

Trabajo Final

Juan Rubio Cobeta

December 1, 2025

Contents

1 Formulación del Problema de Minimización en MDS Métrico	2
2 Fundamentos de la Teoría de Mayorización	3
2.1 Definición y Propiedades de la Función Subrogada	3
2.2 El Algoritmo Iterativo y la Desigualdad Sándwich	3
3 Derivación Algebraica del Algoritmo SMACOF	4
3.1 Descomposición del Stress en Componentes Convexos y Cóncavos	4
3.2 La Función Mayorizante y la Desigualdad de Cauchy-Schwarz	4
3.3 La Transformada de Guttman	5
4 Generalizaciones y Restricciones Configuracionales	6
4.1 Optimización bajo Restricciones Lineales Externas	6
4.2 Escalamiento de Diferencias Individuales (3-Way SMACOF)	6
4.3 Unfolding Métrico para Matrices Rectangulares	6
5 Variantes No Métricas y el Problema de la Regresión Isotónica	7
5.1 Formulación del Problema de Optimización Ordinal	7
5.2 El Algoritmo de Relajación por Bloques	7
5.3 Tratamiento Riguroso de Empates (Ties)	7
5.3.1 Enfoque Secundario (Restricción Fuerte)	8
5.3.2 Enfoque Primario (Restricción Débil)	8

1 Formulación del Problema de Minimización en MDS Métrico

El problema central del Escalamiento Multidimensional (MDS) métrico puede definirse como la búsqueda de una inmersión isométrica aproximada de un conjunto finito de objetos en un espacio vectorial Euclídeo de baja dimensión.

Sea $\mathcal{O} = \{o_1, o_2, \dots, o_n\}$ un conjunto de n entidades o estímulos bajo estudio. Asumimos la existencia de una función de disimilitud empírica $\delta : \mathcal{O} \times \mathcal{O} \rightarrow \mathbb{R}_{\geq 0}$, que cuantifica la divergencia o lejanía perceptual entre pares de objetos. La información de entrada se sistematiza en una matriz de proximidad $\Delta \in \mathbb{R}^{n \times n}$, cuyos elementos $\delta_{ij} = \delta(o_i, o_j)$ satisfacen los axiomas de una semimétrica:

1. *No negatividad*: $\delta_{ij} \geq 0$ para todo $1 \leq i, j \leq n$.
2. *Simetría*: $\delta_{ij} = \delta_{ji}$ para todo $1 \leq i, j \leq n$.
3. *Propiedad hueca (Hollow)*: $\delta_{ii} = 0$ para todo $1 \leq i \leq n$.

El objetivo del análisis es determinar una configuración espacial representada por la matriz $X \in \mathbb{R}^{n \times p}$, donde $p \ll n$ denota la dimensionalidad del espacio latente. Cada fila x_i^\top de X representa el vector de coordenadas del objeto o_i en el espacio Euclídeo \mathbb{R}^p .

La estructura métrica del espacio de representación viene dada por la distancia Euclídea estándar. Definimos la función de distancia $d_{ij} : \mathbb{R}^{n \times p} \rightarrow \mathbb{R}_{\geq 0}$ como la norma ℓ_2 de la diferencia de los vectores de posición:

$$d_{ij}(X) = \|x_i - x_j\|_2 = \sqrt{\sum_{s=1}^p (x_{is} - x_{js})^2} \quad (1)$$

El problema de optimización se formula como la minimización de una función de pérdida, denominada *Raw Stress*, que evalúa la bondad de ajuste entre las disimilitudes observadas y las distancias configuracionales. Analíticamente, buscamos X^* tal que:

$$X^* = \arg \min_{X \in \mathbb{R}^{n \times p}} \sigma(X) \quad (2)$$

donde la función objetivo $\sigma(X)$ se define mediante la suma ponderada de los residuos cuadráticos:

$$\sigma(X) = \sum_{1 \leq i < j \leq n} w_{ij} (\delta_{ij} - d_{ij}(X))^2 \quad (3)$$

La matriz de ponderación $W = (w_{ij}) \in \mathbb{R}^{n \times n}$ juega un papel crucial en la especificación del modelo. Asumimos que W es simétrica, no negativa y hueca. Esta matriz permite incorporar información a priori sobre la fiabilidad de las disimilitudes o manejar estructuras de datos incompletos imponiendo $w_{ij} = 0$ si δ_{ij} es desconocido.

Desde una perspectiva de teoría de grafos, definimos el grafo de adyacencia asociado a W con vértices $\mathcal{V} = \{1, \dots, n\}$ y aristas $\mathcal{E} = \{(i, j) : w_{ij} > 0\}$. Para garantizar que la solución sea única (salvo traslaciones y rotaciones rígidas) y que el problema no se escinda en subproblemas independientes, es condición necesaria y suficiente que dicho grafo sea conexo. En términos algebraicos, esto implica que la matriz W debe ser *irreducible*.

Finalmente, para evitar soluciones triviales ($X = 0$) y fijar la escala de la solución respecto a los datos de entrada, se impone, sin pérdida de generalidad, la siguiente condición de normalización sobre las disimilitudes ponderadas:

$$\eta_\delta^2 = \sum_{1 \leq i < j \leq n} w_{ij} \delta_{ij}^2 = \text{constante} \quad (4)$$

Típicamente, esta constante se fija en $n(n - 1)/2$, lo que simplifica la derivación de los términos constantes en la descomposición del Stress.

2 Fundamentos de la Teoría de Mayorización

En un sentido estricto, la mayorización no debe concebirse como un algoritmo singular, sino como un paradigma metodológico para la construcción de algoritmos de optimización iterativa. Este enfoque, que generaliza procedimientos como el algoritmo EM (Expectation-Maximization) y se inscribe en la clase de los algoritmos MM (Majorize-Minimize), transforma un problema de optimización complejo en una sucesión de problemas elementales más manejables.

Sea $f : \mathcal{D} \subseteq \mathbb{R}^p \rightarrow \mathbb{R}$ una función objetivo escalar, continua y acotada inferiormente que deseamos minimizar. En el contexto de SMACOF, $f(\cdot)$ corresponde a la función de Stress $\sigma(\cdot)$. La dificultad analítica para minimizar $f(x)$ directamente motiva la introducción de una función auxiliar.

2.1 Definición y Propiedades de la Función Subrogada

Definimos una función subrogada o mayorizante $g : \mathcal{D} \times \mathcal{D} \rightarrow \mathbb{R}$, la cual depende de una variable libre x y un punto de soporte fijo y . Para que $g(x, y)$ sea una mayorización válida de $f(x)$ en y , debe satisfacer las siguientes condiciones axiomáticas para todo $x, y \in \mathcal{D}$:

1. *Condición de Dominancia:* La función auxiliar acota superiormente a la función objetivo en todo el dominio.

$$g(x, y) \geq f(x) \quad (5)$$

2. *Condición de Tangencia:* La función auxiliar coincide con la función objetivo en el punto de soporte.

$$g(y, y) = f(y) \quad (6)$$

Geométricamente, esto implica que la superficie $z = g(x, y)$ es tangente a la superficie $z = f(x)$ en el punto $x = y$ y permanece por encima de ella en cualquier otro punto.

2.2 El Algoritmo Iterativo y la Desigualdad Sándwich

El principio de mayorización prescribe un procedimiento iterativo donde, dado un punto inicial $x^{(0)}$, se genera una sucesión $\{x^{(t)}\}_{t \geq 0}$ mediante la regla de actualización:

$$x^{(t+1)} = \arg \min_{x \in \mathcal{D}} g(x, x^{(t)}) \quad (7)$$

La convergencia monótona del algoritmo está garantizada por la denominada *Desigualdad Sándwich*. Si definimos $y = x^{(t)}$ como el punto de soporte en la iteración actual, y $x^{(t+1)}$ como el minimizador de la función subrogada, se establece la siguiente cadena de desigualdades:

$$f(x^{(t+1)}) \leq g(x^{(t+1)}, x^{(t)}) \leq g(x^{(t)}, x^{(t)}) = f(x^{(t)}) \quad (8)$$

Analicemos la implicación de cada término:

- La primera desigualdad, $f(x^{(t+1)}) \leq g(x^{(t+1)}, x^{(t)})$, es consecuencia directa de la condición de dominancia (11).
- La segunda desigualdad, $g(x^{(t+1)}, x^{(t)}) \leq g(x^{(t)}, x^{(t)})$, se deriva de la definición de $x^{(t+1)}$ como el minimizador de $g(\cdot, x^{(t)})$.
- La igualdad final, $g(x^{(t)}, x^{(t)}) = f(x^{(t)})$, resulta de la condición de tangencia (12).

Por consiguiente, $f(x^{(t+1)}) \leq f(x^{(t)})$, lo que demuestra que la sucesión de valores de la función objetivo es monótona decreciente. Dado que f está acotada inferiormente (por ser una suma de cuadrados en el caso del Stress), la sucesión converge a un límite.

La simplicidad de este enfoque radica en la elección inteligente de $g(x, y)$. En SMACOF, $g(x, y)$ se construye como una función cuadrática convexa en x , cuya minimización analítica es trivial (resolución de un sistema lineal), evitando así el uso de métodos de gradiente numérico complejos sobre la función original $f(x)$.

3 Derivación Algebraica del Algoritmo SMACOF

La robustez del algoritmo SMACOF reside en la descomposición analítica del funcional de Stress y en la linealización de su componente cóncavo mediante desigualdades matriciales.

3.1 Descomposición del Stress en Componentes Convexos y Cóncavos

Expandiendo el término cuadrático de la Ecuación (6), el Stress puede descomponerse en tres componentes aditivos:

$$\sigma(X) = \sum_{i < j} w_{ij} \delta_{ij}^2 + \sum_{i < j} w_{ij} d_{ij}^2(X) - 2 \sum_{i < j} w_{ij} \delta_{ij} d_{ij}(X) \quad (9)$$

Denotamos el primer término constante como η_δ^2 . Los dos términos restantes dependen de la configuración X . Para expresarlos en notación matricial, introducimos las matrices elementales $A_{ij} \in \mathbb{R}^{n \times n}$, definidas como matrices de rango 1 de la forma:

$$A_{ij} = (e_i - e_j)(e_i - e_j)^\top \quad (10)$$

donde e_i es el i -ésimo vector de la base canónica de \mathbb{R}^n .

Definimos la matriz de rigidez $V \in \mathbb{R}^{n \times n}$ como la combinación lineal ponderada de estas matrices elementales:

$$V = \sum_{i < j} w_{ij} A_{ij} \quad (11)$$

Dado que $d_{ij}^2(X) = \text{tr}(X A_{ij} X^\top)$, el segundo término del Stress (componente cuadrático) se reescribe como:

$$\eta^2(X) = \sum_{i < j} w_{ij} d_{ij}^2(X) = \text{tr}(X^\top V X) \quad (12)$$

Observamos que V es una matriz Laplaciana ponderada, semidefinida positiva, lo que implica que $\eta^2(X)$ es una función convexa en X .

El tercer término, denominado componente cruzado o $\rho(X)$, encapsula la no-linealidad del problema:

$$\rho(X) = \sum_{i < j} w_{ij} \delta_{ij} d_{ij}(X) = \text{tr}(X^\top B(X) X) \quad (13)$$

donde la matriz $B(X)$ depende de la configuración actual y se define como:

$$B(X) = \sum_{i < j} w_{ij} s_{ij}(X) A_{ij}, \quad \text{con } s_{ij}(X) = \begin{cases} \delta_{ij}/d_{ij}(X) & \text{si } d_{ij}(X) > 0 \\ 0 & \text{si } d_{ij}(X) = 0 \end{cases} \quad (14)$$

En consecuencia, la función objetivo adopta la forma compacta:

$$\sigma(X) = \eta_\delta^2 + \text{tr}(X^\top V X) - 2\text{tr}(X^\top B(X) X) \quad (15)$$

Dado que $d_{ij}(X)$ es convexo, $-\rho(X)$ es cóncavo. La dificultad de la minimización directa radica en este término.

3.2 La Función Mayorizante y la Desigualdad de Cauchy-Schwarz

Para aplicar el principio de mayorización, necesitamos construir una función cuadrática $\tau(X, Y)$ que acote superiormente a $\sigma(X)$. Esto equivale a encontrar una minorización lineal para el término convexo $\rho(X)$.

Aplicando la desigualdad de Cauchy-Schwarz sobre los vectores de diferencia $(x_i - x_j)$ y $(y_i - y_j)$, obtenemos la relación fundamental:

$$d_{ij}(X) \geq \frac{1}{d_{ij}(Y)} \text{tr}((x_i - x_j)(y_i - y_j)^\top) \quad \forall d_{ij}(Y) > 0 \quad (16)$$

Sumando ponderadamente sobre todos los pares, se deduce que:

$$\rho(X) = \sum_{i < j} w_{ij} \delta_{ij} d_{ij}(X) \geq \text{tr}(X^\top B(Y) Y) \quad (17)$$

Sustituyendo esta desigualdad en la ecuación del Stress, obtenemos la función subrogada $\tau(X, Y)$:

$$\sigma(X) \leq \eta_\delta^2 + \text{tr}(X^\top V X) - 2\text{tr}(X^\top B(Y) Y) \equiv \tau(X, Y) \quad (18)$$

Obsérvese que $\tau(X, Y)$ es una forma cuadrática en X , cuya minimización es directa.

3.3 La Transformada de Guttman

Para encontrar el minimizador $X^{(t+1)}$ de $\tau(X, X^{(t)})$, calculamos el gradiente respecto a X e igualamos a cero:

$$\nabla_X \tau(X, Y) = 2VX - 2B(Y)Y = 0 \quad (19)$$

Esto conduce al sistema de ecuaciones normales $VX = B(Y)Y$. Dado que V es singular (tiene al menos un autovalor nulo correspondiente al vector constante, reflejando la invariancia traslacional), utilizamos la inversa generalizada de Moore-Penrose V^+ .

La solución, conocida como la *Transformada de Guttman*, define el paso iterativo del algoritmo:

$$X^{(t+1)} = V^+ B(X^{(t)}) X^{(t)} \quad (20)$$

Esta actualización garantiza que $\sigma(X^{(t+1)}) \leq \sigma(X^{(t)})$, asegurando la convergencia hacia un mínimo local estacionario.

4 Generalizaciones y Restricciones Configuracionales

El marco teórico de SMACOF admite una flexibilidad notable para incorporar información **a priori** o manejar estructuras de datos no estándar. Analizamos aquí tres extensiones críticas: la imposición de restricciones lineales, el modelado de diferencias individuales y el análisis de matrices rectangulares.

4.1 Optimización bajo Restricciones Lineales Externas

En el MDS confirmatorio, se puede restringir el espacio de búsqueda de la configuración X a un subespacio lineal generado por un conjunto de variables predictoras. Sea $Z \in \mathbb{R}^{n \times q}$ ($q \geq p$) una matriz de diseño fija. Imponemos la restricción dura:

$$X = ZC \quad (21)$$

donde $C \in \mathbb{R}^{q \times p}$ es la matriz de coeficientes a estimar.

El problema de optimización se transforma en minimizar $\sigma(ZC)$ respecto a C . La función mayorizante $\tau(X, Y)$ derivada en la Sección 3 sigue siendo válida, pero ahora debe minimizarse sobre el conjunto restringido $\Omega = \{X \mid X = ZC\}$. Esto equivale a resolver un problema de mínimos cuadrados generalizados en cada iteración.

La actualización se obtiene proyectando la Transformada de Guttman no restringida, $\bar{X}^{(t)}$, sobre el espacio columna de Z respecto a la métrica inducida por V . La regla de actualización para los coeficientes es:

$$C^{(t+1)} = (Z^\top VZ)^{-1} Z^\top V \bar{X}^{(t)} \quad (22)$$

Y la configuración actualizada resulta $X^{(t+1)} = ZC^{(t+1)}$. Nótese que si V es singular, la inversa debe interpretarse en el sentido de Moore-Penrose, aunque típicamente $Z^\top VZ$ es invertible si Z tiene rango completo y V es conexa.

4.2 Escalamiento de Diferencias Individuales (3-Way SMACOF)

Consideremos ahora el caso de K matrices de disimilitud $\{\Delta_k\}_{k=1}^K$, correspondientes a diferentes jueces o condiciones. Buscamos K configuraciones X_k que compartan una estructura común subyacente. Definimos la supermatriz de configuraciones $X^* = [X_1^\top, \dots, X_K^\top]^\top$.

El modelo INDSCAL asume que cada configuración individual es una deformación lineal de un espacio común Z mediante pesos dimensionales específicos C_k (diagonal):

$$X_k = ZC_k, \quad C_k = \text{diag}(c_{11,k}, \dots, c_{pp,k}) \quad (23)$$

La función de Stress global es la suma ponderada de los stresses individuales:

$$\sigma(X^*) = \sum_{k=1}^K \sum_{i < j} w_{ij,k} (\delta_{ij,k} - d_{ij}(ZC_k))^2 \quad (24)$$

La minimización requiere un procedimiento de bloques alternados, actualizando iterativamente el espacio común Z y las matrices de pesos C_k .

4.3 Unfolding Métrico para Matrices Rectangulares

Cuando la matriz de disimilitud Δ es rectangular ($n_1 \times n_2$), representando, por ejemplo, preferencias de n_1 sujetos sobre n_2 objetos, el problema se modela como un grafo bipartito. La matriz de configuración se partitiona en $X_1 \in \mathbb{R}^{n_1 \times p}$ (sujetos) y $X_2 \in \mathbb{R}^{n_2 \times p}$ (objetos).

La matriz de pesos W adopta una estructura de bloques con ceros en la diagonal principal de bloques, reflejando la ausencia de disimilitudes intra-conjunto:

$$W = \begin{bmatrix} 0 & W_{12} \\ W_{12}^\top & 0 \end{bmatrix} \quad (25)$$

Consecuentemente, las matrices V y $B(X)$ heredan esta estructura de bloques particionada. La Transformada de Guttman se aplica sobre la configuración conjunta $X = [X_1^\top, X_2^\top]^\top$, utilizando la inversa generalizada V^+ adaptada a la conectividad del grafo bipartito.

5 Variantes No Métricas y el Problema de la Regresión Isotónica

En la formulación métrica clásica, se asume que las disimilitudes δ_{ij} residen en una escala de razón o intervalo. Sin embargo, en numerosas aplicaciones psicométricas y de ciencias sociales, Δ representa información puramente ordinal. El MDS no métrico relaja la hipótesis de linealidad, transformando el problema en la búsqueda de una configuración que preserve el orden de rango de las observaciones.

5.1 Formulación del Problema de Optimización Ordinal

Sea $\Omega = \{(i, j) : 1 \leq i < j \leq n\}$ el conjunto de pares de índices. El MDS no métrico postula la existencia de una transformación monótona desconocida $f : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$ tal que las *disparidades* generadas, $\hat{d}_{ij} = f(\delta_{ij})$, aproximan las distancias Euclídeas $d_{ij}(X)$.

El funcional de pérdida se generaliza para depender simultáneamente de la configuración X y de la matriz de disparidades \hat{D} :

$$\sigma_{nm}(X, \hat{D}) = \sum_{(i,j) \in \Omega} w_{ij} (\hat{d}_{ij} - d_{ij}(X))^2 \quad (26)$$

La minimización de σ_{nm} está sujeta a que \hat{D} pertenezca al cono convexo de matrices monótonas \mathcal{K}_Δ , definido por la restricción de orden débil:

$$\mathcal{K}_\Delta = \{\hat{D} \in \mathbb{R}^{n \times n} \mid \forall (i, j), (k, l) \in \Omega : \delta_{ij} < \delta_{kl} \implies \hat{d}_{ij} \leq \hat{d}_{kl}\} \quad (27)$$

Para evitar la solución trivial $X = 0, \hat{D} = 0$, se impone una restricción de normalización sobre la norma de las disparidades, típicamente $\|\hat{D}\|_W^2 = \sum w_{ij} \hat{d}_{ij}^2 = n(n-1)/2$.

5.2 El Algoritmo de Relajación por Bloques

La estrategia de optimización SMACOF para el caso no métrico emplea un enfoque de *Alternating Least Squares* (ALS), iterando entre dos subproblemas convexos condicionales:

- Actualización de la Configuración (Paso Guttman):** Dada la matriz de disparidades $\hat{D}^{(t)}$ fija en la iteración t , minimizamos respecto a X . Esto se reduce al problema métrico estándar, donde $\hat{D}^{(t)}$ actúa como las pseudo-disimilitudes objetivo.

$$X^{(t+1)} = \arg \min_X \sigma(X, \hat{D}^{(t)}) \implies X^{(t+1)} = V^+ B(\hat{D}^{(t)}) X^{(t)} \quad (28)$$

- Actualización de las Disparidades (Regresión Isotónica):** Dada la configuración actualizada $X^{(t+1)}$ y sus distancias asociadas $D^{(t+1)}$, buscamos la matriz $\hat{D}^{(t+1)} \in \mathcal{K}_\Delta$ que minimice el error cuadrático ponderado. Esto constituye un problema de proyección ortogonal sobre un cono convexo:

$$\hat{D}^{(t+1)} = \arg \min_{\hat{D} \in \mathcal{K}_\Delta} \sum_{(i,j) \in \Omega} w_{ij} (\hat{d}_{ij} - d_{ij}(X^{(t+1)}))^2 \quad (29)$$

Este subproblema de mínimos cuadrados restringidos por orden se resuelve mediante *Regresión Isotónica*. El algoritmo estándar implementado en SMACOF es el PAVA (*Pooled-Adjacent-Violators Algorithm*). El PAVA opera identificando bloques de violaciones de monotonicidad (donde $\hat{d}_{ij} > \hat{d}_{kl}$ a pesar de que $\delta_{ij} < \delta_{kl}$) y sustituyendo dichos valores por su media ponderada, garantizando la satisfacción de las restricciones de Karush-Kuhn-Tucker.

5.3 Tratamiento Riguroso de Empates (Ties)

La estructura discreta de las disimilitudes puede inducir relaciones de equivalencia $\delta_{ij} = \delta_{kl}$. El tratamiento de estos empates define la topología exacta del cono \mathcal{K}_Δ . Distinguimos dos enfoques principales:

5.3.1 Enfoque Secundario (Restricción Fuerte)

Bajo este enfoque, la transformación f debe preservar la igualdad estricta. Si $\delta_{ij} = \delta_{kl}$, entonces se impone $\hat{d}_{ij} = \hat{d}_{kl}$. Esto es equivalente a realizar la regresión isotónica sobre los promedios de los bloques de empates, reduciendo la granularidad del ajuste.

$$\mathcal{K}_\Delta^{sec} = \mathcal{K}_\Delta \cap \{\hat{D} \mid \delta_{ij} = \delta_{kl} \implies \hat{d}_{ij} = \hat{d}_{kl}\} \quad (30)$$

5.3.2 Enfoque Primario (Restricción Débil)

Este es el enfoque predeterminado en la mayoría de las implementaciones modernas. Permite que la transformación rompa los empates para optimizar el ajuste. La condición se relaja a una desigualdad no estricta dentro de los bloques de empates. Si $\delta_{ij} = \delta_{kl}$, el orden relativo entre \hat{d}_{ij} y \hat{d}_{kl} es libre y determinado únicamente por la magnitud de las distancias $d_{ij}(X)$ y $d_{kl}(X)$. Formalmente, el algoritmo PAVA se aplica solo entre bloques de disimilitudes distintas, permitiendo variación intra-bloque.

La convergencia global del algoritmo SMACOF no métrico está asegurada, ya que tanto el paso de actualización de configuración (mayorización) como el paso de actualización de disparidades (proyección) reducen monótonamente el valor de la función de Stress.