Álgebra Lineal

Resumen

Matías Palumbo - 2020

Basado en:

Linear Algebra and its Applications, G. Strang

Apuntes teóricos de la cátedra de Álgebra Lineal (1° cuat. 2020)

Índice

Unidad I

Matrices y Eliminación Gaussiana

1. Preliminares

Dadas las matrices A $m \times n$ y B $n \times p$ y los vectores x $n \times 1$ e y $1 \times n$:

- El producto de una matriz por un vector es una combinación lineal de los vectores columna de la matriz:
$$Ax = \sum_{j=1}^{n} x_j A^j.$$
- El producto de un vector por una matriz es una combinación lineal de las filas de la matriz:
$$yB = \sum_{i=0}^{n} y_i B_i.$$
- Cada entrada de AB es el producto de un vector fila de A y un vector columna de B:
$$(AB)_{ij} = A_i B^j.$$
- Cada columna de AB es el producto de A por cada columna de B:
$$(AB)^j = AB^j.$$
- Cada fila de AB es el producto de cada fila de A por B:
$$(AB)_i = A_i B.$$

Proposición 1. *Si A es una matriz tal que $AB = BA$ para toda matriz B, entonces $A = \lambda I$, $\lambda \in \mathbb{R}$.*

2. Método de Eliminación de Gauss

Definición 1. Una **matriz elemental** $E_{ij}(\ell)$ de tamaño $m \times m$ es la matriz identidad cambiando su entrada i, j, con $i \neq j$ por el valor $\ell \neq 0$.

Lema 1. *Dada una matriz A, la matriz $E_{ij}(\ell)A$ es la matriz que se obtiene a partir de A, cambiando la fila A_i por $A_i + \ell A_j$.*

Definición 2. Una **matriz de permutación** P_{ij} de tamaño $n \times n$ es una matriz obtenida de la permutación de filas de la identidad $n \times n$. Una matriz de permutación es simple si solo han sido permutadas dos filas entre sí. Toda matriz de permutación es producto de matrices de permutación simples.

Método de Eliminación Gaussiana: Dado un sistema $m \times n$, se trabaja con la matriz asociada A $m \times n$ de coeficientes:

1. Si el primer elemento de la primera columna no nula de A es 0, se intercambia la primera fila con alguna que tenga un elemento no nulo en esa columna (si no, se deja como está).
2. Se obtienen ceros debajo de este elemento restando a las filas de abajo múltiplos de la fila en cuestión.
3. Se aplica este procedimiento a la submatriz obtenida quitando la primera fila y columna no nulas.

2.1. Costo Computacional

Para matrices considerablemente grandes, una buena estimación del número de operaciones necesarias en el método de eliminación gaussiana es $\frac{n^3}{3}$ operaciones, y n^2 operaciones en el lado derecho del sistema.

Para hallar la inversa de una matriz utilizando el método de Gauss-Jordan explicado más adelante, la cantidad de operaciones tiende a n^3 (considerablemente bajo; sin embargo, a menos que explícitamente se deba calcular A^{-1}, no se debe usar Gauss-Jordan en la resolución de sistemas).

2.2. Factorización LU y LDV

Lema 2. *La matriz $E_{ij}(\ell)A$ es la matriz que se obtiene a partir de A, cambiando la fila A_i por $A_i + \ell A_j$.*

Dada una matriz cuadrada A, si se puede terminar el proceso de Eliminación Gaussiana (es decir, si existe una permutación de las filas de A que «evita» los ceros en las posiciones pivots), entonces se termina con una matriz triangular superior U con los pivots en la diagonal, en donde por cada operación realizada sobre las filas de A, se multiplica a A por la matriz elemental o de permutación que simboliza dicha operación. Es decir, con E el producto de las matrices elementales usadas y P la matriz de permutación usada, resulta $EPA = U$. Entonces, considerando E^{-1} (cuya existencia se puede asegurar debido a la inversibilidad de las matrices elementales), se tiene que $PA = E^{-1}U = LU$, en donde L es una matriz triangular inferior con 1's en la diagonal y con los multiplicadores de las operaciones realizadas sobre las filas debajo de ella.

Observación 1. Si, por ejemplo, se resta ℓ veces la fila j a la fila i (es decir, se suma $-\ell$ veces), entonces la entrada i, j de L es ℓ.

Teorema 1. *(La descomposición LU es única) Sea la matriz $n \times n$ A tal que $A = L_1U_1$ y $A = L_2U_2$, donde, para $i = 1, 2$, L_i es triangular inferior con 1's en la diagonal, y U_i es triangular superior sin ceros en la diagonal. Entonces, $L_1 = L_2$ y $U_1 = U_2$.*

Demostración. Como L_2 y U_1 son inversibles (al ser triangulares sin ceros en la diagonal),

$$L_1U_1 = L_2U_2 \implies L_2^{-1}L_1U_1U_1^{-1} = L_2^{-1}L_2U_2U_1^{-1} \implies L_2^{-1}L_1 = U_2U_1^{-1}.$$

Luego, como L_1 y L_2^{-1} son triangulares inferiores, $L_2^{-1}L_1$ también lo es. Además, como U_2 y U_1^{-1} son triangulares superiores, $U_2U_1^{-1}$ también lo es. Entonces, la matriz $L_2^{-1}L_1 = U_2U_1^{-1}$ es diagonal, y

$$(L_2^{-1}L_1)_{ii} = \sum_{k=1}^{n}(L_2^{-1})_{ik}(L_1)_{ki} = (L_2^{-1})_{ii}(L_1)_{ii} = 1, \text{ pues si } i \neq k,\ (L_2^{-1})_{ik}(L_1)_{ki} = 0.$$

$\therefore\ L_2^{-1}L_1 = I \implies L_1 = L_2$, y $U_2U_1^{-1} = I \implies U_1 = U_2$. □

Análogamente, si existe una permutación de las filas de A que «evita» los ceros en las posiciones pivots, A admite una factorización LDV, donde L es la misma matriz L de la factorización LU, D es una matriz diagonal con los pivots, y V (a veces llamada también U) es una matriz triangular superior con 1's en la diagonal, resultante de escalar las filas con pivots de la matriz U de la fact. LU como corresponda para hacer 1's en la diagonal.

Teorema 2. *(La descomposición LDV es única) Sea $A = L_1D_1V_1$ y $A = L_2D_2V_2$ donde, para $i = 1, 2$, L_i es triangular inferior con 1's en la diagonal, U_i es triangular superior con 1's en la diagonal y D_i es matriz diagonal sin ceros en la diagonal. Entonces $L_1 = L_2$, $U_1 = U_2$ y $D_1 = D_2$.*

Demostración. Al igual que en la factorización LU, L es una matriz triangular inferior con unos en la diagonal, y como ya se probó la unicidad de esa descomposición, tenemos que, dada una matriz A, $A = LU$. Ahora, supongo que $U = D_1V_1 = D_2V_2$. Se probará que $D_1 = D_2$ y $V_1 = V_2$.

Como D_1 es diagonal sin ceros en la diagonal, es inversible, y como V_2 es triangular superior con unos en la diagonal, también lo es. Luego, $D_1V_1 = D_2V_2 \implies V_1V_2^{-1} = D_1^{-1}D_2$. Por otro lado, sabemos que $D_1^{-1}D_2$ es diagonal, y por ende $V_1V_2^{-1}$ lo es. Además,

$$(V_1V_2^{-1})_{ii} = \sum_{k=1}^{n}(V_1)_{ik}(V_2^{-1})_{ki} = (V_1)_{ii}(V_2^{-1})_{ii} = 1.$$

$\therefore\ D_1^{-1}D_2 = I \implies D_1 = D_2$, y $V_1V_2^{-1} = I \implies V_1 = V_2$. □

3. Matrices Inversibles y Gauss-Jordan

Una matriz A es no singular si existe una permutación de sus filas tal que no aparecen ceros en las posiciones pivots. Equivalentemente a esto, las siguientes ocurrencias son equivalentes e implican que A

es no singular:

- El sistema $Ax = b$ tiene única solución para cualquier vector b ;
- $\det(A) \neq 0$;
- A es inversible ($\exists B/AB = BA = I$).

Lema 3. *Toda matriz tiene a lo sumo una matriz inversa.*

Demostración. Si una matriz no es inversible, no tiene inversa. Sea A una matriz invertible. Supongo a B y C como inversas de A. Entonces, $BA = I$ y $AC = I$.

Luego, $B = BI = B(AC) = (BA)C = IC = C$. Es decir, la inversa es única. □

Luego de probar la unicidad, se nota A^{-1} a la inversa de una matriz invertible A. Además, considerando a las matrices inversibles A, B $n \times n$, tienen las siguientes propiedades:

- A^{-1} es inversible, y $(A^{-1})^{-1} = A$.
- AB es inversible, y $(AB)^{-1} = B^{-1}A^{-1}$.

3.1. Gauss-Jordan

Dada una matriz no singular A $n \times n$, para hallar su inversa se debe buscar una matriz X tal que $AX = I$. Equivalentemente, se pueden resolver n sistemas $Ax^k = e_k$, con $e_k$el vector canónico con 1 en la k-ésima coordenada, para $k = 1, \ldots, n$. El método de Gauss-Jordan permite resolver los n sistemas simultáneamente. En vez de finalizar al obtener la matriz triangular superior U, se siguen restando múltiplos de una fila *a las filas de arriba*. Esto produce ceros arriba de la diagonal, y al llegar a la matriz identidad (dividiendo los pivotes por sí mismos como último paso), se ha obtenido A^{-1}. Se trabaja con la matriz A y la matriz identidad I lado a lado.

Ejemplo:

$$\begin{bmatrix}\mathbf{A} & e_1 & e_2 & e_3\end{bmatrix} = \begin{bmatrix}\mathbf{2} & \mathbf{1} & \mathbf{1} & 1 & 0 & 0\\ \mathbf{4} & \mathbf{-6} & \mathbf{0} & 0 & 1 & 0\\ \mathbf{-2} & \mathbf{7} & \mathbf{2} & 0 & 0 & 1\end{bmatrix}$$

$$\textbf{pivot} = \mathbf{2} \longrightarrow \begin{bmatrix}\mathbf{2} & \mathbf{1} & \mathbf{1} & 1 & 0 & 0\\ \mathbf{0} & \mathbf{-8} & \mathbf{-2} & -2 & 1 & 0\\ \mathbf{0} & \mathbf{8} & \mathbf{3} & 1 & 0 & 1\end{bmatrix}$$

$$\textbf{pivot} = \mathbf{-8} \longrightarrow \begin{bmatrix}\mathbf{2} & \mathbf{1} & \mathbf{1} & 1 & 0 & 0\\ \mathbf{0} & \mathbf{-8} & \mathbf{-2} & -2 & 1 & 0\\ \mathbf{0} & \mathbf{0} & \mathbf{1} & -1 & 1 & 1\end{bmatrix} = \begin{bmatrix}\mathbf{U} & L^{-1}\end{bmatrix}$$

$$\begin{bmatrix}\mathbf{U} & L^{-1}\end{bmatrix} \longrightarrow \begin{bmatrix}\mathbf{2} & \mathbf{1} & \mathbf{0} & 2 & -1 & -1\\ \mathbf{0} & \mathbf{-8} & \mathbf{0} & -4 & 3 & 2\\ \mathbf{0} & \mathbf{0} & \mathbf{1} & -1 & 1 & 1\end{bmatrix}$$

$$\textbf{ceros arriba de los pivotes} \longrightarrow \begin{bmatrix}\mathbf{2} & \mathbf{0} & \mathbf{0} & \frac{12}{8} & -\frac{5}{8} & -\frac{6}{8}\\ \mathbf{0} & \mathbf{-8} & \mathbf{0} & -4 & 3 & 2\\ \mathbf{0} & \mathbf{0} & \mathbf{1} & -1 & 1 & 1\end{bmatrix}$$

$$\textbf{dividir por los pivotes} \longrightarrow \begin{bmatrix}\mathbf{1} & \mathbf{0} & \mathbf{0} & \frac{12}{16} & -\frac{5}{16} & -\frac{6}{16}\\ \mathbf{0} & \mathbf{1} & \mathbf{0} & \frac{4}{8} & -\frac{3}{8} & -\frac{2}{8}\\ \mathbf{0} & \mathbf{0} & \mathbf{1} & -1 & 1 & 1\end{bmatrix} = \begin{bmatrix}\mathbf{I} & A^{-1}\end{bmatrix} \quad \textit{(Fin del ejemplo)}$$

Observación 2. Aunque aparentemente solo se halló una inversa por derecha ($AX = I$), X también verifica $XA = I$, pues, nombrando M a la gran sucesión de operaciones elementales, de permutación y escalamiento realizadas sobre las filas de A en Gauss y Gauss-Jordan, tenemos que $MA = I$ y $MI = X$, de donde se deduce que $M = X$ y $XA = I$. Luego, $X = A^{-1}$.

3.2. Matrices transpuestas y simétricas

Definición 3. Dada una matriz A $m \times n$, la **matriz transpuesta** A^T es la matriz $n \times m$ tal que $A^T_{ij} = A_{ji}$ para todo $i = 1, \dots n$ y todo $j = 1, \dots, m$. Equivalentemente, $(A^T)_i = A^i$ para todo $i = 1, \dots n$ y $(A^T)^j = A_j$ para todo $j = 1, \dots, m$.

Definición 4. Una matriz A es **simétrica** si $A^T = A$.

Notar que si una matriz A es simétrica, entonces es cuadrada. Además, equivalentemente a la definición, A de tamaño $n \times n$ es simétrica si $A_i = A^i$ para todo $i = 1, \dots, n$.

Por otro lado, teniendo en cuenta las definiciones de matrices transpuestas y simétricas, una matriz

A posee las siguientes propiedades:

- $(A^T)^T = A$
- $(AB)^T = B^T A^T$ (*Demostración.* $(AB)^T_{ij} = (AB)_{ji} = \sum_{k=1}^{n} A_{jk}B_{ki} = \sum_{k=1}^{n} (B^T)_{ik}(A^T)_{kj} = (B^T A^T)_{ij}$)
- AA^T y $A^T A$ son simétricas.

Lema 4. *Si A es inversible, A^T también lo es y $(A^T)^{-1} = (A^{-1})^T$.*

Demostración. $A^T(A^{-1})^T = (A^{-1}A)^T = I^T = I$, y $(A^{-1})^T A^T = (AA^{-1})^T = I^T = I$. □

Lema 5. *Si A es simétrica e inversible, entonces A^{-1} es simétrica.*

Demostración. $(A^{-1})^T = (A^T)^{-1} = A^{-1}$. □

Lema 6. *Si A es simétrica y no singular, entonces admite una descomposición LDL^T, donde L es triangular inferior con 1's en la diagonal y D es diagonal sin ceros en la diagonal.*

Demostración. $A = LDV \implies A^T = V^T D^T L^T = A$. Luego, por unicidad de la descomposición LDV, $V = L^T$. □

Observación 3. Si A es simétrica, la eliminación Gaussiana de puede hacer en $\frac{n^3}{6}$ en vez de $\frac{n^3}{3}$ operaciones.

3.3. Más propiedades

- Si D_1 y D_2 son matrices diagonales, entonces $D_2D_1 = D_1D_2$.
- Si D es una matriz diagonal con entradas no nulas $d_1, \ldots, d_n$, entonces D^{-1} es diagonal con entradas $\frac{1}{d_1}, \ldots, \frac{1}{d_n}$.

  Demostración. Como D ya se encuentra en su forma escalonada, si consideramos la matriz ampliada con la identidad al lado de D y aplicamos Gauss-Jordan, el único paso restante es hacer unos en los pivots. Así, dividiendo la fila de cada pivot por su recíproco, la matriz identidad pasa a ser la inversa de D, la cual es una matriz diagonal cuyas entradas son $\frac{1}{d_1}, \ldots, \frac{1}{d_n}$. □

- $E_{ij}(\ell)E_{kj}(\ell) = E_{kj}(\ell)E_{ij}(\ell)$.
- $E_{ij}(\ell)^{-1} = E_{ij}(-\ell)$.
- Si P_{ij} es una matriz de permutación simple, entonces:
  - $(P_{ij})^{-1} = (P_{ij})^T = P_{ij}$.
  - El determinante de P_{ij} es ± 1.
- Si A y B son matrices triangulares, entonces:

- A es invertible si y solo si ningún elemento de la diagonal es nulo.
- Si A y B son triangulares inferiores, AB es triangular inferior, y si A es invertible, entonces A^{-1} es triangular inferior.

Unidad II

Espacios Vectoriales

1. Interpretación Geométrica de Sistemas Lineales

Dado el siguiente sistema de ecuaciones $A\bar{x} = b$: $\begin{cases} x + 2y = 3 \\ 4x + 5y = 6 \end{cases}$ en donde

$$A = \begin{bmatrix} 1 & 2 \\ 4 & 5 \end{bmatrix}, \ \bar{x} = \begin{bmatrix} x \\ y \end{bmatrix} \text{ y } b = \begin{bmatrix} 3 \\ 6 \end{bmatrix},$$

éste se puede interpretar de dos formas: por filas o por columnas. En el caso de la geometría por filas, se tienen dos rectas (r_1) $x + 2y = 3$ y (r_2) $4x + 5y = 6$, y la solución del sistema representa la intersección de las rectas en $\mathbb{R}^2$ (en caso de haberla). Esto en dimensiones más grandes se extiende a n planos $(n-1)-$dimensionales en un espacio $n-$dimensional.

Por otro lado, mediante la geometría por columnas el sistema se puede representar como

$$\begin{bmatrix} 1 \\ 4 \end{bmatrix} x + \begin{bmatrix} 2 \\ 5 \end{bmatrix} y = \begin{bmatrix} 3 \\ 6 \end{bmatrix}.$$

Es decir, se busca una combinación lineal de los vectores $u = (1, 4)$ y $v = (2, 5)$ que dé el vector $b = (3, 6)$.

1.1. Casos Singulares

En general, un sistema es singular si y solo si los n planos no tienen puntos en común o tienen infinitos puntos en común. Equivalentemente, esto sucede si los n vectores columna subyacen en un mismo plano $(n-1)-$dimensional o si los n vectores columna no son linealmente independientes.

2. Espacios Vectoriales

Definición 5. $(V, +, \cdot)$ es un **espacio vectorial sobre** $\mathbb{K}$ si para todo $u, v, w \in V$ y todo $\alpha, \beta \in \mathbb{K}$ se verifica:

1. $u + v \in V$ (suma cerrada en V),

2. $u + (v + w) = (u + v) + w$ y $u + v = v + u$ (suma asociativa y conmutativa),

3. $\exists 0 \in V / v + 0 = v$ (elem. neutro de la suma),

4. $\exists v^* \in V / v + v^* = 0$ (elem. opuesto para la suma),

5. $\alpha \cdot V \in V$ (prod. por escalar cerrado en V),

6. Si $1 \in \mathbb{K}$ es el elem. neutro del producto en $\mathbb{K}$, entonces $1 \cdot v = v$,

7. $(\alpha\beta) \cdot v = \alpha(\beta \cdot v)$,

8. $\alpha \cdot (u + v) = \alpha \cdot u + \alpha \cdot v$,

9. $(\alpha + \beta) \cdot v = \alpha \cdot v + \beta \cdot v$.

Ejemplos de espacios vectoriales:

- $\mathbb{R}^n$ para todo $n \in \mathbb{N}_0$
- $\mathbb{R}^\infty$ (sucesiones reales)
- $\mathbb{R}^{m \times n}$ (matrices reales $m \times n$)
- Funciones reales definidas en $[a, b]$
- Funciones reales continuas
- Funciones derivables en x_0
- $\{$polinomios a coeficientes reales$\} \cup \{$polinomio nulo$\}$
- $\{$polinomios de grado a lo sumo $n\} \cup \{$polinomio nulo$\}$

2.1. Propiedades

- Unicidad del elemento neutro

  Demostración. Sea $0' \in V$ tal que $0' + x = x$ para todo $x \in V$. Luego, teniendo en cuenta que $0 + x = x$ para todo $x \in V$, en particular, $0 + 0' = 0$ y $0' + 0 = 0$. Por conmutatividad de la suma, $0 = 0 + 0' = 0' + 0 = 0'$. □

- Unicidad del opuesto

  Demostración. Dado $x \in V$, sea $\bar{x} \in V$ tal que $x + \bar{x} = \bar{x} + x = 0$. Considerando que existe un x^*tal que $x + x^* = 0$, tenemos que, por asociatividad de la suma, $\bar{x} = \bar{x} + (x + x^*) = (\bar{x} + x) + x^* = 0 + x^* = x^*$. □

- $0 \cdot v = 0$ para todo $v \in V$

  Demostración. Sea $v \in V$ y $w = 0 \cdot v$. Por las propiedades 6 y 9, $w + v = 0 \cdot v + v = 0 \cdot v + 1 \cdot v = (0 + 1) \cdot v = 1 \cdot v = v$. Es decir, w resulta el elemento neutro de la suma, y por la unicidad del mismo, $w = 0 \cdot v = 0$. □

- $(-1) \cdot x$ es el opuesto de x

  Demostración. Dado $x \in V$, existe $-x$ tal que $x+(-x) = 0$. Por las propiedades 6 y 9, $x+(-1)\cdot x = 1 \cdot x + (-1) \cdot x = (1 + (-1)) \cdot x = 0 \cdot x = 0$. Luego $(-1) \cdot x = -x$. □

- $\alpha \cdot 0 = 0 \ \ \forall \alpha \in \mathbb{K}$

  Demostración. Dado $x \in V$ y $\alpha \in \mathbb{K}$, $\alpha \cdot 0 = \alpha \cdot (x + (-x)) = \alpha \cdot x + \alpha \cdot (-x) = \alpha \cdot x + ((-1)\alpha) \cdot x = \alpha \cdot x + (-\alpha) \cdot x = (\alpha + (-\alpha)) \cdot x = 0 \cdot x = 0$. □

- $\alpha \cdot v = 0 \implies \alpha = 0 \vee v = 0$

  Demostración. Si $\alpha = 0$ vale por propiedad anterior. Sea $\alpha \neq 0$. Luego, $\alpha \cdot v = 0 \implies \alpha^{-1} \cdot (\alpha \cdot v) = \alpha^{-1} \cdot 0 = 0 \implies (\alpha^{-1}\alpha) \cdot v = 1 \cdot v = v = 0$. □

- $z + x = z + y \implies x = y$ (Propiedad cancelativa de la suma)

  Demostración. Dados $x, y, x \in V$, $z+x = z+y \implies z+x+(-z) = z+y+(-z) \implies (z+(-z))+x = (z + (-z)) + y \implies 0 + x = 0 + y \implies x = y$. □

2.2. Subespacios Vectoriales

Definición 6. Sea $(V, +, \cdot)$ un espacio vectorial y $U \subset V$. Entonces, U es un **subespacio vectorial** de V si $(U, +, \cdot)$ es un espacio vectorial.

Lema 7. *Sea $(V, +, \cdot)$ un espacio vectorial y $U \subset V$. Entonces, U es un subespacio vectorial de V si y solo si toda combinación lineal de elementos de U pertenece a U; es decir, $\forall u_1, u_2 \in U,\ \alpha, \beta \in \mathbb{K}$, resulta $\alpha u_1 + \beta u_2 \in U$.*

Demostración. **Ida.** Por hipótesis, U es un subespacio vectorial de V. Entonces, se verifica la clausura de la suma y el producto, por lo que dados $\forall u_1, u_2 \in U,\ \alpha, \beta \in \mathbb{K}$, resulta $\alpha u_1 + \beta u_2 \in U$.

Vuelta. Por hipótesis, toda combinación lineal de elementos de U pertenece a U. Entonces, dado $u \in V$, $(-1) \cdot u = -u \in V$, y consecuentemente $u + (-u) = 0 \in V$. El resto de las condiciones se heredan de V. □

Definición 7. Dado un espacio vectorial V y un subconjunto de vectores $U \subset V$, llamamos **subespacio generado por** U y lo notamos $\langle U \rangle$ al subespacio determinado por todas las combinaciones lineales de elementos de U.

Observación 4. Por el lema anterior, sabemos que $\langle U \rangle$ es efectivamente un subespacio vectorial de U.

Definición 8. Sea V un espacio vectorial sobre $\mathbb{K}$, y U y W subespacios de V. Se define como **subespacio suma** al subespacio

$$U + W = \{u + w : u \in U, w \in W\}.$$

Definición 9. Sea V un espacio vectorial sobre $\mathbb{K}$, y U y W subespacios de V. Entonces, si para todo $v \in V$ existen únicos $u \in U$, $w \in W$ tal que $v = u + w$, se dice que V es **suma directa** de U y W, y se nota

$$V = U \oplus W.$$

Teorema 3. $V = U \oplus W \iff V = U + W \wedge V \cap W = \{0\}$.

Demostración. **Ida.** Por hipótesis, $V = U_1 \oplus U_2$. Por definición de suma directa, $V = U_1 + U_2$. Resta ver que $U_1 \cap U_2 = \{0\}$.

Sea $v \in U_1 \cap U_2$. Luego, v puede descomponerse como $v = v+0$, con $v \in U_1$ y $0 \in U_2$, y como $v = 0+v$, con $0 \in U_1$ y $v \in U_2$. Entonces, como $V = U_1 \oplus U_2$, por unicidad de la descomposición $(v + 0 = 0 + v)$, tenemos que $v = 0$.

Vuelta. Sea $v \in V$. Como $V = U + W$, existen $u_1 \in U_1, u_2 \in U_2$ tales que $v = u_1 + u_2$. Ahora, supongo que existen $\bar{u}_1, \bar{u}_2$ tales que $v = \bar{u}_1 + \bar{u}_2$. Entonces,

$$v = u_1 + u_2 = \bar{u}_1 + \bar{u}_2 \implies u_1 + u_2 = \bar{u}_1 + \bar{u}_2 \implies u_1 - \bar{u}_1 = \bar{u}_2 - u_2.$$

Ahora, como $u_1 - \bar{u}_1 \in U_1$ y $\bar{u}_2 - u_2 \in U_2$ (pues son subespacios), se tiene que $u_1 - \bar{u}_1, \bar{u}_2 - u_2 \in U_1 \cap U_2$, y, por hipótesis, $u_1 - \bar{u}_1 = \bar{u}_2 - u_2 = 0$, por lo que $u_1 = \bar{u}_1$ y $\bar{u}_2 = u_2$.

Por lo tanto $v \in V$ se puede escribir de manera única como $v = u_1 + u_2$, con $u_1 \in U_1$ y $u_2 \in U_2$. Luego, $V = U_1 \oplus U_2$. □

Lema 8. *Dado un espacio vectorial V:*

1. $S \subseteq T \subseteq V \implies \langle S \rangle \subseteq \langle T \rangle$.
2. $S \subseteq \langle S \rangle$.
3. Si $S \subseteq T$ y T es un subespacio de V, entonces $\langle S \rangle \subseteq T$. A partir de esta propiedad se deduce que $\langle S \rangle$ es el menor subespacio de V que contiene a S.
4. S es un subespacio de V si y solo si $\langle S \rangle = S$.
5. $\langle S \rangle = U \implies \langle U \rangle = U$.

6. Si $W \subseteq V$, entonces $\langle S \cap W \rangle \subset \langle S \rangle \cap \langle W \rangle$ y $\langle S \cup W \rangle \subset \langle S \rangle + \langle W \rangle$.

Demostración.

1. Sea $v \in \langle S \rangle$. Luego, v es una combinación lineal de los vectores de S. Como $S \subseteq T$, los vectores de S también son vectores de T, por lo que v también es una combinación lineal de vectores de T (si $S \subset T$ estrictamente, los escalares correspondientes a vectores que no están en S son nulos). Luego, $v \in \langle T \rangle$.

2. Sea $s \in S$. Luego, s puede expresarse como combinación lineal de los vectores de S asociando el elemento neutro del producto por escalar (1) al vector s y el elemento neutro de la suma (0) a los otros vectores. Entonces, $s \in \langle S \rangle$.

3. Por hipótesis, $S \subseteq T$ y T es un subespacio de V. Sea $s \in \langle S \rangle$. Luego, como s es combinación lineal de los vectores de S y $S \subseteq T$,

$$\exists\ I \subseteq \mathbb{N}\ /\ x = \sum_{i \in I} \alpha_i s_i,\ \alpha_i \in \mathbb{R}(\text{o } \mathbb{C}),\ s_i \in S \implies \exists\ I \subseteq \mathbb{N}\ /\ x = \sum_{i \in I} \alpha_i s_i,\ \alpha_i \in \mathbb{R}(\text{o } \mathbb{C}),\ s_i \in T.$$

   Es decir, como x es combinación lineal de vectores de T y T es un subespacio de V, tenemos que $x \in T$.

4. La vuelta es trivial. En cuanto a la ida, considerando la doble contención entre $\langle S \rangle$ y S, por **2.** tenemos que $S \subseteq \langle S \rangle$. Por otro lado, se probó antes que S es un subespacio de V si y solo si cualquier combinación lineal de sus elementos pertenece a S. Entonces, dado $x \in \langle S \rangle$ (una combinación lineal de los vectores de S), se tiene que $x \in S$. Luego, $S = \langle S \rangle$.

5. $\langle U \rangle$ está conformado por todas las combinaciones lineales de vectores de U. Sin embargo, como U es el subespacio generado por S, cualquier combinación lineal de vectores de U pertenece a U. Entonces, $\langle U \rangle = U$.

6. Primero pruebo que $\langle S \cap W \rangle \subset \langle S \rangle \cap \langle W \rangle$ y luego que $\langle S \cup W \rangle \subset \langle S \rangle + \langle W \rangle$:

   Sea $x \in \langle S \cap W \rangle$. Luego, x es combinación lineal de elementos de $S \cap W$, por lo que también es combinación lineal de elementos de $\langle S \rangle$ y de $\langle W \rangle$. Luego, $x \in \langle S \rangle$ y $x \in \langle W \rangle$, entonces $x \in \langle S \rangle \cap \langle W \rangle$.

   Por otro lado, considerando $x \in \langle S \cup W \rangle$, x es combinación lineal de vectores de S o de W. Sea $S' = S - W \subset S$. Luego,

$$\exists\ I_1, I_2 \subseteq \mathbb{N}\ /\ x = \sum_{i \in I_1} \alpha_i s^i + \sum_{i \in I_2} \beta_i w^i,\ \alpha_i, \beta_i \in \mathbb{R}(\text{o } \mathbb{C}),\ s_i \in S',\ w^i \in W.$$

   Es decir, se pudo expresar a x como la suma de $\sum_{i \in I_1} \alpha_i s^i \in S$ y $\sum_{i \in I_2} \beta_i w^i$, por lo que $x \in S + W$.

□

3. Espacio Columna y Espacio Nulo

Definición 10. Dada una matriz A $m \times n$, se define:

- El **espacio columna** de A: es el subespacio de $\mathbb{R}^m$ generado por los vectores columna de A. Se nota $C(A)$.
- El **espacio nulo** de A: es el subespacio de $\mathbb{R}^n$ definido como $N(A) = \{x \in \mathbb{R}^n : Ax = 0\}$.

Dada una matriz A $m \times n$ y un vector $b \in \mathbb{R}^m$, $b \in C(A)$ si existe una combinación lineal de las columnas de A que den b. Es decir, si existen $x_1, \ldots, x_n \in \mathbb{R}$ tales que $\sum_{i=1}^n A^i x_i = b$. Equivalentemente, si existe $x \in \mathbb{R}^n$ tal que $Ax = b$ (si existe solución del sistema).

Lema 9. *Si A es una matriz (cuadrada) no singular $n \times n$, entonces $C(A) = \mathbb{R}^n$ y $N(A) = \{0\}$. (También vale la recíproca)*

Demostración. Si A es no singular, sabemos que el Método de Eliminación Gaussiana encuentra siempre una solución al sistema $Ax = b$ para cualquier $b \in \mathbb{R}^n$. Por otro lado, A es no singular si y solo si el sistema $Ax = 0$ admite solo la solución $x = 0$, por lo que $N(A) = \{0\}$. □

Observación 5. Si A es la matriz nula $m \times n$, $C(A) = \{0\} \subset \mathbb{R}^m$ y $N(A) = \mathbb{R}^n$. En general, si A es una matriz $m \times n$, $C(A)$ es un subespacio de $\mathbb{R}^m$, y puede ser de cualquier dimensión entre 0 y m.

Lema 10. *Dada una matriz A $m \times n$, sea A' la matriz que se obtiene de agregar una columna A^{n+1} a A, donde A^{n+1} es una combinación lineal de las columnas de A. Luego, $C(A) = C(A')$.*

Demostración. Como A^{n+1} es combinación lineal de las columnas de A, existen escalares $\beta_i,\ i = 1, \ldots, n$ tales que $A^{n+1} = \sum\limits_{i=1}^{n} \beta_i A^i$. Luego,

$$\begin{aligned}
C(A') &= \left\{ \sum_{i=1}^{n+1} \alpha_i (A')^i : \alpha_i \in \mathbb{R}, i = 1, \ldots, n+1 \right\} \\
&= \left\{ \sum_{i=1}^{n} \alpha_i A^i + \alpha_{n+1} A^{n+1} : \alpha_i \in \mathbb{R}, i = 1, \ldots, n+1 \right\} \\
&= \left\{ \sum_{i=1}^{n} \alpha_i A^i + \alpha_{n+1} \sum_{i=1}^{n} \beta_i A^i : \alpha_i, \beta_i \in \mathbb{R}, i = 1, \ldots, n;\ \alpha_{n+1} \in \mathbb{R} \right\} \\
&= \left\{ \sum_{i=1}^{n} \alpha_i A^i + \sum_{i=1}^{n} \alpha_{n+1} \beta_i A^i : \alpha_i, \beta_i \in \mathbb{R}, i = 1, \ldots, n;\ \alpha_{n+1} \in \mathbb{R} \right\} \\
&= \left\{ \sum_{i=1}^{n} \alpha_i \alpha_{n+1} \beta_i A^i : \alpha_i, \beta_i \in \mathbb{R}, i = 1, \ldots, n;\ \alpha_{n+1} \in \mathbb{R} \right\} \\
&= \left\{ \sum_{i=1}^{n} \gamma_i A^i : \gamma_i \in \mathbb{R}, i = 1, \ldots, n \right\} = C(A)
\end{aligned}$$

□

Lema 11. *Sean A y B matrices tales que $AB = 0$. Luego, $C(B) \subseteq N(A)$.*

Demostración. Considero A $m \times n$, B $n \times p$.

Como $AB = 0$, en particular $(AB)^i = AB^i = 0\ \forall i/1 \leq i \leq p$. Luego, $B^i \in N(A)$ para todo $i = 1, \ldots, p$, por lo que cualquier combinación lineal de ellas pertenece a $N(A)$. Por lo tanto, $C(B) \subseteq N(A)$. □

4. Soluciones de sistemas de ecuaciones lineales

Teorema 4. *Sea A una matriz $m \times n$ y $b \in \mathbb{R}^m$. Sea $\mathcal{S}$ el conjunto de soluciones del sistema $Ax = b$. Entonces, $\mathcal{S} = \emptyset \iff b \notin C(A)$, y si $b \in C(A)$ y $x_P \in \mathcal{S}$, entonces $\mathcal{S} = \{x_P + x_N : x_N \in N(A)\}$.*

Demostración. $\mathcal{S} = \emptyset \iff b \notin C(A)$ se deduce de la definición de $C(A)$.

Pruebo la doble contención para llegar a que $\mathcal{S} = \{x_P + x_N : x_N \in N(A)\}$.

Sea $x_N \in N(A)$. Se debe probar que $x = x_P + x_N \in \mathcal{S}$. Efectivamente, $Ax = A(x_P + x_N) = Ax_P + Ax_N = b + 0 = b$.

$\therefore x \in \mathcal{S}$.

Ahora, sea $x \in \mathcal{S}$. Defino $x_N = x - x_P$. Debo probar que $x_N \in N(A)$. Efectivamente, $Ax_N = A(x - x_P) = Ax - Ax_P = b - b = 0$.

$\therefore x_N \in \mathcal{S}$ y $x = x_P + x_N$. □

Entonces, para describir el conjunto de soluciones de un sistema $Ax = b$, se debe:

1. describir $N(A)$, y

2. conocer una solución particular x_P del sistema $Ax = b$, para algún $b \in C(A)$ (o decidir que no existe tal solución).

Definición 11. Una matriz A no nula $m \times n$ es una **matriz escalonada** si verifica:

1. Si existe $i \in \{1, \ldots, m\}$ tal que $A_i = 0$, entonces $A_j = 0$ para todo j tal que $i < j \leq m$ (las final nulas están al final).

2. Sea $m^* = \min\{i \in \{1, \ldots, m-1\} : A_{i+1} = 0\}$ si A tiene alguna fila nula, y $m^* = m$ en caso contrario. Para todo $i = 1, \ldots, m^*$, existe $j(i) \in \{1, \ldots, n\}$ tal que:

   a) $j(i) < j(i+1)$ para todo $i = 1, \ldots, m^* - 1$,

   b) $A_{ij(i)} \neq 0$,

   c) $A_{ik} = 0$ para todo $k < j(i)$ y $A_{kj(i)} = 0$ para todo $k > i$.

Para $i = 1, \ldots, m^*$, a las entradas (no nulas) $A_{ij(i)}$ se las denomina **pivots** de A. A las filas A_i **filas pivots**, y a las columnas $A^{j(i)}$, **columnas pivots**.

Observación 6. A través de matrices elementales y de permutación, se puede transformar a toda matriz $m \times n$ en una matriz escalonada.

Teorema 5. *Para cualquier matriz A $m \times n$, existe una matriz de permutación P, una matriz triangular inferior L con 1's en la diagonal y una matriz escalonada U tal que $PA = LU$.*

Demostración. Si la primera columna de A es nula, pasamos a la segunda columna. Si no, por medio de matrices de permutación podemos asegurarnos que haya un pivot no nulo en este paso. Luego, aplicando una sucesión de matrices elementales podemos hacer ceros debajo de dicho pivot. Repitiendo estos pasos, llegamos a una matriz escalonada U (con ceros debajo de todos los pivots). En símbolos, considerando k la cantidad de pivots no nulos que aparecen en el proceso de eliminación, L_i el producto de matrices elementales aplicadas a A, y P_i la matriz de permutación aplicada a A, tenemos que

$$L_k P_k \dots L_2 P_2 L_1 P_1 A = U \iff (L_k \dots L_2 L_1)(P_k \dots P_2 P_1)A = U \iff PA = LU,$$

con $P = P_k \dots P_2 P_1$ y $L = (L_k \dots L_2 L_1)^{-1}$. Todo esto es posible por la asociatividad del producto de las matrices y la inversibilidad de las matrices elementales (y por ende su producto). Se pueden agrupar las P_i entre sí debido a que permutar las filas antes de comenzar con la eliminación es lo mismo que hacerlo en cada paso. Nótese que como las matrices elementales fueron utilizadas para eliminar elementos debajo de los pivots, todos los elementos no nulos que no están en la diagonal se encuentran debajo de la misma, por lo que es triangular inferior. Además, se conservan los unos en la diagonal. Entonces, se encontraron una matriz de permutación P y una matriz triangular inferior L con unos en la diagonal tales que $PA = LU$, con U escalonada. □

Definición 12. Una matriz R es **escalonada reducida** si R es escalonada con todos los pivots iguales a 1 y todas las entradas por encima de los pivots son nulas.

Observación 7. Si R es la matriz escalonada reducida que se obtiene a partir de A con permutaciones y operaciones elementales, entonces $N(A) = N(R)$.

4.1. Describiendo $N(A)$

Considero el sistema

$$Rx = \begin{bmatrix} 1 & 3 & 0 & -1 & -2 \\ 0 & 0 & 1 & 1 & \frac{4}{3} \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = 0.$$

Como las filas nulas corresponden a ecuaciones de tipo $0x = 0$, se pueden descartar en la resolución del sistema. Por otro lado, es claro que las variables pivots son x_1 y x_3, y las libres x_2, x_4 y x_5. Asignando valores arbitrarios a las variables libres, por ejemplo, $x_2 = t$, $x_4 = s$ y $x_5 = w$, se tiene que las variables pivots satisfacen

$$\begin{aligned} x_1 &= -\,3t + s + 2w \\ x_3 &= -\,s - \frac{4}{3}w \end{aligned}$$

Es decir, las soluciones del sistema del sistema $Rx = 0$ son de la forma

$$x = \begin{bmatrix} -3t+s+2w \\ t \\ -s-\frac{4}{3}w \\ s \\ w \end{bmatrix} = t\begin{bmatrix} -3 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + s\begin{bmatrix} 1 \\ 0 \\ -1 \\ 1 \\ 0 \end{bmatrix} + w\begin{bmatrix} 2 \\ 0 \\ -\frac{4}{3} \\ 0 \\ 1 \end{bmatrix}, \text{ con } t, s, w \in \mathbb{R}.$$

Por lo tanto, $N(A) = N(R)$ es el espacio generado por los vectores que multiplican a las variables libres, denominados **soluciones especiales** del sistema $Rx = 0$. Cada una de ellas se obtiene fijando una variable libre en 1 y las restantes en 0. Naturalmente, hay tantas soluciones especiales como variables libres.

Observación 8. Si llamamos M a la matriz cuyas columnas son las soluciones especiales de $Rx = 0$, tenemos que $N(R) = C(M)$ y los valores de las variables pivots en las soluciones especiales corresponden a los valores opuestos de las entradas de las filas pivots en las columnas de las variables libres (esto proviene de obtener el valor de cada variable pivot restando los valores libres en función de los cuales se encuentra).

En resumen, para describir $N(A)$, se siguen los siguientes pasos:

1. Encontrar R e identificar variables libres y pivots,
2. Encontrar las soluciones especiales,
3. $N(A)$ es el espacio generado por las soluciones especiales encontradas.

Observación 9. Dada una matriz A $m \times n$, con $m < n$, entonces el sistema $Ax = 0$ siempre admite soluciones distintas de la trivial ($N(A) \neq 0$).

Para cualquier matriz, $N(A)$ es un subespacio vectorial de $\mathbb{R}^n$ cuya dimensión coincide con el número de variables libres. Por otro lado, $C(A)$ es el subespacio vectorial de $\mathbb{R}^m$ cuya dimensión coincide con el número de variables pivots (y es el subespacio generado por las columnas pivots).

Definición 13. El **rango** de una matriz A es el número de variables pivots que posee.

Observación 10. El rango r de una matriz es el número de filas no nulas que posee su forma escalonada.

4.2. Resolviendo el sistema $Ax = b$

Una vez encontrado $N(A)$, el conjunto $\mathcal{S}$ de soluciones del sistema $Ax = b$ se puede escribir a partir de una solución particular del mismo más cualquier elemento de $N(A)$. Esta solución puede encontrarse aplicando a b las mismas operaciones aplicadas para llevar a A a su forma escalonada U o a su forma reducida R. Luego, si aparecen filas nulas en U, queda restringido $C(A)$ a los vectores b que verifiquen que las ecuaciones resultantes de aplicar las operaciones sean iguales a cero.

Ejemplo:

$$Ax = \begin{bmatrix} 1 & 3 & 3 & 2 & 2 \\ 2 & 6 & 6 & 4 & 4 \\ 2 & 6 & 9 & 7 & 8 \\ -1 & -3 & 3 & 4 & 6 \end{bmatrix} x = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} = b \iff Ux = \begin{bmatrix} 1 & 3 & 3 & 2 & 2 \\ 0 & 0 & 6 & 6 & 8 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} x = \begin{bmatrix} b_1 \\ b_1 + b_4 \\ b_3 - \frac{5}{2}b_1 - \frac{1}{2}b_4 \\ b_2 - 2b_1 \end{bmatrix} = b'$$

Luego, $b \in C(A)$ si y solo si $b_3 - \frac{5}{2}b_1 - \frac{1}{2}b_4 = 0$ y $b_2 - 2b_1 = 0$. Si eso se satisface, solo queda resolver

$$\begin{aligned} x_1 + 3x_2 + 3x_3 + 2x_4 + 2x_5 &= b_1 \\ 6x_3 + 6x_4 + 8x_5 &= b_1 + b_4 \end{aligned}$$

y asignando el valor cero a todas las variables libres, llegamos al siguiente sistema triangular sin ceros en la diagonal, que puede resolverse por sustitución para atrás.

$$\begin{aligned} x_1 + 3x_3 &= b_1 \\ 6x_3 &= b_1 + b_4 \end{aligned}$$

Resolviendo con $b = (0, 0, 1, 2) \in C(A)$, se tiene que $x_P = (-1, 0, \frac{1}{3}, 0, 0)$, y el conjunto solución $\mathcal{S}$ del sistema $Ax = b$ queda dado por los $x \in \mathbb{R}^5$ tales que

$$x = \begin{bmatrix} -1 \\ 0 \\ \frac{1}{3} \\ 0 \\ 0 \end{bmatrix} + t \begin{bmatrix} -3 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + s \begin{bmatrix} 1 \\ 0 \\ -1 \\ 1 \\ 0 \end{bmatrix} + w \begin{bmatrix} 2 \\ 0 \\ -\frac{4}{3} \\ 0 \\ 1 \end{bmatrix}, \text{ con } t, s, w \in \mathbb{R}. \quad \textit{(Fin del ejemplo)}$$

Lema 12. *Sea A una matriz 5×4 de rango 4 y $b \in \mathbb{R}^5$. Entonces, el sistema $Ax = b$ no tiene soluciones si y solo si la matriz 5×5 $[A\ b]$ es invertible.*

Demostración. Como A es 5×4 de rango 4, al escalonarla se llega a una matriz cuya última fila es nula. Sea E la matriz producto de todas las matrices de operaciones elementales aplicadas a A para escalonarla. Luego, existe solución al sistema $Ax = b$ (es decir, $b \in C(A)$) si y solo si $(Eb)_5 = 0$. Entonces, considerando la matriz 5×5 $A' = [A\ b]$, tenemos que $b \in C(A)$ si y solo si la quinta fila de EA' es nula. Entonces, si no lo es, no existe solución al sistema $Ax = b$, y la A' tiene rango 5 (rango completo), por lo que es no singular y por ende invertible. □

5. Independencia Lineal

Sea $(V, \oplus, \odot)$ un espacio vectorial sobre el cuerpo $\mathbb{K}$ y U un conjunto finito de vectores de V, $U = \{v^i : i = 1, \ldots, k\} \subseteq V$.

Si un vector v^i de U puede expresarse como combinación lineal de los restantes, se dice que los vectores de U son linealmente dependientes.

Observación 11. Decir que un vector no nulo $v \in U$ se puede expresar como combinación lineal de los otros vectores de U es equivalente a decir que el vector nulo $0 \in V$ puede expresarse una combinación lineal no nula de los vectores de U.

Definición 14. Sea $(V, \oplus, \odot)$ un espacio vectorial sobre $\mathbb{K}$ y $U = \{v^i : i = 1, \ldots, k\} \subseteq V$. Luego, los vectores de U son **linealmente independientes** (**l.i.**) si

$$\sum_{i=1}^{k} \alpha_i v^i = 0 \implies \alpha_i = 0,\ i = 1, \ldots, k.$$

De esto deriva, por ejemplo, que para todo $U \subset V$, si $0 \in V$, entonces los vectores de U son linealmente dependientes.

Dado un conjunto de vectores $U = \{v^i : i = 1, \ldots, k\} \subset \mathbb{R}^n$, se define como $M(U)$ a la matriz $n \times k$ cuya columna j es v^j, $j = 1, \ldots, k$.

Lema 13. *Sea $U = \{v^j : j = 1, \ldots, k\} \subset \mathbb{R}^n$. Entonces, los vectores de U son l.i. si y solo si $N(M(U)) = \{0\}$.*

Demostración. Teniendo en cuenta que toda combinación lineal de las columnas de una matriz A $m \times n$ puede expresarse como Ax, con $x \in \mathbb{R}^n$, entonces los vectores de U son l.i. si y solo si $M(U)x = 0 \implies x = 0$, o, equivalentemente, el único elemento de $N(M(U))$ es 0. □

Corolario 1. *Sea $U = \{v^j : j = 1, \ldots, n\} \subset \mathbb{R}^n$. Entonces, los vectores de U son l.i. si y solo si $M(U)$ es no singular.*

Demostración. **Ida.** Se probó que $N(M(U)) = \{0\}$. Es decir, el sistema homogéneo asociado a $M(U)$ solo admite la solución trivial, y por lo tanto no hay variables libres presentes al reducir la matriz. Entonces, todas las columnas son pivots, y la matriz escalonada asociada a $M(U)$ no posee elementos nulos en la diagonal, por lo que dado $b \in \mathbb{R}^n$, el sistema es compatible, y luego $C(M(U)) = \mathbb{R}^n$. Entonces, por lema ya probado, como $C(M(U)) = \mathbb{R}^n$ y $N(M(U)) = \{0\}$, $M(U)$ es no singular. □

Lema 14. *Si A es una matriz $n \times n$ no singular, sus vectores fila son l.i..*

Demostración. Si A es no singular, entonces A^T es no singular, donde las columnas de A^T son las filas de A, y $N(A^T) = \{0\}$. Luego, por el Lema 8, tenemos que las columnas de A^T (es decir, la filas de A) son l.i.. □

Corolario 2. *Los vectores columna y los vectores fila de toda matriz triangular $n \times n$ sin ceros en la diagonal son l.i..*

Demostración. Si una matriz A es triangular sin ceros en la diagonal, es no singular. Luego, la prueba es inmediata del Lema 8 y Lema 9. □

Corolario 3. *En todo conjunto de más de n vectores de $\mathbb{R}^n$, sus elementos son l.d..*

Demostración. Si se construye una matriz cuyas columnas son los $m > n$ vectores de el conjunto en cuestión, esta matriz $n \times m$ posee más columnas que filas, por lo que al reducirla aparece al menos una variable libre y luego su espacio nulo contiene elementos distintos de 0. □

5.1. Teorema Fundamental del Álgebra Lineal - Primera Parte

Lema 15. *Sean $\{v^1, \dots, v^k\} \subset \mathbb{R}^n$, $\{w^1, \dots, w^k\} \subset \mathbb{R}^p$ y $\{z^1, \dots, z^k\} \subset \mathbb{R}^{n+p}$ tales que, para todo $j = 1, \dots, k$, $z^j = \begin{bmatrix} v^j \\ w^j \end{bmatrix}$. Luego:*

1. *Si $v^1, \dots, v^k$ son vectores l.i., entonces $z^1, \dots, z^k$ son vectores l.i..*
2. *Si $w^j = 0$ para todo $j = 1, \dots, k$, entonces $z^1, \dots, z^k$ son vectores l.i. si y solo si $v^1, \dots, v^k$ son vectores l.i..*

Demostración. Observemos que

$$\sum_{j=1}^{k} \alpha_i z^j = \begin{bmatrix} \sum_{j=1}^{k} \alpha_i v^j \\ \sum_{j=1}^{k} \alpha_i w^j \end{bmatrix}.$$

1. Sean $\alpha^1, \dots, \alpha^k \in \mathbb{R}$ tales que $\sum_{j=1}^{k} \alpha_i z^j = 0 \in \mathbb{R}^{n+p}$. Entonces, $\sum_{j=1}^{k} \alpha_i v^j = 0 \in \mathbb{R}^n$, y como $v^1, \dots, v^k$ son l.i., se tiene que $\alpha_i = 0$ para todo $i = 1, \dots, k$. Luego, $z^1, \dots, z^k$ son vectores l.i..
2. Falta probar la ida. Sean $\alpha^1, \dots, \alpha^k \in \mathbb{R}$ tales que $\sum_{j=1}^{k} \alpha_i z^j = 0 \in \mathbb{R}^n$. Entonces,

$$\begin{bmatrix} \sum_{j=1}^{k} \alpha_i v^j \\ 0 \end{bmatrix} = \sum_{j=1}^{k} \alpha_i \begin{bmatrix} v^j \\ 0 \end{bmatrix} = \sum_{j=1}^{k} \alpha_i z^j = 0 \in \mathbb{R}^{n+p}.$$

Como $z^1, \dots, z^k$ son vectores l.i., tenemos $\alpha_j = 0$ para todo $j = 1, \dots, k$. Entonces, $v^1, \dots, v^k$ son vectores l.i..

□

Lema 16. *Si U es una matriz escalonada, el número máximo de filas l.i. coincide con el número máximo de columnas l.i., y ese número es r, el rango de U.*

Demostración. Considero a un conjunto de vectores que se corresponden con los vectores columna de U. Claramente, se pueden ordenar las columnas de manera que las columnas pivots estén todas al principio, y esto no cambia la lineal independencia de las columnas ni de las filas.

Por otro lado, si $\tilde{U}$ es la submatriz de U que se obtiene borrando sus filas nulas, el número de vectores fila l.i. y el número de vectores columna l.i. coinciden. Efectivamente, por el Lema anterior, un conjunto de columnas de U es l.i. si y solo si los vectores que se obtienen borrando las entradas correspondientes a las filas nulas de U lo son, y además, ningún subconjunto de vectores que incluya al vector nulo es l.i., por lo que ningún conjunto de filas l.i. de U contiene a una fila nula.

Entonces, como los vectores columnas de $\tilde{U}$ son vectores de $\mathbb{R}^r$, sabemos que a lo sumo se tendrán r vectores l.i. en $\tilde{U}$ (y por lo tanto en U), y esto se cumple ya que las r primera columnas de $\tilde{U}$ son l.i.. pues la submatriz correspondiente a ellas es triangular sin ceros en la diagonal.

Además, si los vectores fila de la submatriz correspondiente a las primeras r columnas de $\tilde{U}$ son l.i., por el mismo Lema también lo serán los vectores fila de $\tilde{U}$. Esto se cumple, pues $\tilde{U}$ es triangular sin ceros en la diagonal y por ende las columnas de su matriz transpuesta (también triangular) son l.i., por lo que sus filas son l.i.. Por lo tanto, el número máximo de vectores fila l.i. de $\tilde{U}$ es r, y lo mismo sucede para U. □

Lema 17. *Sean $U = \{v^j : j = 1, \dots, n\} \subset \mathbb{R}^n$, w^1 una combinación lineal de los vectores de U con el coeficiente correspondiente a v^1no nulo, y $W = (U - \{v^1\}) \cup \{w^1\}$. Entonces, los vectores de U son l.i. si y solo si los vectores de W son l.i..*

Demostración. **Ida.** Sabemos que existen $\beta_j \in \mathbb{R}, j = 1, \dots, k$, tales que $w^1 = \sum_{j=1}^{k} \beta_j v^j$, con $\beta_1 \neq 0$. Considero una combinación lineal de los elementos de W que den el vector nulo. Luego,

$$0 = \alpha_1 w^1 + \sum_{j=2}^{k} \alpha_j v^j = \alpha_1 \sum_{j=1}^{k} \beta_j v^j + \sum_{j=2}^{k} \alpha_j v^j = \alpha_1 \beta_1 v^1 + \sum_{j=2}^{k} (\alpha_1 \beta_j + \alpha_j) v^j.$$

Como los vectores de U son l.i., entonces $\alpha_1 \beta_1 = 0$ y $\alpha_1 \beta_j + \alpha_j = 0$ para todo $j = 2, \dots, k$. Como $\beta_1 \neq 0$, entonces $\alpha_1 = 0$, y luego $\alpha_1 \beta_j + \alpha_j = \alpha_j = 0$ para todo $j = 2, \dots, k$. Por lo tanto, los vectores de W son l.i..

Vuelta. Sean $\alpha_j \in \mathbb{R}^n, j = 1, \dots, k$ tales que $\sum_{j=1}^{k} \alpha_j v^j = 0$. Como $\sum_{j=1}^{k} \beta_j v^j$, con $\beta_1 \neq 0$,

$$v_1 = \frac{1}{\beta_1} \left(w^1 - \sum_{j=1}^{k} \beta_j v^j \right).$$

Entonces,

$$\sum_{j=1}^{k} \alpha_j v^j = \alpha_1 v^1 + \sum_{j=2}^{k} \alpha_j v^j = \frac{\alpha_1}{\beta_1}\left(w^1 - \sum_{j=2}^{k} \beta_j v^j\right) + \sum_{j=2}^{k} \alpha_j v^j = \frac{\alpha_1}{\beta_1} w^1 + \sum_{j=2}^{k}\left(\alpha_j - \frac{\alpha_1 \beta_j}{\beta_1}\right) v^j = 0.$$

Luego, si los vectores de W son l.i., $\frac{\alpha_1}{\beta 1} = 0 \implies \alpha_1 = 0$ y $\alpha_j - \frac{\alpha_1\beta_j}{\beta_1} = \alpha_j = 0$ para todo $j = 2, \ldots, k$. Entonces, los vectores de U son l.i.. □

Teorema 6. ***(Teorema Fundamental del Álgebra Lineal - Parte I)*** *En toda matriz, el número máximo de vectores columna l.i. coincide con el número máximo de filas l.i., y ese número es el rango de la matriz.*

Demostración. En el proceso de reducción que se realiza para llevar a una matriz a su forma escalonada, los vectores fila son modificados por una combinación lineal de vectores filas donde participan con coeficiente no nulo. Entonces, todo conjunto de filas l.i. en la matriz reducida es l.i. en la matriz original. Por esto, dada una matriz A $m \times n$, su número máximo de filas l.i. es la cantidad de filas pivots que tiene, es decir, su rango.

Por otro lado, analizando la lineal independencia de las columnas de A, llamemos T a cualquier submatriz por columnas de A. Se sabe que los vectores columna de T son l.i. si y solo si $N(T) = \{0\}$, o equivalentemente si el sistema $Tx = 0$ admite una única solución.

Analizo la forma escalonada de T, llamémosla T'. Sea E la matriz de operaciones elementales tal que $EA = U$. Claramente, $T' = ET$, y las filas nulas en U son también nulas en T', y las columnas pivots de A que estén en T serán columnas pivots de T'. Ahora, $N(T) = \{0\}$ si y solo si todas las columnas de T son pivots, y esto sucede si y solo si los vectores columna de T son l.i. Entonces, un conjunto de columnas A es l.i. si es un subconjunto de sus columnas pivots, por lo que A tiene a lo sumo r vectores columna l.i., donde r es su rango. □

Lema 18. *Sea $(V, \oplus, \odot)$ un espacio vectorial y U un conjunto de vectores l.i. de V de cardinal k. Entonces, el número máximo de vectores l.i. en $\langle U \rangle$ es k.*

Demostración. Sea $U' \subset \langle U \rangle$ un conjunto de vectores l.i. de cardinal k (sabemos que existe porque U lo es). Supongo que existe $x \in \langle U \rangle$ tal que $U' \cup \{x\}$ es un conjunto de vectores l.i.. Luego, x no se puede expresar como combinación lineal de los vectores de U. Entonces, $x \notin \langle U \rangle$. Contradicción. Por lo tanto, el número máximo de vectores l.i. en $\langle U \rangle$ es k. □

Lema 19. *Sea $(V, \oplus, \odot)$ un espacio vectorial. Sea $U = \{v^1, \ldots, v^k\}$ un conjunto de vectores l.i. de V. Entonces:*

1. *Para todo $U' \subset U$, los vectores de U' son l.i..*

2. *Sea $W = \{w^1, \ldots, w^p\}$ un conjunto de vectores l.i. de V, con $p > k$. Entonces, existe $j \in \{1, \ldots, p\}$ tal que $U \cup \{w^j\}$ es un conjunto de vectores l.i..*

Demostración.

1. Supongo $|U'| = r < k$. Sea $\alpha_i = 0$ para $r+1 \leq i \leq k$. Luego, $\sum_{i=1}^{r} \alpha_i u^i = \sum_{i=1}^{k} \alpha_i u^i = 0$, y como los vectores de U son l.i., $\alpha_i = 0$ para todo $i = 1, \ldots, k$, por lo que $\alpha_i = 0$ para todo $i = 1, \ldots, r$.

2. Para cualquier $j \in \{1, \ldots, p\}$ tal que los vectores de $U \cup w^j$ son l.d., w^j puede escribirse como combinación lineal de vectores de U, y por lo tanto $w^j \in \langle U \rangle$. Además, sabemos que el número máximo de vectores l.i. que puede haber en $\langle U \rangle$ es k. Como $|W| = p > k$, entonces no es posible que todos los vectores de W sean elementos de $\langle U \rangle$. Equivalentemente, no es posible que para todo $j \in \{1, \ldots, p\}$, los vectores de $U \cup \{w^j\}$ sean l.d.. Por lo tanto, existe $j \in \{1, \ldots, p\}$ tal que los vectores de $U \cup w^j$ son l.i..

$\square$

5.2. Bases

Definición 15. Dado un espacio vectorial $(V, \oplus, \odot)$ sobre $\mathbb{K}$ y un subconjunto U de V, se dice que U es un **conjunto generador** de V si $\langle U \rangle = V$.

Definición 16. Dado un espacio vectorial $(V, \oplus, \odot)$ sobre $\mathbb{K}$ y un subconjunto U de V, se dice que U es un **conjunto generador minimal** de V si $\langle U \rangle = V$ y no existe un subconjunto U' de U tal que $\langle U' \rangle = V$.

Lema 20. *Sea un espacio vectorial $(V, \oplus, \odot)$ sobre $\mathbb{K}$ y un subconjunto U de V tal que sus vectores son l.d.. Entonces, existe $u \in U$ tal que $\langle U - \{u\} \rangle = \langle U \rangle$.*

Demostración. Sea $u \in U$ tal que u puede escribirse como combinación lineal de los elementos de $U \setminus \{u\}$ (u existe porque los vectores de U son l.d.). Es decir, si U tiene k vectores y $U = \{u, u^1, u^2, \ldots, u^{k-1}\}$, existen escalares $\alpha_i, i = 1, \ldots, k-1$ tales que $u = \sum_{i=1}^{k-1} \alpha_i u^i$. Probaré la doble contención entre $\langle U \setminus \{u\} \rangle$ y $\langle U \rangle$.

($\subset$) Sea $v \in \langle U \setminus \{u\} \rangle$. Luego, existen $\beta_i,\ i = 1, \ldots, k-1$ tales que

$$v = \sum_{i=1}^{k-1} \beta_i u^i = 0 \cdot u + \sum_{i=1}^{k-1} \beta_i u^i.$$

Luego, como v se puede expresar como combinación lineal de los vectores de U, $v \in \langle U \rangle$.

($\supset$) Sea $v \in \langle U \rangle$. Luego, existen escalares $\beta_i,\ i = 0, \ldots, k-1$ tales que

$$v = \beta_0 \cdot u + \sum_{i=1}^{k-1} \beta_i u^i = \sum_{i=1}^{k-1} \beta_0 \alpha_i u^i + \sum_{i=1}^{k-1} \beta_i u^i = \sum_{i=1}^{k-1} (\beta_0 \alpha_i + \beta_i) u^i.$$

Entonces, como v puede expresarse como combinación lineal de los vectores de $U \setminus \{u\}$, $v \in \langle U \setminus \{u\} \rangle$. $\square$

Corolario 4. *Dado un espacio vectorial* $(V, \oplus, \odot)$ *sobre* $\mathbb{K}$ *y un conjunto generador* U *de* V*,* U *es un conjunto generador minimal si y solo si es un conjunto de vectores l.i..*

Demostración. **Ida.** Por hipótesis, U es un conjunto generador minimal de V. Entonces, U genera a V y no existe un subconjunto de U que genere a V. Por el lema anterior, como no existe $\{u\}$ tal que $U \setminus \{u\}$ genere a V, entonces, los vectores de U son l.i. (no son l.d.).

Vuelta. Supongo que U genera a V y los vectores de U son l.i.. Sea $U' \subset U$. Sabemos que los vectores de U son l.i., por lo que ningún vector de U puede expresarse como combinación lineal de los otros vectores. Además, los vectores de U' también son l.i.. Sea $u \in U \setminus U'$. Sólo resta considerar un vector $v \in \langle U \rangle$ en donde el escalar de la combinación lineal correspondiente a u sea no nulo. En ese caso, como u no puede expresarse como combinación lineal de los vectores de $U \setminus \{u\}$ (y por tanto de U'), $v \notin \langle U' \rangle$. Entonces, $\langle U \rangle \neq \langle U' \rangle$, y luego U es conjunto generador minimal. □

Definición 17. Dado un espacio vectorial $(V, \oplus, \odot)$ sobre $\mathbb{K}$, una **base** de V es un conjunto generador de V cuyos vectores son l.i..

Lema 21. *Si* $U = \{v^i : i = 1, \dots, n\}$ *y* $W = \{w^j : j = 1, \dots, m\}$ *son bases de un espacio vectorial* $(V, \oplus, \odot)$ *sobre* $\mathbb{R}$*, entonces* $m = n$*.*

Demostración. Sin pérdida de generalidad, supongo que $m \leq n$. Como W genera a V, para todo $i = 1, \dots, n$ existen $a^i_j \in \mathbb{R}$, con $j = 1, \dots, m$, tales que $v^i = \sum\limits_{j=1}^{m} a^i_j w^j$. Como los vectores de U son l.i.,

$$\sum_{i=1}^{n} \alpha_i v^i = 0 \iff \alpha_i = 0 \ \forall i = 1, \dots n.$$

Además, como los vectores de W son l.i., resulta

$$\sum_{i=1}^{n} \alpha_i v^i = \sum_{i=1}^{n} \alpha_i \sum_{j=1}^{m} a^i_j w^j = \sum_{j=1}^{m} \left(\sum_{i=1}^{n} \alpha_i a^i_j \right) w^j = 0 \iff \sum_{i=1}^{n} \alpha_i a^i_j = 0 \ \forall j = 1, \dots, m,$$

y juntando esto con lo anterior, se tiene que

$$\sum_{i=1}^{n} \alpha_i a^i_j = 0 \ \forall j = 1, \dots, m \iff \alpha_i = 0 \ \forall i = 1, \dots n.$$

Es decir, el sistema de ecuaciones $\sum\limits_{i=1}^{n} \alpha_i a^i_j = 0 \ \forall j = 1, \dots, m$ tiene como única solución a $\alpha_i = 0$ para todo $i = 1, \dots, n$. Se puede ver que la matriz A de coeficientes del sistema es la matriz $m \times n$ tal que para todo $i = 1, \dots, n$, $A^i = (a^i_1, \dots, a^i_m)^T$. Además, que el sistema tenga única solución es equivalente a que $N(A) = \{0\}$, para lo que debe suceder que $n \leq m$. Entonces, $n = m$. □

Definición 18. Se denomina **dimensión** de un espacio vectorial al cardinal de sus bases.

Corolario 5. *Cualquier conjunto de* $n + 1$ *vectores en un espacio vectorial* V *de dimensión* n *es l.d..*

Demostración. Sea B una base de V. Luego, $|B| = n$, y por el lema anterior, como los elementos de B son l.i y $\langle B \rangle = V$, el número máximo de vectores l.i. en V es n, y cualquier subconjunto de V con más de n vectores es l.d.. □

Lema 22. *Sea V un espacio vectorial de dimensión n y U un subconjunto de k elementos de V. Entonces:*

1. *Si los elementos de U son l.i. y $k = n$, entonces U es una base de V.*
2. *Si los elementos de U son l.i. y $k < n$, entonces existe una base B de V tal que $U \subset B$.*
3. *Si U es un conjunto generador de V y $k > n$, entonces existe una base B de V tal que $B \subset U$.*

Demostración.

1. Como U posee n elementos l.i.. de V, resta ver que U genera a V. Supongo que U no genera a V. Entonces, existe v tal que $U \cup \{v\}$ es un conjunto de vectores l.i. en V. Es decir, se encontró un conjunto de $n+1$ vectores l.i. en un espacio vectorial de dimensión n. Contradicción. Por lo tanto, U genera a V, y luego es base de V.
2. Sin pérdida de generalidad, sea $k = n - 1$. Luego, como el cardinal de cualquier base de V es n, existe al menos un vector x que no es combinación lineal de los vectores de U. Es decir, $U \cup \{x\}$ es un conjunto de vectores l.i, con $|U \cup \{x\}| = n$ (para valores menores de k deben agregarse exactamente $n - k$ vectores). Por el apartado anterior, $U \cup \{x\}$ es una base de V, y $U \subset U \cup \{x\}$.
3. Sin pérdida de generalidad, sea $k = n + 1$. Por lema probado anteriormente, sabemos que los elementos de U son l.d. (puede haber como máximo n elementos l.d. en V). Luego, como $\langle U \rangle = V$, tenemos que hay n elementos l.i. en U, pues si no los hubiera, por el mismo lema podría haber un máximo de n elementos l.i. en V y existirían elementos de V que no podrían ser generados por U. Entonces, existe un elemento $x \in U$ que es combinación lineal de los otros elementos de U. Considerando $U' = U \setminus \{x\}$ (para valores mayores de k se le deben restarse exactamente $k - n$ vectores a U), U' es un conjunto de n vectores l.i. en V, por lo que es base de V y $B \subset U$.

□

De estos resultados se puede concluir que las bases de un espacio vectorial son conjuntos independientes maximales y conjuntos generadores minimales.

Observación 12. Si $\mathbb{K}$ es infinito, a partir de una base de un espacio vectorial V sobre $\mathbb{K}$ se pueden construir infinitas bases.

Lema 23. *Sea B una base del espacio vectorial $(V, \oplus, \odot)$ sobre $\mathbb{K}$, y $v \in B$. Sea w una combinación lineal de vectores de B en los que el coeficiente asociado a v es no nulo. Entonces, $(B - \{v\}) \cup \{w\}$ es una base de V.*

Demostración. Sea n la dimensión de V. Por lema, como los vectores de B son l.i. y el coeficiente asociado a v en w es no nulo, entonces los vectores de $B \setminus \{v\} \cup \{w\}$ son l.i., y como la cantidad de vectores de $B \setminus \{v\} \cup \{w\}$ es igual a la dimensión del espacio, por lema tenemos que el conjunto es una base de V. □

Definición 19. El **espacio fila** de una matriz A $m \times n$ es el espacio generado por sus vectores filas. Nótese que el espacio fila de A es $C(A^T) \in \mathbb{R}^n$.

Definición 20. El **espacio nulo a izquierda** de una matriz A $m \times n$ es $N(A^T) \in \mathbb{R}^m$. Observar que $y \in N(A^T) \iff A^T y = 0 \iff y^T A = 0$.

Como $C(A)$ y $C(A^T)$ son los espacios generados por los vectores columna y vectores fila, respectivamente, tenemos que los r vectores columnas l.i. y los r vectores fila l.i. son conjuntos generadores minimales de $C(A)$ y $C(A^T)$, y luego son bases de estos espacios, por lo que la dimensión de ambos es r. Notar que aunque su dimensión es la misma, viven en espacios diferentes.

Al buscar bases de los cuatro espacios fundamentales de una matriz A, es conveniente escalonarla a una matriz U. Luego, las columnas pivot en la matriz original A forman una base de $C(A)$, y las soluciones especiales, una vez detectadas las variables libres, forman una base de $N(A)$. Para encontrar una base del espacio fila de A $C(A^T)$, no es necesario volver a la matriz original para obtener un conjunto generador; basta con las filas pivot de la matriz U, pues en el proceso de reducción se reemplazan las filas por combinaciones lineales de ellas, en donde el coeficiente asociado a la fila en cuestión siempre es no nulo. Por esta misma razón, se conserva la lineal independencia y, al ser un conjunto l.i. maximal, forman una base de $C(A^T)$. Para obtener una base de $N(A^T)$ es necesario resolver el sistema $A^T y = 0 \iff y^T A = 0$.

Lema 24. *Dada una matriz A $m \times n$, la suma de las dimensiones de los espacios $C(A)$ y $N(A)$ es n, el número e columnas de A.*

Demostración. Sea r el rango de A, es decir, la cantidad de columnas pivot que tiene. Entonces, la dimensión de $C(A)$ es r. Por otro lado, A posee $n - r$ variables libres, y sabemos que una base de $N(A)$ está dada por las soluciones especiales (una por cada variable libre), por lo que la dimensión de $N(A)$ es $n - r$. Luego, $\dim(\mathrm{C(A)}) + \dim(\mathrm{N(A)}) = r + (n - r) = n$. □

Lema 25. *Sea $B = \{w^1, \ldots, w^k\}$ una base del espacio vectorial $(V, \oplus, \odot)$. Entonces, para todo $v \in V$ existe una única combinación lineal de elementos de B que generan a v.*

Demostración. Se nota $\alpha v = \alpha \odot v$ y $u + v = u \oplus v$ para todo $\alpha \in \mathbb{K},\ u, v \in V$.

Sean $\alpha_i, \beta_i,\ i = 1, \ldots, k$ tales que

$$v = \sum_{i=1}^{k} \alpha_i w^i = \sum_{i=1}^{k} \beta_i w^i.$$

Hay que probar que $\alpha_i = \beta_i,\ i = 1, \ldots, k$. Efectivamente,

$$\sum_{i=1}^{k} \alpha_i w^i = \sum_{i=1}^{k} \beta_i w^i \iff \sum_{i=1}^{k} (\alpha_i - \beta_i) = 0 \iff \alpha_i = \beta_i\ \forall i = 1, \ldots, k.$$

□

Definición 21. Dado un espacio vectorial de dimensión n y base B y un vector $v \in V$, llamamos **representación de** v **en** B a la única n-upla de escalares que permite obtener a v como combinación lineal de elementos de B.

5.2.1. Matrices de Cambio de Base

Como se vio antes, toda combinación lineal de los vectores de la base puede expresarse a partir de un producto matricial. Es decir, si llamamos B a la matriz $n \times n$ que tiene por columnas a los vectores de $\mathcal{B}_1 = \{v^i : i = 1, \ldots n\}$ (esto es, $B^j = v^j$, $j = 1, \ldots n$), toda combinación lineal de estos vectores puede expresarse como Bx, con $x \in \mathbb{R}^n$. Entonces, la representación de w en $\mathcal{B}_1$ es la solución del sistema $Bx = w$. Como B tiene por columnas a los vectores de una base, B es no singular y por lo tanto siempre existirá una única solución a este sistema (lo cual es consistente con la unicidad de representación de un vector en una base).

Por otro lado, consideremos además a la base $\mathcal{B}_2 = \{w^i : i = 1, \ldots n\}$ y a su matriz asociada $\tilde{B}$. Si x e y son los vectores tales que $Bx = w$ y $\tilde{B}y = w$, tenemos que

$$Bx = \tilde{B}y \implies y = \tilde{B}^{-1}(Bx) = (\tilde{B}^{-1}B)x = Tx, \quad \text{con } T = \tilde{B}^{-1}B.$$

La matriz T que transforma la representación de un vector en la base $\mathcal{B}_1$ en su representación en la base $\mathcal{B}_2$ se llama **matriz de cambio de base** (de $\mathcal{B}_1$ a $\mathcal{B}_2$).

Observación 13. Considerando las bases y matrices anteriores, la matriz T de cambio de base de $\mathcal{B}_1$ a $\mathcal{B}_2$ se puede obtener teniendo en cuenta que $T^i = (\tilde{B}^{-1}B)^i = \tilde{B}^{-1}B^i = \tilde{B}^{-1}v^i$. Esto es equivalente a que T^i sea la solución del sistema $\tilde{B}x = v^i$. Coloquialmente, *la columna i-ésima de la matriz de cambio de base de* $\mathcal{B}_1$ *a* $\mathcal{B}_2$ *es la representación del vector i-ésimo de* $\mathcal{B}_1$ *en la base* $\mathcal{B}_2$.

Definición 22. Si V es un espacio vectorial de dimensión p sobre $\mathbb{R}$, y $\mathcal{B}$ es una base ordenada de V, la notación $[z]_{\mathcal{B}}$ indica el vector en $\mathbb{R}^p$ correspondiente a representación de z en la base $\mathcal{B}$.

Definición 23. Dadas dos bases ordenadas $\mathcal{B}_1, \mathcal{B}_2$ de V, la matriz de cambio de base M de $\mathcal{B}_1$ a $\mathcal{B}_2$ es la matriz que verifica, para todo $v \in V$, $M[v]_{\mathcal{B}_1} = [v]_{\mathcal{B}_2}$.

Observación 14. Si $\mathcal{B} = \{v^i : i = 1, \ldots n\}$ es una base ordenada del e.v. V, entonces $[v^i]_{\mathcal{B}} = e_i$.

6. Inversas a Izquierda y a Derecha

Definición 24. Dada una matriz A $m \times n$, B es una **inversa a izquierda** de A si $BA = I_n$. Si $AC = I_m$, C es una **inversa a derecha** de A.

Observación 15. Si A tiene inversa a derecha B e inversa a izquierda C, entonces A es una matriz cuadrada inversible, y $B = C = A^{-1}$. Además, ninguna matriz rectangular tiene matrices inversibles a ambos lados; tendrá a lo sumo una de las inversas: a izquierda o a derecha.

Definición 25. Notando al rango de una matriz A $m \times n$ como $rg(A)$, se dice que A es de **rango completo** si $rg(A) = \min(m, n)$.

Lema 26. *Sea A una matriz $m \times n$ tal que $rg(A) = m$. Entonces, A tiene inversa a derecha.*

Demostración. Como $rg(A) = m$, sabemos que $C(A) = \mathbb{R}^n$ al no tener filas nulas su matriz reducida. Es decir, el sistema $Ax = b$ tiene solución para todo $b \in \mathbb{R}^n$. En particular, si C^i es una solución del sistema $Ax = e_i$ (con e_i el i-ésimo vector canónico en $\mathbb{R}^n$), la matriz C que tiene por i-ésimo vector columna a C^i, con $i = 1, \ldots, n$, es inversa a derecha de A. □

Observación 16. Si $m < n$, la inversa a derecha de A no es única. Eso sucede ya que, como $m < n$, las columnas de A son l.d. y aparecen variables libres en la resolución del sistema, y por esto, dado un vector $b \in \mathbb{R}^n$ el sistema $Ax = b$ tiene infinitas soluciones.

Si $rg(A) = n < m$, A no tiene inversa a derecha, pues al menos la última fila (fila m) de la matriz reducida de A es nula, y siguiendo el mismo razonamiento que antes, tenemos que el sistema $Ax = e_m$ no admite solución.

Por otro lado, si A es una matriz $m \times n$ con $rg(A) = n$ y $n < m$, entonces A^T es $n \times m$ y tiene inversa a derecha C, es decir, $A^T C = I_n$. Luego,

$$C^T A = (A^T C)^T = I_n^T = I_n,$$

y resulta que $B = C^T$ es una inversa a izquierda de A.

Lema 27. *Sea A una matriz $m \times n$ y $b \in \mathbb{R}^m$:*

1. *Si $rg(A) = n$ y el sistema $Ax = b$ tiene solución, entonces esa solución es única.*

2. *Si $rg(A) = m$ entonces el sistema $Ax = b$ tiene una o infinitas soluciones.*

Demostración. Si $rg(A) = n$, las n columnas de A son l.i., por lo que considerando a la matriz R reducida de A, la submatriz por filas de la primeras n filas R es no singular (ya que es triangular superior), y como $b \in C(A)$, el sistema $Ax = b$ tiene única solución.

Por otro lado, si $rg(A) = m$, entonces $rg(A) \leq n$. Si $rg(A) = n$, A es una matriz cuadrada no singular, por lo que existe única solución al sistema $Ax = b$ admite una única solución. Si $rg(A) < n$, las columnas de A son l.d. y por lo tanto al haber al menos una variable libre, el sistema $Ax = b$ admite infinitas soluciones. □

Definición 26. En caso de tener una matriz A infinitas inversas a izquierda o a derecha, se denomina **pseudo-inversa** (a derecha o izquierda) a la matriz entre ellas donde los valores arbitrarios son iguales a cero.

Observación 17. Dada una matriz A $m \times n$, AA^T y A^TA son matrices cuadradas $m \times m$ y $n \times n$ respectivamente. Si $rg(A) = n$, entonces A^TA es inversible, y la matriz $B = (A^TA)^{-1}A^T$ es la pseudo-inversa a izquierda de A. Por otro lado, se $rg(A) = m$, AA^T es inversible, y $C = A^T(AA^T)^{-1}$ es la pseudo-inversa a derecha de A.

7. Transformaciones Lineales

Sea A una matriz $n \times n$. Cuando A multiplica a cualquier vector $x \in \mathbb{R}^n$, se dice que lo *transforma* en el vector Ax, por lo que podemos decir que A transforma o mapea a $\mathbb{R}^n$ en sí mismo. Tomando como ejemplo a $\mathbb{R}^2$, algunos ejemplos de transformaciones que provienen de matrices son:

1. $A = \begin{bmatrix} c & 0 \\ 0 & c \end{bmatrix}$: A es un múltiplo de la matriz identidad, $A = cI$, y **expande** o **contrae** a cualquier vector, dependiendo de si $c > 1$ o $c < 1$, y cambia su sentido si $c < 0$.

2. $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$: A es una matriz de **rotación**, rotando cualquier vector 90°, transformando (x, y) en $(-y, x)$.

3. $A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$: A es una matriz de **proyección**, transformando cada vector del plano (x, y) es el punto más cercano $(x, 0)$ en el eje horizontal.

4. $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$: A es una matriz de **reflexión** a través de la recta $y = x$, transformando cada vector (x, y) en (y, x).

Es importante reconocer que no toda transformación de $\mathbb{R}^n$ puede expresarse a partir de una matriz. Por ejemplo, cualquier transformación que mueve el origen no puede expresarse mediante una matriz, ya que, cualquiera sea A, $A0 = 0$.

Mas específicamente, las transformaciones que realizadas por matrices verifican ser *lineales*, es decir,

$$A(\alpha v + \beta w) = \alpha Av + \beta Aw.$$

Extendiendo este concepto a cualquier transformación de un espacio vectorial en otro, llegamos a la siguiente definición:

Definición 27. Sean $(V, \oplus, \odot)$ y $(W, +, \cdot)$ espacios vectoriales sobre $\mathbb{K}$. Una **transformación lineal** de V en W es una función $T : V \longrightarrow W$ que verifica que, para todo $v, w \in V$ y todo $\alpha, \beta \in \mathbb{K}$,

$$T((\alpha \odot v) \oplus (\beta \odot w)) = (\alpha \cdot T(v)) + (\beta \cdot T(w)).$$

Entonces, cualquier función definida a partir del producto por una matriz nos lleva a una transformación lineal. Es interesante ver que la recíproca también vale: cualquier transformación lineal nos lleva a una matriz. Esta es la base de un enfoque distinto del álgebra lineal; empezar por la linealidad y desarrollar sus consecuencias, aunque es mucho más abstracto que empezar directamente con matrices y ver cómo representan transformaciones lineales.

Otros ejemplos de transformaciones lineales son, dados V y W espacios vectoriales, la función nula que asigna a todo vector de V el vector nulo de W; la función identidad definida en V sobre sí mismo, la cual a cada vector $v \in V$ le asigna el propio v. También, el operador derivación es lineal sobre el espacio vectorial de funciones derivables en un subconjunto de $\mathbb{R}$, y la integración puede ser vista como un operador lineal sobre el espacio vectorial de funciones continuas en $\mathbb{R}$.

Lema 28. *Dado un operador lineal T entre los espacios V y W, el recorrido de T es un subespacio vectorial de W y $N(T) = \{v \in V : T(v) = 0\}$ es un subespacio vectorial de V.*

Demostración. En primer lugar, dados $u, v \in V$, por la linealidad, $\alpha T(u) + \beta T(v) = T(\alpha u + \beta v)$, con $\alpha u + \beta w \in V$, por lo que $\alpha T(u) + \beta T(v) \in rec(T)$. Por lo tanto, $rec(T)$ es un subespacio vectorial de W.

Por otro lado, considerando $u, v \in N(T)$, nuevamente por la linealidad tenemos que $T(\alpha u + \beta v) = \alpha T(u) + \beta T(v) = \alpha \cdot 0 + \beta \cdot 0 = 0 + 0 = 0$. Luego, $\alpha u + \beta v \in N(T)$, y por lo tanto $N(T)$ es un subespacio vectorial de V. □

Definición 28. Si T es una transformación lineal de V en W, al subespacio $N(T)$ se lo llama **espacio nulo** de T o **kernel** de T.

Observación 18. Si T es un operador lineal asociado a una matriz A $m \times n$, el espacio nulo de T es $N(A)$ y el recorrido de T coincide con $C(A)$.

Definición 29. Las transformaciones lineales de espacios vectoriales se suelen llamar **homomorfismos** de espacios vectoriales, y de esta nominación surge:

1. **Monomorfismo**, para las transformaciones lineales inyectivas,

2. **Epimorfismo**, para las transformaciones lineales sobreyectivas,

3. **Isomorfismo**, para las transformaciones lineales biyectivas.

Lema 29. *Una transformación lineal T de V en W es un monomorfismo si y solo si $N(T) = \{0\}$.*

Demostración. Claramente, para toda transformación lineal T, $T(0) = 0$, pues el origen siempre está fijo. Luego, si T es un monomorfismo, es inyectiva y no existe otro vector en V cuya imagen sea $0 \in W$.

Ahora, suponiendo que $N(T) = \{0\}$, debemos probar que T es inyectiva, es decir, que $Tv = Tw \implies v = w$. Efectivamente,

$$Tv = Tw \implies Tv - Tw = 0 \implies T(v - w) = 0 \implies v - w \in N(T) \implies v - w = 0 \implies v = w.$$

□

Sabiendo que una transformación lineal definida por la matriz A $m \times n$ es un homomorfismo si $N(A) = \{0\}$, y es un epimorfismo si $C(A) = \mathbb{R}^n$, tenemos que la transformación es un isomorfismo si se verifican estas dos condiciones, es decir, si A es una matriz no singular.

Lema 30. *Sea T un monomorfismo de V en W y sean $\{v^1, \ldots, v^k\}$ vectores l.i. en V. Entonces, $\{Tv^1, \ldots, Tv^k\}$ son vectores l.i. en W.*

Demostración. Consideramos una combinación lineal nula en W de los vectores en $\{Tv^1, \ldots, Tv^k\}$, es decir,

$$\alpha_1 Tv^1 + \cdots + \alpha_k Tv^k = 0 \in W. \text{ Luego, } T(\alpha_1 v^1 + \cdots + \alpha_k v^k) = 0,$$

y por lo tanto $\alpha_1 v^1 + \cdots + \alpha_k v^k \in N(T)$. Como T es inyectiva, $N(T) = \{0\}$ y luego $\alpha_1 v^1 + \cdots + \alpha_k v^k = 0 \in V$. Finalmente, como los vectores de $\{v^1, \ldots, v^k\}$ son l.i. por hipótesis, necesariamente $\alpha_i = 0$ para todo $i = 1, \ldots, k$. Entonces, los vectores de $\{Tv^1, \ldots, Tv^k\}$ son l.i. en W. □

Corolario 6. *Si T es un monomorfismo de V en W y $\{v^1, \ldots, v^k\}$ es una base de V, entonces, $\{Tv^1, \ldots, Tv^k\}$ es una base del subespacio imagen de T.*

Demostración. Por el lema anterior, sabemos que los vectores de $\{Tv^1, \ldots, Tv^k\}$ son l.i.. Resta ver que generan $rec(T)$. Efectivamente, dado $x \in V$, $x = \sum_{i=1}^{k} \alpha_i v^i$, con $\alpha_i \in \mathbb{K}$, tenemos que

$$T(x) = T\left(\sum_{i=1}^{k} \alpha_i v^i\right) = \sum_{i=1}^{k} \alpha_i T(v^i),$$

por lo que $\{Tv^1, \ldots, Tv^k\}$ genera a $rec(T)$. □

Observación 19. Toda transformación lineal de un espacio vectorial V en otro W queda unívocamente definida por las imágenes a través de T de los elementos de una de sus bases. Es decir, si $\mathcal{B}$ es una base de V, una base de W es $T(\mathcal{B}) = \{Tv : v \in \mathcal{B}\}$. Además, si $z = \sum_{i=1}^{t} \alpha_i v^i$ con $\{v^1, \ldots, v^t\} \subset \mathcal{B}$, entonces $Tz = \sum_{i=1}^{t} \alpha_i Tv^i$.

Lema 31. *Sean V y W dos espacios vectoriales de dimensión finita. Entonces, V y W son isomorfos si y solo si V y W tienen la misma dimensión.*

Demostración. **Ida.** Si V y W son isomorfos, existe una transformación lineal $T : V \longrightarrow W$ biyectiva. En particular, T es inyectiva si $\{v^1, \dots, v^k\}$ es una base de V, entonces $\{Tv^1, \dots, Tv^k\}$ es una base de $rec(T)$. Como T es también sobreyectiva, $\{Tv^1, \dots, Tv^k\}$ es una base de W, y ambos espacios tienen la misma dimensión.

Vuelta. Sean $\{v^1, \dots, v^k\}$ una base de V y $\{w^1, \dots, w^k\}$ una base de W. Defino $T : V \longrightarrow W$ tal que $T(v^i) = w^i$ para todo $i = 1, \dots, k$. Probaré que T es biyectiva.

Dados $v = \sum_{i=1}^{k} \alpha_i v^i, w = \sum_{i=1}^{k} \beta_i v^i \in V$, con $\alpha_i, \beta_i \in \mathbb{K},\ i = 1, \dots, k$ tales que $v \neq w$ (es decir, los escalares no son iguales entre sí), tenemos que

$$v \neq w \implies Tv = T\left(\sum_{i=1}^{k} \alpha_i v^i\right) = \sum_{i=1}^{k} \alpha_i w^i \neq \sum_{i=1}^{k} \beta_i w^i = \sum_{i=1}^{k} \beta_i T v^i = T\left(\sum_{i=1}^{k} \beta_i v^i\right) = Tw,$$

por lo que T es un inyectiva.

Por otro lado, dado $w = \sum_{i=1}^{k} \alpha_i w^i,\ \alpha_i \in \mathbb{K}$ para todo $i = 1, \dots, k$, existe $v = \sum_{i=1}^{k} \alpha_i v^i$ tal que

$$Tv = T\left(\sum_{i=1}^{k} \alpha_i v^i\right) = \sum_{i=1}^{k} \alpha_i T v^i = \sum_{i=1}^{k} \alpha_i w^i = w,$$

quedando demostrado que T es sobreyectiva, y luego un isomorfismo. □

Corolario 7. *Todo espacio vectorial sobre $\mathbb{R}$ de dimensión n es isomorfo a $\mathbb{R}^n$.*

Ya se vio que una matriz A $m \times n$ define una transformación lineal de $\mathbb{R}^n$ en $\mathbb{R}^m$ a través de la ley $A : \mathbb{R}^n \longrightarrow \mathbb{R}^m \;/\; x \longrightarrow Ax$. Recíprocamente, se puede ver que cualquier transformación lineal de $\mathbb{R}^n$ en $\mathbb{R}^m$ puede ser definida por una matriz $m \times n$. Considerando una transformación lineal T de $\mathbb{R}^n$ en $\mathbb{R}^m$ y los vectores canónicos de $\mathbb{R}^n$, podemos definir la matriz A $m \times n$ tal que $A^i = Te^i$. Así, $Tv = Av\ \forall v \in \mathbb{R}^n$, pues

$$v = \sum_{i=1}^{n} v_i e^i \implies Tv = \sum_{i=1}^{n} v_i A^i = Av.$$

Definición 30. La matriz asociada a T con bases $\mathcal{B}_V$ en V y $\mathcal{B}_W$ en W es la matriz A_T de tamaño $m \times n$ que para todo $v \in V$, $A_T[v]_{\mathcal{B}_V} = [Tv]_{\mathcal{B}_W}$.

Luego, tenemos que para todo $i = 1, \dots n$, $A_T^i = [Tv^i]_{\mathcal{B}_W} \in \mathbb{R}^m$, con $\mathcal{B}_V = \{v^1, \dots v^m\}$.

Algoritmo para construir A_T:

1. Calcular Tv^i para $i = 1, \dots n$
2. Calcular $[Tv^i]_{\mathcal{B}_W}$ para $i = 1, \dots n$
3. $A_T^i := [Tv^i]_{\mathcal{B}_W}$

7.1. Propiedades

Lema 32. *Sea T una transformación lineal entre espacios vectoriales V y W de dimensión finita y A_T su matriz asociada. Entonces, el recorrido de T es isomorfo a $C(A_T)$ y T es un monomorfismo si $N(A_T) = \{0\}$.*

Demostración. Como $T(v) = A_T v$ para cualquier $v \in V$, $rec(T) = \{A_T v : v \in V\} = C(A_T)$. Es decir, $v \in rec(T) \iff v \in C(A_T)$, por lo que tienen la misma dimensión, y luego son isomorfos. □

Teorema 7. *Dados dos espacios vectoriales V y W sobre un mismo cuerpo $\mathbb{K}$, las transformaciones lineales de V en W definen un espacio vectorial con la suma y producto por escalar estándares en funciones.*

Lema 33. *Sean V y W espacios vectoriales sobre $\mathbb{K}$, y T un isomorfismo de V en W, con T^{-1} su función inversa. Entonces, T^{-1} es un isomorfismo de W en V.*

Demostración. Si $T : V \longrightarrow W$ es un isomorfismo, su matriz asociada A_T es no singular. Dado $v' = A_T v \in W$, la transformación lineal T^{-1} está dada por $T^{-1}(Tv) = T^{-1}(A_T v) = v$. Nombrando a su matriz asociada $A_{T^{-1}}$, $T^{-1}(A_T v) = A_{T^{-1}} A_T v = v \implies A_{T^{-1}} A_T = I \implies A_{T^{-1}} = A_T^{-1}$. Luego, T^{-1} es un isomorfismo, pues su matriz asociada es no singular. □

Lema 34. *Sean V, W y U tres espacios vectoriales sobre $\mathbb{K}$, T^1 una transformación lineal de V en W y T^2 una transformación lineal de W en U. Entonces, $T = T^2 \circ T^1$ es una transformación lineal de V en U.*

Demostración. $T(\alpha v+\beta w) = T^2(T^1(\alpha v+\beta w)) = T^2(\alpha T^1(v)+\beta T^1(w)) = \alpha T^2(T^1(v))+\beta T^2(T^1(w))$. □

Observación 20. Habiendo visto que la transformación lineal definida por una matriz A es un isomorfismo si y solo si A es no singular, entonces se puede probar que los isomorfismos definidos por A y A^{-1} son isomorfismos inversos.

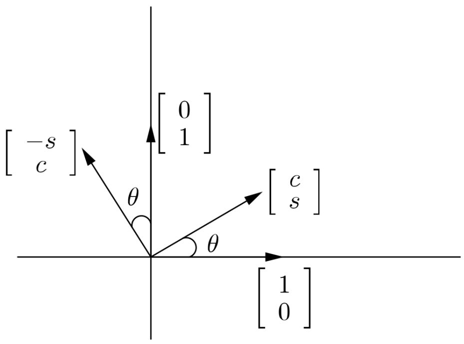
Además, si B es $m \times n$ y C es $n \times p$, con ambas matrices definiendo homomorfismos de $\mathbb{R}^n$ a $\mathbb{R}^m$, y de $\mathbb{R}^n$ a $\mathbb{R}^p$, respectivamente, entonces la matriz BA define al homomorfismo $B \circ A$.

7.2. Rotaciones Q, Proyecciones P y Reflexiones H

Ya se consideraron las matrices de rotación de 90°, proyección sobre el eje x y reflexión sobre la recta $y = x$ en $\mathbb{R}^2$. Sin embargo, las rotaciones de otros ángulos, proyecciones sobre otras rectas y reflexiones a través de otros ejes también son transformaciones lineales, siempre y cuando el origen esté fijo ($A0 = 0$). Para ver la forma de dichas matrices se usará la base canónica compuesta por los vectores $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ y $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$, y dado un ángulo θ, llamamos $c = \cos\theta$ y $s = \sin\theta$.

7.2.1. Rotación

No es difícil ver que el vector (c, s) es obtenido luego de rotar en un ángulo θ al primer vector de la base. Además, la longitud de ambos es 1, y al rotar el segundo vector de la base en un ángulo θ, se obtiene el vector $(-s, c)$.



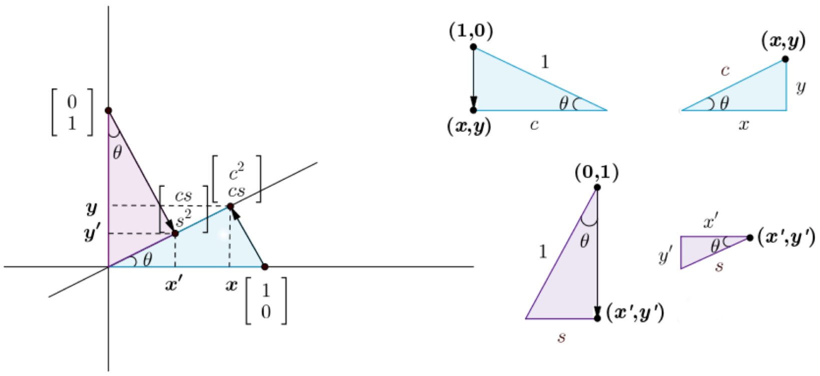
Ahora, para describir la matriz asociada a la transformación lineal rotación, consideramos $\mathbb{R}^2$ y su base canónica. Siendo T la transformación lineal correspondiente a la rotación en un ángulo θ, sabemos que la j-ésima columna de la matriz Q_θ que representa dicha transformación se obtiene aplicando la transformación al j-ésimo vector de la base. Luego,

$$\begin{cases} T(e_1) = ce_1 + se_2 \\ T(e_2) = -se_1 + ce_2 \end{cases} \implies Q_\theta = \begin{bmatrix} c & -s \\ s & c \end{bmatrix}.$$

Efectivamente, se verifican todas las propiedades asociadas a la rotación de vectores con factor θ.

7.2.2. Proyección

En este caso, el primer vector de la base canónica $(1, 0)$ se transforma mediante la proyección sobre la recta θ en $(x, y) = (c^2, cs)$, y el segundo vector de la base canónica, al proyectarlo sobre la misma recta se transforma en $(x', y') = (cs, s^2)$.



Luego, como consideramos la base canónica, la matriz de la transformación lineal termina siendo

$$P_\theta = \begin{bmatrix} c^2 & cs \\ cs & s^2 \end{bmatrix},$$

y cabe destacar que no tiene inversa, y $P_\theta^2 = P_\theta$.

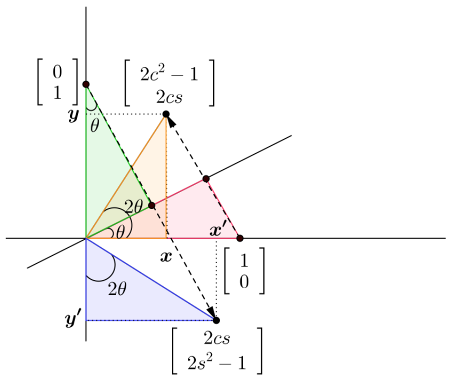
7.2.3. Reflexión

La reflexión del vector $(1,0)$ sobre la recta θ es el vector (x,y). Luego, mirando la figura,

$$x = \cos 2\theta \implies x = \cos^2\theta - \sin^2\theta \implies x = 2\cos^2 - 1 \implies x = 2c^2 - 1$$

$$y = \sin 2\theta \implies y = 2\cos\theta\sin\theta \implies y = 2cs.$$

Análogamente, la reflexión del vector $(0,1)$ sobre la recta θ es el vector $(x', y') = (2cs, 2s^2 - 1)$.



Luego, la matriz H_θ asociada a la transformación reflexión termina siendo

$$H_\theta = \begin{bmatrix} 2c^2 - 1 & 2cs \\ 2cs & 2s^2 - 1 \end{bmatrix}.$$

Esta matriz posee la propiedad $H_\theta^2 = I$, en concordancia con la naturaleza de la reflexión de un vector sobre una recta.

Unidad III

Ortogonalidad

Es de interés en el estudio del Álgebra Lineal la simplificación de cálculos en cuanto a bases y espacios vectoriales; por esta razón es de utilidad el concepto de ortogonalidad entre vectores (y luego, entre bases y subespacios vectoriales).

Para extender las nociones de perpendicularidad entre vectores y longitud de un vector en, por ejemplo, tres dimensiones ($u \perp v \iff u^T v = 0$, $|u| = \sqrt{u^T u}$) a cualquier espacio vectorial, es necesario tener definido un producto interno.

Definición 31. Sea V un espacio vectorial sobre el cuerpo $\mathbb{K}$ ($\mathbb{R}$ o $\mathbb{C}$). Un **producto interno** sobre V es una función $\langle \cdot, \cdot \rangle$ que a cada par de vectores $u, v \in V$ le asigna un escalar $\langle u, v \rangle \in \mathbb{K}$ y verifica las siguientes propiedades para todo $\alpha \in \mathbb{K}$, y todo $u, v, w \in V$:

1. $\langle u, v \rangle = \overline{\langle v, u \rangle}$
2. $\langle u + v, w \rangle = \langle u, w \rangle + \langle v, w \rangle$
3. $\langle \alpha u, v \rangle = \alpha \langle u, v \rangle$
4. $\langle u, u \rangle \geq 0$
5. $\langle u, u \rangle = 0 \iff u = 0$

Otras propiedades que se derivan de un producto interno bien definido son:

- $\langle u, 0 \rangle = \langle 0, v \rangle = 0$

  Demostración. $\langle u, v \rangle = \langle u + 0, v \rangle = \langle u, v \rangle + \langle 0, v \rangle \implies \langle 0, v \rangle = 0$. Por otro lado, $\langle u, v \rangle = \overline{\langle v, u \rangle} = \overline{\langle v + 0, u \rangle} = \overline{\langle v, u \rangle + \langle 0, u \rangle} = \overline{\langle v, u \rangle} + \overline{\langle 0, u \rangle} = \langle u, v \rangle + \langle u, 0 \rangle$, por lo que $\langle u, 0 \rangle = 0$. □

- $\langle u, \alpha v \rangle = \overline{\alpha} \langle u, v \rangle$

  Demostración. $\langle u, \alpha v \rangle = \overline{\langle \alpha v, u \rangle} = \overline{\alpha \langle v, u \rangle} = \overline{\alpha} \langle u, v \rangle$. □

- $\langle u, v + w \rangle = \langle u, v \rangle + \langle u, w \rangle$

  Demostración. $\langle u, v + w \rangle = \overline{\langle v + w, u \rangle} = \overline{\langle v, u \rangle + \langle w, u \rangle} = \overline{\langle v, u \rangle} + \overline{\langle w, u \rangle} = \langle u, v \rangle + \langle u, w \rangle$. □

Algunos ejemplos de productos internos en diferentes subespacios son:

- En $\mathbb{C}^n$, si con $\overline{z}$ indicamos la n-upla de números complejos correspondientes a los conjugados de $z \in \mathbb{C}^n$, se tiene que
$$\langle z, w \rangle = z\overline{w}$$
es un producto interno. Se conoce como *producto interno canónico.*

- En $\mathbb{K}_n[x]$, con $t_0, \ldots, t_n$ escalares distintos, para $p, q \in \mathbb{K}_n[x]$,
$$\langle p, q \rangle = \sum_{i=1}^{n} p(t_i)q(t_i)$$
es un producto interno bien definido.

- En V, donde V es el espacio vectorial de las funciones reales continuas en el intervalo $[0,1]$, con $f, g \in V$, se puede verificar que
$$\langle f, g \rangle_1 = \int_0^1 f(x)g(x)dx \quad \text{y} \quad \langle f, g \rangle_2 = \int_1^e \ln(x)f(x)g(x)dx$$
son definiciones válidas de productos internos.

- En $\mathbb{R}^{m\times n}$, con $A = (a_{ij}) \in \mathbb{R}^{m\times n}$ y $B = (b_{ij}) \in \mathbb{R}^{m\times n}$, se puede definir el siguiente producto interno:
$$\langle A, B \rangle = \sum_{i,j} a_{ij}b_{ij}.$$

Definición 32. Sea V un espacio vectorial con producto interno $\langle \cdot, \cdot \rangle$. Para todo $u, v \in V$:

1. u es **perpendicular** u **ortogonal a** v (y se nota $u \perp v$), si $\langle u, v \rangle = 0$.
2. La norma de u (inducida por $\langle \cdot, \cdot \rangle$) se nota $||u||$ y su valor es $||u|| = \sqrt{\langle u, u \rangle}$.

En cualquier espacio vectorial, la norma proveniente de un producto escalar cumple con las siguientes cuatro propiedades:

- $||x|| \geq 0$

Demostración. $||x|| = \sqrt{\langle x, x \rangle} \geq 0$. □

- $||x|| = 0 \iff x = 0$

Demostración. $||x|| = \sqrt{\langle x, x \rangle} = 0 \iff \langle x, x \rangle = 0 \iff x = 0$. □

- $||\alpha x|| = |\alpha| \cdot ||x||$

Demostración. $||\alpha x|| = \sqrt{\langle \alpha x, \alpha x \rangle} = \sqrt{\alpha \langle x, \alpha x \rangle} = \sqrt{\alpha \cdot \overline{\alpha} \langle x, x \rangle} = \sqrt{|\alpha|^2 \langle x, x \rangle} = |\alpha| \cdot ||x||$. □

- $|\langle x, y\rangle| \leq ||x|| \cdot ||y||$ (desigualdad de Cauchy-Schwartz)

Demostración. La igualdad vale si $y = 0$. Sea $y \neq 0$, $\alpha = \frac{\langle x,y\rangle}{||y||^2}$. Luego,

$$\begin{aligned}
0 \leq ||x-\alpha y||^2 &= \langle x-\alpha y, x-\alpha y\rangle \\
&= \langle x, x\rangle + \langle -\alpha y, x\rangle + \langle x, -\alpha y\rangle + \langle -\alpha y, -\alpha y\rangle \\
&= ||x||^2 - \alpha\langle y, x\rangle - \overline{\alpha}\langle x, y\rangle + |\alpha|^2||y||^2 \\
&= ||x||^2 - \alpha\overline{\langle x, y\rangle} - \overline{\alpha}\langle x, y\rangle + |\alpha|^2||y||^2 \\
&= ||x||^2 - \frac{\langle x, y\rangle}{||y||^2}\overline{\langle x, y\rangle} - \overline{\frac{\langle x, y\rangle}{||y||^2}}\langle x, y\rangle + |\alpha|^2||y||^2 \\
&= ||x||^2 - \frac{|\langle x, y\rangle|^2}{||y||^2} - \frac{|\langle x, y\rangle|^2}{||y||^2} + \left|\frac{\langle x, y\rangle}{||y||^2}\right|^2 ||y||^2 \\
&= ||x||^2 - \frac{|\langle x, y\rangle|^2}{||y||^2}
\end{aligned}$$

Es decir,

$$0 \leq ||x||^2 - \frac{|\langle x, y\rangle|^2}{||y||^2} \implies \frac{|\langle x, y\rangle|^2}{||y||^2} \leq ||x||^2 \implies |\langle x, y\rangle|^2 \leq ||x||^2 \cdot ||y||^2 \implies |\langle x, y\rangle| \leq ||x|| \cdot ||y||.$$

□

- $||x+y|| \leq ||x|| + ||y||$ (desigualdad triangular)

Demostración.

$$\begin{aligned}
||x+y||^2 &= \langle x, x\rangle + \langle x, y\rangle + \langle y, x\rangle + \langle y, y\rangle \\
&= ||x||^2 + 2 \cdot \mathfrak{Re}(\langle x, y\rangle) + ||y||^2 \\
&\leq ||x||^2 + 2 \cdot |\langle x, y\rangle| + ||y||^2 \\
&\leq ||x||^2 + 2 \cdot ||x|| \cdot ||y|| + ||y||^2 = (||x|| + ||y||)^2.
\end{aligned}$$

□

Definición 33. Sea V un espacio vectorial sobre $\mathbb{R}$, $x, y \in V$, $x \neq 0, y \neq 0$. Se define el **ángulo entre** x **e** y como

$$x\hat{\ }y = \arccos \frac{\langle x, y\rangle}{||x||\ ||y||}.$$

Observación 21. El ángulo entre dos elementos de V está bien definido por la desigualdad de Cauchy-Schwartz. Además, es consistente con el producto escalar para dos vectores en $\mathbb{R}^3$, es decir, $\langle x, y\rangle = ||x||\ ||y|| \cos(x\hat{\ }y)$.

Por otro lado, el ángulo (y en particular, la ortogonalidad) entre dos vectores no depende de su norma, es decir, si $x, y \in V$ y $\alpha, \beta \in \mathbb{R}$, entonces $|\cos(x\hat{\ }y)| = |\cos((\alpha x)\hat{\ }(\beta y))|$.

Lema 35. *Sea V un espacio vectorial con producto interno y $W \subset V$ un conjunto de vectores no nulos mutuamente ortogonales. Entonces, los vectores de W son linealmente independientes.*

Demostración. Considero una combinación lineal nula de los vectores $v^1, \ldots, v^k$ de W: $\sum_{i=1}^{k} \alpha_i v^i = 0$. Para $j \in \{1, \ldots, k\}$, realizo el producto escalar

$$\left\langle v^j, \sum_{i=1}^{k} \alpha_i v^i \right\rangle = \sum_{i=1}^{k} \overline{\alpha_i} \langle v^j, v^i \rangle.$$

Como $v^j \perp v^i$ para todo $i \neq j$, y como $v^j \neq 0$, tenemos que

$$\sum_{i=1}^{k} \overline{\alpha_i} \langle v^j, v^i \rangle = \overline{\alpha_j} \langle v^j, v^j \rangle = 0 \implies \alpha_j = 0.$$

Luego, como $\alpha_j = 0$ para todo $j \in \{1, \ldots, k\}$, los vectores de W son l.i.. □

Observación 22. Si V es un espacio con dimensión infinita, W en el lema anterior puede ser un conjunto infinito de vectores ortogonales.

Ejemplo: Sea V el espacio vectorial de las funciones reales continuas en el intervalo $[0, 1]$ con el producto interno $\langle f, g \rangle = \int_0^1 f(t)g(t)dt$. Ahora, para cada $n \in \mathbb{N}$, sean

$$f_n(x) = \sqrt{2}\cos(2\pi n x), \quad g_n(x) = \sqrt{2}\sin(2\pi n x).$$

Puede probarse que el conjunto $\{f_n : n \in \mathbb{N}\} \cup \{g_n : n \in \mathbb{N}\}$ es un conjunto infinito de vectores ortogonales de V. *(Fin del ejemplo)*

Si V es un espacio vectorial con producto interno $\langle \cdot, \cdot \rangle$ y $\mathcal{B} = \{u^1, \ldots, u^k\}$ es una base de vectores mutuamente ortogonales, le descomposición de cualquier vector en $\mathcal{B}$ es sencilla de calcular.

Sea $v = \sum_{i=1}^{k} \alpha_i u^i$. Luego, realizando el producto interno de v con cualquier elemento de $\mathcal{B}$,

$$\langle v, u^j \rangle = \sum_{i=1}^{k} \alpha_j \langle u^i, u^j \rangle = \alpha_j \langle u^j, u^j \rangle = \alpha_j ||u^j||^2.$$

Entonces, $\alpha_j = \frac{\langle v, u^j \rangle}{||u^j||^2}$ para $j = 1, \ldots, k$. Nótese que si los vectores de la base tienen norma 1,

$$v = \sum_{i=1}^{k} \langle v, u^i \rangle u^i.$$

Definición 34. Dado un espacio vectorial V con producto interno, una base de V es **ortogonal** si sus vectores son mutuamente ortogonales, y es **ortonormal** si es base ortogonal y sus vectores tienen norma 1.

Con el ánimo de simplificar cálculos, dada una base ortogonal $\mathcal{B}$ de un e.v. V, se puede definir $\mathcal{B}' = \left\{ \frac{w^j}{||w^j||} : w^j \in \mathcal{B} \right\}$, la cual resulta ser una base ortonormal de V.

Observación 23. Cualquier rotación de los vectores de la base canónica en $\mathbb{R}^n$ conforman también una base ortonormal en $\mathbb{R}^n$. Así, en particular, dado un ángulo θ, los vectores $(\cos\theta, \sin\theta)$ y $(-\sin\theta, \cos\theta)$ conforman una base ortonormal en $\mathbb{R}^2$.

1. Subespacios Ortogonales

Definición 35. Sea un espacio vectorial V con producto interno y dos subespacios W_1 y W_2 de V. Se dice que W_1 y W_2 son **subespacios ortogonales** si para todo $u \in W_1$ y todo $v \in W_2$, se tiene que $u \perp v$.

De esta definición se deduce, por ejemplo, que el subespacio nulo es ortogonal a cualquier otro subespacio, y que dos planos no pueden ser subespacios ortogonales entre sí.

Lema 36. *(Condición necesaria) Sea V un espacio vectorial con producto interno, y W_1 y W_2 subespacios ortogonales de V. Entonces, $W_1 \cap W_2 = \{0\}$.*

Demostración. Supongo que existe $u \in W_1 \cap W_2,\ u \neq 0$. Como W_1 y W_2 son ortogonales, $v \perp w$ para todo $v \in W_1, w \in W_2$. En particular, tomando $v = u \in W_1$ y $w = u \in W_2$, debe verificarse $u \perp u$, sin embargo, $\langle u, u\rangle = ||u||^2 \neq 0$. Contradicción. Luego, $W_1 \cap W_2 = \{0\}$. $\square$

Lema 37. *(Condición suficiente) Sea V un espacio vectorial con producto interno, $W_1 = \langle U_1 \rangle$ y $W_2 = \langle U_2 \rangle$ dos subespacios de V con $U_1 = \{u^1, \ldots, u^{k_1}\}$ y $U_2 = \{v^1, \ldots, v^{k_2}\}$. Entonces, W_1 y W_2 son subespacios ortogonales si y solo si $u \perp w$ para todo $u \in U_1$ y todo $w \in U_2$.*

Demostración. **Ida.** Como $W_1 \perp W_2$ cualquier elemento de W_1 es ortogonal a cualquier elemento de W_2. En particular, dado $j \in \{1, \ldots, k_1\}$, $u^j \perp v^k$ para todo $k \in \{1, \ldots, k_2\}$.

Vuelta. Sean $u = \sum_{i=1}^{k_1} \alpha_i u^i \in W_1$, $v = \sum_{i=1}^{k_2} \beta_i v^i \in W_2$. Luego,

$$\langle u, v\rangle = \sum_{i=1}^{k_1} \alpha_i \langle u^i, v\rangle = \sum_{i=1}^{k_1} \alpha_i \langle u^i, \sum_{i=1}^{k_2} \beta_i v^i \rangle = \sum_{i=1}^{k_1} \sum_{i=1}^{k_2} \overline{\beta_i} \alpha_i \langle u^i, v^i\rangle = 0,$$

pues, por hipótesis, $u \perp w$ para todo $u \in U_1$ y todo $w \in U_2$. Luego, $W_1 \perp W_2$. $\square$

Teorema 8. *Sea A una matriz $m \times n$. Entonces:*

1. *El espacio fila y el espacio nulo de A son subespacios ortogonales de $\mathbb{R}^n$. ($C(A^T) \perp N(A)$)*
2. *El espacio columna y el espacio nulo a izquierda de A son subespacios ortogonales de $\mathbb{R}^m$. ($C(A) \perp N(A^T)$)*

Demostración.

1. Sea $v \in N(A)$. Como $Av = 0$, para todo $i = 1, \ldots m$ resulta $A_i v = 0$. Por lo tanto, para todo $i = 1, \ldots, m$, $A_i \perp v$. Luego, como las filas de A generan el espacio fila, los espacios son ortogonales.

2. Sea $v \in N(A^T)$. Como $A^T v = 0 \iff v^T A = 0$, para todo $i = 1, \ldots, n$ resulta $v^T A^i = 0$. Luego, para todo $i = 1, \ldots, n$, $v^T \perp A^i$, y como las columnas de A generan $C(A)$, los espacios son ortogonales.

□

Lema 38. *Sea V un espacio vectorial con producto interno y W_1, W_2 dos subespacios ortogonales. Sean U_1 y U_2 conjuntos de vectores l.i. de W_1 y W_2, respectivamente. Entonces, $U_1 \cup U_2$ es un conjunto de vectores l.i. de V.*

Demostración. Sean $U_1 = \{u_1^i : i = 1, \ldots, k\}$ y $U_2 = \{u_2^j : j = 1, \ldots, t\}$. Considero una combinación lineal nula de vectores de $U = U_1 \cup U_2$:

$$\sum_{i=1}^{k} \alpha_i u_1^i + \sum_{j=1}^{t} \beta_i u_2^j = 0.$$

Sea $w = \sum\limits_{i=1}^{k} \alpha_i u_1^i = \sum_{j=1}^{t} (-\beta_i) u_2^j$. Nótese que $w \in \langle U_1 \rangle \cap \langle U_2 \rangle$. Luego, como W_1 y W_2 son ortogonales y por ende $W_1 \cap W_2 = \{0\}$,

$$w = \sum_{i=1}^{k} \alpha_i u_1^i = \sum_{j=1}^{t} (-\beta_i) u_2^j = 0.$$

Como los vectores de U_1 y de U_2 son l.i., resulta $\alpha_i = 0$ para todo $i = 1, \ldots, k$ y $\beta_j = 0$ para todo $j = 1, \ldots, t$. □

Teorema 9. *Sea V un espacio vectorial con producto interno y sean W_1 y W_2 subespacios ortogonales de V. Entonces, $\dim(W_1) + \dim(W_2) \leq \dim(V)$.*

Demostración. Para $i = 1, 2$, sea $\mathcal{B}_i$ una base de W_i. Como son espacios ortogonales, $\mathcal{B}_1 \cap \mathcal{B}_2 = \emptyset$ (pues ningún vector no nulo es ortogonal a sí mismo). Por el lema anterior, $\mathcal{B}_1 \cup \mathcal{B}_2$ es un conjunto de vectores l.i.. Por lo tanto, $\dim(V) \geq |\mathcal{B}_1 \cup \mathcal{B}_2| = |\mathcal{B}_1| + |\mathcal{B}_2| = \dim(W_1) + \dim(W_2)$. □

1.1. Teorema Fundamental del Álgebra Lineal - Segunda Parte

De los lemas y teoremas anteriores se inicia la idea de "complemento ortogonal" de un subespacio:

Definición 36. Dado un espacio vectorial V producto interno, y W un subespacio de V, se denomina **complemento ortogonal** de W (y se nota $W^\perp$) al subespacio de V determinado por todos los vectores que son ortogonales a todos los vectores de W. Es decir,

$$W^\perp = \{u \in V : u \perp v \quad \forall v \in W\}.$$

Definición 37. Algunas propiedades que se deducen de la definición son:

1. $W^\perp$ es un subespacio vectorial.

   Demostración. Dados $u, v \in W^\perp$, $\alpha, \beta \in \mathbb{K}$ y $w \in W$, $\langle \alpha u + \beta v, w\rangle = \alpha\langle u, w\rangle + \beta\langle v, w\rangle = 0 + 0 = 0$, por lo que $\alpha u + \beta v \in W^\perp$, y luego es un subespacio vectorial. □

2. W y $W^\perp$ son espacios ortogonales.

   Demostración. Se deduce de la definición de $W^\perp$; cualquier elemento de $W^\perp$ es perpendicular a todo elemento de W. □

3. $(W^\perp)^\perp = W$.

   Demostración. $W \subset (W^\perp)^\perp$, ya que, dado $w \in W$, sabemos que $w \perp v$ $\forall v \in W^\perp$, por lo que $w \in (W^\perp)^\perp$.

   Ahora, sea $x \in (W^\perp)^\perp$. Entonces, $x = u + v$, donde $u \in W$ y $v \in W^\perp$ (por el Teorema de Descomposición). Tomando el producto interno de x con v, como $x \perp v$, $\langle x, v\rangle = 0$. Además,

   $$\langle x, v\rangle = \langle u + v, v\rangle = \langle u, v\rangle + \langle v, v\rangle = 0 + ||v||^2 = ||v||^2 = 0 \iff v = 0.$$

   Entonces, $x = u + v = u + 0 + u$, con $u \in W$. Es decir, $x \in W$, por lo que $(W^\perp)^\perp \subset W$, con lo que llegamos a que $(W^\perp)^\perp = W$. □

Teorema 10. *Sea V un espacio vectorial de dimensión n, con producto interno y sean W_1 y W_2 subespacios ortogonales de V. Entonces,*

$$\dim(W_1) + \dim(W_2) = n \iff W_2 = W_1^\perp.$$

Demostración. **Ida.** Es trivial que $W_2 \subset W_1^\perp$, pues W_1 y W_2 son subespacios ortogonales. Ahora, supongo que existe $u \in W_1^\perp \setminus W_2$. Entonces, $u \neq 0$ (porque W_2 es un subespacio vectorial). Además $u \notin W_1$ ya que $W_1 \perp W_1^\perp$ y por ende $W_1 \cap W_2 = \{0\}$. Entonces, se puede construir un conjunto de $n+1$ vectores l.i. con u y las bases de W_1 y W_2. Contradicción, al ser n la dimensión de v. Entonces, $W_1^\perp \setminus W_2 = \emptyset$, y luego $W_1^\perp \subset W_2$. Por lo tanto, $W_2 = W_1^\perp$.

Vuelta. (Esta prueba requiere que W_1 tenga una base ortogonal. Por el proceso de ortogonalización de Gram-Schmidt, esta prueba es válida para cualquier W_1). Sean W_1 y W_2 tales que $W_2 = W_1^\perp$. Como son ortogonales, $\dim(W_1) + \dim(W_2) \leq \dim(V)$. Más aún, si $\mathcal{B}_i$ es una base de W_i para $i = 1, 2$, entonces $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2$ es un conjunto de vectores l.i.. Sea $\mathcal{B}_1 = \{w^1, \ldots, w^p\}$ una base ortogonal de W_1 y, con $\mathcal{B}_2$ base de $\mathcal{B}_2$, sea $W = \langle \mathcal{B}\rangle$.

Supongo que $\dim(W_1)+\dim(W_2)\leq \dim(V)-1$. Entonces, existe $v\in V\setminus W$. Construyo

$$z = v - \sum_{i=1}^{p} \frac{\langle v, w^i\rangle}{||w^i||^2} w^i.$$

Se verá que $z\perp w^j$ para todo $j=1,\dots,p$. Dado $j\in\{1,\dots,p\}$,

$$\langle z, w^j\rangle = \left\langle v - \sum_{i=1}^{p} \frac{\langle v, w^i\rangle}{||w^i||^2} w^i, w^j \right\rangle = \langle v, w^j\rangle - \sum_{i=1}^{p} \frac{\langle v, w^i\rangle}{||w^i||^2} \langle w^i, w^j\rangle.$$

Como $\mathcal{B}_1$ es una base ortogonal de W_1, $\langle w^i, w^j\rangle = 0$ si $i\neq j$. Por lo tanto,

$$\langle z, w^j\rangle = \langle v, w^j\rangle - \frac{\langle v, w^j\rangle}{||w^j||^2}\langle w^j, w^j\rangle = 0.$$

Como $z\in V\setminus W$ y $0\in W_1$, $z\neq 0$. Además, como $z\perp w^i$ para todo $i=1,\dots,p$, resulta que $z\perp w$ para todo $w\in W_1$. Luego, $z\in W_1^{\perp}=W_2$. Contradicción, pues $W_2\subset W$ (ya que $\mathcal{B}_2\subset\mathcal{B}$). Por lo tanto, $\dim(W_1)+\dim(W_2)=\dim(V)$. □

Teorema 11. ***(Teorema Fundamental del Álgebra Lineal - Parte II)*** *Sea A una matriz $m\times n$. Entonces:*

1. *El espacio fila y el espacio nulo de A son complementos ortogonales en $\mathbb{R}^n$. ($C(A^T)=N(A)^{\perp}$)*

2. *El espacio columna y el espacio nulo a izquierda de A son complementos ortogonales en $\mathbb{R}^m$. ($C(A)=N(A^T)^{\perp}$)*

Demostración.

1. Sea $x\in N(A)$. Luego, $Ax=0$, es decir, $(Ax)_i = A_i x = 0$ para todo $i=1,\dots,m$. Entonces, los vectores fila de A son ortogonales a x. Entonces, x es ortogonal a cualquier combinación de las filas de A, y por ende $x\perp v$ para todo $v\in C(A^T)$. Es decir, $C(A^T)=N(A)^{\perp}$.

2. Sea $y\in N(A^T)$. Luego, $y^TA=0$; es decir, $(y^TA)^i = y^TA^i = 0$ para todo $i=1,\dots,n$. Entonces, y es ortogonal a todas las columnas de A, por lo que será ortogonal a cualquier combinación lineal de ellas, y por ende $y\perp v$ para todo $v\in C(A)$. Es decir, $C(A)=N(A^T)^{\perp}$.

□

Como consecuencia del lema anterior, se tiene que dos subespacios que son complementos ortogonales descomponen al espacio vectorial de la siguiente manera:

Teorema 12. ***(Teorema de Descomposición)*** *Sea V un espacio vectorial de dimensión finita n, con producto interno, y sea W un subespacio de V. Entonces, para todo $v\in V$ existen únicos $w\in W$ y $w^{\perp}\in W^{\perp}$ tales que $v=w+w^{\perp}$.*

Demostración. Sea $\mathcal{B}_1 = \{v^1, \dots v^k\}$ una base de W y $\mathcal{B}_2 = \{w^1, \dots w^{n-k}\}$ una base de $W^\perp$. Luego, como los vectores de $\mathcal{B}_1 \cup \mathcal{B}_2$ son l.i. y $|\mathcal{B}_1| + |\mathcal{B}_2| = n$, entonces $\mathcal{B}_1 \cup \mathcal{B}_2$ es una base de V. Sea $v \in V$ tal que

$$v = \sum_{i=1}^{k} \alpha_i v^i + \sum_{j=1}^{n-k} \beta_i w^j.$$

Entonces, si $w = \sum_{i=1}^{k} \alpha_i v^i$ y $w^\perp = \sum_{j=1}^{n-k} \beta_i w^j$, tenemos que $v = w + w^\perp$.

Ahora, para probar su unicidad, sean $z \in W$ y $z^\perp \in W^\perp$ tales que $v = z + z^\perp = w + w^\perp$. Entonces, $w - z = z^\perp - w^\perp$. Como $w - z \in W$ y $z^\perp - w^\perp \in W^\perp$, resulta que $w - z = z^\perp - w^\perp = 0 \in W \cap W^\perp$, y por lo tanto $z = w$ y $z^\perp = w^\perp$. $\square$

Corolario 8. *Sea A una matriz $m \times n$. Entonces:*

1. *Para todo $x \in \mathbb{R}^n$ existen únicos $x_F \in C(A^T)$ y $x_N \in N(A)$ tales que $x = x_F + x_N$ ($\mathbb{R}^n = C(A^T) \oplus N(A)$).*

2. *Para todo $y \in \mathbb{R}^n$ existen únicos $y_C \in C(A)$ y $y_I \in N(A^T)$ tales que $y = y_C + y_I$ ($\mathbb{R}^m = C(A) \oplus N(A^T)$).*

Teniendo en cuenta esto último, dado $x \in \mathbb{R}^n$, sabemos que existen únicos $x_F \in C(A^T)$ y $x_N \in N(A)$ tal que $x = x_F + x_N$. Por lo tanto,

$$Ax = Ax_F + Ax_N + Ax_F + 0 = Ax_F \in C(A) \subset \mathbb{R}^m.$$

Es decir, la transformación lineal definida por A lleva todo el espacio fila de A en su espacio columna, y el espacio nulo en $\{0\}$.

Lema 39. *Para toda matriz A $m \times n$, A define un isomorfismo entre su espacio fila y su espacio columna.*

Demostración. Como tienen la misma dimensión (el rango de la matriz) sabemos que $C(A^T)$ y $C(A)$ son isomorfos. Sea $T : C(A^T) \longrightarrow C(A)$ la transformación lineal dada por $T(x) = Ax$. Veré que es biyectiva.

Sea $b \in C(A)$. Sabemos que existe al menos un $x \in \mathbb{R}^n$ tal que $Ax = b$. Ahora, como $\mathbb{R}^n = C(A^T) \oplus N(A)$, existen $x_F \in C(A^T)$ y $x_N \in N(A)$ tales que $x = x_F + x_N$. Entonces, $Ax = A(x_F + x_N) = Ax_F + Ax_N = Ax_F + 0 = Ax_F = b$. Es decir, se encontró $x_F \in C(A^T)$ tal que $T(x_F) = Ax_F = b$, por lo que T es sobreyectiva.

Por otro lado, $N(A) = \{x \in \mathbb{R}^m : Ax = 0\}$. Sea $x \in N(A)$. Pero además, como para que x pertenezca al dominio debe pertenecer al espacio fila, resulta que $x \in N(A) \cap C(A^T)$, y como son espacios ortogonales, $x = 0$. Entonces, $N(A) = \{0\}$, y, por lema, T es un inyectiva, y por lo tanto, tenemos que T es un isomorfismo. $\square$

1.2. Proyección sobre Subespacios

Dados W_1 y W_2 complementos ortogonales en V, con $\mathcal{B}_1 = \{w^i : i = 1, \ldots, p\}$ y $\mathcal{B}_2 = \{z^i : i = 1, \ldots, k\}$ bases ortogonales de W_1 y W_2, respectivamente. Entonces, por el Teorema de Descomposición, para todo $v \in V,\ v = v_1 + v_2$, con $v_1 \in W_1$ y $v_2 \in W_2$. Además,

$$\begin{aligned} v_1 &= \sum_{i=1}^{p} \frac{\langle v_1, w^i\rangle}{||w^i||^2} w^i = \sum_{i=1}^{p} \left(\frac{\langle v_1, w^i\rangle}{||w^i||^2} w^i + \frac{\langle v_2, w^i\rangle}{||w^i||^2} w^i \right) \\ &= \sum_{i=1}^{p} \frac{\langle v_1 + v_2, w^i\rangle}{||w^i||^2} w^i = \sum_{i=1}^{p} \frac{\langle v, w^i\rangle}{||w^i||^2} w^i = \sum_{i=1}^{p} \frac{||v||}{||w^i||} \cos(v \hat{\ } w^i) w^i, \end{aligned}$$

y con el mismo razonamiento,

$$v_2 = \sum_{j=1}^{k} \langle v, z^j\rangle z^j = \sum_{j=1}^{k} \frac{||v||}{||z^j||} \cos(v \hat{\ } z^j) z^j.$$

Observación 24. Si las bases son ortonormales,

$$v = v_1 + v_2 = \sum_{i=1}^{p} ||v|| \cos(v \hat{\ } w^i) w^i + \sum_{j=1}^{k} ||v|| \cos(v \hat{\ } z^j) z^j.$$

Definición 38. Sea V un espacio vectorial con producto interno y W un subespacio de V con base ortogonal $\{w^1, \ldots, w^p\}$. Para todo $v \in V$, el **vector proyección de** v **sobre** W es

$$proy_{s/W} v = \sum_{i=1}^{p} \frac{||v||}{||w^i||} \cos(v \hat{\ } w^i) w^i = \sum_{i=1}^{p} \frac{\langle v, w^i\rangle}{||w^i||^2} w^i.$$

De esta manera, la descomposición de un vector $v \in V$ es $v = proy_{s/W} v + proy_{s/W^\perp} v$.

Observación 25. Si la base es ortonormal,

$$proy_{s/W} v = \sum_{i=1}^{p} ||v|| \cos(v \hat{\ } w^i) w^i = \sum_{i=1}^{p} \langle v, w^i\rangle w^i.$$

2. Proceso de Ortogonalización de Gram-Schmidt

Sea V un espacio vectorial con producto interno. Luego, el proceso de ortogonalización de Gram-Schmidt genera p vectores ortogonales no nulos en V, a partir de p vectores l.i. de V.

Nombremos $\{v^i : i = 1, \ldots, p\}$ al conjunto de los p vectores l.i. de V que constituyen la entrada del algoritmo, y $\{w^i : i = 1, \ldots, p\}$ a los p vectores ortogonales no nulos en V resultantes del proceso. Los pasos son los siguientes:

1. $w^1 := v^1$.

2. Para $j = 2, \ldots, p$, hacer
$$w^j = v^j - \sum_{i=1}^{j-1} \frac{\langle v^j, w^i \rangle}{||w^i||^2} w^i.$$

Lema 40. *Dados los conjuntos especificados anteriormente, para todo $i = 1, \ldots, p$, $w^i \neq 0$ y para todo $i \neq j$, $w^i \perp w^j$.*

Demostración. Como los vectores de $\{v^i : i = 1, \ldots, p\}$ son l.i., $v^1 \neq 0$ para todo $i = 1, \ldots, p$. Además, es suficiente probar que, para todo $i = 2, \ldots p$, $w^i \neq 0$ y $w^i \perp w^j$ $\forall j : 1 \leq j \leq i - 1$.

Se prueba por inducción lo siguiente: Para todo $i = 2, \ldots, p$,

1. w^i es una combinación lineal de v^j, $1 \leq j \leq i$;

2. $w^i \neq 0$,

3. $w^i \perp w^j$ $\forall j : 1 \leq j \leq i - 1$.

Considero el caso base $i = 2$. Efectivamente,

$$w^2 = v^2 - \frac{\langle v^2, w^1 \rangle}{||w^1||^2} w^1 = v^2 - \frac{\langle v^2, w^1 \rangle}{||w^1||^2} v^1.$$

Es decir, w^2 es una combinación lineal de v^1 y v^2, y por la lineal independencia de v^1 y v^2 podemos concluir que $w^2 \neq 0$. Por otro lado,

$$\langle w^2, w^1 \rangle = \langle w^2, v^1 \rangle = \langle v^2, v^1 \rangle - \frac{\langle v^2, v^1 \rangle}{||v^1||^2} \langle v^1, v^1 \rangle = 0,$$

por lo que también se verifica que $w^2 \perp w^1$.

Ahora, sea k tal que $2 \leq k \leq p - 1$ y w^j verifica las tres condiciones del enunciado para todo $j \leq k$. Sabemos que

$$w^{k+1} = v^{k+1} - \sum_{i=1}^{k} \frac{\langle v^{k+1}, w^i \rangle}{||w^i||^2} w^i.$$

Por hipótesis de inducción, para todo $i = 1, \ldots, k$, w^i es combinación lineal de v^t, $t = 1, \ldots, i$. Por lo tanto, w^{k+1} es combinación lineal de v^t, $t = 1, \ldots, k+1$. Además, la lineal independencia de v^i, $i = 1, \ldots, k+1$ asegura que $w^{k+1} \neq 0$. Solo falta probar que $w^{k+1} \perp w^j$ $\forall j = 1, \ldots, k$. Dado $j \in \{1, \ldots, k\}$,

$$\langle w^{k+1}, w^j \rangle = \langle v^{k+1}, w^j \rangle - \sum_{i=1}^{k} \frac{\langle v^{k+1}, w^i \rangle}{||w^i||^2} \langle w^i, w^j \rangle.$$

Por hipótesis de inducción, $\langle w^i, w^j \rangle = 0$ para todo $1 \leq j \leq k$, $1 \leq i \leq k - 1$. Por lo tanto,

$$\langle w^{k+1}, w^j \rangle = \langle v^{k+1}, w^j \rangle - \langle v^{k+1}, w^j \rangle = 0.$$

□

Nótese que $\sum_{i=1}^{j-1} \frac{\langle v^j, w^i \rangle}{||w^i||^2} w^i$ es la proyección de v^j sobre el subespacio $W^{(j-1)}$ generado por los vectores ortogonales $w^1, \ldots, w^{j-1}$. Entonces, $w^j = v^j - proy_{s/W^{(j-1)}} v^j$ es el **vector error** en la aproximación de v^j en $W^{(j-1)}$. También es la componente de v^j sobre el complemento ortogonal de $W^{(j-1)}$.

Teorema 13. *Todo espacio vectorial con producto interno y de dimensión finita tiene una base ortonormal.*

Demostración. Dada una base B de un espacio vectorial V de dimensión finita n, podemos aplicarle el proceso de ortogonalización de Gram-Schmidt a B, el cual resulta en un nuevo conjunto B' de n vectores ortogonales. Como son ortogonales, son l.i., y al ser $|B'| = n$, tenemos que B' es una base ortogonal de V. Basta considerar el conjunto $B'' = \{\frac{v}{||v||} : v \in B'\}$, el cual resulta ser una base ortonormal de V. □

Definición 39. Dado un espacio vectorial V con producto interno, W un subespacio de V y $v \in V$, se dice que $\tilde{w}$ es una **mejor aproximación de** v **en** W si, para todo $w \in W$, $||\tilde{w} - v|| \leq ||w - v||$.

Ejemplo: Para conocer la velocidad (constante) v con la que un satélite viaja hacia Marte, se realizan $k+1$ mediciones en tiempos $t_0 = 0 < t_1 < \cdots < t_k$ y se obtienen las distancias $b_i, i = 0, \ldots, k$. Claramente, v debe verificar $t_i v = b_i - b_0,\ i = 1, \ldots, k$.

Aunque sea un sistema con k ecuaciones y una variable, los errores de medición permiten que no se pueda asegurar que el sistema tenga solución.

Sea $T = (t_1, \ldots, t_k)^T$ y $b = (b_1 - b_0, \ldots, b_k - b_0)^T$. Considerando a T como una matriz $k \times 1$ y $b \in \mathbb{R}^k$, estamos ante un sistema de la forma $Tv = b$, con $v \in \mathbb{R}$. Si $b \in C(T)$, el sistema tiene solución. En caso contrario, se busca la mejor aproximación de b en $C(T)$, es decir, se busca $\tilde{b} \in C(T)$ tal que $||b - \tilde{b}||$ sea mínima, y luego se elige v tal que $Tv = \tilde{b}$.

Como $\tilde{b} \in C(T)$, existe un $\alpha \in \mathbb{R}$ tal que $\tilde{b} = \alpha T$. Ahora, si se hacen cuatro mediciones ($k = 4-1 = 3$), $C(T)$ es una recta en $\mathbb{R}^3$ que pasa por el origen, y $b \in \mathbb{R}^3$ es un punto fuera de la recta. El punto más cercano a b en la recta (la mejor aproximación) será el pie de la perpendicular a la recta que pasa por b. Es decir, se busca $\tilde{b}$ tal que $(b - \tilde{b}) \perp T$. Entonces:

$$\langle b - \tilde{b}, T \rangle = \langle b - \alpha T, T \rangle = \langle b, T \rangle - \langle \alpha T, T \rangle = \langle b, T \rangle - \alpha \langle T, T \rangle = 0 \implies \alpha = \frac{\langle b, T \rangle}{||T||^2}.$$

Luego, $\tilde{b} = \frac{\langle b, T \rangle}{||T||^2} T = proy_{s/C(T)} b$. *(Fin del ejemplo)*

Teorema 14. *Sea W un subespacio vectorial del e.v. V sobre $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ con producto interno. Sea $v \in V$. Entonces:*

1. *$\tilde{w} \in W$ es una mejor aproximación de v en W si y solo si $v - \tilde{w}$ es ortogonal a W.*

2. Si existe una mejor aproximación de v en W, entonces es única.

3. Si W es de dimensión finita y $\{z^i : i = 1, \ldots, k\}$ es una base ortonormal de W, entonces el vector

$$proy_{s/W}v = \sum_{i=1}^{k} \langle v, z^i \rangle z^i$$

es la única mejor aproximación a v en W.

Demostración.

1. Para todo $w \in W$ se verifica

$$\begin{aligned} ||w - v||^2 &= ||(w - \bar{w}) + (\bar{w} - v)||^2 = \langle (w - \bar{w}) + (\bar{w} - v), (w - \bar{w}) + (\bar{w} - v) \rangle \\ &= ||w - \bar{w}||^2 + 2\,\mathfrak{Re}(\langle w - \bar{w}, \bar{w} - v \rangle) + ||\bar{w} - v||^2 \end{aligned}$$

Sea $\bar{w} \in W$ tal que $(v - \bar{w}) \perp W$. Entonces, $||w - v||^2 = ||w - \bar{w}||^2 + ||\bar{w} - v||^2 \geq ||\bar{w} - v||^2$. Por lo tanto, $\bar{w}$ es una mejor aproximación de v en W.

Sea ahora $\bar{w} \in W$ tal que, para todo $w \in W$, $||w - v|| \geq ||\bar{w} - v||$. Debemos probar que $(v - \bar{w}) \perp W$, o, equivalentemente, $\langle v - \bar{w}, w \rangle = 0$ para todo $w \in W$.

Como para todo $w \in W$, $||w - v||^2 = ||w - \bar{w}||^2 + 2\,\mathfrak{Re}(\langle w - \bar{w}, \bar{w} - v \rangle) + ||\bar{w} - v||^2$, y $||w - v|| \geq ||\bar{w} - v||$, obtenemos

$$||\bar{w} - w||^2 + 2\,\mathfrak{Re}(\langle w - \bar{w}, \bar{w} - v \rangle) \geq 0,$$

o, equivalentemente, para todo $u \in W$,

$$||u||^2 + 2\,\mathfrak{Re}(\langle u, \bar{w} - v \rangle) \geq 0.$$

Esto sucede porque cualquier vector $u \in W$ puede ser expresado como la resta de otro vector $u' \in W$ y $\bar{w}$ (basta tomar $u' = u + \bar{w} \in W$). Supongamos que existe $u^* \in W$ tal que $\langle u^*, \bar{w} - v \rangle \neq 0$. Sin pérdida de generalidad, supongo que $\mathfrak{Re}(\langle u^*, \bar{w} - v \rangle) < 0$. Nuevamente, esto sucede porque $\mathfrak{Re}(\langle u^*, \bar{w} - v \rangle) = 0 \implies \mathfrak{Im}(\langle u^*, \bar{w} - v \rangle) \neq 0$. Luego, tomando $u^{**} = iu^*$, se tiene que $\mathfrak{Re}(\langle u^{**}, \bar{w} - v \rangle) \neq 0$, y si $\mathfrak{Re}(\langle u^{**}, \bar{w} - v \rangle) > 0$, tomando $u^{***} = -u^{**}$ nos aseguramos que $\mathfrak{Re}(\langle u^{***}, \bar{w} - v \rangle) < 0$. Sea

$$\alpha = \frac{\mathfrak{Re}(\langle u^*, v - \bar{w} \rangle)}{||u^*||^2} > 0,$$

y consideremos $\frac{\alpha}{2}u^* \in W$. Entonces, por lo anterior vale

$$\begin{aligned} \left|\left|\frac{\alpha}{2}u^*\right|\right|^2 + 2\,\mathfrak{Re}\left(\left\langle \frac{\alpha}{2}u^*, \bar{w} - v \right\rangle\right) \geq 0 &\iff \frac{\alpha}{4}||u||^2 - \mathfrak{Re}\left(\langle u^*, v - \bar{w} \rangle\right) \geq 0 \\ &\iff \frac{\alpha}{4} \geq \frac{\mathfrak{Re}\left(\langle u^*, v - \bar{w} \rangle\right)}{||u||^2} = \alpha, \end{aligned}$$

una contradicción. Entonces, $\forall u \in W,\ \langle u, \bar{w} - v \rangle = 0$, y $(\bar{w} - v) \perp W$.

2. Sean $z, \bar{w} \in W$ mejores aproximaciones de v en W. Entonces, $\bar{w} - z \in W$. Además, $\bar{w} - z = (\bar{w} - v) + (v - z)$. Como z y $\bar{w}$ son mejores aproximaciones de v en W, por **1.** sabemos que $(\bar{w} - v) \in W^{\perp}$ y $(v - z) \in W^{\perp}$. Por lo tanto, $\bar{w} - z \in W^{\perp}$. Tenemos, entonces, que $\bar{w} - z \in W \cap W^{\perp}$, por lo que $\bar{w} - z = 0 \iff \bar{w} = z$.

3. Por los ítems anteriores, solo falta probar que

$$\bar{w} = \sum_{i=1}^{k} \langle v, z^i \rangle z^i$$

verifica $(v - \bar{w}) \perp W$. Para ello, basta probar $\langle v - \bar{w}, z^j \rangle = 0$ para todo $j = 1, \ldots, n$. Efectivamente, para todo $j = 1, \ldots, n$,

$$\langle \bar{w}, z^j \rangle = \sum_{i=1}^{k} \langle v, z^i \rangle \langle z^i, z^j \rangle = \langle v, z^j \rangle \langle z^j, z^j \rangle = \langle v, z^j \rangle,$$

y luego, para todo $j = 1, \ldots, n$,

$$\langle v - \bar{w}, z^j \rangle = \langle v, z^j \rangle - \langle \bar{w}, z^j \rangle = \langle v, z^j \rangle - \langle v, z^j \rangle = 0,$$

resultando $\bar{w}$ la mejor aproximación de v en W.

□

3. Matrices de proyección

Generalmente, si se tiene un sistema inconsistente $Ax = b$ con A matriz 3×2 y $b \in \mathbb{R}^2$, y se quiere encontrar la mejor solución $\hat{x}$ del sistema, se busca la mejor aproximación $\tilde{b}$ de b sobre $C(A)$.

Por otro lado, la proyección de un vector $b \in \mathbb{R}^n$ sobre la recta generada por otro vector $a \in \mathbb{R}^n$ es

$$proy_{s/<a>} b = \frac{a^T b}{a^T a} a = Pb,$$

donde $P = \frac{1}{a^T a} a a^T$. Observar que $(a^T b)a = a(a^T b) = (a a^T)b$. Debido a esto, es posible ver que la proyección de un vector de $\mathbb{R}^n$ sobre una recta puede ser pensada como una transformación lineal de $\mathbb{R}^n$ en $\mathbb{R}^n$. Además, tenemos que $P^2 = P$ y P es simétrica.

Ahora, vemos las proyecciones sobre un subespacio general W de $\mathbb{R}^n$. Sea $\{a^i : i = 1, \ldots, k\}$ una base de W, y sea A la matriz $n \times k$ cuya columna i-ésima es a^i tal que $W = C(A)$. Luego, dado $b \in \mathbb{R}^n$, se busca $\hat{b}$, la proyección de b en $W = C(A)$. Sabemos que existe un $\hat{x} \in \mathbb{R}^n$ tal que $A\hat{x} = \hat{b}$.

También, sabemos que la proyección es la mejor aproximación de b en W, y luego el error $e = b - A\hat{x}$

es perpendicular a $C(A)$. Por TFDAL, $b - A\hat{x} \in N(A^T)$. Entonces,

$$A^T(b - A\hat{x}) = 0 \implies A^Tb - A^T(A\hat{x}) = 0 \implies (A^TA)\hat{x} = A^Tb.$$

Es decir, si el sistema $Ax = b$ es inconsistente, la mejor solución $\hat{x}$ que minimiza el error es solución del sistema de ecuaciones normales $(A^TA)\hat{x} = A^Tb$. Esta solución es única cuando A^TA es inversible, y en dicho caso está dada por

$$\hat{x} = (A^TA)^{-1}A^Tb,$$

y la proyección $\hat{b}$ de b en $C(A)$ es $\hat{b} = A\hat{x} = A(A^TA)^{-1}A^Tb$.

Se puede ver también que la matriz de proyección $P = A(A^TA)^{-1}A^T$ define a la proyección como una transformación lineal, y que como $Pb \in C(A)$ y $b - Pb \in N(A^T)$, P da la componente de b en el espacio $C(A)$ y $b - Pb$ es la componente sobre su complemento ortogonal ($N(A^T)$).

Lema 41. *Sea A una matriz $m \times n$ cuyas columnas son l.i.. Entonces, A^TA es inversible.*

Demostración. Considero el sistema $A^TAx = 0$. Nótese que $Ax \in C(A)$. Además, $A^T(Ax) = 0$, por lo que $Ax \in N(A^T)$. Como $C(A)$ y $N(A^T)$ son complementos ortogonales, entonces $Ax = 0$. Ahora, como Ax es una combinación lineal de las columnas de A, y las mismas son l.i., tenemos que $Ax = 0 \implies x = 0$. Entonces, como $(A^TA)x = 0 \implies x = 0$, $N(A^TA) = \{0\}$, y siendo A^TA una matriz cuadrada, luego es inversible. □

Teorema 15. *Sea A una matriz $n \times k$ con columnas l.i. y $P = A(A^TA)^{-1}A^T$. Entonces, $P^2 = P$ y $P^T = P$. Recíprocamente, toda matriz simétrica P que verifica $P^2 = P$ representa una proyección.*

Demostración. **Ida.** Como P es la matriz de proyección sobre $C(A)$, $Pb = proy_{s/C(A)}b \in C(A)$, tenemos que $P(Pb) = Pb$, de lo que se deduce que $PP = P^2 = P$. Por otro lado, $P^T = (A(A^TA)^{-1}A^T)^T = (A^T)^T(A(A^TA)^{-1})^T = A(A^TA)^{-1}A^T = P$.

Vuelta. Si P es simétrica, entonces $N(P) = N(P^T)$. Luego, como son el mismo espacio, su complemento ortogonal es el mismo. Sin embargo, por TFDAL2, el complemento ortogonal de $N(P)$ es $C(P^T)$ y el de $N(P^T)$ es $C(P)$. Por lo tanto, $C(P^T) = C(P)$. Ahora, si $b \in \mathbb{R}^n$, b se puede descomponer en una componente $b_N \in N(P)$ y una $b_C \in C(P)$. Luego,

$$Pb = P(b_N + b_C) = Pb_N + Pb_C = Pb_C.$$

Como $b_C \in C(P)$, existe x tal que $Px = b_C$, y reemplazando,

$$Pb = Pb_C = P(Px) = (PP)x = Px = b_C.$$

Resulta entonces que P devuelve la componente de b en $C(P)$, y por lo tanto proyecta sobre $C(P)$. □

3.1. Matrices Ortogonales

Si A es inversible, podemos ver que $P = I$. Efectivamente, $P = A(A^TA)^{-1}A^T = (AA^{-1})((A^T)^{-1}A^T) = II = I$. Esto tiene sentido, ya que si A (de tamaño $n \times n$) es inversible, $C(A) = \mathbb{R}^n$, y para cualquier vector $b \in \mathbb{R}^n$ el sistema es consistente.

Por otro lado, si A es inversible, sus columnas forman una base de $\mathbb{R}^n$. Si esa base es ortonormal, entonces $A^TA = I$. Más aún, para cualquier matriz A, si sus columnas son ortonormales, $A^TA = I$.

Definición 40. Una **matriz ortogonal** es una matriz cuadrada con columnas ortonormales.

Lema 42. *Si Q es una matriz ortogonal $n \times n$, entonces $Q^T = Q^{-1}$. Además, para todo $x \in \mathbb{R}^n$, $||Qx|| = ||x||$.*

Demostración. Considerando, $Q = [q_1, \ldots, q_n]$, con $q_i \in \mathbb{R}^n$ para todo $i = 1, \ldots, n$, planteo el producto Q^TQ:

$$Q^TQ = \begin{bmatrix} q_1^T \\ q_2^T \\ \vdots \\ q_n^T \end{bmatrix} \begin{bmatrix} q_1 & q_2 & \cdots & q_n \end{bmatrix} = \begin{bmatrix} q_1^Tq_1 & q_1^Tq_2 & \cdots & q_1^Tq_n \\ q_2^Tq_1 & q_2^Tq_2 & \cdots & q_2^Tq_n \\ q_3^Tq_1 & q_3^Tq_2 & \cdots & q_3^Tq_n \\ \vdots & \vdots & \ddots & \vdots \\ q_n^Tq_1 & q_n^Tq_2 & \cdots & q_1^Tq_n \end{bmatrix}.$$

Como Q es una matriz ortogonal, tenemos que $q_i^Tq_j = \langle q_i, q_j \rangle = 0$ si $i \neq j$. Además, $q_i^Tq_i = \langle q_i, q_i \rangle = ||q_i||^2 = 1$. Entonces,

$$Q^TQ = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} = I_n \implies Q^T = Q^{-1}.$$

Por otro lado, como la norma es no negativa, $||Qx|| = ||x|| \iff ||Qx||^2 = ||x||^2$. Trabajando sobre la expresión y utilizando lo recién probado,

$$||Qx||^2 = \langle Qx, Qx \rangle = (Qx)^T(Qx) = (x^TQ^T)(Qx) = x^T(Q^TQ)x = x^TIx = x^Tx = \langle x, x \rangle = ||x||^2.$$

□

Volviendo a la proyección de un vector $b \in \mathbb{R}^n$ sobre el espacio columna de A de tamaño $n \times k$; de lo visto anteriormente, se puede simplificar la fórmula de la mejor solución si A tiene columnas ortonormales. En dicho caso, la ecuación $A^TA\hat{x} = A^Tb$ se reduce a $\hat{x} = A^Tb$. Luego, $\hat{b} = A\hat{x} = AA^Tb$, y la matriz de proyección es $P = AA^T$.

4. Factorización QR

Sea A una matriz $n \times k$ con columnas l.i. y sean $Q^i, i = 1, \ldots, k$ los vectores ortonormales obtenidos luego de aplicar G-S a las columnas de A y luego de normalizarlos. Sea Q la matriz $n \times k$ cuyas columnas son los vectores $Q^i, i = 1, \ldots, k$. Entonces, se puede probar que $A = QR$, donde R es una matriz triangular superior de tamaño $k \times k$. Esto se denomina **factorización** QR de una matriz. Esta factorización simplifica la resolución de sistemas inconsistentes por aproximación.

Lema 43. *Sea A una matriz $n \times k$ de columnas l.i. Sea Q la matriz que se obtiene ortogonalizando las columnas de A. Entonces, $R = Q^T A$ es triangular superior e inversible.*

Demostración. Considerando $Q = [q_1 \ldots q_k]$ y $A = [a_1 \ldots a_k]$, planteo el producto $Q^T A$:

$$Q^T A = \begin{bmatrix} q_1^T \\ q_2^T \\ \vdots \\ q_k^T \end{bmatrix} \begin{bmatrix} a_1 & a_2 & \cdots & a_k \end{bmatrix} = \begin{bmatrix} q_1^T a_1 & q_1^T a_2 & \cdots & q_1^T a_k \\ q_2^T a_1 & q_2^T a_2 & \cdots & q_2^T a_k \\ q_3^T a_1 & q_3^T a_2 & \cdots & q_3^T a_k \\ \vdots & \vdots & \ddots & \vdots \\ q_k^T a_1 & q_k^T a_2 & \cdots & q_1^T a_k \end{bmatrix}.$$

En la demostración de la correctitud del proceso de ortogonalización de Gram-Schmidt, se probó que si los vectores de entrada son l.i., para todo $i = 1, \ldots, k$ vale:

1. $q_i \perp q_j$ para todo j tal que $1 \leq j \leq i - 1$.

2. q_i es una combinación lineal de a_j, con $1 \leq j \leq i$.

3. $q_i \neq 0$.

Si combinamos las dos primeras proposiciones, tenemos que para todo $i = 1, \ldots, n$,

$$q_i \perp \langle \{q_1, \ldots, q_{i-1}\} \rangle = \langle \{a_1, \ldots, a_{i-1}\} \rangle \implies q_i^T a_j = \langle q_i, a_j \rangle = 0 \text{ para } 1 \leq j \leq i - 1.$$

Por lo tanto, R resulta triangular superior. Ahora, para ver que es invertible resta probar que $\langle q_i, a_i \rangle \neq 0$ para todo $i = 1, \ldots, n$:

$$q_i = \sum_{j=1}^{i} c_j a_j \implies \langle q_i, q_i \rangle = \sum_{j=1}^{i} c_j \langle a_j, q_i \rangle = c_i \langle a_i, q_i \rangle, \quad c_j \in \mathbb{R}.$$

Por (3) resulta que $\langle q_i, q_i \rangle \neq 0$, lo que implica que $\langle a_i, q_i \rangle \neq 0$, y luego R es invertible. □

Teorema 16. *Toda matriz A de tamaño $n \times k$ con columnas l.i. puede ser factorizada $A = QR$, con Q matriz $n \times k$ con columnas ortonormales y R matriz triangular superior e inversible de tamaño $k \times k$.*

Demostración. Sea $Q = [q_1 q_2 \dots q_k]$. Si existe una matriz R tal que $A = QR$, entonces

$$A = QR \implies Q^T A = (Q^T Q)R \implies R = Q^T A.$$

Es decir, si existe R, debe valer que $R = Q^T A$. Se probó en el lema anterior que $Q^T A$ es triangular superior e inversible. Resta ver que realmente $A = QR$.

Como las columnas de Q forman una base ortonormal, resulta que

$$A^i = \sum_{j=1}^{n} \langle Q^j, A^i \rangle Q^j = \sum_{j=1}^{i} ((Q^j)^T A^i) Q^j,$$

pues por Gram-Schmidt sabemos que si $j > i$, $\langle Q^j, A^i \rangle = 0$. Luego,

$$(QR)^i = \sum_{j=1}^{n} Q^j ((Q^j)^T A^i) = \sum_{j=1}^{n} \langle Q^j, A^i \rangle Q^j = \sum_{j=1}^{i} \langle Q^j, A^i \rangle Q^j = A^i.$$

$\square$

Unidad IV

Determinantes

El determinante de una matriz cuadrada puede ser visto como una función de las matrices $n \times n$ en los reales. Sabemos que el determinante determina la singularidad o no de una matriz (es singular si y solo si su determinante es nulo), y también otorga una fórmula explícita para las entradas de A^{-1} y para la solución del sistema $Ax = b$ en caso de ser no singular. A pesar de todo esto, el método más eficiente para la resolución de sistemas sigue siendo la Eliminación Gaussiana, y también lo es para el cálculo del determinante.

El determinante de una matriz cuadrada A se nota $\det(A)$, y a lo largo de la unidad todas las matrices se asumen ser cuadradas de tamaño $n \times n$.

Las siguientes tres propiedades permiten definir al determinante:

1. El determinante de la matriz identidad es 1.
2. El determinante cambia su signo al intercambiar dos filas de la matriz.
3. El determinante es lineal en la primera fila. Esto quiere decir, si A es una matriz cuya primer fila A_1 se como combinación lineal de vectores $u, v \in \mathbb{R}^n$ ($A_1 = \alpha u + \beta v$) y A_u y A_v son las matrices obtenidas a partir de A reemplazando A_1 por u y por v respectivamente, entonces

$$\det(A) = \alpha \det(A_u) + \beta \det(A_v).$$

   Combinando esta propiedad con la propiedad 2, se tiene que el determinante es lineal en *todas* sus filas.

1. Propiedades

Las siguientes propiedades del determinante se deducen de las tres anteriores:

4. Si A tiene dos filas iguales, $\det(A) = 0$.

   Demostración. Sean i, j tal que $A_i = A_j$. Luego,

$$\det(P_{ij}A) = -\det(A) = \det(A) \iff \det(A) = 0.$$

□

5. Si a una fila de A se le resta un múltiplo de otra de sus filas, el determinante no cambia.

   Demostración. Sea A' la matriz resultante de A al reemplazar A_i por $A_i - \alpha A_j$, con $i, j \in \{1, \ldots, n\}$, $i \neq j$. Entonces, $\det(A') = -\det(P_{1i}A') = -\det(P_{1i}A) + \alpha \det(P_{1i}A'') = \det(A) + \alpha \det(P_{1i}A'')$, donde A'' es la matriz igual a A con la excepción de que $A_1 = A_j$, por lo que $\det(A'') = 0$. Entonces,

   $$\det(A') = \det(A) + \alpha \det(P_{1i}A'') = \det(A).$$

   □

6. Si A tiene una fila nula, $\det(A) = 0$.

   Demostración. Sea i tal que $A_i = 0$, y $j \neq i$. Por la propiedad 5, $\det(A)$ se mantiene igual al sumarle una vez la fila j a la fila i, y como la matriz resultante tiene dos filas iguales (al ser la fila i nula), se concluye que $\det(A) = 0$. □

7. Si A es una matriz diagonal, su determinante es el producto de los elementos de la diagonal.

   Demostración. Sea $A^{(i)}$, $i = 1, \ldots, n$ la matriz igual tal que $A_j^{(i)} = I_j$ si $j \leq i$, y $A_j^{(i)} = A_j$ si $j > i$. Entonces, por las propiedades 1, 3 y 6,

   $$\det(A) = A_1^1 \det(A^{(1)}) = A_1^1 A_2^2 \det(A^{(2)}) = \cdots = \prod_{i=1}^{n} A_i^i \cdot \det(A^{(n)}) = \prod_{i=1}^{n} A_{ii} \cdot \det(I) = \prod_{i=1}^{n} A_{ii}.$$

   □

Teorema 17. *Para todo $n \in \mathbb{N}$, existe una única función de $\mathbb{R}^{n\times n}$ en $\mathbb{R}$ que verifica las propiedades 1, 2 y 3.*

Demostración. Sea $n \in \mathbb{N}$ y f una función del conjunto de matrices $n \times n$ en $\mathbb{R}$ que verifica las propiedades 1, 2 y 3. Sea A una matriz $n \times n$, y sea D la matriz obtenida de aplicar el método de Eliminación de Gauss. Notemos con $A^{(k)}$ la matriz obtenida en el paso k del proceso de eliminación de Gauss. Claramente, $A^{(0)} = A$. Sea $p \in \mathbb{N}$ tal que $A^{(p)} = D$. Para $k = 1, \ldots, p$ vale lo siguiente:

1. Por la propiedad 5, si $A^{(k)}$ se obtuvo en un paso de eliminación, $f\left(A^{(k)}\right) = f\left(A^{(k-1)}\right)$.

2. Por propiedad 2, si $A^{(k)}$ se obtuvo en un paso de permutación de filas, $f\left(A^{(k)}\right) = -f\left(A^{(k-1)}\right)$.

Por lo tanto, $f(A) = f(D)$ si el número de pasos de permutación de filas es par, y $f(A) = -f(D)$ si es impar. Por la propiedad 7, $f(D)$ queda unívocamente determinado y es el producto de las entradas de su diagonal. Por lo tanto, $f(A)$ también queda unívocamente determinado. □

Definición 41. Dado $n \in \mathbb{N}$, llamamos **determinante** a la función det de $\mathbb{R}^{n\times n}$ en $\mathbb{R}$ que verifica las propiedades 1, 2 y 3.

Más propiedades:

8. Si A es una matriz triangular, $\det(A)$ es el producto de las entradas de A en su diagonal.

   Demostración. Haciendo ceros arriba de la diagonal, podemos llegar desde A a la matriz diagonal D sin con las mismas entradas en la diagonal, sin realizar permutaciones. Por lo tanto, $\det(A) = \det(D) = A_1^1 A_2^2 \cdots A_n^n$. □

9. $\det(A) = 0 \iff$ A es singular.

   Demostración. **Ida.** Sea A una matriz $n \times n$ tal que $\det(A) = 0$. Sea U la matriz triangular superior obtenida de A luego de aplicar el proceso de Eliminación de Gauss. Sea k la cantidad de permutaciones de filas hechas en dicho proceso. Entonces, por las propiedades 2 y 3, $\det(A) = (-1)^k \det(U)$. Ahora, por la propiedad 8, $\det(U) = \prod_{i=0}^{n} U_i^i = \prod_{i=0}^{n} A_i^i$. Entonces,

   $$\det(A) = 0 = (-1)^k \prod_{i=0}^{n} A_i^i \implies \exists i \in \{1, \ldots, n\} / A_i^i = 0.$$

   Ahora, si algún elemento de la diagonal de A es nulo, entonces la forma escalonada de A tiene (al menos) una fila nula, y luego $C(A) \neq \mathbb{R}^n$, y por lo tanto, A es singular.

   Vuelta. Supongo que A es singular. Luego, por lo anterior, $\det(A) = (-1)^k \det(U)$. Ahora, como A es singular, $C(A) \neq \mathbb{R}^n$, es decir, existen $b \in \mathbb{R}^n$ tal que el sistema $Ax = b$ no tiene solución. Entonces, algún elemento de la diagonal de U es nulo. Si todos fueran no nulos, dado $b \in \mathbb{R}^n$, el sistema $Ax = b$ puede resolverse aplicando Gauss a A (y las mismas operaciones a b) y luego resolviendo por sustitución hacia atrás el sistema $Ux = \tilde{b}$. Entonces, efectivamente, existe un $i \in \{1, \ldots, n\}$ tal que $U_i^i = 0$. Por la propiedad 8,

   $$\det(U) = \prod_{i=1}^{n} A_i^i = 0 \implies \det(A) = (-1)^k \det(U) = 0.$$

   □

10. $\det(AB) = \det(A)\det(B)$.

    Demostración. Sabemos que si B es singular, AB también lo es, para toda A. Luego, si B es singular, por Propiedad 9, $\det(AB) = \det(B) = 0$, y vale la tesis. Sea entonces B una matriz $n \times n$ no singular.

    Defino una función f del conjunto de matrices $n \times n$ en $\mathbb{R}$ tal que para toda A matriz $n \times n$, $f(A) = \frac{\det(AB)}{\det(B)}$. Veré que $f = \det$.

    - $f(I) = \frac{\det(IB)}{\det(B)} = \frac{\det(B)}{\det(B)} = 1$, por lo que se verifica la Propiedad 1.

- Sea A' una matriz que se obtiene intercambiando las filas i y j de A. Como $A'B$ se obtiene intercambiando las filas i y j de AB, por la Propiedad 2 sabemos que $\det(A'B) = -\det(AB)$. Luego, $f(A') = -f(A)$.
- Sea A una matriz tal que $A_1 = \alpha u + \beta v$, $u, v \in \mathbb{R}^n$ y $\alpha, \beta \in \mathbb{R}$. Sean $A[u]$ y $A[v]$ las matrices obtenidas de A al reemplazar A_1 por u y v respectivamente. Debo probar que

$$f(A) = \alpha f(A[u]) + \beta f(A[v]).$$

Nótese que la primera fila de AB es $\alpha(uB) + \beta(vB)$. Llamo $AB[uB]$ y $AB[vB]$ a las matrices obtenidas de AB al reemplazar $(AB)_1$ por uB y por vB, respectivamente. Por la Propiedad 3, tenemos que

$$\det(AB) = \alpha \det(AB[uB]) + \beta \det(AB[vB]).$$

Nótese, además, que $AB[uB] = A[u]B$ y $AB[vB] = A[v]B$. Por lo tanto,

$$f(A) = \frac{\det(AB)}{\det(B)} = \alpha \frac{\det(A[u]B)}{\det(B)} + \beta \frac{\det(A[v]B)}{\det(B)} = \alpha f(A[u]) + \beta f(A[v]).$$

$\square$

11. $\det(A^T) = \det(A)$. Esta propiedad permite extender a las columnas todas las propiedades asociadas a las filas.

Demostración. Si A es singular, A^T también lo es y luego ambas matrices tienen determinante nulo. Sea A una matriz no singular. Considero la descomposición LDV de A, esto es, $PA = LDV$. Entonces, por Propiedad 10,

$$\det(PA) = \det(P)\det(A) = \det(L)\det(D)\det(V).$$

Por otro lado, $A^T P^T = (PA)^T = (LDV)^T = V^T D^T L^T$, y luego

$$\det(A^T)\det(P^T) = \det(V^T)\det(D^T)\det(L^T).$$

Como L, L^T, V y V^T son matrices triangulares con 1's en la diagonal, por Propiedad 8 sus determinantes son iguales a 1. Además, como es diagonal, $D = D^T$. Entonces,

$$\det(P)\det(A) = \det(D) = \det(A^T)\det(P^T).$$

Por propiedad 2, los determinantes, de P y P^T son 1 o -1. Como P y P^T son matrices inversas, $PP^T = I$ y luego $\det(P)\det(P^T) = 1$, por lo que ambos determinantes tienen el mismo signo, y $\det(P) = \det(P^T)$. Luego, queda probado que $\det(A) = \det(A^T)$. $\square$

2. Cálculo por Cofactores

También se puede establecer una definición recursiva para el cálculo del determinante. Teniendo en cuenta su linealidad por filas, el determinante de una matriz $n \times n$ queda expresado como suma de n determinantes de matrices que reemplazan la primer fila en A por $A_1^j e^j$, para $j = 1, \dots, n$. Luego, cada uno de esos determinantes resultará igual al producto de A_1^j por una expresión que no depende de las entradas de la columna A^j. Es decir, el determinante asociado a A_1^j dependerá de la información de una submatriz M_{ij} de A que se obtiene borrando la fila 1 y la columna j. Más aun, esta expresión se denomina **cofactor** $C_{1j} = (-1)^{1+j} \det(M_{1j})$.

Entonces, para todo $i, j = 1, \dots, n$, definiendo M_{ij} como la submatriz de A obtenida borrando la fila i y la columna j, y el cofactor $C_{ij} = (-1)^{i+j} \det(M_{ij})$, la fórmula recursiva del determinante es la siguiente:

Teorema 18. *Para todo $i = 1, \dots, n$,*

$$\det(A) = \sum_{j=1}^{n} A_i^j C_{ij}.$$

Definición 42. Dada una matriz A, se denomina **matriz adjunta** de A a la matriz C cuyas entradas son los cofactores de A.

Lema 44. *Si A es una matriz inversible y C su matriz adjunta, entonces $AC^T = \det(A)I$.*

Demostración. Considerando la definición del determinante por cofactores, si $i \neq j$, $(AC^T)_i^j = \sum_{k=1}^{n} A_i^k C_{jk} = \det(A')$, donde A' es la matriz resultante de A al reemplazar la fila j por la fila i. En dicho caso, A' tiene dos filas iguales, y $\det(A') = (AC^T)_i^j = 0$. Por otro lado, si $i = j$, $(AC^T)_i^j = \sum_{k=1}^{n} A_i^k C_{ik} = \det(A)$. Por lo tanto, queda probado que $AC^T = \det(A)I$. □

Lema 45. *Sea A una matriz inversible. Entonces,*

$$A^{-1} = \frac{1}{\det(A)} C^T.$$

Demostración. $A\left(\frac{1}{\det(A)} C^T\right) = \frac{1}{\det(A)}(AC^T) = \frac{\det(A)}{\det(A)} I = I$, y como A es inversible, queda probado que $A^{-1} = \frac{1}{\det(A)} C^T$. □

Como consecuencia de esto, se obtiene la *Regla de Cramer* de utilidad en la resolución de un sistema no singular:

Teorema 19. ***(Regla de Cramer)*** *Sea A una matriz no singular y $b \in \mathbb{R}^n$. Para $j = 1, \dots, n$, sea B_j la matriz que se obtiene reemplazando la columna A^j por b. Entonces, la solución $\hat{x}$ del sistema $Ax = b$ verifica*

$$\hat{x}_j = \frac{\det(B_j)}{\det(A)}.$$

Demostración. Se probó en un lema anterior que $A^{-1} = \frac{1}{\det(A)} C^T$. Luego, la solución del sistema $Ax = b$ está dada por $\hat{x} = A^{-1}b = \frac{1}{\det(A)}(C^T b)$, donde

$$\hat{x}_j = \frac{1}{\det(A)}(C^T b)_j = \frac{1}{\det(A)}(C^T)_j b = \frac{1}{\det(A)} \sum_{i=1}^{n} C_{ij} b_i = \frac{1}{\det(A)} \sum_{i=1}^{n} C_{ij} (B_j)_i^j = \frac{\det(B_j)}{\det(A)}.$$

□

Unidad V

Autovalores y Autovectores

Son de interés en este capítulo las ecuaciones matriciales de la forma

$$Ax = \lambda x,$$

donde A es una matriz cuadrada $n \times n$, y $\lambda \in \mathbb{K}$ y $x \in \mathbb{R}^n$ son variables. Pensando a A como una transformación lineal de $\mathbb{R}^n$ en $\mathbb{R}^n$, dado $x \in \mathbb{R}^n$, Ax es el vector imagen de x a través de A. Si existe $\lambda \in \mathbb{R}$ tal que $Ax = \lambda x$, entonces se puede decir que A *no cambió la dirección* de x.

Lema 46. *Sea $x \in \mathbb{R}^n$ y $\lambda \in \mathbb{R}$ tales que $Ax = \lambda x$. Entonces, para todo $\alpha \in \mathbb{R}$, $A(\alpha x) = \lambda(\alpha x)$.*

Demostración. $A(\alpha x) = \alpha(Ax) = \alpha(\lambda x) = \lambda(\alpha x)$. □

Ejemplo: Este tipo de ecuaciones matriciales toma un rol protagónico en el contexto de las ecuaciones diferenciales. Considerando la ecuación diferencial $v'(x) = av(x)$, con $a \in \mathbb{R}$, se puede probar que todas las posibles soluciones son de la forma $v(x) = ce^{ax}$, con $c \in \mathbb{R}$. Por otro lado, un sistema lineal de ecuaciones diferenciales de dos funciones incógnitas u, v es de la forma

$$\begin{cases} u'(x) = au(x) + bv(x) \\ v'(x) = cu(x) + dv(x) \end{cases}, \text{ con } a, b, c, d \in \mathbb{R}.$$

Luego, si $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, $w(x) = \begin{bmatrix} u(x) \\ v(x) \end{bmatrix}$ y $w'(x) = \begin{bmatrix} u'(x) \\ v'(x) \end{bmatrix}$, el sistema de ecuaciones diferenciales lineales pasa a tener la forma

$$w'(x) = Aw(x).$$

Ahora, la similitud con la ecuación $v'(x) = av(x)$ permite proponer las soluciones exponenciales $u(x) = \alpha e^{\lambda x}$ y $v(x) = \beta e^{\lambda x}$. Se busca determinar $\begin{bmatrix} \alpha \\ \beta \end{bmatrix} \in \mathbb{R}^2$ y $\lambda \in \mathbb{R}$ para que u y v así construidas sean solución del sistema. Entonces, sustituyendo,

$$\begin{cases} \alpha\lambda e^{\lambda x} = a\alpha e^{\lambda x} + b\beta e^{\lambda x} \\ \beta\lambda e^{\lambda x} = c\alpha e^{\lambda x} + d\beta e^{\lambda x} \end{cases} \iff \begin{cases} \alpha\lambda = a\alpha + b\beta \\ \beta\lambda = c\alpha + d\beta \end{cases} \iff \lambda \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = A \begin{bmatrix} \alpha \\ \beta \end{bmatrix}.$$

Esto puede generalizarse a un sistema lineal de ecuaciones diferenciales con n funciones incógnitas, y las soluciones exponenciales de este sistema se construirán resolviendo un sistema del tipo $Ax = \lambda x$. (*Fin* *del ejemplo*)

La siguiente definición se induce de que la ecuación $Ax = \lambda x$, con A $m \times n$, puede ser pensada como un caso particular de la ecuación $Tv = \lambda v$, para T un operador lineal cualquiera de un espacio vectorial en sí mismo.

Definición 43. Sea V un espacio vectorial sobre $\mathbb{K}$ y T una transformación lineal de V en sí mismo. Entonces, $\lambda \in \mathbb{K}$ es un **autovalor** (o **valor propio**) de T si existe $0 \neq v \in V$ tal que $Tv = \lambda v$.

Si $\lambda \in \mathbb{K}$ es un autovalor de T, todo vector $0 \neq v \in V$ tal que $Tv = \lambda v$ se denomina **autovector** (o **vector propio**) de T *asociado* a λ. El conjunto de todos los autovectores asociados a λ, en conjunto con el vector nulo, se denomina **autoespacio** de T *asociado* a λ, y se nota $V(T, \lambda)$.

Observación 26. Cuando se trabaja con transformaciones lineales definidas por una matriz cuadrada, se considera válido hablar de autovectores, autovalores y autoespacios de la matriz.

Lema 47. *Sea V un espacio vectorial sobre $\mathbb{K}$, T una transformación lineal de V en sí mismo y $\lambda \in \mathbb{K}$ un autovalor de T. Entonces, el autoespacio $V(T, \lambda)$ de T asociado a λ es un subespacio vectorial de V.*

Demostración. Sean $v_1, v_2 \in V(T, \lambda)$. Luego, $Av_1 = \lambda v_1$ y $Av_2 = \lambda v_2$. Ahora,

$$A(\alpha v_1 + \beta v_2) = \alpha(Av_1) + \beta(Av_2) = \alpha\lambda v_1 + \beta\lambda v_2 = \lambda(\alpha v_1 + \beta v_2),$$

por lo que $V(T, \lambda)$ es un subespacio vectorial de V. $\square$

Definición 44. *(Definición alternativa de autovalor)* Sea V un espacio vectorial sobre $\mathbb{K}$ y T una transformación lineal de V en sí mismo. Entonces, $\lambda \in \mathbb{K}$ es un autovalor de T si la transformación lineal $T - \lambda I$ no es un isomorfismo.

Equivalentemente, teniendo en cuenta que T es un isomorfismo si y solo si su matriz asociada A es inversible, se tiene que λ es un autovalor de A si y solo si

$$\det(A - \lambda I) = 0 \quad (\textit{Ecuación característica de } A).$$

En general, los autovalores de una matriz A resultan ser las raíces del *polinomio característico* que se obtienen con el desarrollo de $\det(A - \lambda I)$. Además, $V(T, \lambda)$ está formado por los vectores $v \in \mathbb{R}^n$ tales que $(A - \lambda I)v = 0$. Es decir, el autoespacio termina siendo $N(A - \lambda I)$. Como aclaración, al decir que una matriz $n \times n$ tiene n autovalores, se infiere a $\mathbb{R}^n$ como un subespacio vectorial de $\mathbb{C}^n$ sobre $\mathbb{C}$.

- Los autovalores de una matriz diagonal A coinciden con los elementos de la diagonal de A. Además, se puede probar que, con A matriz diagonal $n \times n$, para todo $i = 1, \ldots, n$, $x^i = e_i$ es un autovector asociado a $\lambda_i = A^i_i$.

- Una matriz cualquiera tiene un autovalor nulo si y solo si es no inversible.

- Los autovalores de la matriz proyección de un vector $a \in \mathbb{R}^n$ son $\lambda = 1$ (con autovector a que se proyecta a sí mismo) y $\lambda = 0$ de multiplicidad $n-1$ (con autovectores en $N(P)$ que se proyectan sobre el vector nulo). Esta multiplicidad coincide con la dimensión de $N(P)$. Por lo tanto, se tienen $n-1$ autovectores l.i. del autovalor $\lambda = 0$ que se proyectan sobre el vector nulo.
- En la matriz $K = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ de rotación de 90° en $\mathbb{R}^2$ no hay autovalores reales. Naturalmente, todo vector no nulo de $\mathbb{R}^2$ cambia de dirección bajo el efecto de una rotación.
- Los autovalores de una matriz triangular son los elementos de su diagonal.

Teorema 20. *Sea A una matriz $n \times n$ y $\lambda_i, i = 1, \ldots, n$ sus n autovalores. Entonces,*

1. $\sum_{i=1}^{n} \lambda_i = \sum_{i=1}^{n} A_i^i = tr(A)$.
2. $\prod_{i=1}^{n} \lambda_i = \det(A)$.

Demostración.

1. Más adelante se prueba que para toda matriz cuadrada A existe una matriz ortogonal Q tal que $Q^{-1}AQ = T$, con T matriz triangular. Más aún, estas matrices resultan ser semejantes, y por ende tienen los mismos autovalores. Juntando esto con el hecho de que los autovalores de una matriz triangular son los elementos de su diagonal, y con la conmutatividad de la traza ($tr(AB) = tr(BA)$ para todas A, B) tenemos que

$$\sum_{i=1}^{n} \lambda_i = \sum_{i=1}^{n} T_i^i = tr(Q^{-1}AQ) = tr(AQ^{-1}Q) = tr(A).$$

2. Como los autovalores de A son las raíces de su polinomio característico $\det(A - \lambda I)$, éste puede factorizarse en términos de sus autovalores. precisamente,

$$\det(A - \lambda I) = \prod_{i=1}^{n} (\lambda_i - \lambda),$$

y como λ es una variable, tomando $\lambda = 0$ tenemos

$$\det(A) = \prod_{i=1}^{n} \lambda_i.$$

□

1. Diagonalización

Dada una matriz A $n \times n$, con $\lambda_i, = 1, \ldots, n$, sus autovalores, notamos con Λ a la matriz diagonal $n \times n$ donde $\Lambda_i^i = \lambda_i$.

Lema 48. *Sea A una matriz $n \times n$, con $x^i, = 1, \ldots, n$ autovectores l.i. de A. Sea S la matriz $n \times n$ tal que $S^i = x^i, i = 1, \ldots, n$. Entonces,*

$$S^{-1}AS = \Lambda.$$

Demostración. Para todo $i = 1, \ldots, n$, sea e_{λ_i} el vector donde la componente i es λ_i y el resto de las componentes nulas. Claramente, $\Lambda^i = e_{\lambda_i}$. Entonces, $(AS)^i = Ax^i = \lambda_i x^i = Se_{\lambda_i} = S\Lambda^i = (S\Lambda)^i$. Por lo tanto $AS = S\Lambda$, y por la lineal independencia de los autovectores, tenemos que S es inversible. Entonces,

$$AS = S\Lambda \iff S^{-1}AS = \Lambda.$$

□

Definición 45. Una matriz A $m \times n$ es **diagonalizable** si existe una matriz inversible S tal que $S^{-1}AS$ es una matriz diagonal.

Lema 49. *Sea A una matriz $n \times n$, $\lambda_i,\ i = 1, \ldots, r$ autovalores distintos de A y para $i = 1, \ldots, r$, x^i un autovector asociado a λ_i. Entonces, $\{x^1, \ldots, x^r\}$ son vectores l.i..*

Demostración. Como son autovectores, $x^i \neq 0,\ i = 1, \ldots, r$. La prueba es por inducción en r. El caso $r = 1$ es trivial.

Supongo que la tesis vale para $r = k$. Debo probar su validez para $r = k + 1$.

Como $\lambda_i,\ i = 1, \ldots, k$ son autovalores diferentes, por hipótesis de inducción sabemos que $x^i,\ i = 1, \ldots, k$ son l.i.. Supongo que $\{x^i : i = 1, \ldots, k+1\}$ no son l.i.. Entonces, existen $\alpha_i \in \mathbb{R},\ i = 1, \ldots, k$ tales que

$$x^{k+1} = \sum_{i=1}^{k} \alpha_i x^i.$$

Premultiplicando por A,

$$Ax^{k+1} = \lambda_{k+1}x^{k+1} = \sum_{i=1}^{k} \alpha_i Ax^i = \sum_{i=1}^{k} \alpha_i \lambda_i x^i.$$

Por otro lado, premultiplicando por λ_{k+1} en lugar de A,

$$\lambda_{k+1}x^{k+1} = \sum_{i=1}^{k} \alpha_i \lambda_{k+1} x^i,$$

e igualando las dos expresiones, resulta

$$\sum_{i=1}^{k} \alpha_i \lambda_i x^i = \sum_{i=1}^{k} \alpha_i \lambda_{k+1} x^i \iff \sum_{i=1}^{k} \alpha_i(\lambda_i - \lambda_{k+1})x^i = 0.$$

Como los vectores $\{x^i : i = 1, \ldots, k\}$ son l.i., para todo $i = 1, \ldots, k$ vale $\alpha_i(\lambda_i - \lambda_{k+1}) = 0$. Como $x^{k+1} \neq 0$, existe $j \in \{1, \ldots, k\}$ tal que $\alpha_j \neq 0$. Entonces, $\lambda_j - \lambda_{k+1} = 0 \iff \lambda_j = \lambda_{k+1}$, una contradicción. Por lo tanto, los vectores de $\{x^i : i = 1, \ldots, k+1\}$ son l.i.. □

Teorema 21. *Sea A una matriz $n \times n$. Si los n autovalores de A son diferentes, A es diagonalizable.*

Demostración. Por lo probado anteriormente, si los autovalores de A son diferentes, las columnas de la matriz S (que tiene por columnas a los autovectores) son l.i., y luego es inversible, y $S^{-1}AS = \Lambda$. S diagonaliza a A, por lo que A es diagonalizable. □

Lema 50. *Sea A una matriz $n \times n$. Entonces, existen S inversible y D matriz diagonal tal que $S^{-1}AS = D$ si y solo si A tiene n autovectores l.i..*

Demostración. **Ida.** Sean S y D tal que $S^{-1}AS = D$. Entonces, $AS = SD$. Para todo $i = 1, \dots, n$, sea $S^i = x^i$. Entonces, $(AS)^i = AS^i = Ax^i$. Por otro lado, como D es diagonal, $(SD)^i = SD^i = D^i_i x^i$.

Como $AS = DS$, resulta que, para todo $i = 1, \dots, n$, $Ax^i = D^i_i x^i$ y por lo tanto D^i_i es un autovalor de A y x^i un autovector. Como S es inversible, sus columnas son n autovectores de A l.i. y luego A tiene n autovectores l.i.

Vuelta. Supongo ahora que A tiene n autovectores l.i. x^i asociados a los autovalores λ_i, $i = 1, \dots, n$. Construyo la matriz $n \times n$ S donde $S^i = x^i$. Claramente, S es inversible, y además $AS = S\Lambda$. Por lo tanto, $S^{-1}AS = \Lambda$, y Λ es una matriz diagonal. □

Lema 51. *Sean A, B, S matrices $n \times n$ tal que S diagonaliza a A y B, con $A \neq B$. Entonces, A y B no tienen los mismos n autovalores.*

Demostración. Sean Λ_A y Λ_B las matrices diagonales con los autovalores de A y B, respectivamente. Entonces,

$$A \neq B \implies S^{-1}AS \neq S^{-1}BS \implies \Lambda_A \neq \Lambda_B.$$

Es decir, las matrices diagonales con los autovalores de A y B son distintas, siendo A y B diagonalizadas por la misma matriz (entonces, los autovalores iguales aparecen en las mismas entradas de la matriz). Luego, $\Lambda_A \neq \Lambda_B \implies \exists i \in \{1, \dots, n\} / \lambda_{A_i} \neq \lambda_{B_i}$, o, equivalentemente, A y B no tienen los mismos n autovalores. □

Observación 27. No todas las matrices son diagonalizables. Si los n autovalores de A son diferentes, A es diagonalizable, pero si existen autovalores iguales, no se puede asegurar que A no es diagonalizable. Por ejemplo, la matriz

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

tiene a 0 como autovalor de multiplicidad 2, por lo que S debería ser una matriz inversible tal que $AS = S\Lambda = S0 = 0$, de donde queda $A = 0S^{-1} = 0$, una contradicción. Sin embargo, la matriz identidad I_2 tiene a 1 como autovalor de multiplicidad 2, y resulta diagonalizable.

1.1. Diagonalización de Inversa y Productos

Teorema 22. *Sea A una matriz $n \times n$ inversible, $\{\lambda_i : i = 1, \ldots, n\}$ el conjunto de sus autovalores y $x^i, i = 1, \ldots, n$ sus autovectores asociados. Entonces, para todo $k \in \mathbb{N}$, $\{\lambda_i^k, i = 1, \ldots, n\}$ es el conjunto de autovalores de A^k. Además, x^i es un autovector correspondiente a λ_i^k, para todo $i = 1, \ldots, n$.*

Demostración. Por inducción en k. Vale si $k = 1$. Asumo que $A^t x^i = \lambda_i^t x^i$ para todo $i = 1, \ldots, n$, y quiero probar que $A^{t+1} x^i = \lambda_i^{t+1} x^i$. Para cada $i = 1, \ldots n$,

$$A^{t+1} x^i = A(A^t x^i) = A(\lambda_i^t x^i) = \lambda_i^t (A x^i) = \lambda_i^t (\lambda_i x^i) = \lambda_i^{t+1} x^i.$$

□

Teorema 23. *Sea A una matriz $n \times n$ inversible, $\{\lambda_i : i = 1, \ldots, n\}$ el conjunto de sus autovalores y para todo $i = 1, \ldots, n$, sea x^i el autovector asociado a λ_i. Entonces, $\{\lambda_i^{-1}, i = 1, \ldots, n\}$ es el conjunto de autovalores de A^{-1}. Además, x^i es un autovector correspondiente a λ_i^{-1}, para todo $i = 1, \ldots, n$.*

Demostración. Basta observar que si $Ax = \lambda x$, entonces $x = A^{-1}\lambda x = \lambda(A^{-1}x)$. Así, $A^{-1}x = \lambda^{-1}x$. □

Corolario 9. *Si S diagonaliza a A, entonces S diagonaliza a A^k para todo $k \in \mathbb{N}$. Además, si A es inversible, entonces S diagonaliza a A^{-1}.*

Para analizar los autovalores (y autovectores asociados) del producto de dos matrices, se consideran dos matrices $n \times n$ A y B, y $(\lambda_A, x), (\lambda_B, x)$ autovalores y autovectores asociados de A y B, respectivamente (observar que comparten autovector). Entonces,

$$(AB)x = A(Bx) = A(\lambda_B x) = \lambda_B(Ax) = \lambda_B(\lambda_A x) = (\lambda_B \lambda_A)x,$$

resultando $\lambda_B \lambda_A$ autovalor de AB y x su autovector asociado.

Teorema 24. *Sean A, B dos matrices $n \times n$, y S matriz que diagonaliza a A. Si S diagonaliza a B, entonces $AB = BA$. Más aun, si A tiene n autovalores diferentes y $AB = BA$, S diagonaliza a B.*

Demostración. Si S diagonaliza a A y a B, tenemos $A = S\Lambda_1 S^{-1}$ y $B = S\Lambda_2 S^{-1}$. Entonces,

$$AB = (S\Lambda_1 S^{-1})(S\Lambda_2 S^{-1}) = S(\Lambda_1 \Lambda_2)S^{-1}.$$

Como Λ_1 y Λ_2 son matrices diagonales, $\Lambda_1\Lambda_2 = \Lambda_1\Lambda_2$. Por lo tanto,

$$AB = S(\Lambda_1\Lambda_2)S^{-1}. = S(\Lambda_2\Lambda_1)S^{-1} = BA.$$

Supongo ahora que $AB = BA$ y A tiene todos sus autovalores diferentes. Entonces, dado λ autovalor de A, el autoespacio de λ tiene dimensión 1, pues, si hubiera algún autoespacio de dimensión $k > 1$, existe un conjunto de k vectores l.i. contenido en dicho autoespacio, y, en conjunto con los $n-1$ vectores l.i. asociados a los otros autoespacios, se tendría un conjunto de $n+k$ vectores l.i. en $\mathbb{R}^n$ (contradicción).

Sea λ, x tales que $Ax = \lambda x$. Luego,

$$A(Bx) = (AB)x = (BA)x = B(\lambda x) = \lambda(Bx).$$

Por lo tanto Bx también es autovector de A asociado a λ. Como el autoespacio correspondiente a λ tiene dimensión 1, Bx es un múltiplo de x, es decir, existe λ_B tal que $Bx = \lambda_B x$, y x es un autovector de B. Entonces, S diagonaliza a B. $\square$

Observación 28. La conmutatividad de las matrices en el producto es condición suficiente y necesaria para que compartan sus autovectores. Sin embargo, la condición de que una de ellas tenga todos sus autovalores diferentes no es necesaria, aunque la demostración es compleja.

Teorema 25. *Sean A, B dos matrices $n \times n$ y S, matriz que diagonaliza a A. Entonces, S diagonaliza a B si y solo si $AB = BA$.*

2. Aplicación: Ecuaciones en Diferencias

Muchos problemas de aplicación se resuelven a través de soluciones de un sistema de *ecuaciones en diferencias* de la forma

$$u_{k+1} = Au_k,\ k \in \mathbb{N},\ u_0 \in \mathbb{R}^n,$$

donde A es una matriz $n \times n$. La solución es una sucesión infinita $\{u_k : k \in \mathbb{N}\}$. Sin embargo, muchas veces solo es de interés conocer un término de la sucesión, por ejemplo, el término $u_{\hat{k}}$, y para valores grandes de $u_{\hat{k}}$ deja de ser preferible aplicar la recursión. Por otro lado, es fácil ver que $u_k = A^k u_0$.

Si A es diagonalizable, los cálculos se simplifican. Sea S su matriz diagonalizante de A y $x^i = S^i, i = 1, \ldots, n$ los autovectores columnas de S. Entonces,

$$u_k = A^k u_0 = (S\Lambda^k S^{-1})u_0 = (S\Lambda^k)(S^{-1}u_0).$$

Nombrando $c = S^{-1}u_0$, se puede resolver el sistema $Sc = u_0$ mediante Eliminación Gaussiana, y luego

$$u_k = (S\Lambda^k)c = \sum_{i=1}^{n} c_i \lambda_i^k x^i.$$

Ejemplo: La sucesión de Fibonacci se construye con la ecuación

$$F_{k+2} = F_{k+1} + F_k,\ k \geq 2,\ F_0 = 0,\ F_1 = 1.$$

Entonces, para $k \geq 2$, la ley $F_{k+2} = F_{k+1} + F_k$ es equivalente al sistema

$$\begin{aligned} F_{k+2} &= F_{k+1} + F_k \\ F_{k+1} &= F_{k+1} \end{aligned},$$

y si llamamos $u_k = (F_{k+1}, F_k)$, el sistema puede expresarse como ecuaciones en diferencias al ser reescrito como

$$u_{k+1} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} u_k = Au_k. \quad \textit{(Fin del ejemplo)}$$

3. Matrices Complejas

Se especificó antes que cuando se trata de autovalores, es necesario trabajar en el campo de los complejos y ver a $\mathbb{R}^n$ como un subconjunto de vectores el espacio vectorial $\mathbb{C}^n$ sobre $\mathbb{C}$. Considerando el producto interno $\langle z, w\rangle = \bar{z}^T w$, la norma que se obtiene es $||z|| = \sqrt{\sum_{i=1}^{n} |z_i|^2}$, donde $|z_i|$ es el módulo de $z_i \in \mathbb{C}$. De esta manera, el producto interno y la norma coinciden con las habituales al considerar el subespacio $\mathbb{R}^n$.

Definición 46. Dada una matriz compleja $A = (a_{ij})$ $m \times n$, se define a $\bar{A} = (b_{ij})$ $m \times n$ tal que $b_{ij} = \overline{a_{ij}}$.

Definición 47. Dada una matriz compleja A $m \times n$, se define la matriz A^H **hermitiana** de A como $A^H = \bar{A}^T$.

Teniendo en cuenta la definición de matriz hermitiana, tenemos que si A es una matriz real, $A^H = A^T$. Por otro lado, si $z, w \in \mathbb{C}^n$, $\langle z, w\rangle = \bar{z}^T w = z^H w$.

Definición 48. Una matriz A es **hermitiana** si $A^H = A$.

- Si A es una matriz real, A es hermitiana si y solo si A es simétrica.
- Las matrices hermitianas son cuadradas.
- La diagonal de una matriz hermitiana tiene entradas reales.

Lema 52. *Si A es una matriz hermitiana, entonces para todo $x \in \mathbb{C}^n$, $x^H Ax \in \mathbb{R}$.*

Demostración. $x^H Ax = x^H(Ax) = x^H \sum_{i=1}^{n} x_i A^i = \sum_{i=1}^{n} x_i(x^H A^i) = \sum_{i=1}^{n} x_i \sum_{j=1}^{n} \bar{x}_j A_j^i = \sum_{i=1}^{n}\sum_{j=1}^{n} (x_i\bar{x}_j) A_j^i$. Luego,

$$\overline{x^H Ax} = \sum_{i=1}^{n}\sum_{j=1}^{n} (x_j\bar{x}_i)\overline{A_j^i} = \sum_{i=1}^{n}\sum_{j=1}^{n} (x_j\bar{x}_i) A_i^j = \sum_{j=1}^{n}\sum_{i=1}^{n} (x_j\bar{x}_i) A_i^j = x^H Ax,$$

por lo que $x^H Ax \in \mathbb{R}$. □

Teorema 26. *Si A es una matriz compleja hermitiana, sus autovalores son reales.*

Demostración. Sea $\lambda \in \mathbb{C}$ un autovalor de A y $x \in \mathbb{C}^n$ tal que $Ax = \lambda x$. Entonces,

$$x^H Ax = x^H \lambda x = \lambda ||x||^2.$$

Como $x^H Ax$ y $||x||^2$ son valores reales, $\lambda \in \mathbb{R}$. □

Teorema 27. *Si A es una matriz hermitiana y $\lambda_1 \neq \lambda_2$ son autovalores de A, entonces el autoespacio de λ_1 es ortogonal al autoespacio de λ_2.*

Demostración. Para $i = 1, 2$, sea z^i un autovector asociado a λ_i. Debemos probar que $z^1 \perp z^2$, o, equivalentemente, que $(z^1)^H z^2 = 0$.

$$(\lambda_1 z^1)^H z^2 = (Az^1)^H z^2 = (z^1)^H A^H z^2 = (z^1)^H Az^2 = (z^1)^H \lambda_2 z^2.$$

Como $\lambda_{1,2}$ son reales, tenemos que $\lambda_1 (z_1)^H z^2 = \lambda_2 (z_1)^H z^2$, y como $\lambda_1 \neq \lambda_2$, resulta $(z^1)^H z^2 = 0$. □

Corolario 10. *Si A es hermitiana y S diagonaliza a A, S puede ser elegida con sus columnas ortonormales.*

Demostración. Si existen p autovalores distintos de A, hay exactamente p autoespacios asociados a A, y tenemos probado que cada uno de esos autoespacios tiene una base ortonormal, y al ser A hermitiana, son ortogonales entre sí. Luego, $B = \bigcup_{i=1}^{p} B_i$ es un conjunto de vectores ortonormales, y si se construye la matriz S donde sus columnas son los vectores de B, se tiene que S diagonaliza a A (pues los vectores de B son todos autovectores asociados a los autovalores de A) y sus columnas son ortonormales. □

Corolario 11. *Si A es real, simétrica y diagonalizable, $A = Q\Lambda Q^T$, con Q matriz ortogonal.*

Demostración. Si A es real y simétrica, $A^H = A$, y por el corolario anterior, existe una matriz Q con columnas ortonormales tal que $A = Q\Lambda Q^T$. □

Definición 49. Extendiendo el concepto de matrices ortogonales reales, una matriz compleja (cuadrada) con columnas ortonormales se denomina **matriz unitaria**. Una matriz U de tamaño $n \times n$ unitaria verifica:

1. $U^H U = UU^H = I \implies U^{-1} = U^H$. Más aún, U es unitaria si y solo si $U^H U = UU^H = I$.

   Demostración. **Ida.** Sea $U = [q_1, \ldots, q_n]$ una matriz unitaria, con $q_i \in \mathbb{R}^n$ para todo $i = 1, \ldots, n$. Planteo el producto $U^H U$:

$$U^H U = \begin{bmatrix} \bar{q}_1^T \\ \bar{q}_2^T \\ \vdots \\ \bar{q}_n^T \end{bmatrix} \begin{bmatrix} q_1 & q_2 & \cdots & q_n \end{bmatrix} = \begin{bmatrix} \bar{q}_1^T q_1 & \bar{q}_1^T q_2 & \cdots & \bar{q}_1^T q_n \\ \bar{q}_2^T q_1 & \bar{q}_2^T q_2 & \cdots & \bar{q}_2^T q_n \\ \bar{q}_3^T q_1 & \bar{q}_3^T q_2 & \cdots & \bar{q}_3^T q_n \\ \vdots & \vdots & \ddots & \vdots \\ \bar{q}_3^T q_1 & \bar{q}_3^T q_2 & \cdots & \bar{q}_3^T q_n \end{bmatrix}.$$

Como U es unitaria, tenemos que $q_i^H q_j = \langle q_i, q_j \rangle = 0$ si $i \neq j$. Además, $q_i^H q_i = \langle q_i, q_i \rangle = ||q_i||^2 = 1$. Entonces,

$$U^H U = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} = I_n \implies U^H = U^{-1}.$$

Vuelta. Sea $U = (u_{ij})$ una matriz tal que $U^H = U^{-1}$. Como U es no singular, sus columnas son l.i.. Por otro lado,

$$(U^H U)_{ij} = \sum_{k=1}^{n} \bar{u}_{ki} u_{kj} = (U^i)^H U^j = \langle U^i, U^j \rangle = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}.$$

Entonces, si $i \neq j$, $\langle U^i, U^j \rangle = 0$. Es decir, las columnas de U son ortogonales. Además, $\langle U^i, U^i \rangle = ||U^i||^2 = 1 \implies ||U^i|| = 1$, por lo que las columnas de U tienen norma 1. Por lo tanto, las columnas de U son ortonormales, y U resulta unitaria. □

2. Para todo $x \in \mathbb{C}^n$, $||Ux|| = ||x||$.

3. Todos los autovalores de U tienen módulo 1. Además, a autovalores diferentes le corresponden autovectores ortogonales.

Demostración. Sean λ y x, autovalores de U y autovector correspondiente. Entonces,

$$||x|| = ||Ux|| = ||\lambda x|| = |\lambda| \, ||x|| \implies |\lambda| = 1.$$

Para $i = 1, 2$, sean λ_i y x^i autovalores de U con sus autovectores correspondientes, con $\lambda_1 \neq \lambda_2$. Entonces,

$$(x^1)^H x^2 = (x^1)^H (U^H U) x^2 = (Ux^1)^H (Ux^2) = (\lambda x^1)^H (\lambda_2 x^2) = \bar{\lambda}_1 \lambda_2 (x^1)^H x^2.$$

Por lo tanto, $\bar{\lambda}_1 \lambda_2 = 1$ o $(x^1)^H x^2 = 0$. Supongamos que $\bar{\lambda}_1 \lambda_2 = 1$. Como $|\lambda_1|^2 = \bar{\lambda}_1 \lambda_1 = 1$, tenemos que $\bar{\lambda}_1 \lambda_2 = \bar{\lambda}_1 \lambda_1 = 1$ y entonces $\lambda_1 = \lambda_2$, una contradicción. Por lo tanto, $(x^1)^H x^2 = 0$ y $x^1 \perp x^2$. □

Definición 50. Una matriz K es **hermitiana sesgada** si $K^H = -K$.

Lema 53. *Sea K una matriz sesgada hermitiana. Entonces,*

1. $K = iA$, *con A matriz hermitiana.*

Demostración. Dados $i, j = 1, \ldots, n$, $(K^H)_i^j = -K_i^j$. Esto sucede sólo si K_i^j es un número imaginario (Si $a, b \in \mathbb{R}$, $a + bi = -(a - bi) \iff a = 0$). Por lo tanto, $K_i^j = ik_{ij}$, con $k_{ij} \in \mathbb{R}$. Entonces, si $A = (k_{ij})$, $K = iA$. Resta probar que A es hermitiana. Equivalentemente, probaré que $-iK$ es hermitiana. Efectivamente,

$$(-iK)^H = iK^H = -iK.$$

□

2. Los autovalores de K son imaginarios.

   Demostración. $K = iA$, con A hermitiana. Entonces, sus autovalores son reales. Luego, $Kx = iAx = A(ix) = (i\lambda)x$. Es decir, los autovalores de K son de la forma $i\lambda$, donde λ es un autovalor de A, y como $\lambda \in \mathbb{R}$, luego $i\lambda$ es imaginario. □

3. Los autovectores de K y A coinciden.

   Demostración. Se vio que $Kx = iAx = A(ix) = (i\lambda)x$. Sea x un autovector de A asociado a lambda. Luego, $Ax = \lambda x$ y cualquier múltiplo de x es un autovector. Sea $x' = ix$. Luego,

$$Ax' = A(ix) = (iA)x = Kx = \lambda(ix) = (i\lambda)x,$$

   resultando ser un autovector de K.

   Recíprocamente, sea x un autovector de K asociado a $\lambda = i\lambda'$, con $\lambda' \in \mathbb{R}$. Nótese que λ' es un autovalor de A. Además, cualquier múltiplo de x es un autovector de K, por lo que tomando $x' = \frac{1}{i}x$, x' es un autovector de K asociado a λ. Entonces,

$$Kx' = K\left(\frac{1}{i}x\right) = (iA)\left(\frac{1}{i}x\right) = Ax = \lambda x' = i\lambda'\left(\frac{1}{i}x\right) = \lambda x,$$

   resultando ser un autovector de A asociado a λ. □

3.1. Matrices Semejantes

Definición 51. Dada una matriz inversible M, la transformación que a toda matriz A la lleva a $M^{-1}AM$ es una **transformación de similaridad** o **semejanza**. Se dice que A **es semejante a** B si existe M inversible tal que $B = M^{-1}AM$.

Observación 29. La relación de semejanza entre matrices es una relación de equivalencia.

Las transformaciones de semejanza aparecen naturalmente en los cambios de variables en sistemas lineales. Por ejemplo, sea el sistema de diferencias $u_{k+1} = Au_k,\ k \in \mathbb{N}$. Haciendo el cambio de variables $u = Mv$ (con M inversible), el sistema se transforma en $Mv_{k+1} = AMv_k$, o, equivalentemente, $v_{k+1} = (M^{-1}AM)v_k$.

Los cambios de variables se realizan cuando el nuevo sistema resulta más fácil de resolver que el inicial. Por ejemplo, si S diagonaliza a A, el cambio de variables con $M = S$ nos lleva a un sistema *desacoplado*, que es la forma más sencilla a la que podemos aspirar.

Lema 54. *Sean A, M y $B = M^{-1}AM$ matrices $n \times n$. Entonces, A y B tienen los mismos autovalores. Además, si x es un autovector de A correspondiente a λ, entonces $M^{-1}x$ es un autovector de B correspondiente a λ.*

Demostración. Sea λ un autovalor de A y x un autovector asociado. Entonces, $Ax = \lambda x$. Luego,

$$Ax = MBM^{-1}x = \lambda x \implies M^{-1}MBM^{-1}x = M^{-1}\lambda x \implies B(M^{-1}x) = \lambda(M^{-1}x).$$

Por lo tanto, λ es un autovalor de B y $M^{-1}x$ su autovector asociado. □

Lema 55. *Las matrices semejantes tienen el mismo polinomio característico.*

Demostración. Sean A y $B = M^{-1}AM$, con M inversible. Sea λ un autovalor de A y B. Entonces,

$$\begin{aligned}
\det(A - \lambda I) &= \det(M^{-1}) \det(A - \lambda I) \det(M) \\
&= \det(M^{-1}(A - \lambda I)M) \\
&= \det(M^{-1}AM - \lambda(M^{-1}IM)) \\
&= \det(B - \lambda I).
\end{aligned}$$

□

Lema 56. *Sea T una transformación lineal de un espacio vectorial de dimensión finita n V en sí mismo, y $\mathcal{B}_1, \mathcal{B}_2$ dos bases ordenadas de V. Sean A y B las matrices asociadas a T considerando las bases $\mathcal{B}_1$ y $\mathcal{B}_2$, respectivamente. Entonces, A es semejante a B.*

Demostración. Dado $v \in V$, del enunciado tenemos que $A[v]_{\mathcal{B}_1} = [Tv]_{\mathcal{B}_1}$ y $B[v]_{\mathcal{B}_2} = [Tv]_{\mathcal{B}_2}$. Sea M la matriz de cambio de base de $\mathcal{B}_2$ a $\mathcal{B}_1$. Por definición, M verifica $M[v]_{\mathcal{B}_2} = [v]_{\mathcal{B}_1}$. Entonces,

$$A[v]_{\mathcal{B}_1} = [Tv]_{\mathcal{B}_1} \implies AM[v]_{\mathcal{B}_2} = M[Tv]_{\mathcal{B}_2} \implies M^{-1}AM[v]_{\mathcal{B}_2} = [Tv]_{\mathcal{B}_2}.$$

Dado que B verifica $B[v]_{\mathcal{B}_2} = [Tv]_{\mathcal{B}_2}$ y la matriz asociada a una transformación por una base es única, tenemos que

$$B[v]_{\mathcal{B}_2} = M^{-1}AM[v]_{\mathcal{B}_2} \implies B = M^{-1}AM.$$

Por lo tanto, como existe M inversible tal que $B = M^{-1}AM$, A y B son matrices semejantes. □

3.2. Lema de Schur

Teorema 28. *(Lema de Schur) Sea A una matriz $n \times n$. Entonces, existe una matriz unitaria U tal que $U^{-1}AU = T$, con T matriz triangular.*

Demostración. Por inducción en n. Si $n = 1$, la propiedad es trivial y la matriz diagonalizante es I_1. Supongo que para toda matriz A de tamaño $k \times k$ existe una matriz unitaria U_k tal que $U_k^{-1}AU_k = T_k$, con T_k triangular.

Sea A una matriz de tamaño $(k+1) \times (k+1)$. Tomamos λ_1 autovalor de A y x^1 un autovector normalizado asociado a λ_1. Sea U_1 una matriz unitaria $(k+1) \times (k+1)$ cuya primer columna es x^1 (existe, por G-S), y sea $M = U_1^{-1}AU_1$. Si M^1 es la primer columna de M, tenemos

$$M^1 = Me^1 = (U_1^H AU_1)e^1 = U_1^H A(U_1 e^1) = U_1^H Ax^1 = \lambda_1(U_1^H x^1).$$

Como las filas de U_1^H son las columnas de U_1 conjugadas, y las columnas de U son ortonormales, la primer componente de $U_1^H x^1$ es $||x||^2 = 1$ y las componentes restantes son nulas. Así, existen M_k matriz $k \times k$ y $b = (b_1 \cdots b_k)$ tales que

$$U_1^{-1}AU_1 = \begin{bmatrix} \lambda_1 & b_1 & \cdots & b_k \\ 0 & & & \\ \vdots & & M_k & \\ 0 & & & \end{bmatrix}.$$

Por hipótesis de inducción, existe U_k unitaria tal que $U_k^{-1}M_kU_k = T_k$, con T_k triangular. Sea U_2 la matriz $(k+1) \times (k+1)$ cuya primer columna y primera fila son el vector $e^1 \in \mathbb{R}^{k+1}$ y U_k es su submatriz $k \times k$ en su esquina inferior derecha. Es fácil verificar que U_2 es unitaria. Además,

$$U_2^{-1} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & U_k^{-1} & \\ 0 & & & \end{bmatrix}.$$

Finalmente, si $U_{k+1} = U_1U_2$, entonces U_{k+1} es unitaria. Solo resta probar que $U_{k+1}^{-1}AU_{k+1} = T_{k+1}$, con T_{k+1} triangular. Tenemos

$$U_{k+1}^{-1}AU_{k+1} = (U_2^{-1}U_1^{-1})A(U_1U_2) = U_2^{-1}(U_1^{-1}AU_1)U_2 = U_2^{-1}MU_2.$$

Por lo tanto,

$$U_{k+1}^{-1}AU_{k+1} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & U_k^{-1} & \\ 0 & & & \end{bmatrix} \begin{bmatrix} \lambda_1 & b_1 & \cdots & b_k \\ 0 & & & \\ \vdots & & M_k & \\ 0 & & & \end{bmatrix} \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & U_k & \\ 0 & & & \end{bmatrix} =$$

$$= \begin{bmatrix} \lambda_1 & b_1 & \cdots & b_k \\ 0 & & & \\ \vdots & & U_k^{-1}M_k & \\ 0 & & & \end{bmatrix} \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & U_k & \\ 0 & & & \end{bmatrix} = \begin{bmatrix} \lambda_1 & bU_k \\ 0 & \\ \vdots & U_k^{-1}M_kU_k \\ 0 & \end{bmatrix} = \begin{bmatrix} \lambda_1 & bU_k \\ 0 & \\ \vdots & T_k \\ 0 & \end{bmatrix}.$$

Por lo tanto, $U_{k+1}^{-1}AU_{k+1} = T_{k+1}$ es una matriz triangular. □

Teorema 29. *Toda matriz compleja hermitiana (resp. real simétrica) A puede ser diagonalizable por una matriz unitaria U (resp. ortogonal Q).*

Demostración. Sea A una matriz hermitiana. Por el Lema de Schur, existen U matriz unitaria y T matriz triangular tales que $U^{-1}AU = T$. Entonces,

$$T^H = (U^{-1}AU)^H = U^H A^H (U^{-1})^H = U^H AU = T.$$

Así, T es una matriz triangular tal que $T = T^H$. Por lo tanto, T es diagonal. □

Este teorema introduce la *descomposición espectral* de las matrices hermitianas. Es decir, toda matriz hermitiana puede descomponerse en una suma de matrices de rango 1. Más específicamente, puede ser escrita como combinación lineal de las matrices de proyección sobre los autoespacios de sus autovectores diferentes. Por lo tanto, si A es una matriz hermitiana, con $\lambda_1,\ i = 1, \ldots, k$ sus autovalores diferentes, tenemos que

$$A = \sum_{i=1}^{k} \lambda_i P_i,$$

con P_i matriz de proyección sobre autoespacio asociado a λ_i.

3.3. Matrices Normales

Existe una clase de más amplia de matrices (es decir, que contiene a las matrices hermitianas) las cuales resultan ser diagonalizables por una matriz unitaria.

Definición 52. N es una **matriz normal** si conmuta con su hermitiana. Es decir, si $NN^H = N^H N$.

Observación 30. Las matrices hermitianas y unitarias son normales.

Lema 57. *Sea T una matriz triangular $n \times n$. Entonces, T es normal si y solo si T es diagonal.*

Demostración. **Ida.** Sin pérdida de generalidad, considero T triangular superior. Luego,

$$(T^H T)_{11} = \sum_{k=1}^{n} \overline{T}_{k1} T_{k1} = |T_{11}|^2, \text{ y } (TT^H)_{11} = \sum_{k=1}^{n} T_{1k} \overline{T}_{1k} = \sum_{k=1}^{n} |T_{1k}|^2.$$

Entonces, como es normal, $|T_{11}|^2 = \sum_{k=1}^{n} |T_{1k}|^2$, de donde se deduce que $T_{1k} = 0$ para todo $k = 2, \ldots, n$. Repitiendo el procedimiento en la segunda fila, tenemos que

$$(T^H T)_{22} = |T_{22}|^2, \text{ y } (TT^H)_{22} = \sum_{k=1}^{n} T_{2k} \overline{T}_{2k} = \sum_{k=1}^{n} |T_{2k}|^2,$$

y como A es normal, se deduce nuevamente que $T_{2k} = 0$ para todo $k = 3, \ldots, n$. Continuando de esta manera hasta la última fila, se llega a la conclusión de que $T_{ij} = 0$ si $i \neq j$, y por lo tanto T es diagonal. □

Teorema 30. *Sea A matriz $n \times n$. Entonces, A es diagonalizable por una matriz unitaria U si y solo si A es normal.*

Demostración. **Vuelta.** Sea A una matriz normal. Por el Lema de Schur, existe U unitaria tal que $U^{-1}AU = U^H AU = T$, con T triangular superior. Luego, se tiene que

$$TT^H = (U^H AU)(U^H A^H U) = U^H (AA^H) U = U^H (A^H A) U = U^H A^H (UU^H) AU = (U^H A^H U)(U^H AU) = T^H T.$$

Por lo tanto, T es normal, y como T es triangular, T es diagonal, y $T = \Lambda$.

Ida. Sea U matriz unitaria tal que $U^H AU = \Lambda$. Luego, como las matrices diagonales conmutan,

$$AA^H = (U\Lambda U^H)(U\Lambda^H U^H) = U\Lambda\Lambda^H U^H = U\Lambda^H \Lambda U^H = (U\Lambda^H U^H)(U\Lambda U^H) = A^H A,$$

y por lo tanto A es normal. □

Corolario 12. *A es normal si y solo si tiene n autovectores ortogonales.*

Demostración. A es normal si y solo si es diagonalizable por una matriz unitaria, y como las columnas de esa matriz son autovectores de A, al ser unitaria son ortogonales. □

3.4. Forma de Jordan

El Lema de Schur permite llegar hasta una matriz triangular a través de una transformación de semejanza definida por una matriz unitaria. Si se permiten transformaciones de semejanza con cualquier matriz no singular M, la forma más *rala* posible a la que se puede llevar a una matriz es la denominada **Forma de Jordan**.

Definición 53. Si A es una matriz $n \times n$ con s autovectores l.i., entonces A es semejante a una matriz *diagonal por bloques* J_s, con s bloques de Jordan $J_1, \ldots, J_s$.

Cada bloque de Jordan $J_i,\ i = 1, \ldots, s$ es una matriz triangular superior que tiene un solo autovalor λ_i con autoespacio de dimensión 1. Las entradas de la diagonal son iguales a λ_i, las entradas en la segunda diagonal superior son iguales a 1, y el resto de las entradas son nulas. Es decir, cada bloque de Jordan es de la forma

$$J_i = \begin{bmatrix} \lambda_i & 1 & & & \\ & \lambda_i & 1 & & \\ & & \ddots & \ddots & \\ & & & \lambda_i & 1 \\ & & & & \lambda_i \end{bmatrix}.$$

Se dice que J es diagonal por bloques porque sus bloques se ubican en la diagonal, y fuera de ellos las entradas son todas nulas. Es decir, toda matriz con s autovectores l.i. es semejante a una matriz J de la forma

$$J = \begin{bmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_s \end{bmatrix}.$$

Es importante destacar que la forma de Jordan *caracteriza* a las matrices semejantes. Es decir, dos matrices son semejantes si y solo si tienen la misma forma de Jordan.

3.5. Multiplicidades Aritméticas y Geométricas

Teorema 31. *Sea V un espacio vectorial sobre $\mathbb{K}$ y T una transformación lineal de V en sí mismo. Para $i = 1, \ldots, p$, sean $\lambda_i \in \mathbb{K}$, autovalores diferentes de T y $\mathcal{V}_i$ un conjunto de vectores l.i. en el autoespacio correspondiente a λ_i. Entonces, $\mathcal{V} = \bigcup\limits_{i=1}^{p} \mathcal{V}_i$ es un conjunto de vectores l.i. de V.*

Demostración. Considero una combinación lineal nula de vectores de $\mathcal{V}$:

$$\sum_{i=1}^{p} \sum_{j=1}^{t(i)} \alpha_{ij} v_j^i = 0,$$

donde para cada $i = 1, \ldots, p$, $v_j^i \in \mathcal{V}_i$ para todo $j = 1, \ldots, t(i)$.

Para cada $i = 1, \ldots, p$, sea $u^i = \sum\limits_{j=1}^{t(i)} \alpha_{ij} v_j^i$. Claramente, u^i es un elemento del autoespacio asociado a λ_i. Luego, la combinación lineal nula anterior puede expresarse como

$$\sum_{i=1}^{p} u^i = 0.$$

Ahora, como los autovectores correspondientes a autovalores diferentes son l.i., no es posible que su suma sea el vector nulo. Por lo tanto, $u^i = 0$ para todo $i = 1, \ldots, p$. Entonces, se tiene la nueva combinación lineal nula

$$u^i = \sum_{j=1}^{t(i)} \alpha_{ij} v_j^i = 0.$$

Como los vectores v_j^i, $j = 1, \ldots, t(i)$ son l.i., $\alpha_{ij} = 0$ para todo $j = 1, \ldots, t(i)$. Por lo tanto, los vectores de $\mathcal{V}$ son l.i.. □

Definición 54. Dada una matriz A y un autovalor λ de A, se denomina **multiplicidad algebraica** a la multiplicidad del autovalor como raíz del polinomio característico, y **multiplicidad geométrica** a la dimensión del autoespacio asociado a λ.

Teorema 32. *La multiplicidad geométrica de un autovalor es siempre a lo sumo su multiplicidad algebraica.*

Juntando el teorema y corolario anteriores, se concluye que una matriz A es diagonalizable si y solo si la multiplicidad geométrica de todo autovalor es igual a la algebraica.

Lema 58. *Todo bloque de Jordan tiene un único autovalor con multiplicidad geométrica 1.*

Observación 31. Existen matrices no diagonalizables ni semejantes que tienen el mismo polinomio característico.

Teorema 33. *Sea A una matriz $n \times n$ y λ_i, $i = 1, \ldots, p$ los autovalores distintos de A, con multiplicidad algebraica $ma(i)$, multiplicidad geométrica $mg(i)$ y $\mathcal{B}_i$ una base de su autoespacio. Entonces, las siguientes condiciones son equivalentes:*

1. A es diagonalizable.
2. Para todo $i = 1, \ldots, p$, $ma(i) = mg(i)$.
3. $\sum\limits_{i=1}^{p} mg(i) = n$.
4. $\bigcup\limits_{i=1}^{p} \mathcal{B}_i$ es una base de $\mathbb{K}^n$.

Demostración. **(1) $\implies$ (2):** Sea S diagonalizante de A. Luego, para cada $i = 1, \ldots, p$ existen $ma(i)$ columnas de S que corresponden a autovectores asociados a λ_i. Como S es inversible, esos autovectores son l.i.. Por lo tanto, existen $ma(i)$ vectores l.i. en el autoespacio asociado a λ_i, lo que implica que $ma(i) \leq mg(i)$. Como sabemos que $ma(i) \geq mg(i)$, resulta $ma(i) = mg(i)$.

(2) $\implies$ (3): Trivial.

(3) $\implies$ (4): Ya se probó que $\bigcup\limits_{i=1}^{p} \mathcal{B}_i$ es un conjunto de vectores l.i. de $\mathbb{K}^n$. Además, $\mathcal{B}_i \cap \mathcal{B}_j = \emptyset$ si $i \neq j$, pues si no fuera así, existiría un vector $x \in \langle \mathcal{B}_i \rangle \cap \langle \mathcal{B}_j \rangle$ para algunos i, j, y así $Ax = \lambda_i x = \lambda_j x \implies \lambda_i = \lambda_j$, una contradicción. Luego,

$$\left| \bigcup_{i=1}^{p} \mathcal{B}_i \right| = \sum_{i=1}^{p} |\mathcal{B}i| = \sum_{i=1}^{p} mg(i) = n.$$

Por lo tanto, $\bigcup\limits_{i=1}^{p} \mathcal{B}_i$ es una base de $\mathbb{K}^n$.

(4) $\Longrightarrow$ (1): Si $\bigcup\limits_{i=1}^{p} \mathcal{B}_i$ es una base de $\mathbb{K}^n$, entonces $\left|\bigcup\limits_{i=1}^{p} \mathcal{B}_i\right| = \sum\limits_{i=1}^{p} |\mathcal{B}i| = n$. Por lo tanto,

$$n = \sum_{i=1}^{p} |\mathcal{B}i| = \sum_{i=1}^{p} mg(i) \leq \sum_{i=1}^{p} ma(i) = n,$$

y para todo $i = 1, \ldots, p$, $|\mathcal{B}_i| = ma(i)$ (pues $mg(i) = ma(i)$).

Sea S la matriz $n \times n$ que en sus columnas tiene los n vectores l.i. de $\bigcup\limits_{i=1}^{p} \mathcal{B}_i$. Luego, $(AS)^i = AS^i = Ax^i = \lambda_i x^i = S\Lambda^i = (S\Lambda)^i$, de donde resulta que $AS = S\Lambda$, y como los n autovectores son l.i., S es inversible y luego diagonaliza a A. □

Método para determinar si una matriz A $n \times n$ es diagonalizable:

1. Calcular los autovalores diferentes de A: $\lambda_i,\ i = 1, \ldots, p$.

2. Para todo $i = 1, \ldots, p$, calcular una base $\mathcal{B}_i$ de $N(A - \lambda_i I)$.

3. Si $\sum\limits_{i=1}^{p} |\mathcal{B}_i| < n$, A no es diagonalizable, y termina el proceso.

4. A es diagonalizable, y la matriz S que tiene por columnas los vectores en $\bigcup\limits_{i=1}^{p} \mathcal{B}_i$ es matriz diagonalizante de A.

El paso más costoso computacionalmente es el paso 1, ya que el cálculo de los autovalores se realiza resolviendo el polinomio característico, el cual requiere el desarrollo de un determinante (que en grandes matrices es complejo). Para la búsqueda de raíces del polinomio no existen fórmulas explícitas a partir de polinomios de grado 5, y se debe hacer uso de métodos numéricos.

Sin embargo, para simplificar cálculos, los autovalores se calculan por una secuencia de transformaciones de semejanza simples, llevándola a su fórmula triangular, que ya sabemos que existe por el Lema de Schur.

Unidad VI

El Algoritmo PageRank

Al pensar en cómo debería funcionar un motor de búsqueda como Google, a primera vista parece razonable imaginar que todo lo que hace el algoritmo es mantener indexadas a todas las páginas Web, y cuando el usuario inserta un comando de búsqueda, el algoritmo navega a través del índice y cuenta las ocurrencias de las palabras a buscar. Luego, los primeros resultados son las páginas con el mayor número de ocurrencias de las palabras de interés.

Este era el método considerado correcto en los años 90, cuando los motores de búsqueda usaban *sistemas de clasificación basados en texto*. Sin embargo, este enfoque conlleva algunos problemas. Por ejemplo, la búsqueda de un término común como "Internet" fue problemática. La primera página mostrada por un motor de búsqueda de este estilo estaba escrita en chino, con repetidas ocurrencias de la palabra, sin contener información adicional. Como otro ejemplo, supongamos que se quiere buscar información sobre la Universidad de Cornell. Luego, se busca la palabra "Cornell", y se espera que "www.cornell.edu" sea el primer resultado. Sin embargo, al contar el número de ocurrencias en las páginas de la palabra buscada, este puede no ser el caso: ¿qué pasa si alguien decide diseñar una página Web con la palabra "Cornell" un millón de veces? No tendría sentido que esa página sea el primer resultado en la búsqueda.

La utilidad de un motor de búsqueda depende de la *relevancia* del conjunto de resultados que retorna. Podría haber un millón de páginas Web que incluyan una palabra o frase concreta, pero inevitablemente algunas de ellas serán más relevantes, populares o fidedignas. Un usuario no posee la habilidad de revisar todas las páginas que contienen las palabras de interés. Por esta razón, lo esperable es que las páginas relevantes se encuentren entre las primeras 20-30 páginas retornadas.

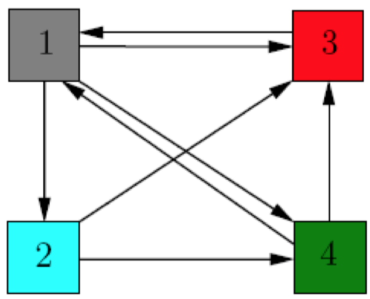
Los motores de búsqueda modernos utilizan otros métodos de clasificación de resultados para proveer las páginas más relevantes en una búsqueda dada. Uno de los más conocidos e influyentes algoritmos para calcular la relevancia de páginas Web es el Algoritmo Page Rank usado por Google. Fue inventado por Larry Page y Sergey Brin mientras realizaban sus estudios en Stanford, y se convirtió en una patente de Google en 1998. La idea sobre la que se centra este algoritmo es que la importancia de cualquier página puede ser determinada al mirar las páginas que poseen vínculos (links) a ella. Si creamos una página Web i e incluimos un vínculo a la página j, esto significa que consideramos a j como importante y relevante para el tópico particular de nuestra página. Por otro lado, si solo una página tiene vínculos a la página j, digamos, la página k, pero k es una página relevante, podemos decir que k afirma que j es relevante. Independientemente de si hablamos de popularidad o confianza, podemos asignar iterativamente un rango a cada página basado en los rangos de las páginas que apuntan a ella.

Con este objetivo, comenzamos visualizando la Web como un grafo dirigido, donde los nodos representan páginas Web y las aristas representan los vínculos entre ellas.

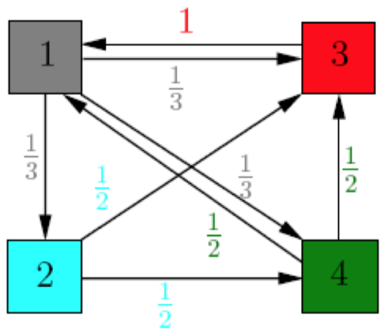
Supongamos, por ejemplo, que nuestra Internet consiste de cuatro páginas: página 1, 2, 3 y 4, referidas entre sí de la siguiente manera:

- La página 1 tiene vínculos a las páginas 2, 3, y 4.
- La página 2 tiene vínculos a las páginas 3, y 4.
- La página 3 tiene vínculos a la página 1.
- La página 4 tiene vínculos a las páginas 1 y 3.

Podemos plasmar la información en un grafo dirigido con cuatro nodos, uno por cada página. Cuando la página i refiere a la página j, añadimos una arista dirigida del nodo i al j. Naturalmente, con el objetivo de calcular el rango de cada página, se ignoran links de navegación como los botones ”Atrás” y ”Adelante”, pues sólo nos importa conocer las conexiones entre diferentes páginas. Luego de analizar las páginas y sus conexiones, tenemos el siguiente grafo:



En nuestro modelo, cada página debería transferir equitativamente su importancia a las páginas a las que refiere. El nodo 1 tiene 3 aristas salientes, por lo que pasará $\frac{1}{3}$ de su importancia a cada uno de esos nodos. El nodo 3 tiene solo una arista saliente, por lo que pasará toda su importancia al nodo 1. En general, si un nodo tiene k aristas salientes, pasará $\frac{1}{k}$ de su importancia a cada uno de los nodos a los que refiere. Visualizamos mejor este proceso mediante la asignación de un peso a cada arista.



Llamemos A a la matriz de transición del grafo. Es decir, $A = \begin{bmatrix} 0 & 0 & 1 & \frac{1}{2} \\ \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} & 0 & 0 \end{bmatrix}$.

1. Enfoque desde Sistemas Dinámicos

Supongamos que inicialmente la importancia está uniformemente distribuida entre los 4 nodos, es decir, cada uno tiene $\frac{1}{4}$. Llamemos v al vector de rango inicial, con todas sus entradas iguales a $\frac{1}{4}$. Cada link entrante aumenta la importancia de una página, por lo que, como primer paso, actualizamos el rango de cada página sumando al valor actual la importancia de los vínculos entrantes. Esto es equivalente a multiplicar la matriz A con v. Entonces, en el primer paso, el nuevo vector de importancia es $v_1 = Av$. Repitiendo el proceso, en el paso dos, el vector de importancia es $v_2 = A(Av) = A^2v$. Luego, mediante cálculo computacional, tenemos:

$$v = \begin{pmatrix} 0{,}25 \\ 0{,}25 \\ 0{,}25 \\ 0{,}25 \end{pmatrix}, \quad Av = \begin{pmatrix} 0{,}37 \\ 0{,}08 \\ 0{,}33 \\ 0{,}20 \end{pmatrix}, \quad A^2v = \begin{pmatrix} 0{,}43 \\ 0{,}12 \\ 0{,}27 \\ 0{,}16 \end{pmatrix},$$

$$A^3v = \begin{pmatrix} 0{,}35 \\ 0{,}14 \\ 0{,}29 \\ 0{,}20 \end{pmatrix}, \quad A^4v = \begin{pmatrix} 0{,}39 \\ 0{,}11 \\ 0{,}29 \\ 0{,}19 \end{pmatrix}, \quad A^5v = \begin{pmatrix} 0{,}39 \\ 0{,}13 \\ 0{,}28 \\ 0{,}19 \end{pmatrix},$$

$$A^6v = \begin{pmatrix} 0{,}38 \\ 0{,}13 \\ 0{,}29 \\ 0{,}19 \end{pmatrix}, \quad A^7v = \begin{pmatrix} 0{,}38 \\ 0{,}12 \\ 0{,}29 \\ 0{,}19 \end{pmatrix}, \quad A^8v = \begin{pmatrix} 0{,}38 \\ 0{,}12 \\ 0{,}29 \\ 0{,}19 \end{pmatrix}.$$

Nótese que las iteraciones $v, Av, \ldots, A^k v$ tienden al valor de equilibrio $v^* = \begin{pmatrix} 0{,}38 \\ 0{,}12 \\ 0{,}29 \\ 0{,}19 \end{pmatrix}$. Llamamos a este el vector PageRank de nuestro grafo.

2. Enfoque desde el Álgebra Lineal

Llamemos x_1, x_2, x_3 y x_4 a la importancia de las cuatro páginas. Analizando la situación de cada nodo, tenemos el siguiente sistema:

$$\begin{cases} x_1 = x_3 + \frac{1}{2}x_4 \\ x_2 = \frac{1}{3}x_1 \\ x_3 = \frac{1}{3}x_1 + \frac{1}{2}x_2 + \frac{1}{2}x_4 \\ x_4 = \frac{1}{3}x_1 + \frac{1}{2}x_2 \end{cases}$$

Esto es equivalente a preguntarse las soluciones de la ecuación

$$A \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}.$$

Es decir, debemos buscar los autovectores asociados al autovalor 1 de la matriz. Realizando los cálculos, tenemos que el autoespacio asociado al autovalor 1 es $\left\langle \begin{bmatrix} 12 \\ 4 \\ 9 \\ 6 \end{bmatrix} \right\rangle$. Como PageRank solo debería reflejar la importancia relativa de los nodos, y como los autovectores son múltiplos por un escalar entre sí, podemos elegir cualquiera de ellos como nuestro vector PageRank. Elegimos el vector v^* como el único autovector con la suma de sus entradas igual a 1 (en ocasiones se referirá a él como el autovector probabilístico correspondiente al autovalor 1). Entonces, el vector

$$v^* = \frac{1}{31} \begin{bmatrix} 12 \\ 4 \\ 9 \\ 6 \end{bmatrix} \sim \begin{bmatrix} 0{,}38 \\ 0{,}12 \\ 0{,}29 \\ 0{,}19 \end{bmatrix}$$

es nuestro vector PageRank.

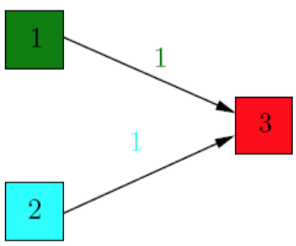
3. Enfoque Probabilístico

Como la importancia de una página Web es medida a través de su popularidad (la cantidad de vínculos hacia ella que tienen otras páginas), podemos ver la importancia de la página i como la probabilidad de que una persona aleatoria que abre un navegador de Internet y comienza a visitar vínculos, visite la página i. Podemos interpretar los pesos de cada arista de manera probabilística: la persona actualmente visitando la página 2 tiene probabilidad $\frac{1}{2}$ de ir a la página 3 y probabilidad $\frac{1}{2}$ de ir a la página 4. Podemos modelar este proceso como un camino aleatorio sobre grafos. Cada página tiene probabilidad $\frac{1}{4}$ de ser elegida como página inicial. Entonces, la probabilidad inicial es dada por el vector $x = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right)^T$. La probabilidad de que la página i sea visitada luego de k pasos es igual a $A^k x$. En este caso, la sucesión $Ax, A^2x, \ldots, A^kx$ converge a un único vector probabilístico v^*. En este contexto, v^* es denominado *distribución estacionaria* y será nuestro vector PageRank. La i-ésima entrada de v^* es la probabilidad de que en un momento dado la persona visite la página i.

El vector PageRank v^* calculado indica que la página 1 es la más relevante. Esto puede parecer inesperado, ya que dos vínculos apuntan a la página 1 mientras que tres vínculos apuntan a la página 3. Sin embargo, el nodo 3 tiene solo una arista saliente que apunta al nodo 1, transfiriendo *toda* su importancia a ese nodo. Es importante notar, además, que el rango de cada página no es solo la suma ponderada de las aristas entrantes de cada nodo. Intuitivamente, en el primer paso, un nodo recibe un voto de importancia de sus vecinos directos, en el segundo paso de los vecinos de sus vecinos, y así sucesivamente.

4. Nodos sin Aristas Salientes (Nodos Colgantes)

Consideremos el siguiente ejemplo:



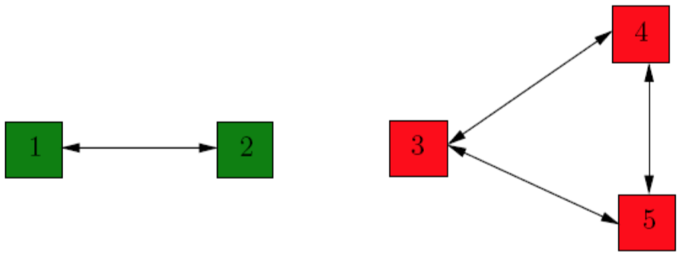
En este caso, la matriz de transición es $A = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 0 \end{bmatrix}$, y realizando los cálculos llegamos al vector

PageRank $v^* = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$. Es decir, el rango de todas las páginas es 0. Claramente, esto no tiene sentido, ya que intuitivamente la página 3 debería tener mayor importancia.

Una solución fácil para este problema es reemplazar la columna correspondiente al nodo colgante 3 con una columna cuyas entradas son $\frac{1}{3}$. De esta manera, la importancia del nodo 3 estaría igualmente distribuida entre los otros nodos del grafo, en vez de perderse.

5. Componentes Desconectados

Veamos el siguiente ejemplo:



Si una persona que navega por las páginas comienza en el primer componente conectado, no tiene forma de visitar la página 5, pues los nodos 1 y 2 no tienen vínculos a ninguna página conectada con la página 5. En este caso, la matriz de transición es

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}.$$

Nótese que $v = (1, 1, 0, 0, 0)^T$ y $u = (0, 0, 1, 1, 1)^T$ ambos son autovectores asociados al autovalor 1, y son linealmente independientes. Entonces, clasificar páginas de acuerdo al primer componente conectado relativamente al segundo componente conectado es ambiguo.

Sea n la cantidad de nodos del grafo. Para resolver estos problemas, se fija una constante positiva p entre 0 y 1, llamada el **factor *damping*** (un valor típico es $p = 0{,}15$). Luego, se define la matriz

PageRank M (también llamada la matriz Google) del grafo de la siguiente manera:

$$M = (1-p) \cdot A + p \cdot B, \text{ con } B = \frac{1}{n} \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix}.$$

Definición 55. Una matriz M de tamaño $n \times n$ es **estocástica por columnas** si para todo $i = 1, \dots, n$, $\sum_{j=1}^{n} M_j^i = 1$.

Problema 1. Probar que M es una matriz estocástica por columnas con entradas positivas.

La matriz M representa el modelo de la persona que navega a través de las páginas de la siguiente manera: la mayoría del tiempo, la persona seguirá los vínculos desde páginas; desde la página i, sigue los links salientes y va hacia los vecinos de la página i. Con un porcentaje menor de probabilidad, la persona dejará la página actual y elegirá arbitrariamente una página diferente de la Web, e irá allí. Esta probabilidad es reflejada por p, y como puede ir a cualquier página, cada página tiene probabilidad $\frac{1}{n}$ de ser elegida.

Problema 2. Realizar nuevamente los cálculos para el Page Rank reemplazando la matriz de transición A por la matriz M para los grafos explicativos de nodos colgantes y de componentes desconectados. Continúan ocurriendo en los grafos los problemas mencionados al utilizar M?

Intuitivamente, la matriz M conecta los grafos y se deshace de los nodos colgantes. Un nodo sin aristas salientes ahora tiene probabilidad $\frac{p}{n}$ de ir a cualquier otro nodo.

Teorema 34. ***(Teorema de Perron-Frobenius)*** *Si M es una matriz estocástica por columnas con entradas positivas, entonces:*

1. 1 *es un autovalor de multiplicidad 1 de M.*

2. 1 *es el mayor autovalor: todos los otros autovalores son de valor absoluto menor a* 1.

3. *Los autovectores correspondientes al autovalor* 1 *tienen solo entradas positivas o solo entradas negativas. En particular, para el autovalor* 1 *existe un único autovector donde la suma de sus entradas es igual a* 1.

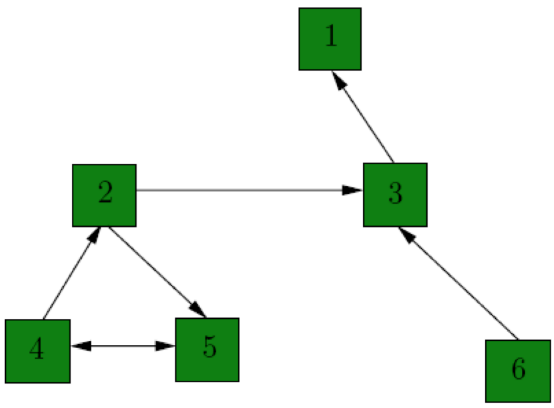
Teorema 35. ***(Convergencia del Método de Potencia)*** *Sea M una matriz estocástica por columnas de tamaño $n \times n$. Sea v^* el vector probabilístico correspondiente al autovalor* 1. *Sea z el vector columna con entradas iguales a $\frac{1}{n}$. Entonces, la sucesión $z, Mz, \dots, M^k z$ converge al vector v^*.*

A luz de lo visto anteriormente, se puede concluir que el vector PageRank de un grafo de páginas Web con matriz de transición A y factor *damping* p, en el único autovector probabilístico de la matriz M, correspondiente al autovalor 1.

Desde el punto de vista matemático, calcular el autovector probabilístico es, en teoría, simple. Sin embargo, es ineficiente realizar estos cálculos para tamaños grandes. Una forma alternativa de calcular este vector es mediante el Método de Potencia. Computacionalmente hablando, es considerablemente más sencillo multiplicar los vectores $x, Mx, \ldots, M^n x$ hasta la convergencia que calcular los autovectores de M para un tamaño en el orden de, por ejemplo, 10^{10}. De hecho, es posible calcular solo los primeros términos de la sucesión para obtener una buena aproximación.

Para una matriz aleatoria, se conoce que el Método de Potencia converge lentamente. Sin embargo, lo que hace que funcione en este caso es el hecho de que el grafo de la Web es poco denso: un nodo i tiene un número pequeño de links salientes (a lo sumo unos cientos, en contraposición con las 3^{10} páginas de la Web). Entonces, la matriz de transición A tiene muchas entradas nulas.

Problema 3. Calcular el vector Page Rank del siguiente grafo, con la constante *damping* p tomando los valores $p = 0$, $p = 0{,}15$, $p = 0{,}5$ y $p = 1$.



Problema 4. Calcular el vector Page Rank del siguiente árbol dirigido, considerando $p = 0{,}15$. Interpretar los resultados en términos de la relación entre el número de vínculos entrantes y el rango de cada nodo.