

Resumen de la Demostración del Teorema de MDS Clásico

Juan Rubio Cobeta

October 3, 2025

1 Idea Central e Hilo Conductor

La demostración establece una equivalencia fundamental: una matriz de distancias \mathbf{D} es euclídea si y solo si una matriz transformada \mathbf{B} , conocida como la matriz de productos escalares centrados, es semidefinida positiva (p.s.d.).

El hilo conductor es el siguiente:

1. Se parte de las distancias al cuadrado d_{rs}^2 .
2. Se convierten en una matriz \mathbf{A} (que está relacionada con los productos escalares, aunque no directamente).
3. Se centra doblemente esta matriz para obtener $\mathbf{B} = \mathbf{H}\mathbf{A}\mathbf{H}$. Este paso es crucial, ya que elimina los efectos de la traslación del origen de coordenadas.
4. La demostración prueba que los elementos de \mathbf{B} son exactamente los productos escalares de los vectores de coordenadas ya centrados respecto a su media (centro de gravedad).
5. A partir de ahí, se utiliza la propiedad de que una matriz de productos escalares (matriz de Gram) es siempre semidefinida positiva. El camino inverso se basa en la descomposición espectral de la matriz \mathbf{B} para reconstruir las coordenadas.

2 Demostración del Teorema

El teorema afirma: **Una matriz de distancias \mathbf{D} es euclídea si y solo si la matriz $\mathbf{B} = \mathbf{H}\mathbf{A}\mathbf{H}$ es semidefinida positiva**, donde $a_{rs} = -\frac{1}{2}d_{rs}^2$ y $\mathbf{H} = \mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}'$ es la matriz de centrado.

La prueba tiene dos partes, correspondientes a las dos implicaciones.

2.1 Parte 1: Si \mathbf{D} es euclídea $\implies \mathbf{B}$ es p.s.d.

- **Hipótesis y Objetivo:** Partimos de que \mathbf{D} es euclídea. Esto significa que existen puntos (coordenadas) $\mathbf{z}_1, \dots, \mathbf{z}_n$ en un espacio \mathbb{R}^p tales que $d_{rs}^2 = (\mathbf{z}_r - \mathbf{z}_s)'(\mathbf{z}_r - \mathbf{z}_s)$. El objetivo es demostrar que \mathbf{B} debe ser p.s.d.
- **Paso 1: Relacionar \mathbf{A} con las coordenadas.** Expandiendo la distancia al cuadrado y usando la definición $a_{rs} = -\frac{1}{2}d_{rs}^2$, obtenemos:

$$a_{rs} = -\frac{1}{2}(\mathbf{z}'_r \mathbf{z}_r - 2\mathbf{z}'_r \mathbf{z}_s + \mathbf{z}'_s \mathbf{z}_s)$$

- **Paso 2: La transformación de doble centrado.** La matriz $\mathbf{B} = \mathbf{H}\mathbf{A}\mathbf{H}$ realiza una operación de "doble centrado" sobre \mathbf{A} . Un elemento b_{rs} de \mathbf{B} se puede escribir como:

$$b_{rs} = a_{rs} - \bar{a}_{r.} - \bar{a}_{.s} + \bar{a}_{..}$$

donde $\bar{a}_{r.}$ es la media de la fila r , $\bar{a}_{.s}$ la media de la columna s , y $\bar{a}_{..}$ es la media total de \mathbf{A} .

- **Paso 3: El resultado clave.** Al sustituir la expresión de a_{rs} en términos de las coordenadas \mathbf{z} dentro de la fórmula de b_{rs} , y tras una simplificación algebraica (que es el núcleo de esta parte de la prueba), se llega a una conclusión notable:

$$b_{rs} = (\mathbf{z}_r - \bar{\mathbf{z}})'(\mathbf{z}_s - \bar{\mathbf{z}})$$

donde $\bar{\mathbf{z}} = \frac{1}{n} \sum_{i=1}^n \mathbf{z}_i$ es el centro de gravedad de la configuración de puntos.

- **Paso 4: Conclusión.** La ecuación anterior significa que la matriz \mathbf{B} es la matriz de productos escalares de los vectores de coordenadas centrados. Si definimos una matriz $\tilde{\mathbf{Z}}$ cuyas filas son $(\mathbf{z}_r - \bar{\mathbf{z}})'$, entonces \mathbf{B} se puede escribir como $\mathbf{B} = \tilde{\mathbf{Z}}\tilde{\mathbf{Z}}'$. Una matriz de esta forma (conocida como matriz de Gram) es, por definición, semidefinida positiva. Esto completa la primera parte de la prueba.

2.2 Parte 2: Si \mathbf{B} es p.s.d. $\implies \mathbf{D}$ es euclídea.

- **Hipótesis y Objetivo:** Ahora partimos de que \mathbf{B} es simétrica y p.s.d. El objetivo es construir una configuración de puntos $\mathbf{x}_1, \dots, \mathbf{x}_n$ tal que las distancias entre ellos correspondan a las de la matriz \mathbf{D} original.
- **Paso 1: Descomposición espectral de \mathbf{B} .** Como \mathbf{B} es simétrica y p.s.d., admite una descomposición espectral: $\mathbf{B} = \mathbf{\Gamma}\mathbf{\Lambda}\mathbf{\Gamma}'$, donde $\mathbf{\Lambda}$ es una matriz diagonal con los autovalores no negativos ($\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p > 0, \dots, \lambda_n = 0$) y $\mathbf{\Gamma}$ es la matriz de autovectores ortonormales correspondientes.

- **Paso 2: Construcción de las coordenadas.** Se construye una matriz de coordenadas \mathbf{X} de tamaño $n \times p$ (donde p es el rango de \mathbf{B} , i.e., el número de autovalores positivos) de la siguiente manera:

$$\mathbf{X} = \mathbf{\Gamma}_p \mathbf{\Lambda}_p^{1/2}$$

Aquí, $\mathbf{\Gamma}_p$ contiene los primeros p autovectores y $\mathbf{\Lambda}_p^{1/2}$ es la matriz diagonal con las raíces cuadradas de los p autovalores positivos. La fila r de \mathbf{X} , denotada \mathbf{x}'_r , son las coordenadas del punto r .

- **Paso 3: Verificar que la matriz de productos escalares de \mathbf{X} es \mathbf{B} .** Calculamos la matriz de productos escalares de la configuración \mathbf{X} :

$$\mathbf{X}\mathbf{X}' = (\mathbf{\Gamma}_p \mathbf{\Lambda}_p^{1/2})(\mathbf{\Gamma}_p \mathbf{\Lambda}_p^{1/2})' = \mathbf{\Gamma}_p \mathbf{\Lambda}_p^{1/2} \mathbf{\Lambda}_p^{1/2} \mathbf{\Gamma}_p' = \mathbf{\Gamma}_p \mathbf{\Lambda}_p \mathbf{\Gamma}_p' = \mathbf{B}$$

Esto confirma que los elementos de \mathbf{B} , b_{rs} , son los productos escalares $\mathbf{x}'_r \mathbf{x}_s$.

- **Paso 4: Reconstruir la distancia original.** Ahora, calculamos la distancia euclídea al cuadrado entre dos puntos construidos \mathbf{x}_r y \mathbf{x}_s :

$$\|\mathbf{x}_r - \mathbf{x}_s\|^2 = (\mathbf{x}_r - \mathbf{x}_s)'(\mathbf{x}_r - \mathbf{x}_s) = \mathbf{x}'_r \mathbf{x}_r - 2\mathbf{x}'_r \mathbf{x}_s + \mathbf{x}'_s \mathbf{x}_s$$

Usando el resultado del paso anterior, esto es igual a $b_{rr} - 2b_{rs} + b_{ss}$. Al revertir la definición de b_{rs} en términos de a_{rs} y sus medias, se demuestra que esta expresión se simplifica a:

$$b_{rr} - 2b_{rs} + b_{ss} = -2a_{rs}$$

- **Paso 5: Conclusión Final.** Recordando que $a_{rs} = -\frac{1}{2}d_{rs}^2$, obtenemos:

$$\|\mathbf{x}_r - \mathbf{x}_s\|^2 = -2 \left(-\frac{1}{2}d_{rs}^2 \right) = d_{rs}^2$$

Hemos construido con éxito un conjunto de puntos cuyas distancias inter-punto coinciden con las dadas en \mathbf{D} . Por lo tanto, \mathbf{D} es una matriz de distancias euclídea.

La gran utilidad de este teorema es que funciona como un puente entre un concepto abstracto, como una tabla de distancias entre diferentes objetos, y algo concreto y visual, como un mapa de puntos. Permite hacer dos cosas fundamentales. Primero, nos da un criterio claro para saber si es posible representar esas distancias de forma perfecta en un mapa sin distorsión. Segundo, y más importante, nos ofrece el método exacto, paso a paso, para calcular las coordenadas de cada punto en ese mapa.