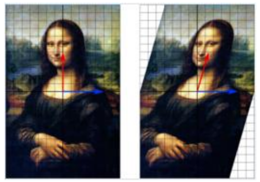
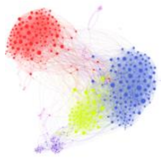
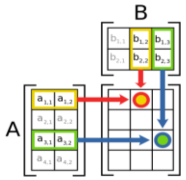
Tipos de matrices II

"Vectores, matrices y tensores: Una perspectiva desde ciencia e ingeniería de datos"

Reading Group organizado por el "Grupo de Investigación de Alto Rendimiento en Ciencia de Datos y Procesamiento de Señal de la URJC"







Esta presentación utiliza material cuya autoría corresponde a: Marques, Buciulea, Strang, Wikipedia; así como contribuciones menores de distintas autorías.

Índice de la clase

- Matrices persimétricas, centrosimétricas, circulantes por bloque, Hankel
- Matrices idempotentes, tripotentes, nilpotentes, unipotentes
- Matrices de Hessenberg y tridiagonales
- Matriz de Hadamard
- Matrices de Vandermonde y de Fourier
- Matrices positivas, no-negativas e irreducibles
- Matrices indescomponibles y de Perron
- Matrices estables, estables positivas, P, Z, M y ML
- Matrices de Krylov y Gram
- Matrices estocásticas y doblemente estocásticas
- Matriz de Carleman

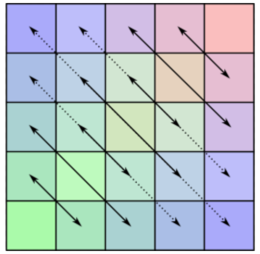
Tipos de matrices en función de sus simetrías

- Simetría ➔ redundancia (dada parte de la matriz, podemos reconstruir la otra parte)
  - En la clase anterior vimos matrices: simétricas, antisimétricas, Toeplitz y circulante

Persmétrica (*Persymmetric*):

Simétrica respecto a la diagonal secundaria

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{14} \\ a_{31} & a_{32} & a_{33} & a_{23} & a_{13} \\ a_{41} & a_{42} & a_{32} & a_{22} & a_{12} \\ a_{51} & a_{41} & a_{31} & a_{21} & a_{11} \end{bmatrix}$$



$$a_{ij} = a_{(n-j+1),(n-i+1)}, \ \forall i,j \qquad \mathbb{R}^{n\times n} \Rightarrow (n+1)n/2$$

- Algunas propiedades

$$\mathbf{A} = \mathbf{E}\mathbf{A}^\top\mathbf{E} \qquad \mathbf{E}\mathbf{E} = \mathbf{I} \qquad \mathbf{E} = \begin{bmatrix} 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & \dots & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & \dots & 0 & 0 \end{bmatrix}$$

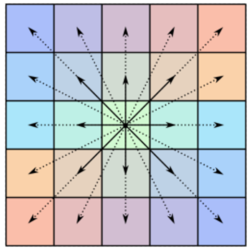
Si $\mathbf{A}^{-1}$ existe, es persimétrica

Si $\mathbf{T} \in \mathbb{R}^{n\times n}$ es Toeplitz $\rightarrow$ $\mathbf{T}$ es persimétrica

Centrosimétrica (*Cross-symmetric*):

Simétrica respecto al centro de la matriz

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{31} & a_{32} & a_{33} & a_{32} & a_{31} \\ a_{25} & a_{24} & a_{23} & a_{22} & a_{21} \\ a_{15} & a_{14} & a_{13} & a_{12} & a_{11} \end{bmatrix}$$



$$a_{ij} = a_{(n+1-i),(n+1-j)}, \ \forall i,j \qquad \mathbb{R}^{n\times n} \Rightarrow (n^2 + n\%2)/2$$

- Algunas propiedades

$$\mathbf{A} = \mathbf{E}\mathbf{A}\mathbf{E}$$

$$\mathbf{A} + \mathbf{B}, \ c\mathbf{A}, \ \mathbf{A}\mathbf{B}$$

Son centrosimétricas

Matriz de autocorr. de AR(1)

$$\mathbf{B} = \begin{bmatrix} 1 & p & p^2 & \cdot & p^{n-1} \\ p & 1 & p & \cdot & p^{n-2} \\ p^2 & p & 1 & \cdot & p_{n-3} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ p^{n-1} & p^{n-2} & p^{n-3} & \cdot & 1 \end{bmatrix}$$

Toeplitz simétrica $\rightarrow a_{ij} = a_{(i+k),(j+k)} = a_{ji}, \ \forall i,j$

Tipos de matrices en función de sus simetrías

- Simetría ➔ redundancia (dada parte de la matriz, podemos reconstruir la otra parte)

- Circulante por bloques (*Block circulant*):

$\mathbf{A} \in \mathbb{R}^{nk \times nk} \Rightarrow nk^2$ ➔ matriz circulante por bloques

$$\mathbf{A} = \begin{bmatrix} \mathbf{C}_0 & \mathbf{C}_1 & \mathbf{C}_2 & \dots & \mathbf{C}_{n-1} \\ \mathbf{C}_{n-1} & \mathbf{C}_0 & \mathbf{C}_1 & \dots & \mathbf{C}_{n-2} \\ \mathbf{C}_{n-2} & \mathbf{C}_{n-1} & \mathbf{C}_0 & \dots & \mathbf{C}_{n-3} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \mathbf{C}_1 & \mathbf{C}_2 & \mathbf{C}_3 & \dots & \mathbf{C}_0 \end{bmatrix}$$

$\mathbf{C} \in \mathbb{R}^{k \times k}$ ➔ matriz circulante

$$\mathbf{C} = \begin{bmatrix} c_0 & c_1 & c_2 & \dots & c_{k-1} \\ c_{k-1} & c_0 & c_1 & \dots & c_{k-2} \\ c_{k-2} & c_{k-1} & c_0 & \dots & c_{k-3} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ c_1 & c_2 & c_3 & \dots & c_0 \end{bmatrix}$$

$$\mathbf{A} = \mathbf{I}_n \otimes \mathbf{C}_0 + \mathbf{P}_0 \otimes \mathbf{C}_1 + \mathbf{P}_0^2 \otimes \mathbf{C}_2 + \dots + \mathbf{P}_0^{n-1} \otimes \mathbf{C}_{n-1} \in \mathcal{B}_{n,k}$$

$$\text{Si } \mathbf{A}, \mathbf{B} \in \mathcal{B}_{n,k} \rightarrow \mathbf{A}^\top,\ \alpha_1\mathbf{A} + \alpha_2\mathbf{B},\ \mathbf{AB},\ \mathbf{A}^{-1} \in \mathcal{B}_{n,k}$$

¡Todo esto es cierto también para las circulantes!

- Hankel:

Diagonales paralelas numéricamente de dcha. a izda.

$\mathbf{A}_H \in \mathbb{R}^{L \times K}$

$$\mathbf{A}_H = \begin{bmatrix} a_0 & a_1 & a_2 & \dots & a_{K-1} \\ a_1 & a_2 & a_3 & \dots & a_K \\ a_2 & a_3 & a_4 & \dots & a_{K+1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{L-1} & a_L & a_{L+1} & \dots & a_{K+L-2} \end{bmatrix}$$

$$a_{ij} = a_{(i-1),(j+1)},\ \forall i,j \in \mathcal{A}_{\mathcal{H}}$$

➢ Algunas propiedades

Si $\mathbf{T}$ es Toeplitz $\rightarrow \mathbf{ET} = \mathbf{A}_H,\ \mathbf{EA}_H = \mathbf{T}$

Si $\mathbf{T}$ es real y simétrica $\rightarrow \lambda_i(\mathbf{A}_H) = \lambda_i(\mathbf{T})$

La matriz de Hilbert es una matriz de Hankel

Matrices idempotentes, tripotentes, nilpotentes y unipotentes

Idempotente:

$$\mathbf{A}\cdot\mathbf{A}=\mathbf{A}^2=\mathbf{A} \quad \mathbf{R}^{-1}\mathbf{A}\mathbf{R}=\begin{bmatrix}\mathbf{I}_r & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}$$

$$\det(\mathbf{A}^2)=\det(\mathbf{A})=0 \text{ } o \text{ } 1$$

$$\operatorname{rank}(\mathbf{A})=\operatorname{trace}(\mathbf{A})=r$$

$$\lambda_i(\mathbf{A})\in\{0,1\} \qquad \mathbf{A}=\begin{bmatrix}2 & -2 & -4\\ -1 & 3 & 4\\ 1 & -2 & -3\end{bmatrix}$$

Tripotente:

$$\mathbf{A}\cdot\mathbf{A}\cdot\mathbf{A}=\mathbf{A}^3=\mathbf{A}$$

$$\mathbf{A}^2 \text{ es idempotente}$$

$$\operatorname{rank}(\mathbf{A})=\operatorname{trace}(\mathbf{A}^2)=r$$

$$\lambda_i(\mathbf{A})\in\{-1,0,1\}$$

Unipotente:

$$\mathbf{A}^2=\mathbf{I}_n \qquad \mathbf{A}=\begin{bmatrix}\mathbf{I} & \mathbf{B}\\ \mathbf{0} & -\mathbf{I}\end{bmatrix}$$

$$\mathbf{A}\mathbf{A}=\begin{bmatrix}\mathbf{I} & \mathbf{B}\\ \mathbf{0} & -\mathbf{I}\end{bmatrix}\begin{bmatrix}\mathbf{I} & \mathbf{B}\\ \mathbf{0} & -\mathbf{I}\end{bmatrix}=\begin{bmatrix}\mathbf{II} & \mathbf{B}-\mathbf{B}\\ \mathbf{0} & (-\mathbf{I})(-\mathbf{I})\end{bmatrix}$$

$$\mathbf{A}=\begin{bmatrix}1 & 0 & 3 & 4\\ 0 & 1 & 2 & 4\\ 0 & 0 & -1 & 0\\ 0 & 0 & 0 & -1\end{bmatrix}$$

Nilpotente:

$$\mathbf{A}^k=\mathbf{0} \text{ con } k=1,2,3,\ldots$$

$$\text{Índice de nilpotencia: } \min(k) \text{ tal que } \mathbf{A}^k=\mathbf{0}$$

$$\lambda_i(\mathbf{A})\in\{0\} \qquad \mathbf{A}=\begin{bmatrix}0 & 3 & 2 & 1\\ 0 & 0 & 2 & 2\\ 0 & 0 & 0 & 3\\ 0 & 0 & 0 & 0\end{bmatrix}$$

Jordan Block:

Jordan block de tamaño k y autovalor λ : $\mathbf{J}_{\lambda,k}$

$$\mathbf{J}_{i,2}=\begin{bmatrix}i & 1\\ 0 & i\end{bmatrix} \qquad \mathbf{J}_{7,3}=\begin{bmatrix}7 & 1 & 0\\ 0 & 7 & 1\\ 0 & 0 & 7\end{bmatrix}$$

Jordan Matrix:

Una matriz block diagonal con
Jordan block es una Jordan matrix

$$\mathbf{J}=\mathbf{J}_{0,3}\otimes\mathbf{J}_{i,2}\otimes\mathbf{J}_{i,2}\otimes\mathbf{J}_{7,3}$$

$$J=\left[\begin{array}{ccc|cc|cc|ccc}0&1&0&0&0&0&0&0&0&0\\0&0&1&0&0&0&0&0&0&0\\0&0&0&0&0&0&0&0&0&0\\\hline 0&0&0&i&1&0&0&0&0&0\\0&0&0&0&i&0&0&0&0&0\\\hline 0&0&0&0&0&i&1&0&0&0\\0&0&0&0&0&0&i&0&0&0\\\hline 0&0&0&0&0&0&0&7&1&0\\0&0&0&0&0&0&0&0&7&1\\0&0&0&0&0&0&0&0&0&7\end{array}\right]$$

Factorización de **J**

$$\mathbf{J}=\mathbf{D}+\mathbf{N}$$

$$\mathbf{N}^k=\mathbf{0}, \text{ nilplotente de orden } k$$

$$\mathbf{D}=\operatorname{diag}(\operatorname{diag}(\mathbf{J}))$$

Útil para descomposición de matrices

$$\mathbf{A}=\mathbf{W}^{-1}\mathbf{J}\mathbf{W},\ \mathbf{W}^{-1}\mathbf{W}=\mathbf{I}$$

Matrices de Hessenberg y tridiagonales

- Una matriz $\mathbf{H} \in \mathbb{R}^{n\times n}$ es *lower Hessenberg* si los elementos por encima de la subdiagonal son cero. Su traspuesta es una *upper Hessenberg*

$$\mathbf{H}_{low} = \begin{bmatrix} h_{11} & h_{12} & 0 & 0 & 0 \\ h_{21} & h_{22} & h_{23} & 0 & 0 \\ h_{31} & h_{32} & h_{33} & h_{34} & 0 \\ h_{41} & h_{42} & h_{43} & h_{44} & h_{45} \\ h_{51} & h_{52} & h_{53} & h_{54} & h_{55} \end{bmatrix}$$

$$(h_{ij} = 0\ \forall\ i,j \text{ tal que } i > j+1)$$

$$\mathbf{H}_{up} = \mathbf{H}_{low}^{\top} = \begin{bmatrix} h_{11} & h_{12} & h_{13} & h_{14} & h_{15} \\ h_{21} & h_{22} & h_{23} & h_{24} & h_{25} \\ 0 & h_{32} & h_{33} & h_{34} & h_{35} \\ 0 & 0 & h_{43} & h_{44} & h_{45} \\ 0 & 0 & 0 & h_{54} & h_{55} \end{bmatrix}$$

$$(h_{ij} = 0\ \forall\ i,j \text{ tal que } j > i+1)$$

- Útil para reducir el coste computacional de los algoritmos
  - Segunda mejor opción después de las matrices triangulares
  - Método para transformar una matriz cualquiera en Hessenberg
    - Householder's transformation
  - Son utilizadas como entrada para distintos métodos eficientes
    - *Schur decomposition* $\mathbf{A} = \mathbf{QUQ}^{-1},\ \mathbf{Q}^{-1} = \mathbf{Q}^{\mathrm{H}}$
    - *QR factorization* $\mathbf{A} = \mathbf{QU},\ \mathbf{Q}^{-1} = \mathbf{Q}^{\top}$

- Una matriz $\mathbf{A} \in \mathbb{R}^{n\times n}$ es tridiagonal si los elementos de fuera de las tres diagonales centrales son distintos cero $(a_{ij} \neq 0 \text{ si } |i-j| \leq 1 \text{ y } a_{ij} = 0 \text{ si } |i-j| > 1)$

$$\mathbf{A} = \begin{bmatrix} a_1 & b_1 & 0 & 0 & 0 \\ c_1 & a_2 & b_2 & 0 & 0 \\ 0 & c_2 & \ddots & \ddots & 0 \\ 0 & 0 & \ddots & \ddots & b_{n-1} \\ 0 & 0 & 0 & c_{n-1} & a_n \end{bmatrix}$$

$$\det(\mathbf{A}) = f_n = a_n f_{n-1} - c_{n-1} b_{n-1} f_{n-2}$$
$$f_1 = |a_1|,\ f_0 = 1,\ f_{-1} = 0$$

- Muy útil para
  - Descomposición y factorización de matrices
  - Análisis de estabilidad → Schwartz matrix **C**
    - **C** es estable positiva si $c_1 c_2 \cdots c_n > 0$

$$\mathbf{C} = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 & 0 & 0 \\ -c_n & 0 & 1 & \dots & 0 & 0 & 0 \\ 0 & -c_{n-1} & 0 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -c_3 & 0 & 1 \\ 0 & 0 & 0 & \dots & 0 & -c_2 & c_1 \end{bmatrix}$$

Matrices de Hadamard

- Una matriz de Hadamard **H**, todos sus elementos son ± 1 tal que:

Estructura:

$$\begin{bmatrix} \mathbf{H} & \mathbf{H} \\ \mathbf{H} & -\mathbf{H} \end{bmatrix}$$

Ejemplo:

$$\mathbf{H}_1 = [1]$$

$$\mathbf{H}_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \quad \mathbf{H}_4 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}$$

Fórmula general:

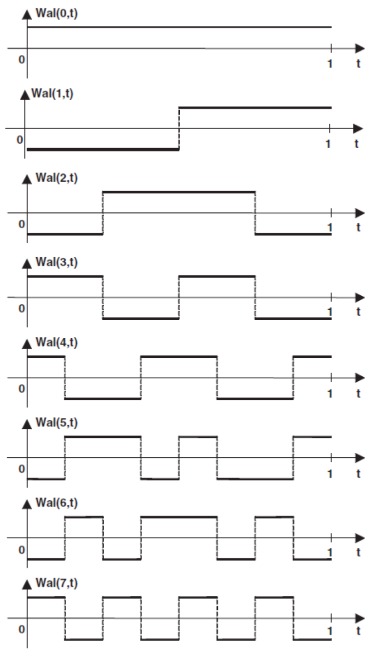
$$\mathbf{H}_{2^k} = \begin{bmatrix} \mathbf{H}_{2^{k-1}} & \mathbf{H}_{2^{k-1}} \\ \mathbf{H}_{2^{k-1}} & -\mathbf{H}_{2^{k-1}} \end{bmatrix} = \mathbf{H}_2 \otimes \mathbf{H}_{2^{k-1}}$$

- Algunas propiedades de una matriz de Hadamard $\mathbf{H}$ de orden *n*
  - La matriz traspuesta muy relacionada con su inversa ➔ $\mathbf{H}\mathbf{H}^\top = n\mathbf{I}_n$
  - Dado que $\det(\mathbf{H}\mathbf{H}^\top) = n^n$, tenemos que $\det(\mathbf{H}) = n^{n/2}$
  - El producto de Kronecker de dos matrices de Hadamard es otra matriz de Hadamard

$$\mathbf{H} = \mathbf{H}^{(A)} \otimes \mathbf{H}^{(B)} \in \mathbb{R}^{n_1 n_2 \times n_1 n_2}, \;\; \text{con } \mathbf{H}^{(A)} \in \mathbb{R}^{n_1 \times n_1} \text{ y } \mathbf{H}^{(B)} \in \mathbb{R}^{n_2 \times n_2}$$

- Aplicaciones
  - Generación de sets de funciones Walash ortogonales.
  - Generación de códigos ortogonales para la transmisión CDMA

$$H_8 = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & - & 1 & - & 1 & - & 1 & - \\ 1 & 1 & - & - & 1 & 1 & - & - \\ 1 & - & - & 1 & 1 & - & - & 1 \\ 1 & 1 & 1 & 1 & - & - & - & - \\ 1 & - & 1 & - & - & 1 & - & 1 \\ 1 & 1 & - & - & - & - & 1 & 1 \\ 1 & - & - & 1 & - & 1 & 1 & - \end{pmatrix} \begin{matrix} Wal(0,t) \\ Wal(7,t) \\ Wal(3,t) \\ Wal(4,t) \\ Wal(1,t) \\ Wal(6,t) \\ Wal(2,t) \\ Wal(5,t) \end{matrix}$$



Matriz de Vandermonde

- Una matriz de Vandermonde contiene los términos de una progresión geométrica en cada columna

$$\mathbf{V}(a_1, a_2, ..., a_n)$$

$$\mathbf{V} = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ a_1 & a_2 & a_3 & \dots & a_n \\ a_1^2 & a_2^2 & a_3^2 & \dots & a_n^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_1^{n-1} & a_2^{n-1} & a_3^{n-1} & \dots & a_n^{n-1} \end{bmatrix}$$

$$\det(\mathbf{V}) = \prod_{1 \le i \le j \le n} (a_j - a_i)$$

$$\mathrm{rank}(\mathbf{V}) = \mathrm{r}$$

$$\mathrm{r} \to \text{número de } a_i \text{ distintos}$$

- Matrices de Vandermonde $\mathbf{V}(a_1, a_2, ..., a_n), \ a_i \in \mathbb{R}$
  - Mal condicionadas $\kappa(\mathbf{V}) = \|\mathbf{V}\|_2 \|\mathbf{V}^{-1}\|_2 \gg n$
  - Los autovalores aumentan exponencialmente con *n*
  - **V** se vuelve singular a medida que $n \to \infty$

```
>> V = vander([1,2,3,4,5,6,7])

V =

        1        1        1        1        1        1        1
       64       32       16        8        4        2        1
      729      243       81       27        9        3        1
     4096     1024      256       64       16        4        1
    15625     3125      625      125       25        5        1
    46656     7776     1296      216       36        6        1
   117649    16807     2401      343       49        7        1
```

```
>> eig(V)

ans =

  556.4501
 -457.4998
  141.9610
  -35.7140
    5.1654
   -0.3727
    0.0100
```

La matriz **V** mapea el vector de coeficientes de un polinomio a un vector de valores del polinomio

Útil en interpolación de polinomios (invirtiendo **V**)

Ejemplo: $p(x) = a_3x^3 + a_2x^2 + a_1x + a_0$

Expresar *p(x)* en forma matricial

$$\begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix}^\top \begin{bmatrix} 1 & 1 & 1 & 1 \\ x_0 & x_1 & x_2 & x_3 \\ x_0^2 & x_1^2 & x_2^2 & x_3^2 \\ x_0^3 & x_1^3 & x_2^3 & x_3^3 \end{bmatrix} = \begin{bmatrix} y_0 \\ y_1 \\ y_2 \\ y_3 \end{bmatrix}^\top$$

Encontrar un polinomio *p(x)*, que pase por los puntos (1,-6), (2,2), (4,12) y (3,-10)

$$\begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix}^\top \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 4 & 3 \\ 1^2 & 2^2 & 4^2 & 3^2 \\ 1^3 & 2^3 & 4^3 & 3^3 \end{bmatrix} = \begin{bmatrix} -6 \\ 2 \\ 12 \\ -10 \end{bmatrix}^\top$$

$$p(x) = 9x^3 + -64x^2 + 137x - 88$$

Matriz de Fourier I

Matriz de Fourier

- Una matriz de Vandermonde específica

$$w = e^{2\pi i/n}, \ \overline{w} = e^{-2\pi i/n}$$

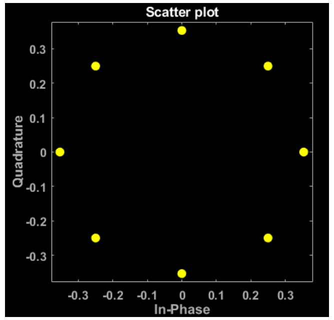
$$\mathbf{F} = n^{-1/n}\mathbf{V}(1, \overline{w}, \overline{w}^2, ..., \overline{w}^{n-1})$$

$$\mathbf{F}^* = \overline{\mathbf{F}} = \frac{1}{\sqrt{n}}\begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & w & w^2 & \dots & w^{n-1} \\ 1 & w^2 & w^4 & \dots & w^{2n-2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & w^{n-1} & w^{2n-2} & \dots & w^{(n-1)(n-1)} \end{bmatrix}$$

DFT de ocho puntos

- Matriz de DFT (**F**) y representación de los valores de **F**

$$\mathbf{F} = \frac{1}{\sqrt{8}}\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & \omega & -i & -i\omega & -1 & -\omega & i & i\omega \\ 1 & -i & -1 & i & 1 & -i & -1 & i \\ 1 & -i\omega & i & \omega & -1 & i\omega & -i & -\omega \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & -\omega & -i & i\omega & -1 & \omega & i & -i\omega \\ 1 & i & -1 & -i & 1 & i & -1 & -i \\ 1 & i\omega & i & -\omega & -1 & -i\omega & -i & \omega \end{bmatrix}$$

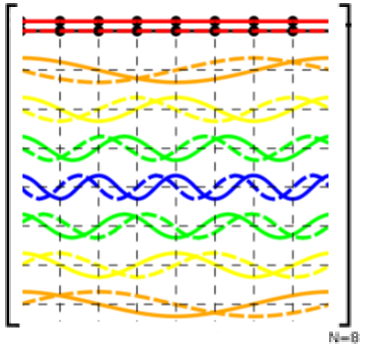


$$w = \frac{1}{\sqrt{2}} - \frac{i}{\sqrt{2}}$$

La DFT como multiplicación de matrices

- **F** como muestras de exponenciales complejas

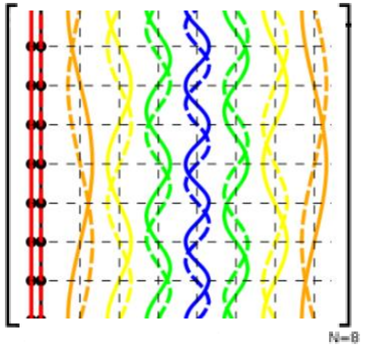
$$\begin{bmatrix} X[0] \\ X[1] \\ X[2] \\ X[3] \\ X[4] \\ X[5] \\ X[6] \\ X[7] \end{bmatrix} =$$



$$\begin{bmatrix} x[0] \\ x[1] \\ x[2] \\ x[3] \\ x[4] \\ x[5] \\ x[6] \\ x[7] \end{bmatrix}$$

- $\mathbf{F}^H$ como muestras de coeficientes

$$\begin{bmatrix} x[0] \\ x[1] \\ x[2] \\ x[3] \\ x[4] \\ x[5] \\ x[6] \\ x[7] \end{bmatrix} =$$



$$\begin{bmatrix} X[0] \\ X[1] \\ X[2] \\ X[3] \\ X[4] \\ X[5] \\ X[6] \\ X[7] \end{bmatrix}$$

Matriz de Fourier II

- Propiedades de la matriz de Fourier
  - Matriz “especial”, con una estructura muy potente

$$\mathbf{F}\mathbf{F}^{\mathsf{H}} = \mathbf{I}_n \qquad \mathbf{F}^{-1} = \bar{\mathbf{F}} \qquad \mathbf{F}^{-1} = \mathbf{F}^{\circ -1}$$

  - Matriz simétrica

$$\mathbf{F} = \mathbf{F}^{\top} \text{ y } \mathbf{F}^{\mathsf{H}} = (\mathbf{F}^{\mathsf{H}})^{\top}$$

  - Propiedades de las potencias de **F**

$$\mathbf{F}^2 = \left(\mathbf{F}^{\mathsf{H}}\right)^2 = \mathbf{P} \qquad \mathbf{F}^4 = \left(\mathbf{F}^{\mathsf{H}}\right)^4 = \mathbf{I}_n$$

  - Transformada discreta de Fourier (DFT)

$\mathbf{y} = \mathbf{F}\mathbf{z}$ ➔ DFT de los elementos de **z**

$\mathbf{z} = \mathbf{F}^{-1}\mathbf{y} = \mathbf{F}^{\mathsf{H}}\mathbf{y}$ ➔ IDFT de los elementos de **y**

  - Autovalores de **F**

$$\lambda_k(\mathbf{F}) \in \{-1, 1, -i, i\}$$

  - Matrix **F** “idealmente” condicionada

$$\kappa(\mathbf{F}) = \|\mathbf{F}\|_2 \|\mathbf{F}^{-1}\|_2 = 1$$

- Diagonalización de matrices circulantes mediante DFT: $\mathbf{c} \in \mathbb{C}^n, \; \mathbf{C}, \; \mathbf{F} \in \mathbb{C}^{n\times n}$

$$\boxed{\mathbf{F}\mathbf{C}\mathbf{F}^{\mathsf{T}} = \mathrm{diag}(\sqrt{n}\mathbf{F}\mathbf{c})}$$

  - Ejemplo

$$\mathbf{c} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} \quad \mathbf{C} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 4 & 1 & 2 & 3 \\ 3 & 4 & 1 & 2 \\ 2 & 3 & 4 & 1 \end{bmatrix} \quad \mathbf{F} = \frac{1}{\sqrt{4}} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -i & -1 & i \\ 1 & -1 & 1 & -1 \\ 1 & i & -1 & -i \end{bmatrix} \quad \mathbf{F}\mathbf{C}\mathbf{F}^{\mathsf{T}} = \begin{bmatrix} 10 & 0 & 0 & 0 \\ 0 & -2-2i & 0 & 0 \\ 0 & 0 & -2 & 0 \\ 0 & 0 & 0 & -2+2i \end{bmatrix} \quad \sqrt{4}\mathbf{F}\mathbf{c} = \begin{bmatrix} 10 \\ -2-2i \\ -2 \\ -2+2i \end{bmatrix}$$

Matrices positivas, no-negativas e irreducibles

- Una matriz $\mathbf{A} \in \mathbb{R}^{n\times n}$ es positiva si todos sus valores son mayores que cero $\mathbf{A} > \mathbf{0}$
  - Todos sus menores son positivos (de determinante de cualquier submatriz de **A**)
  - Todas sus potencias son mayores que cero $\mathbf{A}^k > \mathbf{0}$ para $k = 1, 2, 3, \ldots$

- Una matriz $\mathbf{A} \in \mathbb{R}^{n\times n}$ es no-negativa si todos sus valores son iguales o mayores a cero $\mathbf{A} \geq \mathbf{0}$

- Una matriz $\mathbf{A} \geq \mathbf{0} \in \mathbb{R}^{n\times n}$ es reducible si existe una matriz de permutación **P** tal que

$$\mathbf{A} = \mathbf{P}\mathbf{A}\mathbf{P}^\top = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{0} \\ \mathbf{A}_{12} & \mathbf{A}_{22} \end{bmatrix}$$

$$\mathbf{A}_{11} \in \mathbb{R}^{n1\times n1}, \ \mathbf{A}_{22} \in \mathbb{R}^{n2\times n2}$$

- **A** es irreducible si

$$(\mathbf{I}_n + \mathbf{A})^{n-1} > \mathbf{0}$$

- Si **A** es irreducible
  - Tiene un autovalor real y positivo ρ, que se corresponde con el radio espectral de **A**
  - Los autovectores asociados al autovalor ρ son estrictamente positivos.
  - Todo par de índices de **A** se pueden comunicar.
- **A** es reducible si
  - Tiene al menos una fila (columna) de ceros
  - $\mathbf{A}^k$ es reducible para $k = [1, 2, \ldots]$

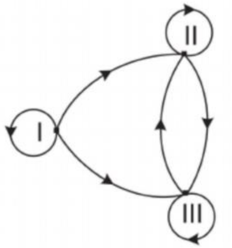
Radio espectral de **A**

$$\rho = \max\{|\lambda_1|, \ldots, |\lambda_n|\}$$
$$\lambda_i \rightarrow \text{autovalores de } \mathbf{A}$$

Matrices indescomponibles y de Perron

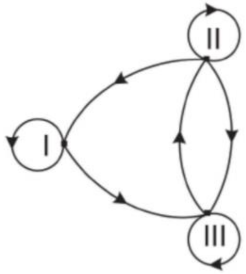
- Ejemplo de matrices irreducibles en grafos
  - **B** es reducible y G(**B**) no está fuertemente conectado

$$B = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 3 & 4 \\ 5 & 6 & 7 \end{pmatrix} \Leftrightarrow$$



  - **B** es irreducible y G(**B**) está fuertemente conectado

$$B = \begin{pmatrix} 1 & 2 & 0 \\ 0 & 3 & 4 \\ 5 & 6 & 7 \end{pmatrix} \Leftrightarrow$$



- Una matriz $\mathbf{A} \in \mathbb{R}^{n\times n}$ es parcialmente descomponible si existen dos matrices de permutación tales que

$$\mathbf{A} = \mathbf{P}_1\mathbf{A}\mathbf{P}_2 = \begin{bmatrix} \mathbf{B}_{11} & \mathbf{0} \\ \mathbf{B}_{12} & \mathbf{B}_{22} \end{bmatrix}$$

Si $\mathbf{A} \geq \mathbf{0}$ y es indescomponible $\to \mathbf{A}^{n-1} > \mathbf{0}$

Si $\mathbf{A},\ \mathbf{B} \geq \mathbf{0}$ e indescomponibles, $\mathbf{AB}$ también lo es

- Dada $\mathbf{A} \in \mathbb{R}^{n\times n}$, si existe un polinomio f tal que $f(\mathbf{A}) > \mathbf{0}$ ➔ **A** es Perron

  Si **A** > **0** es Perron y si $\mathbf{A} \geq \mathbf{0}$ e irreducible es Perron

  Si **A** es L-matrix irreducible entonces es una Perron

Veremos más adelante su definición

Si **A** es Perron, existe un autovalor ρ cuyos autovectores (izquierdo y derecho) asociados son estrictamente positivos

$$\mathbf{A}\mathbf{v} = \rho\mathbf{v},\ \ \mathbf{u}\mathbf{A} = \rho\mathbf{u} \quad \mathbf{v}, \mathbf{u} > \mathbf{0}$$

Matrices asociadas a operadores diferenciales

- En muchas aplicaciones tenemos un sistema de ecuaciones (variable multidimensional) y queremos analizar su comportamiento (sistema dinámico) o propiedades de una función definida sobre esas variables

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t)$$

Dinámica lineal y continua ➔ solución de sistema de ecuaciones diferenciales (Ce^{ct})

$$\mathbf{x}_{n+1} - \mathbf{x}_n = \mathbf{A}\mathbf{x}_n$$

Dinámica lineal y discreta ➔ solución de sistema de ecuaciones en diferencias (Ca^n)

$$x_{n+1} = f(x_n) \Rightarrow x_{n+1} = x_n + f'(x_n)$$

Dinámica no lineal ➔ aproximación lineal de primer orden

$$\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^n \qquad \mathbf{f}(\mathbf{x}) = [f_1(\mathbf{x}), ..., f_n(\mathbf{x})]$$

Podemos definir matrices donde sus entradas correspondan a derivadas de la función

$$\mathbf{J} = \begin{bmatrix} \frac{\partial f_1(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial f_1(\mathbf{x})}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial f_n(\mathbf{x})}{\partial x_n} \end{bmatrix}$$

Jacobiano, pero también pueden considerarse matrices con derivadas de mayor orden

- Si la función o la dinámica del sistema tiene ciertas propiedades, esas propiedades se tienen que reflejar también en las matrices **A**, **J**, …

- Otra forma de verlo es que las matrices **A**, **J**, … pueden utilizarse para estudiar las propiedades del sistema
  - Matrices con una estructura especial estarán asociadas a funciones/dinámicas especiales

- Ejemplos de interés: Matrices estables, matrices Z, matrices M…

Matrices estables

- Una matriz $\mathbf{A} \in \mathbb{C}^{n \times n}$ es estable si la parte real de todos sus autovalores es negativa

$$\Re(\lambda_i(\mathbf{A})) < 0 \;\; i = 1, 2, ..., n$$

En análisis de sistemas continuos, un sistema lineal e invariante es estable si los polos de su función de transferencia (Laplace) tienen parte real negativa

$$\Re(a_{ii}) < -\sum_{j=1; j \neq i}^{n} |a_{ij}|, \;\; i = 1, 2, ..., n$$

Si satisface esto, **A** es estable (no es un iff).
Intuición: "diagonal dominante" con signo negativo

- Intuición solución ecuaciones diferenciales lineales: $x(t)=C \cdot e^{at}$

En función de $\mathbf{A}$: $\exp(\mathbf{A}t) = \mathbf{I}_n + t\mathbf{A} + \frac{t^2}{2!}\mathbf{A}^2 + \cdots + \frac{t^r}{r!}\mathbf{A}^r + \ldots, -\infty < t < \infty$

En función de $\mathbf{V}$ y $\boldsymbol{\Lambda}$: $\exp(\mathbf{A}t) = \exp(\mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{-1}t) = \mathbf{V}\exp(\boldsymbol{\Lambda}t)\mathbf{V}^{-1} =$

$$= \mathbf{V}\left(\exp(\Re(\boldsymbol{\Lambda})t) + \exp(\Im(\boldsymbol{\Lambda})t)\right)\mathbf{V}^{-1}, -\infty < t < \infty$$

$\exp(\mathbf{A}t) \rightarrow \mathbf{0}$ cuando $t \rightarrow \infty$

- Ejemplo: Matriz de Hurwitz

$$p(z) = a_0 z^n + a_1 z^{n-1} + \cdots + a_{n-1} z + a_n$$

$$\mathbf{H} = \begin{bmatrix} a_1 & a_3 & a_5 & \ldots & 0 & 0 & 0 \\ a_0 & a_2 & a_4 & \ldots & 0 & 0 & 0 \\ 0 & a_1 & a_3 & \ldots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & a_{n-2} & a_n & 0 \\ 0 & 0 & 0 & \ldots & a_{n-3} & a_{n-1} & 0 \\ 0 & 0 & 0 & \ldots & a_{n-4} & a_{n-2} & a_n \end{bmatrix}$$

Cuadrada, real y cuyas entradas se corresponden a los coeficientes de un polinomio

Relacionada con la matriz de control de sistemas de control polinómicos

¿Es el sistema de control estable?

Si $a_0 > 0$ y los menores principales de **H** son positivos, *p(z)* es estable, es decir, sus raíces son estrictamente negativas

$$\Delta_1(p) = |a_1|$$

$$\Delta_2(p) = \begin{vmatrix} a_1 & a_3 \\ a_0 & a_2 \end{vmatrix}$$

$$\Delta_3(p) = \begin{vmatrix} a_1 & a_3 & a_5 \\ a_0 & a_2 & a_4 \\ 0 & a_1 & a_3 \end{vmatrix}$$

Matrices estables positivas

- $\mathbf{A} \in \mathbb{C}^{n\times n}$ era estable si la parte real de todos sus autovalores es negativa, ahora caso opuesto

- Una matriz $\mathbf{A} \in \mathbb{C}^{n\times n}$ es estable positiva si la parte real de todos sus autovalores es positiva

$$\Re(\lambda_i(\mathbf{A})) > 0 \;\; i = 1, 2, ..., n$$

  - Propiedades: $\mathbf{A}^{-1},\ \mathbf{A}^*,\ \mathbf{A}^\top$ $\quad \Re(\mathrm{trace}(\mathbf{A})) > \mathbf{0}$ $\quad \det(\mathbf{A}^k) > 0 \text{ para } k = 1, 2, 3, ...$

Son todas estables positivas

- Relacionadas con matrices definidas positivas $\mathbf{x}^H\mathbf{A}\mathbf{x} > 0$
  - Si la matriz **A** es hermítica (simétrica): definida positiva implica estable positiva
  - Si la matriz **A** no es hermítica: podemos considerar la ecuación

$$\mathbf{XA} + \mathbf{A}^*\mathbf{X} = \mathbf{C}$$

Si existe una matriz hermítica $\mathbf{C} \succ 0$ tal que al resolver con respecto a **X** obtenemos como solución una matriz hermítica $\mathbf{X} \succ 0$, entonces **A** es estable positiva

- Tanto las matrices definidas positivas como las matrices estables positivas pertenecen a una clase "más grande": las matrices P

Matrices P, Z, L y M

- $\mathbf{A} \in \mathbb{C}^{n\times n}$ es una P-matrix si todos sus $k \times k$ principales menores son positivos para $k = 1, 2, ..., n$
  - También son P-matrix $\mathbf{A}^{\mathsf{H}},\ \mathbf{AD},\ \mathbf{DA},\ \mathbf{A}+\mathbf{D},\ \mathbf{P}^{\mathsf{H}}\mathbf{AP}$ permutación $\mathbf{P}$ diagonal $\mathbf{D} > \mathbf{0}$
  - Otras propiedades $\Re(\lambda_i(\mathbf{A})) > 0 \quad \mathbf{x}^\top\mathbf{DAx} > 0$
  - Casos particulares: def. positivas; estables positivas; totalmente positivas (todos los menores positivos), M-matrix
- ¿Qué es una M-matrix? Una matriz que es: 1) una Z-matrix y 2) estable positiva ➔ ¿Qué es una Z-matrix?
- $\mathbf{A} \in \mathbb{C}^{n\times n}$ es una Z-matrix si $a_{ij} \leq 0$ fuera de la diagonal (si, además, $a_{ii} > 0$ ➔ **A** es una L-matrix)
- Definiciones equivalentes para que $\mathbf{A} \in \mathbb{C}^{n\times n}$ sea una M-matrix (y propiedades equivalentes)

$$\mathbf{A} = s\mathbf{I}_n - \mathbf{B} \text{ con } \mathbf{B} \geq \mathbf{0},\ s > \rho(\mathbf{B}) \qquad \mathbf{A}^{-1} \geq \mathbf{0} \qquad \mathbf{Ax} \geq \mathbf{0} \Rightarrow \mathbf{x} \geq \mathbf{0} \qquad \mathbf{A} = \mathbf{LU} \text{ con } l_{ii} > 0\ u_{ii} > 0$$

Las M-matrix aparecen en el análisis de métodos iterativos de sistemas de ecuaciones esparsos, en econometría, para el análisis de Laplacianos, en procesos de Markov, en estabilidad de Lyapunov

Matrices de Krylov y Gramm

- Matriz de Krylov $\mathbf{K} \in \mathbb{R}^{n\times r}, \ \mathbf{A} \in \mathbb{R}^{n\times n}, \ \mathbf{b} \in \mathbb{R}^{n}$

  Subespacio de Krylov:

  $$\mathcal{K}_r(\mathbf{A},\mathbf{b}) = \mathrm{span}\{\mathbf{b}, \mathbf{A}\mathbf{b}, \mathbf{A}^2\mathbf{b}, ..., \mathbf{A}^{r-1}\mathbf{b}\}$$

  Matriz de Krylov:

  $$\mathbf{K}_r(\mathbf{A},\mathbf{b}) = [\mathbf{b}, \mathbf{A}\mathbf{b}, \mathbf{A}^2\mathbf{b}, ..., \mathbf{A}^{r-1}\mathbf{b}]$$

Ejemplo:

$$\mathbf{A} = \begin{bmatrix} 10 & -1 & 2 & 0 \\ -1 & 11 & -1 & 3 \\ 2 & -1 & 10 & -1 \\ 0 & 3 & -1 & 8 \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \quad \mathbf{K}_3 = \begin{bmatrix} 1 & 11 & 118 \\ 1 & 12 & 141 \\ 1 & 10 & 100 \\ 1 & 10 & 106 \end{bmatrix}$$

  - Surge del método de Lanczos, aproximaciones de autovalores y autovectores de matrices dispersas grandes

- Una matriz $\mathbf{G} \in \mathbb{R}^{n\times n}$ es Gram si $\mathbf{G} = \mathbf{V}\mathbf{V}^\top \rightarrow (g_{ij}) = \langle \mathbf{v}_i, \mathbf{v}_j \rangle$ con $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n\} \in \mathbb{R}^n$
  - **G** es hermítica y semidefinida positiva
  - **G** no es singular si sus vectores son linealmente independientes
  - El rango de **G** es igual al número de vectores linealmente independientes

Determinante de **G**:

$$|G(v_1, \ldots, v_n)| = \|v_1 \wedge \cdots \wedge v_n\|^2$$

La función ***K*** es un kernel si su “kernel matrix” es simétrica y definida positiva

$$\mathbf{X} = \{\mathbf{x}_1, ..., \mathbf{x}_m\}$$

$$\boldsymbol{K}(\mathbf{x}_i, \mathbf{x}_j) : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$$

Las entradas de la matriz **G** son el kernel de los “data points” correspondientes

$$\mathbf{G}_{ij} = \boldsymbol{K}(\mathbf{x}_i, \mathbf{x}_j)$$

$$\mathbf{G} = \mathbf{X}\mathbf{X}^\top$$

Solución de los problemas de regresión lineal

$$\min_{\mathbf{w}} \sum_{i=1}^{n} (\mathbf{w}^\top \mathbf{x}_i - y_i)^2$$

$$\mathbf{w} = (\mathbf{X}\mathbf{X}^\top)^{-1}\mathbf{X}\mathbf{y}$$

Matrices estocásticas y doblemente estocásticas

- Una matriz $\mathbf{P} \in \mathbb{R}^{n\times n}$ no negativa cuyas filas (columnas) suman 1 se llama matriz estocástica

Row stochastic

$$\sum_{j=1}^{n} P_{i,j} = 1, \ \forall i$$

Column stochastic

$$\sum_{i=1}^{n} P_{i,j} = 1, \ \forall j$$

Double stochastic

$$\mathbf{P}\mathbf{1}_n = \mathbf{1}_n \text{ y } \mathbf{1}_n^\top \mathbf{P} = \mathbf{1}_n^\top$$

¿Autovalor?
¿Autovector?

$$\mathbf{P} = \frac{1}{n}\mathbf{1}_{n\times n} \qquad \boldsymbol{\mu}_{\mathbf{x}} = \frac{1}{n}\mathbf{1}_{n\times n}\mathbf{x}$$

- Ejemplos: promedios, permutaciones, matrices de transición de una cadena de Markov finita…
  - Cadena de Markov tenemos: n estados, matriz de transición $(\mathrm{n} \times \mathrm{n})$, probabilidad de cada estado por instantes

Distribución de prob. instante inicial

$$\mathbf{p}_{(0)}^\top = [p_1(0), p_2(0), ..., p_n(0)]^\top$$

Matriz de transición

$$\mathbf{P} = \begin{bmatrix} P_{1,1} & \dots & P_{1,n} \\ \vdots & \dots & \vdots \\ P_{n,1} & \dots & P_{n,n} \end{bmatrix}$$

$P_{x,y}$= prob de, estando en x, transicionar a y

Distribución de prob. tras k transiciones

$$\mathbf{p}_{(k)}^\top = \mathbf{p}_{(0)}^\top \mathbf{P}^k \qquad \mathbf{p}_{(k)} = (\mathbf{P}^\top)^k \mathbf{p}_{(0)}$$

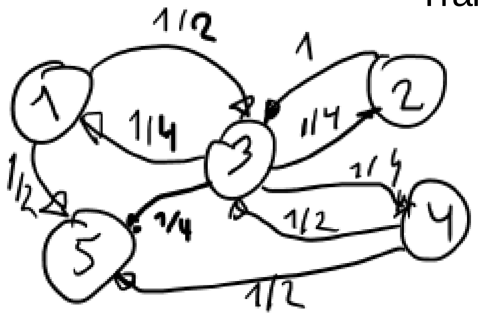
EJEMPLO DEL GATO Y EL RATÓN

El ratón empieza en la caja 1 y el gato en la 5

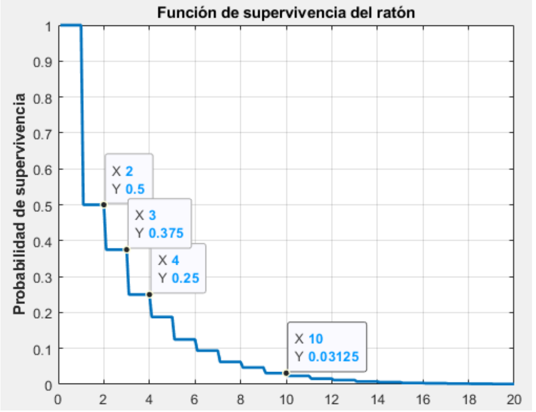
$$\mathbf{p}_{(0)}^R = [1,0,0,0,0]^\top$$

$$\mathbf{p}_{(0)}^G = [0,0,0,0,1]^\top$$

Transiciones



$$\mathbf{P} = \begin{bmatrix} 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 1 & 0 & 0 \\ 1/4 & 1/4 & 0 & 1/4 & 1/4 \\ 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$



Matriz de Carleman

- Realmente hay muchas más matrices con estructura. Su estudio/interés depende de la rama científica en la que trabajemos ➔ A continuación mostramos algunos ejemplos
- La matriz de Carleman de una función infinitamente diferenciable *f(x)* se define como

$$M[f]_{jk} = \frac{1}{k!}\left[\frac{d^k}{dx^k}(f(x))^j\right]_{x=0}$$

Bell matrix:

$$B[f]_{jk} = \frac{1}{j!}\left[\frac{d^j}{dx^j}(f(x))^k\right]_{x=0}$$

Sumatorio de productos:

$$f(x)^j = \sum_{k=0}^{\infty} \mathbf{M}[f]_{(j,k)} x^k$$

Producto matricial:

$$\mathbf{M}[f]\begin{bmatrix}1 & x & x^2 & x^3 & \dots\end{bmatrix}^\top = \begin{bmatrix}1 & f(x) & f(x)^2 & f(x)^3 & \dots\end{bmatrix}^\top$$

$$\mathbf{B} = \mathbf{M}^\top$$

- Propiedades
  - **M** es una representación directa de *f(x)*

$$\mathbf{M}[f \circ g] = \mathbf{M}[f]\mathbf{M}[g]$$

  - **B** es una anti-representación de *f(x)*

$$\mathbf{B}[f \circ g] = \mathbf{B}[g]\mathbf{B}[f]$$

Composición de funciones como multiplicación de matrices

  - Potencias e inversa de **M**

$$\mathbf{M}[f^n] = \mathbf{M}[f]^n, \ \mathbf{M}[f^{-1}] = \mathbf{M}[f]^{-1}$$

Ejemplo: Matriz de Carleman de una función lineal

$$\begin{array}{ccccc} \mathbf{M}[a+cx] & & \mathbf{x} & & \mathbf{y} \\ \begin{bmatrix} 1 & 0 & 0 & \dots \\ a & c & 0 & \dots \\ a^2 & 2ac & c^2 & \dots \\ \vdots & \vdots & \vdots & \dots \end{bmatrix} & & \begin{bmatrix} 1 \\ x \\ x^2 \\ \vdots \end{bmatrix} & = & \begin{bmatrix} 1 \\ a+cx \\ a^2+2acx+c^2x^2 \\ \vdots \end{bmatrix} \end{array}$$

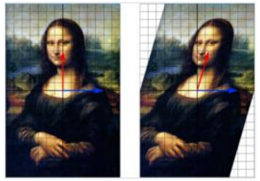
Recapitulación de conceptos

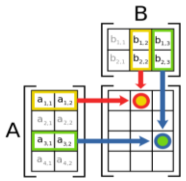
- Si las matrices tienen estructura adicional, son más fáciles de entender y, por lo tanto, existen más resultados específicos para ese tipo de matriz
- Tipos de matrices que hemos visto:
  - Persimétricas, centrosimétricas, circulante por bloques, Hankel
  - Idempotentes, tripotentes, nilpotentes, unipotentes
  - Matrices de Hessenberg y tridiagonales
  - Matriz de Hadamard
  - Matrices de Vandermonde y de Fourier
  - Matrices positivas, no-negativas e irreducibles
  - Matrices indescomponibles y de Perron
  - Estables, estables positivas, P, Z, L y M
  - Matrices de Krylov y Gram
  - Estocásticas y doblemente estocásticas
  - Matriz de Carleman

Cálculo con matrices I

“Vectores, matrices y tensores: Una perspectiva desde ciencia e ingeniería de datos”

Reading Group organizado por el “Grupo de Investigación de Alto Rendimiento en Ciencia de Datos y Procesamiento de Señal de la URJC”





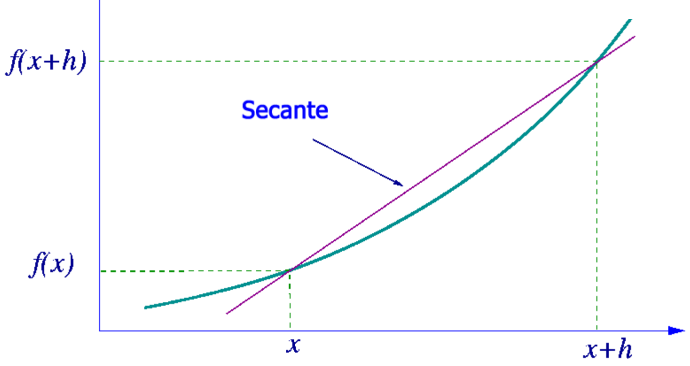
Esta presentación utiliza material cuya autoría corresponde a: Marques, Buciulea, Wikipedia; así como contribuciones menores de distintas autorías.

Índice de la clase

- Derivada de una función c.r.a. un escalar
- Derivada de una función c.r.a un vector
- Derivada de una función c.r.a una matriz
- Reglas de transformación
- Diferenciales de matrices
- Perturbaciones usando diferenciales
- Derivadas de normas de vectores/matrices
- Derivadas de matrices con estructura
- Ecuaciones diferenciales lineales
- Derivadas de segundo orden

¿Qué es una derivada?

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$



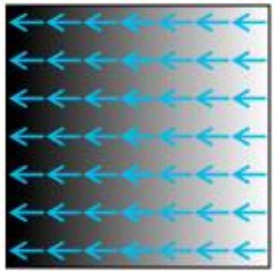
- **La derivada** de f en x es el límite del valor del cociente diferencial, conforme las líneas secantes se aproximan a la línea tangente

- **El gradiente** es una generalización de la derivada. Si $f : \mathbb{R}^n \to \mathbb{R}$ entonces el gradiente de f en $\mathbf{r}$ se define como una función vectorial $\nabla f : \mathbb{R}^n \to \mathbb{R}^n$

$$\nabla f(\mathbf{r}) = \left(\frac{\partial f(\mathbf{r})}{\partial x_1}, \ldots, \frac{\partial f(\mathbf{r})}{\partial x_n} \right)$$

Ejemplo: $f(x,y,z) = 2x + 3y^2 - \operatorname{sen}(z)$

$$\nabla f = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z} \right) = (2, 6y, -\cos(z))$$



- **La matriz jacobiana** de una función de varias variables es la matriz cuyos elementos son las derivadas parciales de primer orden de dicha función

$$\mathbf{J} = \begin{bmatrix} \dfrac{\partial \mathbf{F}}{\partial x_1} & \cdots & \dfrac{\partial \mathbf{F}}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \nabla^T f_1 \\ \vdots \\ \nabla^T f_m \end{bmatrix} = \begin{bmatrix} \dfrac{\partial f_1}{\partial x_1} & \cdots & \dfrac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial f_m}{\partial x_1} & \cdots & \dfrac{\partial f_m}{\partial x_n} \end{bmatrix}$$

Ejemplo: $\mathbf{F}(x_1, x_2, x_3) = (x_1, 5x_3, 4x_2^2 - 2x_3)$

$$f_1 = x_1$$
$$f_2 = 5x_3$$
$$f_3 = 4x_2^2 - 2x_3$$

$$\mathbf{J_F}(x_1, x_2, x_3) = \begin{bmatrix} \dfrac{\partial f_1}{\partial x_1} & \dfrac{\partial f_1}{\partial x_2} & \dfrac{\partial f_1}{\partial x_3} \\ \dfrac{\partial f_2}{\partial x_1} & \dfrac{\partial f_2}{\partial x_2} & \dfrac{\partial f_2}{\partial x_3} \\ \dfrac{\partial f_3}{\partial x_1} & \dfrac{\partial f_3}{\partial x_2} & \dfrac{\partial f_3}{\partial x_3} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 5 \\ 0 & 8x_2 & -2 \end{bmatrix}$$

Derivada de funciones con respecto a un escalar

- Derivada de un vector o matriz con respecto a un escalar

$$\mathbf{a}(t) = a_i(t) \rightarrow \frac{\partial \mathbf{a}(t)}{\partial t} = \frac{\partial a_i(t)}{\partial t} \qquad \mathbf{A}(t) = a_{ij}(t) \rightarrow \frac{\partial \mathbf{A}(t)}{\partial t} = \frac{\partial a_{ij}(t)}{\partial t}$$

Derivada de cada uno de sus elementos

- Propiedades

Por definición

$$\frac{\partial\{\mathbf{A}\mathbf{X}(t)\mathbf{C}\}}{\partial t} = \mathbf{A}\frac{\partial \mathbf{X}(t)}{\partial t}\mathbf{C}$$

$$\frac{\partial \mathrm{vec}\,\mathbf{X}(t)}{\partial t} = \mathrm{vec}\,\frac{\partial \mathbf{X}(t)}{\partial t}$$

Derivada del producto de matrices

$$\frac{\partial \mathbf{A}(t)\mathbf{B}(t)\mathbf{C}(t)}{\partial t} = \frac{\partial \mathbf{A}}{\partial t}\mathbf{B}\mathbf{C} + \mathbf{A}\frac{\partial \mathbf{B}}{\partial t}\mathbf{C} + \mathbf{A}\mathbf{B}\frac{\partial \mathbf{C}}{\partial t}$$

$$\frac{\partial(\mathbf{A}(t) \otimes \mathbf{B}(t))}{\partial t} = \frac{\partial \mathbf{A}}{\partial t} \otimes \mathbf{B} + \mathbf{A} \otimes \frac{\partial \mathbf{B}}{\partial t}$$

Derivada de $\boldsymbol{A}^{-1}$ y $\boldsymbol{A}^{-}$ weak inverse

$$\frac{\partial \mathbf{A}^{-1}(t)}{\partial t} = -\mathbf{A}^{-1}\frac{\partial \mathbf{A}(t)}{\partial t}\mathbf{A}^{-1}$$

$$\mathbf{A}\frac{\partial \mathbf{A}^{-}(t)}{\partial t}\mathbf{A} = -\mathbf{A}\mathbf{A}^{-}\frac{\partial \mathbf{A}}{\partial t}\mathbf{A}^{-}\mathbf{A}$$

Derivada del determinante

$$\frac{\partial \det \mathbf{A}}{\partial t} = \det \mathbf{A}\frac{\partial \log(\det \mathbf{A})}{\partial t}$$

Derivada de la traza

$$\frac{\partial[\mathrm{trace}(\mathbf{A}(t))]}{\partial t} = \mathrm{trace}\left[\frac{\partial \mathbf{A}(t)}{\partial t}\right]$$

Derivada de una exponencial

$$\frac{\partial e^{\mathbf{A}t}}{\partial t} = \mathbf{A}e^{\mathbf{A}t}$$

Derivada del logdet

$$\frac{\partial \log(\det \mathbf{A})}{\partial t} = \mathrm{trace}\left[\mathbf{A}^{-1}\frac{\partial \mathbf{A}}{\partial t}\right]$$

$$\boxed{\frac{\partial \det(\mathbf{A})}{\partial \mathbf{H}} := D(\det)|_{\mathbf{A}}(\mathbf{H}) = \lim_{\epsilon \to 0}\frac{\det(\mathbf{A}+\epsilon\mathbf{H}) - \det(\mathbf{A})}{\epsilon} = \mathrm{tr}\left(\mathrm{adj}(\mathbf{A})\,\mathbf{H}\right)}$$

Derivada escalar II

- Derivada de una función con respecto a los elementos de un vector

$$\mathbf{F}(\mathbf{x}) \rightarrow \text{función vectorial del vector } \mathbf{x} = (x_1, x_2, \ldots, x_n)^{\mathsf{T}}$$

Pasamos de derivar con respecto a t a derivar con respecto a x_i

- Propiedades

Derivada de la traza

$$\frac{\partial\, \mathrm{trace}[\mathbf{F}(\mathbf{x})]}{\partial x_i} = \mathrm{trace}\left(\frac{\partial \mathbf{F}}{\partial x_i}\right)$$

Derivada del producto de Kronecker

$$\frac{\partial(\mathbf{F}\otimes\mathbf{G})}{\partial x_i} = \mathbf{F}\otimes\frac{\partial \mathbf{G}}{\partial x_i} + \frac{\partial \mathbf{F}}{\partial x_i}\otimes\mathbf{G}$$

$$\frac{\partial \mathrm{vec}\,(\mathbf{F}\otimes\mathbf{G})}{\partial x_i} = (\mathbf{I}_q\otimes\mathbf{K}_{sp}\otimes\mathbf{I}_r)\frac{\partial(\mathrm{vec}\,\mathbf{F}\otimes\mathrm{vec}\,\mathbf{G})}{\partial x_i}$$

Matriz de conmutación

Derivada de $\boldsymbol{F}^{-1}$ y $\boldsymbol{F}^{-}$

$$\frac{\partial \mathbf{F}^{-1}}{\partial x_i} = -\mathbf{F}^{-1}\frac{\partial \mathbf{F}}{\partial x_i}\mathbf{F}^{-1}$$

$$\mathbf{F}\frac{\partial \mathbf{F}^{-}}{\partial x_i}\mathbf{F} = -\mathbf{F}\mathbf{F}^{-}\frac{\partial \mathbf{F}}{\partial x_i}\mathbf{F}^{-}\mathbf{F}$$

Derivada del determinante y el logdet

$$\frac{\partial \det[\mathbf{F}(\mathbf{x})]}{\partial x_i} = (\det \mathbf{F})\,\mathrm{trace}\left(\mathbf{F}^{-1}\frac{\partial \mathbf{F}}{\partial x_i}\right)$$

$$\frac{\partial \log\det[\mathbf{F}(\mathbf{x})]}{\partial x_i} = \frac{1}{\det \mathbf{F}}\frac{\partial \det \mathbf{F}}{\partial x_i} = \mathrm{trace}\left(\mathbf{F}^{-1}\frac{\partial \mathbf{F}}{\partial x_i}\right)$$

$$\frac{\partial \log\det(\mathbf{A}\mathbf{F}^{-1}\mathbf{B})}{\partial x_i} = -\,\mathrm{trace}\left[\mathbf{F}^{-1}\mathbf{B}(\mathbf{A}\mathbf{F}^{-1}\mathbf{B})^{-1}\mathbf{A}\mathbf{F}^{-1}\frac{\partial \mathbf{F}}{\partial x_i}\right]$$

Autovalores y autovectores

$$\mathbf{F}\mathbf{u} = \lambda\mathbf{u} \qquad \mathbf{H}^{+} = \mathbf{F} - \lambda\mathbf{I}$$

$$\frac{\partial \lambda}{\partial x_i} = \mathbf{u}'\frac{\partial \mathbf{F}}{\partial x_i}\mathbf{u} \qquad \frac{\partial \mathbf{u}}{\partial x_i} = -\mathbf{H}^{+}\frac{\partial \mathbf{F}}{\partial x_i}\mathbf{u}$$

Modelado no lineal

$$\frac{\partial \det(\mathbf{X}'\mathbf{X})}{\partial \phi_i} = \det(\mathbf{X}'\mathbf{X})\,\mathrm{trace}\left(\mathbf{X}^{+}\frac{\partial \mathbf{X}}{\partial \phi_i}\right)$$

Derivada escalar III

- Derivada de una matriz con respecto a los elementos de la matriz

$$\mathbf{X} \in \mathbb{R}^{m\times n}$$

$$\mathbf{E}_{ij} \in \mathbb{R}^{m\times n}$$

$$\frac{\partial \mathbf{X}}{\partial x_{ij}} = \begin{cases} \mathbf{E}_{ij}, & \mathbf{X} \text{ unconstrained}, \\ \mathbf{E}_{ij} + \mathbf{E}'_{ij} - \delta_{ij}\mathbf{E}_{ii}, & \mathbf{X} \text{ symmetric}, \end{cases}$$

$$\mathbf{E}_{ij} = \mathbf{e}_{i,m}\mathbf{e}_{n,j}^{\mathsf{T}}$$

1 en la posición (*i,j*) y cero en el resto

$$\delta_{ij} = 1 \text{ cuando } i = j \text{ y } \delta_{ij} = 0 \text{ cuando } i \neq j$$

- Ejemplos

Derivada del producto de matrices

$$\frac{\partial(\mathbf{AXB})}{\partial x_{ij}} = \mathbf{A}\frac{\partial \mathbf{X}}{\partial x_{ij}}\mathbf{B} = \mathbf{A}\mathbf{E}_{ij}\mathbf{B}$$

$$\frac{\partial(\mathbf{AX'B})}{\partial x_{ij}} = \mathbf{A}\mathbf{E}'_{ij}\mathbf{B}$$

$$\frac{\partial(\mathbf{X'AXB})}{\partial x_{ij}} = \frac{\partial \mathbf{X}'}{\partial x_{ij}}\mathbf{AXB} + \mathbf{X}'\frac{\partial(\mathbf{AXB})}{\partial x_{ij}} = \mathbf{E}'_{ij}\mathbf{AXB} + \mathbf{X'A}\mathbf{E}_{ij}\mathbf{B}.$$

$$\frac{\partial(\mathbf{XAX'B})}{\partial x_{ij}} = \mathbf{E}_{ij}\mathbf{AX'B} + \mathbf{XA}\mathbf{E}'_{ij}\mathbf{B}$$

$$\frac{\partial(\mathbf{AXX'B})}{\partial x_{ij}} = \mathbf{A}\frac{\partial \mathbf{XX}'}{\partial x_{ij}}\mathbf{B} = \mathbf{A}\mathbf{E}_{ij}\mathbf{X'B} + \mathbf{AX}\mathbf{E}'_{ij}\mathbf{B}$$

$$\frac{\partial(\mathbf{XAXB})}{\partial x_{ij}} = \mathbf{E}_{ij}\mathbf{AXB} + \mathbf{XA}\mathbf{E}_{ij}\mathbf{B}$$

$$\frac{\partial \mathbf{XX'X}}{\partial x_{ij}} = \mathbf{E}_{ij}\mathbf{X'X} + \mathbf{X}\mathbf{E}'_{ij}\mathbf{X} + \mathbf{XX'}\mathbf{E}_{ij}$$

$$\frac{\partial(\mathbf{X'AX'B})}{\partial x_{ij}} = \mathbf{E}'_{ij}\mathbf{AX'B} + \mathbf{X'A}\mathbf{E}'_{ij}\mathbf{B}$$

Derivada escalar III

- Más propiedades
  - Derivada de la inversa de $\mathbf{X} \in \mathbb{R}^{n\times n}$

La columna *i* de $\mathbf{X}^{-1}$

$$\frac{\partial \mathbf{X}^{-1}}{\partial x_{ij}} = -\mathbf{X}^{-1}\mathbf{E}_{ij}\mathbf{X}^{-1} = -\mathbf{X}^{-1}\mathbf{e}_{i,m}\mathbf{e}'_{j,m}\mathbf{X}^{-1} = -\mathbf{y}_i\mathbf{z}'_j$$

Es simétrica

La fila *j* de $\mathbf{X}^{-1}$

$$\frac{\partial \mathbf{X}^{-1}}{\partial x_{ij}} = \begin{cases} -\mathbf{y}_i\mathbf{y}'_i, & \text{if } i=j \\ -\mathbf{y}_i\mathbf{y}'_j - \mathbf{y}_j\mathbf{y}'_i, & \text{if } i>j \end{cases}$$

Expresión general $\quad (\mathbf{BXC})(\mathbf{BXC})^{-1} = \mathbf{I}$

$$\frac{\partial\{\mathbf{A}(\mathbf{BXC})^{-1}\mathbf{D}\}}{\partial x_{ij}} = -\mathbf{A}(\mathbf{BXC})^{-1}\mathbf{B}\mathbf{E}_{ij}\mathbf{C}(\mathbf{BXC})^{-1}\mathbf{D}$$

- Derivada de Funciones de matrices

$$\frac{\partial \det \mathbf{Y}}{\partial x_{ij}} = \operatorname{trace}\left[(\operatorname{adj}\mathbf{Y})\left(\frac{\partial \mathbf{Y}}{\partial x_{ij}}\right)\right] = (\det \mathbf{Y})\operatorname{trace}\left[\mathbf{Y}^{-1}\frac{\partial \mathbf{Y}}{\partial x_{ij}}\right]$$

$$\frac{\partial \operatorname{vec}\mathbf{Y}}{\partial x_{ij}} = \operatorname{vec}\frac{\partial \mathbf{Y}}{\partial x_{ij}} \qquad \frac{\partial \operatorname{trace}\mathbf{Y}}{\partial x_{ij}} = \operatorname{trace}\left(\frac{\partial \mathbf{Y}}{\partial x_{ij}}\right) \qquad \frac{\partial \mathbf{Y}^{-1}}{\partial x_{ij}} = -\mathbf{Y}^{-1}\frac{\partial \mathbf{Y}}{\partial x_{ij}}\mathbf{Y}^{-1}$$

- Derivada del determinante

Cofactor (adjunto) de x_{ij}

$$\frac{\partial \det \mathbf{X}}{\partial x_{ij}} = \begin{cases} \xi_{ij}, & \mathbf{X} \text{ unconstrained}, \\ (2-\delta_{ij})\xi_{ij}, & \mathbf{X} = \mathbf{X}', \end{cases}$$

$\delta_{ij} = 1$ cuando $i = j$ y $\delta_{ij} = 0$ cuando $i \neq j$

- Derivada de potencias de **X**

$$\frac{\partial \mathbf{X}^k}{\partial x_{ij}} = \sum_{r=0}^{k-1}\mathbf{X}^r\mathbf{E}_{ij}\mathbf{X}^{k-r-1}$$

para $\mathbf{X}^k\mathbf{X}^{-k} = \mathbf{I}$

$$\frac{\partial \mathbf{X}^{-k}}{\partial x_{ij}} = -\mathbf{X}^{-k}\frac{\partial \mathbf{X}^k}{\partial x_{ij}}\mathbf{X}^{-k}$$

- Derivada de autovalores y autovectores

$$\frac{\partial \lambda}{\partial x_{ij}} = 2\mathbf{u}\mathbf{u}' - \operatorname{diag}(u_1^2, u_2^2, \ldots, u_n^2)$$

$$\mathbf{X}\mathbf{u} = \lambda\mathbf{u}$$

$$\mathbf{H}^+ = \mathbf{X} - \lambda\mathbf{I}_n$$

$$\frac{\partial \mathbf{u}}{\partial x_{ij}} = \begin{cases} -u_i\mathbf{g}_i, & \text{if } j=i \\ -(u_j\mathbf{g}_i + u_i\mathbf{g}_j), & \text{if } j<i \end{cases}$$

La columna *j* de $\mathbf{H}^+$

Derivada vectorial de funciones escalares I

- Derivada de una función escalar respecto a un vector

$$\frac{\partial f}{\partial \mathbf{x}} = \left(\frac{\partial f}{\partial x_i}\right) \qquad \frac{\partial \mathbf{x}^\mathsf{T}\mathbf{a}}{\partial \mathbf{x}} = \mathbf{a} = \frac{\partial \mathbf{a}^\mathsf{T}\mathbf{x}}{\partial \mathbf{x}} \qquad \frac{\partial \mathbf{x}^\mathsf{T}\mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = (\mathbf{A}+\mathbf{A}^\mathsf{T})\mathbf{x} = 2\mathbf{A}\mathbf{x}$$

Si **A** es simétrica

En forma de vector fila

$$\frac{\partial z}{\partial \mathbf{x}^\mathsf{T}} = \frac{\partial z}{\partial \mathbf{y}^\mathsf{T}} \cdot \frac{\partial \mathbf{y}}{\partial \mathbf{x}^\mathsf{T}}$$

En forma de vector columna

$$\frac{\partial z}{\partial \mathbf{x}} = \left(\frac{\partial \mathbf{y}}{\partial \mathbf{x}^\mathsf{T}}\right)^\mathsf{T} \frac{\partial z}{\partial \mathbf{y}}.$$

- Propiedades

Función escalar de **X**

$$\frac{\partial f(\mathbf{X})}{\partial \operatorname{vec}\mathbf{X}} = \operatorname{vec}\left(\frac{\partial f(\mathbf{X})}{\partial \mathbf{X}}\right) = \left(\frac{\partial f(\mathbf{X})}{\partial(\operatorname{vec}\mathbf{X})'}\right)'$$

- Derivada de la traza

$$\frac{\partial \operatorname{trace}(\mathbf{AXB})}{\partial \operatorname{vec}\mathbf{X}} = \operatorname{vec}(\mathbf{A}'\mathbf{B}')$$

$$\operatorname{trace}(\mathbf{AXB}) = \operatorname{trace}(\mathbf{BAX}) = \operatorname{vec}(\mathbf{A}'\mathbf{B}')'\operatorname{vec}\mathbf{X}$$

$$\frac{\partial \operatorname{trace}(\mathbf{X}'\mathbf{AXB})}{\partial \operatorname{vec}\mathbf{X}} = [(\mathbf{B}'\otimes\mathbf{A}) + (\mathbf{B}\otimes\mathbf{A}')]\operatorname{vec}\mathbf{X}$$

- Derivada del determinante y el logdet

$$\frac{\partial \det\mathbf{X}}{\partial \operatorname{vec}\mathbf{X}} = \operatorname{vec}[(\operatorname{adj}\mathbf{X})'] = (\det\mathbf{X})\operatorname{vec}(\mathbf{X}^{-1\prime})$$

Si **Y** es una función de **X**

$$\frac{\partial(\log\det\mathbf{Y})}{\partial \operatorname{vec}\mathbf{X}} = \frac{1}{\det\mathbf{Y}}\frac{\partial\det\mathbf{Y}}{\partial \operatorname{vec}\mathbf{X}}$$

- Ejemplo $\mathbf{Y} = \mathbf{XBX}'$

$$\frac{\partial\det\mathbf{Y}}{\partial \operatorname{vec}\mathbf{X}} = \det\mathbf{Y}[\mathbf{B}\otimes(\mathbf{Y}^{-1})' + (\mathbf{B}'\otimes\mathbf{Y}^{-1})]\operatorname{vec}\mathbf{X}$$

Si **B** es simétrica → **Y** es simétrica

$$\frac{\partial\det\mathbf{Y}}{\partial \operatorname{vec}\mathbf{X}} = 2\det\mathbf{Y}[\mathbf{B}\otimes(\mathbf{Y}^{-1})]\operatorname{vec}\mathbf{X}$$

Derivada vectorial de funciones escalares I: ejemplo I

- Consideramos muestras Gaussianas $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ cuya PDF viene dada por

$$f(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \det(2\pi\boldsymbol{\Sigma})^{-\frac{1}{2}} \exp(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}))$$

- Queremos encontrar el vector $\boldsymbol{\mu}$ que mejor describa nuestras observaciones
  - Calculamos la función de *"log-likelihood" asociada a* $f(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$

$$\mathcal{L}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \log(f(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})) = -\frac{1}{2}\log\det(2\pi\boldsymbol{\Sigma}) - \frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}))$$

Recordatorio

$$\frac{\partial \mathbf{x}^{\mathsf{T}}\mathbf{a}}{\partial \mathbf{x}} = \mathbf{a} = \frac{\partial \mathbf{a}^{\mathsf{T}}\mathbf{x}}{\partial \mathbf{x}}$$

$$\frac{\partial \mathbf{x}^{\mathsf{T}}\mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = (\mathbf{A}+\mathbf{A}^{\mathsf{T}})\mathbf{x} = 2\mathbf{A}\mathbf{x}$$

  - Derivamos con respecto al vector $\boldsymbol{\mu}$

$$\frac{\partial \mathcal{L}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})}{\partial \boldsymbol{\mu}} = -\frac{1}{2}\frac{\partial(\mathbf{x}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\mathbf{x} - \mathbf{x}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \boldsymbol{\mu}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\mathbf{x} + \boldsymbol{\mu}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu})}{\partial \boldsymbol{\mu}} = \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})$$

Derivada vectorial de funciones escalares I: ejemplo II

- Consideramos que queremos encontrar el valor de **x** de la siguiente función

$$f(\mathbf{x}) = \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2$$ "Least squares problem"

- Queremos encontrar el vector **x** que se adapte mejor a nuestras observaciones en **y**

Recordatorio

$$\frac{\partial \mathbf{x}^\top \mathbf{a}}{\partial \mathbf{x}} = \mathbf{a} = \frac{\partial \mathbf{a}^\top \mathbf{x}}{\partial \mathbf{x}}$$

$$\frac{\partial \mathbf{x}^\top \mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = (\mathbf{A} + \mathbf{A}^\top)\mathbf{x} = 2\mathbf{A}\mathbf{x}$$

  - Derivamos $f(\mathbf{x})$ con respecto al vector **x**

$$\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} = \frac{\partial \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2}{\partial \mathbf{x}} = \frac{\partial[(\mathbf{y} - \mathbf{A}\mathbf{x})^\top(\mathbf{y} - \mathbf{A}\mathbf{x})]}{\partial \mathbf{x}} = \frac{\partial(\mathbf{y}^\top\mathbf{y} - \mathbf{y}^\top\mathbf{A}\mathbf{x} - \mathbf{x}^\top\mathbf{A}^\top\mathbf{y} - \mathbf{x}^\top\mathbf{A}^\top\mathbf{A}\mathbf{x})}{\partial \mathbf{x}}$$

$$= -\mathbf{A}^\top\mathbf{y} - \mathbf{A}^\top\mathbf{y} + 2\mathbf{A}^\top\mathbf{A}\mathbf{x} = 0 \Rightarrow \mathbf{x} = (\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{A}^\top\mathbf{y}$$

Derivada vectorial de funciones escalares II

- Derivada de una función de funciones respecto a un vector
  - Partimos de $y = \mathbf{w}^{\mathsf{T}}\mathbf{A}\mathbf{z}$ donde $\mathbf{A} \in \mathbb{R}^{m\times n}$, $\mathbf{w} \in \mathbb{R}^{m}$, $\mathbf{z} \in \mathbb{R}^{n}$ son funciones de **x**
  - Queremos encontrar el vector fila $\partial y/\partial \mathbf{x}^{\mathsf{T}}$

$$y = \mathrm{vec}\, y = (\mathbf{z}' \otimes \mathbf{w}')\mathrm{vec}\,\mathbf{A} = (\mathbf{z} \otimes \mathbf{w})'\mathrm{vec}\,\mathbf{A} = [\mathrm{vec}\,(\mathbf{w}\mathbf{z}')]'\mathrm{vec}\,\mathbf{A}$$

$$\frac{\partial y}{\partial \mathbf{x}'} = \mathbf{z}'\mathbf{A}'\frac{\partial \mathbf{w}}{\partial \mathbf{x}'} + [\mathrm{vec}\,(\mathbf{w}\mathbf{z}')]'\frac{\partial\, \mathrm{vec}\,\mathbf{A}}{\partial \mathbf{x}'} + \mathbf{w}'\mathbf{A}\frac{\partial \mathbf{z}}{\partial \mathbf{x}'}$$

- Derivada de la traza de una función de funciones

$$y = \mathrm{trace}[\mathbf{F}(\mathbf{Z})]$$

F es una función matricial de **Z**

Z es una función matricial de **x**

  - Aplicando la regla de la cadena

$$\frac{\partial y}{\partial \mathbf{x}'} = \frac{\partial y}{\partial (\mathrm{vec}\,\mathbf{Z})'} \cdot \frac{\partial\, \mathrm{vec}\,\mathbf{Z}}{\partial \mathbf{x}'}$$

  - Ejemplos

$$\frac{\partial\, \mathrm{trace}(\mathbf{AZB})}{\partial \mathbf{x}'} = \mathrm{vec}\,(\mathbf{A}'\mathbf{B}')'\frac{\partial\, \mathrm{vec}\,\mathbf{Z}}{\partial \mathbf{x}'}$$

$$\frac{\partial\, \mathrm{trace}(\mathbf{AZ}'\mathbf{BZ})}{\partial \mathbf{x}'} = (\mathrm{vec}\,\mathbf{Z})'(\mathbf{A}' \otimes \mathbf{B} + \mathbf{A} \otimes \mathbf{B}')\frac{\partial\, \mathrm{vec}\,\mathbf{Z}}{\partial \mathbf{x}'}$$

$$\frac{\partial \det \mathbf{Z}}{\partial \mathbf{x}'} = \mathrm{vec}\,[(\mathrm{adj}\mathbf{Z})']'\frac{\partial\, \mathrm{vec}\,\mathbf{Z}}{\partial \mathbf{x}'}$$

Derivada vectorial de funciones vectoriales I

- Derivada de una función vectorial respecto a un vector
  - La derivada de cada una de las entradas de **y** con respecto a cada una de las entradas de **x**

$\mathbf{X} \in \mathbb{R}^{m\times n}$ $\mathbf{A} \in \mathbb{R}^{m\times m}$ $\mathbf{B} \in \mathbb{R}^{n\times n}$ $\mathbf{x} \in \mathbb{R}^{m}$ $\mathbf{y} \in \mathbb{R}^{m}$ $\mathbf{a} \in \mathbb{R}^{n}$

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}'} = \left(\frac{\partial y_i}{\partial x_j}\right)$$

$$\mathbf{y} = \mathbf{A}\mathbf{x} \qquad \partial \mathbf{y}/\partial \mathbf{x}' = \mathbf{A}$$

$$\frac{\partial \operatorname{vec} \mathbf{X}}{\partial (\operatorname{vec} \mathbf{X})'} = \mathbf{I}$$

$$\operatorname{vec} \mathbf{Y} = \mathbf{B} \operatorname{vec} \mathbf{X} \qquad \frac{\partial \operatorname{vec} \mathbf{Y}}{\partial (\operatorname{vec} \mathbf{X})'} = \mathbf{B}$$

Vector de $mn \times 1$

$$\frac{\partial (\mathbf{x} \otimes \mathbf{a})}{\partial \mathbf{x}'} = \mathbf{I}_n \otimes \mathbf{a}$$

Matriz de $mn \times m$

$$\frac{\partial (\mathbf{a} \otimes \mathbf{x})}{\partial \mathbf{x}'} = \mathbf{a} \otimes \mathbf{I}_n$$

$$\mathbf{Y} = \mathbf{A}\mathbf{X}\mathbf{B} \qquad \mathbf{Y} = (\mathbf{B}' \otimes \mathbf{A}) \operatorname{vec} \mathbf{X}$$

$$\frac{\partial \operatorname{vec} \mathbf{Y}}{\partial (\operatorname{vec} \mathbf{X})'} = \mathbf{B}' \otimes \mathbf{A}$$

$$\mathbf{Y} = \mathbf{A}\mathbf{X}^{-1}\mathbf{B}$$

$$\frac{\partial \operatorname{vec} \mathbf{Y}}{\partial (\operatorname{vec} \mathbf{X})'} = -(\mathbf{X}^{-1}\mathbf{B})' \otimes (\mathbf{A}\mathbf{X}^{-1})$$

$$\mathbf{Y} = \mathbf{X}'\mathbf{A}\mathbf{X}$$

$$\frac{\partial \operatorname{vec} \mathbf{Y}}{\partial (\operatorname{vec} \mathbf{X})'} = (\mathbf{X}'\mathbf{A}' \otimes \mathbf{I}_n)\mathbf{I}_{(n,m)} + (\mathbf{I}_n \otimes \mathbf{X}'\mathbf{A})$$

$$(\mathbf{I}_{n^2} + \mathbf{I}_{(n,n)})(\mathbf{I}_n \otimes \mathbf{X}'\mathbf{A})$$

$$\mathbf{Y} = \mathbf{A}\mathbf{X}'\mathbf{B} \qquad \mathbf{Y} = (\mathbf{B}' \otimes \mathbf{A}) \operatorname{vec} \mathbf{X}'$$

$$\frac{\partial \operatorname{vec} \mathbf{Y}}{\partial (\operatorname{vec} \mathbf{X})'} = (\mathbf{B}' \otimes \mathbf{A})\mathbf{I}_{(n,m)}$$

Matriz de conmutación

$$\mathbf{Y} = \mathbf{X}^k$$

$$\frac{\partial \operatorname{vec} \mathbf{Y}}{\partial (\operatorname{vec} \mathbf{X})'} = \sum_{i=1}^{k} ((\mathbf{X}')^{k-i} \otimes \mathbf{X}^{i-1})$$

$$\mathbf{Y} = \mathbf{X}\mathbf{B}\mathbf{X}'$$

$$\frac{\partial \operatorname{vec} \mathbf{Y}}{\partial (\operatorname{vec} \mathbf{X})'} = (\mathbf{X}\mathbf{B}' \otimes \mathbf{I}_m) + (\mathbf{I}_m \otimes \mathbf{X}\mathbf{B})\mathbf{I}_{(n,m)}$$

$$(\mathbf{I}_{m^2} + \mathbf{I}_{(m,m)})(\mathbf{X}\mathbf{B} \otimes \mathbf{I}_m)$$

Derivada vectorial de funciones vectoriales II

- Derivada de una función vectorial **F** genérica con respecto a un vector

  - Multiplicación de matrices

$$\frac{\partial\,\mathrm{vec}\,(\mathbf{AFB})}{\partial\mathbf{x}'} = (\mathbf{B}'\otimes\mathbf{A})\frac{\partial\,\mathrm{vec}\,\mathbf{F}}{\partial\mathbf{x}'}$$

Si **F** es simétrica

$$\frac{\partial\,\mathrm{vech}\,(\mathbf{AFA}')}{\partial\mathbf{x}'} = \mathbf{H}_n(\mathbf{A}\otimes\mathbf{A})\mathbf{G}_n\frac{\partial\,\mathrm{vech}\,\mathbf{F}}{\partial\mathbf{x}'}$$

Matriz identidad de $pq \times pq$

$$\mathbf{F}\in\mathbb{R}^{p\times q} \qquad \mathbf{F}=\mathbf{X} \qquad \frac{\partial\,\mathrm{vec}\,\mathbf{F}}{\partial\mathbf{x}'} = \mathbf{I}_{pq}$$

  - Inversa de F

$$\frac{\partial\,\mathrm{vec}\,\mathbf{F}^{-1}}{\partial\mathbf{x}'} = -(\mathbf{F}^{-1\prime}\otimes\mathbf{F}^{-1})\frac{\partial\,\mathrm{vec}\,\mathbf{F}}{\partial\mathbf{x}'}$$

Si **F** es simétrica

$$\frac{\partial\,\mathrm{vech}\,\mathbf{F}^{-1}}{\partial\mathbf{x}'} = -\mathbf{H}_n(\mathbf{F}^{-1}\otimes\mathbf{F}^{-1})\mathbf{G}_n\frac{\partial\,\mathrm{vech}\,\mathbf{F}}{\partial\mathbf{x}'}$$

$$\mathbf{y} = \mathbf{Az}$$

A y **z** son funciones de **x**

$$\frac{\partial\mathbf{y}}{\partial\mathbf{x}'} = (\mathbf{z}'\otimes\mathbf{I}_m)\frac{\partial\,\mathrm{vec}\,\mathbf{A}}{\partial\mathbf{x}'} + \mathbf{A}\frac{\partial\mathbf{z}}{\partial\mathbf{x}'}$$

  - Potencias de F

$$\frac{\partial\,\mathrm{vec}\,(\mathbf{F}^k)}{\partial\mathbf{x}'} = \sum_{i=1}^{k}[(\mathbf{F}^{i-1})'\otimes\mathbf{F}^{k-i}]\frac{\partial\,\mathrm{vec}\,\mathbf{F}}{\partial\mathbf{x}'} \qquad (\mathbf{F}^0=\mathbf{I}_n)$$

  - Producto de Hadamard

$$\frac{\partial\,\mathrm{vec}\,(\mathbf{F}\circ\mathbf{G})}{\partial\mathbf{x}'} = \mathbf{D}(\mathbf{F})\frac{\partial\,\mathrm{vec}\,\mathbf{G}}{\partial\mathbf{x}'} + \mathbf{D}(\mathbf{G})\frac{\partial\,\mathrm{vec}\,\mathbf{F}}{\partial\mathbf{x}'}$$

$$\mathbf{D}(\mathbf{A}) = \mathrm{diag}(a_{11},a_{12},\ldots,a_{1n},a_{21},a_{22},\ldots,a_{2n},\ldots,a_{m1},a_{m2},\ldots,a_{mn})$$

  - Producto de Kronecker

$$\frac{\partial(\mathrm{vec}\,\mathbf{F}\otimes\mathrm{vec}\,\mathbf{G})}{\partial\mathbf{x}'} = \left(\mathrm{vec}\,\mathbf{F}\otimes\frac{\partial\mathrm{vec}\,\mathbf{G}}{\partial\mathbf{x}'}\right) + \left(\frac{\partial\mathrm{vec}\,\mathbf{F}}{\partial\mathbf{x}'}\otimes\mathrm{vec}\,\mathbf{G}\right)$$

$$\frac{\partial\,\mathrm{vec}\,(\mathbf{F}\otimes\mathbf{G})}{\partial\mathbf{x}'} = (\mathbf{I}_n\otimes\mathbf{I}_{(m,q)}\otimes\mathbf{I}_p)\frac{\partial(\mathrm{vec}\,\mathbf{F}\otimes\mathrm{vec}\,\mathbf{G})}{\partial\mathbf{x}'}$$

Derivada matricial de funciones escalares I

- La derivada de una función escalar $y = f(\mathbf{X})$ con $\mathbf{X} \in \mathbb{R}^{m\times n}$ es una matriz con elementos $\partial y/\partial x_{ij}$

Derivada matricial de una función escalar

$$\frac{\partial f(\mathbf{X})}{\partial \mathbf{X}} = \left(\frac{\partial f(\mathbf{X})}{\partial x_{ij}}\right)$$

Regla de la cadena

$$y = f(\mathbf{X}) \qquad z = g(y) \qquad \frac{\partial z}{\partial \mathbf{X}} = \frac{\partial z}{\partial y}\cdot\frac{\partial y}{\partial \mathbf{X}}$$

Ejemplo

$$\mathbf{Y} = \mathbf{F}(\mathbf{X}) \qquad z = g(\mathbf{Y}) \qquad \frac{\partial z}{\partial x_{ij}} = \sum_r\sum_s \frac{\partial z}{\partial y_{rs}}\frac{\partial y_{rs}}{\partial x_{ij}} = \mathrm{trace}\left(\frac{\partial z}{\partial \mathbf{Y}}\cdot\frac{\partial \mathbf{Y}'}{\partial x_{ij}}\right)$$

- Derivada de la traza $\frac{\partial \mathrm{Tr}(F(\mathbf{X}))}{\partial \mathbf{X}} = f(\mathbf{X})^T$

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{X}) = \mathbf{I} \qquad \frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{XA}) = \mathbf{A}^T$$

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{AXB}) = \mathbf{A}^T\mathbf{B}^T \qquad \frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{AX}^T\mathbf{B}) = \mathbf{BA}$$

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{AXBX}) = \mathbf{A}^T\mathbf{X}^T\mathbf{B}^T + \mathbf{B}^T\mathbf{X}^T\mathbf{A}^T$$

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}\left[\mathbf{X}^T\mathbf{BXC}\right] = \mathbf{BXC} + \mathbf{B}^T\mathbf{XC}^T$$

Traza de la inversa de **X**

Si **X** es simétrica

$$\frac{\partial\, \mathrm{trace}\, \mathbf{X}^{-1}}{\partial \mathbf{X}} = -(\mathbf{X}^{-2})' = -2(\mathbf{X}^{-2}) + \mathrm{diag}(\mathbf{X}^{-2})$$

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{AX}^{-1}\mathbf{B}) = -(\mathbf{X}^{-1}\mathbf{BAX}^{-1})^T = -\mathbf{X}^{-T}\mathbf{A}^T\mathbf{B}^T\mathbf{X}^{-T}$$

Traza de potencias de **X**

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{X}^k) = k(\mathbf{X}^{k-1})^T \qquad \frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{AX}^k) = \sum_{r=0}^{k-1}(\mathbf{X}^r\mathbf{AX}^{k-r-1})^T$$

Traza de exponenciales de **X**

$$\frac{\partial\, \mathrm{trace}\, e^{\mathbf{X}}}{\partial \mathbf{X}} = (e^{\mathbf{X}})' \qquad \frac{\partial e^{\mathrm{trace}(\mathbf{X}^2)}}{\partial \mathbf{X}} = 2e^{\mathrm{trace}(\mathbf{X}^2)}\mathbf{X}'$$

Derivada matricial de funciones escalares II

- Derivada matricial del determinante

$$\frac{\partial \det \mathbf{X}}{\partial \mathbf{X}} = \begin{cases} (\operatorname{adj}\mathbf{X})' = (\det \mathbf{X})(\mathbf{X}^{-1})', & \mathbf{X} \text{ unconstrained}, \\ \det(\mathbf{X})[2\mathbf{X}^{-1} - \operatorname{diag}(\mathbf{X}^{-1})], & \mathbf{X} \text{ symmetric}. \end{cases}$$

$$\frac{\partial \det(\mathbf{X}'\mathbf{A}\mathbf{X})}{\partial \mathbf{X}} = \det(\mathbf{X}'\mathbf{A}\mathbf{X})\{\mathbf{A}\mathbf{X}(\mathbf{X}'\mathbf{A}\mathbf{X})^{-1} + \mathbf{A}'\mathbf{X}[(\mathbf{X}'\mathbf{A}\mathbf{X})^{-1}]'\}$$

$$= 2[\det(\mathbf{X}'\mathbf{X})]\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = 2\det(\mathbf{X}'\mathbf{X})\,\mathbf{X}^{+}$$

Si $\mathbf{A} = \mathbf{I}_n$

Inversa de Moore-Penrose de $\mathbf{X}$

$\mathbf{C} = \mathbf{B}(\mathbf{A}\mathbf{X}\mathbf{B})^{-1}\mathbf{A}$

$$\frac{\partial \det(\mathbf{A}\mathbf{X}\mathbf{B})}{\partial \mathbf{X}} = \begin{cases} \det(\mathbf{A}\mathbf{X}\mathbf{B})\,\mathbf{C}', & \mathbf{X} \text{ unconstrained}, \\ \det(\mathbf{A}\mathbf{X}\mathbf{B})[\mathbf{C} + \mathbf{C}' - \operatorname{diag}\mathbf{C}], & \mathbf{X} \text{ symmetric}, \end{cases}$$

$\mathbf{C} = [\mathbf{X}^{-1}\mathbf{B}(\mathbf{A}\mathbf{X}^{-1}\mathbf{B})^{-1}\mathbf{A}\mathbf{X}^{-1}]$

$$\frac{\partial \det(\mathbf{A}\mathbf{X}^{-1}\mathbf{B})}{\partial \mathbf{X}} = \begin{cases} -\det(\mathbf{A}\mathbf{X}^{-1}\mathbf{B})\mathbf{C}', & \mathbf{X} \text{ unconstrained}, \\ -\det(\mathbf{A}\mathbf{X}^{-1}\mathbf{B})[\mathbf{C} - \mathbf{C}' - \operatorname{diag}\mathbf{C}], & \mathbf{X} \text{ symmetric}, \end{cases}$$

$$\boxed{\frac{\partial \det(\mathbf{X}\mathbf{X}')}{\partial \mathbf{X}} = 2[\det(\mathbf{X}\mathbf{X}')](\mathbf{X}\mathbf{X}')^{-1}\mathbf{X}}$$

Si $\mathbf{A} = \mathbf{I}_n$ y $\mathbf{X} = \mathbf{X}^T$

- Derivada matricial del logdet

det(**F**) >0

$$\frac{\partial \log \det \mathbf{F}}{\partial \mathbf{X}} = (\det \mathbf{F})^{-1} \frac{\partial \det \mathbf{F}}{\partial \mathbf{X}}$$

$$\frac{\partial \log(\det \mathbf{X})}{\partial \mathbf{X}} = (\det \mathbf{X})^{-1} \frac{\partial \det \mathbf{X}}{\partial \mathbf{X}} = \begin{cases} (\mathbf{X}^{-1})', & \mathbf{X} \text{ unconstrained}, \\ 2\mathbf{X}^{-1} - \operatorname{diag}(\mathbf{X}^{-1}), & \mathbf{X} \text{ symmetric}. \end{cases}$$

Derivada matricial de funciones escalares II: ejemplo I

- Consideramos muestras Gaussianas $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ cuya PDF viene dada por

$$f(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \det(2\pi\boldsymbol{\Sigma})^{-\frac{1}{2}} \exp(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}))$$

- Queremos encontrar la matriz $\boldsymbol{\Sigma} \in \mathbb{R}^{n\times n}$ que mejor describa nuestras observaciones
  - Calculamos la función de *"log-likelihood" asociada a* $f(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$

$$\mathcal{L}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \log(f(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})) = -\frac{1}{2}\log\det(2\pi\boldsymbol{\Sigma}) - \frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})$$

Recordatorio

$$\frac{\partial \log(\det \mathbf{X})}{\partial \mathbf{X}} = \begin{cases} (\mathbf{X}^{-1})', & \mathbf{X} \text{ uncon} \\ 2\mathbf{X}^{-1} - \operatorname{diag}(\mathbf{X}^{-1}), & \mathbf{X} \text{ symm} \end{cases}$$

$$\frac{\partial \operatorname{trace}(\mathbf{A}\mathbf{X}^{-1}\mathbf{B})}{\partial \mathbf{X}} = -(\mathbf{X}^{-1}\mathbf{B}\mathbf{A}\mathbf{X}^{-1})'$$

  - Derivamos con respecto a la matriz $\boldsymbol{\Sigma}$

$$\frac{\partial\mathcal{L}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})}{\partial\boldsymbol{\Sigma}} = -\pi n\frac{\partial(\log\det(\boldsymbol{\Sigma}) + (\mathbf{x}-\boldsymbol{\mu})^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}))}{\partial\boldsymbol{\Sigma}} = \pi n\left(\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})^{\mathsf{T}}(\mathbf{x}-\boldsymbol{\mu})\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\right)$$

Derivada matricial de funciones escalares II: ejemplo II

- Consideramos que queremos encontrar el valor de **A** de la siguiente función

$$f(\mathbf{A}) = \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2$$ “Least squares problem”

- Queremos encontrar la matriz **A** que mejor se adapte a nuestras observaciones en **y** dado **x**

Recordatorio

$$\frac{\partial}{\partial \mathbf{X}} \mathrm{Tr}(\mathbf{AXB}) = \mathbf{A}^T\mathbf{B}^T$$

$$\frac{\partial}{\partial \mathbf{X}} \mathrm{Tr}(\mathbf{AX}^T\mathbf{B}) = \mathbf{BA}$$

  - Derivamos $f(\mathbf{A})$ con respecto a la matriz **A**

$$\frac{\partial f(\mathbf{A})}{\partial \mathbf{A}} = \frac{\partial \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2}{\partial \mathbf{A}} = \frac{\partial[(\mathbf{y} - \mathbf{A}\mathbf{x})^\top(\mathbf{y} - \mathbf{A}\mathbf{x})]}{\partial \mathbf{A}} = \frac{\partial(\mathbf{y}^\top\mathbf{y} - \mathbf{y}^\top\mathbf{A}\mathbf{x} - \mathbf{x}^\top\mathbf{A}^\top\mathbf{y} - \mathbf{x}^\top\mathbf{A}^\top\mathbf{A}\mathbf{x})}{\partial \mathbf{A}}$$

$$= -\mathbf{y}\mathbf{x}^\top - \mathbf{y}\mathbf{x}^\top + \mathbf{A}(\mathbf{x}\mathbf{x}^\top)^\top + \mathbf{A}\mathbf{x}\mathbf{x}^\top = 0 \Rightarrow \mathbf{A} = \mathbf{y}\mathbf{x}^\top(\mathbf{x}\mathbf{x}^\top)^{-1}$$

Derivada matricial de funciones escalares III

$$\mathbf{X} \in \mathbb{R}^{m\times n} \qquad \mathbf{E}_{ij} \in \mathbb{R}^{m\times n}$$

$$\frac{\partial(\mathbf{AXB})_{rs}}{\partial\mathbf{X}} = \mathbf{A}'\mathbf{E}_{rs}\mathbf{B}'$$

$$\frac{\partial(\mathbf{AX'B})_{rs}}{\partial\mathbf{X}} = \mathbf{BE}'_{rs}\mathbf{A}$$

$$\frac{\partial(\mathbf{AX^{-1}B})_{rs}}{\partial\mathbf{X}} = -(\mathbf{X}^{-1})'\mathbf{A}'\mathbf{E}_{rs}\mathbf{B}'(\mathbf{X}^{-1})'$$

$$\frac{\partial(\mathbf{X'AX})_{rs}}{\partial\mathbf{X}} = \mathbf{AXE}'_{rs} + \mathbf{A}'\mathbf{XE}_{rs}$$

$$\frac{\partial(\mathbf{XAX})_{rs}}{\partial\mathbf{X}} = \mathbf{E}_{rs}\mathbf{X}'\mathbf{A}' + \mathbf{A}'\mathbf{X}'\mathbf{E}_{rs}$$

$$\frac{\partial(\mathbf{X}^k)_{rs}}{\partial\mathbf{X}} = \sum_{j=0}^{k-1}(\mathbf{X}')^j\mathbf{E}_{rs}(\mathbf{X}')^{n-j-1}$$

$$\frac{\partial(\mathbf{X'AX'})_{rs}}{\partial\mathbf{X}} = \mathbf{AX'E}'_{rs} + \mathbf{E}'_{rs}\mathbf{X}'\mathbf{A}'$$

- Derivada de los autovalores de **X**

$$\mathbf{Xu} = \lambda\mathbf{u}$$
$$\mathbf{vX} = \lambda\mathbf{v}$$

Autovalor no repetido

$$\frac{\partial\lambda}{\partial\mathbf{X}} = \mathbf{v}(\mathbf{v}'\mathbf{u})^{-1}\mathbf{u}'$$

Si **X** es simétrica, sus autovectores están normalizados y coinciden ($\boldsymbol{u}^T\boldsymbol{u} = 1$)

$$\frac{\partial\lambda}{\partial\mathbf{X}} = \mathbf{u}_0\mathbf{u}'_0$$

Reglas de transformación

- Las reglas de transformación nos permiten utilizar los resultados de un tipo de diferenciación para obtener otro tipo de resultados.
  - **Y** es una función de $\mathbf{X} \in \mathbb{R}^{m \times n}$
  - Se aplican para cualquier matrices $\mathbf{A}_t, \mathbf{B}_t, \mathbf{C}_v$ y $\mathbf{D}_v$
  - Incluidas las funciones de **X** (son distintas formas de escribir $\partial y_{rs}/\partial x_{ij}$)
  - Si a la hora de derivar obtenemos algo como en (1) o (2), podemos obtener inmediatamente (3)
  - En algunos casos puede ser complicado obtener (3) directamente

$$\frac{\partial y_{rs}}{\partial \mathbf{X}} = \sum_t \mathbf{A}_t' \mathbf{E}_{rs} \mathbf{B}_t' + \sum_v \mathbf{C}_v \mathbf{E}_{rs}' \mathbf{D}_v \qquad \frac{\partial \mathbf{Y}}{\partial x_{ij}} = \sum_t \mathbf{A}_t \mathbf{E}_{ij} \mathbf{B}_t + \sum_v \mathbf{D}_v \mathbf{E}_{ij}' \mathbf{C}_v \qquad \frac{\partial \operatorname{vec} \mathbf{Y}}{\partial (\operatorname{vec} \mathbf{X})'} = \sum_t \mathbf{B}_t' \otimes \mathbf{A}_t + \sum_v \mathbf{I}_{(m,n)} (\mathbf{D}_v \otimes \mathbf{C}_v')$$

Derivada matricial de funciones matriciales

- La derivada de una función matricial **Y** = **F**(**X**) respecto a una matriz

$$\mathbf{Y} \in \mathbb{R}^{p\times q} \qquad \mathbf{X} \in \mathbb{R}^{m\times n}$$

- Método 1:
  - Derivar cada componente de **Y** con respecto a **X**

$$\frac{\mathrm{d}\mathbf{Y}}{\mathrm{d}\mathbf{X}} = \begin{pmatrix} \frac{\partial y_{11}}{\partial \mathbf{X}} & \frac{\partial y_{12}}{\partial \mathbf{X}} & \cdots & \frac{\partial y_{1q}}{\partial \mathbf{X}} \\ \frac{\partial y_{21}}{\partial \mathbf{X}} & \frac{\partial y_{22}}{\partial \mathbf{X}} & \cdots & \frac{\partial y_{2q}}{\partial \mathbf{X}} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial y_{p1}}{\partial \mathbf{X}} & \frac{\partial y_{p2}}{\partial \mathbf{X}} & \cdots & \frac{\partial y_{pq}}{\partial \mathbf{X}} \end{pmatrix} = \mathbf{Y} \otimes \frac{\partial}{\partial \mathbf{X}}$$

- Método 2:
  - Derivar **Y** con respecto a cada componente de **X**

$$\frac{\mathrm{d}\mathbf{Y}}{\mathrm{d}\mathbf{X}} = \begin{pmatrix} \frac{\partial \mathbf{Y}}{\partial x_{11}} & \frac{\partial \mathbf{Y}}{\partial x_{12}} & \cdots & \frac{\partial \mathbf{Y}}{\partial x_{1n}} \\ \frac{\partial \mathbf{Y}}{\partial x_{21}} & \frac{\partial \mathbf{Y}}{\partial x_{22}} & \cdots & \frac{\partial \mathbf{Y}}{\partial x_{2n}} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial \mathbf{Y}}{\partial x_{m1}} & \frac{\partial \mathbf{Y}}{\partial x_{m2}} & \cdots & \frac{\partial \mathbf{Y}}{\partial x_{mn}} \end{pmatrix} = \frac{\partial}{\partial \mathbf{X}} \otimes \mathbf{Y}$$

Diferenciales de matrices

- Definimos d**X** como la matriz de diferenciales de la matriz **X** → $\mathrm{d}\mathbf{X} = (\mathrm{d}x_{ij})$
  - Es otro método muy utilizado para el cálculo de derivadas

$$y = f(\mathbf{x}) \qquad \mathbf{x} = (x_1, x_2, \ldots, x_n)'$$

$$\mathrm{d}y = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}\mathrm{d}x_i = \sum_{i=1}^{n} \frac{\partial y}{\partial x_i}\mathrm{d}x_i$$

- Propiedades básicas

$$\mathrm{d}(\mathbf{A}\mathbf{X}) = \mathbf{A}\mathrm{d}\mathbf{X} \qquad \mathrm{d}(\mathbf{X} \pm \mathbf{Y}) = \mathrm{d}\mathbf{X} \pm \mathrm{d}\mathbf{Y} \qquad \mathrm{d}(\mathbf{X}') = (\mathrm{d}\mathbf{X})' \qquad \mathrm{d}\,\mathrm{vec}\,\mathbf{X} = \mathrm{vec}\,\mathrm{d}\mathbf{X} \qquad \mathrm{d}\,\mathrm{vec}\,\mathbf{X}' = \mathrm{vec}\,(\mathrm{d}\mathbf{X}') = \mathbf{I}_{(n,m)}\mathrm{vec}\,(\mathrm{d}\mathbf{X})$$

  - Traza

$$\mathrm{d}(\mathrm{trace}\,\mathbf{X}) = \mathrm{trace}(\mathrm{d}\mathbf{X}) = \mathrm{trace}(\mathbf{I}_n\mathrm{d}\mathbf{X}) = \mathrm{vec}\,(\mathbf{I}_n)'\mathrm{d}(\mathrm{vec}\,\mathbf{X})$$

  - Determinante

$$\mathrm{d}(\det\mathbf{X}) = (\det\mathbf{X})\,\mathrm{trace}(\mathbf{X}^{-1}\mathrm{d}\mathbf{X})$$

  - Producto de Kronecker

$$\mathrm{d}(\mathbf{X} \otimes \mathbf{Y}) = (\mathrm{d}\mathbf{X}) \otimes \mathbf{Y} + \mathbf{X} \otimes \mathrm{d}\mathbf{Y}$$

  - Producto de Hadamard

$$\mathrm{d}(\mathbf{X} \circ \mathbf{Y}) = (\mathrm{d}\mathbf{X}) \circ \mathbf{Y} + \mathbf{X} \circ \mathrm{d}\mathbf{Y}$$

  - Inversa

$$\mathrm{d}\mathbf{X}^{-1} = -\mathbf{X}^{-1}(\mathrm{d}\mathbf{X})\mathbf{X}^{-1}$$

Diferenciales de matrices

A vector transformation rule
A matrix transformation rule…

- *A scalar transformation rule*
  - Si y es una función escalar de $\mathbf{X}$ entonces $\mathrm{d}y = \mathrm{trace}(\mathbf{A}'\mathrm{d}\mathbf{X})$ “iff“ $\frac{\partial f}{\partial \mathbf{X}} = \mathbf{A}$

- Ejemplo de la derivada de la traza

$$y = \mathrm{trace}(\mathbf{XAX}'\mathbf{B})$$

$$\begin{aligned} \mathrm{d}y &= \mathrm{trace}[\mathrm{d}(\mathbf{XAX}'\mathbf{B})] = \mathrm{trace}[(\mathrm{d}\mathbf{X})\mathbf{AX}'\mathbf{B}] + \mathrm{trace}[\mathbf{XA}(\mathrm{d}\mathbf{X})'\mathbf{B}] \\ &= \mathrm{trace}[(\mathbf{AX}'\mathbf{B} + \mathbf{A}'\mathbf{X}'\mathbf{B}')\mathrm{d}\mathbf{X}] \end{aligned}$$

$$\frac{\partial y}{\partial \mathbf{X}} = (\mathbf{AX}'\mathbf{B} + \mathbf{A}'\mathbf{X}'\mathbf{B}')'$$

- Ejemplo de la derivada del determinante

$$y = \det(\mathbf{AXB}) \qquad \mathbf{Y} = \mathbf{AXB}$$

$$\begin{aligned} \mathrm{d}(\det \mathbf{Y}) &= \det \mathbf{Y}\, \mathrm{trace}(\mathbf{Y}^{-1}\mathrm{d}\mathbf{Y}) \\ &= \det \mathbf{Y}\, \mathrm{trace}[\mathbf{Y}^{-1}\mathbf{A}(\mathrm{d}\mathbf{X})\mathbf{B}] \\ &= \det \mathbf{Y}\, \mathrm{trace}[\mathbf{B}(\mathbf{AXB})^{-1}\mathbf{A}\mathrm{d}\mathbf{X}] \\ &= \det \mathbf{Y}\, \mathrm{trace}(\mathbf{C}\mathrm{d}\mathbf{X}), \quad \text{say}, \end{aligned}$$

$$\frac{\partial y}{\partial \mathbf{X}} = (\det \mathbf{Y})\mathbf{C}' \qquad \mathbf{C} = \mathbf{B}(\mathbf{AXB})^{-1}\mathbf{A}$$

Derivadas de normas de vectores/matrices

- Derivada de la norma l_2 de un vector

$$\frac{\partial}{\partial \mathbf{x}}\|\mathbf{x}-\mathbf{a}\|_2 = \frac{\mathbf{x}-\mathbf{a}}{\|\mathbf{x}-\mathbf{a}\|_2} \qquad \frac{\partial}{\partial \mathbf{x}}\frac{\mathbf{x}-\mathbf{a}}{\|\mathbf{x}-\mathbf{a}\|_2} = \frac{\mathbf{I}}{\|\mathbf{x}-\mathbf{a}\|_2} - \frac{(\mathbf{x}-\mathbf{a})(\mathbf{x}-\mathbf{a})^T}{\|\mathbf{x}-\mathbf{a}\|_2^3} \qquad \frac{\partial\|\mathbf{x}\|_2^2}{\partial \mathbf{x}} = \frac{\partial\|\mathbf{x}^T\mathbf{x}\|_2}{\partial \mathbf{x}} = 2\mathbf{x}$$

- Derivada de la norma Frobenius de una matriz

$$\frac{\partial}{\partial \mathbf{X}}\|\mathbf{X}\|_F^2 = \frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{X}\mathbf{X}^H) = 2\mathbf{X}$$

- Derivada de la norma nuclear de una matriz

$$\|\mathbf{X}\|_* = \mathrm{trace}(\sqrt{\mathbf{X}^*\mathbf{X}}) \qquad \mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\mathsf{T}$$

$$\frac{\partial \mathrm{tr}(\mathbf{\Sigma})}{\partial \mathbf{X}} = \frac{\mathrm{tr}(\partial \mathbf{\Sigma})}{\partial \mathbf{X}} = \cdots = \mathbf{U}\mathbf{V}^\mathsf{T}$$

- Derivada de la norma l_1

The differential of the Holder 1-norm (h) of a matrix (Y) is

$$dh = \mathrm{sign}(Y) : dY$$

where the sign function is applied element-wise and the colon represents the Frobenius product.

Now substitute $Y = (X - AX)$

$$\begin{aligned} dY &= (I - A)\,dX \\ dh &= \mathrm{sign}(Y) : (I - A)\,dX \\ &= (I - A)^T \mathrm{sign}(Y) : dX \\ &= (I - A)^T \mathrm{sign}(X - AX) : dX \end{aligned}$$

Since $dh = \left(\frac{\partial h}{\partial X} : dX\right)$, the gradient must be

$$\frac{\partial h}{\partial X} = (I - A)^T \mathrm{sign}(X - AX)$$

The result is unchanged if the matrices $\{X, Y\}$ are replaced by vectors $\{x, y\}$.

Derivadas de matrices con estructura

- Se aprovecha la estructura de las matrices para calcular sus derivadas

$$\frac{df}{dA_{ij}} = \sum_{kl} \frac{\partial f}{\partial A_{kl}} \frac{\partial A_{kl}}{\partial A_{ij}} = \mathrm{Tr}\left[\left[\frac{\partial f}{\partial \mathbf{A}}\right]^T \frac{\partial \mathbf{A}}{\partial A_{ij}}\right] \qquad \frac{\partial \mathbf{A}}{\partial A_{ij}} = \mathbf{S}^{ij}$$

Estructura de **A**

- Derivada de matrices simétricas

$$\mathbf{S}^{ij} = \mathbf{J}^{ij} + \mathbf{J}^{ji} - \mathbf{J}^{ij}\mathbf{J}^{ij}$$

$$\frac{df}{d\mathbf{A}} = \left[\frac{\partial f}{\partial \mathbf{A}}\right] + \left[\frac{\partial f}{\partial \mathbf{A}}\right]^T - \mathrm{diag}\left[\frac{\partial f}{\partial \mathbf{A}}\right]$$

$$\frac{\partial \mathrm{Tr}(\mathbf{AX})}{\partial \mathbf{X}} = \mathbf{A} + \mathbf{A}^T - (\mathbf{A} \circ \mathbf{I})$$

$$\frac{\partial \mathrm{Tr}(\mathbf{AX})}{\partial \mathbf{X}} = \mathbf{A} \circ \mathbf{I}$$

Si **X** es diagonal

$$\frac{\partial \det(\mathbf{X})}{\partial \mathbf{X}} = \det(\mathbf{X})(2\mathbf{X}^{-1} - (\mathbf{X}^{-1} \circ \mathbf{I}))$$

$$\frac{\partial \ln\det(\mathbf{X})}{\partial \mathbf{X}} = 2\mathbf{X}^{-1} - (\mathbf{X}^{-1} \circ \mathbf{I})$$

- Derivada de matrices Toeplitz

$$\frac{\partial \mathrm{Tr}(\mathbf{AT})}{\partial \mathbf{T}} = \frac{\partial \mathrm{Tr}(\mathbf{TA})}{\partial \mathbf{T}} = \boldsymbol{\alpha}(\mathbf{A}) = \begin{bmatrix} \mathrm{Tr}(\mathbf{A}) & \mathrm{Tr}([\mathbf{A}^T]_{n1}) & \mathrm{Tr}([[\mathbf{A}^T]_{1n}]_{n-1,2}) & \cdots & A_{n1} \\ \mathrm{Tr}([\mathbf{A}^T]_{1n})) & \mathrm{Tr}(\mathbf{A}) & \ddots & \ddots & \vdots \\ \mathrm{Tr}([[\mathbf{A}^T]_{1n}]_{2,n-1}) & \ddots & \ddots & \ddots & \mathrm{Tr}([[\mathbf{A}^T]_{1n}]_{n-1,2}) \\ \vdots & \ddots & \ddots & \ddots & \mathrm{Tr}([\mathbf{A}^T]_{n1}) \\ A_{1n} & \cdots & \mathrm{Tr}([[\mathbf{A}^T]_{1n}]_{2,n-1}) & \mathrm{Tr}([\mathbf{A}^T]_{1n})) & \mathrm{Tr}(\mathbf{A}) \end{bmatrix}$$

Perturbaciones usando diferenciales

- Encontrar la expansión de Taylor de una función de **X**+d**X**

Función vectorial de **X** e **Y**

$$\mathbf{f}(\mathbf{X}+\epsilon\mathbf{Y}) = \mathbf{f}(\mathbf{X}) + \sum_{i=1}^{\infty} \epsilon^i \mathbf{g}_i(\mathbf{X},\mathbf{Y})$$

Función matricial de **X** e **Y**

$$\mathbf{F}(\mathbf{X}+\epsilon\mathbf{Y}) = \mathbf{F}(\mathbf{X}) + \sum_{i=1}^{\infty} \epsilon^i \mathbf{G}_i(\mathbf{X},\mathbf{Y})$$

- Perturbaciones en la inversa de **X**

$$\begin{aligned}(\mathbf{X}+\mathrm{d}\mathbf{X})^{-1} &= \mathbf{X}^{-1} - \mathbf{X}^{-1}(\mathrm{d}\mathbf{X})\mathbf{X}^{-1} + \mathbf{X}^{-1}(\mathrm{d}\mathbf{X})\mathbf{X}^{-1}(\mathrm{d}\mathbf{X})\mathbf{X}^{-1} \\ &\quad -\mathbf{X}^{-1}(\mathrm{d}\mathbf{X})\mathbf{X}^{-1}(\mathrm{d}\mathbf{X})\mathbf{X}^{-1}(\mathrm{d}\mathbf{X})\mathbf{X}^{-1} + \cdots.\end{aligned}$$

W es pequeño

- Perturbaciones en los autovalores de **X** $\quad \mathbf{X} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T \qquad \mathbf{X}\mathbf{q}_i = \lambda_i \mathbf{q}_i \qquad \mathrm{d}\mathbf{X} = \mathbf{Q}^T\mathbf{W}\mathbf{Q}$

$$\lambda_i(\mathbf{X}+\mathrm{d}\mathbf{X}) = \lambda_i + \mathbf{q}_i'\mathbf{W}\mathbf{q}_i + \cdots.$$

$$\gamma_i(\mathbf{X}+\mathrm{d}\mathbf{X}) = \mathbf{q}_i - (\mathbf{Z}-\lambda_i\mathbf{I}_n)^{+}\mathbf{W}\mathbf{q}_i + \cdots.$$

Ecuaciones diferenciales lineales

- Definición
  - $\boldsymbol{x} = \boldsymbol{x}(t)$ es un vector cuyos elementos son funciones de t y $\boldsymbol{A}$ es una matriz constante

$$\frac{\partial \mathbf{x}(t)}{\partial t} = \mathbf{A}\mathbf{x}(t), \qquad \mathbf{x}(0) = \mathbf{x}_0 \qquad \mathbf{x} = e^{\mathbf{A}t}\,\mathbf{x}_0$$

- Ejemplo $\mathbf{A} = \mathbf{S}\mathbf{J}_O\mathbf{S}^{-1}$ siendo $\mathbf{J}_O$ la forma canónica de Jordan de la matriz $\boldsymbol{A}$

$$\frac{\partial \mathbf{y}(t)}{\partial t} = \mathbf{J}_O\mathbf{y}(t), \qquad \mathbf{y}(0) = \mathbf{y}_0 \qquad \mathbf{x}(t) = \mathbf{S}\mathbf{y}(t) \qquad \mathbf{y}_0 = \mathbf{S}^{-1}\mathbf{x}_0$$

  - Si A es diagonalizable

$$\mathbf{J}_O = \operatorname{diag}(\lambda_1, \ldots, \lambda_n) \qquad y_i(t) = y_i(0)e^{\lambda_i t}, \quad i = 1, 2, \ldots, n$$

Derivadas de segundo orden

- Definición de Hessiano
  - Si $f(\boldsymbol{X})$ una función escalar de la matriz $\boldsymbol{X}$

$$\begin{aligned}\nabla^2 f(\mathbf{X}) &= \frac{\partial^2 f(\mathbf{X})}{\partial \operatorname{vec}\mathbf{X}\,\partial(\operatorname{vec}\mathbf{X})'} \\ &= \frac{\partial}{\partial \operatorname{vec}\mathbf{X}}\left(\frac{\partial f(\mathbf{X})}{\partial(\operatorname{vec}\mathbf{X})'}\right)\end{aligned}$$

Si **x** es un vector

$$\nabla^2 f(\mathbf{x}) = \frac{\partial^2 f}{\partial \mathbf{x}\partial \mathbf{x}'} = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}\right)$$

B puede depender de **X** pero no de d**X**

- *Identification Rules* → $\mathrm{d}^2 f(\mathbf{X}) = (\operatorname{vec}\mathrm{d}\mathbf{X})'\mathbf{B}(\operatorname{vec}\mathrm{d}\mathbf{X})$ sii $\nabla^2 f(\mathbf{X}) = \frac{1}{2}(\mathbf{B}+\mathbf{B}')$

- Ejemplo:

$$f(\mathbf{X}) = \operatorname{trace}(\mathbf{AXBX'C})$$

$$\mathrm{d}f(\mathbf{X}) = \operatorname{trace}[\mathbf{A}(\mathrm{d}\mathbf{X})\mathbf{BX'C} + \mathbf{AXB}(\mathrm{d}\mathbf{X})'\mathbf{C}]$$

$$\boxed{\begin{array}{c}\mathrm{d}^2\mathbf{X} = \mathbf{0} \quad \mathrm{d}(\mathrm{d}\mathbf{X})' = \mathbf{0} \\ \mathbf{F} = \operatorname{trace}\mathbf{F}'\end{array}}$$

$$\begin{aligned}\mathrm{d}^2 f(\mathbf{X}) &= 2\operatorname{trace}[\mathbf{A}(\mathrm{d}\mathbf{X})\mathbf{B}(\mathrm{d}\mathbf{X})'\mathbf{C}] \\ &= 2\operatorname{trace}[(\mathrm{d}\mathbf{X})'\mathbf{CA}(\mathrm{d}\mathbf{X})\mathbf{B} \\ &= 2(\operatorname{vec}\mathrm{d}\mathbf{X})'(\mathbf{B}'\otimes\mathbf{CA})(\operatorname{vec}\mathrm{d}\mathbf{X})\end{aligned}$$

$$\mathrm{d}^2 f(\mathbf{X}) = \operatorname{trace}[\mathbf{A}(\mathrm{d}\mathbf{X})\mathbf{B}(\mathrm{d}\mathbf{X})'\mathbf{C}]$$

sii

$$\nabla^2 f(\mathbf{X}) = \frac{1}{2}(\mathbf{B}'\otimes\mathbf{CA} + \mathbf{B}\otimes\mathbf{A}'\mathbf{C}')$$

Ecuaciones en diferencia de vectores

- Definición de ecuaciones en diferencia de vectores

$$\mathbf{A}_0\mathbf{y}_t + \mathbf{A}_1\mathbf{y}_{t-1} + \ldots + \mathbf{A}_r\mathbf{y}_{t-r} = \mathbf{g}(t)$$

- Reducción a ecuación de primer orden

$$\mathbf{z}_t = (\mathbf{y}_t', \mathbf{y}_{t-1}', \ldots, \mathbf{y}_{t-r+1}')'$$

$$\mathbf{z}_t = \mathbf{B}\mathbf{z}_{t-1} + \mathbf{e}_1 \otimes \mathbf{g}(t)$$

$$\mathbf{B} = \begin{pmatrix} -\mathbf{A}_1 & -\mathbf{A}_2 & \ldots & -\mathbf{A}_{r-1} & -\mathbf{A}_r \\ \mathbf{I}_d & \mathbf{0} & \ldots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ldots & \mathbf{I}_d & \mathbf{0} \end{pmatrix}$$

- Algunos detalles

$$\mathbf{x}_t = \mathbf{A}\mathbf{x}_{t-1} + \mathbf{d}$$

$\mathbf{A}$ y $\mathbf{x}_0$ *son conocidos*

⇨

$\mathbf{A}^t \rightarrow \mathbf{0}$ si $t \rightarrow \infty$

$(\mathbf{I} - \mathbf{A})^{-1}$ existe

⇨

$$\mathbf{x}_t = \mathbf{A}^t\mathbf{x}_0 + (\mathbf{I} - \mathbf{A}^t)(\mathbf{I} - \mathbf{A})^{-1}\mathbf{d}$$

Si $t \rightarrow \infty$ → $\mathbf{x}_t \rightarrow (\mathbf{I} - \mathbf{A})^{-1}\mathbf{d}$

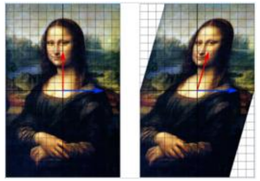
Recapitulación de conceptos

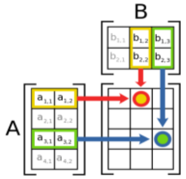
- Xxxxx1
  - Yyyyy 1.1
  - Yyyyy 1.2
  - …
- Xxxxx2
  - Yyyyy 2.1
  - Yyyyy 2.2
  - …
- Xxxxx3
  - Yyyyy 3.1
  - Yyyyy 3.2
  - …

Cálculo con matrices II

"Vectores, matrices y tensores: Una perspectiva desde ciencia e ingeniería de datos"

Reading Group organizado por el "Grupo de Investigación de Alto Rendimiento en Ciencia de Datos y Procesamiento de Señal de la URJC"





Esta presentación utiliza material cuya autoría corresponde a: Marques, Buciulea, Wikipedia; así como contribuciones menores de distintas autorías.

Índice de la clase

Recordatorio

- **La derivada** de f en x es el límite del valor del cociente diferencial, conforme las líneas secantes se aproximan a la línea tangente

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

- **El gradiente** es una generalización de la derivada. Si $f : \mathbb{R}^n \to \mathbb{R}$ entonces el gradiente de f en $\mathbf{r}$ se define como una función vectorial $\nabla f : \mathbb{R}^n \to \mathbb{R}^n$

$$\nabla f(\mathbf{r}) = \left(\frac{\partial f(\mathbf{r})}{\partial x_1}, \ldots, \frac{\partial f(\mathbf{r})}{\partial x_n} \right)$$

- **La matriz Jacobiana** de una función vectorial de varias variables es la matriz cuyos elementos son las derivadas parciales de primer orden de dicha función

$$\mathbf{J} = \begin{bmatrix} \dfrac{\partial \mathbf{F}}{\partial x_1} & \cdots & \dfrac{\partial \mathbf{F}}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \nabla^T f_1 \\ \vdots \\ \nabla^T f_m \end{bmatrix} = \begin{bmatrix} \dfrac{\partial f_1}{\partial x_1} & \cdots & \dfrac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial f_m}{\partial x_1} & \cdots & \dfrac{\partial f_m}{\partial x_n} \end{bmatrix}$$

- **La matriz hessiana** de una función escalar de varias variables es la matriz cuyos elementos son las derivadas parciales de segundo orden de dicha función

$$(H_f)_{i,j} = \frac{\partial^2 f}{\partial x_i \, \partial x_j} \qquad H_f = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x_1^2} & \dfrac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \dfrac{\partial^2 f}{\partial x_1 \partial x_n} \\ \dfrac{\partial^2 f}{\partial x_2 \partial x_1} & \dfrac{\partial^2 f}{\partial x_2^2} & \cdots & \dfrac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial^2 f}{\partial x_n \partial x_1} & \dfrac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \dfrac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

Introducción I

- Dada una función $\boldsymbol{y} = f(\boldsymbol{x})$ donde $\mathbf{x},\, \mathbf{y} \in \mathbb{R}^n$ diferenciable y con inversa $g = f^{-1}$
  - Matriz Jacobiana de la transformación $\mathbf{x} \rightarrow \mathbf{y}$

$$\mathbf{J} = \frac{\partial \mathbf{x}}{\partial \mathbf{y}^{\mathsf{T}}}$$

  - Determinante de la matriz Jacobiana

$$J_{\mathbf{x}\rightarrow\mathbf{y}} = \det\left(\frac{\partial \mathbf{x}}{\partial \mathbf{y}^{\mathsf{T}}}\right)$$

- Si tenemos acceso a $J_{\mathbf{y}\rightarrow\mathbf{x}}$ podemos calcular $J_{\mathbf{x}\rightarrow\mathbf{y}} = J^{-1}_{\mathbf{y}\rightarrow\mathbf{x}}$
  - Se puede demostrar que si $\mathbf{x} = g(f(\mathbf{x}))$

$$\mathbf{I}_n = \left(\frac{\partial x_i}{\partial x_j}\right) = \left(\sum_r \frac{\partial g_i}{\partial y_r} \cdot \frac{\partial y_r}{\partial x_j}\right) = \frac{\partial \mathbf{x}}{\partial \mathbf{y}'} \cdot \frac{\partial \mathbf{y}}{\partial \mathbf{x}'}$$

  - De manera similar para el determinante

$$1 = \det\left\{\frac{\partial \mathbf{x}}{\partial \mathbf{y}'} \cdot \frac{\partial \mathbf{y}}{\partial \mathbf{x}'}\right\} = \det\left\{\frac{\partial \mathbf{x}}{\partial \mathbf{y}'}\right\} \cdot \det\left\{\frac{\partial \mathbf{y}}{\partial \mathbf{x}'}\right\} = J_{\mathbf{x}\rightarrow\mathbf{y}} J_{\mathbf{y}\rightarrow\mathbf{x}}$$

Mayor interés en el $J_{\mathbf{x}\rightarrow\mathbf{y}} = \det(\mathbf{J})$

$J_{\mathbf{x}\rightarrow\mathbf{y}} \neq 0$	f es invertible
$J_{\mathbf{x}\rightarrow\mathbf{y}} > 0$	f mantiene orientación
$J_{\mathbf{x}\rightarrow\mathbf{y}} < 0$	f invierte orientación
$\lvert J_{\mathbf{x}\rightarrow\mathbf{y}} \rvert$	factor de expans./compres.

Introducción II: Ejemplos

- Aproximación lineal de orden 2 de funciones no lineales $\mathrm{y} = \mathrm{f}(\mathbf{x})\colon\ \mathbb{R}^n \rightarrow \mathbb{R}$

Si n=1 ➔ f es escalar y aplicamos el desarrollo en serie de Taylor clásico

$$f(\mathbf{x}) \approx f(\mathbf{a}) + f'(\mathbf{a})(\mathbf{x}-\mathbf{a}) + \tfrac{1}{2} f''(\mathbf{a})(\mathbf{x}-\mathbf{a})^2$$

Cuando n>1, podemos encontrar la aproximación de segundo orden como

$$f(\mathbf{x}) \approx f(\mathbf{a}) + \begin{bmatrix} \partial f/\partial x_1 & \partial f/\partial x_2 \end{bmatrix}_{\mathbf{a}} (\mathbf{x}-\mathbf{a}) + \frac{1}{2}(\mathbf{x}-\mathbf{a})^T \begin{bmatrix} \partial f/\partial x_1^2 & \partial f/\partial x_1 \partial x_2 \\ \partial f/\partial x_2 \partial x_1 & \partial f/\partial x_2^2 \end{bmatrix}_{\mathbf{a}} (\mathbf{x}-\mathbf{a})$$

- Transformaciones de FDP $f_{\mathbf{y}}(\mathbf{y}) = f_{\mathbf{x}}(g(\mathbf{y}))|J_{\mathbf{x}\rightarrow\mathbf{y}}|$
  - Sea un vector aleatorio $\mathbf{x} = (x_1, x_2, x_3)$ con FDP $f_{\mathbf{x}}(x_1, x_2, x_3) = 6(1 + x_1 + x_2 + x_3)^{-4};\ x_1, x_2, x_3 > 0$
  - Calcular la función de densidad de probabilidad de $x_1 + x_2 + x_3$

Realizamos el cambio de variable

$$\begin{aligned} y_1 &= x_1 + x_2 + x_3 & x_1 &= y_3 \\ y_2 &= x_1 + x_2 & x_2 &= y_2 - y_3 \\ y_3 &= x_1 & x_3 &= y_1 - y_2 \end{aligned}$$

Calculamos el determinante de la matriz Jacobiana

$$J_{\mathbf{x}\rightarrow\mathbf{y}} = \begin{vmatrix} 0 & 0 & 1 \\ 0 & 1 & -1 \\ 1 & -1 & 0 \end{vmatrix} = -1$$

Aplicamos la transformación de PDF

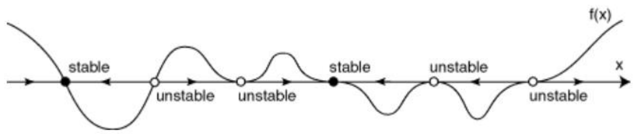
$$f_{\mathbf{y}}(y_1, y_2, y_3) = f_{\mathbf{x}}(y_3, y_2 - y_3, y_1 - y_2)|-1| = 6(1 + y_1)^{-4};\ y_1 > y_2 > y_3 > 0$$

Introducción III: Ejemplos

- Cambio de variable al integrar

$$\int \cdot \int h(x_1, x_2, \ldots, x_n) \mathrm{d}x_1 \mathrm{d}x_2 \cdots \mathrm{d}x_n = \int \cdot \int h(\mathbf{g}(\mathbf{y})) |J_{\mathbf{x} \to \mathbf{y}}| \mathrm{d}y_1 \mathrm{d}y_2 \cdots \mathrm{d}y_n$$

- Análisis de estabilidad en sistemas dinámicos $\dot{\mathbf{x}} = F(\mathbf{x}) \quad F: \mathbb{R}^n \to \mathbb{R}^n$



  - Si $F(\mathbf{x}_0) = 0$ entonces $\mathbf{x}_0$ es un punto estacionario
  - El comportamiento del sistema cerca de $\mathbf{x}_0$ está relacionado con los autovalores del Jacobiano $\lambda_i\{J_F(\mathbf{x}_0)\}$

$\mathcal{R}\{\lambda_i\} < 0, \forall i$	Si la parte real de los autovalores es negativa, el sistema es estable.
$\mathcal{R}\{\lambda_i\} > 0$	Si algún autovalor tiene parte real positiva el sistema es inestable
$\mathcal{R}\{\max(\lambda_i)\} = 0$	Si la mayor parte real de los autovalores es cero, no se puede evaluar la estabilidad

Introducción IV: Ejemplos

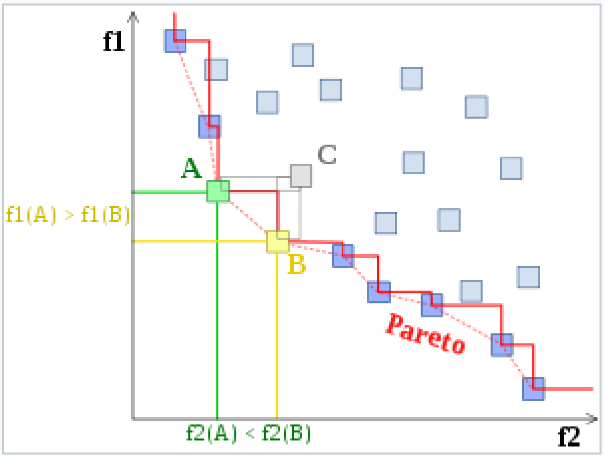
- Optimización multiobjetivo

$$\min_{x \in X} (f_1(x), f_2(x), \ldots, f_k(x))$$

  - X es el set de posibles vectores que engloba el conjunto de restricciones
  - El Jacobiano se emplea para el cálculo de máximos y mínimos
  - Normalmente no existe una solución que minimiza todas las funciones simultáneamente

1. $\forall i \in \{1, \ldots, k\}, f_i(x_1) \leq f_i(x_2)$
2. $\exists i \in \{1, \ldots, k\}, f_i(x_1) < f_i(x_2)$

$x_1 \rightarrow$ solución Pareto dominante

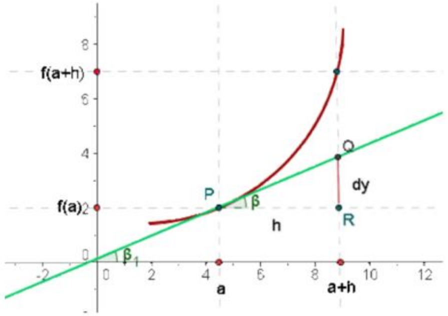


Example of a Pareto frontier (in red), the set of Pareto optimal solutions (those that are not dominated by any other feasible solutions). The boxed points represent feasible choices, and smaller values are preferred to larger ones. Point *C* is not on the Pareto frontier because it is dominated by both point *A* and point *B*. Points *A* and *B* are not strictly dominated by any other, and hence do lie on the frontier.

- Aplicaciones
  - **Ámbito económico:** maximizar las ganancias de diversos bienes sujeto a ingresos disponibles y precios de los productos
  - **Finanzas:** dos objetivos en conflicto → maximización de portfolio y minimización de riesgo
  - **Gestión de recursos Radio:** repartir los recursos(tiempo, frecuencia, bloques) sujeto a restricciones de latencia, tasa, potencia

Cálculo de Jacobianos: Método por diferenciales

- ¿Qué es un diferencial?



- ¿Cómo podemos encontrar los jacobianos mediante los métodos por diferenciales?
  - Si **y** es una función de **x** → **y** = **Ax**

$$d\mathbf{y} = \mathbf{A}d\mathbf{x}$$

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}'} = \mathbf{A}$$

$$J_{\mathbf{y}\to\mathbf{x}} = \det \mathbf{A}$$

  - Si **Y** es una función de **X** → **Y** = **AX**

$$d\,\mathrm{vec}\,\mathbf{Y} = \mathbf{B}\,d\,\mathrm{vec}\,\mathbf{X}$$

$$\frac{\partial\,\mathrm{vec}\,\mathbf{Y}}{\partial(\mathrm{vec}\,\mathbf{X})'} = \mathbf{B}$$

$$J_{\mathbf{Y}\to\mathbf{X}} = \det \mathbf{B}$$

- Con este método se pueden calcular los jacobianos

Queremos calcular el Jacobiano de **Y**

$$\mathbf{Y}_{m\times n} = \mathbf{A}_{m\times m}\mathbf{X}_{m\times n}\mathbf{B}_{n\times n}$$

Aplicamos el método por diferenciales y vectorizamos

$$d\mathbf{Y} = \mathbf{A}(d\mathbf{X})\mathbf{B} \qquad \mathrm{vec}\,(d\mathbf{Y}) = (\mathbf{B}'\otimes\mathbf{A})\mathrm{vec}\,(d\mathbf{X})$$

Jacobiano de transformación $\mathbf{Y}\to\mathbf{X}$

$$J_{\mathbf{Y}\to\mathbf{X}} = \det(\mathbf{B}'\otimes\mathbf{A}) = (\det\mathbf{B})^m(\det\mathbf{A})^n$$

Jacobiano de transformación $\mathbf{X}\to\mathbf{Y}$

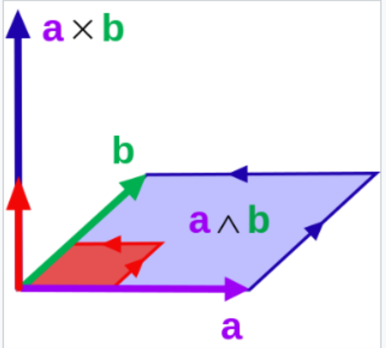
$$J_{\mathbf{X}\to\mathbf{Y}} = (\det\mathbf{B})^{-m}(\det\mathbf{A})^{-n}$$

Cálculo de Jacobianos: Otras técnicas I

- 2. La regla de la cadena
  - Queremos calcular el determinante de la matriz Jacobiana: $J_{\mathbf{x}\to\mathbf{z}} = \det(\mathbf{J})$
  - Donde tenemos varias transformaciones $\mathbf{x} \to \mathbf{y},\ \mathbf{y} \to \mathbf{z}$ entonces $J_{\mathbf{x}\to\mathbf{z}} = J_{\mathbf{x}\to\mathbf{y}} J_{\mathbf{y}\to\mathbf{z}}$
  - Muy útil cuando $J_{\mathbf{x}\to\mathbf{z}}$ es difícil de encontrar
- 3. El producto exterior de diferenciales

$$\mathrm{d}y_1 \wedge \mathrm{d}y_2 \wedge \cdots \wedge \mathrm{d}y_n = \det\left(\frac{\partial \mathbf{y}}{\partial \mathbf{x}'}\right) \mathrm{d}x_1 \wedge \mathrm{d}x_2 \wedge \cdots \wedge \mathrm{d}x_n$$

Recordatorio



The cross product (blue vector) in relation to the exterior product (light blue parallelogram). The length of the cross product is to the length of the parallel unit vector (red) as the size of the exterior product is to the size of the reference parallelogram (light red).

Ejemplo para n=2

$$\begin{aligned}
\mathrm{d}y_1 \wedge \mathrm{d}y_2 &= \left(\frac{\partial y_1}{\partial x_1}\mathrm{d}x_1 + \frac{\partial y_1}{\partial x_2}\mathrm{d}x_2\right) \wedge \left(\frac{\partial y_2}{\partial x_1}\mathrm{d}x_1 + \frac{\partial y_2}{\partial x_2}\mathrm{d}x_2\right) \\
&= \frac{\partial y_1}{\partial x_1}\cdot\frac{\partial y_2}{\partial x_1}\mathrm{d}x_1 \wedge \mathrm{d}x_1 + \frac{\partial y_1}{\partial x_1}\cdot\frac{\partial y_2}{\partial x_2}\mathrm{d}x_1 \wedge \mathrm{d}x_2 \\
&\quad + \frac{\partial y_1}{\partial x_2}\cdot\frac{\partial y_2}{\partial x_1}\mathrm{d}x_2 \wedge \mathrm{d}x_1 + \frac{\partial y_1}{\partial x_2}\cdot\frac{\partial y_2}{\partial x_2}\mathrm{d}x_2 \wedge \mathrm{d}x_2 \\
&= 0 + \left(\frac{\partial y_1}{\partial x_1}\cdot\frac{\partial y_2}{\partial x_2} - \frac{\partial y_1}{\partial x_2}\cdot\frac{\partial y_2}{\partial x_1}\right)\mathrm{d}x_1 \wedge \mathrm{d}x_2 + 0 \\
&= \det\begin{pmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} \end{pmatrix} \mathrm{d}x_1 \wedge \mathrm{d}x_2 \\
&= \det\left(\frac{\partial \mathbf{y}}{\partial \mathbf{x}'}\right)\mathrm{d}x_1 \wedge \mathrm{d}x_2 .
\end{aligned}$$

  - Propiedades del producto exterior de diferenciales

$$\mathrm{d}y_i \wedge \mathrm{d}y_j = -\mathrm{d}y_j \wedge \mathrm{d}y_i \qquad \mathrm{d}y_i \wedge \mathrm{d}y_i = 0$$

  - Ejemplo de cálculo del Jacobiano de la transformación

$$\mathrm{d}_w\mathbf{x} = \wedge_{i=1}^{n}\mathrm{d}x_i \qquad \mathrm{d}_w\mathbf{Y} = (\det\mathbf{C})\mathrm{d}_w\mathbf{X} \qquad |J_{\mathbf{Y}\to\mathbf{X}}| = |\det\mathbf{C}|$$

    - El producto exterior permite construir formas diferenciales de mayor grado a partir de otras de menor grado

Cálculo de Jacobianos: Otras técnicas II

- 4. *Induced Functional Equations*
  - Objetivo: calcular el Jacobiano de $\mathbf{Z} = \mathbf{BYB}'$ donde $\mathbf{Y} = \mathbf{AXA}'$ entonces $\mathbf{Z} = \mathbf{ABX}(\mathbf{AB})'$
  - Aplicamos la propiedad de la multiplicación de Jacobianos $|J_{\mathbf{Z}\to\mathbf{X}}| = |J_{\mathbf{Z}\to\mathbf{Y}}| \cdot |J_{\mathbf{Y}\to\mathbf{X}}|$
  - Transformaciones lineales en **X** → $h(\mathbf{AB}) = h(\mathbf{A})h(\mathbf{B})$ con $h(\mathbf{A}) = |\det \mathbf{A}|^c$

- 5. Jacobianos que involucran traspuestas
  - Queremos calcular el jacobiano de la siguiente transformación

$\mathbf{Y} = \mathbf{AWB}$ $\quad$ $\mathbf{W} = \mathbf{X}'$

$$\mathbf{Y}_{m\times n} = \mathbf{A}_{m\times n}\mathbf{X}'_{m\times n}\mathbf{B}_{n\times n}$$

$$J_{\mathbf{Y}\to\mathbf{X}} = J_{\mathbf{Y}\to\mathbf{W}}\, J_{\mathbf{W}\to\mathbf{X}} = (\det \mathbf{B})^m(\det \mathbf{A})^n(-1)^{\frac{1}{4}m(m-1)n(n-1)}$$

Si no tenemos en cuenta el signo

$|J_{\mathbf{W}\to\mathbf{X}}| = |J_{\mathbf{X}'\to\mathbf{X}}| = 1$

$|J_{\mathbf{Y}\to\mathbf{X}}| = |J_{\mathbf{Y}\to\mathbf{W}}|$

Cálculo de Jacobianos: Otras técnicas III

- 6. Patterned matrices and L-structures
  - Se aprovecha la estructura de las matrices (simétricas, triangulares, diagonales)
  - Si $\mathbf{X} \in \mathbb{R}^{m\times n}$, se define un subespacio lineal $\mathcal{D}_s$ de $\mathbb{R}^{mn}$ que representa la estructura de $\mathbf{X}$

Matriz de vectores base del subespacio $\mathcal{D}_s$

$$\boldsymbol{\Delta}_{mn\times s} = (\boldsymbol{\delta}_1, \boldsymbol{\delta}_2, \ldots, \boldsymbol{\delta}_s)$$

L-structures → colección de matrices mxn que cumplen

$$L(\boldsymbol{\Delta}_{mn\times s}) = \{\mathbf{X} : \mathbf{X} \in \mathbb{R}^{m\times n}, \operatorname{vec}\mathbf{X} \in \mathcal{D}_s\}$$

  - Ejemplo para una matriz simétrica $\mathbf{X} \in \mathbb{R}^{n\times n}$ con $n = 2,\ \mathcal{D}_3 \subset \mathbb{R}^4$

$$\boldsymbol{\Delta}_{4\times 3} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad \boldsymbol{\Delta}_{4\times 3}\begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} a \\ b \\ b \\ c \end{pmatrix} = \operatorname{vec}\begin{pmatrix} a & b \\ b & c \end{pmatrix} = \operatorname{vec}\mathbf{X}$$

  - Ejemplo de Jacobiano de $\mathbf{Y} = \mathbf{F}(\mathbf{X})$
    - Donde $\mathbf{Y} \in L(\boldsymbol{\Delta}_2)$ y $\mathbf{X} \in L(\boldsymbol{\Delta}_1)$

Sí se tiene en cuenta la estructura

$$J_{\mathbf{Y}\to\mathbf{X}} = \det\left[\boldsymbol{\Delta}_2^{+} \frac{\partial \operatorname{vec}\mathbf{Y}}{(\partial \operatorname{vec}\mathbf{X})'} \boldsymbol{\Delta}_1\right]$$

No se tiene en cuenta la estructura

Transformaciones de vectores

- Transformaciones de vectores de funciones simétricas

$$\begin{aligned} y_1 &= x_1 + \cdots + x_n \\ y_2 &= x_1x_2 + x_1x_3 + \cdots + x_{n-1}x_n \\ y_n &= x_1x_2 \cdots x_n \end{aligned}$$

$$|J_{\mathbf{y}\to\mathbf{x}}| = \prod_{i=1}^{n-1} \prod_{j=i+1}^{n} |x_i - x_j|$$

- Jacobianos de las coordenadas polares

$$\begin{aligned} x_1 &= r \sin\theta_1 \sin\theta_2 \cdots \sin\theta_{n-2} \sin\theta_{n-1}, \\ x_2 &= r \sin\theta_1 \sin\theta_2 \cdots \sin\theta_{n-2} \cos\theta_{n-1}, \\ x_3 &= r \sin\theta_1 \sin\theta_2 \cdots \cos\theta_{n-2}, \\ &\vdots \\ x_{n-1} &= r \sin\theta_1 \cos\theta_2, \\ x_n &= r \cos\theta_1, \end{aligned}$$

$$|J_{\mathbf{x} \to r,\boldsymbol{\theta}}| = r^{n-1} |(\sin\theta_1)^{n-2} (\sin\theta_2)^{n-3} \cdots \sin\theta_{n-2}|$$

$\Downarrow$

Si invertimos el orden de $\mathbf{x}_i$

$\Downarrow$

$$|J_{\mathbf{x} \to r,\boldsymbol{\theta}}| = r^{n-1} |(\cos\theta_1)^{n-2} (\cos\theta_2)^{n-3} \cdots \cos\theta_{n-2}|$$

Jacobianos de vectores complejos y matrices

- Queremos calcular el jacobiano de la transformación $\mathbf{y} = \mathbf{A}\mathbf{x}$
  - Donde $\mathbf{y} = \mathbf{y}_1 + i\mathbf{y}_2$ y $\mathbf{x} = \mathbf{x}_1 + i\mathbf{x}_2$ son vectores complejos.

$$\begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{A} \end{pmatrix} \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix} = \mathbf{B}\mathbf{x} \qquad \mathbf{J}_{\mathbf{y}_1,\mathbf{y}_2 \to \mathbf{x}_1,\mathbf{x}_2} = \det(\mathbf{B}) = \det(\mathbf{A})^2$$

  - Si **A** es una matriz compleja $\mathbf{A} = \mathbf{A}_1 + i\mathbf{A}_2$

Transformación **y** = **Ax**

$$\begin{aligned} \mathbf{y} &= \mathbf{y}_1 + i\mathbf{y}_2 = (\mathbf{A}_1 + i\mathbf{A}_2)(\mathbf{x}_1 + i\mathbf{x}_2) \\ &= (\mathbf{A}_1\mathbf{x}_1 - \mathbf{A}_2\mathbf{x}_2) + i(\mathbf{A}_1\mathbf{x}_2 + \mathbf{A}_2\mathbf{x}_1) \end{aligned}$$

Matriz Jacobiana

$$\partial\mathbf{y}_1/\partial\mathbf{x}_1' = \mathbf{A}_1 \qquad \partial\mathbf{y}_1/\partial\mathbf{x}_2' = -\mathbf{A}_2$$

$$\partial\mathbf{y}_2/\partial\mathbf{x}_2' = \mathbf{A}_1 \qquad \partial\mathbf{y}_2/\partial\mathbf{x}_1' = \mathbf{A}_2$$

Determinante de la matriz Jacobiana

$$|\mathbf{J}_{\mathbf{y}_1,\mathbf{y}_2 \to \mathbf{x}_1,\mathbf{x}_2}| = \left|\det\begin{pmatrix} \mathbf{A}_1 & -\mathbf{A}_2 \\ \mathbf{A}_2 & \mathbf{A}_1 \end{pmatrix}\right| = (|\det \mathbf{A}|)^2 = |\det(\mathbf{A}\mathbf{A}^*)|$$

Matrices con elementos funcionalmente independientes

- Aquí vamos a hablar de matrices cuyos elementos son funcionalmente independientes

  - El objetivo es transformar cada una de las funciones para que tengan relaciones "one to one"

Función original

$$\mathbf{Y} = \mathbf{A}\mathbf{X}^{-1}\mathbf{B}$$

Transformación "one to one"

$$\mathbf{Y} = \mathbf{A}\mathbf{Z}\mathbf{B},\ \mathbf{Z} = \mathbf{X}^{-1}$$

Jacobiano de la transformación

$$|J_{\mathbf{Y}\to\mathbf{X}}| = |(\det \mathbf{B})^n (\det \mathbf{X})^{-2n} (\det \mathbf{A})^n|$$

  - Ejemplo de cálculo del Jacobiano de la transformación

$$\mathbf{Y} = \mathbf{A}\mathbf{X}\mathbf{A}' \pm \mathbf{B}\mathbf{X}\mathbf{B}' \qquad \mathrm{vec}\,\mathbf{Y} = (\mathbf{A}\otimes\mathbf{A} \pm \mathbf{B}\otimes\mathbf{B})\mathrm{vec}\,\mathbf{X} \qquad |J_{\mathbf{Y}\to\mathbf{X}}| = \prod_{i=1}^{n}\prod_{j=1}^{n} |\alpha_i\alpha_j \pm \beta_i\beta_j|$$

Autovalores de **A** y **B**

  - Caso práctico del cálculo del Jacobiano de la transformación $\mathbf{Y} = \mathbf{X}^2$

Matrices ortogonales

$$\mathbf{X} = \mathbf{V}\begin{bmatrix}\lambda_1 & 0\\ 0 & \lambda_2\end{bmatrix}\mathbf{U}^{\mathsf{T}} \qquad \mathrm{dvec}(\mathbf{Y}) = (\mathbf{I}\otimes\mathbf{X} + \mathbf{X}\otimes\mathbf{I})\mathrm{dvec}(\mathbf{X}) \qquad |\mathbf{J}_{\mathbf{Y}\to\mathbf{X}}| = ?$$

Jacobiano de matrices simétricas y hermíticas

- Jacobiano de matrices con estructura

Matrices con diagonal nula

Número de variables funcionalmente independientes

Transformación	Si **X** es simétrica $\mathbf{X}=\mathbf{X}^{\mathsf{T}}$	Si **X** es antisimétrica $\mathbf{X}=-\mathbf{X}^{\mathsf{T}}$
$\mathbf{Y}=a\mathbf{X}$	$\lvert J_{\mathbf{Y}\to\mathbf{X}}\rvert = \lvert a^{n(n+1)/2}\rvert$	$\lvert J_{\mathbf{Y}\to\mathbf{X}}\rvert = \lvert a\rvert^{n(n-1)/2}$
$\mathbf{Y}=\mathbf{A}\mathbf{X}\mathbf{A}'$	$[\lvert J_{\mathbf{Y}\to\mathbf{X}}\rvert = (\lvert\det\mathbf{A})\rvert^{n+1}$	$\lvert J_{\mathbf{Y}\to\mathbf{X}}\rvert = (\lvert\det\mathbf{A}\rvert)^{n-1}$
$\mathbf{Y}=\mathbf{A}\mathbf{X}^{-1}\mathbf{A}'$	$\lvert J_{\mathbf{Y}\to\mathbf{X}}\rvert = \lvert(\det\mathbf{A})^{n+1}(\det\mathbf{X})^{-(n+1)}\rvert$	$\lvert J_{\mathbf{Y}\to\mathbf{X}}\rvert = \lvert\det\mathbf{A}\rvert^{n-1}\lvert\det\mathbf{X}\rvert^{-(n-1)}$

- Si **X** es hermítica

$$\mathbf{X}=\mathbf{X}_1+i\mathbf{X}_2$$

$$\mathbf{Y}=\mathbf{Y}_1+i\mathbf{Y}_2=\mathbf{A}\mathbf{X}\mathbf{A}^*$$

$$\lvert J_{\mathbf{Y}_1,\mathbf{Y}_2\to\mathbf{X}_1,\mathbf{X}_2}\rvert = \lvert\det\mathbf{A}\rvert^{2n}$$

Si **X** es una matriz simétrica

Si **A** y **B** son matrices triangulares

$$\mathbf{Y}=\mathbf{A}\mathbf{X}\mathbf{A}'\pm\mathbf{B}\mathbf{X}\mathbf{B}'$$

$$\operatorname{vech}\mathbf{Y}=[\mathbf{H}(\mathbf{A}\otimes\mathbf{A}\pm\mathbf{B}\otimes\mathbf{B})\mathbf{G}]\operatorname{vech}\mathbf{X}$$

$$\lvert J_{\mathbf{Y}\to\mathbf{X}}\rvert = \Big\lvert\prod_{i=1}^{n}\prod_{j=i}^{n}(a_{ii}a_{jj}\pm b_{ii}b_{jj})\Big\rvert$$

Jacobiano de matrices triangulares

- Determinante de la matriz Jacobiana de matrices triangulares $\mathbf{Y} = \mathbf{PXQ}$ $\mathbf{P}, \mathbf{X}, \mathbf{Q} \in \mathbb{R}^{n\times n}$

$$|J_{\mathbf{Y}\to\mathbf{X}}| = \left|\prod_{i=1}^{n} p_{ii}^{i} q_{ii}^{n-i+1}\right|$$

$$\mathbf{P} = \begin{bmatrix} 2 & 0 \\ 1 & 1 \end{bmatrix} \quad \mathbf{Q} = \begin{bmatrix} 2 & 0 \\ 1 & 3 \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} x_{11} & 0 \\ x_{21} & x_{22} \end{bmatrix}$$

$$\mathbf{Y} = \begin{bmatrix} 4x_{11} & 0 \\ 2(x_{11}+x_{21}) + x_{22} & 3x_{22} \end{bmatrix}$$

$$\mathbf{J} = \begin{bmatrix} 4 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 1 & 3 \end{bmatrix}$$

- Transformaciones no lineales con matrices triangulares

Transformaciones cuadráticas

$$\tilde{\mathbf{Y}} = \tilde{\mathbf{X}}\tilde{\mathbf{P}}\tilde{\mathbf{X}}$$

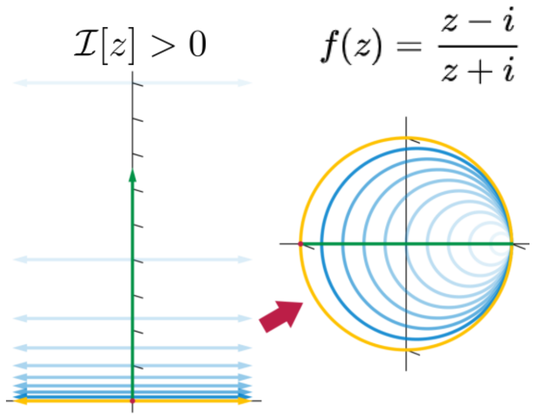
$$|J_{\tilde{\mathbf{Y}}\to\tilde{\mathbf{X}}}| = 2^n \left|(\det\tilde{\mathbf{P}})(\det\tilde{\mathbf{X}}) \prod_{i=1}^{n}\prod_{j=i+1}^{n}(p_{ii}x_{ii} + p_{jj}x_{jj})\right|$$

Transformación que involucra la inversa de **X**

$$\tilde{\mathbf{Y}} = \tilde{\mathbf{X}}^{-1}$$

$$|J_{\tilde{\mathbf{Y}}\to\tilde{\mathbf{X}}}| = |\det\tilde{\mathbf{X}}|^{-(n+1)}| = |\prod_{i=1}^{n} x_{ii}^{-(n+1)}|$$

Descomposiciones basadas en matrices antisimétricas



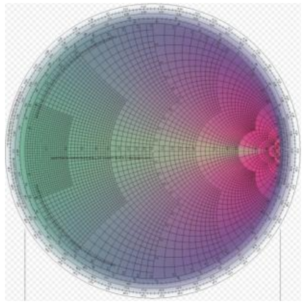
- Descomposiciones basadas en matrices antisimétricas
  - Cualquier matriz no singular **Y** puede expresarse como **Y** = **XT**
  - Siendo $\mathbf{X}$ una matriz triangular, $\boldsymbol{S} = -\boldsymbol{S}^T$ una matriz antisimétrica y
  - $\mathbf{T} = (\mathbf{I}_n - \mathbf{S})(\mathbf{I}_n + \mathbf{S})^{-1} = 2(\mathbf{S} + \mathbf{I}_n)^{-1} - \mathbf{I}_n$
    - una matriz ortogonal de transformaciones "one-to-one"

$$\mathbf{Y} = \mathbf{T}\tilde{\mathbf{X}}\mathbf{T}' = [2(\mathbf{S} + \mathbf{I}_n)^{-1} - \mathbf{I}_n]\tilde{\mathbf{X}}[2(\mathbf{S} + \mathbf{I}_n)^{-1} - \mathbf{I}_n]'$$

$$|J_{\mathbf{Y}\to\tilde{\mathbf{X}},\mathbf{S}}| = 2^{n(n-1)/2}\,|\det(\mathbf{I}_n + \mathbf{S})^{-(n-1)}\prod_{i=1}^{n}\prod_{j=i+1}^{n}(x_{ii} - x_{jj})|$$

- Aplicaciones
  - Adaptar polinomios de Legendre de funciones reales positivas a funciones racionales
  - Relación entre modelo de semiplano y modelo de disco de Poincaré
  - Adaptación de impedancias de las líneas de transmisión

- *Caley transformation*
  - Mapeo entre matrices antisimétricas y matrices ortogonales



$$\mathbf{Y} = 2(\mathbf{A} + \mathbf{X})^{-1}\mathbf{A} - \mathbf{I}_n \qquad |J_{\mathbf{Y}\to\mathbf{X}}| = 2^{n^2}\,(|\det \mathbf{A})|^n(|\det(\mathbf{A} + \mathbf{X})|^{-2n}$$

Transformaciones basadas en la estructura de Y

- Transformaciones basadas en la estructura de $\mathbf{Y} \in \mathbb{R}^{n\times n}$
  - Matrices triangular inferior no singulares $\tilde{\mathbf{X}}, \tilde{\mathbf{P}}, \tilde{\mathbf{Q}}, \tilde{\mathbf{R}} \in \mathbb{R}^{n\times n}$

- Matriz simétrica

$$\mathbf{Y} = \tilde{\mathbf{X}}\tilde{\mathbf{P}} + \tilde{\mathbf{P}}'\tilde{\mathbf{X}}'$$

$$|J_{\mathbf{Y}\to\tilde{\mathbf{X}}}| = 2^n \prod_{i=1}^{n} |p_{ii}|^{n-i+1}$$

- Matriz definida positiva

$$\mathbf{Y} = \tilde{\mathbf{X}}\tilde{\mathbf{X}}'$$

$$|J_{\mathbf{Y}\to\tilde{\mathbf{X}}}| = 2^n \prod_{i=1}^{n} x_{ii}^{n-i+1}$$

- Matriz definida positiva y hermítica

$$\mathbf{Y} = \tilde{\mathbf{X}}\tilde{\mathbf{X}}^*$$

$$|J_{\mathbf{Y}_1,\mathbf{Y}_2\to\tilde{\mathbf{X}}_1,\tilde{\mathbf{X}}_2}| = 2^n \prod_{i=1}^{n} x_{ii}^{2(n-i)+1}$$

- Matriz antisimétrica

$$\mathbf{Y} = \tilde{\mathbf{R}}'\tilde{\mathbf{Q}}\tilde{\mathbf{X}}\tilde{\mathbf{P}} - \tilde{\mathbf{P}}'\tilde{\mathbf{X}}'\tilde{\mathbf{Q}}'\tilde{\mathbf{R}}$$

$$|J_{\mathbf{Y}\to\tilde{\mathbf{X}}}| = |\prod_{i=1}^{n} (q_{ii}r_{ii})^{i-1} p_{ii}^{n-i}|$$

- Descomposición LU

$$\mathbf{Y} = \tilde{\mathbf{L}}\mathbf{U}$$

$$|J_{\mathbf{Y}\to\tilde{\mathbf{L}},\mathbf{U}}| = |\prod_{i=1}^{n} l_{ii}^{n-i+1} u_{ii}^{n-i}|$$

Descomposiciones que involucran matrices diagonales

- Matrices cuadradas con valores unitarios en la diagonal $\mathbf{X},\ \mathbf{Y} \in \mathbb{R}^{n\times n}$

$$\mathbf{Y} = \mathbf{D_w X}$$

$$|J_{\mathbf{Y}\to\mathbf{w},\mathbf{X}}| = \prod_{i=1}^{n} |w_i|^{n-1}$$

$$\mathbf{Y} = \mathbf{D_w X D_w}$$

$$|J_{\mathbf{Y}\to\mathbf{w},\mathbf{X}}| = 2^n \prod_{i=1}^{n} w_i^{2n-1}$$

- Matrices triangulares con valores unitarios en la diagonal $\tilde{\mathbf{X}}, \tilde{\mathbf{Y}} \in \mathbb{R}^{n\times n}$

$$\tilde{\mathbf{Y}} = \mathbf{D_w}\tilde{\mathbf{X}}$$

$$|J_{\tilde{\mathbf{Y}}\to\mathbf{w},\tilde{\mathbf{X}}}| = |\det \tilde{\mathbf{X}}| \prod_{i=1}^{n} |w_i|^{i-1}$$

$$\mathbf{Y} = \tilde{\mathbf{X}}\mathbf{D_w}\tilde{\mathbf{X}}'$$

$$|J_{\mathbf{Y}\to\mathbf{w},\tilde{\mathbf{X}}}| = (\det \tilde{\mathbf{X}})^2 \prod_{i=1}^{n} |(w_i x_{ii})|^{n-i}$$

- Matrices simétricas con valores unitarios en la diagonal $\mathbf{X},\ \mathbf{Y} \in \mathbb{R}^{n\times n}$

$$\mathbf{Y} = \mathbf{D_w X D_w}$$

$$|J_{\mathbf{Y}\to\mathbf{w},\mathbf{X}}| = 2^n \prod_{i=1}^{n} |w_i|^n$$

Descomposiciones que involucran matrices definidas positivas

- Matrices definidas positivas $\mathbf{X},\, \mathbf{Y} \in \mathbb{R}^{n\times n}$

$$\mathbf{Y} = (\det \mathbf{X})\mathbf{X}^{-1}$$

$$|J_{\mathbf{Y}\to\mathbf{X}}| = (n-1)|(\det \mathbf{X})^{(n+1)(n-2)/2}|$$

$$\mathbf{Y} = \mathbf{XAX}$$

$$|J_{\mathbf{Y}\to\mathbf{X}}| = \prod_{i=1}^{n}\prod_{j=i}^{n}(\lambda_i + \lambda_j)$$

Autovalores de **X**

- Matrices diagonalizables

$$\mathbf{Y} = \mathbf{X}^k \qquad |J_{\mathbf{Y}\to\mathbf{X}}| = \prod_{i=1}^{n}\prod_{j=1}^{n}\sum_{r=1}^{k}\lambda_i^{k-r}\lambda_j^{r-1}$$

- Pares de matrices

$$\mathbf{Y}_1 = \mathbf{X}_2^{-1/2}\mathbf{X}_1\mathbf{X}_2^{-1/2}$$

$$\mathbf{Y}_2 = \mathbf{X}_2$$

$$|J_{\mathbf{Y}_1,\mathbf{Y}_2\to\mathbf{X}_1,\mathbf{X}_2}| = |\det \mathbf{X}_2|^{-(n+1)/2}$$

$$\mathbf{Y}_1 = (\mathbf{X}_1+\mathbf{X}_2)^{-1/2}\mathbf{X}_1(\mathbf{X}_1+\mathbf{X}_2)^{-1/2}$$

$$\mathbf{Y}_2 = \mathbf{X}_1 + \mathbf{X}_2$$

$$|J_{\mathbf{Y}_1,\mathbf{Y}_2\to\mathbf{X}_1,\mathbf{X}_2}| = |\det(\mathbf{X}_1+\mathbf{X}_2)|^{-(n+1)/2}$$

Recapitulación de conceptos

- Hemos visto qué es el Jacobiano
- Ejemplos de aplicaciones donde se usa el Jacobiano
- Distintos métodos para el cálculo del Jacobiano
- Casos particulares en los que el Jacobiano tiene una expresión “sencilla”
- Jacobiano de transformaciones que involucran matrices con estructura