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

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

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**

$$|(v,w)| \leq ||v||\,||w|| \tag{19.1}$$

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 $V = \mathbb{R}^{m\times n}$,
$$\text{Tr}[A^tB] \leq \sqrt{\text{Tr}[A^tA]}\sqrt{\text{Tr}[B^tB]}$$
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 $v \in V$,

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

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 $(e_i, v) = (e_i, \sum_{j=1}^{n} x_j e_j) = \sum_{j=1}^{n} x_j (e_i, e_j) = x_i$ 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,

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

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 $||v||^2 = \sum_{i=1}^{n} ||x_i e'_i||^2 = \sum_{i=1}^{n} x_i^2 ||e'_i||^2$ 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 $[F]_{\tilde{e}}^{e} \in \mathbb{R}^{m \times n}$ 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

$$(e'_j, e'_k) = \sum_{i=1}^{n} S_{ij} S_{ik} = (S^t S)_{jk}$$

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 $|S| \equiv \mathrm{Det} S = \pm 1$, 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

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

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:

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

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 $d_{\rm min}(v,S)$.

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 $(AX)_i=A_iX$ 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

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

de donde

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

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

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

La matriz que representa al proyector sobre S es entonces

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

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^tR.
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

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

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 $u=\sum_{i=1}^n \beta_i v_i$,

$$(u,u)=\beta^tG\beta=\beta^t0=0$$

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 defición 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, $F_{ij} = (\tilde{e}_i, F(e_j)) = (F^\dagger(\tilde{e}_i), e_j) = (e_j, F^\dagger(\tilde{e}_i)) = F_{ji}^\dagger$.

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 $S^tS = I_n$, 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 $n \times n$

$$A^t A$$

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*:

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

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

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

que son ortonormales:

$$u_j^t u_i = \frac{1}{\sqrt{\lambda_j\lambda_i}} v_j^t A^t A v_i = \frac{\lambda_i v_j^t v_i}{\sqrt{\lambda_i\lambda_j}} = \delta_{ij}$$

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 $F : V \to W$ 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 $B=R^tAS$, con $R^tR=I_m$, $S^tS=I_n \Rightarrow B^tB=S^tA^tRR^tAS=S^tA^tAS$, y los autovalores de B^tB son entonces idénticos a los de A^tA.

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

$$A^t=VA'^tU^t$$

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^tA 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 $|U|=\pm1$, $|V|=\pm1$, por lo que $|\mathrm{Det}[A]|=\mathrm{Det}[A']$.

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* $(A^t=A)\Rightarrow A^tA=A^2$, por lo que $\lambda_i=(\lambda_i^A)^2$, con λ_i^A los autovalores de A. Se obtiene entonces

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

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 $U=(s_1v_1,\ldots,s_nv_n)$, 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 $v_iv_i^t$ 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

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

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:

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

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

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

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

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

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)$= n \Rightarrow \tilde{A} = (A^t A)^{-1}A^t$, 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 ($b \notin EC(A)$) 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 $[G]_{e'} = S^t [G]_e S = [G]_e$) 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 ($ss' = -1$).