perp$. Si $(w_1, \ldots, w_m)$ es una base de S, que podemos escogerla ortogonal utilizando el método de Gram-Schmidt, entonces

$$v_s = \sum_{i=1}^{m} \alpha_i w_i, \quad \alpha_i = (w_i, v_s)/||w_i||^2$$

La condición $v - v_s \in S_\perp$ implica entonces

$$0 = (w_i, v - v_s) = (w_i, v) - (w_i, v_s) = (w_i, v) - \alpha_i ||w_i||^2, \ \ i = 1, \ldots, m$$

o sea, $\alpha_i = (w_i, v)/||w_i||^2$. En tal caso, $(v - v_s)$ será también ortogonal a cualquier vector de S (pues estos serán combinaciones lineales de los w_i), por lo que $v - v_w \in S_\perp$. Además $S \cap S_\perp = \{0\}$, pues si $u \in S$ y $u \in S_\perp \Rightarrow (u, u) = 0$ y por lo tanto $u = 0$. Queda probado entonces que $V = S \oplus S_\perp$. Si V es de dimensión n y S de dimensión $m \Rightarrow \dim S_\perp = n - m$.
El vector v_s así construido es *la proyección ortogonal* de v sobre el subespacio S, y puede escribirse como

$$v_s = \sum_{i=1}^{m} P_{w_i}(v) = P_S(v)$$

donde $P_{w_i}(v) = \frac{(w_i, v)}{||w_i||^2} w_i$ es el proyector sobre w_i y

$$P_S = \sum_{i=1}^{m} P_{w_i}$$

el proyector otrogonal sobre S. En esta expresión los w_i deben formar una base *ortogonal* de S.

El vector v_s es el vector $\in S$ que posee distancia **mínima** a v: Si $u_s \in S$,

$$||v - u_s||^2 = ||v - v_s + (v_s - u_s)||^2 = ||v - v_s||^2 + ||v_s - u_s||^2 + 2(v - v_s, v_s - u_s) = ||v - v_s||^2 + ||v_s - u_s||^2 \geq ||v - v_s||^2$$

Esta distancia mínima define la distancia de v a S:

$$d_{\min}(v, S) = ||v - v_s|| = ||v - P_S(v)||$$

Al disponer de una métrica, en un espacio euclídeo podemos pues no sólo determinar si un vector v pertence al subespacio S generado por un conjunto de vectores $\{w_1, \ldots, w_m\}$, sino también determinar que tan lejos está v de este subespacio, a través de la distancia $d_{\min}(v, S)$.

El método de **Gram-Schmidt** puede entonces expresarse en forma aún más concisa como

$$w_1 = v_1, \ \ w_i = v_i - P_{\overline{\{w_1, \ldots, w_{i-1}\}}}(v_i), \ \ i = 2, \ldots, m$$

donde $P_{\overline{\{w_1, \ldots, w_{i-1}\}}} = P_{w_1} + \ldots + P_{w_{i-1}}$ es el proyector ortogonal sobre el subespacio generado por los $i - 1$ vectores anteriores.

Ejemplo 1: El espacio nulo de una matriz A de $m \times n$, $N(A) = \{X | AX = 0\}$ con X vectores de $n \times 1$, es el *complemento ortogonal* de las filas de A, es decir, del espacio fila de A ($EF(A)$), ya que $(AX)_i = A_i X$ es el producto escalar de la fila i de A por X.

Se cumple por lo tanto dim $EF(A)$+dim $N(A) = n$.

Ejemplo 2: Encontrar $S_\perp$ si S es el espacio generado por los vectores $(1,1,1), (1,1,-1)$.

Una manera es resolver el sistema homogéneo $\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, que da como resultado el conjunto $\{x(1,-1,0),\ x \in \mathbb{R}\}$. $S_\perp$ es entonces el espacio generado por $(1,-1,0)$.
Se cumple dim S+dim $S_\perp$=2+1=3.

Ejemplo 3: a) Proyectar el vector $v = (1,2,3)$ sobre el plano generado por los vectores ortogonales $w_1 = (1,0,1)$ y $w_2 = (0,1,0)$.
Tenemos $(v, w_1) = 4$, $(v, w_2) = 2$, y

$$v_s = P_S(v) = P_{w_1}(v) + P_{w_2}(v) = \frac{4}{2}(1,0,1) + \frac{2}{1}(0,1,0) = (2,2,2)$$

b) Hallar la distancia mínima de v a S.
Tenemos $v - v_s = (-1,0,1)$ y $d_{\min} = ||v - v_s|| = \sqrt{2}$. Además, el ángulo entre v y S puede obtenerse a partir de $\cos\theta = ||v_s||/||v|| = 2\sqrt{3}/\sqrt{14}$.
c) Hallar la matriz que representa el proyector ortogonal sobre S en la base canónica de $\mathbb{R}^3$.

$$[P_S]_e = [P_{w_1}]_e + [P_{w_2}]_e = \frac{1}{2}[w_1]_e[w_1]_e^t + [w_2]_e[w_2]_e^t = \begin{pmatrix} 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \end{pmatrix}$$

Se verifica $[v_s]_e = [P_S]_e[v]_e$.

19.8 Representación general del operador de proyección

Es posible dar la expresión general de la matriz que representa al proyector ortogonal sobre un subespacio S generado por un conjunto LI de m vectores $\{w_1, \ldots, w_m\}$ no necesariamente ortogonales. Definamos la matriz

$$R = ([w_1]_e, \ldots, [w_m]_e)$$

de $n \times m$, con $m \leq n$, que contiene las coordenadas de los vectores en una base canónica e (con $n =$ dim V). Como $v_s \in S$ podemos escribir $v_s = \sum_{i=1}^m \alpha_i w_i$ y por lo tanto

$$[v_s]_e = \sum_{i=1}^m \alpha_i [w_i]_e = R\alpha$$

donde $\alpha = (\alpha_1, \ldots, \alpha_m)^t$. La condición $(w_i, v - v_s) = 0$ para $i = 1, \ldots, m$ implica

$$0 = R^t([v]_e - [v_s]_e) = R^t[v]_e - R^t[v_s]_e = R^t[v]_e - R^t R\alpha$$

de donde

$$\alpha = (R^t R)^{-1} R^t [v]_e$$

Por lo tanto, $[v_s]_e = R\alpha$ estará dado por

$$[v_s]_e = R(R^t R)^{-1} R^t [v]_e$$

La matriz que representa al proyector sobre S es entonces

$$[P_S]_e = R(R^t R)^{-1} R^t$$

Notar que $[P_S]_e^2 = [P_S]_e$, y que la expresión anterior no se puede simplificar, pues R no es cuadrada.
Discutiremos luego las propiedades de la matriz $R^t R$.
Ejemplo: Proyectar el vector $v = (1,2,3)$ sobre el plano generado por los vectores $w_1 = (1,1,1)$ y $w_2 = (2,1,2)$. Utilizando el método anterior, tenemos en este caso

$$R = \begin{pmatrix} 1 & 2 \\ 1 & 1 \\ 1 & 2 \end{pmatrix}$$

con $R^tR = \binom{3\ 5}{5\ 9}$, $R^{-1} = \binom{9\ -5}{-5\ 3}/2$ y

$$[P_S]_e = R(R^tR)^{-1}R^t = \begin{pmatrix} 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \end{pmatrix}$$

que coincide con el resultado del último ejercicio. La razón es que el espacio generado por $(1,1,1)$ y $(2,1,2)$ coincide con el generado por los vectores ortogonales $(1,0,1)$ y $(0,1,0)$ ($(1,1,1) = (1,0,1) + (0,1,0)$, $(2,1,2) = 2(1,0,1) + (0,1,0)$). Una forma general de obtener el resultado anterior es precisamente ver si los proyectores sobre el espacio generado son idénticos.

Solución de cuadrados mínimos
Consideremos nuevamente el sistema de m ecuaciones lineales con n incógnitas

$$AX = b$$

donde $A \in \mathbb{R}^{m\times n}$, $X \in \mathbb{R}^{n\times 1}$ $b \in \mathbb{R}^{m\times 1}$. Si A representa un monomorfismo $\Rightarrow$ rango$(A) = n \leq m$ y la matriz $A^tA \in \mathbb{R}^{n\times n}$ es no singular. De poseer solución, el sistema tiene entonces una solución única dada por

$$X = (A^tA)^{-1}A^t\,b$$

donde $(A^tA)^{-1}A^t$ es una inversa a izquierda de A. Esta solución se obtiene al multiplicar ambos miembros de $AX = B$ por $(A^tA)^{-1}A^t$, y es válida cuando b pertenece al espacio columna de A (EC(A)), es decir, cuando el sistema es compatible.

Cabe destacar, no obstante, que la expresión anterior para X tiene sentido aún si el sistema no tiene solución: En tal caso

$$AX = A(A^tA)^{-1}A^tb$$

es la **proyección ortogonal** de b sobre el espacio generado por las columnas de A, es decir, $AX = P_{EC(A)}(b)$, de modo que AX es el vector de EC(A) **más cercano a** b. En otras palabras, es el X que minimiza la distancia $||AX - b|| = \sqrt{\sum_{i=1}^m (AX-b)_i^2}$.

Ejemplo: Dado un conjunto de m puntos (x_i, y_i), $i = 1,\ldots,m$, con $x_i \neq x_j$ si $i \neq j$, hallar el polinomio de grado $n-1$ $p(x) = \sum_{j=0}^{n-1} c_jx^j$ tal que $\sum_{i=1}^m (p(x_i) - y_i)^2$ es mínimo. Considerar el caso $m \geq n$.

Tenemos un sistema de m ecuaciones $p(x_i) = y_i$, $i = 1,\ldots,m$, con n incógnitas c_j, $j = 0,\ldots,n-1$:

$$\begin{pmatrix} 1 & x_1 & \ldots & x_1^{n-1} \\ \ldots & & & \\ 1 & x_m & \ldots & x_m^{n-1} \end{pmatrix} \begin{pmatrix} c_0 \\ \ldots \\ c_{n-1} \end{pmatrix} = \begin{pmatrix} y_1 \\ \ldots \\ y_m \end{pmatrix}$$

Este sistema es en general incompatible si $m \geq n$. No obstante, el objetivo es buscar la solución que minimiza la distancia $||p(X) - Y||$ o equivalentemente $||p(X) - Y||^2$, donde $Y = (y_1,\ldots,y_m)^t$, $X = (x_1,\ldots,x_m)^t$ y $p(X) = AC$, con A la matriz de $m \times n$ de elementos $A_{ij} = x_i^{j-1}$ y C el vector columna de coeficientes c_i. Tal solución estará dada entonces por $C = (A^tA)^{-1}A^tY$, tal que $AC = A(A^tA)^{-1}A^tY$ es la proyección ortogonal de Y sobre $EC(A)$.

19.9 Matriz de Gram

Dado un conjunto de vectores $v_1,\ldots,v_m$ pertenecientes a un espacio vectorial euclídeo V de dimensión finita, la matriz simétrica G de $m \times m$ productos escalares, de elementos

$$G_{ij} = (v_i, v_j) = G_{ji}$$

se denomina matriz de Gram y posee importantes propiedades.
Notemos que en términos de la matriz $R = ([v_1]_e,\ldots,[v_m]_e)$ de $n \times m$, con e una base canónica,

$$G = R^tR$$

1) El producto escalar (w,u) de combinaciones lineales $w=\sum_{i=1}^n \alpha_i v_i$, $u=\sum_{i=1}^n \beta_i v_i$, con $[w]_e=R\alpha$, $[u]_e=R\beta$, y $\alpha=(\alpha_1,\ldots,\alpha_m)^t$, $\beta=(\beta_1,\ldots,\beta_m)^t$, puede expresarse como

$$(w,u)=[v]_e^t[u]_e=(R\alpha)^t(R\beta)=\alpha^t R^t R\beta=\alpha^t G\beta$$

2) La matriz G es no singular sii los vectores v_i son LI:
Si G es singular, existe un vector columna no nulo β de $m\times 1$ tal que $G\beta=0$ y por lo tanto, si $u=\sum_{i=1}^n \beta_i v_i$,

$$(u,u)=\beta^t G\beta=\beta^t 0=0$$

por lo que necesariamente $u=0$. Por lo tanto, existe una combinación lineal nula u con coeficientes no todos nulos. Esto implica que los v_i son LD.
Análogamente, si existe una combinación lineal nula $u=\sum_{i=1}^n \beta_i v_i=0$, con los β_i no todos nulos, entonces $0=(v_j,u)=\sum_{i=1}^n G_{ji}\beta_i$ para cualquier j por lo que $G\beta=0$ y por lo tanto G es necesariamente singular.
Un método sencillo de determinar si los m vectores v_i son LI es pues evaluar el determinante $|G|=|R^tR|$: $\{v_1,\ldots,v_m\}$ es LI sii $|G|\neq 0$.
Para $m=n$, R es de $n\times n$ y $|G|=|R|^2$, por lo que se reobtiene la condición conocida $|R|\neq 0$ para n vectores en $\mathbb{R}^n$.
3) Si $w_i=\sum_{j=1}^n S_{ji}v_j$, $i=1,\ldots,m$, entonces $G'_{ij}\equiv(w_i,w_j)=\sum_{k,l}S_{ki}S_{lj}G_{kl}$, o sea,

$$G'=S^tGS$$

con $|G'|=|S|^2|G|$. En particular, si los v_i son LI, podemos ortogonalizarlos con el método de Gram-Schmidt, generando vectores ortogonales w_i. La correspondiente matriz S cumple, por construcción, $|S|=1$ y por lo tanto $|G|=|G'|=\prod_{i=1}^m ||w_i||^2$.
Este último producto representa *el cuadrado del volumen m dimensional del paralelepípedo formado por los vectores* $w_1,\ldots,w_m$, y por lo tanto, por $v_1,\ldots,v_m$. El volumen generado por estos m vectores es pues

$$Vol_{v_1,\ldots,v_m}=\sqrt{|G|}$$

Si $m=n$, $|G|=|R^tR|=|R|^2$, y $Vol_{v_1,\ldots,v_m}=|\mathrm{Det}(R)|$.
4) La matriz G es diagonalizable, por ser real y simétrica, y los autovalores de G son *positivos o nulos*. Los autovectores asociados a autovalores no nulos corresponden a vectores *ortogonales*, y los correspondientes a autovalores nulos a combinaciones lineales *nulas* de los vectores v_i.
En efecto, si $G\alpha=\lambda_\alpha\alpha$, con $\alpha\neq 0$, para $w=\sum_{i=1}^m \alpha_i v_i$ se obtiene

$$0\leq(w,w)=\alpha^tG\alpha=\lambda_\alpha\alpha^t\alpha$$

Como $\alpha^t\alpha>0$ entonces $\lambda_\alpha\geq 0$. Si $\lambda_\alpha>0\Rightarrow w$ es no nulo, mientras que si $\lambda_\alpha=0\Rightarrow w=0$, siendo pues una combinación lineal *nula* de los v_i. Además, si $G\beta=\lambda_\beta\beta$, con $\beta\neq 0$, tenemos, para $u=\sum_{i=1}^m \beta_i v_i$,

$$(w,u)=\alpha^tG\beta=\lambda_\beta\alpha^t\beta=0 \ \text{ si } \ \lambda_\alpha\neq\lambda_\beta$$

por ser α,β autovectores de una matriz simétrica. La diagonalización de G proporciona pues un método directo de extraer un conjunto ortogonal de k vectores LI de los m vectores w_i, que son los determinados por los autovectores asociados a los autovalores no nulos.
El número k de autovalores no nulos de G es precisamente el *rango* de G y determina entonces la *dimensión* del subespacio generado por los m vectores v_i: $k=r(G)=\dim\overline{\{v_1,\ldots,v_m\}}$.

Ejemplo: Consideremos los vectores $v_1=(1,1,1,1)$, $v_2=(1,1,-1,1)$, $v_3=(0,0,1,0)$ de $\mathbb{R}^4$. Tenemos

$$R=\begin{pmatrix}1&1&0\\1&1&0\\1&-1&1\\1&1&0\end{pmatrix},\quad G=R^tR=\begin{pmatrix}4&2&1\\2&4&-1\\1&-1&1\end{pmatrix}$$

Como $|G|=0$ los vectores son LD. Además, los autovalores de G son $\lambda=6,3,0$, con autovectores $(1,1,0)$, $(1,-1,1)$, $(-1,1,2)$.
Por lo tanto, los vectores $w'_1=v_1+v_2=(2,2,0,2)$, $w_2=v_1-v_2+v_3=(0,0,3,0)$, son ortogonales y $w'_3=-w_1+w_2+2w_3=(0,0,0,0)$ es la combinación lineal nula.
Pueden obtenerse resultados similares utilizando Gram-Schmidt. El determinante del primer menor de O, $16-4=12$, representa el cuadrado del área del paralelogramo determinado por w_1 y w_2.

19.10 Operadores adjuntos y autoadjuntos en espacios euclídeos

Sea $F: V \to V$ un operador lineal en un espacio euclídeo V. El operador *adjunto* $F^\dagger$ se define por

$$(v, F(w)) = (F^\dagger(v), w)$$

$\forall\, v, w \in V$. Si V es de dimensión finita y e denota una base canónica de V, la definición anterior implica

$$[v]_e^t [F]_e [w]_e = ([F^\dagger]_e [v]_e)^t [w]_e = [v]_e^t [F^\dagger]_e^t [w]_e$$

por lo que la matriz $[F^\dagger]_e \equiv [F^\dagger]_e^e$ que representa a $F^\dagger$ en dicha base es la traspuesta de la matriz que representa a F:

$$[F^\dagger]_e = [F]_e^t$$

Esto también muestra que $(F^\dagger)^\dagger = F$ (pues $[(F^\dagger)^\dagger]_e = ([F]_e^t)^t = [F]_e$) y que si $G: V \to V$ es otro operador lineal, $(FG)^\dagger = G^\dagger F^\dagger$ (pues $[(FG)^\dagger]_e = [FG]_e^t = [G]_e^t [F]_e^t$). Estas dos últimas propiedades pueden también demostrarse a partir de la defínción de operador adjunto (se deja como ejercicio).
Notemos que $(F(v), w) = (v, F^\dagger(w))$.
Operador autoadjunto: Si $F^\dagger = F$ el operador se dice *autoadjunto*. En este caso debe cumplirse

$$[F]_e^t = [F]_e$$

por lo que F será autoadjunto si y sólo si es representado por una matriz *simétirca* en una base canónica.

Notemos que en una base arbitraria B, no necesariamente ortogonal, con $(b_i, b_j) = g_{ij} = g_{ji}$, tendríamos $(v, F(w)) = [v]_B G [F]_B [w]_B$, $(F^\dagger(v), w) = [v]_B^t [F^\dagger]_B^t G [w]_B$ y por lo tanto, $[F^\dagger]_B^t G = G [F]_B$, por lo que

$$[F^\dagger]_B = G^{-1} [F]_B^t G$$

Si F es autoadjunto $\Rightarrow [F]_B = G^{-1} [F]_B^t G$.

En general, si $F: V \to W$ es una transformación lineal entre espacios euclídeos V, W, podemos definir $F^\dagger: W \to V$ de la misma forma: $(w, F(v)) = (F^\dagger(w), v)\ \forall\, v,\, w$. Para espacios V, W de dimensión finita n y m respectivamente, esto implica $[F^\dagger]_e^{\tilde{e}} = ([F]_{\tilde{e}}^e)^{\rm tr}$ en bases ortonormales e y $\tilde{e}$ de V y W. De esta forma, $F_{ij} = (\tilde{e}_i, F(e_j)) = (F^\dagger(\tilde{e}_i), e_j) = (e_j, F^\dagger(\tilde{e}_i)) = F^\dagger_{ji}$.

Diagonalización de operadores autoadjuntos

Si $F: V \to V$ es un operador lineal autoadjunto en un espacio V de dimensión finita, demostraremos que *existe siempre una base canónica e' en la que $[F]_{e'}$ es diagonal* (Ya habíamos demostrado que los autovalores de matrices reales simétricas son todos reales y que los autovectores corresp. a autovalores distintos son ortogonales). Un resultado aún más general será demostrado luego para espacios complejos.
Para $n = \dim\ V = 1$, el resultado es trivial. Asumimos ahora que es válido para $\dim\ V = n-1$. Si e'_1 es un autovector normalizado de F con autovalor λ_1, tal que $F(e'_1) = \lambda_1 e'_1$, $(e'_1, e'_1) = 1$, se puede construir, por Gram-Schmidt, una base ortonormal $\tilde{e}$ de V tal que $\tilde{e}_1 = e'_1$, definida por una matriz de cambio de base S ortonormal ($S^t = S$). En tal caso $[F]_{\tilde{e}} = S^{-1}[F]_e S = S^t [F]_e S$ será también *simétrica*: $[F]_{\tilde{e}}^t = S^t [F]_e S = [F]_{\tilde{e}}$. Pero como e'_1 es autovector, $[F]_{\tilde{e}}$ tendrá entonces la forma $\left(\begin{smallmatrix}\lambda_1 & 0\\ 0 & \tilde{F}\end{smallmatrix}\right)$, con $\tilde{F}$ una matriz *simétrica* de $(n-1)\times(n-1)$ que representa a un operador autoadjunto en un subespacio de dimensión $n-1$. Por hipótesis inductiva, existe una base ortonormal de este subespacio en la que el operador será representado por una matriz diagonal F'. Por lo tanto, agregando a esta base el autovector e'_1, tendremos una base ortonormal e' de V en la que $[F]_{e'} = \left(\begin{smallmatrix}\lambda_1 & 0\\ 0 & F'\end{smallmatrix}\right)$ será también diagonal.
Resumiendo, dado F autoadjunto ($[F]_e$ simétrica en una base canónica e) existe una base ortonormal definida por una matriz de cambio de base S, con $S^{-1} = S^t$, tal que

$$[F]_{e'} = S^t [F]_e S$$

es diagonal

19.11 Isometrías

Las isometrías son operadores $U : V \to V$ que conservan el producto escalar. Ejemplos comunes en $V = \mathbb{R}^n$ son rotaciones y reflexiones. Si U es una isometría,

$$(U(v), U(w)) = (v, w) \ \ \forall \ \ v, w \in V$$

Por lo tanto, si e es una base canónica, $(U(v), U(w)) = [v]_e^t [U]_e^t [U]_e [w]_e = [v]_e^t [w]_e \ \forall \ v, w \in V$, por lo que

$$[U]_e^t [U]_e = I_n$$

con I_n la matriz identidad, es decir, $[U]_e^{-1} = [U]_e^t$. Esto implica a su vez $[U]_e [U]_e^t = I_n$. Las matríces $[U]_e$ que representan a una isometría en una base canónica e son pues matrices *ortonormales*, y tanto las filas como las columnas de $[U]_e$ serán por lo tanto *ortonormales*, como se vió anteriormente: Si $U_{ij} = ([U]_e)_{ij}$,

$$\sum_{j=1}^{n} U_{ji} U_{jk} = \delta_{ik}, \ \ \sum_{j=1}^{n} U_{ij} U_{kj} = \delta_{ik}$$

En términos de operadores adjuntos, $(U(v), U(w)) = (v, U^\dagger U(w))$, por lo que U será una isometría si y sólo si

$$U^{-1} = U^\dagger$$

Demostraremos luego que toda isometría puede ser descompuesta en rotaciones y/o reflexiones.

Las isometrías transforman bases ortogonales en bases ortogonales. En efecto, al conservar todos los productos escalares, si $e'_i = U(e_i) = \sum_{j=1}^n U_{ji} e_j$, entonces

$$(e'_i, e'_j) = (U(e_i), U(e_j)) = (e_i, e_j) = \delta_{ij}$$

La recíproca es obviamente también válida: Cualquier par de bases canónicas e, e' de V estarán relacionadas por una isometría $e'_i = U(e_i)$. Cualquier matriz de cambio de base S que represente una isometría debe pues satisfacer $S^t S = I_n$, como se vió anteriormente.
Ejemplo: Si

$$[U]_e = \begin{pmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

entonces U es una isometría ya que $[U]_e^t [U]_e = I_3$. Tanto las filas como las columnas de $[U]_e$ son ortonormales (ortogonales y de longitud 1). Esta matriz representa una rotación de ángulo α antihoraria en el plano xy, compuesta con una reflexión respecto a este plano:

$$[U]_e = \begin{pmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

Isomorfismo Euclídeo

Dados dos espacios euclídeos V, V' de la misma dimensión, podemos siempre elegir bases canónicas $e = (e_1, \ldots, e_n)$ en V y $e' = (e'_1, \ldots, e'_n)$ en V' tal que tales que $(e_i, e_j) = \delta_{ij}$, $(e'_i, e'_j) = \delta_{ij}$. Definiendo un isomorfismo $Q : V \to V'$ tal que $Q(e_i) = e'_i$, $i = 1, \ldots, n$, se tiene

$$(e'_i, e'_j) = (Q(e_i), Q(e_j)) = (e_i, e_j) = \delta_{ij}$$

Por lo tanto, si $v' = \sum_{i=1}^n \alpha_i e'_i$, $w' = \sum_{i=1}^n \beta_i e'_i \Rightarrow v' = Q(v)$, $w' = Q(w)$, con $v = \sum_{i=1}^n \alpha_i e_i$, $w = \sum_{i=1}^n \beta_i e_i$ y

$$(v', w') = (Q(v), Q(w)) = (v, w) = \sum_{i=1}^{n} \alpha_i \beta_i$$

Un isomorfismo $Q : V \to V'$ de este tipo (que conserva todos los productos escalares) se lo denomina isomorfismo euclídeo. La existencia de Q muestra que todas las propiedades geométricas de $\mathbb{R}^n$ pueden extenderse directamente a cualquier espacio euclídeo V' de dimensión n.

20 Descomposición en valores singulares (DVS)

Consideremos una matriz real A de $m\times n$. Podemos formar la matriz de $n\times n$

$$A^tA$$

la cual es simétrica ($(A^tA)^t=A^tA$) y tiene la mismas propiedades que la matriz de Gram. Por lo tanto, tiene un conjunto de n autovectores $v_i\in\mathbb{R}^{n\times 1}$ *ortonormales* asociados a autovalores λ_i*positivos o nulos*:

$$A^tAv_i=\lambda_i v_i,\quad i=1,\ldots,n\ \text{ con }\ v_j^tv_i=\delta_{ij},\ \ \lambda_i\geq 0$$

Sean $\lambda_1,\ldots,\lambda_k$, $k\leq n$, los autovalores no nulos de O. Podemos definir los k vectores de $m\times 1$

$$u_i=\frac{1}{\sqrt{\lambda_i}}Av_i,\quad i=1,\ldots,k\,,\ \ \lambda_i\neq 0$$

que son ortonormales:

$$u_j^tu_i=\frac{1}{\sqrt{\lambda_j\lambda_i}}v_j^tA^tAv_i=\frac{\lambda_i v_j^tv_i}{\sqrt{\lambda_i\lambda_j}}=\delta_{ij}$$

Si $k<m$, podemos completar estos k vectores con $m-k$ vectores obtenidos por el método de Gram-Schmidt, tal que $(u_1,\ldots,u_m)$ forme un conjunto ortonormal (base de $\mathbb{R}^{m\times 1}$). Además, para $i=k+1,\ldots n$ se cumple $A^tAv_i=0$ y entonces $(Av_i)^t(Av_i)=v_i^tA^tAv_i=0$, es decir $||Av_i||=0$, lo que implica $Av_i=0$. Tenemos entonces

$$A(v_1,\ldots,v_n)=(\sqrt{\lambda_1}u_1,\ldots,\sqrt{\lambda_k}u_k,0\ldots,0)=(u_1,\ldots,u_m)A'$$

donde A' es una matriz "diagonal" de $m\times n$ de la forma

$$A'=\begin{pmatrix}\sigma_1 & 0 & \ldots & 0 & \ldots & 0\\ 0 & \sigma_2 & \ldots & 0 & \ldots & 0\\ & & \ldots & & & \\ 0 & \ldots & \sigma_k & 0 & \ldots & 0\\ 0 & \ldots & 0 & 0 & \ldots & 0\\ & & \ldots & & & \\ 0 & \ldots & 0 & 0 & \ldots & 0\end{pmatrix},\qquad \sigma_i=\sqrt{\lambda_i},\ \ i=1,\ldots,k$$

Por consiguiente, definiendo las matrices ortonormales $V=(v_1,\ldots,v_n)$, $U=(u_1,\ldots,u_m)$ (que satisfacen $V^tV=I_n$, $U^tU=I_m$), se tiene $AV=UA'$ y por lo tanto

$$A=UA'V^t$$

Esta representación de A se denomina *descomposición en valores singulares* (del inglés *singular value decomposition*) y los elementos σ_i de A' *los valores singulares* de A, que son las raíces de los autovalores no nulos de A^tA (necesariamente positivos). Vemos así que rango(A) =rango$(A')=k$, por lo que $k\leq \text{Min}[m,n]$. Además, por construcción, los primeros k vectores u_j, $j=1,\ldots,k$ forman una base del espacio columna de A y los últimos $n-k$ vectores $v_{k+1},\ldots,v_n$ una base del espacio nulo de A (el subespacio ortogonal al espacio fila de A).

Notemos también que si $A=UA'V^t$, con A' "diagonal" de $m\times n$ con elementos positivos o nulos y U, V matrices ortonormales, entonces necesariamente los elementos diagonales no nulos de A' son los valores singulares, pues

$$A^tA=VA'^tU^tUA'V^t=V(A'^tA')V^t$$

con A'^tA' diagonal de $n\times n$. Esto implica $V^tA^tAV=A'^tA'$, lo que muestra que V es necesariamente una matriz ortonormal de autovectores de A^tA y A'^tA' la correspondiente matriz diagonal de autovalores.

Desde el punto de vista operacional, A puede considerarse como la representación $[F]_{\tilde{e}}^{e}$ de una transformación lineal $F:V\rightarrow W$ entre espacios euclídeos V y W de dimensión n y m respectivamente, en bases canónicas $e=(e_1,\ldots,e_n)$ y $\tilde{e}=(e_1,\ldots,e_m)$ de V y W, siendo A^tA *la matriz de Gram* del conjunto de imágenes $\{F(e_1),\ldots,F(e_n)\}$: $(A^tA)_{ij}=(F(e_i),F(e_j))$.

La descomposición anterior muestra que es siempre posible encontrar *bases ortonormales* e' y $\tilde{e}'$ de V y W en la que F tiene una representación "diagonal", con elementos diagonales reales *positivos* o nulos, es decir

$$[F]_{\tilde{e}'}^{e'}=U^t[F]_{\tilde{e}}^{e}V=A'$$

con $V = [I]_e^{e'}$, $U = [I]_{\tilde{e}}^{\tilde{e}'}$ y $F(e'_i) = \sigma_i \tilde{e}'_i$, $i = 1, \ldots, k$, con $F(e'_i) = 0$ si $i > k$. Los primeros k vectores de $\tilde{e}'$ forman pues una base ortonormal de $Im(F) = F(V)$, y los últimos $n - k$ vectores de e' una base ortonormal de $N(F)$. Notemos que los valores singulares son independientes de las bases canónicas elegidas: Si $B = R^t AS$, con $R^t R = I_m$, $S^t S = I_n \Rightarrow B^t B = S^t A^t R R^t AS = S^t A^t AS$, y los autovalores de $B^t B$ son entonces idénticos a los de $A^t A$.

Otro comentario importante es que si $A = UA'V^t \Rightarrow$

$$A^t = VA'^t U^t$$

que es necesariamente la descomposición singular de A^t. Esto muestra que los valores singulares son también las raíces de los autovalores no nulos de AA^t (matriz real simétrica de $m \times m$) y U una matriz ortonormal de autovectores de AA^t. Para la obtención de los valores singulares se puede pues diagonalizar la menor de las matrices $A^t A$ y AA^t.

Se ve también que si A es de $n \times n$ y no singular,

$$A^{-1} = VA'^{-1} U^t$$

lo que muestra que los valores singulares de A^{-1} son los inversos de los valores singulares de A (y que si A es no singular estos son necesariamente no nulos). Notemos que para A de $n \times n$, $|A| = |U||A'||V^t| = \pm|A'|$, donde $|U| = \pm 1$, $|V| = \pm 1$, por lo que $|\text{Det}[A]| = \text{Det}[A']$.

Si A representa un monomorfismo $\rightarrow$ rango$(A) = n$, por lo que $k = n \leq m$. En tal caso, conociendo la descomposición singular de A, una **inversa a izquierda** $\tilde{A}$ (de $n \times m$) puede obtenerse como

$$\tilde{A} = V\tilde{A}' U^t$$

con $\tilde{A}'$ una matriz "diagonal" de $n \times m$ de elementos $\tilde{\sigma}_i = 1/\sigma_i$, $i = 1, \ldots, n$, ya que se verifica $\tilde{A}'A' = I_n$ y por tanto $\tilde{A}A = V\tilde{A}'A'V^t = I_n$. Esto muestra asimismo que los valores singulares de $\tilde{A}$ son los inversos de los de A. En forma análoga, si A representa un epimorfismo, rango$(A) = m$, por lo que $k = m \leq n$ y una **inversa a derecha** de A estará dada por $\tilde{A} = V\tilde{A}'U^t$, pues en este caso $A'\tilde{A}' = I_m$ y $A\tilde{A} = UA'\tilde{A}'U^t = I_m$.

Una última observación general muy importante es que la descomposición singular de A permite expandir a esta como

$$A = \sum_{i=1}^{k} \sigma_i u_i v_i^t$$

lo que constituye la generalización de la expansión de una matriz simétrica A de $n \times n$ en teŕminos de autovalores y autovectores ortonormales (ver siguiente comentario). En el caso de matrices de grandes dimensiones, un método general de compresión de información (utilizado en la compresión de imágenes digitales) consiste precisamente en conservar de la expansión anterior los términos con σ_i mayor a cierto valor inferior umbral.

En el caso especial de que A sea de $n \times n$ y *simétrica* ($A^t = A$) $\Rightarrow A^t A = A^2$, por lo que $\lambda_i = (\lambda_i^A)^2$, con λ_i^A los autovalores de A. Se obtiene entonces

$$\sigma_i = |\lambda_i^A|, \quad i = 1, \ldots, k$$

es decir, los valores singulares son los *valores absolutos* de los autovalores no nulos de A. La matriz V puede entonces elegirse como la matriz de autovectores de A y U como la matriz $U = (s_1 v_1, \ldots, s_n v_n)$, con s_i el signo de λ_i. En este caso la expansión anterior se reduce a

$$A = \sum_{i=1}^{n} \lambda_i v_i v_i^t$$

con $v_i v_i^t$ la representación matricial del *proyector ortogonal* sobre el espacio generado por v_i.

Ejemplo : Consideremos

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

Tenemos

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

Los autovalores de A^tA son entonces $\lambda_\pm = 2 \pm 1$ por lo que los valores singulares son $\sigma_1 = \sqrt{3}$, $\sigma_2 = 1$. Se obtiene $v_1 = (1,1)^t/\sqrt{2}$, $v_2 = (-1,1)^t/\sqrt{2}$, y $u_1 = Av_1/\sigma_1 = (1,2,1)^t/\sqrt{6}$, $u_2 = Av_2/\sigma_2 = (-1,0,1)/\sqrt{2}$. u_3 puede elegirse, utilizando GS a partir de u_1, u_2 y $(1,0,0)$, como $(1,-1,1)/\sqrt{3}$. Se obtiene entonces

$$A = \begin{pmatrix} 1/\sqrt{6} & -1/\sqrt{2} & 1/\sqrt{3} \\ 2/\sqrt{6} & 0 & -1/\sqrt{3} \\ 1/\sqrt{6} & 1/\sqrt{2} & 1/\sqrt{3} \end{pmatrix} \begin{pmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} /\sqrt{2}$$

Algunas aplicaciones

20.1 Norma inducida de una matriz

Primeramente, consideremos una forma cuadrática real $\tilde{B}(v) = X^tBX$, con B de $n \times n$ real simétrica y $X = (x_1, \ldots, x_n)^t = [v]_e$ de $n \times 1$. Diagonalizando B, tenemos $S^tBS = B'$, con B' diagonal ($B'_{ij} = \lambda_i \delta_{ij}$) y $S = (X_1, \ldots, X_n)$ una matriz ortonormal de autovectores ($S^tS = I_n$). Por lo tanto, definiendo $X' = S^tX = (x'_1, \ldots, x'_n)$, tal que $X = SX'$, se obtiene

$$\tilde{B}(v) = X^tBX = X'^tS^tBSX' = X'^tB'X' = \sum_{i=1}^{n} \lambda_i x'^2_i$$

Como $||v||^2 = X^tX = X'^tS^tSX' = X'^tX'$, se obtiene, para $v \neq 0$,

$$Q(v) \equiv \frac{\tilde{B}(v)}{||v||^2} = \frac{X^tBX}{X^tX} = \frac{X'^tB'X'}{X'^tX'} = \frac{\sum_{i=1}^n \lambda_i x'^2_i}{\sum_{i=1}^n x'^2_i}$$

Si $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$, vemos entonces que

$$\lambda_1 \leq \frac{X^tBX}{X^tX} \leq \lambda_n$$

con el valor máximo λ_n alcanzado si $X = X_n$, con $BX_n = \lambda_n X_n$ y el mínimo λ_1 si $X = X_1$, con $BX_1 = \lambda_1 X_1$. Hemos pues demostrado que el valor máximo (mínimo) que toma la forma cuadrática X^tBX en la esfera unidad ($X^tX = 1$) es el máximo (mínimo) autovalor de B.

El cociente $Q(v)$ se denomina en contextos físicos cociente de Rayleigh y proporciona un *método variacional* para la determinación del autovalor máximo y mínimo de una matriz simétrica B:

$$\lambda_1 = \mathrm{Min}_{v\neq 0} Q(v), \quad \lambda_n = \mathrm{Max}_{v \neq 0} Q(v)$$

Consideremos ahora una transformación $F : \mathbb{R}^n \to \mathbb{R}^m$, representada en las bases canónicas por una matriz A de $m \times n$. Tenemos, para un vector no nulo $v \in \mathbb{R}^n$ tal que $[v]_e = X$,

$$\frac{||F(v)||^2}{||v||^2} = \frac{||AX||^2}{||X||^2} = \frac{(AX)^tAX}{X^tX} = \frac{X^tA^tAX}{X^tX}$$

y por lo tanto, utilizando el resultado anterior,

$$\sigma_m^2 \leq \frac{||AX||^2}{||X||^2} \leq \sigma_M^2$$

donde σ_M^2 y σ_m^2 denotan aquí el máximo y mínimo autovalor de A^tA (σ_M y σ_m son entonces los valores singulares extremos si son no nulos). Por lo tanto,

$$\sigma_m \leq \frac{||F(v)||}{||v||} \leq \sigma_M$$

Los valores σ_M y σ_m indican pues la máxima y mínima "dilatación" que puede experimentar un vector v al ser transformado por F. Si $m < n$ necesariamente $\sigma_n = 0$.

La **norma** de una matriz A de $m \times n$ (o de la transformación asociada F) **inducida por la norma del vector** se define como

$$||A|| = \mathrm{Max}_{\{X, X\neq 0\}} \frac{||AX||}{||X||} = \mathrm{Max}_{\{X, ||X||=1\}} ||AX||$$

El resultado anterior implica entonces

$$||A|| = \sigma_M$$

es decir, la norma es el mayor valor singular de A. Este resultado se denomina en realidad norma 2 de la matriz, pues está derivado de la norma $||X|| \equiv \sqrt{X^t X} = \sqrt{x_1^2 + \ldots + x_n^2}$.
Una consecuencia inmediata pero importante de esta norma es que se cumple

$$||AX|| \leq ||A||\,||X|| \;\; \forall X \in \mathbb{R}^n$$

Esta norma satisface las cuatro propiedades básicas siguientes:
1) $||A|| > 0$, con $||A|| = 0$ si y sólo si $A = 0$
2) $||\alpha A|| = |\alpha|||A||$
3) $||A + B|| \leq ||A|| + ||B||$
(pues $||A + B|| = ||(A + B)X_M||/||X_M|| \leq (||AX_M|| + ||BX_M||)/||X_M|| \leq ||A| + ||B||)$.
4) $||AB|| \leq ||A||\,||B||$ ($B \in \mathbb{R}^{m\times n}$, $A \in \mathbb{R}^{p\times m}$)
(pues $||ABX|| = ||A(BX)|| = ||A||\,||BX|| \leq ||A||\,||B||\,||X|| \;\forall\; X \in \mathbb{R}^{n\times 1}$.

20.2 Imagen de la esfera unidad
Consideremos ahora la imagen por $F : \mathbb{R}^n \to \mathbb{R}^m$ de la esfera unidad C de $\mathbb{R}^n$, es decir $F(C) = \{F(v)|\,||v|| = 1\}$. La descomposición en valores singulares permite encontrar bases canónicas e' y $\tilde{e}'$ de $\mathbb{R}^n$ y $\mathbb{R}^m$ en las que la matriz A' que representa a F es "diagonal", con elementos diagonales $\sigma_i \geq 0$. Si $[v]_{e'} = X' = (x'_1, \ldots, x'_n)^t$, con $X'^t X' = 1 \Rightarrow Y' = [F(v)]_{\tilde{e}'} = A'X' = (\sigma_1 x'_1, \ldots, \sigma_k x'_k, 0, \ldots, 0)^t$. Por lo tanto, si $k = n \leq m$ las k componentes no nulas $y'_i = x'_i \sigma_i$ de Y' satisfacen

$$\sum_{i=1}^{n} y'^2_i/\sigma_i^2 = 1$$

lo que indica que la imagen en la base $\tilde{e}'$ es la superficie de un elipsoide de dimensión $k = n$ con ejes principales en la dirección de los $\tilde{e}'_i$ y radios de longitud σ_i. Si $k < n \Rightarrow$ al menos uno de los radios es nulo y la superficie del elipsoide degenera en el interior y borde de un elipsoide de dimensión $k < n$ (en este caso $\sum_{i=1}^{k} y'^2_i/\sigma_i^2 \leq 1$). En resumen, los valores singulares determinan los radios del elipsoide obtenido como imagen por F de la esfera unidad.

20.3 Número de condición de una matriz
Consideremos un sistema de n ecuaciones con n incógnitas representado por la ecuación matricial

$$AX = Y$$

con A de $n \times n$, y X, Y de $n \times 1$. Si A es no singular la única solución está dada por $X = A^{-1}Y$. Estudiemos ahora la estabilidad de esta solución frente a variaciones δY de Y. Tenemos $\delta X = A^{-1}\delta Y$ y por lo tanto

$$\frac{||\delta X||}{||X||} = \frac{||A^{-1}\delta Y||}{||X||} \leq \frac{||A^{-1}||||\delta Y||}{||X||} \leq ||A^{-1}||\,||A||\frac{||\delta Y||}{||Y||}$$

donde en la última expresión hemos utilizado la desigualdad $||Y|| = ||AX|| \leq ||A||\,||X||$. El número de condición de una matriz se define entonces como

$$n_c(A) = ||A||\,||A^{-1}||$$

y acota la inestabilidad de la solución del sistema asociado frente a variaciones en la inhomogeneidad Y:

$$\frac{||\delta X||}{||X||} \leq n_c(A)\frac{||\delta Y||}{||Y||}$$

En virtud del resultado previo, se tiene, utilizando la norma 2, $||A|| = \sigma_M$, $||A^{-1}|| = 1/\sigma_m$, con σ_M y σ_m el máximo y mínimo valor singular, y por lo tanto

$$n_c(A) = \sigma_M/\sigma_m \geq 1$$

El número de condición es entonces adimensional y queda determinado por el cociente entre los valores singulares extremos. Para matrices reales simétricas, $\sigma_M = |\lambda_M|$, $\sigma_m = |\lambda_m|$, con λ_M y λ_m los autovalores de mayor y menor valor absoluto respectivamente. Nótese que si la matriz A es singular, $\sigma_m = 0$ y en tal caso $n_c(A) = \infty$. Números de condición grandes indican matrices "cuasi singulares" (o mal condicionadas), para las que no se puede asegurar estabilidad en la solución del sistema asociado.

Es importante destacar que la estabilidad frente a variaciones en la matriz A queda también determinada por el mismo número de condición. Si $AX = Y$ y $(A + \delta A)(X + \delta X) = Y$, entonces, a primer orden en δX y δA, se obtiene $(\delta A)X + A\delta X = 0$ y

$$\delta X = -A^{-1}(\delta A)X$$

Por lo tanto

$$||\delta X|| = ||A^{-1}(\delta A)X|| \leq ||A^{-1}||\,||\delta A||\,||X||$$

de donde

$$\frac{||\delta X||}{||X||} \leq ||A^{-1}||\,||\delta A|| = n_c(A)\frac{||\delta A||}{||A||}$$

Ejemplo: Si

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

entonces

$$A^t A = \begin{pmatrix} \varepsilon^2 & 0 \\ 0 & 1 \end{pmatrix}$$

por lo que los valores singulares son $|\varepsilon|$ y 1 y el número de condición es

$$n_c(A) = 1/|\varepsilon|$$

si $|\varepsilon| \leq 1$. Notemos que $\mathrm{Det}[A] = -\varepsilon$ y que $n_c(A) \to \infty$ si $\varepsilon \to 0$. La solución al sistema $AX = Y$ es $X = (y_2/\varepsilon, y_1)^t$ con $\delta X = (\delta y_2/\varepsilon, \delta y_1)^t$ y $||\delta X||^2/||X||^2 = (\delta y_2^2/\varepsilon^2 + \delta y_1^2)/(y_2^2/\varepsilon^2 + y_1^2)$. Si por ejemplo $y_2 = 0$, $y_1 = 1$ y $\delta y_1 = 0 \Rightarrow ||\delta X||/||X|| = |\delta y_2|/|\varepsilon| = n_c(A)||\delta Y||/||Y||$, por lo que $||\delta X||/||X||$ puede ser mucho mayor que $||\delta Y||/||Y||$ cuando ε es suf. pequeño.

Notemos en cambio que la matriz

$$B = \begin{pmatrix} \varepsilon & 0 \\ 0 & \varepsilon \end{pmatrix}$$

tiene número de condición 1 a pesar de que $\mathrm{Det}[B] = \varepsilon^2 \ll 1$ para $|\varepsilon| \ll 1$.

20.4 Pseudoinversa

Sea $A \in \mathbb{R}^{m\times n}$ con $A = UA'V^t = \sum_{i=1}^k \sigma_i u_i v_i^t$ su DVS. La pseudoinversa de A (denominada también pseudoinversa de Moore-Penrose) es una matriz $\tilde{A} \in \mathbb{R}^{n\times m}$ definida como

$$\tilde{A} = V\tilde{A}'U^{\mathrm{tr}} = \sum_{i=1}^{k} \frac{1}{\sigma_i} v_i u_i^t$$

con $\tilde{A}'$ una matriz de $n \times m$ de elementos diagonales $1/\sigma_i$ ($A'_{ij} = \delta_{ij}/\sigma_i$ si $i \leq k$ y 0 en caso contrario). Dado que $u_i^t u_j = \delta_{ij}$, $v_i^t v_j = \delta_{ij}$, se verifica que $A\tilde{A} = \sum_{i=1}^k u_i u_i^t$ es el **proyector ortogonal sobre el espacio columna de la matriz**, mientras que $\tilde{A}A = \sum_{i=1}^k v_i v_i^t$ es el proyector ortogonal sobre el espacio fila (es decir, sobre el espacio columna de A^t). Se verifica entonces

$$\tilde{A}A\tilde{A} = \tilde{A}\,, \quad A\tilde{A}A = A$$

Es facil ver que si rango(A)$= n \Rightarrow \tilde{A} = (A^tA)^{-1}A^t$, coincidiendo con una inversa a izquierda de A, mientras que si rango(A)$= m \Rightarrow \tilde{A} = A^t(AA^t)^{-1}$, coincidiendo con una inversa a derecha de A. Si rango(A)$= n = m$ $\Rightarrow \tilde{A} = A^{-1}$ es la inversa de A.

Consideremos ahora el sistema de ecuaciones lineales de $m \times n$

$$AX = b$$

donde $X \in \mathbb{R}^{n\times 1}$ y $b \in \mathbb{R}^{m\times 1}$. Si el sistema es compatible, $b = A\tilde{A}b$ (pues $b \in EC(A)$) y entonces una solución particular del sistema es

$$X = \tilde{A}b$$

pues $AX = A\tilde{A}b = b$. Si no existe solución ($b \not\in EC(A)$) entonces $X = \tilde{A}b$ es el vector que minimiza la diferencia $||AX - b||$, pues $A\tilde{A}b$ es la proyección ortogonal de b sobre $EC(A)$.

En el caso compatible, la solución general del sistema $AX = b$ puede expresarse como

$$X = \tilde{A}b + (I_n - \tilde{A}A)v$$

con v un vector arbitrario de $\mathbb{R}^n$. El segundo término es un vector general del núcleo de A, pues $I_n - \tilde{A}A$ es el proyector ortogonal sobre $\mathrm{Nu}(A)$ ($A(I - \tilde{A}A) = (A - A) = 0$), y representa una solución general del sistema homogéneo $AX = 0$. El primer término $\tilde{A}b$ es una solución particular de $AX = b$, y es **la solución particular de norma mínima**, pues es ortogonal a $(I - \tilde{A}A)w\ \forall\ w$ (ya que pertence al espacio fila de A).

En el caso general no necesariamente compatible, $X = \tilde{A}b$ es el vector de norma mínima que minimiza $||AX - b||$.

21. Espacios semieuclídeos y pseudoeuclídeos

Resumen. Para $\dim V = 2$ estos espacios quedan definidos por una forma bilineal $(v, w)_G = [v]_e^t[G]_e[w]_e$, con

$$[G]_e = \left(\begin{smallmatrix}0&0\\0&1\end{smallmatrix}\right)$$

en el caso semieuclídeo, tal que $(v, w)_G = yy'$, $(v, v)_G = y^2$ si $[v]_e = \left(\begin{smallmatrix}x\\y\end{smallmatrix}\right)$, $[w]_e = \left(\begin{smallmatrix}x'\\y'\end{smallmatrix}\right)$, y

$$[G]_e = \left(\begin{smallmatrix}1&0\\0&-1\end{smallmatrix}\right)$$

en el pseudoeuclídeo, tal que $(v, w)_G = xx' - yy'$, $(v, v)_G = x^2 - y^2$. En estos casos $(v, v)_G$ puede ser 0 aun si $v \neq 0$, y en el caso pseudoeuclídeo puede ser también negativo.
Se demostró en clase que las transformaciones reales $\left(\begin{smallmatrix}x\\y\end{smallmatrix}\right) = S\left(\begin{smallmatrix}x'\\y'\end{smallmatrix}\right)$ que preservan estas formas bilineales (tales que $[G]_{e'} = S^t[G]_eS = [G]_e$) corresponden en el caso semieuclídeo a

$$S = \left(\begin{smallmatrix}a&b\\0&d\end{smallmatrix}\right)$$

con $d = \pm 1$, y a, b arbitrarios, $a \neq 0$, y en el caso pseudoeuclídeo a

$$S = \left(\begin{smallmatrix}s\cosh(z)&s'\sinh(z)\\s\sinh(z)&s'\cosh(z)\end{smallmatrix}\right)$$

con $s = \pm 1$, $s' = \pm 1$ y z arbitrario.
En particular, estas transformaciones comprenden las transformaciones de Galileo

$$\left(\begin{smallmatrix}x\\t\end{smallmatrix}\right) = \left(\begin{smallmatrix}1&v\\0&1\end{smallmatrix}\right)\left(\begin{smallmatrix}x'\\t'\end{smallmatrix}\right)$$

en el caso semieuclídeo ($a = d = 1$, $b = v$) y las transformaciones de Lorentz

$$\left(\begin{smallmatrix}x\\ct\end{smallmatrix}\right) = \left(\begin{smallmatrix}\cosh z&\sinh z\\\sinh z&\cosh z\end{smallmatrix}\right)\left(\begin{smallmatrix}x'\\ct'\end{smallmatrix}\right)$$

en el caso pseudoeuclídeo, con $\tanh(z) = v/c$, $s = s' = 1$, tal que $\cosh z = \frac{1}{\sqrt{1-v^2/c^2}}$, $\sinh z = \frac{v/c}{\sqrt{1-v^2/c^2}}$.
Para $v/c \to 0$, las transformaciones de Lorentz en las variables (x, t) se reducen a las de Galileo:

$$\left(\begin{smallmatrix}x\\t\end{smallmatrix}\right) = \left(\begin{smallmatrix}\cosh z&c\sinh z\\\frac{1}{c}\sinh z&\cosh z\end{smallmatrix}\right)\left(\begin{smallmatrix}x'\\t'\end{smallmatrix}\right) \underset{v/c\to 0}{\longrightarrow} \left(\begin{smallmatrix}1&v\\0&1\end{smallmatrix}\right)\left(\begin{smallmatrix}x'\\t'\end{smallmatrix}\right)$$

Recordemos que para $n = 2$, las transformaciones que dejan invariante el producto escalar euclídeo son de la forma

$$S = \left(\begin{smallmatrix}s\cos\theta&-s'\sin\theta\\s\sin\theta&s'\cos\theta\end{smallmatrix}\right)$$

con $s = \pm 1$, $s' = \pm 1$, que representan rotaciones (si $|S| = ss' = 1$) o reflexiones ($ss' = -1$).

22 Formas bilineales complejas

Sea V un espacio vectorial sobre el cuerpo de los complejos $\mathbb{C}$. Una función $A : V \times V \to \mathbb{C}$ se dice que es una forma bilineal *hermítica* si

$$A(v_1 + v_2, w) = A(v_1, w) + A(v_2, w), \quad A(v, w_1 + w_2) = A(v, w_1) + A(v, w_2)$$

$$A(v, \alpha w) = \alpha A(v, w), \quad A(\alpha v, w) = \alpha^* A(v, w)$$

$\forall\ v, v_1, v_2, w, w_1, w_2 \in V$ y $\alpha \in \mathbb{C}$. Nótese que α sale como conjugado cuando está en el primer miembro.
Si V es de dimensión finita n y $e = (e_1, \ldots, e_n)$ es una base de V, escribiendo $v = \sum_{1=1}^{n} \alpha_i e_i$, $w = \sum_{j=1}^{n} \beta_j e_j$, con $[v]_e = (\alpha_1, \ldots, \alpha_n)^t$, $[w]_e = (\beta_1, \ldots, \beta_n)^t$, se obtiene

$$A(v, w) = \sum_{i,j=1}^{n} \alpha_i^* \beta_j A(e_i, e_j) = [v]_e^\dagger [A]_e [w]_e$$

donde el símbolo $\dagger$ denota traspuesto conjugado ($[v]_e^\dagger \equiv ([v]_e^t)^*$) y $[A]_e$ es la matriz de $n \times n$ de elementos

$$([A]_e)_{ij} = A(e_i, e_j)$$

Ejemplo: La siguiente es una forma bilineal de $\mathbb{C} \times \mathbb{C} \to \mathbb{C}$:

$$\begin{aligned} A(v, w) &= \alpha_1^* \beta_1 + (1 + i)\alpha_1^* \beta_2 + (1 - i)\alpha_2^* \beta_1 + 2\alpha_2^* \beta_2 \\ &= (\alpha_1^*, \alpha_2^*) \begin{pmatrix} 1 & 1 + i \\ 1 - i & 2 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} \end{aligned}$$

donde hemos escrito en la base canónica $v = (\alpha_1, \alpha_2) = \alpha_1 e_1 + \alpha_2 e_2$, $w = (\beta_1, \beta_2) = \beta_1 e_1 + \beta_2 e_2$. A queda entonces representada en esta base por la matriz

$$[A]_e = \begin{pmatrix} 1 & 1 + i \\ 1 - i & 2 \end{pmatrix}$$

con $A(e_1, e_1) = 1$, $A(e_1, e_2) = 1 + i$, $A(e_2, e_1) = 1 - i$, $A(e_2, e_2) = 2$.

Si $A(v, w) = A(w, v)^*\ \forall v, w \Rightarrow$ la forma bilineal se dice que es hermíticamente simétrica y si $A(v, w) = -A(w, v)^*$, hermíticamente antisimétrica En el primer caso, la matriz que la representa es *hermítica*: $[A]_e^\dagger = [A]_e$, ya que $A(e_i, e_j) = A(e_j, e_i)^*$, y en el segundo caso *antihermítica*: $[A]_e^\dagger = -[A]_e$. Análogamente, si $[A]_e^\dagger = \pm[A]_e$, A es herm. simétrica ($+$) o antisimétrica ($-$). El ejemplo anterior corresponde a una forma bilineal herm. simétrica.
Notemos que una forma bilineal compleja que satisface las 4 condiciones no puede ser simplemente simétrica o antisimétrica a no ser que sea nula: Si $A(v, w) = \pm A(w, v)\ \forall\ v, w, \Rightarrow A(\alpha v, w) = \alpha^* A(v, w) = \pm A(w, \alpha v) = \pm\alpha A(w, v) = \alpha A(v, w)\ \forall\ \alpha,\ v, w$, lo que implica $(\alpha - \alpha^*)A(v, w) = 0$, es decir $A(v, w) = 0\ \forall\ v, w$.
Notemos también que toda forma bilineal puede expresarse como suma de una forma bilineal herm. simétrica y una forma bilineal herm. antisimétrica:

$$A(v, w) = \frac{1}{2}[A(v, w) + A(w, v)^*] + \frac{1}{2}[A(v, w) - A(w, v)^*]$$

22.1 Formas cuadráticas complejas

En forma análoga al caso real, la función $Q : V \to \mathbb{C}$ definida por

$$Q(v) = A(v, v)$$

se denomina *forma cuadrática* y satisface

$$Q(\alpha v) = A(\alpha v, \alpha v) = \alpha^* \alpha A(v, v) = |\alpha|^2 Q(v)$$

Una diferencia importante con las formas bilineales reales es que ahora la forma cuadrática determina completamente la forma bilineal (y no solamente la parte simétrica, como en el caso real). En efecto, podemos expandir $Q(v + w) = A(v + w, v + w)$ y $Q(v + iw) = A(v + iw, v + iw)$ como

$$Q(v + w) = Q(v) + Q(w) + A(v, w) + A(w, v),$$

$$Q(v+iw) = Q(v)+Q(w)+i[A(v,w)-A(w,v)]$$

de donde

$$\begin{array}{rcl} A(v,w) &=& Q(v+w) - iQ(v+iw) - (1-i)(Q(v)+Q(w)) \\ A(w,v) &=& Q(v+w) + iQ(v+iw) - (1+i)(Q(v)+Q(w)) \end{array}$$

De aquí se deduce también una propiedad fundamental:
$Q(v)$ es *real* $\forall\ v$ si y sólo si $A(v,w) = [A(w,v)]^*\ \forall\ v, w$, es decir, sii la forma bilineal asociada es hermíticamente simétrica.
En efecto, de las expresiones anteriores se ve que si $Q(v) \in \mathbb{R}\ \forall\ v \in V \Rightarrow A(w,v) = [A(v,w)]^*\ \forall\ v, w \in V$.
Y si $A(w,v) = [A(v,w)]^*\ \forall\, v, w \Rightarrow Q(v) = A(v,v) = [A(v,v)]^*$ es real. Formas cuadráticas *reales* determinan pues *formas bilineales hermíticamente simétricas y viceversa.*
Por otro lado, si $A(v,w)$ es hermíticamente simétrica, $A'(v,w) = iA(v,w)$ es hermíticamente antisimétrica. La forma cuadrática asociada $Q'(v) = A'(v,v)$ es obviamente imaginaria.
Ejemplo: La forma bilineal del ej. anterior origina la forma cuadrática

$$\begin{array}{rcl} Q(v) = A(v,v) = (\alpha_1^*, \alpha_2^*)\begin{pmatrix} 1 & 1+i \\ 1-i & 2 \end{pmatrix}\begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} &=& \alpha_1^*\alpha_1 + 2\alpha_2^*\alpha_2 + (1+i)\alpha_1^*\alpha_2 + (1-i)\alpha_2^*\alpha_1 \\ &=& |\alpha_1|^2 + 2|\alpha_2|^2 + 2\mathrm{Re}[(1+i)\alpha_1^*\alpha_2] \end{array}$$

que es obviamente *real*.

22.2 Cambio de base

Si efectuamos un cambio de base $e'_i = \sum_{j=1}^n S_{ji} e_j$, con $|S| \neq 0$,

$$A(e'_i, e'_j) = A(\sum_{k=1}^n S_{ki} e_k, \sum_{l=1}^n S_{lj} e_l) = \sum_{k,l} S_{ki}^* S_{lj} A(e_k, e_l)$$

por lo que

$$[A]_{e'} = S^\dagger [A]_e S$$

donde $\dagger$ denota por su puesto la operación de traspuesto+conjugado.
Notemos que $\mathrm{Det}([A]_{e'}) = |\mathrm{Det}(S)|^2 \mathrm{Det}([A]_e)$, por lo que la fase del determinante es la misma en cualquier base. Obtenemos entonces

$$A(v,w) = [v]_e^\dagger [A]_e [w]_e = [v]_{e'}^\dagger [A]_{e'} [w]_{e'}$$

donde $[w]_{e'} = R[w]_e$, $[v]_{e'}^\dagger = [v]_e R^\dagger$ y $R = S^{-1}$.

22.3 Base canónica: Si A es herm. simétrica, $\exists$ *una base* e' (base canónica) donde $[A]'_e$ es diagonal:

$$[A]_{e'} = S^\dagger [A]_e S = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ & & \dots & \\ 0 & 0 & \dots & \lambda_n \end{pmatrix}$$

En esta base, si $v = \sum_{i=1}^n \alpha'_i e'_i$, $w = \sum_{i=1}^n \beta'_i e'_i$, tenemos

$$A(v,w) = \sum_{i=1}^n \lambda_i {\alpha'_i}^* \beta'_i$$

La demostración de la existencia de esta base puede efectuarse en forma similar al caso real, completando ahora módulos cuadrados, y se deja comos ejercicio. Sug.: Llamando $a_{ij} = ([A]_e)_{ij}$ (con $a_{ji} = a_{ij}^*$) y asumiendo $a_{nn} \neq 0$, escribir la parte que contiene α_n y α_n^* en $A(v,v)$ como

$$a_{nn}\alpha_n^*\alpha_n + \sum_{j=1}^{n-1}(a_{nj}\alpha_n^*\alpha_j + a_{nj}^*\alpha_j^*\alpha_n) = a_{nn}{\alpha'_n}^*\alpha'_n - (\sum_{j=1}^{n-1} a_{nj}^*\alpha_j^*)(\sum_{j=1}^{n-1} a_{nj}\alpha_j)/a_{nn}$$

con $\alpha'_n = \alpha_n + \frac{1}{a_{nn}} \sum_{j=1}^{n-1} a_{nj}\alpha_j$, y proceder luego por inducción. Si $a_{nn} = 0$ se comienza con una variable α_i tal que $a_{ii} \neq 0$, y si $a_{ii} = 0$ $\forall i$ se efectúa un cambio de variables simple para que a_{ii} sea no nulo para algún i (por ej., si $a_{ij} = a^*_{ji} \neq 0$, $a_{ij}\alpha^*_i\alpha_j + a_{ji}\alpha^*_j\alpha_i = 2|a_{ij}|^2(|\alpha'_i|^2 - |\alpha'_j|^2)$, con $\alpha_i = a_{ij}(\alpha'_i + \alpha'_j)$, $\alpha_j = \alpha'_i - \alpha'_j$. El cambio $\alpha'_i = \sum_{j=1}^n R_{ij}\alpha_j$ define una base $e'_i = \sum_{j=1}^n S_{ji}e_j$, con $S = R^{-1}$, en la que $[A]_{e'} = S^\dagger[A]S$ es diagonal.

Otra forma de demostrar la existencia es mediante la diagonalización de la matriz $[A]_e$, que es en este caso hermítica y por lo tanto diagonalizable en una base ortonormal, tal que $S^{-1} = S^\dagger$ y $S^\dagger[A]_eS$ es diagonal. No obstante esto supone haber demostrado antes que tales matrices son diagonalizables, lo que en este curso se realizar luego.

La base canónica no es única. Una base canónica puede obtenerse, al igual que en el caso real, completando módulos cuadrados o bien diagonalizando la matriz $[A]_e$. Además, si e es una base canónica, $e' = (e'_1, \ldots, e'_n)$, con $e'_i = \alpha_i e_i$, $\alpha_i \neq 0$, es otra base canónica:

$$A(e'_i, e'_j) = \alpha^*_i\alpha_j A(e_i, e_j) = |\alpha_i|^2\lambda_i\delta_{ij}$$

Eligiendo α_i tal que $|\alpha_i| = 1/\sqrt{|\lambda_i|}$ si $\lambda_i \neq 0$, podemos siempre obtener una base en la que

$$A(e'_i, e'_j) = \lambda'_i\delta_{ij}, \;\; \text{con} \;\; \lambda'_i = 0 \text{ o } \pm 1$$

En algunos contextos se denomina base canónica sólo a estas tipo de bases, es decir, aquellas en las que $A(e_i, e_i)$ es o bien ± 1 o 0, con $A(e_i, e_j) = 0$ si $i \neq j$.

Ejemplo: Hallar una base canónica para el ejemplo previo. Completando módulos cuadrados, obtenemos

$$Q(v) = (\alpha^*_1 + (1-i)\alpha^*_2)(\alpha_1 + (1+i)\alpha_2) + |\alpha_2|^2[2 - (1+i)(1-i)] = |\alpha'_1|^2 + 0|\alpha'_2|^2$$

donde $\alpha'_1 = \alpha_1 + (1+i)\alpha_2$, $\alpha'_2 = \alpha_2$, o sea $\binom{\alpha'_1}{\alpha'_2} = \left(\begin{smallmatrix}1 & 1+i\\ 0 & 1\end{smallmatrix}\right)\binom{\alpha_1}{\alpha_2}$. La matriz de cambio de base es entonces

$$S = \begin{pmatrix} 1 & 1+i \\ 0 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & -1-i \\ 0 & 1 \end{pmatrix}$$

y se verifica

$$[A]_{e'} = S^\dagger[A]_eS = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

Alternativamente, diagonalizando la matriz $[A]_e$ se obtienen los autovalores y autovectores

$$\lambda_1 = 1, \; v'_1 = (1+i, 2), \qquad \lambda_2 = 0, \; v_2 = (-1-i, 1)$$

Normalizando los autovectores, la matriz de cambio de base es entonces

$$S = \begin{pmatrix} (1+i)/\sqrt{6} & -(1+i)/\sqrt{3} \\ 2/\sqrt{6} & 1/\sqrt{3} \end{pmatrix}$$

con $S^{-1} = S^\dagger$ (pues los autovectores en S están normalizados). Se obtiene así la reprentación diagonal

$$[A]_{e'} = S^\dagger[A]_eS = \begin{pmatrix} 3 & 0 \\ 0 & 0 \end{pmatrix}$$

Vemos que el número de coeficientes diagonales positivos y nulos en las dos formas diagonales obtenidas es el mismo. Esta propiedad es general y constituye el

22.4 Teorema de Inercia para formas cuadráticas hermíticas: Si Q es una forma cuadrática herm. simétrica, el número de términos diagonales positivos, negativos, y nulos en una representación diagonal arbitraria es siempre el mismo. Se demuestra igual que en el caso real (Demostrar como ejercicio).
Es importante notar que el teorema de inercia no vale para formas cuadráticas comunes extendidas a los complejos: Si $Q(v) = \alpha_1^2 + \alpha_2^2$, la transformación $\alpha'_1 = i\alpha_1$, $\alpha'_2 = \alpha_2$ la lleva a $-\alpha'^2_1 + \alpha'^2_2$. Tal forma cuadrática no proviene de una forma bilineal hermítica, ya que no cumple $Q(\alpha v) = |\alpha|^2Q(v)$.

22.5 Formas cuadráticas positivas:
Una forma cuadrática se denomina *definida positiva* (o estríctamente positiva) si $Q(v) > 0 \ \forall \ v \neq 0$, y semipositiva (o no negativa) si $Q(v) \geq 0 \ \forall \ v$ (obviamente, en cualquier caso, $Q(0) = 0$). Por ser reales, estas formas cuadráticas están necesariamente asociadas a formas bilineales hermíticamente simétricas.
La forma cuadrática es pues definida positiva si y sólo si los coeficientes diagonales en una base canónica satisfacen $a_{ii} > 0 \ \forall \ i$ y semipositiva sii $a_{ii} \geq 0 \ \forall \ i$. El teorema de inercia asegura que el número de coeficientes diagonales positivos y nulos para estas formas cuadráticas (pero no su valor particular) es siempre el mismo en cualquier base canónica.
En form análoga, una matriz $A \in \mathbb{C}^{n\times n}$ se dice definida positiva si $X^\dagger AX > 0 \ \forall \ X \neq 0$, $X \in \mathbb{C}^n$, en cuyo caso podemos considerarla como la representación en la base canónica de $V = \mathbb{C}^n$ de una forma cuadrática definida positiva. Notemos que necesariamente A debe ser *hermítica* ($A^\dagger = A$), para que $X^\dagger AX$ sea real.
Una matriz hermítica A es pues definida positiva si y sólo si todos sus autovalores son positivos.

Notemos que si $Q_A(v)$ es una forma cuadrática definida positiva $\Rightarrow$ *existe una base canónica e'' donde* $A(e''_i, e''_j) = \delta_{ij}$, es decir,

$$[A]_{e''} = I_n$$

(matriz identidad). En efecto, existirá una base canónica, obtenida completando módulos cuadrados o diagonalizando, en la que $A(e'_i, e'_j) = ([A]_{e'})_{ij} = \lambda_i\delta_{ij}$, con $\lambda_i > 0 \ \forall \ i$. En la nueva base definida por $e''_i = e'_i/\sqrt{\lambda_i}$ tendremos $A(e''_i, e''_j) = A(e'_i, e'_j)/\sqrt{\lambda_i\lambda_j} = \lambda_i\delta_{ij}/\sqrt{\lambda_i\lambda_j} = \delta_{ij}$ para $i, j = 1, \ldots, n$.
Esto implica que existe una matriz $S = [I]_e^{e''}$ no singular tal que

$$[A]_{e''} = S^\dagger [A]_e S = I_n$$

con I_n la matriz identidad.
Esto implica a su vez que toda matriz $A \equiv [A]_e$ definida positiva puede escribirse como

$$A = R^\dagger R$$

con $R = S^{-1}$ no singular. La recíproca es también válida: La matriz $R^\dagger R$ es definida positiva $\forall \ R$ no singular (probar como ejercicio).

Para saber si una forma cuadrática es definida positiva o semipositiva se completan módulos cuadrados o se obtienen los autovalores de la matriz que la representa en alguna base, y se observan lo signos de los coeficientes diagonales resultantes. El determinante de una matriz asociada a una forma cuadrática definida positiva es obviamente positivo en cualquier base ($\text{Det}[A] = \text{Det}[R^\dagger R] = |\text{Det}[R]|^2$), aunque esta condición no garantiza que A sea definida positiva.

Es válido no obstante el siguiente teorema: **Una matriz hermítica A de $n \times n$ es definida positiva si y sólo si todos sus determinantes principales son positivos ($\text{Det}(A_m) > 0$, $m = 1, \ldots, n$).**
Dem.: Si A es definida positiva la forma cuadrática asociada debe ser positiva en cualquier subespacio de V y por lo tanto cualquier submatriz de A (obtenida quitando un conjunto de filas y las resp. columnas) es definida positiva. En particular, todas las submatrices principales ($A_m = [a_{ij}], i, j = 1, \ldots, m$, $m = 1, \ldots, n$) son definidas positivas y sus determinantes por ende positivos.
Recíprocamente, si todos los determinantes principales son positivos, procedemos por inducción sobre n. Para $n = 1$ se cumple trivialmente. Asumiendo la submatriz principal de $(n-1) \times (n-1)$ definida positiva, existirá una base $(e'_1, \ldots, e'_{n-1})$ del subespacio correspondiente en la que $A(e'_i, e'_j) = \delta_{ij}$. Definimos ahora $e'_n = e_n - \sum_{i=1}^{n-1} \alpha_i e'_i$, con $\alpha_i = A(e'_i, e_n)$, tal que $A(e'_i, e'_n) = A(e'_i, e_n) - \alpha_i = 0$ para $i = 1, \ldots, n-1$. En tal caso, $[A]_{e'}$ será diagonal y con todos sus elementos diagonales positivos, pues $A(e'_i, e'_i) = 1$ para $i < n$ y $A(e'_n, e'_n) = \text{Det}([A]_{e'}) > 0$ por hipótesis (el signo del determinante no cambia al cambiar la base).

Recordemos también aqui los **círculos de Gershgorin:** Si $A \in \mathbb{C}^{n\times n}$ es una matriz cuadrada de elementos a_{ij}, entonces sus autovalores se encuentran en la unión de los círculos (en el plano complejo)

$$|\lambda - a_{ii}| \leq |\sum_{j\neq i} |a_{ij}|$$

En efecto, sea $v = (x_1, \ldots, x_n)^t \in \mathbb{C}^{n\times 1}$, $v \neq 0$, un autovector de A asociado al autovalor λ, tal que $Av = \lambda v$, es decir, $\sum_j a_{ij}x_j = \lambda x_i$. Si $|x_i| = \text{Max}[|x_1|, \ldots, |x_n|] \neq 0$ es el módulo de la coordenada de v de módulo máximo, tenemos, dado que $(\lambda - a_{ii})x_i = \sum_{j\neq i} a_{ij}x_j$,

$$|\lambda - a_{ii}| = |\sum_{j\neq i} a_{ij}x_j/x_i| \leq \sum_{j\neq i} |a_{ij}| x_j/x_i| \leq \sum_{j\neq i} |a_{ij}|$$

Una desigualdad simular es válida para sumas sobre columnas, ya que los autovalores de A son idénticos a los de A^t.

En el caso de matrices hermíticas, tanto los elementos diagonales como los autovalores son todos reales. La cota anterior implica entonces la siguiente condición suficiente (aunque no necesaria) de positividad de una matriz hermítica A: Si $a_{ii} > 0 \;\forall\; i$ y $\sum_{j\neq i} |a_{ij}| < a_{ii} \;\forall\; i$, los autovalores serán todos positivos y por ende A será definida positiva.

23 Espacios Unitarios (Espacios de Hilbert)

Un espacio vectorial V sobre $\mathbb{C}$ se denomina unitario o espacio de Hilbert si está equipado con una operación $V \times V \to \mathbb{C}$, denominada *producto interno* o producto escalar, y denotada por (v, w), que satisface

$$(v, w) = (w, v)^*, \quad (v, \alpha w) = \alpha(v, w), \quad (v, w_1 + w_2) = (v, w_1) + (v, w_2)$$

$$(v, v) > 0 \;\forall\; v \neq 0$$

Es decir, el producto interno no es otra cosa que una forma bilineal hermíticamente simétrica y *definida positiva*. En el caso de dimensión infinita, un espacio de Hilbert debe ser además completo: Si $\{u_n\}$ es una sucesión de vectores tal que $\sum_{n=0}^{\infty} ||u_n||$ es convergente entonces $\lim_{n\to\infty} u_n$ debe pertenecer al espacio.

En en el caso de dimensión finita, en una base arbitraria e tendremos, denotando con $[A]_e$ la matriz de elementos $a_{ij} = (e_i, e_j) = a_{ji}^*$,

$$(v, w) = [v]_e^\dagger [A]_e [w]_e = \sum_{i,j=1}^{n} \alpha_i^* a_{ij} \beta_j$$

donde $v = \sum_{i=1}^{n} \alpha_i e_i$, $w = \sum_{i=1}^{n} \beta_i e_i$
Y si e denota ahora la base canónica en la que $(e_i, e_j) = \delta_{ij}$, obtenemos la forma corriente

$$(v, w) = [v]_e^\dagger [w]_e = \sum_{i=1}^{n} \alpha_i^* \beta_i$$

Esta base es una *base ortonormal* para el producto escalar ($(e_i, e_j) = \delta_{ij}$). En esta base,

$$(v, v) = [v]_e^\dagger [v]_e = \sum_{i=1}^{n} |\alpha_i|^2$$

Ejemplo: En $\mathbb{C}^n$, el producto interno usual en la base canónica está dado por

$$(v, w) = \sum_{i=1}^{n} x_i^* y_i, \quad (w, v) = \sum_{i=1}^{n} y_i^* x_i = (v, w)^*$$

para $v = (x_1, \ldots, x_n)$, $w = (y_1, \ldots, w_n)$, lo que implica $(v, v) = \sum_{i=1}^{n} |x_i|^2$.

Ejemplo: En el espacio de funciones complejas de parte real e imaginaria continua, $C_{[a,b]} = \{f : \mathbb{R} \to \mathbb{C},\;\; f = f^r + if^i,\; f^r, f^i \in R_{[a,b]}\}$, con $a < b$, el producto interno usual está dado por

$$(f, g) = \int_a^b f^*(x) g(x) dx = (g, f)^*$$

con $(f, f) = \int_a^b f^*(x) f(x) dx = \int_a^b |f(x)|^2 dx > 0$ si $f \neq 0$.
Ejemplo: En el caso de matrices complejas de $m \times n$, $V = \mathbb{C}^{m\times n}$, podemos definir el producto escalar

$$(A, B) = \mathrm{Tr}\,[A^\dagger\, B] = \sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij}^* B_{ij}$$

con $(A,A)=\sum_{i,j}|A_{ij}|^2>0\;\forall\;A\neq 0$.

En los espacios unitarios son válidas propiedades similares a las de espacios euclídeos. En particular:

La **norma** de un vector se define por

$$||v||=\sqrt{(v,v)}\geq 0$$

con $||v||=0$ sii $v=0$ y $||\alpha v||=|\alpha|\,||v||$. La distancia entre dos vectores es $d(v,w)=||v-w||$.

La desigualdad de **Cauchy-Schwarz** también se verifica:

$$|(v,w)|\leq ||v||\,||w||$$

donde la igualdad vale si y sólo si $\{v,w\}$ son LD.
Demostración: Si $v=0$ o $w=0$ la igualdad se cumple trivialmente: $0=(v,w)=||v||\,||w||$.
Idem si v y w son LD: En tal caso $w=\alpha v$ (o $v=\alpha w$) y por lo tanto $|(v,w)|=|\alpha(v,v)|=|\alpha|\,||v||^2=||v||\,||w||$.
Si $v\neq 0$ y $w\neq 0$, denotemos con $v_n=v/||v||$, $w_n=w/||w_n||$ los vectores normalizados ($||v_n||=||w_n||=1$), tal que $(v,w)=(v_n,w_n)||v||\,||w||$. Se obtiene, para s un número complejo arbitrario de módulo 1 ($|s|=1$),

$$0\leq(v_n-sw_n,v_n-sw_n)=||v_n||^2+|s|^2||w_n||^2-s(v_n,w_n)-s^*(w_n,v_n)=2-2\mathrm{Re}[s(v_n,w_n)]=2(1-\mathrm{Re}[s(v_n,w_n)])$$

Recordemos ahora que todo número complejo z puede escribirse como $z=|z|e^{i\phi}$, con $|z|=\sqrt{zz^*}$ (módulo) y ϕ reales. Por lo tanto, si $z=(v_n,w_n)=|(v_n,w_n)|e^{i\phi}$, eligiendo $s=e^{-i\phi}$ se obtiene

$$0\leq 1-|(v_n,w_n)|$$

de donde $|(v_n,w_n)|\leq 1$. Por lo tanto, $|(w,v)|=|(v,w)|\leq ||v||\,||w||$, q.e.d.
Además, si $|(w,v)|=||w||||v||\Rightarrow|(w_n,v_n)|=1$ y $(v_n-sw_n,v_n-sw_n)=0$, por lo que $v_n-sw_n=0$, es decir, $v=sw||v_n||/||w_n||$, lo que implica que v,w son L.D.
Las desigualdades triangulares *permanecen válidas* en espacios unitarios, por la vigencia de la desigualdad anterior: $|||v||-||w|||\leq||v+w||\leq||v||+||w||$.
No obstante, no se pueden definir ahora ángulos entre vectores pues $(v,w)/(||v||\,||w||)$, si bien tiene módulo menor que 1, es en general complejo.
Ejemplo: Dados $v=(1+i,i),w=(i,1+i)\in\mathbb{C}^2$, tenemos

$$(v,w)=(1-i)i+(-i)(1+i)=2\leq||v||\,||w||=\sqrt{|1+i|^2+1}\sqrt{1+|1+i|^2}=\sqrt{3}\sqrt{3}=3$$

Notación de Mecánica Cuántica:
La notación empleada en mecánica cuántica para los vectores de estado de un sistema (que pertenecen a un espacio de Hilbert) es $|v\rangle$, y para el producto interno $\langle w|v\rangle$. Es decir,

$$v\rightarrow|v\rangle,\;\;(w,v)\rightarrow\langle w|v\rangle,\;\;\text{con}\;\;\langle v|w\rangle=\langle w|v\rangle^*$$

23.1 Ortogonalidad y Método de ortogonalización Gram-Schmidt

Las propiedades de ortogonalidad son análogas al caso euclídeo. Dos vectores v,w de un espacio unitario son **ortogonales** si $(v,w)=0$.
Al igual que en el caso euclídeo, dado un conjunto de m vectores v_i L.I., es posible construir con el método de Gram-Schmidt un conjunto ortogonal de vectores que genera el mismo espacio que los v_i, dados por:

$$w_1=v_1\,,\;\;\;w_i=v_i-\sum_{j=1}^{i-1}P_{w_j}(v_i)\,,\;\;i=2,\ldots,m$$

donde

$$P_{w_j}(v_i)=\frac{(w_j,v_i)}{||w_j||^2}w_j$$

es la *proyección ortogonal* de v_i sobre w_j. Notemos que en el caso complejo es necesario ser cuidadoso con el orden en el producto escalar, ya que $(w_j,v_i)\neq(v_i,w_j)=(w_j,v_i)^*$. Es fácil verificar que de esta forma,

$(w_i, w_j) = 0$ si $i \neq j$, siendo los w_i no nulos si los vectores originales son L.I.
Dada una base arbitraria de V, es pues siempre posible por este método construir una base ortogonal de V, que puede convertirse en *ortonormal* normalizando los vectores resultantes.
Notemos que el cuadrado de la norma de los w_i está dado, para $i > 1$, por

$$||w_i||^2 = (w_i, w_i) = (v_i, w_i) = ||v_i||^2 - \sum_{j=1}^{i-1} \frac{|(w_j, v_i)|^2}{||w_j||^2} \leq ||v_i||^2$$

Notemos también que la matriz que representa al proyector sobre w_i en la base canónica es

$$[P_{w_i}]_e = \frac{[w_i]_e [w_i]_e^\dagger}{||w_i||^2}$$

Ejemplo 1 : Consideremos los vectores $v_1 = (1+i, i, 0)$, $v_2 = (i, 1+i, 1)$. Tenemos

$$w_1 = v_1 = (1+i, i, 0), \quad w_2 = v_2 - \frac{(w_1, v_2)}{||w_1||^2} w_1 = (i, 1+i, 1) - \frac{2}{3}(1+i, i, 0) = (-2+i, 3+i, 3)/3$$

que verifican $(w_1, w_2) = 0$.

Ejemplo 2: Las funciones $f_k(x) = e^{ikx}$, con k *entero*, son ortogonales con el producto interno $(f, g) = \int_{-\pi}^{\pi} f^*(x) g(x) dx$:

$$(f_{k'}, f_k) = \int_{-\pi}^{\pi} e^{-ik'x} e^{ikx} dx = \int_{-\pi}^{\pi} e^{ix(k-k')} dx = \begin{cases} 2\pi & k = k' \\ \frac{e^{ix(k-k')}}{i(k-k')}|_{-\pi}^{\pi} = 0 & k \neq k' \end{cases}$$

Ejemplo 3 (Transformada de Fourier discreta): Sea $V = \mathbb{C}^n$ y sea $e = (e_1, \ldots, e_n)$ una base canónica ($(e_i, e_j) = \delta_{ij}$). Los n vectores

$$\tilde{e}_k = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} e^{i2\pi kj/n} e_j$$

forman también una base ortonormal: $(\tilde{e}_k, \tilde{e}_l) = \delta_{kl}$.
En efecto, utilizando que $(e_i, e_j) = \delta_{ij}$ obtenemos, para $k, l = 1, \ldots, n$,

$$(\tilde{e}_k, \tilde{e}_l) = \frac{1}{n} \sum_{j=1}^{n} e^{i2\pi j(l-k)/n} = \begin{cases} 1 & k = l \\ \frac{1}{n} \frac{1 - e^{i2\pi(l-k)}}{1 - e^{i2\pi(l-k)/n}} = 0 & k \neq l \end{cases}$$

Ejemplo 4: Obtener una base ortonormal de $\mathbb{C}^{2\times 2}$ (con escalares complejos) para el producto escalar $(A, B) = \mathrm{Tr}\, A^\dagger B$, partiendo de $v_1 = I_2 = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$.
Consideremos las matrices $v_1 = I_2$, $v_2 = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right)$, $v_3 = \frac{1}{2}\left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$, $v_4 = \left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)$, que forman una base no ortogonal de $\mathbb{C}^{2\times 2}$. Obtenemos, $w_1 = v_1 = I_2$,

$$w_2 = v_2 - \tfrac{1}{2}(w_1, v_2) w_1 = v_2 - \tfrac{1}{2} w_1 = \tfrac{1}{2}\left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right), \quad w_3 = v_3 - \tfrac{1}{2}(w_1, v_3) w_1 - 2(w_2, v_3) w_2 = v_3 = \tfrac{1}{2}\left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right),$$

$$w_4 = v_4 - \tfrac{1}{2}(w_1, v_4) w_1 - 2(w_2, v_3) w_2 - 2(w_3, v_4) w_3 = v_4 - w_3 = \tfrac{1}{2}\left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$$

Las **matrices de Pauli** se definen precisamente como

$$\sigma_0 = I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \sigma_x = 2w_3 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = -2iw_4 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_z = 2w_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

y forman una base ortogonal y hermítica de $\mathbb{C}^{2\times 2}$: $\sigma_\mu^\dagger = \sigma_\mu$, $(\sigma_\mu, \sigma_\nu) = \mathrm{Tr}\, \sigma_\mu \sigma_\nu = 2\delta_{\mu\nu}$ para $\mu, \nu = 0, x, y, z$.

Considerando ahora $\mathbb{C}^{2\times 2}$ sobre escalares reales, estas 4 matrices forman también una base del subespacio de matrices hermíticas de 2×2. Las matrices que representan las componentes del espín $\mathbf{s} = (s_x, s_y, s_z)$ en la base estándar de autoestados de s_z son precisamente

$$s_\mu = \frac{1}{2} \hbar \sigma_\mu, \quad \mu = x, y, z$$

23.2 Expansión en una base ortonormal

Si $e = (e_1, \ldots, e_n)$ es una base ortonormal de V ($(e_i, e_j) = \delta_{ij}$) y

$$v = \sum_{i=1}^{n} x_i e_i$$

entonces $(e_i, v) = \sum_{j=1}^{n} x_j (e_i, e_j) = x_j$. Por lo tanto,

$$x_i = (e_i, v)$$

Se cumple entonces

$$v = \sum_{i=1}^{n} P_{e_i}(v)$$

Se verifica también, por la ortogonalidad de los e_i, la generalización del teorema de Pitágoras,

$$||v|| = (v, v) = \sum_{i=1}^{n} ||x_i e_i||^2 = \sum_{i=1}^{n} |x_i|^2$$

Notemos que en la notación de mecánica cuántica, $e_i \to |i\rangle$ y $|v\rangle = \sum_i \alpha_i |i\rangle$, con $\alpha_i = \langle i|v\rangle$.

23.3 Proyectores ortogonales y matriz de Gram

Dado un subespacio $S \subset V$, es posible construir el complemento ortogonal $S_\perp = \{v \in V | (w, v) = 0 \; \forall \; w \in S\}$, cumpliéndose que $V = S \oplus S_\perp$ y por lo tanto, dim $S+$ dim $S_\perp = n$
Si $v \in V$, podemos escribir

$$v = v_s + (v - v_s)$$

con $v_s \in S$ y $v - v_s \in S_\perp$. Si $(w_1, \ldots, w_m)$ es una base *ortogonal* de S, escribiendo $v_s = \sum_{i=1}^{m} \alpha_i w_i$, la condición $(w_i, v - v_s) = 0$ para $i = 1, \ldots, m$ implica $\alpha_i = (w_i, v)/||w_i||^2$ y por lo tanto

$$v_s = \sum_{i=1}^{m} P_{w_i}(v) = P_S(v)\,, \qquad P_S = \sum_{i=1}^{m} P_{w_i}$$

El vector v_s es el vector de S con distancia mínima a v: Si $w_s \in S$,

$$||v - w_s||^2 = ||(v - v_s) + (v_s - w_s)||^2 = ||v - v_s||^2 + ||w_s - v_s||^2 \geq ||v - v_s||^2$$

En general, para una base arbitraria $(w_i, \ldots, w_m)$ de S no necesariamente ortogonal,

$$[P_S]_e = R(R^\dagger R)^{-1} R^\dagger$$

donde $R = ([w_1]_e, \ldots, [w_m]_e)$ es la matriz de $n \times m$ donde cada columna son las coordenadas de los m vectores w_i de la base de S en una base canónica de V ($(e_i, e_j) = \delta_{ij}$). Las fórmulas del caso euclídeo se generalizan pues directamente al caso unitario reemplazando t (traspuesta) por $\dagger$ (traspuesto conjugado).
Recordemos que

$$P_S + P_{S_\perp} = I$$

Notemos también que la matriz de Gram

$$G = R^\dagger R$$

de $m \times m$, con $G_{ij} = (w_i, w_j) = G^*_{ji}$, es ahora una matriz *hermítica*, que posee las mismas propiedades anteriores: $|G| \neq 0$ sii los m vectores w_i son LI, los autovalores λ_i de G son reales y no negativos, los autovectores X_i ($GX_i = \lambda_i X_i$) correspondientes a autovalores no nulos determinan vectores u_i de componentes $[u_i]_e = RX_i$, que son ortogonales ($(u_i, u_j) = 0$ si $\lambda_i \neq \lambda_j$), y aquellos correspondientes a autovalores nulos dan las combinaciones lineales nulas de los w_i (Demostraciones totalmente similares al caso euclídeo).

Ejemplo: Proyectar el vector $v = (1, i, 1+i) \in \mathbb{C}^3$ sobre el espacio generado por los vectores $v_1 = (1+i, i, 0)$, $v_2 = (i, 1+i, 1)$. Aplicando la representación general, tenemos

$$R = \begin{pmatrix} 1+i & i \\ i & 1+i \\ 0 & 1 \end{pmatrix}$$

con $R^\dagger R = \binom{3\ 2}{2\ 4}$, $(R^\dagger R)^{-1} = \binom{4\ -2}{-2\ 3}/8$ y

$$[P_S]_e = R(R^\dagger R)^{-1}R^\dagger = \begin{pmatrix} 7 & 1-i & -2+i \\ 1+i & 6 & 3+i \\ -2-i & 3-i & 3 \end{pmatrix}/8$$

Podemos arribar a este mismo resultado considerando también la base ortogonal de S obtenida previamente al ortogonalizar v_1 y v_2 por Gram-Schmidt, dada por $w_1 = v_1$, $w_2 = (-2+i, 3+i, 3)/3$:

$$[P_S]_e = [P_{w_1}]_e + [P_{w_2}]_e = \frac{[w_1]_e[w_1]_e^\dagger}{||w_1||^2} + \frac{[w_2]_e[w_2]_e^\dagger}{||w_2||^2}$$

$$= \frac{1}{3}\begin{pmatrix} 1+i \\ i \\ 0 \end{pmatrix}(1-i, -i, 0) + \frac{1}{24}\begin{pmatrix} -2+i \\ 3+i \\ 3 \end{pmatrix}(-2-i, 3-i, 3) = \begin{pmatrix} 7 & 1-i & -2+i \\ 1+i & 6 & 3+i \\ -2-i & 3-i & 3 \end{pmatrix}/8$$

Se obtiene finalmente

$$[P_S(v)]_e = [P_S]_e[v]_e = \begin{pmatrix} 5 \\ 3+11i \\ 2+5i \end{pmatrix}/8$$

La distancia mínima al plano es $||v - v_s|| = 3/\sqrt{8}$.

23.4 Operadores adjuntos y autoadjuntos en espacios unitarios

Sea $F : V \to V$ un operador lineal. El operador adjunto $F^\dagger$ se define por la relación

$$(v, F(w)) = (F^\dagger(v), w)$$

$\forall\ v, w \in V$. Considerando una base canónica e de V ($(e_i, e_j) = \delta_{ij}$), y teniendo en cuenta que $[F(v)]_e = [F]_e[v]_e$, y $(v, w) = [v]_e^\dagger[w]_e$, se obtiene $(v, F(w)) = [v]_e^\dagger[F]_e[w]_e$, $(F^\dagger(v), w) = [v]_e^\dagger[F^\dagger]_e^\dagger[w]_e$ y por lo tanto

$$[F^\dagger]_e = [F]_e^\dagger$$

La matriz que representa al operador adjunto de F en una base canónica es pues la traspuesta conjugada de la que representa a F en dicha base. Notemos que:
1) si $G = \alpha F$, con $\alpha \in \mathbb{C} \Rightarrow G^\dagger = \alpha^* F^\dagger$ (pues $(\alpha^* F^\dagger(v), w) = \alpha(F^\dagger(v), w) = \alpha(v, F(w)) = (v, \alpha F(w))$)
2) $(F^\dagger)^\dagger = F$ (pues $((F^\dagger)^\dagger(v), w) = (v, F^\dagger(w)) = (F(v), w)\ \forall\ v, w$)
3) $(FG)^\dagger = G^\dagger F^\dagger$ (pues $(v, FG(w)) = (F^\dagger(v), G(w)) = (G^\dagger F^\dagger(v), w)$).

Un operador F es *autoadjunto* si $F^\dagger = F$. En tal caso la matriz que lo representa en una base canónica es *hermítica*:

$$[F]_e^\dagger = [F]_e$$

Una propiedad importante de operadores adjuntos es que **si S es un subespacio invariante por $F \Rightarrow S_\perp$ es invariante por $F^\dagger$.**
Demostración: si $F(v) \in S\ \forall\ v \in S$, y $w \in S_\perp \Rightarrow (w, F(v)) = 0\ \forall\ w \in S_\perp$ y $v \in S$. Por lo tanto,

$$(F^\dagger(w), v) = (w, F(v)) = 0$$

$\forall\ w \in S_\perp$ y $v \in S$, de modo que $F^\dagger(w) \in S_\perp$
En particular, si F es *autoadjunto* y S es invariante por $F \Rightarrow S_\perp$ *es también invariante por* F.

Comentemos finalmente que en una base B general, donde $(v,w) = [v]_B^\dagger A[w]_B$ con A es una matriz hermítica definida positiva ($A^\dagger = A$, $X^\dagger AX > 0 \;\forall\; X \neq 0$, $X \in \mathbb{C}^{n\times 1}$) la condición $(v, F(w)) = (F^\dagger(v), w)\;\forall\; v, w \in V$ implica $[F^\dagger]_B^\dagger A = A[F]_B$, y por lo tanto,

$$[F^\dagger]_B = A^{-1}[F]_B^\dagger A$$

La matriz que representa el operador adjunto $F^\dagger$ en una base arbitraria es pues semejante (pero no necesariamente igual) a $[F]_B^\dagger$.

23.5 Operadores Unitarios

Un operador lineal $U : V \to V$ que conserva el producto interno en un espacio unitario se denomina unitario:

$$(U(v), U(w)) = (v, w)$$

$\forall\; v, w \in V$. Como $(U(v), U(w)) = (U^\dagger U(v), w) \Rightarrow U^\dagger U = I$ (identidad), por lo que en una base canónica tenemos

$$[U]_e^\dagger [U]_e = I_n$$

y por lo tanto $[U]_e[U]_e^\dagger = I_n$. Las matrices que representan a un operador unitario en una base canónica se denominan unitarias y satisfacen $[U]_e^{-1} = [U]_e^\dagger$, lo que implica filas y columnas ortonormales:

$$\sum_{j=1}^{n} S_{ji}^* S_{jk} = \delta_{ik} \qquad \sum_{j=1}^{n} S_{ij} S_{kj}^* = \delta_{ik}$$

donde aquí $S_{ij} = ([U]_e)_{ij}$. El determinante de un operador unitario tiene módulo 1:

$$1 = \text{Det}[U^\dagger U] = \text{Det}[U]^*\text{Det}[U] = |\text{Det}[U]|^2$$

por lo que

$$|\text{Det}[U]| = 1$$

Podemos entonces escribir $\text{Det}[U] = e^{i\phi}$, con ϕ real.

Debe remarcarse que los operadores unitarios transforman bases ortonormales en bases ortonormales: si $e_i' = U(e_i) = \sum_{j=1}^{m} S_{ji} e_j$, $i = 1, \ldots, n \Rightarrow$

$$(e_i', e_j') = (U(e_i), U(e_j)) = (e_i, e_j) = \delta_{ij}$$

Análogamente, cualquier par de bases ortonormales e, e' de V están relacionadas por una transformación unitaria, es decir, por una matriz de cambio de base S que satisface $S^\dagger S = SS^\dagger = I_n$, como es fácil verificar: Si $e_i' = \sum_j S_{ji} e_j$ y $(e_i', e_j') = (e_i, e_j) = \delta_{ij}$ entonces

$$(e_i', e_j') = \sum_{k,l} (S_{ki} e_i, S_{lj} e_l) = \sum_{k,l} S_{ki}^* S_{lj} (e_k, el) = \sum_k S_{ik}^\dagger S_{kj} = (S^\dagger S)_{ij} = \delta_{ij}$$

Remarquemos también que el *producto* (pero no la suma) de operadores unitarios es *unitario*: Si U, W son unitarios $\Rightarrow (UW)^{-1} = W^{-1}U^{-1} = W^\dagger U^\dagger = (UW)^\dagger$, por lo que UW es unitario. Está propiedad es también obvia a partir de la definición.

24. Autovalores y Autovectores de operadores autoadjuntos

1) Si $F : V \to V$ es un operador lineal autoadjunto $\Rightarrow$ sus autovalores son todos reales y los autovectores correspondientes a autovalores distintos son ortogonales.

Demostración: Si $F(v) = \lambda v$,

$$(v, F(v)) = (v, \lambda v) = \lambda(v, v)$$

pero por ser F autoadjunto,

$$(v, F(v)) = (F(v), v) = (\lambda v, v) = \lambda^*(v, v)$$

por lo que

$$(\lambda - \lambda^*)(v, v) = 0$$

lo que implica, si $v \neq 0$, $\lambda - \lambda^* = 0$, es decir, λ *real*. Todos los autovalores de F serán pues *reales*.
Además, si $F(v) = \lambda v$ y $F(v') = \lambda' v'$, entonces

$$(v', F(v)) = \lambda(v', v) = (F(v'), v) = \lambda'(v', v)$$

por lo que

$$(v', v)(\lambda - \lambda') = 0$$

lo que implica

$$(v', v) = 0 \text{ si } \lambda \neq \lambda'$$

2) Si $F : V \to V$ es un operador lineal autoadjunto en un espacio V de dimensión finita, existe siempre una base *ortonormal* de V formada por autovectores de F: $\exists\ e' = (e'_1, \ldots, e'_n)$, tal que

$$F(e'_i) = \lambda_i e'_i, \;\; i = 1, \ldots, n, \;\; (e'_i, e'_j) = \delta_{ij}$$

Es decir, F *es siempre diagonalizable* y además lo es en una base ortonormal, la cual estará relacionada con la base canónica original por una transformación unitaria U:

$$[F]_{e'} = S^\dagger [F]_e S = \begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ & & \ldots & \\ 0 & 0 & \ldots & \lambda_n \end{pmatrix}, \quad S^\dagger S = S S^\dagger = I$$

con $S = [U]_e$ y $e'_i = U(e_i)$.
Demostración: Por inducción sobre n. Para $n = 1$ todo F es trivialmente diagonal en cualquier base. Considerando ahora $n > 1$, si los n autovalores de F son todos distintos, entonces esta propiedad es inmediata, ya que por 1) existirán n autovectores ortogonales entre si, que puede ser convertidos en ortonormales luego de normalización ($e'_i \to e'_i/||e'_i||$).
En general, supongamos que e'_1 es un autovector normalizado de F ($F(e'_1) = \lambda_1 e'_1$, $((e'_1, e'_1) = 1)$ y sea S_1 el subespacio de V generado por e'_1. En tal caso S_1 es invariante por F y por lo tanto, el complemento ortogonal $S_{1\perp}$, de dimensión $n-1$, **será también invariante por** $F^\dagger = F$. F restringido a $S_{1\perp}$ es obviamente también autoadjunto. Por lo tanto, por hipótesis inductiva, existe una base ortonormal de $S_{1\perp}$ en la que F es diagonal. F resulta así diagonal en la base ortonormal de V formada por e'_1 y la base anterior de $S_{1\perp}$. F será entonces diagonalizable $\forall\ n$ en una base ortonormal.

3) Si F y G son dos operadores autoadjuntos y $[F, G] = 0$ (o sea, $FG = GF$) $\Rightarrow$ *existe una base ortonormal común* e' en la que ambos operadores son *simultáneamente diagonales:*

$$F(e'_i) = \lambda_i^F e'_i\,, \;\; G(e'_i) = \lambda_i^G e'_i, \;\; i = 1, \ldots, n$$

Demostración: Como F es autoadjunto, existe una base ortonormal donde F es diagonal. Como $[G, F] = 0$ $\Rightarrow$ si $F(e'_i) = \lambda_i^F e'_i$, $FG(e'_i) = GF(e'_i) = \lambda_i^F G(e'_i)$, por lo que $G(e'_i) \in V_F(\lambda_i^F)$ (espacio propio). $V_F(\lambda_i^F)$ es pues también invariante por G. Pero G restringido a $V_F(\lambda_i^F)$ es asimismo autoadjunto, por lo que es siempre posible elegir una base ortonormal de $V_F(\lambda_i^F)$ en la que G será también diagonal, con autovalores λ_i^G. Los elementos de dicha base serán, por pertenecer a $V_F(\lambda_i^F)$, también autovectores de F. Repitiendo esto para todos los autovalores, vemos que existirá una base ortonormal de V en la que tanto F y G serán diagonales.

24.1 Operadores normales

Un operador lineal $A : V \to V$ se dice *normal* si $A^\dagger A = A A^\dagger$, es decir, si

$$[A, A^\dagger] = 0$$

donde $[A, B] = AB - BA$ denota el conmutador.
Así, los operadores autoadjuntos ($F^\dagger = F$) son obviamente normales, y también son normales los unitarios ($U^\dagger = U^{-1}$, con $UU^\dagger = U^\dagger U = I$). Otro caso de operador normal son las antiautadjuntos ($F^\dagger = -F$).

Teorema de diagonalización para operadores normales:
Si $A: V \to V$ es un operador normal, entonces existe una base *ortonormal* e' en la cual $[A]_{e'}$ es *diagonal*:

$$[A]_{e'} = S^\dagger [A]_e S = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ & & \dots & \\ 0 & 0 & \dots & \lambda_n \end{pmatrix}, \quad S^\dagger S = S S^\dagger = I_n$$

Además si $A: V \to V$ es diagonal en una base ortonormal $\Rightarrow$ es normal.

Demostración: Hemos ya demostrado que para todo operador autoadjunto existe una base ortonormal donde es diagonal. La extensión para todo operador normal se basa en la descomposición

$$A = A^r + iA^i, \quad A^r = \frac{A + A^\dagger}{2}, \quad A^i = \frac{A - A^\dagger}{2i}$$

válida para *cualquier* operador A, donde A^r y A^i son claramente operadores *autoadjuntos*: $(A^r)^\dagger = A^r$, $(A^i)^\dagger = A^i$. Esta descomposición del operador es similar a la de un número complejo $z = x + iy$ en parte real x e imaginaria iy (caso particular $n = 1$).
Si A es *normal* $\Rightarrow$

$$[A^r, A^i] = \frac{1}{4i}[A + A^\dagger, A - A^\dagger] = 0$$

y por lo tanto, *existe una base ortonormal común e' donde A^r y A^i son simultáneamente diagonales.* Los autovalores de A serán entonces de la forma

$$\lambda_j = \lambda_j^r + i\lambda_j^i$$

con λ_j^r y λ_j^i *reales* y autovalores de A^r y A^i respect., por lo que λ_j será en general complejo.
Si A es autoadjunto ($A^\dagger = A$) $\Rightarrow A_i = 0$ y por lo tanto $\lambda_j^i = 0$. Los autovalores de A son entonces todos reales, como ya habíamos demostrado.
Si A es antiautoadjunto ($A^\dagger = -A$) $\Rightarrow A_r = 0$ y por lo tanto $\lambda_j^r = 0$. Los autovalores de A son entonces todos *imaginarios puros*.
Finalmente, si A es unitario, $[A]_{e'}^\dagger [A]_{e'} = I_n$, lo que implica $\lambda_j \lambda_j^* = |\lambda_j|^2 = 1$, es decir $|\lambda_j| = 1$. Esto implica

$$\lambda_j = e^{i\phi_j} = \cos\phi_j + i \sin\phi_j$$

con $\lambda_j^r = \cos\phi_j$, $\lambda_j^i = \sin\phi_j$.

Por otro lado, si A es diagonal en una base e' ortonormal $\Rightarrow A^\dagger$ es también diagonal en dicha base, con

$$[A^\dagger]_{e'} = [A]_{e'}^\dagger = \begin{pmatrix} \lambda_1^* & 0 & \dots & 0 \\ 0 & \lambda_2^* & \dots & 0 \\ & & \dots & \\ 0 & 0 & \dots & \lambda_n^* \end{pmatrix}$$

Por lo tanto

$$[AA^\dagger - A^\dagger A]_{e'} = [A]_{e'}[A^\dagger]_{e'} - [A^\dagger]_{e'}[A]_{e'} = 0$$

lo que implica $AA^\dagger - A^\dagger A = 0$. A es entonces normal.

En resumen, el teorema implica que en un espacio unitario, *un operador tiene representación diagonal en una base ortonormal si y sólo si es un operador normal.* En términos matriciales, si A es una matriz de $n \times n$, entonces *existe una matriz unitaria S tal que $A' = S^\dagger A S$ es diagonal si y sólo si A es normal* ($[A^\dagger, A] = 0$). Esto comprende en particular las matrices hermíticas ($A^\dagger = A$), antihermíticas ($A^\dagger = -A$) y unitarias ($A^\dagger = A^{-1}$). Destaquemos también que todo $v \in V$ puede expandirse en la base e' de autovectores de un operador normal A,

$$v = \sum_{i=1}^{n} \alpha_i e'_i = \sum_{i=1}^{n} P_{e'_i}(v)$$

donde $\alpha_i = (e'_i, v)$ y $P_{e'_i}(v) = (e'_i, v)e'_i = \alpha'_i e'_i$. Por lo tanto

$$A(v) = \sum_{i=1}^{n} A(\alpha_i e'_i) = \sum_{i=1}^{n} \alpha_i \lambda_i e'_i = \sum_{i=1}^{n} \lambda_i P_{e'_i}(v)$$

Como v es arbitrario, esto implica

$$A = \sum_{i=1}^{n} \lambda_i P_{e'_i}$$

Un operador normal puede pues expresarse como *combinación lineal de proyectores ortogonales sobre sus espacios propios.*

De lo anterior se desprende además que todo operador unitario U puede escribirse en la forma

$$U = \exp[iF]$$

con F autoadjunto: Como los autovalores de U son de la forma $e^{i\phi_j}$, podemos definir F como el operador autoadjunto que es también diagonal en la base ortonormal e' en que U es diagonal y que tiene autovalores reales ϕ_j. En tal caso, $[U]_{e'} = \exp[i[F]_{e'}] = [\exp[iF]]_{e'}$, lo que implica $[U]_e = [\exp(iF)]_e$ en cualquier base. Esto conduce a $U = \exp[iF]$.

Ejercicio: Utilizando la representación diagonal, mostrar que si $F : V \to V$ es autoadjunto, entonces $\forall$ $v \in V$, con $v \neq 0$, se tiene

$$\lambda_m \leq \frac{(v, F(v))}{(v, v)} \leq \lambda_M$$

donde λ_m y λ_M denotan resp. el menor y mayor autovalor de F.

24.2 Isometrías en espacios euclideos

Hemos visto que los autovalores de un operador unitario U son necesariamente de la forma $\lambda = e^{i\phi} = \cos(\phi) + i\sin(\phi)$, con ϕ real. Mediante el "embedding" de un espacio euclídeo en un espacio unitario discutido en clase, esto permite demostrar que las isometrías U en espacios euclídeos sólo pueden ser rotaciones ($\text{Det}U = 1$) o rotaciones seguidas o precedidas de una reflexión ($\text{Det}U = -1$).
En efecto, si $S \equiv [U]_e$ es una matriz real que representa una isometría U en una base ortonormal de un espacio euclideo ($S^t = S^{-1}$), considerada en un espacio complejo representa una transformación unitaria ($S^\dagger = S^{-1}$). Dado que S es real, los autovalores vendrán de a pares conjugados con autovectores conjugados:

$$SX = \lambda X, \quad SX^* = \lambda^* X^*$$

Escribiendo $\lambda = \lambda_r + i\lambda_i$, $X = X_r + iX_i$, con $\lambda_r = \cos\phi$, $\lambda_i = \sin\phi$ y X_r, X_i reales, esto implica

$$SX_r = \lambda_r X_r - \lambda_i X_i\,, \quad SX_i = \lambda_r X_i + \lambda_i X_r$$

Si λ no es real ($\lambda_i \neq 0$) $\Rightarrow$ $X_i \neq 0$ (pues S es real) y la ortogonalidad de los autovectores para autovalores distintos (válido para cualquier matriz normal $[S^\dagger, S] = 0$) implica $(X^*)^\dagger X = 0$, o sea,

$$(X_r + iX_i)^t (X_r + iX_i) = X_r^t X_r - X_i^t X_i + 2iX_i^t X_r = 0$$

de donde $X_r^t X_r = X_i^t X_i$ y $X_i^t X_r = 0$. Por lo tanto, vemos que en el subespacio generado por X^*, X, existe una base *real* y *ortonormal con el producto escalar euclideo*, formada por (X_i, X_r), en la que el bloque correspondiente de $[U]_{e'}$ tiene la forma

$$S'_\phi = \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix}$$

que representa una rotación de ángulo ϕ ($\text{Det}[S'_\phi] = 1$). Y en el espacio euclideo completo, vemos entonces que existe una base ortonormal e' donde $S' \equiv [U]_{e'}$ tiene la forma

$$S' = \begin{pmatrix} {S'}_{\phi_1} & 0 & \dots & 0 & 0 \\ 0 & {S'}_{\phi_2} & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & \pm 1 & 0 \\ 0 & 0 & \dots & 0 & \pm 1 \end{pmatrix}$$

donde S'_{ϕ_i} son bloques de la forma anterior que representan rotaciones en subespacios de dimensión 2, y los elementos ± 1 representan los posibles autovalores reales. U representa pues rotaciones ($\text{Det}[S']=1$) o rotaciones compuestas con reflexiones ($\text{Det}[S']=-1$). Por ej., en $\mathbb{R}^3$, las posibilidades son un bloque A_ϕ seguido de $+1$ (rotación) o -1 (rotación compuesta con reflexión).

24.3 Elementos de matriz de un operador lineal en una base ortonormal

Recordemos que si $F: V \to V$ es un operador lineal $\Rightarrow$ la matriz $T = [F]_e$ ($\equiv [F]_e^e$) que lo representa en una base e de V queda definida por

$$F(e_i) = \sum_{j=1}^{n} T_{ji} e_j$$

En un espacio unitario y en una base ortonormal e, los elementos de matriz $T_{ji} = ([F]_e)_{ji}$ pueden entonces obtenerse, por ortonormalidad de los e_i, como

$$T_{ji} = (e_j, F(e_i))$$

De esta forma,

$$v = \sum_{i=1}^{n} \alpha_i e_i, \quad \alpha_i = (e_i, v)$$

y

$$F = \sum_{i,j=1}^{n} T_{ji} E_{ji}, \quad T_{ji} = (e_j, F(e_i))$$

con E_{ji} el operador lineal definido por

$$E_{ji}(v) = (e_i, v) e_j$$

ya que $F(e_i) = \sum_{j=1}^{n} T_{ji} e_j = \sum_{j,k=1}^{n} T_{jk}(e_k, e_i) e_j = (\sum_{j,k=1}^{n} T_{jk} E_{jk})(e_i)$.
Notemos que $([E_{ji}]_e)_{kl} = \delta_{kj}\delta_{il}$.

Notación de Mecánica cuántica:

$$T_{ji} = (e_j, F(e_i)) \to \langle j|F|i\rangle, \quad E_{ji} \to |j\rangle\langle i|$$

donde $|i\rangle \equiv e_i$ y $\langle i| \equiv f^i$ (vector asociado del espacio dual). Por lo tanto

$$F = \sum_{i,j} F_{ij} |i\rangle\langle j|, \quad F_{ij} = \langle i|F|j\rangle$$

Por ej., el proyector ortogonal sobre e_i se escribe como $P_{e_i} = |i\rangle\langle i|$ (ya que $(e_j, P_{e_i}(e_i)) = (e_j, e_i) = \delta_{ji}$) mientras que el operador identidad es $I = \sum_i |i\rangle\langle i|$. En general, para todo operador normal $F: V \to V$ existe entonces una base ortonormal $\{|i\rangle\}$ de V formada de autovectores de F en la que $\langle i|F|j\rangle = \delta_{ij}\lambda_i$ y

$$F = \sum_i \lambda_i |i\rangle\langle i|$$

25 Descomposición en valores singulares (DVS)

Sea $F: V \to W$ una transformación lineal arbitraria entre espacios unitarios V y W de dimensiónes n y m respectivamente, y e_V, e_W bases ortonormales de V y W. Entonces existen bases ortonormales e'_V, e'_W en las que F queda representado por una matriz *diagonal* de elementos no negativos. En otras palabras, dada una matriz A de $m \times n$, existen matrices unitarias U de $m \times m$ y V de $n \times n$ tales que

$$A = U A' V^\dagger$$

con $U^\dagger U = I_m$, $V^\dagger V = I_n$ y A' de $m \times n$ *diagonal* de elementos $A'_{kj} = \sigma_j \delta_{kj}$, con $\sigma_j \geq 0$. Aquí $A = [F]_{e_W}^{e_V}$, $A' = [F]_{e'_W}^{e'_V}$, $U = [I]_{e_W}^{e'_W}$ y $V = [I]_{e_V}^{e'_V}$, con $V^\dagger = [I]_{e'_V}^{e_V}$. Los σ_j no nulos se denominan *valores singulares* y son las *raíces* de los autovalores no nulos de la matriz hermítica $A^\dagger A$, de $n \times n$ (que posee autovalores no negativos). V es la correspondiente matriz de autovectores normalizados (tal que $V^\dagger (A^\dagger A) V$ es diagonal).

La demostración es similar al caso de matrices reales (espacios euclideos) y se deja como ejercicio. Recordemos que si k es el número de autovalores no nulos de $A^\dagger A$, las primeras k columnas de U son los vectores $u_i = Av_i/\sigma_i$, $i = 1, \ldots, k$, $\sigma_i \neq 0$, obteniéndose las restantes $m-k$ columnas ortonormales de U por el método de Gram-Schmidt complejo.

Ejercicios: Para A de $m \times n$ compleja general, demostrar (en forma similar al caso euclideo) que:
0) Las matrices $A^\dagger A$ y $AA^\dagger$ son ambas hermíticas.
1) Los autovalores de $A^\dagger A$ son todos no negativos.
2) El número de autovalores no nulos de $A^\dagger A$ es igual al rango de A.
3) Los autovalores no nulos de las matrices $A^\dagger A$ y $AA^\dagger$ son iguales.
4) $||A||_2 = \sigma_M$, siendo σ_M el máximo valor singular y $||A||_2 \equiv \text{Max}_{v\neq 0}||Av||/||v||$, con $||v|| = \sqrt{v^\dagger v}$.
5) Si $m = n$ y λ es autovalor de $A \Rightarrow |\lambda| \leq \sigma_M$.
6) Si $m = n$ y A es invertible $\Rightarrow n_c(A) = \sigma_M/\sigma_m$, donde σ_m es el mínimo valor singular de A y $n_c(A)$ es el número de condición.

25.1 Forma polar de un operador lineal

En el caso de $V = W$, la DVS permite obtener en forma inmediata la denominada *forma polar* de un operador: Si $F : V \to V$ es un operador lineal en un espacio unitario V entonces F puede escribirse como

$$F = WM = \tilde{M}W$$

donde W es un operador *unitario* y M, $\tilde{M}$ operadores autoadjuntos *positivos*.
Dem.: Utilizando la DVS para la representación $A = [F]_e$ de F en una base ortonormal e de V, se tiene

$$\begin{aligned} A &= UA'V^\dagger \\ &= (UV^\dagger)(VA'V^\dagger) = WM, \quad W = UV^\dagger, \;\; M = VA'V^\dagger = \sqrt{A^\dagger A} \\ &= (UA'U^\dagger)(UV^\dagger) = \tilde{M}W, \quad \tilde{M} = UA'U^\dagger = \sqrt{AA^\dagger} \end{aligned}$$

donde W es unitario ($W^\dagger = VU^\dagger = W^{-1}$) y $A^\dagger A = VA'^2V^\dagger$, $AA^\dagger = UA'^2U^\dagger$.
Ejercicio: Discutir la DVS y la descomposición polar de una matriz hermítica.

26 Desigualdad de Cauchy Schwarz y relaciones de incerteza

Consideremos dos operadores autoadjuntos F, G. El valor medio de un operador F en un estado normalizado $|\psi\rangle$ ($\langle\psi|\psi\rangle = 1$) es

$$\langle F\rangle_\psi = \langle\psi|F|\psi\rangle$$

(o sea $\langle F\rangle_\psi = (\psi, F(\psi))$ en notación de A.L.). Si F es autoadjunto, $\langle F\rangle_\psi$ es *real*, ya que $\langle\psi|F|\psi\rangle = \langle\psi|F^\dagger|\psi\rangle^* = \langle\psi|F|\psi\rangle^*$ (o sea, $(\psi, F(\psi)) = (F(\psi), \psi)^* = (\psi, F^\dagger(\psi))^* = (\psi, F(\psi))^*$).

La varianza de un operador F en el estado $|\psi\rangle$ se define como el valor medio del cuadrado de la diferencia entre F y $\langle F\rangle_\psi$ y es una medida de la dispersión alrededor de la media:

$$\Delta^2 F = \langle(F - \langle F\rangle_\psi I)^2\rangle_\psi = \langle\psi|(F - \langle F\rangle_\psi I)^2|\psi\rangle = \langle\psi|F^2|\psi\rangle - \langle\psi|F|\psi\rangle^2$$

La desviación estándar es la raíz de la varianza: $\Delta F = \sqrt{\Delta^2 F} = \sqrt{\langle(F - \langle F\rangle_\psi I)^2\rangle_\psi}$.

Definamos ahora

$$\tilde{F} = F - \langle F\rangle_\psi I,\ \ \tilde{G} = G - \langle G\rangle_\psi I$$

tal que $\Delta F = \sqrt{\langle\psi|\tilde{F}^2|\psi\rangle}$, $\Delta G = \sqrt{\langle\psi|\tilde{G}^2|\psi\rangle}$, y consideremos el producto escalar $(\tilde{F}(\psi), \tilde{G}(\psi)) = (\psi, \tilde{F}\tilde{G}(\psi))$, es decir $\langle\psi|\tilde{F}\tilde{G}|\psi\rangle$ en notación cuántica. La desigualdad de Cauchy-Schwarz implica $|(\tilde{F}(\psi), \tilde{G}(\psi))| \leq ||\tilde{F}(\psi)||\,||\tilde{G}(\psi)||$, con $||\tilde{F}(\psi)||^2 = (\tilde{F}(\psi), \tilde{F}(\psi)) = (\psi, \tilde{F}^2(\psi))$, o sea,

$$|\langle\psi|\tilde{F}\tilde{G}|\psi\rangle| \leq \sqrt{\langle\psi|\tilde{F}^2|\psi\rangle}\sqrt{\langle\psi|\tilde{G}^2|\psi\rangle} = (\Delta F)(\Delta G)$$

Por otro lado, si $[F, G] = FG - GF$ denota el conmutador de F y G, entonces

$$\langle\psi|[F,G]|\psi\rangle = \langle\psi|[\tilde{F},\tilde{G}]|\psi\rangle = \langle\psi|\tilde{F}\tilde{G}|\psi\rangle - \langle\psi|\tilde{G}\tilde{F}|\psi\rangle = 2i\,\mathrm{Im}[\langle|\psi|\tilde{F}\tilde{G}|\psi\rangle]$$

donde Im denota la parte imaginaria, ya que $\langle\psi|\tilde{G}\tilde{F}|\psi\rangle = \langle\psi|(\tilde{G}\tilde{F})^\dagger|\psi\rangle^* = \langle\psi|\tilde{F}\tilde{G}|\psi\rangle^*$. Por lo tanto

$$\tfrac{1}{2}|\langle\psi|[F,G]|\psi\rangle| \leq |\langle\psi|\tilde{F}\tilde{G}|\psi\rangle| \leq (\Delta F)(\Delta G)$$

es decir,

$$(\Delta F)(\Delta G) \geq \tfrac{1}{2}|\langle[F,G]\rangle_\psi|$$

Esta es la denominada relación de incerteza entre dos operadores: Si el conmutador es no nulo entonces el producto de sus "incertezas" $(\Delta F)(\Delta G)$ en un estado $|\psi\rangle$ no puede ser menor que el módulo del valor medio del conmutador en dicho estado.

Como ejemplo fundamental, consideremos el espacio L^2 de funciones $\psi(x)$ de $\mathbb{R} \to \mathbb{C}$ de norma finita ($||\psi||^2 = \int_{-\infty}^{\infty} |\psi(x)|^2 dx < \infty$) y que tienden a 0 para $x \to \pm\infty$, tal que el producto escalar

$$(\psi, \phi) = \int_{-\infty}^{\infty} \psi^*(x)\phi(x)dx$$

esté bien definido. Los operatores X y $P = -i\hbar\partial_x = -i\hbar\frac{\partial}{\partial x}$, donde $\hbar = h/(2\pi)$, con h la constante de Planck, son autoadjuntos en este espacio: $(\psi, X\phi) = \int_{-\infty}^{\infty} \psi^*(x)x\phi(x)dx = (X\psi, \phi)$, y

$$(\psi, P\phi) = -i\hbar\int_{-\infty}^{\infty}\psi^*(x)\phi'(x)dx = -i\hbar[\ \psi^*(x)\phi(x)\,|_{-\infty}^{\infty} - \int_{-\infty}^{\infty}\psi^{*\prime}(x)\phi(x)dx] = \int_{-\infty}^{\infty}[-i\hbar\psi'(x)]*\phi(x)dx = (P(\psi), \phi)$$

Dado que $[X,P]\psi(x) = -i\hbar(x\psi'(x) - (x\psi(x))') = i\hbar\psi(x)\ \forall\ \psi$, es decir, $[X,P] = i\hbar I$, obtenemos $|\langle[X,P]\rangle_\psi| = \hbar\ \forall\ \psi$ y el resultado anterior implica entonces

$$(\Delta P)(\Delta X) \geq \frac{\hbar}{2}$$

El operador P representa en Mecánica Cuántica el operador impulso de una partícula (en una dimensión). Por lo tanto, en cualquier estado cuántico el producto de las desviaciones estándar de X y P es no nulo y mayor que $\hbar/2$.

Algebra Lineal: Aplicaciones a la Física (Curso 2014)

27 Tensores (Resumen)

27. 1 Notación tensorial

Mediante la convención de Einstein para sumas, el cambio de base $e'_i = \sum_{j=1}^n S_{ji} e_j$, con $S = [I]_e^{e'}$ una matriz de $n \times n$ no singular, se escribe

$$e'_i = S_i^j e_j$$

donde $S_i^j e_j \equiv \sum_{j=1}^n S_i^j e_j$ y n es la dimensión del espacio. El índice superior en S denota fila y el inferior columna. En forma matricial, la relación anterior equivale pues a

$$(e'_1, \ldots, e'_n) = (e_1, \ldots, e_n) S$$

Por otro lado, la transformación $x'^i = \sum_{j=1}^n S^{-1}_{ij} x^j$ de las componentes de un vector $v = \sum_{i=1}^n x^i e_i = \sum_{i=1}^n x'^i e'_i$, se escribe en la forma

$$x'^i = R_j^i x^j, \quad R = S^{-1}$$

donde $R_j^i x^j \equiv \sum_{j=1}^n R_j^i x^j$. En forma matricial, la relación previa equivale pues a

$$\begin{pmatrix} x'^1 \\ \ldots \\ x'^n \end{pmatrix} = R \begin{pmatrix} x^1 \\ \ldots \\ x^n \end{pmatrix}$$

lo que está también de acuerdo con el supraíndice como índice de fila. Notemos que

$$R_j^i S_k^j = S_j^i R_k^j = \delta_k^i$$

que es la expresión tensorial de la relación matricial $RS = SR = I$. El vector v se escribe entonces como

$$v = x^i e_i = x'^i e'_i$$

Como verificación, reemplazando $x'^i = R_j^i x^j$, $e'_i = S_i^k e_k$, se tiene $x'^i e'_i = R_j^i S_i^k x^j e_k = \delta_j^k x^j e_k = x^j e_j$.

En general, n coordenadas a_i que se transforman como

$$a'_i = S_i^j a_j$$

se denominan *covariantes*, mientras que n cordenadas b^i que se transforman como

$$b'^i = R_j^i b^j$$

con $R_j^i S_k^j = \delta_k^i$ (o sea, $R = S^{-1}$) se denominan *contravariantes*. En tal caso, el producto

$$b'^i a'_i = b^i a_i$$

(donde la suma sobre i está implícita) permanece invariante frente a cambios de base.

Notemos finalmente que las relaciones inversas están dadas por

$$a_i = R_i^j a'_j, \quad b^i = S_j^i b'^j$$

Transformación de las derivadas parciales:
Dado el cambio de variables lineal $x'^i = R_j^i x^j$ y su relación inversa $x^j = S_i^j x'^i$, con $S = R^{-1}$, y R, S *independientes de las coordenadas*, tenemos

$$S_i^j = \frac{\partial x^j}{\partial x'^i}, \quad R_j^i = \frac{\partial x'^i}{\partial x^j}$$

En virtud de la regla de la cadena, se obtiene entonces

$$\frac{\partial}{\partial x'^i} = \sum_{j=1}^n \frac{\partial x^j}{\partial x'^i} \frac{\partial}{\partial x^j}$$

o sea, en notación covariante,

$$\partial'_i = S_i^j \partial_j$$

donde $\partial'_i \equiv \frac{\partial}{\partial x'^i}$, $\partial_j \equiv \frac{\partial}{\partial x^j}$. Las derivadas respecto de componentes contravariantes se transforman pues de manera covariante.

27.2 Transformación de vectores del dual

Dada una base $e = (e_1, \ldots, e_n)$ de V, los elementos de la base dual $e^* = (e^1, \ldots, e^n)$ del espacio dual V^* (el conjunto de formas lineales de V en K) quedan definidos por

$$(e^i, e_j) = \delta^i_j$$

(utilizamos la notación $e^i(v) = (e^i, v)$. Esto implica la ley de transformación contravariante

$$e'^i = R^i_j e^j$$

de forma que

$$(e'^i, e'_j) = R^i_k S^l_j (e^k, e_l) = R^i_k S^l_j \delta^k_l = R^i_l S^l_j = \delta^i_j$$

donde $e'_j = S^i_j e_i$. Un elemento arbitrario $h \in V^*$ puede entonces ser escrito como

$$h = a_i e^i = a'_i e'^i$$

donde

$$a'_i = S^j_i a_j$$

Notemos que si $v = x^i e_i$, $h = a_i e^i$,

$$a_i = (h, e_i), \quad x^i = (e^i, v)$$

Finalmente, mencionemos que si $(e_1, \ldots, e_n)$, $(e'^1, \ldots, e'^n)$ son bases arbitrarias de V y V^* respect., con

$$R^i_j = (e'^i, e_j)$$

una matriz no singular, la base dual de V asociada a la base e' de V^* está dada por

$$e'_i = S^j_i e_j$$

con $S = R^{-1}$, ya que $(e'^k, e'_i) = (e'^k, e_j)S^j_i = R^k_j S^j_i = \delta^k_i$. Análogamente, la base dual de V^* asociada a la base e de V está formada por

$$e^i = S^i_j e'^j$$

ya que $(e^i, e_k) = S^i_j (f'^j, e_k) = S^i_j R^j_k = \delta^i_k$.

Ejemplo: Sea $V = \mathbb{R}^2$ y sean (e'^1, e'^2) las formas lineales definidas por $(e'^1, v) = 3x + y$, $(e'^2, v) = 2x + y$, donde $v = (x, y) = xe_1 + ye_2$, con (e_1, e_2) la base canónica de $\mathbb{R}^2$. Hallar la base dual de V asociada a e'^1, e'^2.

Podemos escribir $e'^i = R^i_k e^k$, con $R = \left(\begin{smallmatrix} 3 & 1 \\ 2 & 1 \end{smallmatrix}\right)$ y (e^1, e^2) la base dual asociada a (e_1, e_2) ($(e^1, v) = x$, $(e^2, v) = y$). Es claro que (e'^1, e'^2) es base de V^* pues $|R| = 1 \neq 0$. La base dual asociada de V está entonces dada por $e'_i = S^j_i e_j$, con $S = R^{-1} = \left(\begin{smallmatrix} 1 & -1 \\ -2 & 3 \end{smallmatrix}\right)$:

$$e'_1 = e_1 - 2e_2, \; e'_2 = -e_1 + 3e_2$$

verificándose que $(e'^i, e'_j) = \delta^i_j$.

27.3 Tensor métrico

Dado un espacio euclideo V de dimensión finita, con el producto escalar denotado por (v, w), y dada una base arbitraria $e = (e_1, \ldots, e_n)$ de V, el tensor métrico se define como

$$g_{ij} = (e_i, e_j)$$

Es una matriz simétrica ($g_{ij} = g_{ji}$) no singular ($|g| \neq 0$). En tal caso, la norma al cuadrado de un vector $v = x^i e_i$ (es decir, la distancia al cuadrado del extremo del vector al origen) está dada por

$$||v||^2 = (x^i e_i, x^j e_j) = x^i (e_i, e_j) x^j = x^i g_{ij} x^j$$

Podemos escribir lo anterior también en la forma

$$||v||^2 = x^i x_i, \quad x_i \equiv g_{ij} x^j$$

Frente a un cambio de base, el tensor métrico se transforma como

$$g'_{ij} = (e'_i, e'_j) = S_i^k S_j^l (e_k, e_l) = S_i^k S_j^l g_{kl}$$

que corresponde a un tensor de rango $(2,0)$ (dos veces covariante), como se verá en breve.
Las componentes x_i se transforman pues en forma covariante:

$$x'_i = g'_{ij} x'^j = S_i^k S_j^l R_m^j g_{kl} x^m = S_i^k g_{kl} x^l = S_i^k x_k$$

En espacios euclideos V de dimensión finita, podemos identificar con cada elemento h del dual V^* uno y sólo un vector $w_h \in V$ tal que

$$(h, v) = (w_h, v)$$

$\forall\, v \in V$, donde el segundo paréntesis denota producto escalar: Si $h = a_i e^i$ y $w_h = a^i e_i$, con $(e^i, e_j) = \delta^i_j$,

$$(h, e_j) = a_j = (w_h, e_j) = a^i (e_i, e_j) = a^i g_{ij}$$

de modo que $a^i g_{ij} = a_j$. Por lo tanto,

$$a^i = g^{ji} a_j$$

donde g^{ij} denota los elementos de la *matriz inversa* de la matriz de elementos g_{ij}:

$$g^{ik} g_{kj} = \delta^i_j$$

En lo sucesivo denotaremos a w_h directamente como h. Por consiguiente, podemos escribir los elementos de la base dual como combinación lineal de los e_i. En notación tensorial,

$$e^i = g^{ik} e_k$$

con

$$(e^i, e_j) = g^{ik}(e_k, e_j) = g^{ik} g_{kj} = \delta^i_j$$

Notemos también que

$$(e^i, e^j) = g^{jk}(e^i, e_k) = g^{ji}$$

por lo que g^{ji} es el tensor métrico en la base dual. Un vector v puede pues escribirse en las formas

$$v = x^i e_i = x_i e^i$$

donde $x_i = g_{ij} x^j$, $e^i = g^{ik} e_k$, ya que $x_i e^i = g^{ik} g_{ij} x^j e_k = \delta^k_j x^j e_k = x^j e_j$. Para el producto escalar de dos vectores $v = x^i e_i$, $w = y^j e_j$ se tienen pues las expresiones

$$(v, w) = x^i g_{ij} y^j = x^i y_i = x_i y^i = x_i g^{ij} y_j$$

27.4 Tensores

Un tensor general de p índices covariantes y q indices contravariantes (que denotaremos aquí como tensor $\binom{q}{p}$) en un espacio de dimensión n, es un conjunto de n^{p+q} números $T^{j_1 \ldots j_q}_{i_1 \ldots i_p}$ dependientes de una base ordenada $B = (e_1, \ldots, e_n)$ de un espacio vectorial V, que se transforman frente a cambios de base $e'_i = S_i^j e_j$ en la forma

$$T'^{j'_1 \ldots j'_q}_{i'_1 \ldots i'_p} = S^{i_1}_{i'_1} \ldots S^{i_p}_{i'_p} R^{j'_1}_{j_1} \ldots R^{j'_q}_{j_q} T^{j_1 \ldots j_q}_{i_1 \ldots i_p}$$

con $R = S^{-1}$. Por ejemplo, para un tensor $\binom{1}{1}$, $T'^l_k = R^l_j S^i_k T^j_i$, que involucra una suma sobre i y j.

Una posible realización de un tensor $\binom{q}{p}$ es una forma *multilineal* $T : V^p \times (V^*)^q \to K$ de p vectores de V y q vectores del espacio dual V^* (una función es multilineal si es lineal en cada uno de sus argumentos:

$T(\alpha_1 v_1 + \alpha'_1 v'_1, v_2, \ldots, v_p, w^1, \ldots, w^q) = \alpha_1 T(v_1, v_2, \ldots, v_p, w^1, \ldots, w^q) + \alpha'_1 T(v'_1, v_2, \ldots, v_p, w^1, \ldots, w^q)$, y similar para los restantes argumentos). En tal caso, si $v_i = x_i^j e_j$ y $w^i = a_j^i e^j$,

$$T(v_1, \ldots, v_p, w^1, \ldots, w^q) = x_1^{i_1} \ldots x_p^{i_p} a_{j_1}^1 \ldots a_{j_q}^q T(e_{i_1}, \ldots, e_{i_p}, e^{j_1}, \ldots, e^{j_q})$$

Si los f^i son los vectores de la base dual ($(f^i, e_j) = \delta_j^i$), los elementos

$$T_{i_1 \ldots i_p}^{j_1 \ldots j_q} \equiv T(e_{i_1}, \ldots, e_{i_p}, e^{j_1}, \ldots, e^{j_q})$$

se transforman como un tensor $\binom{q}{p}$ frente a cambios de base: Si $e'_i = S_i^j e_j$, entonces $e'^i = R_j^i e^j$ y

$$\begin{aligned} T'^{j'_1 \ldots j'_q}_{i'_1 \ldots i'_p} &= T(e'_{i'_1}, \ldots, e'_{i'_p}, e'^{j'_1}, \ldots, e'^{j'_q}) = T(S_{i'_1}^{i_1} e_{i_1}, \ldots, S_{i'_p}^{i_p} e_{i_p}, R_{j_1}^{j'_1} e^{j_1}, \ldots, R_{j_q}^{j'_q} e^{j_q}) \\ &= S_{i'_1}^{i_1} \ldots S_{i'_p}^{i_p} R_{j_1}^{j'_1} \ldots R_{j_q}^{j'_q} T_{i_1 \ldots i_p}^{j_1 \ldots j_q} \end{aligned}$$

Otra posibilidad es considerar a $T_{i_1 \ldots i_p}^{j_1 \ldots j_q}$ como las coordenadas de un vector T perteneciente al **producto tensorial** de espacios $\underbrace{V \otimes \ldots \otimes V}_{\text{q veces}} \otimes \underbrace{V^* \otimes \ldots \otimes V^*}_{\text{p veces}}$ en una base $B = \{e_{j_1} \otimes \ldots \otimes e_{j_q} \otimes e^{i_1} \otimes \ldots \otimes e^{i_p}\}$, donde nuevamente $(e^i, e_j) = \delta_j^i$:

$$T = T_{i_1 \ldots i_p}^{j_1 \ldots j_q} e_{j_1} \otimes \ldots \otimes e_{j_q} \otimes e^{i_1} \otimes \ldots \otimes e^{i_p}$$

Si $e_i = R_i^j e'_j$ y $e^i = S_j^i e'^j$ (tal que $e'_i = S_i^j e_j$, $e'^i = R_j^i e^j$, con $R = S^{-1}$), tenemos

$$\begin{aligned} T &= T_{i_1 \ldots i_p}^{j_1 \ldots j_q} R_{j_1}^{j'_1} \ldots R_{j_q}^{j'_q} S_{i'_1}^{i_1} \ldots S_{i'_p}^{i_p} e'_{j'_1} \otimes \ldots \otimes e'_{j'_q} \otimes e'^{i'_1} \otimes \ldots \otimes e'^{i'_p} \\ &= T'^{j'_1 \ldots j'_q}_{i'_1 \ldots i'_p} e'_{j'_1} \otimes \ldots \otimes e'_{j'_q} \otimes e'^{i'_1} \otimes \ldots \otimes e'^{i'_p} \end{aligned}$$

por lo que

$$T'^{j'_1 \ldots j'_q}_{i'_1 \ldots i'_p} = S_{i'_1}^{i_1} \ldots S_{i'_p}^{i_p} R_{j_1}^{j'_1} \ldots R_{j_q}^{j'_q} T_{i_1 \ldots i_p}^{j_1 \ldots j_q}$$

Un tensor $\binom{0}{0}$ es un *escalar*. Permanece invariante frente a cambios de base:

$$T' = T$$

Un tensor $\binom{1}{0}$ representa el conjunto de coordenadas *contravariantes* de un vector. Se transforman como

$$T'^i = R_j^i T^j$$

En forma matricial esto corresponde a $T' = RT$, con T un vector columna.
Por ejemplo, las coordendas x^i de un vector $v = x^i e_i \in V$ se transforman como $x'^i = R_j^i x^j$.

Un tensor $\binom{0}{1}$ representa el conjunto de coordenadas *covariantes* de un vector. Se transforman como

$$T'_i = S_i^j T_j$$

En forma matricial esto corresponde a $T' = TS$, con T un vector fila.
Por ejemplo, las coordenadas a_i de un vector $h = a_i e^i \in V^*$ se transforman como $a'_i = S_i^j a_j$.

Un tensor $\binom{1}{1}$ se transforma como

$$T'^j_i = R_l^j S_i^k T_k^l$$

En forma matricial, esto corresponde a $T_i^j = (RTS)_i^j$, es decir, $T' = RTS$, con $R = S^{-1}$. Un ejemplo son pues las matrices que representan operadores lineales $F : V \to V$. Estos pueden expresarse como $F = F_i^j e_j e^i$, de forma que $F(e_k) = F_i^j e_j (e^i, e_k) = F_k^j e_j$, siendo $F_i^j = [F(e_i)]^j = ([F]_e^e)_i^j$ la matriz que lo representa en la base e. Recordemos que esta matriz se transforma precisamente como $F' = RFS$ con $R = S^{-1}$, o sea, $F'^j_i = R_l^j S_i^k F_k^l$.

Un tensor $\binom{0}{2}$ se transforma como

$$T'_{ij} = S_i^k S_j^l T_{kl}$$

En forma matricial, esto equivale a $T_{ij} = (S^t T S)_{ij}$, es decir, $T' = S^t T S$. Un ejemplo son pues las matrices que representan formas cuadráticas (funciones de $V \times V \to K$), de elementos $A_{ij} = A(e_i, e_j)$, las que se transforman como $A' = S^t A S$, es decir, $A'_{ij} = S_i^k A_{kl} S_j^l$. En forma análoga se ve el caso de un tensor $\binom{2}{0}$ (funciones de $V^* \times V^*$ en K).

27.5 Producto Tensorial de Espacios Vectoriales. Recordemos aquí que el **producto tensorial** $V \otimes W$ de dos espacios vectoriales V, W sobre el mismo cuerpo K, de dimensiones n y m respectivamente, es el espacio generado por los productos $\{e_i \otimes \tilde{e}_j\}$, $i = 1, \ldots, n$, $j = 1, \ldots, m$, donde $\{e_1, \ldots, e_n\}$ es una base de V y $\{\tilde{e}_1, \ldots, \tilde{e}_m\}$ una base de W. Se verifica, $\forall\, v \in V$, $w \in W$ y $\alpha \in K$,

$$\alpha(v \otimes w) = (\alpha v) \otimes w = v \otimes (\alpha w)$$

$$(v_1 + v_2) \otimes w = v_1 \otimes w + v_2 \otimes w, \quad v \otimes (w_1 + w_2) = v \otimes w_1 + v \otimes w_2$$

$$0 \otimes w = v \otimes 0 = 0$$

Si $u \in V \otimes W \Rightarrow$

$$u = \sum_{i=1}^{n} \sum_{j=1}^{m} c_{ij}\, e_i \otimes \tilde{e}_j, \quad c_{ij} \in K$$

Destaquemos que esto incluye vectores producto $u = v \otimes w$, con $v \in V$ y $w \in W$, como así también vectores que son combinaciones lineales de productos pero que no pueden ser escritos como un único producto. La dimensión de $V \times W$ es $n \times m$ (y no $n + m$, como sucede con $V \times W$).

En mecánica cuántica, el espacio de estados de un sistema compuesto por dos subistemas distinguibles es justamente el producto tensorial de los espacios de estados de cada subsistema, siendo estos últimos espacios de Hilbert ($K = \mathbb{C}$). Para $e_i \otimes \tilde{e}_j$ se emplea la notación $|i\rangle \otimes |\tilde{j}\rangle$ o directamente $|i\rangle|\tilde{j}\rangle$ o $|i\tilde{j}\rangle$.
Los estados producto $|u\rangle = |v\rangle \otimes |w\rangle$ se denominan estados **separables**, mientras que los estados que no pueden ser escritos como producto se denominan correlacionados o **entrelazados**.

27.6 Producto y Suma de tensores

Sea T un tensor $\binom{q}{p}$ y U un tensor $\binom{q'}{p'}$ sobre el mismo espacio. Su producto es un tensor $\binom{p+p'}{q+q'}$ dado por

$$(TU)^{j_1 \ldots j_{q+q'}}_{i_1 \ldots i_{p+p'}} = T^{j_1 \ldots j_q}_{i_1 \ldots i_p}\, U^{j_{q+1} \ldots j_{q+q'}}_{i_{p+1} \ldots i_{p+p'}}$$

La suma está definida para tensores del mismo rango $\binom{p}{q}$: $(T+U)^{j_1 \ldots j_q}_{i_1 \ldots i_p} = T^{j_1 \ldots j_q}_{i_1 \ldots i_p} + U^{j_1 \ldots j_q}_{i_1 \ldots i_p}$.

27.7 Producto tensorial de operadores. Si $F : V \to V$ y $G : W \to W$ son operadores lineales en espacios V, W, entonces $F \otimes G : V \otimes W \to V \otimes W$ es un operador lineal en el espacio producto tensorial $V \otimes W$, definido por

$$(F \otimes G)(v \otimes w) = F(v) \otimes G(w)$$

Si $F(v_i) = \lambda_i^F v_i$, $G(w_j) = \lambda^G w_j$, entonces

$$(F \otimes G)(v_i \otimes w_j) = \lambda_i^F \lambda_j^G v_i \otimes w_j$$

por lo que si F y G son diagonalizables, $(F \otimes G)$ también lo es, con $n \times m$ autovalores $\lambda_i^F \lambda_j^G$, $i = 1, \ldots, n$, $j = 1, \ldots, m$. Además, $\mathrm{Det}(F \otimes G) = \mathrm{Det}(F)^m \mathrm{Det}(G)^n$. Notemos finalmente que $(F \otimes G)^k = F^k \otimes G^k$, válido para $k \in \mathbb{N}$ y también $k \in \mathbb{Z}$ si F y G son invertibles.

Si $F = F_j^i e_i e^k$, $G = G_l^k \tilde{e}_k \tilde{e}^l$, $\Rightarrow F \otimes G = F_j^i G_l^k (e_i \otimes \tilde{e}_k)(e^j \otimes \tilde{e}^l)$, por lo que $(F \otimes G)^{ik}_{jl} = F_k^i G_l^j$. Esto corresponde pues al producto tensorial de las matrices que representan a F y G, denominado también producto Kronecker: Ordenando la base en la forma $b = (e_i \otimes \tilde{e}_1, e_1 \otimes \tilde{e}_2, \ldots, e_n \otimes \tilde{e}_m)$, la matriz de $nm \times nm$ que representa a $F \otimes G$ en esta base es

$$[F \otimes G]_b = [F]_e \otimes [G]_{\tilde{e}} = \begin{pmatrix} F_1^1 [G]_{\tilde{e}} & \ldots & F_n^1 [G]_{\tilde{e}} \\ \vdots & \ddots & \vdots \\ F_1^n [G]_{\tilde{e}} & \ldots & F_n^n [G]_{\tilde{e}} \end{pmatrix}$$

En notación de Mecánica Cuántica, $e_i \to |i\rangle$, $e^j \to \langle j|$ y $F \to \sum_{i,j} F_{ij} |i\rangle\langle j|$, $G = \sum_{k,l} G_{kl} |\tilde{k}\rangle\langle\tilde{l}|$, con $F \otimes G = \sum_{i,j,k,l} F_{ij} G_{kl} |i\tilde{j}\rangle\langle k\tilde{l}|$.

27.8 Contracción de tensores

La contracción de un tensor $\binom{p}{q}$, con $p \geq 1$, $q \geq 1$, queda definida por una suma de la forma

$$T^{j_1 \ldots k \ldots j_q}_{i_1 \ldots \ldots k \ldots i_p}$$

(donde la suma es sobre el índice repetido k), la cual se transforma como un tensor $\binom{p-1}{q-1}$, pues $S^i_k R^k_j = \delta^i_j$. Por ejemplo, si

$$U^j_i = T^{kj}_{ik}$$

entonces

$$U'^{j'}_{i'} = T'^{k'j'}_{i'k'} = S^i_{i'} S^l_{k'} R^{k'}_k R^{j'}_j T^{kj}_{il} = S^i_{i'} \delta^l_k R^{j'}_j T^{kj}_{il} = S^i_{i'} R^{j'}_j T^{kj}_{ik} = S^i_{i'} R^{j'}_j U^j_i$$

donde hemos utilizado $S^l_{k'} R^{k'}_k = \delta^l_k$. Vemos pues que se transforma como un tensor $\binom{1}{1}$.

Así, dado un tensor T^{ij}_{kl} (tensor $\binom{2}{2}$) son posibles las 4 contracciones

$$T^{kj}_{ki}\ ,\ \ T^{jk}_{ik}\ ,\ \ T^{kj}_{ik}\ ,\ \ T^{jk}_{ki}$$

que originan 4 tensores $\binom{1}{1}$ (en general distintos). Por otro lado, las dos posibles contracciones dobles que dan lugar a un escalar (tensor $\binom{0}{0}$) son

$$T^{kj}_{kj}\ ,\ \ T^{jk}_{kj}$$

Por ejemplo, dado el tensor T^j_i, la única contracción posible es el escalar T^i_i. Este representa la *traza* de la matriz T:

$$\mathrm{Tr}\, T = T^i_i$$

Esta es, como hemos visto, invariante frente a cambios de base.

Dado el tensor producto $T^{jk}_{il} = F^j_i G^k_l$, el escalar $T^{jk}_{jk} = F^j_j G^k_k$ representa, matricialmente, el producto de trazas: $(\mathrm{Tr}F)(\mathrm{Tr}G) = F^i_i G^k_k$, mientras que el escalar $T^{jk}_{kj} = F^j_k G^k_j$ representa la traza del producto: $\mathrm{Tr}(FG) = F^j_k G^k_j$.

Además, la contracción $T^{jk}_{ki} = F^j_k G^k_i$ es un tensor $\binom{1}{1}$, que representa el producto matricial FG.

Un tensor es simétrico respecto a dos índices del mismo tipo si $T^{\ldots i \ldots j \ldots}_{\ldots} = T^{\ldots j \ldots i \ldots}_{\ldots}$, y es antisimétrico si $T^{\ldots i \ldots j \ldots}_{\ldots} = -T^{\ldots j \ldots i \ldots}_{\ldots}$ (Definición similar respecto de índices inferiores). Esta propiedad es independiente de la base: Por ejemplo, si $T^{ij}_{kl} = T^{ji}_{kl}$,

$$T'^{i'j'}_{k'l'} = R^{i'}_i R^{j'}_j S^k_{k'} S^j_{j'} T^{ij}_{kl} = R^{i'}_i R^{j'}_j S^k_{k'} S^j_{j'} T^{ji}_{kl} = T'^{j'i'}_{k'l'}$$

Un tensor es completamente simétrico (antisimétrico) si es simétrico (antisimétrico) respecto de todo par de índices del mismo tipo.

27.9 Determinante: Consideremos una forma multilineal completamente antisimétrica de $V^n \to K$. En tal caso, si $v_i = x^j_i e_j$,

$$F(v_1, \ldots, v_n) = x^{i_1}_1 \ldots x^{i_n}_n F_{i_1, \ldots, i_n}$$

donde $F_{i_1,\ldots,i_n} = F(e_{i_1}, \ldots, e_{i_n})$. Se tiene $F_{\ldots,i,\ldots,j,\ldots} = -F_{\ldots,j,\ldots,i,\ldots}$ para cualquier par de índices i, j. Es claro entonces que $F_{\ldots,i,\ldots,j,\ldots} = 0$ si $i = j$, es decir, si dos (o más) índices coinciden, y que si los índices son todos distintos, $F_{i_1,\ldots,i_n} = (-1)^{n_{i_1,\ldots,i_n}} F_{1,2\ldots,n}$, donde $n_{i_1,\ldots,i_n}$ es el número de permutaciones necesarias para llevar $(i_1, \ldots, i_n)$ al orden normal $(1, 2, \ldots, n)$. Podemos pues escribir

$$F_{i_1,\ldots,i_n} = \lambda \epsilon_{i_1,\ldots,i_n}$$

donde $\lambda = F_{1,2,\ldots,n}$ y $\epsilon_{i_1,\ldots,i_n}$ es el símbolo completamente antisimétrico que satisface $\epsilon_{1,2,\ldots,n} = 1$ (símbolo de Levi-Civita). Por lo tanto,

$$F(v_1, \ldots, v_n) = \lambda x^{i_1}_1 \ldots x^{i_n}_n \epsilon_{i_1 \ldots i_n} = \lambda\, \mathrm{Det}[X]$$

donde

$$\mathrm{Det}[X] = x^{i_1}_1 \ldots x^{i_p}_n \epsilon_{i_1 \ldots i_n}$$

es el determinante de la matriz de elementos x^i_j (la cual es una función multilineal completamente antisimétrica de las columnas de la matriz, que vale 1 para la matriz identidad). Por ejemplo, para $n=2$, $\mathrm{Det}[X]=x^i_1x^j_2\epsilon_{ij}=x^1_1x^2_2\epsilon_{12}+x^2_1x^1_2\epsilon_{21}=x^1_1x^2_2-x^2_1x^1_2$, mientras que para $n=3$,
$\mathrm{Det}[X]=x^i_1x^j_2x^k_3\epsilon_{ijk}=x^1_1x^2_2x^3_3\epsilon_{123}+x^1_1x^3_2x^2_3\epsilon_{132}+x^2_1x^3_2x^1_3\epsilon_{231}+x^2_1x^1_2x^3_3\epsilon_{213}+x^3_1x^1_2x^2_3\epsilon_{312}+x^3_1x^2_2x^1_3\epsilon_{321}$
$\quad=x^1_1x^2_2x^3_3-x^1_1x^3_2x^2_3+x^2_1x^3_2x^1_3-x^2_1x^1_2x^3_3+x^3_1x^1_2x^2_3-x^3_1x^2_2x^1_3$.
Notemos también que $x^i_1x^j_2\epsilon_{ij}=x^1_1x^2_2-x^2_1x^1_2=x^1_1x^2_2-x^1_2x^2_1=x^1_ix^2_j\epsilon^{ij}$, donde $\epsilon^{ij}=\epsilon_{ij}$, y en general, $\mathrm{Det}[X]=x^{j_1}_1\ldots x^{j_n}_n\epsilon_{j_1\ldots j_n}=\frac{1}{n!}x^{j_1}_{i_1}\ldots x^{j_n}_{i_n}\epsilon_{j_1\ldots j_n}\epsilon^{i_1\ldots i_n}=x^1_{i_1}\ldots x^n_{i_n}\epsilon^{i_1\ldots i_n}$,
donde $\epsilon^{i_1\ldots i_n}=\epsilon_{i_1\ldots i_n}$.

Observemos que frente a un cambio de base general, $F_{i_1,\ldots,i_n}=F(e_{i_1},\ldots,e_{i_n})$ transforma como

$$F'_{i'_1\ldots i'_n}=S^{i_1}_{i'_1}\ldots S^{i_n}_{i'_n}F_{i_1\ldots i_n}=\lambda S^{i_1}_{i'_1}\ldots S^{i_n}_{i'_n}\epsilon_{i_1\ldots i_n}=\lambda\mathrm{Det}(S)\epsilon_{i'_1\ldots i'_n}=\mathrm{Det}(S)F_{i'_1,\ldots,i'_p}$$

Suba y baja de índices y tensores cartesianos. En un espacio euclideo, es posible bajar o subir índices de un tensor mediante el tensor métrico$g_{ij}=(e_i,e_j)$, y su inversa $g^{ij}=(e^i,e^j)$, que son tensores simétricos de tipo $\binom{0}{2}$ y $\binom{2}{0}$ respectivamente:

$$\begin{aligned}T^{j_1,\ldots,j_q}_{i_1\ldots,i_p} &= T(e_{i_1},\ldots,e_{i_p},e^{j_1},\ldots,e^{j_q})=T(e_{i_1},\ldots,e_{i_p},g^{j_1j'_1}e_{j'_1},\ldots,g^{j_qj'_q}e_{j'_q})\\ &= g^{j_1j'_1}\ldots g^{j_qj'_q}T(e_{i_1},\ldots,e_{i_p},e_{j_1},\ldots,e_{jq})=g^{j_1j'_1}\ldots g^{j_qj'_q}T_{i_1\ldots,i_p,j'_1,\ldots,j'_q}\end{aligned}$$

Por ejemplo, si T^j_i es un tensor $\binom{1}{1}$, $T^{ji}=g^{ki}T^j_k$ es un tensor $\binom{2}{0}$ y $T_{ji}=g_{jk}T^k_i$ es un tensor $\binom{0}{2}$. *Tensores cartesianos*: En un espacio euclideo V, si nos restringimos a transformaciones isométricas entre bases ortonormales, entonces $g_{ij}=(e_i,e_j)=\delta_{ij}$, $g^{ij}=\delta^{ij}$ y $e^i=g^{ij}e_j=e_i$. En tal caso no se puede distinguir entre índices covariantes y contravariantes y se tiene $T^i=T_i$, $T^i_j=T^{ij}=T_{ij}$, $T^{ij}_{kl}=T_{ijkl}$, etc.

Notemos precisamente que para transformaciones entre bases ortonormales (isometrías) $R=S^{-1}=S^t$, es decir, $R^i_j=S^j_i$. En tal caso, $T'^j=R^j_iT^i=\sum_i S^i_jT^i$, verificándose que T^j se transforma igual que T_j.

Pseudotensores cartesianos: Si frente a un cambio de base isométrico en un espacio euclideo se tiene

$$T'_{i'_1\ldots i'_p}=\mathrm{Det}(S)S^{i_1}_{i'_1}\ldots S^{i_p}_{i'_p}T_{i_1\ldots i_p}$$

se dice que T es un pseudotensor cartesiano de rango p. Se comporta como un tensor de rango p frente a cambios de base que satisfacen $\mathrm{Det}[S]=+1$ (rotaciones) pero exhibe un cambio de signo adicional si $\mathrm{Det}[S]=-1$ (reflexiones).

Por ejemplo, frente a isométrías, el tensor completamente antisimétrico $F_{i_1,\ldots i_n}=F(e_1,\ldots,e_n)$ es un pseudoescalar, mientras que $(\mathbf{a}\times\mathbf{b})_k=a^ib^j\epsilon_{ijk}$ es un pseudovector $(a'^{i'}b'^{j'}\epsilon_{i'j'k}=R^{i'}_iR^{j'}_ja^ib^j\epsilon_{i'j'k}=R^{i'}_iR^{j'}_jR^{k'}_lS^l_ka^ib^j\epsilon_{i'j'k'}=\mathrm{Det}(R)S^l_ka^ib^j\epsilon_{ijl}=\mathrm{Det}(R)S^l_k(\mathbf{a}\times\mathbf{b})_l)$.

28 Campos tensoriales, símbolos de Christoffel y derivada covariante

Consideremos un cambio general de coordenadas $x'^i(x^1,\ldots,x^n)$ en $V=\mathbb{R}^n$. Tenemos

$$dx'^i=R^i_jdx^j,\quad R^i_j=\frac{\partial x'^i}{\partial x^j}=\partial_jx'^i$$

La matriz inversa es

$$S^i_j=\frac{\partial x^i}{\partial x'^j}=\partial'_jx^i$$

y satisface

$$S^i_jR^j_k=R^i_jS^j_k=\delta^i_k$$

Tanto S como R *dependen ahora de las coordenadas.* Podemos considerar en c/punto la base definida por

$$e'_i=S^j_ie_j$$

siendo aquí $e=(e_1,\ldots,e_n)$ una base de V *independiente* de las coordenadas, y $e'=(e'_1,\ldots,e'_n)$ *dependiente* de las coordenadas.

Si e es la base canónica, el tensor métrico original es $g_{ij} = (e_i, e_j) = \delta_{ij}$ mientras que en la nueva base, $g'_{ij} = (e'_i, e'_j) = S^k_i S^l_j g_{kl} = S^k_i S^l_j \delta_{kl}$, es decir, $g' = S^T S$ en notación matricial. Se obtiene entonces

$$ds^2 \equiv dx_i dx^i = dx'_i dx'^i = dx'^i dx'^j g'_{ij}$$

Un campo vectorial v dependiente de las coordenadas puede pues escribirse como

$$v = v^i(x^1, \ldots, x^n)e_i = v'^i(x'^1, \ldots, x'^n)e'_i, \quad v'^i = R^i_j v^j$$

Generalizando, si $D \subset V$, un campo tensorial real $\binom{q}{p}$ es una función $T : D \to \underbrace{V \otimes \ldots \otimes V}_{\text{q veces}} \otimes \underbrace{V^* \otimes \ldots \otimes V^*}_{\text{p veces}}$:

$$T = T^{j_1,\ldots,j_q}_{i_1,\ldots,i_p}(x^1, \ldots, x^n)e_{j_1} \otimes \ldots \otimes e_{j_q} \otimes f^{i_1} \otimes f^{i_p}$$

Frente a un cambio general de coordenadas, se obtiene

$$T = T'^{j'_1,\ldots,j'_q}_{i'_1,\ldots,i'_p}(x'^1, \ldots, x'^n)e'_{j'_1} \otimes \ldots \otimes e'_{j'_q} \otimes f'^{i_1} \otimes f'^{i_p}$$

con

$$T'^{j'_1,\ldots,j'_q}_{i'_1,\ldots,i'_p}(x'^1, \ldots, x'^n) = S^{i_1}_{i'_1} \ldots S^{i_p}_{i'_p} R^{j'_1}_{j_1} \ldots R^{j'_q}_{j_q} T^{j_1,\ldots,j_q}_{i_1,\ldots,i_p}(x^1, \ldots, x^n)$$

Por ejemplo, un campo vectorial es un campo tensorial $\binom{1}{0}$.

Consideremos ahora la derivada de un campo tensorial $\binom{1}{0}$,

$$\partial'_j v = \partial'_j(v'^i e'_i) = (\partial'_j v'^i)e'_i + v'^i(\partial'_j e'_i)$$

El segundo término da cuenta de la dependencia de la base de las coordenadas. Dado que $e'_i = S^k_i e_k$, se tiene $\partial'_j e'_i = (\partial'_j S^l_i)e_l = (\partial'_j S^l_i)R^k_l e'_k$ y por lo tanto

$$\partial'_j e'_i = \Gamma^k_{ij} e'_k$$

donde $\Gamma^k_{ij} = (\partial'_j S^l_i)R^k_l = -S^l_i \partial'_j R^k_l$ son los *símbolos de Christoffel*, que dan cuenta de la variación de los elementos de la base. Como $S^i_j = \partial'_j x^i \Rightarrow \Gamma^k_{ij} = \Gamma^k_{ji}$, pues $\partial'_j S^l_i = \partial'_j \partial'_i x^l = \partial'_i \partial'_j x^l = \partial'_i S^l_j$.

Se obtiene entonces

$$\partial'_j v = [(\partial'_j v'^k) + v'^i \Gamma^k_{ij}]e'_k$$

La expresión

$$v'^k_{;j} \equiv v'^k_{,j} + v'^i \Gamma^k_{ij}$$

donde $v'^k_{,j} \equiv \partial'_j v'^k$, se denomina *derivada covariante* de las componentes contravariantes, y satisface las reglas correctas de transformación. Tenemos pues

$$\partial'_j v = v'^k_{;j} e'_k$$

En el caso de que la base sea independiente de las coordenadas, $\Gamma^k_{ij} = 0$ y la derivada covariante se reduce a la usual ($v^i_{;j} = v^i_{,j}$).

Por ejemplo, la divergencia de un campo vectorial $v = v^i e_i = v'^i e'_i$ puede entonces expresarse en la forma (demostrar como ejercicio)

$$\partial_i v^i = v^i_{,i} = v'^i_{;i} = (\partial'_i v'^i) + v'^i \Gamma^j_{ij}$$

Para componentes covariantes, tenemos $v = v_i e^i = v'_i e'^i$, con $e'^i = R^i_k e^k$, y e^k independiente de las coordenadas. Por lo tanto,

$$\partial'_j v = (\partial'_j v'_i)e'^i + v'^i(\partial'_j e'^i)$$

Pero $\partial'_j e'^i = (\partial'_j R^i_l)e^l = S^l_k(\partial'_j R^i_l)e'^k = -\Gamma^i_{kj}$ por lo que

$$\partial'_j v = [(\partial'_j v'_k) - v'_i \Gamma^i_{kj}]e'^k$$

La derivada covariante de componentes covariantes debe pues definirse como

$$v'_{k;j} = v'_{k,j} - v'_i \Gamma^i_{kj}$$

para que

$$\partial'_j v = v'_{k;j} e'^k$$

En forma análoga se definen las derivadas covariantes de tensores arbitrarios de rango $\binom{p}{q}$

Dado que $g'_{ik} = S^l_i S^m_k g_{lm}$, tenemos, para g_{lm} independiente de las coordenadas, $\partial'_j g'_{ik} = (\partial'_j S^l_i) S^m_k g_{lm} + S^l_i (\partial'_j S^m_k) g_{lm} = (\partial'_j S^l_i) R^r_l S^s_r S^m_k g_{sm} + (\partial'_j S^m_k) R^r_m S^s_r S^l_i g_{ls} = \Gamma^r_{ij} g'_{rk} + \Gamma^r_{kj} g'_{ir}$, por lo que

$$g'_{ik;j} = g'_{ik,j} - g'_{lk}\Gamma^l_{ij} - g'_{il}\Gamma^l_{kj} = 0$$

De esta forma, si $v'_i = g'_{ik} v'^k$ se verifica que $v'_{i;j} = g'_{ik} v'^k_{;j}$. La última ecuación permite también escribir los símbolos de Christoffel directamente en términos de derivadas del tensor métrico:

$$\Gamma^i_{kl} = \frac{1}{2} g^{im}(g_{mk,l} + g_{ml,k} - g_{kl,m})$$

Ejemplo: Para $V = \mathbb{R}^2$ y coordenadas polares, definidas por

$$x = r\cos\theta\,, \quad y = r\sin\theta$$

se obtiene $dx = dr\cos\theta - r\sin\theta d\theta$, $dy = dr\sin\theta + r\cos\theta d\theta$, de forma que

$$S = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix}, \quad R = \frac{1}{r}\begin{pmatrix} r\cos\theta & r\sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}, \quad g' = \begin{pmatrix} 1 & 0 \\ 0 & r^2 \end{pmatrix}$$

con $dr = dx\cos\theta + dy\sin\theta$, $d\theta = (-dx\sin\theta + dy\cos\theta)/r$,
$e_r = e_x\cos\theta + e_y\sin\theta$, $e_\theta = r(-e_x\sin\theta + e_y\cos\theta)$, y e_x, e_y la base canónica. Obtenemos entonces

$$ds^2 = dx^2 + dy^2 = dr^2 + r^2 d\theta^2$$

En este caso, los únicos símbolos de Christoffel no nulos son $\Gamma^\theta_{r\theta} = \Gamma^\theta_{\theta r} = 1/r$, $\Gamma^r_{\theta\theta} = -r$.

La divergencia de un campo vectorial

$$v = v^x e_x + v^y e_y = v^r e_r + v^\theta e_\theta$$

es entonces

$$\partial_x v^x + \partial_y v^y = \partial_r v^r + \partial_\theta v^\theta + v^r \Gamma^\theta_{r\theta} = \partial_r v^r + \partial_\theta v^\theta + v^r/r$$

El gradiente de un campo escalar ϕ puede escribirse en la forma $(\partial^i\phi)e_i = (\partial'^i\phi)e'_i$, donde $\partial'^i = g'^{ij}\partial'_j$. Por lo tanto,

$$\frac{\partial\phi}{\partial x}e_x + \frac{\partial\phi}{\partial y}e_y = \frac{\partial\phi}{\partial r}e_r + \frac{1}{r^2}\frac{\partial\phi}{\partial\theta}e_\theta$$

Finalmente, el Laplaciano de un campo escalar ϕ (la divergencia del gradiente de ϕ) puede expresarse como

$$\partial_i\partial^i\phi = \partial'_i\partial'^i\phi + \Gamma^i_{ji}\partial'^j\phi = \frac{\partial^2\phi}{\partial r^2} + \frac{1}{r^2}\frac{\partial^2\phi}{\partial\theta^2} + \frac{1}{r}\frac{\partial\phi}{\partial r}$$