

5/154



Dpto. de Estadística e I.O.  
Universidad de Granada

UNIVERSIDAD DE GRANADA
Facultad de Ciencias
Fecha 24/5/00.....
ENTRADA NUM. 1826.....

## Tesis Doctoral

---

# *Dimensionalidad en Multidimensional Scaling y su tratamiento computacional*

---

**Yolanda Román Montoya**  
**Granada, 2000**

UNIVERSIDAD DE GRANADA
11 MAYO 2000
COMISION DE DOCTORADO

Tesis Doctoral

Yolanda Román Montoya

# Dimensionalidad en Multidimensional Scaling y su tratamiento computacional

Memoria presentada para optar al  
grado de Doctor en Ciencias por  
Yolanda Román Montoya



Yolanda Román Montoya

Vº Bº  
Director de Tesis:

Prof. Dr. D. Andrés González Carmona

Vº Bº  
Director de Tesis:

Prof. Dr. D. José Fernando Vera Vera

Departamento de Estadística e Investigación Operativa

Universidad de Granada

*A mi abuela*

Deseo expresar mi más profundo agradecimiento a todos aquellos que me han ayudado durante la realización de este trabajo:

A mis padres, a quienes debo todo lo que soy.

A mis directores de tesis, el prof. Dr. D. Andrés González Carmona y el Dr. D. José Fernando Vera Vera por sus aportaciones en la elaboración de esta memoria.

Al prof. Dr. Wilem Heiser, de la Universidad de Leiden, Holanda, por su seguimiento y colaboración.

A mis compañeros del Departamento de Estadística e I.O. de la Universidad de Granada por su continuo apoyo y, en especial, al director del Departamento, el prof. Dr. D. Ramón Gutiérrez Jáimez, por su desinteresado apoyo y prestación de los medios necesarios para la elaboración de esta investigación

# Índice

<b>I</b>	<b>Multidimensional Scaling</b>	<b>1</b>
1	Introducción . . . . .	1
2	Modelos en MDS . . . . .	3
2.1	Método clásico de MDS . . . . .	4
2.2	MDS no métrico. Algoritmo de Shepard-Kruskal . . . . .	8
2.3	Modelos de diferencias individuales . . . . .	10
2.4	Modelos confirmatorios en MDS . . . . .	14
2.5	Análisis Procrustiano General: GPA . . . . .	17
<b>II</b>	<b>Situación actual de las técnicas MDS</b>	<b>19</b>
1	Nuevas técnicas de minimización. El algoritmo SMACOF . . . . .	19
1.1	Mayorización de la función STRESS . . . . .	21
2	Determinación de mínimos globales. Algoritmo Tunneling . . . . .	25
2.1	Clasificación del campo de la optimización global . . . . .	25
2.2	Algoritmo tunneling . . . . .	28
3	Análisis de grandes conjuntos de datos: CDS . . . . .	29
4	Aplicación de las técnicas con gráficos móviles a MDS . . . . .	35
4.1	Análisis de estructuras en dimensiones superiores. Técnicas Visuales/PXPL . . . . .	36
4.2	Análisis de sensibilidad de las soluciones MDS . . . . .	36
5	Contraste de configuraciones. Técnicas GPA . . . . .	37
5.1	Análisis de la variabilidad en GPA con datos faltantes . . . . .	39
<b>III</b>	<b>Determinación de la dimensionalidad en MDS</b>	<b>45</b>
1	Modelo anidado para MDS métrico según el algoritmo SMACOF . . . . .	51
1.1	Convergencia del modelo anidado para MDS métrico . . . . .	53
2	Modelo anidado para las soluciones MDS no métricas . . . . .	54

2.1	Función de pérdida SSTRESS para el algoritmo anidado . . . . .	55
2.2	Algoritmo para la obtención de soluciones anidadas . . . . .	56
2.3	Convergencia del modelo anidado no métrico . . . . .	63
<b>IV Análisis de los resultados obtenidos en el modelo anidado para MDS</b>		
<b>no métrico</b>		<b>69</b>
1	Análisis de los valores de SSTRESS . . . . .	71
2	Contraste de las soluciones anidadas. Análisis Procrustiano General . . .	73
2.1	Análisis procrustiano de las configuraciones anidadas . . . . .	74
2.2	Análisis procrustiano entre la configuración real y las diferentes configuraciones anidadas . . . . .	74
3	Análisis de la estabilidad del modelo . . . . .	78
3.1	Análisis de robustez del modelo para MDS no métrico anidado . .	82
<b>A Programas</b>		<b>87</b>
1	Programa para la determinación de configuraciones anidadas mediante MDS no métrico . . . . .	87
2	Programa para la determinación de las medidas que establecen la robustez del modelo . . . . .	105

# Capítulo I

## Multidimensional Scaling

### 1 Introducción

El conjunto de técnicas estadísticas englobadas bajo el término *Multidimensional Scaling* (MDS) constituye una parte del análisis exploratorio de datos que permite la obtención de estructuras multidimensionales a partir del análisis de datos de proximidad entre objetos.

Dado un conjunto  $\Omega$  formado por  $n$  objetos o estímulos se llama *disimilaridad* entre dos de ellos,  $i$  y  $j$ , a una medida de proximidad para dicho par,  $\delta_{ij}$ , que representa el grado de *desemejanza* entre esos objetos respecto de un conjunto de características cuantitativas o cualitativas, de forma que valores pequeños de  $\delta_{ij}$  indiquen una gran semejanza y cuanto mayores sean las diferencias, mayor sea el valor de la *disimilaridad*. En general, una *disimilaridad* será una función real, verificando las siguientes condiciones:

1.  $\delta_{ij} \geq \delta_0 \quad \forall i, j$
2.  $\delta_{ii} = \delta_0 \quad \forall i$
3.  $\delta_{ij} = \delta_{ji} \quad \forall i, j$
4.  $\delta_0 \geq 0$ .

Existen gran cantidad de métodos que permiten obtener datos de disimilaridad en MDS, entre ellos podemos citar:

- *Métodos de clasificación directa.* Consisten en utilizar escalas del tipo *muy disimilar-poco disimilar* para objetos que no pueden ser diferenciados por algún procedimiento cuantitativo. Una vez realizadas todas las clasificaciones, la disimilaridad

entre dos objetos  $i$  y  $j$  será, generalmente, la media de los valores que un grupo de sujetos le haya atribuido.

- *Frecuencias de clasificación.* Supongamos que se cuenta con  $K$  individuos para realizar el análisis. Este método consiste en que los  $K$  individuos formen grupos de objetos en función de la semejanza entre estos. Si  $N_{ij}$  es el número de sujetos que han incluido los objetos  $i$  y  $j$  en el mismo grupo, la disimilaridad vendrá dada por

$$\delta_{ij} = 1 - \frac{N_{ij}}{N},$$

siendo  $N$  el número total de objetos que se están considerando.

- *Métodos cuantitativos.* En el caso de que se trate con objetos que pueden ser caracterizados mediante variables de tipo cuantitativo, se pueden utilizar diferentes medidas, por ejemplo las distancias de Minkowski:

$$\delta_{ij} = \left[ \sum_{a=1}^p |x_{ia} - x_{ja}|^q \right]^{1/q}, \quad 1 \leq q \leq \infty.$$

Los datos con los que se trabaja en MDS, relativos a objetos o estímulos e individuos o sujetos, se pueden clasificar atendiendo a distintos criterios:

1. NIVEL DE MEDIDA. Los datos de partida en un estudio con MDS pueden considerarse en cualquiera de los cuatro niveles de medida habituales:
  - Nominal. Los objetos se clasifican solamente en grupos.
  - Ordinal. Los objetos se organizan según un orden de magnitud.
  - Intervalo. Los objetos se sitúan en una escala de forma que el tamaño de las diferencias entre estos queda reflejado por la escala, en la que no se consideran ceros reales.
  - Razón. Los objetos se sitúan en una escala de forma que su posición en ésta representa el valor exacto del atributo que se mide.
2. NÚMERO DE MODOS. Se denomina modo a cada conjunto de elementos subyacente en los datos para el análisis con MDS. Si, por ejemplo, se consideran medidas de juicios de disimilaridad entre pares de objetos emitidas por distintos sujetos, se hablará de datos bimodales (los objetos y los individuos).

3. NÚMERO DE VÍAS. Se denomina vía a cada índice en la medida entre objetos.

Dada una matriz,  $\Delta = (\delta_{ij})_{n \times n}$ , de disimilaridades obtenidas entre  $n$  estímulos u objetos de un conjunto, el problema que resuelven las técnicas MDS consiste fundamentalmente en la representación de los  $n$  objetos analizados, como puntos en un espacio de dimensión  $p$ .

$$X = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix},$$

de manera que la distancia entre el punto  $i$  y el punto  $j$ ,  $d_{ij}$ , sea lo más parecida posible al valor observado de la disimilaridad,  $\delta_{ij}$ .

En general, las técnicas de escalamiento y análisis dimensional se refieren a una amplia variedad de estrategias de investigación y procedimientos que proporcionan disposiciones visuales que son muy útiles para presentar resultados analíticos. El elemento común entre ellas es la obtención de representaciones de los datos cuantitativos y/o geométricos, a partir de dónde conseguir:

1. Reducciones de los datos. Es posible resumir un gran conjunto de variables con un pequeño número de medidas compuestas.
2. Análisis de la dimensionalidad, comprobando las fuentes subyacentes de variación en un conjunto de datos.
3. Mediciones a partir de la representación empírica de las dimensiones subyacentes (y generalmente no observables), que pueden ser empleadas como variables analíticas en otros procedimientos estadísticos.

## 2 Modelos en MDS

Son muchos los algoritmos que se han desarrollado para obtener la configuración que mejor se ajuste a las disimilaridades de partida. Estos algoritmos pueden clasificarse, atendiendo a un enfoque confirmatorio o exploratorio, en dos grandes grupos: modelos probabilísticos y modelos no probabilísticos respectivamente, distinguiendo en ambos casos entre modelos métricos, si las consideraciones entre distancias y disimilaridades son cuantitativas, y modelos no métricos, si son cualitativas.

El método de estimación empleado y las diferentes medidas de bondad de ajuste, son los principales factores que determinan la existencia de los diferentes métodos de MDS que se conocen en la actualidad. A continuación daremos un breve repaso de las técnicas que se han ido desarrollando para MDS desde su origen en 1936. La evolución de estas técnicas, tanto en el desarrollo de los algoritmos que los definen como en el tipo de datos con que se trabaja, está íntimamente ligada a su implementación computarizada ya que sin el uso del ordenador, el desarrollo de MDS sería prácticamente imposible debido a la naturaleza iterativa de los algoritmos.

## 2.1 Método clásico de MDS

Las técnicas que constituyen MDS tuvieron su origen en el campo de la psicología al pretender definir la relación entre la intensidad física de ciertos estímulos y su intensidad subjetiva. No obstante, gracias a los trabajos de Torgerson (1952), que definieron el problema que soluciona MDS y proporcionaron la primera solución métrica en un entorno multidimensional, MDS dejó de tener el matiz tan particular antes descrito para pasar a abordar problemas del análisis de datos y adquirir así un carácter más universal. Torgerson evidenció cómo los métodos psicofísicos de MDS presuponían el conocimiento de las dimensiones de las áreas que estaban siendo investigadas, información de la cual no se dispone en la mayoría de los casos, proporcionando resultados en términos de dimensiones que, en muchas ocasiones, no eran relevantes. La técnica que él propuso permite obtener soluciones en las cuales la dimensión, al igual que los valores de escala de los estímulos, se determina a partir de los datos originales.

El método descrito por Torgerson, conocido como método clásico, es utilizado hoy día por la mayoría de los métodos de MDS para la obtención de una configuración inicial. Puede sintetizarse en los siguientes puntos:

Considérese que se dispone de una matriz de disimilaridades entre  $n$  objetos que denotaremos por

$$\Delta = (\delta_{ij})_{n \times n}.$$

El método clásico interpreta las disimilaridades como distancias euclídeas, transformándolas si es necesario mediante un procedimiento de estimación denominado *método de la constante aditiva*, que proporciona nuevos valores,  $d_{ij}^* = \delta_{ij} + c$ , siendo

$$c = \max_{ijk} \{ \delta_{ij} - \delta_{ik} - \delta_{kj} \},$$

la menor constante que garantiza el cumplimiento de la desigualdad triangular para cualquier tripleta de estímulos del conjunto.

Una vez que se dispone de la matriz de distancias al cuadrado, que denotaremos por  $D$ , se calcula la matriz de productos escalares asociada,  $B$ , mediante la expresión:

$$B = -\frac{1}{2}JDJ,$$

siendo  $J = I - \frac{1}{n}\mathbf{1}\mathbf{1}'$  donde  $\mathbf{1}$  es un vector de unos de dimensión apropiada.

A continuación, el método de Torgerson calcula una matriz de coordenadas a partir de la descomposición en valores característicos de la matriz  $B$  de productos escalares

$$B = U\Lambda U^t$$

siendo  $\Lambda$  la matriz diagonal formada por las raíces características de  $B$ , y  $U$ , la matriz de vectores característicos correspondientes a cada una de las raíces características, normalizados para tener longitud 1.

Obtenida la descomposición, el siguiente paso es determinar la dimensión apropiada para la representación. Es claro que al menos un valor propio es 0, ya que  $B\mathbf{1} = U\Lambda U\mathbf{1} = 0$ . Luego, al menos habrá solución en dimensión  $n - 1$ , que vendrá dada, por construcción, en el sentido de los ejes principales

$$X = U\Lambda^{1/2}.$$

Si deseamos fijar la representación en  $p$  dimensiones, con  $p < \text{rango}(B)$ , el problema tendrá solución. Dado que en la descomposición espectral de la matriz  $B$ , las distancias entre los puntos en el espacio euclídeo  $(n - 1)$ -dimensional vienen dadas por:

$$d_{ij}^2 = \sum_{a=1}^{n-1} \lambda_a (x_{ia} - x_{ja})^2,$$

si los  $p$  primeros valores propios son suficientemente grandes, la contribución del resto de valores propios y por tanto del resto de dimensiones subyacentes será insignificante, por lo que el problema de representación en un espacio de dimensión baja quedaría resuelto.

Exponemos a continuación la demostración de este resultado.

**Teorema 2.1** *La matriz  $\hat{X} = U_p \Lambda_p^{1/2}$ , donde  $\Lambda_p$  es la matriz diagonal formada por las  $p$  mayores raíces características de  $B$ , y  $U_p$  la matriz de orden  $n \times p$  de los vectores característicos correspondientes a cada una de tales raíces características, define el mínimo de la medida*

$$\Psi = \sum_{r,s=1}^n (b_{rs} - \hat{b}_{rs})^2 = \text{tr} [(B - \hat{B})^2]$$

siendo  $B$  la matriz de productos escalares asociada a la matriz de disimilaridades inicial y  $\hat{B}$  la matriz de productos escalares asociada a la estimación realizada.

DEMOSTRACIÓN:

Sean  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$ , los  $n$  valores propios reales de  $B$  (semidefinida positiva),  $\Lambda_{n \times n}$  la matriz diagonal formada por esos valores propios y  $U_{n \times n}$  la matriz de los correspondientes vectores propios normalizados.

Sea  $\hat{X}$  una configuración cualquiera del conjunto de todas las configuraciones en dimensión  $p$ , y sea  $\hat{B}_{n \times n}$  su matriz de productos escalares, con  $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq \hat{\lambda}_n \geq 0$  los  $n$  valores propios de  $\hat{B}$ , y  $\hat{\Lambda}$  la matriz diagonal formada por esos valores propios.

Claramente, por la ortogonalidad de  $U$ , la matriz  $U^t \hat{B} U$  tiene los mismos valores propios que  $\hat{B}$ :

$$|U^t \hat{B} U - \lambda I| = |U^t \hat{B} U - \lambda U^t U| = |U^t ||\hat{B} - \lambda I||U| = |\hat{B} - \lambda I|.$$

Por otra parte,  $(U^t \hat{B} U)^t = U^t \hat{B} U$ , luego, en virtud del teorema de descomposición espectral de una matriz,

$$U^t \hat{B} U = G \hat{\Lambda} G^t,$$

siendo  $G$  la matriz de vectores propios asociados a los valores  $\hat{\lambda}_i$ .

Entonces:

$$\begin{aligned} \Psi &= \text{tr} [(B - \hat{B})^2] = \text{tr} [(B - \hat{B})^2 U U^t] = \text{tr} [U^t (B - \hat{B})^2 U] = \\ &= \text{tr} [U^t (B - \hat{B})(B - \hat{B})U] = \text{tr} [U^t (B - \hat{B})U U^t (B - \hat{B})U] = \\ &= \text{tr} [(U^t (B - \hat{B})U) (U^t (B - \hat{B})U)] = \end{aligned}$$

$$\begin{aligned} &= \text{tr} \left[ (U^t B U - U^t \hat{B} U) (U^t B U - U^t \hat{B} U) \right] = \\ &= \text{tr} \left[ (\Lambda - G \hat{\Lambda} G^t)^2 \right]. \end{aligned}$$

Supuesta fija  $\hat{\Lambda}$ , veamos que  $\Psi$  se minimiza sobre  $G$  cuando  $G = I$ :

$$\text{tr} \left[ (\Lambda - G \hat{\Lambda} G^t)^2 \right] = \text{tr} [\Lambda^2] - 2 \text{tr} [\Lambda G \hat{\Lambda} G^t] + \text{tr} [(G \hat{\Lambda} G^t)^2],$$

por lo que minimizar  $\text{tr} \left[ (\Lambda - G \hat{\Lambda} G^t)^2 \right]$  sobre  $G$ , equivale a maximizar

$$\text{tr} [\Lambda G \hat{\Lambda} G^t] = \sum_{i=1}^n \sum_{j=1}^n \lambda_i \hat{\lambda}_j g_{ij}^2.$$

Ahora bien,  $G$  es una matriz ortogonal,  $G G^t = I$ , por lo que, si llamamos  $C = G G^t$ , se verifica:

$$\begin{aligned} c_{ij} &= \sum_{a=1}^n g_{ia} g_{ja} \quad \forall i \neq j \\ c_{ii} &= \sum_{a=1}^n g_{ia}^2 = 1, \end{aligned}$$

de donde se deduce que  $g_{ij}^2 \leq 1 \quad \forall i, j$ . De modo que se obtiene la cota superior:

$$\sum_{i=1}^n \sum_{j=1}^n \lambda_i \hat{\lambda}_j g_{ij}^2 \leq \sum_{i=1}^n \sum_{j=1}^n \lambda_i \hat{\lambda}_j,$$

para cualquier matriz ortogonal  $G$ . Ahora bien, la matriz identidad  $I$  es ortogonal y verifica que:

$$\text{tr} [\Lambda I \hat{\Lambda} I^t] = \sum_{i=1}^n \sum_{j=1}^n \lambda_i \hat{\lambda}_j$$

luego la expresión se maximiza para  $G = I$ , con lo que  $\Psi$  se minimiza para  $G = I$ . A continuación, se minimiza  $\Psi$  en función de  $\hat{\Lambda}$ , quedando su expresión de la forma:

$$\Psi = \text{tr} [\Lambda - \hat{\Lambda}] = \sum_{i=1}^n (\lambda_i - \hat{\lambda}_i)^2.$$

Puesto que la configuración arbitraria,  $\hat{X}$ , de que partimos tiene dimensión  $p$ , su matriz de productos escalares centrada,  $\hat{B} = J\hat{X}\hat{X}^tJ$ , tendrá al menos  $p$  valores propios no nulos y por tanto positivos. Es claro que  $\Psi$  se minimiza respecto a  $\hat{\Lambda}$ , cuando

$$\hat{\lambda}_i = \begin{cases} \lambda_i, & i = 1, 2, \dots, p \\ 0, & i = p + 1, p + 2, \dots, n \end{cases}$$

Por tanto  $\hat{B}_p = U_p\Lambda_pU_p^t$ , por lo que  $\hat{X}$  puede tomarse igual a  $U_p\hat{\Lambda}_p^{1/2}$ , que es la solución del procedimiento clásico de MDS. Además, el mínimo valor de  $\Psi$  es

$$\min \Psi = \lambda_{p+1}^2 + \dots + \lambda_n^2.$$

C.Q.D.

## 2.2 MDS no métrico. Algoritmo de Shepard-Kruskal

Una limitación importante del modelo propuesto por Torgerson es la hipótesis sobre el nivel de medida utilizado para la determinación de los datos. Torgerson supone que los datos están medidos en una escala tipo intervalo, es decir, supone que el sujeto utiliza una unidad de medida constante cuando realiza los juicios de disimilaridad entre objetos, lo que, cuando menos, es altamente restrictivo.

En la década de los 60 aparecen los trabajos de Shepard (1962) y Kruskal (1964) sobre las técnicas no métricas de MDS. La importancia del trabajo de Shepard reside no sólo en la obtención de configuraciones de objetos a partir de datos procedentes de escalas ordinales, sino principalmente en la corriente investigadora que se originó en torno a este tema. Fruto de esta labor de investigación y de la aplicación de su técnica en campos de la ciencia tan diversos como la Arquitectura, la Zoología, las Ciencias Políticas, etc., los procedimientos de MDS cobraron una amplia popularidad y se difundieron rápidamente. El trabajo de Shepard puede definirse como la obtención de información métrica a partir de resultados no métricos determinando, en un número mínimo de dimensiones del espacio euclídeo, la representación de los objetos sometidos a estudio. La configuración obtenida ha de ser tal que las distancias entre los puntos representados estén relacionadas monótonamente con las medidas de proximidad no métricas dadas inicialmente.

Shepard realizó una aportación brillantísima a las técnicas de MDS, pero su procedimiento presentaba ciertos inconvenientes, ya que requería obtener la representación de

$n$  objetos en  $n - 1$  dimensiones para luego ir disminuyendo la dimensionalidad de la configuración. En esta misma línea destacan los trabajos de Kruskal, en los que se propone una solución al problema de MDS no métrico desde un punto de vista completamente diferente al de Shepard. Kruskal determinó un procedimiento que permite obtener la mejor representación posible de  $n$  objetos en un espacio  $p$ -dimensional especificado por el usuario antes de realizar el análisis, con  $p < n - 1$ . Mientras que Shepard aborda el problema desde una perspectiva de trabajo similar a la definida por el análisis factorial, Kruskal lo ve como un problema de optimización en el que es posible obtener la mejor descripción posible de los datos en un espacio de dimensión conocida. Para ello, Kruskal toma como criterio de minimización el criterio de mínimos cuadrados, definiendo una función de pérdida dada por la suma de los cuadrados de las diferencias entre las distancias obtenidas a partir de los puntos de la configuración y transformaciones monótonas de las proximidades de partida.

En las siguientes líneas describimos brevemente el método de Shepard-Kruskal:

Dada una ordenación estrictamente monótona de las disimilaridades obtenidas a partir del análisis de  $n$  objetos,

$$\delta_{i_1, j_1} < \delta_{i_2, j_2} < \dots < \delta_{i_h, j_h},$$

con  $h = n(n - 1)/2$ , y dada una función monótona creciente  $f$ , el método de Shepard-Kruskal consiste en encontrar una configuración euclídea cuyas distancias,  $d_{ij}$ , se ajusten a las disimilaridades transformadas

$$\hat{d}_{ij} = f(\delta_{ij})$$

denominadas *disparidades*, de manera que conservando la ordenación original, se minimize la función de pérdida,

$$S^* = \frac{\sum_{i < j} (d_{ij} - \hat{d}_{ij})^2}{\sum_{j=1}^n \sum_{i=1}^{j-1} d_{ij}^2},$$

normalizada para evitar imprecisiones por la escala de medida utilizada.

De este modo, para cada función monótona creciente se obtiene una expresión de la medida de ajuste, que puede definirse del siguiente modo:

**Definición 2.2** Se denomina *f*-STRESS asociado a la función *f* a la cantidad

$$S_f = \frac{\sum_{i < j} (d_{ij} - \hat{d}_{ij})^2}{\sum_{i < j} d_{ij}^2},$$

de forma que  $0 \leq S_f \leq 1$  y  $S_f = 0 \Leftrightarrow d_{ij} = f(\delta_{ij}) \quad \forall (i, j)$ .

**Definición 2.3** Se denomina STRESS al valor

$$S = \min_f S_f.$$

Esta medida de ajuste se puede interpretar como una medida de monotonía, que obviamente disminuirá a medida que aumente la dimensión de la configuración euclídea (en dimensión  $n - 2$  es posible obtener una solución MDS no métrica exacta).

Una vez definida la función de pérdida, el siguiente paso es su minimización. Shepard y Kruskal definen un procedimiento iterativo basado en la minimización del STRESS según la función monótona creciente *f* y la configuración euclídea en dimensión *p*. Partiendo de una configuración inicial (arbitraria o no), cada punto se desplaza según la dirección del gradiente negativo del STRESS, repitiéndose el procedimiento hasta que el valor de *S* se estabiliza en un mínimo local. El ajuste de la función *f*, se lleva a cabo aplicando técnicas de regresión.

### 2.3 Modelos de diferencias individuales

A finales de la década de los 60 e inicios de los 70 se desarrolló una nueva vertiente en MDS enfocada a solucionar el problema de las diferencias individuales en los análisis. Una de las principales limitaciones de los procedimientos MDS conocidos hasta esa fecha era que el investigador tan solo podía analizar una única matriz de datos de proximidad. En el caso de que se dispusiera de más información solo había dos alternativas: Hacer un análisis diferente para cada una de las matrices de datos, o promediar los datos y realizar un análisis MDS general con la matriz promedio.

El primer procedimiento de diferencias individuales en MDS fue el procedimiento de los *Puntos de Vista* de Tucker y Messick (1963). Obtenidos los juicios de similitud para cada sujeto, se calcula la matriz de correlaciones entre los juicios de cada individuo con los restantes que participan en el análisis. Esa matriz es analizada factorialmente,

obteniéndose un espacio de representación para los sujetos a partir del cual el investigador interpreta las diferencias individuales. En la mayoría de los casos el investigador intenta identificar clústeres de individuos, de modo que los individuos dentro de un clúster estén altamente correlados. Una vez que el investigador determina tales clústeres, el procedimiento genera conjuntos de juicios hipotéticos para cada uno de ellos, que pueden ser considerados como una media de los juicios hechos por los individuos del clúster correspondiente en un análisis aislado.

En 1968, McGee define un nuevo procedimiento de diferencias individuales en el que no se realiza ningún tipo de promedio de los datos. La solución de McGee se centra fundamentalmente en dos tipos de diferencias individuales conocidas actualmente como diferencias en el estilo de respuesta y diferencias en el estilo perceptivo/cognitivo. El procedimiento de McGee consiste en una técnica no métrica capaz de analizar varias matrices de datos simultáneamente, que permite al usuario elegir entre una transformación monótona separada para cada matriz o una única transformación común a todas las matrices de datos. La propuesta de McGee define un mecanismo con el que se puede controlar el grado de relación entre las representaciones para cada individuo, teniendo como casos extremos la no relación y la identidad. Este mecanismo está bajo el control del usuario, siendo el procedimiento incapaz de determinar el mejor grado de relación. Además, cuando se supone que tanto el estilo de respuesta como el estilo perceptivo/cognitivo son idénticos para todos los individuos, la técnica de McGee es equivalente a promediar todas las matrices de datos y realizar un único MDS no métrico.

El siguiente paso y quizás el más importante en el desarrollo de las diferencias individuales, fue el procedimiento dado por Carroll y Chang (1970). El modelo de estos autores proporciona en primer lugar, un espacio de atributos común que describe los objetos analizados por el grupo total de sujetos y, en segundo lugar, las diferentes apreciaciones de los individuos respecto a las semejanzas entre los atributos. Las diferencias individuales se representan mediante una generalización del modelo euclídeo con ponderaciones de las dimensiones del espacio de atributos. Cada individuo tiene un vector de pesos, cuyas componentes reflejan la importancia que el individuo da a cada uno de los ejes de coordenadas.

El modelo propuesto por Carroll y Chang es conocido como INDSCAL (*INDividual Differences SCALing*). En dicho modelo se parte de  $K$  matrices de disimilaridades;  $\Delta_1, \Delta_2, \dots, \Delta_K$ ; una para cada sujeto, todas ellas entre  $n$  objetos en estudio. La métrica natural asociada a los puntos del espacio común subyacente de dimensión  $p$ , será la euclídea ponderada, denotándose por  $d_{ijk}$ ,

$$d_{ijk} = \left( \sum_{a=1}^p w_{ka} (x_{ia} - x_{ja})^2 \right)^{1/2},$$

la distancia entre el objeto  $i$  y el  $j$  para el individuo  $k$ .

Los datos,  $\delta_{ijk}$ , observados, se relacionarán con las distancias,  $d_{ijk}$ , a través de funciones monótonas en la forma habitual

$$f_k(\delta_{ijk}) \cong d_{ijk}.$$

A partir de la definición de la métrica euclídea ponderada, se obtiene la relación

$$d_{ijk}^2 = \sum_{a=1}^p w_{ka} (x_{ia} - x_{ja})^2 = \sum_{a=1}^p \left( w_{ka}^{1/2} x_{ia} - w_{ka}^{1/2} x_{ja} \right)^2,$$

que puede expresarse como:

$$d_{ijk}^2 = \sum_{a=1}^p \left( y_{ia}^{(k)} - y_{ja}^{(k)} \right)^2,$$

con  $y_{ia}^{(k)} = w_{ka}^{1/2} x_{ia}$ , de donde se deduce una interpretación conjunta de la configuración, a la vez que se ofrece la posibilidad de efectuar una descomposición en productos escalares análoga a la realizada en el método métrico clásico.

De este modo, la solución ofrecida por INDSCAL abarca dos conceptos:

- Una matriz de configuración,  $X_{n \times p}$ , que permite una representación común de los  $n$  objetos (*espacio común de los objetos*) en dimensión  $p$ .
- Una matriz de pesos,  $W_{K \times p}$ , cuyas filas constituyen una representación euclídea de las diferencias individuales entre los  $K$  sujetos (*espacio de los sujetos*) en dimensión  $p$ .

Combinando  $X$  y  $W$ , para cada sujeto,  $k$ , se obtiene  $Y^{(k)} = \left( y_{ia}^{(k)} \right)_{n \times p}$ , que define la representación euclídea de los objetos para dicho individuo (*espacio perceptivo del sujeto  $k$* ).

Muchos de los desarrollos que se han citado fueron resumidos en un único trabajo debido a Takane, Young y De Leeuw (1977) y Young, Takane y Lewyckij (1978): ALSCAL. ALSCAL es un algoritmo que, mediante mínimos cuadrados alternantes, permite desarrollar *Multidimensional Scaling* y *Multidimensional Unfolding*. ALSCAL es un modelo

que permite analizar datos mediante un modelo euclídeo simple o ponderado que puede ser: (a) nominal, ordinal, intervalo o razón; (b) con observaciones faltantes; (c) simétrico o asimétrico; (d) condicional o incondicional; (e) con o sin réplicas y (f) discreto o continuo. Así pues, puede considerarse como el más completo, ya que es el único que permite realizar análisis métricos y no métricos, siendo sus resultados métricos similares, e incluso mejores, en el sentido de proximidad a la obtención de mínimos globales, a los obtenidos por INDSCAL en el caso de diferencias individuales. En el caso de ajustar modelos euclídeos simples sus resultados son equivalentes a los del modelo de Kruskal. Dada su flexibilidad, además de la rutina original de los autores en FORTRAN, este algoritmo ha sido implementado en los paquetes estadísticos SAS y SPSS, de ahí que su utilización se haya difundido pasando a ser el modelo más empleado.

ALSCAL utiliza como función de bondad de ajuste la función SSTRESS, definida como un criterio de mínimos cuadrados sobre los cuadrados de las distancias y las disparidades

$$SS(X, W, D^*) = \sum_{k=1}^N \sum_{i=1}^n \sum_{j=1}^{i-1} (d_{ijk}^{*2} - d_{ijk}^2)^2.$$

El algoritmo ALSICAL se define mediante un proceso iterativo en el que cada iteración está constituida por dos etapas que van alternándose hasta obtener la convergencia global del proceso:

1. *Fase de representación óptima.* Se obtienen las estimaciones por mínimos cuadrados y las observaciones representadas de manera óptima,  $D^*$ , bajo la suposición de que la configuración,  $X$ , y los pesos,  $W$ , son constantes. Es decir, se resuelve el problema de mínimos cuadrados condicionados que minimiza la función SSTRESS bajo la suposición de que  $X$  y  $W$  no son variables. Matemáticamente el problema se formula como

$$\min_{D^*} [SS(D^*/X, W)]$$

2. *Fase de estimación del modelo.* Engloba dos subfases de minimización separadas:

- (a) Cálculo de los pesos de la configuración, supuestas constantes las disparidades y la configuración:

$$\min_W [SS(W/D^*, X)].$$

- (b) Estimación de las coordenadas de la configuración, suponiendo fijos los valores de los pesos estimados en la etapa anterior, así como las disparidades calculadas en la fase de representación óptima:

$$\min_{\mathbf{X}}[SS(X/W, D^*)].$$

## 2.4 Modelos confirmatorios en MDS

A finales de los 70 e inicios de los 80, Ramsay introdujo el procedimiento de estimación de máxima verosimilitud en MDS. Estas técnicas permitieron a los modelos MDS pasar de una orientación simplemente descriptiva al campo de la inferencia estadística. Asociada a este cambio está la introducción de contrastes de hipótesis y las medidas descriptivas AIC y BIC para determinar la convergencia del proceso, la dimensionalidad apropiada de la representación y el modelo de error que siguen los datos. Estas técnicas proporcionan también regiones de confianza para el valor de las coordenadas que representan a los estímulos en la configuración así como para los vectores que definen a los sujetos en el caso ponderado. El origen de los modelos de máxima verosimilitud en MDS es la naturaleza del error en los datos de entrada. Si el modelo de error considerado es una representación apropiada de la situación empírica, los contrastes de significación que se obtendrán podrán ser significativos.

Son cinco los modelos de error propuestos hasta el momento, dos de ellos lo han sido por Ramsay, dos por Takane y el último por Zinnes y MacKay.

Los modelos de Ramsay y Takane son similares en su primera definición. Difieren en las suposiciones que hacen acerca de la influencia de los modelos sobre los datos. Ambos suponen que el error sigue una distribución normal (o lognormal), pero mientras que Ramsay trabaja para el caso métrico suponiendo que el proceso de error introduce valores de disimilaridad que siguen un determinado modelo, Takane (1980a, 1980b) presenta sus resultados para el caso no métrico, suponiendo que el proceso de error introduce violaciones de la monotonía. Ramsay discute una amplia variedad de transformaciones de los datos, concretamente transformaciones de potencias y splines monótonos, mientras que Takane discute restricciones adicionales de las transformaciones de los datos, que le permiten trabajar en cualquiera de los cuatro niveles de medida.

Expondremos a continuación brevemente ambos modelos:

- **Modelo de Ramsay:**

Supongamos que se dispone de una configuración de puntos,  $X$ , en un espacio euclídeo que representan un conjunto de objetos; y supongamos también que conocemos las distancias euclídeas,  $\{d_{ij}\}$ , entre todos los pares de puntos de la configuración.

Supongamos que las disimilaridades,  $\delta_{ij}$ , entre los objetos  $i$  y  $j$  condicionadas a las distancias,  $d_{ij}$ , tienen función de densidad de probabilidad,  $p(\delta_{ij}/d_{ij})$ , independientemente  $\forall (i, j)$ . Entonces, la función de verosimilitud asociada a la configuración, vendría dada por:

$$l = \sum_{i=1}^n \sum_{j=1}^n \ln [p(\delta_{ij}/d_{ij})].$$

Las distancias pueden ser escritas en términos de las coordenadas de los puntos:

$$d_{ij}^2 = (x_i - x_j)^t (x_i - x_j),$$

siendo  $x_i$  la  $i$ -ésima fila de la matriz de coordenadas. Con lo que la verosimilitud podría ser minimizada respecto a  $x_i$  y respecto a cualquiera de los parámetros de la función de densidad de probabilidad,  $p$ , proporcionando las estimaciones de máxima verosimilitud de las coordenadas,  $\hat{x}_i$ .

Dos distribuciones posibles para  $\delta_{ij}/d_{ij}$  son la distribución normal y la lognormal:

$$\delta_{ij} \sim N(d_{ij}, \sigma^2)$$

$$\ln \delta_{ij} \sim N(\ln d_{ij}, \sigma^2)$$

con varianza  $\sigma^2$  constante.

En cualquier caso, en ocasiones, resulta conveniente aplicar una transformación a las disimilaridades antes de considerar la estructura de los errores, en concreto, Ramsay sugiere una transformación basada en splines monótonos.

- **Modelo de Takane:**

Takane presenta un método de máxima verosimilitud para MDS no métrico. Dada una configuración de puntos,  $X$ , y las distancias entre dichos puntos,  $\{d_{ij}\}$ , para un modelo de error aditivo, se considera una variable,  $\lambda_{ij}$ , definida como

$$\lambda_{ij} = d_{ij} + \varepsilon_{ij} \quad \varepsilon_{ij} \sim N(0, \sigma_{ij}^2)$$

de manera que si  $\lambda_{ij} \geq \lambda_{rs}$ , se define la variable:

$$Y_{ijrs} = \begin{cases} 1 & \text{si } \delta_{ij} \geq \delta_{rs} \\ 0 & \text{si } \delta_{ij} < \delta_{rs}. \end{cases}$$

Con lo que, si se considera  $\Phi$ , la función de distribución de una variable aleatoria  $N(0,1)$ ,

$$P[Y_{ijrs} = 1] = P[\lambda_{ij} - \lambda_{rs} \geq 0] = \Phi\left(\frac{d_{ij} - d_{rs}}{(\sigma_{ij}^2 + \sigma_{rs}^2)^{1/2}}\right) = \Phi_{ijrs},$$

la verosimilitud, suponiendo la independencia de los errores  $\varepsilon_{ij}$ , vendrá dada por:

$$L = \prod \Phi_{ijrs}^{Y_{ijrs}} (1 - \Phi_{ijrs})^{1-Y_{ijrs}}.$$

Al igual que con el procedimiento de Ramsay, escribiendo las distancias en términos de las coordenadas, es posible maximizar la verosimilitud y obtener la configuración buscada.

Zinnes y Mackay dan una aproximación diferente al introducir los errores probabilísticos en las técnicas MDS. Suponen que la perturbación de los datos se produce durante el proceso de juicio de los sujetos, lo que implica que las disimilaridades siguen una distribución  $\chi^2$  no centrada en torno a la verdadera distancia. Al igual que Ramsay y Takane; Zinnes y Mackay suponen también que los datos son valores de disimilaridad con un determinado modelo de error.

Si se considera cada estímulo representado por un vector aleatorio,  $X_i = (X_{i_1}, \dots, X_{i_p})$ , cuyas componentes  $X_{i_j}$  se suponen independientes, normalmente distribuidas con media,  $\mu_i$  y varianza,  $\sigma_i^2$ , dicha distribución induce otra distribución para la distancia euclídea  $(X_i - X_j)^t(X_i - X_j)$ . Suponiendo además que la disimilaridad observada es la distancia euclídea, es decir,

$$\delta_{ij} = [(X_i - X_j)^t(X_i - X_j)]^{1/2},$$

y que la verdadera distancia entre los puntos  $i$  y  $j$  viene dada por

$$d_{ij}^2 = (\mu_i - \mu_j)^t(\mu_i - \mu_j),$$

como  $\frac{\delta_{ij}}{(\sigma_i^2 + \sigma_j^2)}$  sigue una distribución  $\chi^2$  no centrada, es posible dar una aproximación de las funciones de densidad de las disimilaridades.

Zinnes y Mackay maximizan la suma de los logaritmos de las funciones de densidad aproximadas (una para cada disimilaridad) respecto a  $\{\mu_r\}$  y  $\{\sigma_r^2\}$ . Los valores  $\{\mu_r\}$  dan las coordenadas de los puntos de la configuración.

## 2.5 Análisis Procrustiano General: GPA

Simultáneamente al desarrollo de las técnicas citadas, surge el problema de establecer relaciones entre diversas representaciones de un mismo conjunto de elementos. Las situaciones que pueden surgir son varias y a todas ellas da solución el análisis procrustiano:

1. Si diversos investigadores recogen información de un mismo conjunto de datos y analizan los resultados mediante diferentes técnicas de MDS o Análisis Multivariante, se obtendrán representaciones diferentes. Ante esta situación, se plantea el comprobar la equivalencia de los resultados.
2. Dadas  $K$  matrices de disimilaridades, en lugar de optar por un análisis MDS con diferencias individuales conjunto de las  $K$  matrices, se puede proceder al análisis individual de cada una de ellas para investigar a continuación las relaciones entre las configuraciones obtenidas. De este análisis se pueden concluir resultados referentes a las diferencias entre una y otra fuente de información.
3. Un mismo conjunto de datos se puede analizar varias veces considerando distintos niveles de medida de los datos y diferentes aproximaciones a las ligaduras. El contraste de los resultados obtenidos permite determinar las diferencias entre uno y otro análisis.

Los orígenes del GPA se deben a Green (1952), Cliff (1966) y Schönemann (1966), quienes resolvieron el problema de transformar mediante rotaciones y reflexiones una representación euclídea de tal modo que la suma de los cuadrados de las diferencias

entre las distancias para cada par de puntos en las configuraciones que se comparan, se minimice. Las soluciones de Green y Cliff sólo funcionan para matrices de rango completo por columnas, mientras que Schönemann generalizó la solución al caso en que las matrices no cumplieran este requisito. Schönemann y Carroll (1970) investigaron y resolvieron el problema de ajustar una configuración a otra usando, además de las transformaciones ortonormales ya citadas, traslaciones y reescalamiento uniforme.

Kristof y Wingersky (1971) extendieron las transformaciones ortonormales al caso de trabajar con más de dos configuraciones. Gower (1975) resolvió nuevamente el problema de ajustar más de dos configuraciones usando transformaciones ortonormales y dilataciones centrales cuando las representaciones tienen la misma dimensionalidad (a Gower se debe también el término GPA). En 1977, Ten Berge rectificó el método de determinación de dilataciones centrales desarrollado por Gower y mejoró el procedimiento de cálculo de transformaciones ortonormales. Posteriormente, y en colaboración con Knol (Ten Berge y Knol, 1984), extendieron el problema al caso en que las configuraciones contrastadas viniesen dadas en un número diferente de dimensiones (aunque considerando sólo transformaciones ortonormales). Este análisis fue completado por Peay (1988) resolviendo el problema de ajustar más de dos configuraciones, en diferentes dimensiones, incluyendo traslaciones y dilataciones centrales.

Los últimos trabajos en GPA, que comentaremos en el siguiente capítulo, se deben a Commandeur (1991), quien ha generalizado las técnicas GPA al caso de trabajar con configuraciones que tienen un número diferente de filas, es decir, configuraciones en las que falta información sobre algún punto. Se encuentran antecedentes de este tema en los trabajos de Everitt y Gower (1981), en lo que ellos denominaron *método procrustiano generalizado ponderado*, aunque no trabajaron con dilataciones centrales ni proporcionaron un tratamiento para las traslaciones. De Leeuw y Meulman (1986) han trabajado también con configuraciones incompletas en el contexto del *jackknife*, investigando cómo igualar una serie de configuraciones cuando cada configuración carece de información de un único estímulo.

## Capítulo II

# Situación actual de las técnicas MDS

En las cuatro primeras décadas de desarrollo de MDS, las líneas de investigación giraron en torno a la creación de técnicas de minimización más potentes y que permitieran trabajar con cualquier tipo de datos. En la actualidad, la investigación en MDS, si bien sigue tratando los problemas de minimización (SMACOF y TUNNELING), se centra más propiamente en problemas concretos: trabajar con un gran número de datos (CDS), incorporación de las técnicas de gráficos móviles en los análisis MDS (VISUALS/Pxpl y ViSta MDS), comparación de análisis diferentes (PROCRUSTES), etc.

### 1 Nuevas técnicas de minimización. El algoritmo SMACOF

El algoritmo *SMACOF* (Heiser y De Leeuw, 1980), basado en las técnicas de minimización por mayorización iterativa, proporciona una estrategia de minimización consistente en el recubrimiento de la superficie de la función objetivo por funciones auxiliares más simples (lineales o bilineales). Al aplicar la técnica de mayorización se aproxima una función complicada por una función auxiliar más simple que se denominará función de mayorización. Esta función *toca* a la función implicada en la estimación actual (*punto base*) y está localizada siempre por encima de ella. En el mínimo de la función de mayorización, el valor de la función original tiene que ser menor o igual al valor que toma en el punto base, dado que la función de mayorización está localizada encima de la función original y este punto se convierte en el punto base de la siguiente función de mayorización. Iterando el procedimiento, se obtienen mejores estimaciones que llevan a un mínimo local.

Aplicando las técnicas de mayorización se obtiene un algoritmo convergente para minimizar el STRESS con una razón de convergencia determinada.

Si  $\phi(x)$  es la función que se pretende minimizar, la idea central del método de mayorización consiste en ir reemplazando iterativamente la función original,  $\phi(x)$ , por una función auxiliar,  $\hat{\phi}(x, y)$ , verificando las siguientes condiciones:

1.  $\hat{\phi}(x, y)$  es más fácil de minimizar que  $\phi(x)$
2.  $\phi(x) \leq \hat{\phi}(x, y), \quad \forall x, y$
3.  $\phi(y) = \hat{\phi}(y, y)$ .

Esta función,  $\hat{\phi}(x, y)$ , recibe el nombre de *función de mayorización de  $\phi(x)$*  y el parámetro que la define se denomina *punto base*.

Para hacer cumplir el tercer principio de la mayorización iterativa, la función de mayorización ha de elegirse de manera que exista su mínimo sobre el dominio y que éste sea único; de este modo se puede definir la aplicación

$$x^* = \arg \min_x \hat{\phi}(x, y)$$

que da origen a la cadena de desigualdades:

$$\phi(x^*) \leq \hat{\phi}(x^*, y) \leq \hat{\phi}(y, y) = \phi(y) \quad \forall x, y,$$

que proporciona una secuencia no creciente de valores de la función, convergente a un punto estacionario de la función objetivo, que en la mayoría de los casos será un mínimo local.

El algoritmo de mayorización queda entonces:

1.  $y \leftarrow y_0$ ; con  $y_0$  un valor inicial.
2. Encontrar  $x^*$  tal que  $\hat{\phi}(x^*, y) = \min_x \hat{\phi}(x, y)$ .
3. Si  $\phi(y) - \phi(x^*) < \epsilon$ , parar.
4.  $y \leftarrow x^*$  y volver al paso 2.

Si el proceso se para, la anulaci3n de la primera derivada de la funci3n de mayorizaci3n en su m3nimo, implica que la derivada de la funci3n de mayorizaci3n de la funci3n objetivo, si existe en ese punto, se anule tambi3n. Esta propiedad se verifica debido a la derivabilidad y continuidad con que se selecciona la funci3n de mayorizaci3n y que conlleva que el minorante y la funci3n de mayorizaci3n tienen la misma derivada en cualquier punto base.

### 1.1 Mayorizaci3n de la funci3n STRESS

Como aplicaci3n del algoritmo SMACOF, expondremos en unas l3neas el proceso de mayorizaci3n de la funci3n STRESS

$$S(X) = \sum_{i < j}^n w_{ij} (\delta_{ij} - d_{ij}(X))^2.$$

Desarrollando el cuadrado, esta funci3n puede ser reescrita como:

$$\begin{aligned} S(X) &= \sum_{i < j}^n w_{ij} (\delta_{ij} - d_{ij}(X))^2 = \\ &= \sum_{i < j}^n w_{ij} \delta_{ij}^2 + \sum_{i < j}^n w_{ij} d_{ij}^2(X) - 2 \sum_{i < j}^n w_{ij} \delta_{ij} d_{ij}(X) = \\ &= \eta_\delta^2 + \eta^2(X) - 2\rho(X). \end{aligned}$$

En todo momento se supondr3 que la matriz de pesos es irreducible, es decir, no existe ninguna partici3n de los objetos en subconjuntos disjuntos tales que  $w_{ij} = 0$ , si los objetos  $i$  y  $j$  est3n en subconjuntos diferentes. En el caso de que la matriz de pesos sea reducible, el problema puede ser descompuesto en peque1os problemas separados de MDS para cada subconjunto. Sin p3rdida de generalidad, se puede considerar que la media de las columnas de la matriz de configuraci3n,  $X$ , vale 0, es decir, que el origen de coordenadas se encuentra en el centroide de la configuraci3n.

Usando la desigualdad de Cauchy-Schwarz

$$\|x_i - x_j\| \|y_i - y_j\| \geq (x_i - x_j)'(y_i - y_j),$$

en t3rminos de la distancia eucl3dea se obtiene que

$$d_{ij}(X)d_{ij}(Y) \geq (x_i - x_j)'(y_i - y_j)$$

y multiplicando el primer miembro por  $w_{ij}\delta_{ij}/w_{ij}\delta_{ij}$  resulta

$$\frac{w_{ij}\delta_{ij}d_{ij}(X)d_{ij}(Y)}{w_{ij}\delta_{ij}} \geq (x_i - x_j)'(y_i - y_j),$$

con lo que reordenando los términos y cambiando de signo en ambos miembros de la desigualdad se obtiene

$$-w_{ij}\delta_{ij}d_{ij}(X) \leq -\frac{w_{ij}\delta_{ij}}{d_{ij}(Y)}(x_i - x_j)'(y_i - y_j).$$

Por tanto,

- Si  $d_{ij}(Y) > 0$ , se da la igualdad en el caso  $X = cY$ .
- Si  $d_{ij}(Y) = 0$ , se sigue cumpliendo

$$-w_{ij}\delta_{ij}d_{ij}(X) \leq 0$$

dándose la igualdad cuando  $d_{ij}(X) = 0$ .

Definiendo entonces

$$b_{ij}(Y) = \begin{cases} w_{ij}\delta_{ij}/d_{ij}(Y) & \text{si } d_{ij}(Y) \neq 0 \\ 0 & \text{si } d_{ij}(Y) = 0 \end{cases}$$

se verifica siempre que

$$-w_{ij}\delta_{ij}d_{ij}(X) \leq -b_{ij}(Y)(x_i - x_j)'(y_i - y_j)$$

con la igualdad si  $X = cY$ .

Sumando en ambos miembros de la desigualdad en  $i < j$  resulta:

$$-\rho(X) = -\sum_{i < j} w_{ij}\delta_{ij}d_{ij}(X) \leq -\sum_{i < j} b_{ij}(Y)(x_i - x_j)'(y_i - y_j) = -\hat{\rho}(X, Y), \quad (\text{II.1})$$

lo que implica que  $-\rho(X)$  puede ser mayorizada por una función lineal  $-\hat{\rho}(X, Y)$  tal que

$$\rho(Y) = \hat{\rho}(Y, Y)$$

resultando de esta manera un caso de mayorización lineal.

Si se considera la expresión

$$\sum_{i < j} a_{ij} (x_i - x_j)' (y_i - y_j) = \text{tr} (X'AY),$$

válida para cualquier matriz simétrica,  $A$ , con elementos fuera de la diagonal principal,  $-a_{ij}$ , y elementos diagonales,  $\sum_{j \neq i} a_{ij}$ , y combinándola con la parte derecha de (II.1) la función de mayorización,  $\hat{\rho}(X, Y)$ , queda expresada como:

$$\hat{\rho}(X, Y) = \text{tr} [X'B(Y)Y]$$

Dado que la igualdad en (II.1) se alcanza cuando  $X = Y$ , la expresión de mayorización adopta la siguiente forma

$$-\rho(X) = -\hat{\rho}(X, X) = -\text{tr} [X'B(X)X] \leq -\text{tr} [X'B(Y)Y] = -\hat{\rho}(X, Y).$$

Análogamente, se obtiene una expresión matricial para  $\eta^2(X)$ . Si se toma  $D^{(2)}$  como la matriz de las distancias al cuadrado  $d_{ij}(X)$ ,

$$D^{(2)} = \mathbf{1}\alpha' + \alpha\mathbf{1}' - 2XX', \tag{II.2}$$

donde  $\alpha$  es un vector con los elementos diagonales de  $XX'$ .

Entonces, como  $D^{(2)}$  tiene todos sus elementos diagonales  $\theta$ , resulta que  $\eta^2(X)$  puede escribirse como

$$\eta^2(X) = -\frac{1}{2} \text{tr}(VD^{(2)}) \tag{II.3}$$

siendo  $V$  una matriz simétrica con elementos no diagonales,  $-w_{ij}$ , y elementos diagonales,  $\sum_{i \neq j} w_{ij}$ . El factor  $1/2$  se introduce debido a la simetría de  $D^{(2)}$ . Uniendo (II.2) y (II.3), y usando el hecho de que  $\mathbf{1}$  está en el espacio anulador de  $V$ , resulta entonces que

$$\begin{aligned} \eta^2(X) &= -\frac{1}{2} \text{tr}(VD^{(2)}) = -\frac{1}{2} \text{tr} [V(\mathbf{1}\alpha' + \alpha\mathbf{1}' - 2XX')] = \\ &= -\frac{1}{2} [\text{tr}(V\mathbf{1}\alpha') + \text{tr}(V\alpha\mathbf{1}' - 2 \text{tr}(VXX'))] = \text{tr}(X'VX). \end{aligned}$$

dado que los dos primeros sumandos son nulos.

Uniendo los resultados anteriores, se obtiene la expresión del STRESS en forma matricial dada por

$$S^2(X) = \eta_\delta^2 + \text{tr}(X'VX) - 2\text{tr}[X'B(X)X],$$

que es mayorizada por

$$\hat{S}^2(X, Y) = \eta_\delta^2 + \text{tr}(X'VX) - 2\text{tr}[X'B(Y)Y]. \quad (\text{II.4})$$

La función de mayorización (II.4) es una función cuadrática en  $X$ , con lo que igualando a 0 su gradiente respecto  $X$ , se obtiene su mínimo

$$\nabla \hat{S}^2(X, Y) = 2VX - 2B(Y)Y = 0 \quad (\text{II.5})$$

Como antes se señalaba, la matriz  $V$  tiene un  $\mathbf{1}$  en su espacio anulador, por lo que para obtener el mínimo indicado es necesario el uso de una inversa generalizada, que, por su unicidad, es la inversa de Moore-Penrose, que viene definida por

$$V^- = (V + \mathbf{1}\mathbf{1}')^{-1} - n^{-2}\mathbf{1}\mathbf{1}'.$$

El último término,  $-n^{-2}\mathbf{1}\mathbf{1}'$ , es irrelevante en SMACOF dado que  $V^-$  es multiplicada a la derecha por una matriz ortogonal a  $\mathbf{1}$  y puesto que  $B(Y)$  también tiene un  $\mathbf{1}$  en su espacio anulador, se obtiene la fórmula actualizada del algoritmo SMACOF, dada por la expresión

$$X = V^-B(Y)Y. \quad (\text{II.6})$$

De Leeuw y Heiser (1980) denominaron a (II.6) *Transformada de Guttman*, en reconocimiento a Guttman (1968). El algoritmo de mayorización garantiza una serie no creciente de valores de STRESS, de forma que cuando el algoritmo para, se verifica la condición estacionaria

$$X = V^-B(X)X.$$

Dado que las distancias no se ven alteradas por una rotación de la configuración,  $X$ , es conveniente, tras alcanzar la convergencia, rotar la configuración a los ejes principales.

## 2 Determinación de mínimos globales. Algoritmo Tunneling

Una de las ventajas del algoritmo de mayorización en MDS métrico es la obtención de una sucesión decreciente de valores de STRESS. Sin embargo, el método de mayorización sigue sin obtener el mínimo global, por lo que pueden existir otras configuraciones con un ajuste mejor. Se conocen diversos procedimientos de investigación local, de hecho, podría utilizarse cualquier técnica estándar de programación no lineal, pero todos estos procedimientos siguen sin facilitar el mínimo global. Se presenta a continuación un resumen y una clasificación de las técnicas de optimización global.

### 2.1 Clasificación del campo de la optimización global

Atendiendo a la clasificación de Törn y Zilinskas (1989), se establece en primer lugar una distinción entre técnicas que tienen una cierta precisión en la solución, garantizada, durante la búsqueda del óptimo global y las que no pueden determinar tal precisión. Al primer grupo pertenecen los métodos de recubrimiento, que excluyen subregiones que no contienen el mínimo global. El segundo grupo se divide a su vez en métodos directos (tales como la investigación aleatoria, métodos *clustering* y métodos descendentes generalizados) que utilizan tan solo la información local, y métodos indirectos, que se basan en la construcción de un modelo sobre los conjuntos de nivel o la función objetivo (aproximando los conjuntos de nivel a la función objetivo).

Queda así la siguiente clasificación de los *Métodos de Optimización Global*:

#### MÉTODOS DE OPTIMIZACIÓN GLOBAL

##### 1. Métodos con precisión garantizada

- Métodos de recubrimiento
  - (a) *Branch & Bound*
  - (b) Programación dinámica

##### 2. Métodos con precisión no garantizada

- Métodos directos
  - (a) Búsqueda aleatoria
    - i. Búsqueda aleatoria pura

- ii. *Singlestart*
- iii. *Multistart*
- (b) Estrategias de intercambio de pares (LOPI)
  - i. LOPI1
  - ii. LOPI2
  - iii. LOPI3
  - iv. LOPI4
- (c) Clustering
  - i. *Multi-level-single-linkage*
- (d) Descenso generalizado
  - i. Búsqueda Tabú
  - ii. Métodos penalti (*Tunneling*)
- Métodos indirectos
  - (a) Aproximación por conjuntos de nivel

El ejemplo más simple que puede darse de los **métodos de recubrimiento** es la enumeración completa en problemas de combinatoria, como es el caso del *scaling* unidimensional. A este grupo corresponden también las técnicas *Branch & Bound*, que se basan en la división del espacio de parámetros total en diversos subconjuntos en los que son conocidos los límites inferior y superior que toma la función. Los subconjuntos que se sabe que no contienen el mínimo global se van excluyendo mientras que los subconjuntos restantes vuelven a ser divididos, de manera que el proceso finaliza con el mínimo global dentro de algún subconjunto preciso. Recientes publicaciones acerca de este tipo de métodos aparecen en Horst y Tuy (1990), quienes sólo usan análisis convexo y no necesitan por tanto de los requerimientos usuales de existencia de la primera y segunda derivada de la función objetivo. Otro tipo de método de recubrimiento fue utilizado por Hubert y Arabie (1986), quienes hicieron uso de la *programación dinámica* para *scaling* unidimensional con resultados satisfactorios. Estos métodos pueden ser considerados como determinísticos, dado que tienen el éxito garantizado, aunque éste no siempre se alcance en un tiempo finito.

Entre los **métodos directos** pueden destacarse los métodos de *búsqueda aleatoria*, entre los que destacan los métodos de *búsqueda aleatoria pura*, *singlestart* y *multistart*. En los tres casos se generan configuraciones aleatorias para las que se determinan los valores de STRESS. En *singlestart* se empieza una búsqueda local a partir de la configuración aleatoria con menor STRESS, mientras que *multistart* lleva a cabo tales búsquedas

locales desde cada configuración aleatoria obtenida. En la búsqueda puramente aleatoria no se realiza ninguna búsqueda local. Los tres métodos consideran el menor valor de STRESS como el mínimo global. Estas estrategias (especialmente *multistart*) son ampliamente utilizadas para investigar si una función presenta o no mínimos locales.

Las estrategias **LOPI**, aplicadas al *scaling* unidimensional persiguen la obtención del mínimo global intercambiando pares de elementos. Según el criterio que se siga para seleccionar estos pares, se definen diferentes procedimientos LOPI.

El cuarto tipo está formado por los **métodos clustering**. Estos métodos refinan el método *multistart* de manera que las búsquedas locales que se llevan a cabo se realizan sólo una vez sobre configuraciones que conducen al mismo mínimo local.

El quinto tipo está formado por los **métodos descendentes generalizados**, donde se modifica la función para garantizar un mínimo menor. Dentro de este grupo hemos de citar en el caso unidimensional la *búsqueda tabú* que es un procedimiento en dos fases, una de búsqueda local y otra de búsqueda global en la que se prohíben direcciones de búsqueda con objeto de evitar volver a puntos estacionarios localizados en pasos anteriores. Los *métodos penalti* modifican la propia función para prevenir la vuelta a alguno de los mínimos locales encontrado en alguna iteración previa. Uno de tales métodos penalti es el *algoritmo tunneling*, que alterna una búsqueda local con un paso *tunneling* con el objetivo de encontrar en el último paso una solución diferente de los anteriores mínimos locales, con el mismo valor de STRESS que el mínimo local. Una importante y atractiva característica del algoritmo *tunneling* es que los sucesivos mínimos locales que se van obteniendo, proporcionan un valor de STRESS cada vez menor.

Los primeros **modelos indirectos** aspiran a una *aproximación mediante los conjuntos de nivel*, conjuntos éstos que se definen como aquella serie de configuraciones con STRESS menor o igual que alguna constante:

$$L_c = \{X \in R^{n \times p} \text{ tal que } \sigma(X) \leq c\}$$

La idea consiste en encontrar un valor,  $c$ , y una solución,  $X$ , para los que la medida de volumen del conjunto  $L_c$  sea 0, de manera que el STRESS tenga en esa configuración su mínimo global. Nuevamente, Törn y Zilinskas observan que esta aproximación no es muy satisfactoria.

## 2.2 Algoritmo tunneling

Entre todos los métodos de minimización global hemos de citar, por los recientes estudios que se han hecho sobre él, el método *Tunneling*. Este método se basa en la obtención de una serie decreciente de mínimos locales, alternando una búsqueda local con un paso *tunneling* en el que se obtiene una nueva configuración con igual o menor STRESS que la anterior pero distinta a ella. El último paso se realiza minimizando la función *tunneling*, que se modeliza de manera que se excluyan soluciones triviales o que hayan sido obtenidas anteriormente.

El método *Tunneling* (Groenen, 1993) propuso una alternativa al problema actual de los métodos iterativos actuales, cuyo objetivo principal es buscar métodos para salir de un punto estacionario. Para ello hace uso de los llamados *polos móviles* que permiten eliminar la atracción de puntos estacionarios no óptimos globalmente, a la vez que proporciona una serie decreciente de mínimos locales. El método *tunneling* viene definido por un procedimiento iterativo en dos pasos:

1. Localización de un mínimo local (*fase de minimización local*).
2. Determinación de una nueva configuración con el mismo STRESS que el mínimo local obtenido (*fase tunneling*).

El paso *tunneling* es la parte esencial del método. Se realiza minimizando una función particular,  $\tau(X)$ , denominada *función tunneling*, que presenta varias características:

1. Toma el valor 0 en configuraciones con STRESS igual al del mínimo local,  $X^*$ ; además, el valor 0, ha de ser el menor valor que pueda tomar la función  $\tau$ .
2. Es invariante ante rotaciones de la configuración,  $X^*$ .
3. En puntos cercanos a la configuración mínimo local,  $X^*$ ,  $\tau(X)$  toma valores elevados.

De este modo, la función *tunneling* se define como:

$$\tau(X) = (S(X) - S(X^*))^{2\lambda} \left( 1 + \frac{1}{\|D(X^*) - D(X)\|^2} \right),$$

siendo  $S(X)$  el STRESS de la configuración,  $X$ , y  $\|D(X)\|^2$ , la norma euclídea de la matriz de distancias asociada a la configuración.

Esta función puede considerarse como una razón de dos funciones en  $X$ , con lo que la minimización de  $\tau(X)$  puede verse como un problema de programación fraccional. Para minimizar esta función, Dinkelbach (1967) propuso un algoritmo que fue utilizado por Groenen (1990) con condiciones menos restrictivas: Mientras que el algoritmo de Dinkelbach supone que en cada iteración ha de obtenerse el mínimo absoluto sobre  $X$  de una función auxiliar  $F(q, X)$ , Groenen y Heiser (1996) demostraron que tal algoritmo sigue siendo convergente si se encuentra el mínimo de una función que mayorice a  $F(q, X)$ .

Si se reescribe la función *tunneling* como

$$\tau(X) = N(X) \left( 1 + \frac{1}{P(X)} \right) = \frac{M(X)}{P(X)}$$

con  $M(X) = N(X)(1 + P(X))$ , y definiendo

$$F(q, X) = M(X) - qP(X),$$

el algoritmo de Dinkelbach equivale a:

1. Inicializar  $q \leftarrow \tau(X_0)$
2.  $X^* \leftarrow \arg \min F(q, X)$
3.  $q^+ \leftarrow \tau(X^+)$ .
4. Si  $q^+ < \epsilon$ , detener el algoritmo (con  $\epsilon$  una constante positiva pequeña).
5. Fijar  $q = q^+$  y volver al paso 2.

En cualquier caso, aunque este procedimiento no garantizase la obtención del mínimo global, en la práctica parecía quedarse bastante cerca de él; sin embargo, estudios posteriores demostraron que el procedimiento de mayorización empleado para la minimización de la función *tunneling* no era correcto y, hasta el momento, no se ha producido ningún nuevo avance en esta línea.

### 3 Análisis de grandes conjuntos de datos: CDS

Cuando se consideran grandes conjuntos de datos se producen diversos problemas en la interpretación de los resultados de MDS, dado que ésta puede llegar a ser bastante complicada. Una solución a este problema podría ser realizar un análisis clúster sobre

las coordenadas de los objetos tras el análisis MDS; sin embargo, esta estrategia ignora la información acerca del buen ajuste de las distancias con las disimilaridades entre los objetos y de ahí que se trabaje sobre un método propuesto por Heiser y Groenen (1997) conocido como *SCALING CON DIFERENCIAS TIPO CLUSTERING: CDS* (*cluster differences scaling*) que realiza simultáneamente *scaling* de los clústeres y *clustering* de los objetos, de manera que las distancias se ajusten a las disimilaridades tan bien como sea posible. En su análisis, Heiser propone una descomposición del STRESS en dos sumandos: STRESS dentro de, y entre los componentes de los clústeres.

La característica básica de CDS es que, en este caso, los elementos que son escalados son los clústeres de objetos y no los propios objetos, lo que permite reducir el número de objetos diferentes en la configuración y proporciona una clasificación de objetos similares. Los objetos se asignan a uno de los  $K$  clústeres, que se representan en un espacio  $p$ -dimensional por medio de la matriz de coordenadas,  $X_{K \times p}$ , de modo que minimice el STRESS. Si se define la matriz  $G$ , indicadora de los componentes de cada clúster

$$g_{ik} = \begin{cases} 1 & \text{si el objeto, } i, \text{ está en el clúster, } k \\ 0 & \text{en caso contrario} \end{cases},$$

el STRESS puede escribirse como:

$$S(G, X) = \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^K \sum_{l=1}^K g_{ik} g_{jl} w_{ij} (\delta_{ij} - d_{kl}(X))^2 \quad (\text{II.7})$$

donde

- $\delta_{ij}$  es la disimilaridad entre los objetos  $i$  y  $j$ ,
- $w_{ij}$  es una ponderación que se hace para cada par de objetos  $(i, j)$ , que puede ser binaria (códigos 0 y 1 para trabajar con datos faltantes) o continua (ponderación basada en la inversa del error estándar para medidas con réplicas), y
- $d_{kl}$  es la distancia euclídea entre el punto clúster  $k$  y el punto clúster  $l$ .

La suma de cuadrados (II.7) puede ser descompuesta, al igual que en un análisis de la varianza (ANOVA) en cuatro sumandos haciendo uso de la distancia de Sokal-Michener, que indica la media de las disimilaridades entre los clústeres  $k$  y  $l$ :

$$\tilde{\delta}_{kl} = \frac{\sum_{i=1}^n \sum_{j=1}^n g_{ik} g_{jl} w_{ij} \delta_{ij}}{\sum_{i=1}^n \sum_{j=1}^n g_{ik} g_{jl} w_{ij}}.$$

Introduciendo la distancia de Sokal-Michener, sumando y restando, la descomposición queda del siguiente modo:

$$\begin{aligned} S(G, X) &= \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^K \sum_{l=1}^K g_{ik} g_{jl} w_{ij} (\delta_{ij} - d_{kl}(X))^2 = \\ &= \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^K \sum_{l=1}^K g_{ik} g_{jl} w_{ij} (\delta_{ij} - \tilde{\delta}_{kl} + \tilde{\delta}_{kl} - d_{kl}(X))^2 = \\ &= \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^K \sum_{l=1}^K g_{ik} g_{jl} w_{ij} (\delta_{ij} - \tilde{\delta}_{kl})^2 + \\ &+ \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^K \sum_{l=1}^K g_{ik} g_{jl} w_{ij} (\tilde{\delta}_{kl} - d_{kl}(X))^2 + \\ &+ 2 \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^K \sum_{l=1}^K g_{ik} g_{jl} w_{ij} (\delta_{ij} - \tilde{\delta}_{kl}) (\tilde{\delta}_{kl} - d_{kl}(X)). \end{aligned}$$

Ahora bien, dado que la media de las desviaciones a la media es 0, resulta

$$\begin{aligned} &\sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^K \sum_{l=1}^K g_{ik} g_{jl} w_{ij} (\delta_{ij} - \tilde{\delta}_{kl}) (\tilde{\delta}_{kl} - d_{kl}(X)) = \\ &= \sum_{k=1}^K \sum_{l=1}^K (\tilde{\delta}_{kl} - d_{kl}(X)) \sum_{i=1}^n \sum_{j=1}^n g_{ik} g_{jl} w_{ij} (\delta_{ij} - \tilde{\delta}_{kl}) = \\ &= \sum_{k=1}^K \sum_{l=1}^K (\tilde{\delta}_{kl} - d_{kl}(X)) \sum_{i=1}^n \sum_{j=1}^n g_{ik} g_{jl} w_{ij} \left( \delta_{ij} - \frac{\sum_{i=1}^n \sum_{j=1}^n g_{ik} g_{jl} w_{ij} \delta_{ij}}{\sum_{i=1}^n \sum_{j=1}^n g_{ik} g_{jl} w_{ij}} \right) = \end{aligned}$$

$$= \sum_{k=1}^K \sum_{l=1}^K (\tilde{\delta}_{kl} - d_{kl}(X)) \sum_{i=1}^n \sum_{j=1}^n g_{ik} g_{jl} w_{ij} \delta_{ij} - \sum_{i=1}^n \sum_{j=1}^n g_{ik} g_{jl} w_{ij} \delta_{ij} = 0.$$

Queda entonces

$$\begin{aligned} S(G, X) &= \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^K \sum_{l=1}^K g_{ik} g_{jl} w_{ij} (\delta_{ij} - \tilde{\delta}_{kl})^2 + \\ &+ \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^K \sum_{l=1}^K g_{ik} g_{jl} w_{ij} (\tilde{\delta}_{kl} - d_{kl}(X))^2. \end{aligned}$$

Partiendo en dos cada sumatoria, de manera que se considere aparte el caso en que  $l = k$ , y teniendo en cuenta que  $d_{kk}(X) = 0$  se obtiene

$$\begin{aligned} S(G, X) &= \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^K \sum_{l=1, l \neq k}^K g_{ik} g_{jl} w_{ij} (\delta_{ij} - \tilde{\delta}_{kl})^2 + \\ &+ \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^K g_{ik} g_{jk} w_{ij} (\delta_{ij} - \tilde{\delta}_{kk})^2 + \\ &+ \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^K \sum_{l=1, l \neq k}^K g_{ik} g_{jl} w_{ij} (\tilde{\delta}_{kl} - d_{kl}(X))^2 + \\ &+ \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^K g_{ik} g_{jk} w_{ij} \tilde{\delta}_{kk}^2. \end{aligned} \quad (\text{II.8})$$

Cada uno de los sumandos obtenidos en la descomposición tiene un significado concreto:

1. STRESS ENTRE PUNTOS EN LOS DISTINTOS CLÚSTERES: Mide el STRESS producido por las disimilaridades entre los objetos y la media de las disimilaridades entre los clústeres a los que pertenecen.
2. STRESS ENTRE PUNTOS DENTRO DE UN CLÚSTER: Mide la desviación de las disimilaridades de los objetos dentro de un clúster respecto la disimilaridad media de dicho clúster.
3. STRESS ENTRE CLÚSTERES: Muestra la similitud entre las distancias y las disimilaridades entre clústeres.

4. STRESS PRODUCIDO POR EL PROPIO CLÚSTER: Al considerar un único clúster necesariamente, el valor de la distancia es 0, pero las disimilaridades no tienen que tomar este valor necesariamente, de ahí que se incluya un término final que mida el STRESS dentro del clúster debido a las disimilaridades no nulas del clúster. El STRESS dentro del clúster mide la desviación a 0 de la distancia de Sokal-Michener de un clúster.

La descomposición (II.8) implica que el STRESS se puede minimizar en dos pasos:

- un paso de ajuste de *scaling* multidimensional y
- un paso de *clustering*.

Este algoritmo externo se resuelve de un modo alternante. Primero se obtiene una configuración sub-óptima  $X$  para los clústeres fijos  $G$ . Entonces, se fija  $X$  y se buscan clústeres  $G$  sub-óptimos, volviendo a continuación al primer paso hasta alcanzar la convergencia.

El único término de (II.8) dependiente de  $X$  es el STRESS entre clústeres. Claramente, para  $G$  fijo, este término es equivalente a un problema MDS con ponderaciones especiales, que puede ser minimizado por el algoritmo SMACOF ordinario. El STRESS entre clústeres puede expresarse así:

$$\begin{aligned}
 S_{BC}(X) &= \sum_{k=1}^K \sum_{l=1}^K \left( \sum_{i=1}^n \sum_{j=1}^n g_{ik} g_{jl} w_{ij} \right) (\tilde{\delta}_{kl} - d_{kl}(X))^2 = \\
 &= \sum_{k=1}^K \sum_{l=1}^K \left( \sum_{i=1}^n \sum_{j=1}^n g_{ik} g_{jl} w_{ij} \right) \tilde{\delta}_{kl}^2 + \\
 &+ \sum_{k=1}^K \sum_{l=1}^K \left( \sum_{i=1}^n \sum_{j=1}^n g_{ik} g_{jl} w_{ij} \right) d_{kl}^2(X) - \\
 &- 2 \sum_{k=1}^K \sum_{l=1}^K \left( \sum_{i=1}^n \sum_{j=1}^n g_{ik} g_{jl} w_{ij} \right) \tilde{\delta}_{kl} d_{kl} = \\
 &= \sum_{k=1}^K \sum_{l=1}^K \left( \sum_{i=1}^n \sum_{j=1}^n g_{ik} g_{jl} w_{ij} \right) \tilde{\delta}_{kl}^2 + \text{tr}(X' V X) - 2 \text{tr}(X' B(X) X)
 \end{aligned}$$

donde la matriz  $B(X)$  tiene elementos no diagonales

$$b_{kl} = \frac{\left( - \sum_{i=1}^n \sum_{j=1}^n g_{ik} g_{jl} w_{ij} \right) \tilde{\delta}_{kl}}{d_{kl}(X)}$$

y elementos diagonales

$$b_{kk} = - \sum_{l=1, l \neq k}^K b_{kl}.$$

Análogamente,  $V$  tiene elementos no diagonales

$$v_{kl} = \sum_{i=1}^n \sum_{j=1}^n g_{ik} g_{jl} w_{ij}$$

y elementos diagonales

$$v_{kk} = - \sum_{l=1, l \neq k}^K v_{kl}.$$

De acuerdo con el procedimiento SMACOF, la actualización,  $X^+$ , dada por la transformada de Guttman, vale

$$X^+ = V^- B(X) X, \quad (\text{II.9})$$

siendo  $V^-$  la inversa de Moore-Penrose de  $V$ . Si cada clúster tiene al menos un punto,  $V$  tendrá rango  $K - 1$ , dado que sus filas y columnas suman 0.

El segundo paso del algoritmo es formar los clústeres óptimos. Centrémonos sólo en el clúster que contiene al objeto  $i$ , manteniendo fijos los otros clústeres. Escribiendo el STRESS como

$$S(G, X) = \sum_{i=1}^n \sum_{k=1}^K g_{ik} \sum_{j=1}^n \sum_{l=1}^K g_{jl} w_{ij} (\delta_{ij} - d_{kl}(X))^2 = \sum_{i=1}^n \sum_{k=1}^K g_{ik} \gamma_{ik}$$

se puede ver que la mayor reducción en el STRESS se obtiene cuando el objeto  $i$  se asigna al clúster  $k$  de modo que

$$\gamma_{ik} = \min_l \{ \gamma_{il} \}.$$

Esto puede hacerse para un objeto cada vez, objeto tras objeto o mediante iteraciones sobre todos los objetos hasta que no se produzcan cambios. Si no hay restricciones sobre los conjuntos, es posible que aparezcan entonces clústeres sin objetos, lo que no es deseable a la vista de (II.8), que muestra que el STRESS se produce siempre que haya dos o más objetos en un clúster. Por tanto, siempre será necesario que cada clúster contenga al menos un objeto y que  $w_{ij}\delta_{ij} > 0$ .

Como el paso *clustering* es sólo uno de los dos pasos del algoritmo de mínimos cuadrados alternos, son posibles varias estrategias para detener el paso *clustering*:

- Si el objeto  $i$  cambia de un clúster a otro, se continua entonces en el paso MDS; y en el siguiente paso *clustering*, se continúa moviendo los objetos a partir del objeto  $i + 1$ .
- Recorrer todos los objetos una vez y relocalizarlos en el clúster que produce el menor STRESS.
- Se puede extender la segunda estrategia continuando el ciclo hasta que no se produzca ningún nuevo reordenamiento.

A las tres estrategias se les puede aplicar la restricción de que cada clúster contenga al menos un objeto, de manera que no se cambiará un objeto si es el último de un clúster. Sea cual sea la estrategia usada siempre hay que basar los cálculos en la partición más reciente de los clústeres, de manera que si el objeto  $i$  se cambia del clúster  $k$  al  $l$ , los cálculos para el objeto  $i + 1$  han de basarse en que el objeto  $i$  está en el clúster  $l$ , de modo que pueda garantizarse la reducción de STRESS.

La cantidad de reducción en la función de pérdida, STRESS, puede variar entre los dos pasos. Por ejemplo, si el paso *clustering* no produce ninguna reordenación, el STRESS no cambia. Es de esperar que durante los primeros pasos *clustering* disminuya más el STRESS, por lo que parece bastante útil tomar un criterio de convergencia débil en el paso *scaling* si se producen grandes descensos del STRESS en el paso clúster. Este criterio se puede ir haciéndose más restrictivo a medida que las reducciones se van haciendo menores.

## 4 Aplicación de las técnicas con gráficos móviles a MDS

La utilización de este tipo de gráficos se ha extendido al campo de las técnicas de MDS aplicándose a dos aspectos diferentes:

- Análisis de estructuras en dimensiones superiores (Visuals/PXPL)
- Análisis de sensibilidad de las soluciones MDS (ViSta MDS)

#### 4.1 Análisis de estructuras en dimensiones superiores. Técnicas Visuals/PXPL

Las técnicas VISUALS/PXPL constituyen una potente herramienta computacional que permite explorar y visualizar datos multivariantes representados en altas dimensiones. El objetivo de estas técnicas es la creación de hipótesis sobre la estructura geométrica de datos en altas dimensiones, aunque sólo seamos capaces de verlas en representaciones tridimensionales a lo sumo.

Los algoritmos desarrollados producen, a partir de una configuración de  $n$  puntos en dimensión  $h$ ,  $X_{n \times h}$ , una serie de gráficos tridimensionales,  $V_p$ ,  $p = 1, 2, \dots$ , obtenidos a partir de las proyecciones ortogonales del espacio  $h$ -dimensional rotado convenientemente. Estas proyecciones, presentadas en rápida sucesión, permiten ayudar a la formulación de hipótesis sobre la estructura geométrica de la configuración inicial, en dimensión  $h$ .

#### 4.2 Análisis de sensibilidad de las soluciones MDS

Aunque ya son muchos los programas estadísticos que incorporan módulos para la realización de análisis MDS, ninguno de ellos facilita, simultáneamente, la posibilidad de llevar a cabo un análisis de la estabilidad de los parámetros estimados en los modelos. El programa ViSta-MDS ofrece esta alternativa al usuario presentando un nuevo entorno de análisis MDS en el que también, al igual que en cualquier otro modelo, se puede proceder a la visualización del espacio de estímulos.

El programa, desarrollado por McFarlane y Young, está basado en lo que los autores denominan gráficos de expansión, que son gráficos dinámicos, conectados algebraicamente, de manera que:

1. Las configuraciones obtenidas son fácilmente manipulables. Se puede proceder a la traslación y rotación arbitraria de los puntos representados.
2. Cuando se lleva a cabo una manipulación de los gráficos, los resultados acerca del nuevo ajuste se obtienen de manera instantánea.

El proceso de estimación de ViSta-MDS se basa en el algoritmo SMACOF-I, tomando como función de pérdidas la función STRESS normalizada:

$$S = \frac{\sum_{i=1}^n \sum_{j=1}^n (\delta_{ij} - d_{ij})^2}{\sum_{i=1}^n \sum_{j=1}^n \delta_{ij}^2},$$

las coordenadas de los nuevos puntos tras cada iteración, vienen dadas por:

$$\hat{x}_{is} = \frac{1}{2n} \sum_{i=1}^n \sum_{k \neq i}^n \frac{2\delta_{ik}}{d_{ik}} x_{is} - \frac{1}{2n} \sum_{i=1}^n \sum_{j \neq i}^n \frac{2\delta_{ij}}{d_{ij}} x_{js}.$$

ViSta-MDS es único en el sentido de que realiza el algoritmo descrito mientras permite al usuario visualizar el movimiento de los puntos y los cambios en los valores de STRESS. El usuario puede ver los caminos descritos por los puntos y determinar, a la vista de los cambios producidos en la función de pérdidas, el número de iteraciones necesario para alcanzar la solución óptima. El número de iteraciones a realizar está bajo el control del usuario y es posible volver a la configuración obtenida en un paso anterior en cualquier momento del proceso.

## 5 Contraste de configuraciones. Técnicas GPA

El problema que resuelven los procedimientos GPA consiste en el cálculo de las rotaciones, traslaciones, reflexiones y ponderaciones que hagan que los cuadrados de las diferencias entre las distancias para cada par de puntos de dos representaciones diferentes de un mismo conjunto de objetos, sea mínima.

Los orígenes de GPA se deben a Green (1952), Cliff (1966) y Schönemann (1966), quienes resolvieron el problema de transformar mediante rotaciones y reflexiones una matriz tal que la suma de los cuadrados de las diferencias entre las distancias entre pares de puntos de las configuraciones que se comparan se minimice (ambas configuraciones han de estar en la misma dimensión y tener el mismo número de puntos representados). Mientras que las soluciones de Green y Cliff sólo funcionan para matrices de rango completo por columnas, Schönemann generalizó la solución al caso en que las matrices no cumplieran este requisito. En 1970, Schönemann y Carroll investigaron y resolvieron el problema de ajustar una matriz a otra usando, además de las transformaciones ortonormales ya citadas, traslaciones y reescalamiento uniforme.

En la década de los 70, los estudios en torno a las técnicas GPA se centraron en trabajar con más de dos configuraciones que podían estar en diferentes dimensiones.

En 1991, Jacques Commandeur generalizó GPA al caso de trabajar con configuraciones en las que falta toda la información sobre algún punto en alguna de las configuraciones analizadas.

El punto de partida en GPA es el siguiente resultado:

**Teorema 5.1** *Dados  $p$  puntos en un espacio  $m$ -dimensional, la suma de los cuadrados de las distancias entre esos  $n$  puntos es igual a  $n$  veces la suma de los cuadrados de las distancias entre los  $n$  puntos y su centroide.*

$$\sum_{j < k}^p (x_j - x_k)' (x_j - x_k) = p \sum_{j=1}^p (x_j - z)' (x_j - z) = p \sum_{j=1}^p x_j' x_j - p^2 z' z.$$

Que en versión matricial se formularía del siguiente modo:

**Teorema 5.2** *Dadas  $n$  configuraciones,  $X_j$ ,  $j = 1, 2, \dots, n$ , de orden  $(p \times m)$ , si se definen las matrices,  $M_j$ , diagonales, de orden  $(p \times p)$  con 1 en la diagonal si la correspondiente fila de  $X_j$  no es faltante, y 0 en otro caso (matrices simétricas e idempotentes); entonces:*

$$\sum_{j=1}^n \text{tr} (X_j - Z)' M_j (X_j - Z) = \sum_{j < k}^n \text{tr} (X_j - X_k)' M_j M_k \left( \sum_{j=1}^n M_j \right)^{-1} (X_j - X_k),$$

con

$$Z = \left( \sum_{j=1}^n M_j \right)^{-1} \left( \sum_{j=1}^n M_j X_j \right),$$

la matriz de centroides.

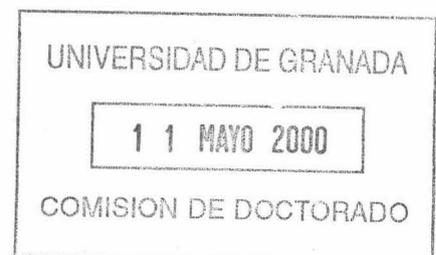
De este modo, son dos las funciones de pérdida que pueden considerarse para realizar un análisis GPA. Sean:

$\mathbf{1}_{p \times 1}$  un vector de unos,

$u_j$  un vector de traslación desconocido de orden  $m \times 1$ ,

$s_j$  un factor de escalamiento uniforme, y

$R_j$  una matriz desconocida de reflexión y rotación de orden  $m \times m$ .



Considerados todos los vectores,  $u_j$ , en una matriz  $U_{(m \times n)}$ , los factores de escalamiento uniforme en el vector  $s_{n \times 1}$  y las matrices ortonormales en la matriz  $R_{(nm \times m)}$  y aplicando la primera parte de la igualdad del teorema, se define la *Aproximación Centroide en GPA*, con la función de pérdida de mínimos cuadrados:

$$f(U, s, R) = \sum_{j=1}^n \text{tr} [s_j(X_j - \mathbf{1}u'_j)R_j - Z]' M_j [s_j(X_j - \mathbf{1}u'_j)R_j - Z], \quad (\text{II.10})$$

siendo

$$Z = \left( \sum_{j=1}^n M_j \right)^{-1} \left( \sum_{j=1}^n s_j M_j (X_j - \mathbf{1}u'_j) R_j \right). \quad (\text{II.11})$$

Análogamente, si se usa el segundo criterio de minimización, se define la *Aproximación Directa en GPA* cuya función de pérdida se formula como:

$$g(U, s, R) = \sum_{j < k}^n \text{tr} [\hat{X}_j - \hat{X}_k]' M_j M_k \left( \sum_{j=1}^n M_j \right)^{-1} [\hat{X}_j - \hat{X}_k], \quad (\text{II.12})$$

con

$$\hat{X}_j = s_j(X_j - \mathbf{1}u'_j)R_j.$$

En la literatura de GPA habitualmente se lleva a cabo la minimización de la función de pérdida de la aproximación directa. Sin embargo, en el caso de trabajar con datos faltantes, resulta de mayor utilidad utilizar la función de pérdida que define la aproximación del centroide, dado que los vectores de traslación y la configuración centroide pueden ser eliminados rápidamente en el proceso de minimización.

## 5.1 Análisis de la variabilidad en GPA con datos faltantes

La aportación de Commandeur a las técnicas GPA no consiste tan solo en la posibilidad de trabajar con configuraciones en las que hay datos faltantes, sino que su método ofrece diferentes descomposiciones de la variabilidad de los datos, que permiten el análisis de la aportación de cada dimensión, cada estímulo y cada configuración a la solución total. En este apartado se describen en líneas generales las medidas de bondad de ajuste utilizadas por Commandeur, así como las descomposiciones mencionadas.

Para simplificar los cálculos, se considerará que las  $n$  configuraciones se han normalizado, de manera que

$$\sum_{j=1}^n \text{tr} (X_j' C_j X_j) = n.$$

En el contexto de GPA sin datos faltantes son dos las medidas de ajuste que se vienen utilizando:

1. *Coefficiente de Correlación* entre las configuraciones óptimamente transformadas y la configuración centroide,

$$r^2 (\hat{X}_j, Z),$$

medida con la que se calcula la contribución relativa de cada configuración  $X_j$  a la solución global GPA y, a partir de donde se construye una medida global,

$$\frac{1}{n} \sum_{j=1}^n r^2 (\hat{X}_j, Z),$$

que es simétrica y no se ve alterada por diferencias en la escala de medida de ambas configuraciones.

2. *Cuadrado de la norma de la configuración centroide*. Esta configuración verifica

$$0 \leq \text{tr}(Z'Z) \leq 1,$$

y por tanto puede ser utilizada como una medida del ajuste global. Además, si las configuraciones  $X_j$  tienen sus columnas normalizadas a la unidad, se verifica:

$$r^2 (\hat{X}_j, Z) = s_j \text{tr}(Z'Z).$$

En el caso de trabajar con configuraciones con elementos faltantes, estas medidas no son equivalentes y se considera más apropiado utilizar como medida para el ajuste global la dada por el cuadrado de la norma de la configuración centroide, que desempeña un papel equivalente al del coeficiente de correlación de Pearson en el análisis de regresión, aunque en GPA no se ha desarrollado ningún contraste estadístico que determine si la proporción de varianza explicada por el modelo GPA es o no significativa.

A continuación expondremos tres descomposiciones diferentes de la cuantificación de la dispersión de los datos en términos de la norma de la matriz centroide,  $Z$ . Estas descomposiciones permitirán medir:

1. La contribución de cada estímulo al ajuste total de la solución.
2. La importancia relativa de cada configuración individual a la proporción de la suma de cuadrados explicada por el modelo GPA.
3. La contribución relativa de cada dimensión a la solución total GPA.

### Descomposición de la suma total de cuadrados respecto a cada estímulo

La descomposición se basa en la relación entre  $n$  puntos,  $x_j$ , y su centroide,  $z = \frac{1}{n} \sum_{j=1}^n x_j$ :

$$\sum_{j=1}^n x_j' x_j = n z' z + \sum_{j=1}^n (x_j - z)' (x_j - z).$$

Usando el modelo GPA,

$$s_j M_j (X_j - 1 u_j)' R_j = M_j Z + M_j E_j, \quad j = 1, \dots, n,$$

se llega a una descomposición del tipo:

$$SS_{total} = SS_{ajuste} + SS_{residual}.$$

Para llegar a ella, dado que se trabaja con  $n$  estímulos, se tiene  $\forall j = 1, \dots, n$

$$\begin{aligned} & \text{diag} \left( \sum_{j=1}^n [s_j M_j (X_j - 1 u_j)' R_j] [s_j M_j ((X_j - 1 u_j)' R_j)] \right) = \\ & = \text{diag} \left( \sum_{j=1}^n M_j Z (M_j Z)' \right) + \text{diag} \left( \sum_{j=1}^n (C_j E_j) (C_j E_j)' \right) \end{aligned}$$

y sumando en todos los términos

$$n = n \text{tr} Z Z' + \sum_{j=1}^n \text{tr} E_j' C_j E_j$$

Tabla II.1: Descomposición de la suma de cuadrados total respecto de cada estímulo

Estímulo	% $SS_{ajuste}$	$SS_{res}$	$SS_{total}$
1	$a_{11}$	$b_{11}$	$d_{11}$
2	$a_{22}$	$b_{22}$	$d_{22}$
3	$a_{33}$	$b_{33}$	$d_{33}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$n$	$a_{nn}$	$b_{nn}$	$d_{nn}$
$\sum_{i=1}^n$	$n \operatorname{tr} Z Z'$	$\sum_{j=1}^n \operatorname{tr} E_j' C_j E_j$	$n$

obteniéndose la tabla de análisis de la varianza II.1; descomposición válida siempre y cuando todas las configuraciones estén completas.

Los elementos que definen la tabla son:

- $a_{ii}$ , elemento diagonal de  $\sum_{j=1}^n M_j Z (M_j Z')$
- $b_{ii}$ , elemento diagonal de  $\sum_{j=1}^n (C_j E_j) (C_j E_j)'$
- $d_{ii}$ , elemento diagonal de  $\sum_{j=1}^n (s_j M_j (X_j - 1u_j') R_j) (s_j M_j (X_j - 1u_j') R_j)'$

### Descomposición respecto a cada configuración

El esquema de descomposición aquí presentado tiene la ventaja de que si se trata con configuraciones completas, la suma de cuadrados explicada por una configuración coincide con el cuadrado de la correlación entre los elementos de  $X_j$  óptimamente transformados y los de  $Z$ .

En el caso de datos faltantes, las componentes de la descomposición no son independientes, por lo que hay que restar los términos de covarianza,  $\operatorname{tr}(Z^t C_j E_j)$ , de la suma total de cuadrados asociada a cada configuración,  $s_j^2 \operatorname{tr} X_j' C_j X_j$ ;  $\forall j = 1, \dots, n$ :

Tabla II.2: Contribución relativa de cada configuración

Estímulo	% $SS_{ajuste}$	$SS_{res}$	$SS_{total}$
1	$s_1 \text{tr } Z'C_1X_1R_1$	$\text{tr } Z'_1C_1E_1$	$s_1^2 \text{tr } X'_1C_1X_1$
2	$s_2 \text{tr } Z'C_2X_2R_2$	$\text{tr } Z'_2C_2E_2$	$s_2^2 \text{tr } X'_2C_2X_2$
n	$s_n \text{tr } Z'C_nX_nR_n$	$\text{tr } Z'_nC_nE_n$	$s_n^2 \text{tr } X'_nC_nX_n$
⋮	⋮	⋮	⋮
$\sum_{j=1}^n$	$\text{tr } Z' CZ$	$\sum_{j=1}^n \text{tr } E'_j C_j E_j$	$n$

$$s_j^2 \text{tr } X'_j C_j X_j - \text{tr } Z' C_j E_j = s_j \text{tr } Z' C_j X_j R_j + \text{tr } E'_j C_j E_j.$$

Sumando en  $j$ , resulta:

$$n = \text{tr } Z' CZ + \sum_{j=1}^n \text{tr } E'_j C_j E_j.$$

Los resultados de dicha descomposición se recogen en la tabla II.2

### Contribución relativa de cada dimensión en GPA

Para cada dimensión,  $r = 1, \dots, p$ ; se verifica la siguiente relación:

$$\begin{aligned} & \text{diag} \left[ \sum_{j=1}^n (s_j C_j X_j R_j K)' (s_j C_j X_j R_j K) \right] = \\ & = \text{diag}(\Lambda) + \text{diag} \left[ \sum_{j=1}^n (s_j X_j R_j K - ZK)' (s_j X_j R_j K - ZK) \right]. \end{aligned}$$

Y, sumando en todas las dimensiones,  $r = 1, \dots, p$ ; se obtiene:

$$n = \text{tr } Z' CZ + \sum_{j=1}^n \text{tr } E'_j C_j E_j$$

Tabla II.3: Contribución relativa de cada dimensión en GPA

Estímulo	$\% SS_{ajuste}$	$SS_{res}$	$SS_{total}$
1	$e_{11}$	$f_{11}$	$g_{11}$
2	$e_{22}$	$f_{22}$	$g_{22}$
3	$e_{33}$	$f_{33}$	$g_{33}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$p$	$e_{pp}$	$f_{pp}$	$g_{pp}$
$\sum_{r=1}^p$	$\text{tr } Z' CZ$	$\sum_{j=1}^n \text{tr } E_j' C_j E_j$	$n$

siendo  $K$  y  $\Lambda$  los vectores y valores propios de  $Z' CZ$ ,

$$Z' CZ = K \Lambda K'.$$

Llamando

- $e_{ii}$ , elemento diagonal de  $\Lambda$
- $f_{ii}$ , elemento diagonal de  $\sum_{j=1}^n (s_j X_j R_j K - ZK)' C_j (s_j X_j R_j K - ZK)$
- $g_{ii}$ , elemento diagonal de  $\sum_{j=1}^n (s_j C_j X_j R_j K)' (s_j C_j X_j R_j K)$

se obtiene la tabla II.3

## Capítulo III

# Determinación de la dimensionalidad en MDS

Todos los procedimientos descritos hasta aquí han sido desarrollados con el propósito de obtener mejores soluciones que las producidas por los algoritmos MDS clásicos. Tal y como se ha indicado, se formulan nuevos algoritmos de minimización que permitan obtener soluciones más próximas al mínimo global y se determinan procedimientos que faciliten el análisis de soluciones en dimensiones superiores, aunque no puedan ser visualizadas. Sin embargo, siguen vigentes problemas clásicos de MDS como es el problema de determinación de la dimensión apropiada para la representación.

En este capítulo nos centraremos precisamente en ese aspecto. El concepto de dimensionalidad o número de dimensiones de la configuración, se refiere al número de coordenadas con que representar los objetos analizados a partir de una determinada medida de disimilaridad. Este valor no tiene que coincidir exactamente con el número de características relevantes en los objetos. Aunque una característica o atributo tenga efecto en la configuración, es posible que no se refleje en un nuevo eje, debido a que esté correlada con otra dimensión, o a que sólo afecte a un subconjunto de estímulos, etc., por lo que, si su presencia se puede hacer efectiva por otros medios, no hay por qué mostrarla como una dimensión adicional, es decir, el número de direcciones interpretables puede ser mayor que la dimensionalidad del espacio. Los procedimientos que se siguen actualmente para la elección de la dimensión de representación pueden resumirse en los siguientes puntos:

1. *Análisis de los cambios producidos en las medidas de bondad de ajuste* al considerar dimensionalidades diferentes (coeficiente de alienación en MINISSA, STRESS en POLYCON y KYST, correlaciones y/o varianzas explicadas en INDSCAL,

SSTRESS en ALSCAL y error estándar y verosimilitud en MULTISCALE). En general, (S)STRESS disminuirá y las correlaciones aumentarán conforme vaya creciendo el número de dimensiones de la configuración.

Si estos valores se representan sobre el eje de ordenadas frente al número de dimensiones que se están considerando en el eje de abscisas, en un gran número de casos aparecerá un "codo" en la curva. El número de dimensiones correspondiente al eje X de dicho codo, se corresponde con el número máximo de dimensiones a considerar en dicha representación.

2. *Utilización de resultados obtenidos mediante las técnicas de MonteCarlo.* Un grupo de investigadores han generado espacios de estímulos usando conjuntos de datos obtenidos a partir de tablas de números aleatorios. Una vez analizados tales datos, han determinado los valores de STRESS según el número de datos de entrada. Como conclusión de estas investigaciones, se han dado una serie de valores críticos que permiten determinar el número de dimensiones apropiado en función del número de objetos que se está analizando.

La idea subyacente del razonamiento empleado es que existe una configuración real de los datos en alguna dimensión,  $p$ . Se supone que las proximidades se han generado a partir de las distancias en la configuración exacta pero con errores aleatorios de algún tipo,  $e$ , que se incorporan al nivel controlado. Cuando las proximidades son representadas en distintas dimensiones, la gráfica de los puntos de STRESS depende de  $p$  y de  $e$ . Así, cada combinación  $(p, e)$  se corresponde con una apariencia particular de la gráfica de puntos. Usando el método de Monte Carlo se generan muchas configuraciones y se representan en cada escala, de forma que las gráficas de puntos se catalogan con arreglo al par  $(p, e)$ . Dado un conjunto de disimilaridades entre puntos y realizado un análisis MDS, se comparan los resultados obtenidos con los dados por las técnicas de MonteCarlo, infiriendo el valor de  $p$  y de  $e$  de manera que se obtenga una gráfica lo más parecida a los datos que se poseen. Si la representación de puntos es muy diferente de las salidas del método de MonteCarlo, podría deberse a una convergencia incompleta o a un mínimo local de STRESS. Igualmente se observa que para grandes valores de  $e$ , las diferencias son prácticamente inapreciables para los puntos representados en diferentes salidas de  $p$ , por lo que se puede concluir que no aporta ninguna información nueva incrementar el nivel de error en los datos para determinar la dimensión de la representación.

3. *Interpretabilidad.* Dado el trabajo y la intuición que requiere la interpretación de una dimensión, es posible que las dimensiones que no puedan ser interpretadas no existan. Tal y como señalábamos en la introducción de este capítulo, en las representaciones obtenidas mediante técnicas de MDS las dimensiones de representación no siempre coinciden con las dimensiones de interpretación.

Un aspecto muy importante de MDS lo constituyen las direcciones en la configuración obtenida, las cuales tienen unas interpretaciones muy interesantes. Dicho de otro modo, las posiciones en la configuración pueden asociarse con alguna característica del elemento que es representado. De hecho, estudiar esas características es una de las razones del empleo de MDS. Una forma práctica de descubrir esas características es simplemente observar la configuración, recordando qué cosas se conocen sobre los objetos. Sin embargo, este método es limitado por cuanto requiere de un gran esfuerzo memorístico y una gran capacidad de representación mental, así como limita las dimensiones en las que se puede trabajar al no imaginar la mente humana más de tres.

Para contrarrestar tales limitaciones, el método que normalmente se utiliza está basado en la regresión lineal. Supongamos que tenemos una variable asociada a un elemento en cuestión, de forma que se sospecha que esa variable tiene alguna relación con la posición de dicho elemento en la configuración. Una forma de poder confirmarlo es mediante regresión lineal múltiple de esa variable sobre las otras variables que constituyen cada coordenada del punto, o sea, si la variable es  $V$  y si suponemos una configuración tridimensional, fijándonos en el  $i$ -ésimo ítem, tendremos por un lado el valor de  $V$  asociado a éste,  $v_i$ , y por otro, sus coordenadas en la configuración,  $(x_{i1}, x_{i2}, x_{i3})$ . Se tratará pues de ajustar

$$v_i = a + b_1x_{i1} + b_2x_{i2} + b_3x_{i3}.$$

El coeficiente de correlación múltiple dará una medida de lo bien que se ajusta el hiperplano a la variable. Así pues, la idea de utilizar regresión para facilitar la interpretabilidad de la configuración consiste en que, una vez representada la nube de puntos en una dimensión adecuada y en un espacio adecuado, de forma que el valor asociado de la función de error haya resultado suficientemente pequeño como para que dicha configuración sea representativa, se consideran una serie de variables,  $V$ , que puedan suministrar alguna información en cuanto a la posición de los puntos en el espacio en que se representan, ya que como antes afirmábamos,

el sistema de referencia en el que estos se representan, no es significativo en cuanto a la interpretación de los resultados. Al efectuar regresión de esas variables con las coordenadas de los puntos, se elegirán de entre ellas, las que resulten con mayor coeficiente de correlación, representando sus correspondientes rectas de regresión en el espacio de la configuración, e interpretando la posición de los puntos, no respecto al sistema de referencia, sino según la posición de éstos respecto de las rectas representadas.

Este método no solo se usa en el sentido descrito (o sea, una vez obtenida la dimensionalidad del problema, según el valor de la función de error, se representan en el espacio obtenido, tanto los puntos, como las rectas de regresión antes mencionadas), sino que además ayuda en muchas ocasiones a elegir igualmente la dimensionalidad del problema. Así pues, una vez obtenidos los puntos, si al efectuar regresión y calcular los cosenos directores de la recta con respecto a los ejes de coordenadas, resulta que sólo algunos de estos cosenos son suficientemente grandes mientras que el resto no lo son, podríamos interpretar este hecho como que, a pesar de que el valor de la función que mide el error (STRESS, SSTRESS, etc.) en la dimensión dada es pequeño, no todas las dimensiones resultan susceptibles de ser interpretadas, por lo que se elegiría una nueva dimensionalidad para el problema, de acuerdo con el número de cosenos directores suficientemente grandes, volviendo por tanto a obtener las coordenadas de los puntos en esa nueva dimensión, así como una nueva representación de la recta de regresión en el subespacio elegido.

En definitiva, una vez que se ha deducido lo más razonablemente posible la dimensionalidad del problema, se pasaría a estudiar las relaciones de los puntos representados con respecto a esos nuevos ejes que se han obtenido. Lógicamente, la elección de las variables,  $V$ , se realizará con arreglo a los factores que interese poner de manifiesto en el problema en cuestión y con arreglo a los cuales queramos interpretar las posiciones de los puntos de la configuración obtenida.

4. *Repetición del análisis.* Junto a la interpretabilidad y el ajuste, la selección de la dimensionalidad depende de la facilidad de uso, la estabilidad de la estructura y la generalidad entre métodos alternativos de recolección de datos. Para analizar la robustez de la dimensionalidad elegida, las líneas de acción empleadas son varias:
  - Repetir el experimento completo y comprobar si el nuevo conjunto de datos proporciona la misma solución obtenida en el primer caso.

- Dividir el conjunto de datos considerado en dos grupos o más y examinar si las soluciones obtenidas a partir de cada subgrupo de datos son o no equivalentes.
- Incluir estímulos repetidos en el conjunto original y analizarlos conjuntamente, considerándolos diferentes. La proximidad entre los puntos de la representación correspondientes a los estímulos repetidos, no solo proporciona una medida del ruido de los datos, sino también una indicación de la dimensionalidad apropiada.

Además de la determinación de la dimensión, en la práctica resulta conveniente analizar soluciones MDS en diferentes dimensiones y comparar la información que aporta cada una de ellas aunque, la forma en que cambia la interpretación de una dimensión a la siguiente puede ser complicada. En MDS es posible, por ejemplo, que en dos dimensiones no se obtenga una interpretación de los resultados y sin embargo, exista una clara interpretación de la solución si se considera la representación en tres dimensiones. Esto es debido a que la relación entre las diferentes dimensionalidades es más complicada en este método que en otros. Por ejemplo, en Análisis de componentes principales, la configuración en  $p - 1$  dimensiones se puede obtener a partir de la configuración en dimensión  $p$  sin más que omitir la última coordenada, ocurriendo algo parecido en Análisis factorial. En MDS, la cuestión es más complicada. Omitir la segunda coordenada en una configuración bidimensional, es tanto como proyectar sobre el eje de abscisas los valores obtenidos. En MDS, la configuración unidimensional se puede obtener algunas veces de forma aproximada, proyectando sobre alguna recta en el espacio bidimensional, aunque en otros casos esto no es posible. En tres dimensiones, a partir de la configuración tridimensional, sólo en algunos casos se puede obtener la bidimensional proyectando aquella sobre algún plano del espacio tridimensional.

Todo ello plantea la necesidad de desarrollar un procedimiento que, simultáneamente, proporcione resultados en dimensiones consecutivas, estableciendo una relación entre las configuraciones que permita mantener en una dimensión superior las propiedades observadas en la representación anterior. La determinación de la dimensión apropiada de la configuración vendrá dada a partir del análisis global de las configuraciones estimadas, en función del porcentaje de variabilidad global de los datos que es explicado por cada una de las dimensiones consideradas.

En esta memoria se plantea una solución similar a la considerada en el problema de regresión mediante polinomios ortogonales: *Definir configuraciones en dimensiones consecutivas que permitan ir almacenando los resultados obtenidos en las etapas anteriores.* Se presenta un algoritmo con el que, a partir de las medidas de disimilaridad entre  $n$

objetos, se determinan  $n - 1$  configuraciones en dimensiones consecutivas, de manera que dada una configuración en dimensión  $p$ , al pasar a la configuración en dimensión  $p + 1$  se pueden mantener las conclusiones obtenidas hasta el momento. Además, el algoritmo proporciona una medida de la bondad del ajuste de cada configuración en función del número de dimensiones consideradas. Es decir, permite medir la aportación que proporciona la nueva dimensión a la configuración que se está estimando, por lo que se tomará como dimensionalidad adecuada, aquella a partir de la cual se produzca un descenso en la tendencia de la curva que representa el ajuste de cada configuración en términos del número de dimensiones.

Tras el estudio de los diferentes procedimientos empleados en MDS, se observa una clara diferencia entre las soluciones proporcionadas por los métodos clásicos (Torgerson, 1952; Gower, 1966) y las de los procedimientos basados en técnicas de mínimos cuadrados (KYST, de Kruskal, Young y Seery, 1973; SMACOF de Heiser y De Leeuw, 1977). Las soluciones clásicas están anidadas, siempre es posible obtener una representación  $p$ -dimensional para  $n$  puntos de manera que la configuración calculada coincida con las  $p$  primeras coordenadas de la solución  $(p + 1)$ -dimensional siendo  $p$  cualquier valor entre 1 y  $n - 1$ . La búsqueda de esta propiedad en las soluciones MDS obtenidas por métodos de mínimos cuadrados queda justificada dentro del propio entorno de este tipo de análisis, ya que los procedimientos mínimo cuadráticos de MDS pueden considerarse, a grandes rasgos, como una aplicación no lineal de un conjunto de disimilaridades en un conjunto de distancias definidas sobre un conjunto discreto de elementos que podemos transformar en una proyección sobre diferentes dimensiones y considerar por tanto una aproximación simultánea mediante diferentes configuraciones. Además, tal y como hemos mencionado anteriormente, los modelos anidados no sólo permiten observar las distintas características que puedan presentar los datos sino que también ofrecen una medida de la variabilidad de los datos explicada por cada una de las dimensiones consideradas, con lo que el procedimiento de elección de la dimensión apropiada resulta inmediato, eligiéndose aquella dimensión que en general explique un porcentaje apropiado de la variabilidad total existente en las proximidades analizadas.

El problema de obtener soluciones anidadas en los modelos de mínimos cuadrados para MDS métrico ya ha sido tratado por Heiser (1995) obteniendo soluciones anidadas con el algoritmo SMACOF.

## 1 Modelo anidado para MDS métrico según el algoritmo SMACOF

En contraste con los modelos de MDS usuales, en los que hay que explicar el comportamiento de  $K$  matrices de disimilaridades,  $\{\Delta_k, k = 1, 2, \dots, K\}$ , mediante una única configuración en dimensión predeterminada, la introducción del anidamiento en las funciones de pérdida de MDS lleva implícita la idea de aproximar una única matriz de disimilaridades,  $\Delta = (\delta_{ij})_{n \times n}$ , mediante varias matrices de distancias  $D(X_q) = (d_{ij}(X_q))_{n \times n}$ , siendo  $q = 1, 2, \dots, n-1$ , la dimensionalidad de la configuración.

El estudio realizado por Heiser, así como el desarrollo de nuestro modelo en secciones posteriores, se centra en el análisis de las distancias euclídeas definidas en términos matriciales como:

$$d_{ij}(X) = [\text{tr}(X' E_{ij} X)]^{1/2},$$

siendo  $E_{ij} = (e_i - e_j)(e_i - e_j)'$  de orden  $n \times n$ , y  $e_i$  la  $i$ -ésima columna de la matriz identidad. Sin pérdida de generalidad, dado que las traslaciones no alteran las distancias, se supondrá que la configuración está centrada en el origen de coordenadas.

El problema de obtener un conjunto de configuraciones anidadas para MDS métrico con mínimos cuadrados para el algoritmo SMACOF, se formula como el problema de obtener un conjunto de matrices

$$X = \{X_1, X_2, \dots, X_{n-1}\}$$

verificando  $X_q = XU_q$ , siendo  $U_q$  una matriz diagonal, de manera que el valor

$$S(X) = \frac{1}{n-1} \sum_{q=1}^{n-1} \|\Delta - D(X_q)\|^2 \quad (\text{III.1})$$

sea mínimo, exigiendo que la configuración esté normalizada por columnas

$$\text{diag}(X'X) = I$$

y que las matrices  $U_q$  verifiquen la relación,

$$U_q = \Phi C_q,$$

siendo  $\Phi$  una matriz  $p \times p$  diagonal que contiene las longitudes óptimas de los ejes de representación y las matrices  $C_q$ ,  $q = 1, 2, \dots, n-1$ , matrices diagonales tales que la  $q$ -ésima matriz sea una matriz con  $q$  unos y  $p - q$  ceros en la diagonal.

Es decir, el algoritmo de Heiser origina un conjunto de configuraciones

$$X_q = X\Phi C_q, \quad q = 1, 2, \dots, n-1$$

con  $X$  un espacio común  $(n-1)$ -dimensional,  $\Phi$  una matriz  $(n-1) \times (n-1)$  con las ponderaciones individuales de las dimensiones, y  $C_q = \text{diag} \left( \sum_{p=1}^q e_p \right)$ .

La obtención de tales configuraciones se lleva a cabo minimizando la función objetivo mediante técnicas de Mayorización Iterativa. La utilización de este procedimiento define un proceso convergente que conduce al cálculo de un mínimo local de  $S(X)$ . Si se denota la función de mayorización en el punto  $\underline{X}$  por  $M(X, \underline{X})$ , resulta:

$$S(X) \leq M(\bar{X}|\underline{X}) + \frac{n}{n-1} \sum_{q=1}^{n-1} \|\bar{X}_q - X\Phi C_q\|^2, \quad (\text{III.2})$$

donde, en el segundo término de la función de mayorización, se ajusta la estructura anidada a un conjunto de transformadas de Guttman, cada una de diferente dimensión  $\bar{X} = \{\bar{X}_1, \bar{X}_2, \dots, \bar{X}_{n-1}\}$ , siendo  $\bar{X}_q$ , la transformada de Guttman; el mejor sucesor sin restricciones de alguna estimación previa  $\underline{X}_q$ . Dicha transformada viene dada por:

$$\bar{X}_q = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n b_{ij} \delta_{ij} \underline{X}_q$$

con

$$b_{ij} = \begin{cases} 0 & \text{si } d_{ij}(\underline{X}_q) = 0 \\ \frac{\delta_{ij}}{d_{ij}(\underline{X}_q)} & \text{en otro caso.} \end{cases}$$

Puede observarse que la transformada de Guttman no se ve alterada ni por el centrado de  $\underline{X}_q$  ni por posibles columnas de ceros que pudieran existir en esta configuración. Por otra parte, dado que la transformada de Guttman depende de anteriores estimaciones,  $\underline{X}_q = X\Phi C_q$ , el proceso de optimización sobre el espacio común,  $X$ , y los pesos dimensionales,  $\Phi$ , tiene que ser repetido hasta la convergencia. Al igual que en los modelos con diferencias individuales, no hay necesidad de una rotación explícita,  $T$ , de la configuración, dado que el proceso de mínimos cuadrados delimitará una implícitamente.

### 1.1 Convergencia del modelo anidado para MDS métrico

Para analizar la convergencia de su modelo, Heiser aplica la descomposición de STRESS dada por Gower (1984, 1986):

$$S(\hat{X}) = \|\Delta\|^2 - \frac{1}{n-1} \sum_{q=1}^{n-1} \|D(\hat{X}_q)\|^2$$

que permite expresar la suma de los cuadrados de las disimilaridades (Dispersión Total) en una componente de pérdida por el ajuste y una componente de la dispersión explicada. A partir de este resultado, se obtiene una descomposición de la componente de la dispersión explicada en  $n - 1$  componentes, correspondientes a cada una de las dimensiones, que reflejan la importancia relativa de cada eje en la solución global.

$$\begin{aligned} \|\Delta\|^2 &= S(\hat{X}) + \frac{1}{n-1} \sum_{q=1}^{n-1} \|D(\hat{X}_q)\|^2 = \\ &= S(\hat{X}) + 2n \operatorname{tr} \left[ \frac{1}{n-1} \sum_{q=1}^{n-1} C_q \right] \hat{\Phi}^2. \end{aligned} \quad (\text{III.3})$$

En su trabajo, Heiser aplica su modelo a dos situaciones diferentes. En la primera analiza las distancias euclídeas entre siete puntos en una configuración bidimensional centrada en el origen de coordenadas. Los puntos que la integran fueron determinados generando valores aleatoriamente de dos distribuciones uniformes. En el segundo caso, considera disimilaridades evaluadas sobre una escala ordinal para medir el grado de semejanza entre diferentes estados emocionales. Esta medición fue llevada a cabo por 30 individuos independientemente, considerándose en el análisis la matriz promedio de las 30 matrices de disimilaridades obtenidas. Además se realizaron varios análisis, tomando diferentes configuraciones iniciales, generadas cada una de ellas, de manera aleatoria. Obtiene las siguientes conclusiones:

1. Por una parte, establece un nuevo procedimiento para resolver el problema de determinación de la dimensionalidad correcta en la solución MDS. Heiser propone como criterio de selección, una vez analizadas las disimilaridades entre  $n$  objetos, el siguiente:

Escoger la dimensión de representación,  $p$ , de tal modo que las últimas  $n - (p + 1)$  componentes de la pérdida de ajuste sean casi iguales; esta

“*igualdad*” indica que las nuevas dimensiones tienen poco o nada que contribuir a una mejor aproximación de las disimilaridades.

2. En segundo lugar, hace referencia a la persistencia del problema de los mínimos locales en MDS anidado. En este aspecto Heiser señala lo apropiado que resulta tomar como configuración inicial la dada por los procedimientos clásicos de MDS.

Señala también que dado que es conocido que la seriedad del problema de los mínimos locales decrece conforme aumenta la dimensión de la configuración, podría ser posible mitigar éste ponderando adecuadamente las  $n - 1$  componentes de STRESS tal y como se ha definido en el análisis realizado,

$$S(X) = \frac{1}{n-1} \sum_{q=1}^{n-1} \| \Delta - D(X_q) \|^2,$$

por un factor  $q$  o  $q^2$ .

## 2 Modelo anidado para las soluciones MDS no métricas

El modelo de Heiser fue diseñado para tratar datos métricos, minimizando las diferencias entre las disimilaridades de partida y las distancias entre los puntos de la configuración estimada en cada etapa del proceso de minimización. En contra del principio de *representación óptima* de Fisher de que, para la representación de los datos,

- (a) el modelo debería ajustarse a los datos tanto como fuese posible en sentido mínimo cuadrático, y
- (b) las características de las medidas que originan los datos deberían ser estrictamente preservadas,

el modelo de Heiser no considera transformaciones sobre la matriz de disimilaridades.

El objetivo de esta sección es solucionar el problema de determinación de la dimensionalidad apropiada en los análisis de MDS para datos no métricos. Para ello se propone la obtención de soluciones anidadas en MDS no métrico, considerando transformaciones monótonas sobre las disimilaridades de partida, es decir, considerando una transformación,  $t : \Delta \rightarrow \hat{D}$ , que sea una función que mantenga las restricciones de medida y que establezca una relación mínimo cuadrática entre las disparidades y las distancias.

El modelo, que se expone a continuación, está basado en la función de pérdida SSTRESS, que será minimizada siguiendo los criterios empleados en ALSCAL. La elección de este algoritmo para el proceso de minimización queda justificada por las ventajas que reporta frente al resto de modelos para MDS y que ya indicamos en el primer capítulo de esta memoria.

## 2.1 Función de pérdida SSTRESS para el algoritmo anidado

Heiser define su modelo, en términos de la función STRESS, a partir de la aproximación de una única matriz de disimilaridades al conjunto de matrices de distancias obtenidas a partir de las configuraciones anidadas. En este caso, al permitir transformaciones de las disimilaridades, la definición de la función de pérdida varía considerablemente. Se define la función SSTRESS para el modelo anidado como la aproximación entre las distancias obtenidas para cada configuración y las disparidades, estimaciones mediante regresión monótona mínimo cuadrática de las disimilaridades iniciales.

De este modo, la función de pérdida para el modelo anidado, vendrá dada por:

$$SS(X) = \frac{1}{n-1} \sum_{q=1}^{n-1} \|\hat{D}^2(X_q) - D^2(X_q)\|^2,$$

siendo  $\hat{D}^2$  la matriz de disparidades, obtenidas mediante técnicas de regresión monótona a partir de las distancias euclídeas  $D^2$ .

Dada la carencia de aditividad de la distancia euclídea al considerar dos dimensiones consecutivas,  $q$  y  $q+1$ , si  $x_{q+1}$  denota la última columna de la matriz con las coordenadas de la configuración en dimensión  $q+1$ ,  $X_{q+1}$ , se tiene entonces

$$d_{ij}(X_{q+1}) = \left[ d_{ij}^2(X_q) + x_{q+1}^t E_{ij} x_{q+1} \right]^{1/2},$$

a partir de donde, la reformulación dada para SSTRESS quedaría

$$SS(X) = \frac{1}{n-1} \sum_{q=1}^{n-1} \sum_{i=1}^n \sum_{j=1}^{i-1} \left( \hat{d}_{ijq} - \sum_{p=1}^q x_p^t E_{ij} x_p \right)^2.$$

Con lo que, dado un procedimiento ALSCAL unidimensional, el proceso de minimización para el modelo no métrico anidado consistirá en un esquema que circule a través de las dimensiones, minimizando cada vez el cuadrado de las diferencias entre las disparidades y las distancias, respecto al actual conjunto de coordenadas activo.

## 2.2 Algoritmo para la obtención de soluciones anidadas

El proceso de determinación de las configuraciones anidadas comienza con el cálculo de una configuración inicial. Debido a que por propia naturaleza presenta la propiedad de anidamiento y dadas las buenas características que presenta al ser considerada como configuración inicial para cualquier procedimiento de MDS, consideraremos la representación proporcionada por el algoritmo de Torgerson. A partir de ella se estiman las disparidades correspondientes a cada una de las dimensiones mediante técnicas mínimo cuadráticas de regresión monótona y comienza entonces el procedimiento iterativo de minimización de la función objetivo. En este proceso se diferencian dos etapas:

- estimación de la configuración, y
- estimación de las disparidades;

que se van alternando hasta que no se produzca una mejora significativa en la función de minimización.

Pasemos entonces a describir cada una de las etapas que definen el algoritmo:

### Fase de inicialización

Se calculan los valores iniciales de  $X$  directamente a partir de la matriz de disparidades,  $\Delta$ , utilizando la solución algebraica de Schönemann (Schönemann, 1972) que, aplicada a una única matriz de disimilaridades, se reduce al procedimiento clásico de MDS (Torgerson, 1952), por lo que la solución inicial obtenida tiene la característica deseada del anidamiento (Heiser, 1995). El procedimiento de cálculo puede resumirse del siguiente modo:

Se define la matriz de productos escalares,  $P = (p_{ij})$ , así:

$$p_{ij} = \frac{-1}{2} (\delta_{ij}^2 - \delta_{i.}^2 - \delta_{.j}^2 + \delta_{..}^2)$$

siendo  $\delta_{.j}$ , la media de las disimilaridades en la primera variable,  $\delta_{i.}$ , la media de las disimilaridades en la segunda variable y  $\delta_{..}$ , la media global de las disimilaridades.

Atendiendo al teorema de Young y Householder (1938), se toma como solución la matriz,  $X$ , de vectores propios (truncada según la dimensión) obtenida a partir de la diagonalización de  $P$  y normalizada para que el cuadrado de la longitud de tales vectores coincida con los valores propios asociados.

### Fase de representación óptima

En esta etapa se calculan de manera óptima los cuadrados de las observaciones,  $\Delta$ , para obtener las disparidades,  $\hat{D}$ . Estos valores han de ser tales que cumplan las restricciones de medida seleccionadas y sean estimaciones por mínimos cuadrados de los cuadrados de las distancias  $D$ . En el desarrollo de esta etapa se distinguen tres fases:

1. *Cálculo del cuadrado de las distancias euclídeas,  $D$* , usando la configuración,  $X$ , obtenida en la etapa anterior

$$d_{ij}^2(X_q) = \sum_{a=1}^q (x_{ia} - x_{ja})^2.$$

2. *Determinación de las disparidades óptimamente escaladas,  $D^*$* , a partir de las distancias,  $D$ , las observaciones,  $\Delta$ , y las restricciones de medidas relevantes. Dado que la configuración  $X$  se mantiene fija, la estimación de las disparidades resulta independiente de las características de anidamiento de la solución. Tan solo habrá que considerar las distintas dimensiones con las que trabajamos simultáneamente, obteniéndose una estimación diferente para cada una de ellas.

El procedimiento de estimación de las disparidades será el algoritmo de transformación monótona de mínimos cuadrados de Kruskal para datos ordinales. Los procedimientos no métricos de MDS pretenden conseguir que las distancias estimadas estén en el mismo orden que los datos de entrada. Para conseguir esto, se precede a la estimación de las disparidades, que son valores que sí verifican dichas restricciones de orden. Los pasos a dar son los siguientes:

- (a) Se ordenan las distancias de acuerdo a las observaciones.
  - (b) Empezando por las primeras distancias, si aparece una secuencia de distancias que no se ajusta al orden, se reemplazan esas distancias por su media aritmética.
  - (c) Se repite el paso anterior hasta que todas las distancias formen una secuencia que sí se ajuste al orden.
3. *Normalización.* Con objeto de que no haya errores de imprecisión por la escala de medida utilizada, se minimiza la expresión normalizada de SSTRESS para el modelo anidado.

$$SS(X) = \frac{1}{n-1} \sum_{q=1}^{n-1} \left[ \frac{\sum_{i=1}^n \sum_{j=1}^{i-1} \left( \hat{d}_{ijq} - \sum_{p=1}^q x_p^t E_{ij} x_p \right)^2}{\sum_{i=1}^n \sum_{j=1}^{i-1} \hat{d}_{ijq}^4} \right].$$

Para ello se reajusta la longitud del vector de disparidades:

$$\hat{D}_q = D_q^* (D_q^t D_q) (D_q^t D_q^*)^{-1}$$

siendo

$D_q^*$ , un vector columna con las  $n(n-1)/2$  disparidades estimadas para la configuración en dimensión  $q$ .

$D_q$ , un vector columna con las  $n(n-1)/2$  distancias para la configuración en dimensión  $q$ .

### Fase de terminación

En esta etapa basta con determinar el valor de SSTRESS de la iteración actual y compararlo con el obtenido en la etapa anterior. Si la razón de mejora de SSTRESS es suficientemente baja como para garantizar la finalización del proceso, se imprimen los resultados y se detiene el estudio; si no lo es, se pasa a la etapa siguiente.

El algoritmo para la obtención de soluciones anidadas en MDS no métrico puede considerarse del tipo de mínimos cuadrados alternantes (ALS). El conjunto de parámetros que se quiere estimar se divide en distintos grupos y se procede a la minimización de la función de pérdida respecto de cada uno de ellos, considerando fijos el resto de los mismos. La función de pérdida es una función acotada inferiormente por 0 y, dado que en cada fase se produce una mejora de la función de pérdida, ya que nuestro esquema de minimización es el mismo que el de ALSICAL cuya convergencia ya ha sido estudiada (Takane et al. 1977), estamos en condiciones de asegurar la obtención de un mínimo local de la función SSTRESS para el algoritmo anidado.

Ante esta argumentación, faltaría demostrar la equivalencia entre el proceso de minimización del algoritmo ALSICAL y el de nuestro algoritmo. Dado que en los dos se procede a la estimación de los puntos determinando las coordenadas que los definen una a una, podría parecer inicialmente que se realiza el mismo procedimiento en los dos

casos, sin embargo esto no es así. En ALSCAL, la determinación de las coordenadas se realiza una a una para cada punto. Una vez que se obtiene la posición de un objeto en la configuración, se pasa a determinar la del siguiente ítem y así hasta obtener la representación completa. En el modelo anidado esta estimación se realiza dimensión a dimensión. Determinada la solución unidimensional, el proceso consiste en ir minimizando la función de pérdidas según un esquema iterativo que va determinando las diferentes configuraciones anidadas. En el modelo anidado, en cada etapa del proceso se estiman las coordenadas que establecen la posición de cada objeto en una dimensión determinada, manteniendo las estimaciones realizadas para las dimensiones inferiores; es decir, en cada etapa se determina una y solo una coordenada para cada objeto, la que lo posiciona en una dimensión superior. Sin embargo, a pesar de las diferencias entre ambos modelos, los resultados que se obtienen en el proceso de minimización son equivalentes en el sentido de que ambos conducen a la obtención de un mínimo local de la función de pérdidas, lo que se demuestra con el siguiente resultado:

**Teorema 2.1** *En el proceso de minimización de la función SSTRESS, según el algoritmo ALSCAL, se verifica:*

$$\min_{i,j} SS(X) = \min_i \left[ \min_j SS(X) \right] = \min_j \left[ \min_i SS(X) \right],$$

siendo

$$SS(X) = \sum_{i=1}^n \sum_{j=1}^{i-1} (\hat{d}_{ij}^2 - d_{ij}^2)^2.$$

DEMOSTRACIÓN:

El resultado que acabamos de enunciar, simplemente dice que, en el proceso de minimización de la función SSTRESS, no importa el orden que se siga al seleccionar las variables que se van a estimar.

Para simplificar la notación, consideraremos la función SSTRESS como una función de un conjunto de  $m = n(n-1)/2$  variables, que reescribiremos así

$$SS(X) = f(z) = f(z_1, z_2, \dots, z_m).$$

Dadas las propiedades de continuidad y derivabilidad de la función SSTRESS, acotada inferiormente por 0, sabemos que existe un mínimo global (aunque no seamos capaces de determinarlo),

$$\exists z^0 = (z_1^0, z_2^0, \dots, z_m^0) \quad \text{t.q.} \quad f(z^0) \leq f(z) \quad \forall z \in R^m.$$

A partir de esta función en  $m$  variables, podemos definir  $m$  nuevas funciones univariantes como:

$$g_i(z) = f(a_1, a_2, \dots, z_i, \dots, a_{m-1}, a_m) \quad i = 1, 2, \dots, m$$

con  $\{a_1, a_2, \dots, a_{m-1}, a_m\}$  escalares.

Por las propiedades de continuidad y derivabilidad de  $f$ , las funciones  $g_i$  también son continuas y diferenciables, y acotadas inferiormente por 0, por lo que también se puede determinar su mínimo. Sea  $x^0$  dicho mínimo para una de estas funciones, entonces

$$g_i(x^0) = f(a_1, a_2, \dots, x_i^0, \dots, a_{m-1}, a_m) \leq f(a_1, a_2, \dots, z_i, \dots, a_{m-1}, a_m)$$

$$\forall z, a_1, a_2, \dots, z_i, \dots, a_{m-1}, a_m \in R$$

por lo que, en particular

$$f(a_1, a_2, \dots, x_i^0, \dots, a_{m-1}, a_m) \leq f(z_1^0, z_2^0, \dots, z_i^0, \dots, z_{m-1}^0, z_m^0).$$

Ahora bien,  $z^0 = (z_1^0, z_2^0, \dots, z_m^0)$  era también un mínimo, por lo que

$$\exists a_2, a_3, \dots, a_m \quad \text{t.q.} \quad f(a_1, \dots, x_i^0, \dots, a_m) = f(z_1^0, \dots, z_i^0, \dots, z_m^0).$$

Para cada variable,  $z_i$ , se determina un mínimo,  $x_i^0$ , que es función del resto de los parámetros,  $x_i^0 = h(a_1, a_2, \dots, a_{i-1}, a_{i+1}, \dots, a_n)$ , independientemente del orden de selección de las variables.

C.Q.D.

### Fase de estimación del modelo

En la fase de estimación se determinan las coordenadas de los puntos que representan a cada uno de los estímulos sometidos al análisis. Para ello se aplican las técnicas de mínimos cuadrados, consiguiendo la estimación de puntos coordenada a coordenada. Tal y como ya se ha expuesto en la sección anterior, el procedimiento se inicia con la determinación de las coordenadas correspondientes a la configuración unidimensional.

Una vez obtenida esta configuración, y considerando los resultados de las iteraciones anteriores, se procede a estimar las coordenadas de cada punto en cada nueva dimensión.

El primer paso por tanto consiste en calcular la derivada de la función de pérdidas respecto de una coordenada cualquiera. Para ello la descomponemos en una suma de funciones auxiliares más simples:

$$SS(X) = \frac{1}{n-1} \sum_{q=1}^{n-1} \sum_{i=1}^n \sum_{j=1}^{i-1} (\hat{d}_{ijq}^2 - d_{ij}^2(X_q))^2 = \frac{1}{n-1} \sum_{q=1}^{n-1} SS_q(X)$$

siendo

$$SS_q(X) = \frac{1}{n-1} \sum_{i=1}^n \sum_{j=1}^{i-1} (\hat{d}_{ijq}^2 - d_{ij}^2(X_q))^2,$$

la función SSTRESS correspondiente a la configuración en dimensión  $q$ . Desarrollando esta expresión en función del valor de la distancia euclídea, resulta:

$$\begin{aligned} SS_q(X) &= \sum_{i=1}^n \sum_{j=1}^{i-1} \left( \hat{d}_{ijq}^2 - \sum_{a=1}^q (x_{ia} - x_{ja})^2 \right)^2 = \\ &= \sum_{i=1}^n \sum_{j=1}^{i-1} \left( \hat{d}_{ijq}^2 - \sum_{a=1, a \neq s}^q (x_{ia} - x_{ja})^2 - (x_{is} - x_{js})^2 \right)^2 = \\ &= \sum_{i>j} (a_{ijs}^2(X_q) - (x_{is} - x_{js})^2)^2 = \sum_{i>j} \gamma_{ijs}^2(X_q). \end{aligned} \quad (\text{III.4})$$

con

$$a_{ijs}^2(X_q) = \hat{d}_{ijq}^2 - \sum_{a=1; a \neq s}^q (x_{ia} - x_{ja})^2$$

y

$$\gamma_{ijs}(X_q) = a_{ijs}^2(X_q) - (x_{is} - x_{js})^2$$

Derivando parcialmente respecto la coordenada del objeto,  $r$ , correspondiente a la dimensión,  $s$ , se obtiene:

$$\frac{\partial SS(X)}{\partial x_{rs}} = \frac{1}{n-1} \sum_{q=1}^{n-1} \frac{\partial SS_q(X)}{\partial x_{rs}}.$$

Y dado que, cuando calculamos las coordenadas, consideramos todo el conjunto de configuraciones anidadas, se tiene que, cuando  $s > q$ , entonces  $\frac{\partial SS_q(X)}{\partial x_{rs}} = 0$ , dado que en  $SS_q(X)$  se trabaja en dimensión  $q$ , por lo que no tiene sentido considerar derivadas parciales respecto a componentes en dimensiones superiores. Resulta por tanto:

$$\begin{aligned} \frac{\partial SS(X)}{\partial x_{rs}} &= \frac{1}{n-1} \sum_{q=1}^{s-1} \frac{\partial SS_q(X)}{\partial x_{rs}} + \frac{1}{n-1} \sum_{q=s}^{n-1} \frac{\partial SS_q(X)}{\partial x_{rs}} = \\ &= \frac{1}{n-1} \sum_{q=s}^{n-1} \frac{\partial SS_q(X)}{\partial x_{rs}} = \\ &= \frac{1}{n-1} \sum_{q=s}^{n-1} \sum_{i=1}^n \sum_{j=1}^{i-1} \frac{\partial \gamma_{ijs}^2(X_q)}{\partial x_{rs}}. \end{aligned} \tag{III.5}$$

Teniendo en cuenta a su vez que

$$\begin{aligned} \frac{\partial \gamma_{ijs}(X_q)}{\partial x_{rs}} &= 0 \\ \frac{\partial \gamma_{rjs}(X_q)}{\partial x_{rs}} &= -2(x_{rs} - x_{js}) \\ \frac{\partial \gamma_{irs}(X_q)}{\partial x_{rs}} &= 2(x_{is} - x_{rs}) \\ \frac{\partial \gamma_{rrs}(X_q)}{\partial x_{rs}} &= 0 \end{aligned}$$

se obtiene:

$$\frac{\partial SS(X)}{\partial x_{rs}} =$$

$$\begin{aligned}
 &= \frac{1}{n-1} \sum_{q=s}^{n-1} \frac{\partial}{\partial x_{rs}} \left[ \sum_{j < i, i=1}^{r-1} \gamma_{ijs}^2(X_q) + \sum_{j=1}^{r-1} \gamma_{rjs}^2(X_q) + \sum_{i=r+1}^n \sum_{j=1; j \neq r}^{i-1} \gamma_{ijs}^2(X_q) + \sum_{i=r+1}^n \gamma_{ijs}^2(X_q) \right] = \\
 &= \frac{1}{n-1} \sum_{q=s}^{n-1} \left[ \sum_{j=1}^{r-1} 2(x_{js} - x_{rs}) 2\gamma_{rjs}(X_q) + \sum_{i=r+1}^n 2(x_{is} - x_{rs}) 2\gamma_{irs}(X_q) \right] = \\
 &= \frac{4}{n-1} \sum_{q=s}^{n-1} \sum_{j=1}^n \gamma_{rjs}(X_q)(x_{js} - x_{rs})
 \end{aligned}$$

e igualando a 0,

$$\frac{\partial SS(X)}{\partial x_{rs}} = \frac{4}{n-1} \sum_{q=s}^{n-1} \sum_{j=1}^n \gamma_{rjs}(X_q)(x_{js} - x_{rs}) = 0$$

se llega a una ecuación de tercer grado que se resuelve por métodos iterativos.

Calculadas todas las coordenadas para una dimensión, pasamos a la dimensión siguiente manteniendo constantes las coordenadas anteriores. De este modo se obtiene

$$X = \{X_1, \dots, X_{n-1}\} \text{ verificando } X_q = XC_q,$$

con  $C_q$  una matriz diagonal con los  $q$  primeros elementos diagonales iguales a uno y los restantes elementos nulos.

### 2.3 Convergencia del modelo anidado no métrico

Los ingredientes computacionales básicos en nuestro programa son el método de mínimos cuadrados alternos, utilizado en la definición de la función de pérdida, y el concepto de escalamiento óptimo.

La función de pérdida utilizada, por construcción, toma el valor 0 en una elección concreta de los parámetros, los correspondientes al mínimo global. Por otra parte, el algoritmo utilizado es de tipo ALS, lo que quiere decir que en cada iteración se alternan dos etapas de minimización: en la primera, se calculan las transformaciones óptimas para la determinación de las disparidades a partir de las disimilaridades de partida y las distancias entre los puntos estimados y en la segunda, se calculan nuevos valores para las coordenadas de los puntos que representan a los objetos analizados, manteniendo fijos los valores de las disparidades obtenidas en la etapa anterior.

Alternar estas dos etapas, tal y como hemos demostrado en la fase de terminación del algoritmo, produce una sucesión decreciente de valores de la función de pérdida que converge por estar acotada inferiormente por 0. Bajo suaves condiciones de regularidad, se puede probar que los valores a los que converge la función coinciden con algún punto estacionario de la función de pérdida (Takane et al., 1977).

### Descomposición de SSTRESS

En los procedimientos clásicos de MDS se consigue una aproximación a la matriz de disimilaridades a partir de la matriz de productos escalares asociada a estos datos. La descomposición en componentes principales de esta segunda matriz proporciona una configuración en  $p$  dimensiones y además, los valores propios asociados a la descomposición proporcionan una medida de la importancia del eje en la representación.

En los procedimientos de mínimos cuadrados, incluido el modelo anidado, se realiza una aproximación directa entre las disimilaridades observadas,  $\delta_{ij}$ , y las distancias,  $d_{ij}$ , en una dimensión determinada o en un conjunto de dimensiones anidadas como es el caso que nos ocupa. Si bien estos métodos no proporcionan información sobre la influencia que cada dimensión tiene en la configuración estimada, Gower (1984, 1986) determinó una descomposición de la varianza de las funciones de pérdida STRESS y SSTRESS que permitía establecer la proporción de variabilidad explicada por la solución y que pasamos a resumir brevemente en las siguientes líneas.

Si consideramos el valor de SSTRESS para una única dimensión en términos de las disimilaridades:

$$SS(X_q) = \sum_{i=1}^n \sum_{j=1}^{i-1} (\delta_{ij}^2 - d_{ij}^2(X_q))^2,$$

cuyas ecuaciones normales vienen dadas por

$$TX_q = 0,$$

con

$$t_{ij} = \delta_{ij}^2 - d_{ij}^2(X_q), \quad \forall i \neq j$$

$$t_{ii} = - \sum_{j=1, j \neq i}^n t_{ij}.$$

Definiendo la matriz de productos escalares asociada a la matriz de disimilaridades (tras la estimación de la constante aditiva, las disimilaridades se comportan como distancias)

$$B = -\frac{1}{2}J\Delta J$$

con  $J = I - \frac{1}{n}\mathbf{1}\mathbf{1}^t$  y  $\mathbf{1}$ , un vector de unos.

Operando sobre las ecuaciones normales,  $TX_q = 0$ , y teniendo en cuenta que por construcción  $T\mathbf{1} = 0$ , se tiene

$$TJX_q = T\left(I - \frac{1}{n}\mathbf{1}\mathbf{1}^t\right)X_q = 0.$$

De modo que

$$T\left(I - \frac{1}{n}\mathbf{1}\mathbf{1}^t\right)X_qX_q^t\left(I - \frac{1}{n}\mathbf{1}\mathbf{1}^t\right)^t = 0,$$

por lo que,

$$\text{tr}\left[T\left(I - \frac{1}{n}\mathbf{1}\mathbf{1}^t\right)X_qX_q^t\left(I - \frac{1}{n}\mathbf{1}\mathbf{1}^t\right)\right] = \text{tr}\left[TJX_qX_q^tJ\right] = 0.$$

Y aplicando las propiedades de la traza

$$\text{tr}\left[TJX_qX_q^tJ\right] = \text{tr}\left[JTJX_qX_q^t\right] = \text{tr}\left[TX_qX_q^t\right] = 0,$$

que escrito en términos de sumatorias

$$\sum_{i=1}^n \sum_{j=1}^{i-1} (\delta_{ij}^2 - d_{ij}(q)^2(X_q))^2 = \sum_{i=1}^n \sum_{j=1}^{i-1} \delta_{ij}^4 - \sum_{i=1}^n \sum_{j=1}^{i-1} d_{ij}^4(X_q),$$

y despejando apropiadamente

$$\sum_{i=1}^n \sum_{j=1}^{i-1} \delta_{ij}^4 = \sum_{i=1}^n \sum_{j=1}^{i-1} (\delta_{ij}^2 - d_{ij}^2(X_q))^2 + \sum_{i=1}^n \sum_{j=1}^{i-1} d_{ij}^4(X_q), \quad \forall q = 1, 2, \dots, n-1$$

se obtiene la descomposición citada, a partir de la cual puede llevarse a cabo un análisis de la variabilidad explicada por la solución: la suma de los cuadrados de las disimilaridades iniciales al cuadrado, se descompone como la suma de los cuadrados de las distancias al

cuadrado entre los puntos de la configuración ajustada más la suma de los cuadrados de los residuos.

Aplicando esta descomposición al modelo anidado, sumando los resultados obtenidos y ponderando por el número de configuraciones, resulta

$$SS(X) = \frac{1}{n-1} \sum_{q=1}^{n-1} \sum_{i=1}^n \sum_{j=1}^{i-1} (\delta_{ij}^2 - d_{ij}^2(X_q))^2 = \frac{1}{n-1} \sum_{q=1}^{n-1} \sum_{i=1}^n \sum_{j=1}^{i-1} \delta_{ij}^4 - \frac{1}{n-1} \sum_{q=1}^{n-1} \sum_{i=1}^n \sum_{j=1}^{i-1} d_{ij}^4(X_q),$$

$$SS(X) = \sum_{i=1}^n \sum_{j=1}^{i-1} \delta_{ij}^4 - \frac{1}{n-1} \sum_{q=1}^{n-1} \sum_{i=1}^n \sum_{j=1}^{i-1} d_{ij}^4(X_q),$$

de donde

$$\sum_{i < j}^n \delta_{ij}^4 = \frac{1}{n-1} \sum_{q=1}^{n-1} \sum_{i < j}^n d_{ij}^4(X_q) + SS(X). \quad (\text{III.6})$$

Considerando la expresión del SSTRESS normalizado, la descomposición para una única configuración resultaría,

$$1 = \frac{\sum_{i=1}^n \sum_{j=1}^{i-1} (\delta_{ij}^2 - d_{ij}^2(X_q))^2}{\sum_{i=1}^n \sum_{j=1}^{i-1} \delta_{ij}^4} + \frac{\sum_{i=1}^n \sum_{j=1}^{i-1} d_{ij}^4(X_q)}{\sum_{i=1}^n \sum_{j=1}^{i-1} \delta_{ij}^4}, \quad \forall q = 1, 2, \dots, n-1.$$

De donde, la expresión para el modelo anidado:

$$1 = \frac{1}{n-1} \sum_{q=1}^{n-1} \left[ \frac{\sum_{i < j}^n d_{ij}^4(X_q)}{\sum_{i < j}^n \delta_{ij}^4} \right] + SS(X).$$

Dada la estructura del modelo utilizado en el algoritmo anidado, es necesario definir una medida de la aportación que cada dimensión proporciona a la solución global dada por el algoritmo. Dicha medida será la norma de los vectores columna de la matriz de configuración:

$$\lambda_i^2 = \|x_i\|^2 = x_i^t x_i \quad i = 1, 2, \dots, n$$

a partir de donde se reescribe la matriz de coordenadas con las columnas normalizadas

$$x_i \rightarrow \hat{x}_i = \frac{x_i}{\|x_i\|} = \frac{x_i}{\lambda_i}, \quad i = 1, 2, \dots, n$$

obteniendo una nueva configuración anidada con vectores de longitud 1 junto con los valores,  $\lambda_i$ ,  $i = 1, 2, \dots, n$ , que dan la medida de la importancia que cada dimensión tiene en la representación.

Para expresar la descomposición (III.6) en términos matriciales, se define  $Y_q$ ,  $q = 1, 2, \dots, n - 1$ , de dimensión  $\frac{n(n-1)}{2} \times q$  formada, por columnas, por las diferencias entre las coordenadas correspondientes a cada dimensión,  $(x_{ia} - x_{ja})$ . Con lo que la suma de la potencia cuarta de las distancias euclídeas vendrá dada por

$$\sum_{i < j} d_{ij}^4(X_q) = \text{tr} [Y_q Y_q^t \text{diag}(Y_q Y_q^t)],$$

a partir de donde se obtiene la descomposición de SSTRESS:

$$\|\Delta\|^2 = \frac{1}{n-1} \sum_{q=1}^{n-1} \text{tr} [\bar{Y}_q \bar{Y}_q^t \text{diag}(\bar{Y}_q \bar{Y}_q^t)] \lambda_q^2 + SS(X),$$

$$\text{con } \bar{Y}_q = \frac{1}{\sqrt{\lambda_q}} Y_q.$$



## Capítulo IV

# Análisis de los resultados obtenidos en el modelo anidado para MDS no métrico

Procederemos a continuación a hacer un análisis práctico del modelo expuesto en el capítulo anterior. Para ilustrar el método se ha procedido al análisis de las distancias entre los vértices de un prisma recto octogonal, perturbadas mediante valores generados aleatoriamente a partir de dos distribuciones uniformes de hasta el 25% de las distancias originales. Se consideraron dos matrices de disimilaridades dado el elevado número de estimaciones que han de hacerse en el modelo. En las figuras IV.1 y IV.2 se presentan los gráficos correspondientes a las representaciones tridimensionales y bidimensionales del prisma real a partir del cual se han generado las disimilaridades junto con los resultados obtenidos tras el análisis de las disimilaridades con el modelo anidado.

Como configuración inicial se tomó la descomposición en valores propios de la matriz de disimilaridades formada por los valores medios de las dos matrices de datos de las que partimos, aplicando el procedimiento de MDS clásico que sabemos proporciona configuraciones anidadas.

El modelo expuesto en el capítulo anterior ha sido implementado en Microsoft Visual Basic, para realizar análisis de datos no métricos con réplicas de la matriz de disimilaridades y a partir de una matriz de configuración inicial previamente calculada por algún otro procedimiento. El análisis que se va a presentar a lo largo de este capítulo puede resumirse en tres etapas:

1. Estudio de las iteraciones y resultados de SSTRESS del modelo anidado, implementado en Visual Basic, y comprobación de la validez de los resultados obtenidos.

Figura IV.1: Representaciones tridimensionales del prisma octogonal real y la representación obtenida mediante el modelo anidado no métrico

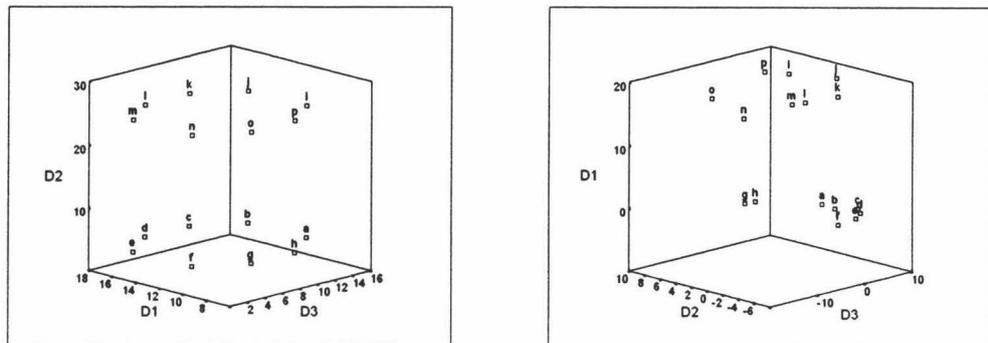
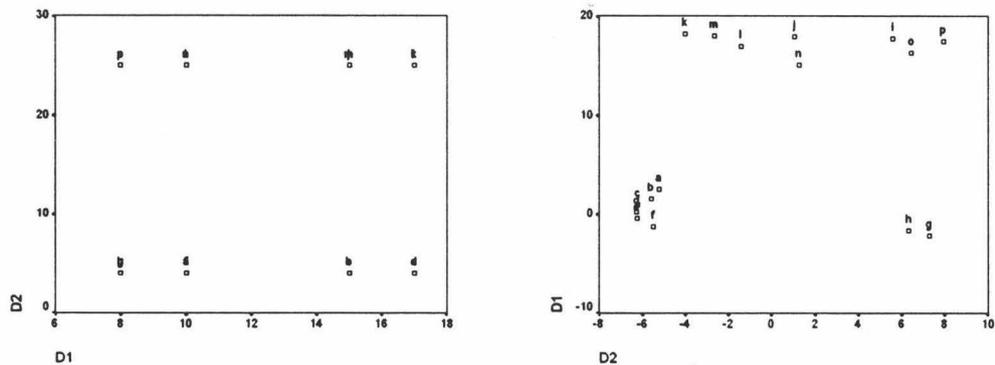


Figura IV.2: Representaciones bidimensionales del prisma octogonal real y la representación obtenida mediante el modelo anidado no métrico



2. Contraste, mediante procedimientos Procrustes, de las soluciones anidadas proporcionadas por el modelo.
3. Análisis, mediante técnicas jackknife, de la robustez del modelo.

## 1 Análisis de los valores de SSTRESS

Introducidos los datos de la configuración inicial, el modelo anidado se detuvo, por no producirse mejoras en la función de pérdida, tras cinco iteraciones del proceso en las que se obtuvieron los valores de SSTRESS, recogidos en en la tabla IV.1, en la que también aparecen los valores de SSTRESS en cada dimensión.

Tabla IV.1: Resultados de las iteraciones del modelo anidado para MDS aplicado a las distancias entre los vértices del prisma octogonal perturbadas aleatoriamente mediante una distribución uniforme.

Dimensión	Iteración 1	Iteración 2	Iteración 3	Iteración 4	Iteración 5
1	0.53417548	0.4925273	0.4435906	0.4424229	0.4424229
2	0.45429975	0.3864509	0.3721599	0.3642555	0.3642555
3	0.47231827	0.3803263	0.2900237	0.2349138	0.2349138
4	0.47231827	0.3772338	0.2361008	0.1991127	0.1991127
5	0.47231827	0.2766900	0.2259733	0.1890265	0.1890265
6	0.47231827	0.2664758	0.2155534	0.1790373	0.1790373
7	0.47231827	0.2643473	0.2157866	0.1751975	0.1751975
8	0.47231827	0.2541766	0.2140329	0.1729739	0.1729739
9	0.47231827	0.2539152	0.2028727	0.1726374	0.1726374
10	0.47231827	0.2539152	0.2028727	0.1726374	0.1726374
11	0.47231827	0.2539152	0.2028727	0.1726374	0.1726374
12	0.47231827	0.2539152	0.2028727	0.1726374	0.1726374
13	0.47231827	0.2539152	0.2028727	0.1726374	0.1726374
14	0.47231827	0.2539152	0.2028727	0.1726374	0.1726374
15	0.47231827	0.2539152	0.2028727	0.1726374	0.1726374
SSTRESS anidado	0.47524085	0.2983784	0.2418887	0.2116939	0.2116939

Del análisis de los valores de la tabla se observa que, en la primera iteración, los valores de SSTRESS permanecen estables en todas las dimensiones en torno a 0.47; no se produce una mejora significativa con respecto a la configuración inicial. A partir de la segunda iteración se observa un descenso de 0.1 unidades en el valor de SSTRESS alcanzado en las cuatro primeras dimensiones; en el resto de iteraciones del proceso, las diferencias entre las tres primeras dimensiones son evidentes. Claramente, se mantienen los resultados generales de cualquier análisis con MDS, conforme aumenta la dimensión de representación de la configuración disminuye el valor de SSTRESS. En este caso, los valores se estabilizan a partir de una determinada dimensión debido a que en el proceso de anidamiento, forzamos la adición de dimensiones que, en algunos casos, como es el que nos ocupa, no aportan nuevas informaciones al análisis.

Se observan unos elevados valores de SSTRESS medio, resultado totalmente predecible ante las fuertes restricciones a las que se ve sometido el modelo. A partir de la configuración inicial hemos fijado las dimensiones de representación, dadas por la descomposición en valores singulares de las matrices de disimilaridades, por lo que, en este sentido, el modelo aquí definido podría considerarse equivalente a los modelos de diferencias individuales con los que habitualmente se trabaja y en los que las dimensiones de representación no pueden ser rotadas.

En este primer análisis se procedió también al cálculo de los módulos dimensionales y de los coeficientes de correlación entre las distancias de la configuración verdadera y las distancias entre los vértices del prisma en cada una de las configuraciones anidadas. Los resultados se resumen en la tabla IV.2. Los elevados valores de los coeficientes de correlación confirman la adecuación del modelo para el ajuste de los datos. Se observa cómo estos valores se van incrementando, hasta llegar a la configuración en cuatro dimensiones; a partir de ahí se produce una oscilación de los valores hasta que se estabilizan. El motivo de esta oscilación es la introducción de factores de ruido al forzar la adición de las nuevas dimensiones, predeterminadas a partir de la configuración inicial. La determinación de la dimensionalidad vendría dada a partir del análisis de la segunda columna de la tabla IV.2, en donde se observa que la mayor parte de la variabilidad viene explicada por las tres primeras dimensiones, resultado éste conforme con los valores dimensionales de SSTRESS. En la última columna de la tabla IV.1, tal y como ya hemos señalado anteriormente, se observa que a partir de la cuarta dimensión los valores dimensionales de SSTRESS también se estabilizan, es decir, las nuevas dimensiones no aportan nada nuevo al ajuste del modelo.

Tabla IV.2: Análisis de los módulos dimensionales de las configuraciones anidadas y los valores de los coeficientes de correlación entre las distancias verdaderas y las estimadas

Dimensión	Módulos dimensionales	Correlación distancias reales-estimadas
1	2399.927	0.606
2	462.9604	0.830
3	406.8931	0.841
4	123.3636	0.841
5	115.0268	0.843
6	46.2337	0.838
7	42.2378	0.789
8	26.9125	0.835
9	10.6472	0.824
10	0.0000	0.824
11	0.0000	0.824
12	0.0000	0.824
13	0.0000	0.824
14	0.0000	0.824
15	0.0000	0.824

## 2 Contraste de las soluciones anidadas. Análisis Procrustiano General

El siguiente paso del análisis es el estudio conjunto de las configuraciones estimadas. Para ello haremos uso del modelo Matchals de Commandeur que ya comentamos extensamente en el segundo capítulo de esta memoria.

En nuestro estudio hemos procedido a realizar dos análisis diferentes:

1. Análisis procrustiano de las quince configuraciones obtenidas mediante el algoritmo anidado.
2. Análisis procrustiano entre cada una de las configuraciones anidadas y la configuración real del prisma.

A continuación exponemos los resultados obtenidos en ambos análisis.

## 2.1 Análisis procrustiano de las configuraciones anidadas

Entre las ventajas del modelo anidado está el hecho de que permite el análisis conjunto de las soluciones obtenidas. Las configuraciones derivadas del método están dadas sobre los mismos ejes de representación, por lo que es posible contrastar qué información aporta cada dimensión.

Para ello se procede en esta sección al análisis conjunto de las quince configuraciones (si bien las últimas cinco configuraciones coinciden como quedará reflejado en los resultados). Tras someter las quince configuraciones anidadas al análisis procrustiano, se obtiene que el 95.26% de la variabilidad de los datos es explicada por el modelo. A continuación estudiaremos la descomposición de la variabilidad dada en términos de las configuraciones y de las dimensiones de representación. Los resultados se resumen en las tablas IV.3 y IV.4.

En la tabla IV.3 se observa, análogamente a los resultados observados en la tabla anterior, que la primera dimensión representa un 98.826% de variabilidad explicada, el segundo eje de representación, explica un 97.58% y el tercero un 87.28%, la proporción de variabilidad explicada por los ejes 4<sub>0</sub>, 5<sub>0</sub> y 6<sub>0</sub> oscila en torno al 70%, la de los tres siguientes, en torno al 50% y el resto de los ejes no aporta nada nuevo al análisis conjunto.

En lo referente al análisis de la variabilidad en términos de la contribución relativa de cada configuración en la tabla IV.4 se observa como la mayor variabilidad explicada se debe a las tres primeras configuraciones (con un 99.2% de variabilidad explicada), en las siete siguientes configuraciones, la proporción de variabilidad explicada se estabiliza entre el 98% y el 99% y, en el resto de las configuraciones se va produciendo un ligero descenso en estos porcentajes hasta llegar al 63.65%, resultados también conformes con el hecho de que toda la información referente a la constitución de la configuración real reside en las tres primeras configuraciones anidadas.

## 2.2 Análisis procrustiano entre la configuración real y las diferentes configuraciones anidadas

En esta etapa del análisis se procedió al contraste directo entre las configuraciones anidadas y la configuración real del prisma. Los resultados se resumen a continuación.

En la tabla IV.5 se resumen las medidas de bondad de ajuste para cada uno de los análisis. Se observa el elevado valor de tales medidas, todas en torno a 0.9, indicativas del buen ajuste entre las configuraciones anidadas y la configuración real. Los mayores

Tabla IV.3: Contribución relativa de cada dimensión en el estudio conjunto de las quince configuraciones anidadas

Dimensión	Variabilidad explicada por el ajuste	Variabilidad residual	Variabilidad total
1	9.1770	0.0316	9.2086
2	2.4640	0.2213	2.6853
3	2.1432	0.2998	2.4430
4	0.3837	0.1015	0.4852
5	0.0828	0.0351	0.1179
6	0.0253	0.0119	0.0372
7	0.0120	0.0093	0.0213
8	0.0007	0.0006	0.0013
9	0.0002	0.0002	0.0003
10	0.0000	0.0000	0.0000
11	0.0000	0.0000	0.0000
12	0.0000	0.0000	0.0000
13	0.0000	0.0000	0.0000
14	0.0000	0.0000	0.0000
15	0.0000	0.0000	0.0000
Total	14.2887	0.7113	15.0000

Tabla IV.4: Contribución relativa de cada configuración en el estudio conjunto de las quince configuraciones anidadas

Configuración	Variabilidad explicada por el ajuste	Variabilidad residual	Variabilidad total
1	0.9921	0.0098	1.0020
2	0.9921	0.0098	1.0020
3	0.9921	0.0098	1.0020
4	0.9921	0.0098	1.0020
5	0.9921	0.0098	1.0020
6	0.9921	0.0098	1.0020
7	0.9921	0.0098	1.0020
8	0.9935	0.0085	1.0020
9	0.9941	0.0080	1.0021
10	0.9944	0.0077	1.0021
11	0.9911	0.0109	1.0019
12	0.9843	0.0173	1.0016
13	0.9551	0.0450	1.0001
14	0.8053	0.1874	0.9927
15	0.6262	0.3576	0.9838
Total	14.2887	0.7113	15.0000

Tabla IV.5: Medidas de bondad de ajuste entre la configuración real y las configuraciones anidadas

Configuración anidada	Medida de bondad de ajuste
1	0.9944
2	0.9749
3	0.9398
4	0.9371
5	0.9353
6	0.9619
7	0.9312
8	0.9299
9	0.9289
10	0.9289
11	0.9289
12	0.9289
13	0.9289
14	0.9289
15	0.9289

valores de la correlación se observan en las tres primeras configuraciones, tal y como era de esperar. También cabe destacar la estabilización de los valores en el momento en que las nuevas dimensiones no aportan ninguna nueva información, tal y como ocurría en los resultados obtenidos en la sección anterior.

En la tabla IV.6 se resumen los resultados correspondientes a la descomposición de la variabilidad de los datos en los quince contrastes con referencia a la información aportada por cada dimensión. Nuevamente, estos resultados muestran la diferencia patente entre la proporción de variabilidad explicada por las tres primeras dimensiones, en torno siempre al 80%, y las restantes, con valores cercanos al 50%.

No se incluyen en la memoria las descomposiciones de la variabilidad correspondientes a las configuraciones y los datos dado que no aportan ningún resultado significativo.

En las descomposiciones de la variabilidad obtenidas, se observa que las configuraciones siempre juegan el mismo papel en términos de relevancia en el análisis. Es de

especial interés fijarnos en los resultados de la descomposición de la variabilidad respecto a cada una de las dimensiones.

### 3 Análisis de la estabilidad del modelo

Para completar el análisis del modelo de soluciones anidadas, presentamos en esta sección un estudio sobre la estabilidad en las soluciones proporcionadas por el método.

Hay muchos tipos de estabilidad y, por lo tanto, hay muchas formas de análisis de estabilidad o sensibilidad. Todas ellas tienen en común que estudian el efecto de una pequeña perturbación de los datos en la solución. Un método es estable en una aplicación particular, si pequeños cambios en los datos producen sólo pequeños cambios en la solución.

En la mayoría de las aplicaciones de MDS, no se estudia la estabilidad de la solución, aunque siempre hay excepciones, como ocurre en el método de máxima verosimilitud (ML) aplicado a MDS (Ramsay, 1982), en el que la teoría estándar para la estimación máximo verosímil con grandes muestras puede ser aplicada para obtener información acerca de la estabilidad estadística. En cualquier caso, es cierto que las propiedades de optimalidad de la estimación ML, a menudo no se aplican directamente a las situaciones MDS debido a aspectos no estándar de tales situaciones, que están llenos de parámetros de incidencia, dependencias, pérdida por replicaciones, etc.

Si hay réplicas independientes en un análisis MDS, se puede obtener información estadística incluso sin suponer modelos de distribución paramétricos o funciones de respuesta específicas, bastaría con usar técnicas de remuestreo como el *bootstrap* (Efron, 1982; Heiser y Meulman, 1983a, 1983b; Weinberg, Carroll y Cohen, 1984).

Si no hay réplicas, el problema puede ser un poco más complicado. Se necesita un modelo paramétrico para obtener información estadística sobre la estabilidad, pero muy a menudo, el conocimiento a priori requerido para hacer suposiciones necesarias con alguna confianza no está disponible. Hay también otras situaciones en las que la idea del muestreo aleatorio no tiene mucho sentido y en las que estudiar la estabilidad en el modo estadístico usual, no es consecuentemente interesante, de hecho, ni siquiera está definido. No obstante, todavía es posible, en tales casos, realizar otros tipos de análisis de estabilidad. La idea de un pequeño cambio en los datos se puede definir de distintos modos y los efectos de tales pequeñas perturbaciones se pueden estudiar. Esta posibilidad ya fue señalada en el contexto de MDS, por Kruskal y Wish (1978).

Tabla IV.6: Análisis de la variabilidad en términos de las dimensiones

Configuración anidada	Dimensiones	Variabilidad explicada por el ajuste	Variabilidad residual	Variabilidad total
1	1	1.9888	0.0112	2.0000
2	1	1.6537	0.0093	1.6631
	2	0.2961	0.0408	0.3369
3	1	1.3586	0.0071	1.3657
	2	0.2830	0.0900	0.3731
	3	0.2380	0.0232	0.2612
4	1	1.3282	0.0080	1.3362
	2	0.2899	0.0720	0.3619
	3	0.2377	0.0293	0.2670
	4	0.0184	0.0166	0.0350
5	1	1.3320	0.0090	1.3409
	2	0.2824	0.0702	0.3526
	3	0.2328	0.0289	0.2618
	4	0.0208	0.0190	0.0398
	5	0.0026	0.0024	0.0050
6	1	1.1219	0.0254	1.1473
	2	0.6497	0.0231	0.6728
	3	0.0885	0.0194	0.1079
	4	0.0550	0.0041	0.0592
	5	0.0045	0.0037	0.0082
	6	0.0042	0.0005	0.0047
7	1	1.3204	0.0099	1.3303
	2	0.2804	0.0692	0.3495
	3	0.2303	0.0293	0.2596
	4	0.0230	0.0214	0.0443
	5	0.0067	0.0063	0.0130
	6	0.0015	0.0015	0.0030
	7	0.0001	0.0001	0.0002

Configuración anidada	Dimensiones	Variabilidad explicada por el ajuste	Variabilidad residual	Variabilidad total
8	1	1.3135	0.0103	1.3238
	2	0.2838	0.0684	0.3522
	3	0.2299	0.0309	0.2608
	4	0.0231	0.0215	0.0446
	5	0.0066	0.0063	0.0129
	6	0.0020	0.0020	0.0040
	7	0.0009	0.0008	0.0017
	8	0.0000	0.0000	0.0001
9	1	1.3117	0.0107	1.3224
	2	0.2834	0.0675	0.3508
	3	0.2277	0.0311	0.2588
	4	0.0250	0.0232	0.0482
	5	0.0065	0.0062	0.0127
	6	0.0020	0.0019	0.0039
	7	0.0015	0.0014	0.0029
	8	0.0000	0.0000	0.0002
	9	0.0000	0.0000	0.0000
10	1	1.3117	0.0107	1.3224
	2	0.2834	0.0675	0.3508
	3	0.2277	0.0311	0.2588
	4	0.0250	0.0232	0.0482
	5	0.0065	0.0062	0.0127
	6	0.0020	0.0019	0.0039
	7	0.0015	0.0014	0.0029
	8	0.0000	0.0000	0.0002
	9	0.0000	0.0000	0.0000
	10	0.0000	0.0000	0.0000
11	1	1.3117	0.0107	1.3224
	2	0.2834	0.0675	0.3508
	3	0.2277	0.0311	0.2588
	4	0.0250	0.0232	0.0482
	5	0.0065	0.0062	0.0127
	6	0.0020	0.0019	0.0039
	7	0.0015	0.0014	0.0029
	8	0.0000	0.0000	0.0002
	9	0.0000	0.0000	0.0000
	10	0.0000	0.0000	0.0000
	11	0.0000	0.0000	0.0000

Una primera sugerencia y un modo bastante natural de análisis, es hacer un pequeño cambio en una disimilaridad de un problema de MDS métrico. Utilizando el término *Jacknife* en un sentido muy amplio, para una clase de técnicas que estudia la estabilidad borrando parte de los datos o varias partes de los datos en modos sistemáticos, se puede considerar una versión discreta de estos procedimientos *Jacknife* (Gray y Schucany, 1972 y Miller, 1974). El modelo propuesto por estos autores consiste en eliminar una disimilaridad y tratarla como faltante, para a continuación reanalizar los datos con la disimilaridad omitida, y combinar las  $n(n-1)/2$  soluciones diferentes obtenidas sucesivamente. Este mismo esquema puede aplicarse a MDS no métrico y puede resultar bastante útil. Sin embargo, cuando se analiza un gran número de objetos, habría que realizar muchos análisis MDS y las perturbaciones por omitir una disimilaridad podrían resultar insignificantes, con lo que los resultados no serían significativos.

En nuestro estudio consideraremos otra versión analítica de *Jacknife* para investigar el efecto de mayores perturbaciones en los datos, estudiada por De Leeuw y Meulman en 1986. Los resultados que se obtendrán carecerán de aplicaciones inferenciales, siendo válidos tan solo desde el punto de vista del análisis de datos. El punto de inicio es preguntarse qué ocurre en las soluciones MDS si se elimina toda la información correspondiente a uno de los puntos que se consideran en el análisis. En lugar de analizar las disimilaridades entre  $n$  estímulos una sola vez, se analizan  $n$  veces las disimilaridades entre  $n-1$  estímulos, eliminando uno cada vez. La cuestión de estabilidad asociada con este esquema es interesante por sí misma, y los resultados de los  $n+1$  análisis pueden reunirse muy bien en un único gráfico. Claramente, esta técnica es aplicable tanto para MDS métrico como para no métrico. Este método también se ha utilizado para determinar la dimensionalidad correcta de la representación.

De Leeuw y Meulman diseñaron este procedimiento de análisis para estudiar la estabilidad de los distintos métodos de MDS, haciendo un contraste entre las técnicas en términos de las medidas de bondad de ajuste empleadas para cada procedimiento. Los procedimientos contrastados fueron los siguientes:

- Procedimiento clásico de MDS de Torgerson y Gower, minimizando los cuadrados de las diferencias entre los productos escalares.
- MDS métrico de Kruskal, tomando como función de pérdida la función STRESS.
- MDS no métrico de Shepard-Kruskal, considerando también como función de pérdida la función STRESS.

- MDS no métrico de Takane, Young y de Leeuw, minimizando la suma de los cuadrados de las diferencias entre las distancias al cuadrado, la función SSTRESS.
- MDS máximo-verosímil de Ramsay, en el que se considera como función de pérdida la suma de los cuadrados sobre los logaritmos de las distancias.

Para minimizar la función STRESS utilizaron SMACOF, ALSCAL lo emplearon en la minimización de la función SSTRESS y por último, para el procedimiento de MDS máximo-verosímil consideraron MULTISCALE, tomando en todos los casos como configuración inicial la proporcionada por los procedimientos clásicos de MDS. Tras su análisis, estos autores concluyeron que las soluciones más robustas frente a alteraciones en la matriz de disimilaridades eran las proporcionadas por ALSCAL, mientras que las que más sensibles frente estas alteraciones fueron las proporcionadas por MULTISCALE.

### 3.1 Análisis de robustez del modelo para MDS no métrico anidado

El procedimiento que hemos seguido en nuestro estudio fue descrito por estos autores en los siguientes pasos:

1. Se determina la solución dada por el modelo considerando toda la información de que se dispone,  $X_0$ .
2. Se procede al cálculo de las  $n$  configuraciones,  $X_i$ ,  $i = 1, \dots, n$ ; obtenidas al eliminar de la matriz de disimilaridades inicial la información referente a cada punto,  $x_i$ ,  $i = 1, \dots, n$ .
3. Se comparan las  $n$  configuraciones obtenidas en la etapa anterior mediante las técnicas MATCHALS de Commandeur, obteniéndose nuevas configuraciones,  $Y_i$ ,  $i = 1, \dots, n$ ; transformaciones ortogonales de las anteriores  $X_i$ .
4. Se determina la configuración centroide  $Y_0$ , tras el proceso de comparación de las  $n$  configuraciones,  $X_i$ .

De este modo, tras el proceso se dispone de  $n + 2$  soluciones para cada una de las dimensiones de análisis, entre las que se procede a realizar un estudio comparativo en lo referente a la dispersión de los resultados.

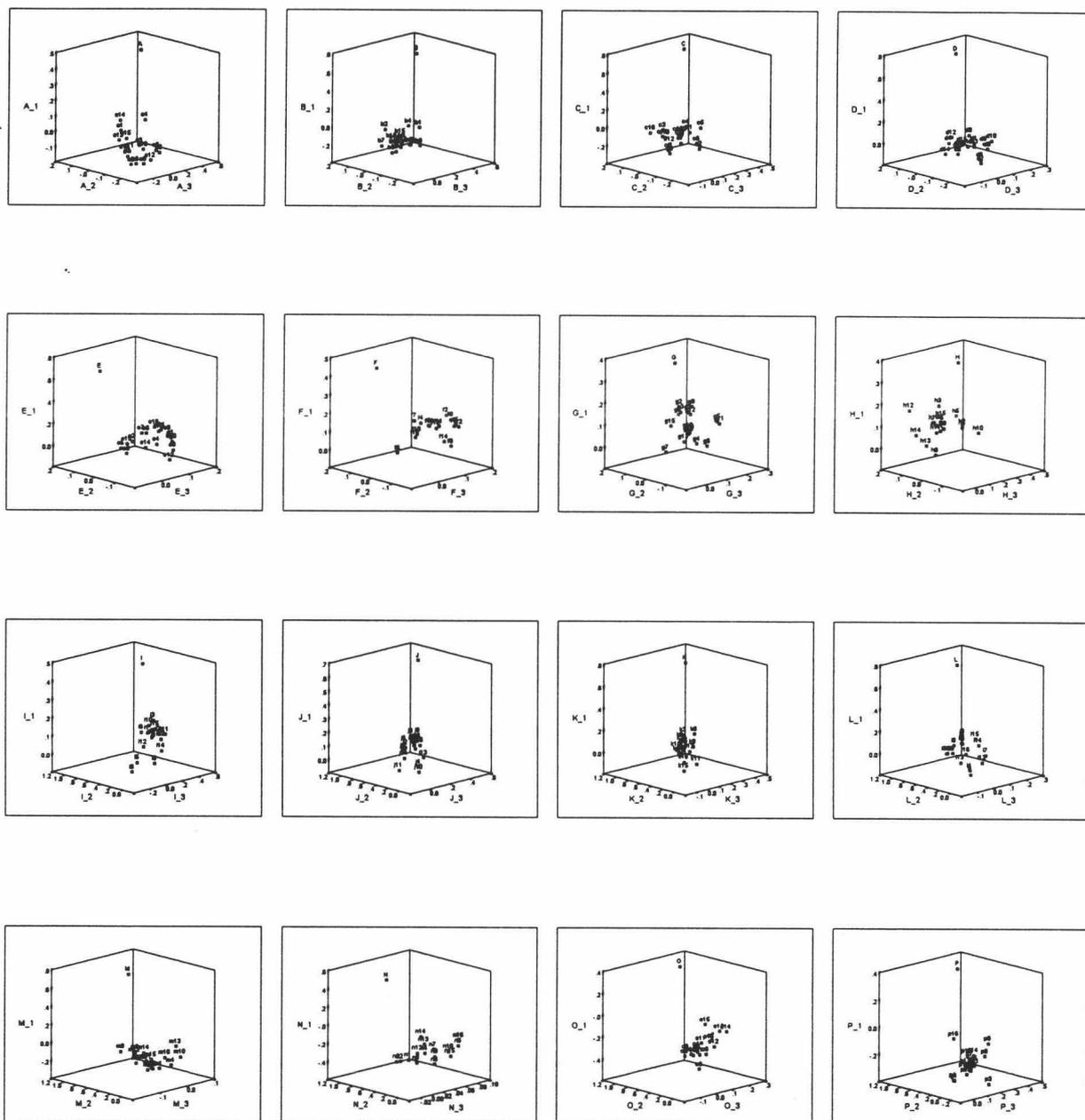
La primera sentencia de estabilidad se obtiene comparando las  $Y_i$ . Este paso puede hacerse gráficamente, dibujando todas las  $Y_i$  como vértices de estrellas centradas en  $Y_0$ .

Junto a las comparaciones gráficas, se puede determinar numéricamente la dispersión de las configuraciones estimadas respecto de la configuración verdadera; para ello basta considerar la siguiente descomposición:

$$\begin{aligned} \sum_{i=1}^n \| Y_i - X_0 \| &= \sum_{i=1}^n \| Y_i - Y_0 + Y_0 - X_0 \| = \\ &= \sum_{i=1}^n \| Y_i - Y_0 \| + \sum_{i=1}^n \| Y_0 - X_0 \| + \sum_{i=1}^n \| Y_i - Y_0 \| \| Y_0 - X_0 \|, \end{aligned}$$

y dado que el último sumando, aplicando la definición de la norma euclídea de matrices, se hace 0, se obtiene una descomposición de la variabilidad total de las configuraciones estimadas respecto de la solución original en términos de la variabilidad del modelo de análisis escogido y del error residual:

Figura IV.3: Gráficos de dispersión



$$\sum_{i=1}^n \| Y_i - X_0 \| = \sum_{i=1}^n \| Y_i - Y_0 \| + n \| Y_0 - X_0 \| .$$

Normalizando resultados,

$$1 = \frac{\sum_{i=1}^n \| Y_i - Y_0 \|}{\sum_{i=1}^n \| Y_i - X_0 \|} + \frac{n \| Y_0 - X_0 \|}{\sum_{i=1}^n \| Y_i - X_0 \|} ,$$

es posible definir una medida, con valores en el intervalo  $[0, 1]$ , indicativa del porcentaje de variabilidad explicada por las nuevas configuraciones:

$$\rho^2 = \frac{\sum_{i=1}^n \| Y_i - Y_0 \|}{\sum_{i=1}^n \| Y_i - X_0 \|} = 1 - \frac{n \| Y_0 - X_0 \|}{\sum_{i=1}^n \| Y_i - X_0 \|} .$$

Los resultados obtenidos en nuestro análisis se resumen en la tabla IV.7.

De los resultados de la tabla se deduce la fuerte estabilidad en cuanto a las soluciones ofrecidas por el procedimiento descrito. Es de destacar la estabilización de los valores del coeficiente  $\rho^2$  a partir de la novena dimensión, así como el hecho de que la solución más estable es la configuración obtenida en tres dimensiones.

Tabla IV.7: Medidas de dispersión en el análisis de robustez del modelo anidado

Dimensión de la configuración	$\rho^2$
1	0.999953
2	0.9999267
3	0.9999275
4	0.9999254
5	0.9999257
6	0.9999234
7	0.9999197
8	0.9999191
9	0.9999169
10	0.9999169
11	0.9999169
12	0.9999169
13	0.9999169
14	0.9999169
15	0.9999169

# APENDICE A

## Programas

### 1 Programa para la determinación de configuraciones anidadas mediante MDS no métrico

VERSION 4.00

Begin VB.Form ALNEST5

```
BackColor      = &H00C0E0FF&
Caption        = "MDS anidado no metrico"
ClientHeight   = 4410
ClientLeft     = 1770
ClientTop      = 1515
ClientWidth    = 6570
Height         = 4815
Left           = 1710
LinkTopic      = "Form1"
ScaleHeight    = 4410
ScaleWidth     = 6570
Top            = 1170
Width          = 6690
```

Begin VB.TextBox ss\_mejora

```
Height        = 375
Left          = 3000
TabIndex      = 12
Top           = 3720
```

```
        Width          = 2175
End

Begin VB.TextBox ss_fin
    Height          = 375
    Left            = 3000
    TabIndex       = 10
    Top             = 3120
    Width          = 2175
End

Begin VB.CommandButton fin
    Caption        = "Fin"
    Height        = 495
    Left          = 5280
    TabIndex     = 8
    Top          = 3720
    Width        = 855
End

Begin VB.TextBox ss_ini
    Height          = 375
    Left            = 2880
    TabIndex       = 7
    Top             = 1920
    Visible        = 0 'False
    Width          = 2295
End

Begin VB.CommandButton iteracion
    Caption        = "Comenzar iteraciones"
    Height        = 375
    Left          = 1800
    TabIndex     = 5
    Top          = 2520
    Visible      = 0 'False
    Width        = 2415
```

End

Begin VB.CommandButton inicio

BackColor = &H00C0FFFF&  
Caption = "Calculo del SSTRESS y las disparidades iniciales"  
Height = 375  
Left = 1080  
TabIndex = 4  
Top = 1320  
Width = 3975

End

Begin VB.TextBox Text2

Height = 375  
Left = 1920  
TabIndex = 3  
Top = 720  
Width = 3975

End

Begin VB.TextBox Text1

Height = 375  
Left = 1920  
TabIndex = 2  
Top = 120  
Width = 3975

End

Begin VB.Label Label5

BackColor = &H00C0E0FF&  
Caption = "Mejora del SSTRESS:"  
Height = 375  
Left = 960  
TabIndex = 11  
Top = 3720  
Width = 1815

End

```
Begin VB.Label Label4
    BackColor      = &H00COE0FF&
    Caption        = "SSTRESS final:"
    Height         = 375
    Left           = 960
    TabIndex       = 9
    Top            = 3120
    Width          = 1695
End

Begin VB.Label Label3
    BackColor      = &H00COE0FF&
    Caption        = "SSTRESS inicial:"
    Height         = 375
    Left           = 960
    TabIndex       = 6
    Top            = 1920
    Visible        = 0 'False
    Width          = 1455
End

Begin MSComDlg.CommonDialog CommonDialog1
    Left           = 5400
    Top            = 1320
    _Version       = 65536
    _ExtentX       = 847
    _ExtentY       = 847
    _StockProps    = 0
End

Begin VB.Label Label2
    BackColor      = &H00COE0FF&
    Caption        = "Configuracion inicial:"
    Height         = 375
    Left           = 240
    TabIndex       = 1
```

```
        Top           = 720
        Width         = 1575
    End

    Begin VB.Label Label1
        BackColor      = &HOOCOE0FF&
        Caption        = "Disimilaridades:"
        Height         = 375
        Left           = 240
        TabIndex       = 0
        Top            = 120
        Width          = 1575
    End
End

Attribute VB_Name = "ALNEST"
Attribute VB_Creatable = False
Attribute VB_Exposed = False

Option Base 1

'Definicion de constantes
Dim objetos As Integer
Dim nt As Integer
Dim SS As Double

'Matrices de datos iniciales
Dim A() As Double
Dim A_ant() As Double
Dim disim1() As Single
Dim disim2() As Single
Dim disim() As Single

'Matrices para el optimal scaling
'orden de las disimilaridades
Dim x() As Integer
```

```
'matriz de distancias  
Dim dist() As Single
```

```
'matriz de disparidades  
Dim dispar() As Double
```

```
'matriz de distancias en la estimacion  
'de los coeficientes de la ecuacion  
Dim d() As Double
```

```
'Coeficientes de la ecuacion cubica  
Dim coef As Double, coef_1 As Double, coef_2 As Double, coef_3 As Double  
Dim coef1 As Double, coef1_1 As Double, coef1_2 As Double, coef1_3 As Double  
Dim coef2 As Double, coef2_1 As Double, coef2_2 As Double, coef2_3 As Double  
Dim cubo As Double
```

```
Private Sub Text1_DblClick()  
    Dim Filtros As String  
    'Si ocurre un error ejecutar ManipularErrorAbrir  
    On Error GoTo ManipularErrorAbrir  
    'Generar un error cuando se pulse Cancelar  
    CommonDialog1.CancelError = True  
    'filtros  
    Filtros = "Ficheros de datos (*.dat)|*.dat|"  
    CommonDialog1.Filter = Filtros  
    'Filtro por defecto  
    CommonDialog1.FilterIndex = 1  
    'Visualizar la caja de dialogo  
    CommonDialog1.ShowOpen  
    'Commondialog1.filetitle contiene el nombre  
    'del fichero seleccionado  
    Text1.Text = CommonDialog1.FileTitle
```

```
SalirAbrir:  
Exit Sub
```

```
ManipularErrorAbrir:
  If Err.Number = cdlCancel Then Exit Sub
  MsgBox Err.Description
  Resume SalirAbrir

End Sub

Private Sub Text2_DblClick()
  Dim Filtros As String
  'Si ocurre un error ejecutar ManipularErrorAbrir
  On Error GoTo ManipularErrorAbrir
  'Generar un error cuando se pulse Cancelar
  CommonDialog1.CancelError = True
  'filtros
  Filtros = "Ficheros de datos (*.dat)|*.dat|"
  CommonDialog1.Filter = Filtros
  'Filtro por defecto
  CommonDialog1.FilterIndex = 1
  'Visualizar la caja de di\ss{}logo
  CommonDialog1.ShowOpen
  'CommonDialog1.filetitle contiene el nombre
  'del fichero seleccionado
  Text2.Text = CommonDialog1.FileTitle

SalirAbrir:
  Exit Sub

ManipularErrorAbrir:
  If Err.Number = cdlCancel Then Exit Sub
  MsgBox Err.Description
  Resume SalirAbrir
End Sub

Private Sub inicio_Click()

  Dim c As Single, k As Integer
```

```
Open Text2.Text For Input As #1
Do Until EOF(1)
    Line Input #1, cb$
    n = n + 1
Loop
Close 1

objetos = n - 1
nt = objetos * (objetos - 1) / 2

'Lectura de la matriz con la configuracion inicial
ReDim A(objetos, objetos)
Open Text2.Text For Input As #1
Line Input #1, b$
For i = 1 To objetos
    For j = 1 To objetos - 1
        Input #1, A(i, j)
    Next
Next
Close 1

'Lectura de las matrices de disimilaridades
ReDim disim1(objetos, objetos), disim2(objetos, objetos)
For i = 1 To objetos
    disim1(i, i) = 0
    disim2(i, i) = 0
Next

Open Text1.Text For Input As #2

Line Input #2, b$
For i = 2 To objetos
    For j = 1 To i - 1
        Input #2, disim1(i, j)
        'disim1(i, j) = disim1(i, j) ^ 2
    Next
Next
```

```
Line Input #2, b$
For i = objetos + 2 To 2 * objetos
  For j = 1 To i - (1 + objetos)
    Input #2, disim2(i - objetos, j)
  Next
Next
Close 2

'Construccion del vector de disimilaridades como la media
'de las disimilaridades de las dos matrices de partida

ReDim disim(nt)
h = 0
For i = 2 To objetos
  For j = 1 To i - 1
    h = h + 1
    disim(h) = (disim1(i, j) + disim2(i, j)) / 2
  Next
Next

'REALIZACION DEL OPTIMAL SCALING

'Ordenacion del vector de disimilaridades
ReDim x(nt)
Dim orden1 As Integer

For j = 1 To nt
  x(j) = j
Next

For j = 1 To nt - 1
  For i = 1 To nt - 1
    If disim(i) >= disim(i + 1) Then
      orden1 = x(i)
      tempor = disim(i)
      x(i) = x(i + 1)
```

```
        disim(i) = disim(i + 1)
        x(i + 1) = orden1
        disim(i + 1) = tempor
    End If
Next i
Next j

Label3.Visible = True
ss_ini.Visible = True
iteracion.Visible = True

Call distancia(A)

Call optimal(dist, dispar)

Open "sstress.txt" For Append As #2
    Print #2, "SSTRESS inicial:"
    Print #2, ""
Close 2
ss_ini.Text = sstress(dist, dispar)

End Sub

Private Sub iteracion_Click()
    ReDim A_ant(objetos, objetos - 1)

    ss0 = Val(ss_ini.Text)

    'Comienzo del ciclo de iteraciones
    iter = 0
    Do
        iter = iter + 1

        For i = 1 To objetos
            For j = 1 To objetos - 1
                A_ant(i, j) = A(i, j)
            Next j
        Next i
    Loop
End Sub
```

```
Next i

For s = 1 To objetos - 1
  For r = 1 To objetos
    'estimar coeficientes de la ecuacion cubica
    Call coeficiente(r, s, dispar, A)
    'resolver la ecuacion cubica
    Call ecuacion(r, s, A, coef, coef1, coef2)
    A(r, s) = cubo
  Next r
Next s

Call distancia(A)

Call optimal(dist, dispar)

Open "sstress.txt" For Append As #2
  Print #2, "ITERACION:", iter
  Print #2, ""
Close 2

SS = sstress(dist, dispar)
dif = ss0 - SS
ss0 = SS
If dif < 0 Then GoTo fin1

Loop Until SS < 0.001

GoTo fin2

'Escritura de los resultados finales
fin1:
ss_fin.Text = ss0

Open "config.mia" For Output As #1
  For i = 1 To objetos
    For j = 1 To objetos - 2
```

```
        Write #1, A_ant(i, j) * 2,
    Next j
    Write #1, A_ant(i, objetos - 1) * 2
Next i
Close 1
ss_mejora.Text = dif
GoTo fin3

'Escritura de los resultados finales en caso
'de detencion de las iteraciones

fin2:
ss_fin.Text = SS

Open "config.mia" For Output As #1
    For i = 1 To objetos
        For j = 1 To objetos - 2
            Write #1, A(i, j) * 2,
        Next j
        Write #1, A(i, objetos - 1) * 2
    Next i
Close 1
ss_mejora.Text = dif

fin3:
Close 2
End Sub

Public Sub distancia(A)
    ReDim dist(nt, objetos - 1)

    For dimension = 1 To objetos - 1
        h = 0
        For i = 2 To objetos
            For j = 1 To i - 1
                h = h + 1
                For k = 1 To dimension
```

```
        dist(h, dimension) = dist(h, dimension) + _
        (A(i, k) - A(j, k)) ^ 2
    Next k
Next j
Next i
Next dimension
End Sub

Public Sub optimal(dist, dispar)

    ReDim dispar(nt, objetos - 1)

    'Asociar las distancias a las disimilaridades ordenadas
    For dimension = 1 To objetos - 1
        For i = 1 To nt
            dispar(i, dimension) = dist(x(i), dimension)
        Next i
    Next dimension

    'Obtencion de las disparidades. Regresion monotona por minimos cuadrados

    For dimension = 1 To objetos - 1
        For i = 2 To nt
            im1 = i - 1
            'determinar si el orden es correcto
            If dispar(i, dimension) > dispar(im1, dimension) Then
                GoTo continuar
            Else
                Sum = dispar(i, dimension)
                fn = 1
                'determinar el n\umero de elementos del bloque
                For j = 1 To im1
                    imj = i - j
                    Sum = Sum + dispar(imj, dimension)
                    fn = fn + 1
                    dispt = Sum / fn
                    If j = im1 Then GoTo escribir
```

```

        If dispt > dispar(imj - 1, dimension) Then GoTo escribir
    Next j
End If
escribir: 'ajustar las disparidades a la media del bloque
    For j = imj To i
        dispar(j, dimension) = dispt
    Next j
continuar:
    Next i

'Normalizacion de las disparidades
    ssq = 0
    spd = 0
    For i = 1 To nt
        ssq = ssq + dist(i, dimension) ^ 2
        spd = spd + dispar(i, dimension) * dist(i, dimension)
    Next i

    ratio = ssq / spd

    For j = 1 To nt
        dispar(j, dimension) = dispar(j, dimension) * ratio
    Next j
Next dimension
End Sub

Public Function sstress(dist, dispar)

'Definicion de la funcion SSSTRESS
sstress = 0
Open "sstress.txt" For Append As #2

For dimension = 1 To objetos - 1
    sstress_d = 0
    For i = 1 To nt
        sstress_d = sstress_d + (dist(i,dimension) - dispar(i,dimension)) ^ 2
    Next i

```

```
'Normalizacion. Se normaliza con las disparidades

sv4sq = 0
For i = 1 To nt
    sv4sq = sv4sq + dispar(i, dimension) ^ 2
Next i
Print #2, "dimension", dimension, "***", sstress_d / sv4sq
sstress = sstress + sstress_d / sv4sq

Next dimension
sstress = sstress / (objetos - 1)
Print #2, "SSTRESS anidado:", sstress
Print #2, "*****"
Close 2
End Function

Public Sub coeficiente(r, s, dispar, A)

'definicion de la nueva matriz de disparidades.
'esta matriz va a ser una matriz en tres dimensiones
'la tercera dimension correpondera a la dimension de la
'configuracion

ReDim d(objetos, objetos, objetos - 1)
For dimension = 1 To objetos - 1
    h = 0
    For i = 2 To objetos
        For j = 1 To i - 1
            h = h + 1
            d(i, j, dimension) = dispar(h, dimension)
            d(j, i, dimension) = d(i, j, dimension)
        Next j
    Next i
Next dimension

For dimension = 1 To objetos - 1
```

```
For i = 1 To objetos
    d(i, i, dimension) = 0
Next i
Next dimension

'Definicion de los coeficientes de la ecuacion

'Coeficiente de x^3 = 1

'Coeficiente de x^2
coef2 = 0

For j = 1 To r - 1
    coef2 = coef2 + A(j, s)
Next j

For j = r + 1 To objetos
    coef2 = coef2 + A(j, s)
Next j

coef2 = -3 * coef2 / (objetos - 1)

'Coeficiente de x
coef1_1 = 0
For dimension = s To objetos - 1
    If dimension = 1 Then GoTo continua2:
    For j = 1 To objetos
        For h = 1 To dimension
            If h = s Then h = h + 1
            coef1_1 = coef1_1 + (A(r, h) - A(j, h)) ^ 2
        Next h
    Next j
Next dimension

continua2:
coef1_2 = 0
For dimension = s To objetos - 1
```

```
For j = 1 To r - 1
    coef1_2 = coef1_2 + d(r, j, dimension)
Next j
For j = r To objetos
    coef1_2 = coef1_2 + d(r, j, dimension)
Next j
Next dimension

coef1_3 = 0
For j = 1 To r - 1
    coef1_3 = coef1_3 + A(j, s) ^ 2
Next j
For j = r + 1 To objetos
    coef1_3 = coef1_3 + A(j, s) ^ 2
Next j
coef1_3 = 3 * (objetos - s) * coef1_3

coef1 = (coef1_1 - coef1_2 + coef1_3) / ((objetos - 1) * (objetos - s))

'Termino independiente
coef_1 = 0
For dimension = s To objetos - 1
    If dimension = 1 Then GoTo continua3:
    For j = 1 To objetos
        coef_1_1 = 0
        For h = 1 To dimension
            If h = s Then h = h + 1
            coef_1_1 = coef_1_1 + (A(r, h) - A(j, h)) ^ 2
        Next h
        coef_1 = coef1 + coef_1_1 * A(j, s)
    Next j
Next dimension

continua3:
coef_2 = 0
For dimension = s To objetos - 1
    For j = 1 To r - 1
```

UNIVERSIDAD DE GRANADA

11 MAYO 2000

COMISION DE DOCTORADO

```
        coef_2 = coef_2 + d(r, j, dimension) * A(j, s)
    Next j
    For j = r + 1 To objetos
        coef_2 = coef_2 + d(r, j, dimension) * A(j, s)
    Next j
Next dimension

coef_3 = 0
For j = 1 To r - 1
    coef_3 = coef_3 + A(j, s) ^ 3
Next j
For j = r + 1 To objetos
    coef_3 = coef_3 + A(j, s) ^ 3
Next j
coef_3 = coef_3 * (objetos - s)

coef = (-coef_1 + coef_2 - coef_3) / ((objetos - 1) * (objetos - s))
End Sub

Public Sub ecuacion(r, s, A, coef, coef1, coef2)
    Dim p As Double, q As Double

    p = coef1 - coef2 / 3
    q = (2 * (coef2 ^ 3) / 27) - (coef1 * coef2 / 3) + coef

    cubo = A(r, s) * 2
    For k = 1 To 50
        cubo = cubo - (cubo ^ 3 + p * cubo + q) / (3 * cubo ^ 2 + p)
    Next k

    cubo = cubo - coef2 / 3
End Sub

Private Sub fin_Click()
    End
End Sub
```

## 2 Programa para la determinación de las medidas que establecen la robustez del modelo

VERSION 4.00

Begin VB.Form Matchals

```
BackColor      = &H00FFFFC0&
Caption        = "Parametros de ajuste del analisis Matchals"
ClientHeight   = 3480
ClientLeft     = 1920
ClientTop      = 1710
ClientWidth    = 6615
Height         = 3885
Left           = 1860
LinkTopic      = "Form1"
ScaleHeight    = 3480
ScaleWidth     = 6615
Top            = 1365
Width          = 6735
```

Begin VB.CommandButton Command1

```
Caption        = "Fin"
Height         = 495
Left           = 3720
TabIndex       = 7
Top            = 2400
Visible        = 0 'False
Width          = 1095
```

End

Begin VB.CommandButton Command3

```
Caption        = "Calculo de los coeficientes de ajuste del modelo"
Height         = 975
Left           = 840
TabIndex       = 6
Top            = 2160
Width          = 2175
```

End

```
Begin VB.TextBox fichero2
    Height      = 375
    Left        = 1560
    TabIndex    = 5
    Top         = 960
    Width       = 4695
End
```

```
Begin VB.TextBox fichero3
    Height      = 375
    Left        = 1560
    TabIndex    = 3
    Top         = 1560
    Width       = 4695
End
```

```
Begin VB.TextBox fichero1
    Height      = 375
    Left        = 1560
    TabIndex    = 0
    Top         = 360
    Width       = 4695
End
```

```
Begin MSComDlg.CommonDialog CommonDialog1
    Left        = 5880
    Top         = 2160
    _Version    = 65536
    _ExtentX    = 847
    _ExtentY    = 847
    _StockProps = 0
End
```

```
Begin VB.Label f
    BackColor   = &H00FFFFC0&
```

```
        Caption      = "Configuracion media:"
        Height       = 375
        Left         = 120
        TabIndex     = 4
        Top          = 960
        Width        = 1215
    End

    Begin VB.Label Label2
        BackColor     = &H00FFFFC0&
        Caption       = "Configuracion ALNEST:"
        Height        = 495
        Left          = 120
        TabIndex      = 2
        Top           = 1560
        Width         = 1215
    End

    Begin VB.Label Label1
        BackColor     = &H00FFFFC0&
        Caption       = "Configuraciones tras el GPA:"
        Height        = 495
        Left          = 120
        TabIndex      = 1
        Top           = 240
        Width         = 1215
    End
End

Attribute VB_Name = "Matchals"
Attribute VB_Creatable = False
Attribute VB_Exposed = False

Option Base 1
Dim datos As Integer
Dim linea As Integer
Dim num1() As Single, num2() As Single, norm() As Single
```

```
Dim singles() As Single, disp_modelo() As Single
Dim cen() As Single

Private Sub fichero1_DblClick()
    'fichero donde se guardan las configuraciones
    'obtenidas tras el jacknife

    Dim Filtros As String
    'Si ocurre un error ejecutar ManipularErrorAbrir
    On Error GoTo ManipularErrorAbrir
    'Generar un error cuando se pulse Cancelar
    CommonDialog1.CancelError = True
    'filtros
    Filtros = "Ficheros de datos (*.jac)|*.jac|"
    CommonDialog1.Filter = Filtros
    'Filtro por defecto
    CommonDialog1.FilterIndex = 1
    'Visualizar la caja de di\ss{}logo
    CommonDialog1.ShowOpen
    'Commondialog1.filetitle contiene el nombre
    'del fichero seleccionado
    fichero1.Text = CommonDialog1.FileTitle

SalirAbrir:
    Exit Sub

ManipularErrorAbrir:
    If Err.Number = cdlCancel Then Exit Sub
    MsgBox Err.Description
    Resume SalirAbrir

End Sub

Private Sub fichero2_DblClick()
    'Fichero con la configuracion obtenida como la media
    'de las configuraciones tras el GPA
```

```
Dim Filtros As String
'Si ocurre un error ejecutar ManipularErrorAbrir
On Error GoTo ManipularErrorAbrir
'Generar un error cuando se pulse Cancelar
CommonDialog1.CancelError = True
'filtros
Filtros = "Ficheros de datos (*.jac)|*.jac|"
CommonDialog1.Filter = Filtros
'Filtro por defecto
CommonDialog1.FilterIndex = 1
'Visualizar la caja de di\ss{}logo
CommonDialog1.ShowOpen
'Commondialog1.filetitle contiene el nombre
'del fichero seleccionado
fichero2.Text = CommonDialog1.FileTitle

SalirAbrir:
Exit Sub

ManipularErrorAbrir:
If Err.Number = cdlCancel Then Exit Sub
MsgBox Err.Description
Resume SalirAbrir

End Sub

Private Sub fichero3_DblClick()

'Fichero con la configuracion alnest
Dim Filtros As String
'Si ocurre un error ejecutar ManipularErrorAbrir
On Error GoTo ManipularErrorAbrir
'Generar un error cuando se pulse Cancelar
CommonDialog1.CancelError = True
'filtros
Filtros = "Ficheros de datos (*.jac)|*.jac|"
CommonDialog1.Filter = Filtros
```

```
'Filtro por defecto
CommonDialog1.FilterIndex = 1
'Visualizar la caja de di\ss{}logo
CommonDialog1.ShowOpen
'Commondialog1.filetitle contiene el nombre
'del fichero seleccionado
fichero3.Text = CommonDialog1.FileTitle

SalirAbrir:
Exit Sub

ManipularErrorAbrir:
If Err.Number = cdlCancel Then Exit Sub
MsgBox Err.Description
Resume SalirAbrir

End Sub

Private Sub Command3_Click()

'Numerador de la dispersion por el modelo
Open fichero1.Text For Input As #1
Do Until EOF(1)
    Line Input #1, a$
    linea = linea + 1
Loop
Close 1

'n datos para cada una de las n configuraciones que se leen

datos = Sqr(linea)

'Se colocan los datos en una matriz tri-dimensional

ReDim Y(datos, datos - 1, datos)
Open fichero1.Text For Input As #1
For k = 1 To datos
```

```
        For i = 1 To datos
            For j = 1 To datos - 1
                Input #1, Y(i, j, k)
            Next j
        Next i
    Next k
Close 1

'lectura de la matriz con la configuracion centroide
ReDim cen(datos, datos - 1)
Open fichero2.Text For Input As #2
    For i = 1 To datos
        For j = 1 To datos - 1
            Input #2, cen(i, j)
        Next j
    Next i
Close 2

ReDim num1(datos - 1)

For q = 1 To datos - 1
    num1(q) = 0
    For k = 1 To datos
        For i = 1 To datos
            For j = 1 To q
                num1(q) = num1(q) + (Y(i, j, k) - cen(i, j)) ^ 2
            Next j
        Next i
    Next k
Next q

'Numerador de la dispersion por los residuos

ReDim num2(datos - 1)
ReDim cen(datos, datos - 1)
Open fichero2.Text For Input As #2
    For i = 1 To datos
```

```
        For j = 1 To datos - 1
            Input #2, cen(i, j)
        Next j
    Next i
Close 2

ReDim X(datos, datos - 1)
Open fichero3.Text For Input As #2
For i = 1 To datos
    For j = 1 To datos - 1
        Input #2, X(i, j)
    Next j
Next i
Close 2

For q = 1 To datos - 1
    num2(q) = 0
    For i = 1 To datos
        For j = 1 To q
            num2(q) = num2(q) + (cen(i, j) - X(i, j)) ^ 2
        Next j
    Next i
    num2(q) = datos * num2(q)
Next q

'Factor de normalizacion

ReDim norm(datos - 1)
ReDim singles(datos - 1)
ReDim disp_modelo(datos - 1)
ReDim X(datos, datos - 1)
Open fichero3.Text For Input As #2
For i = 1 To datos
    For j = 1 To datos - 1
        Input #2, X(i, j)
    Next j
Next i
```

```
Close 2

ReDim Y(datos, datos - 1, datos)
Open fichero1.Text For Input As #1
  For k = 1 To datos
    For i = 1 To datos
      For j = 1 To datos - 1
        Input #1, Y(i, j, k)
      Next j
    Next i
  Next k
Close 1

'factor de normalizacion
For q = 1 To datos - 1
norm(q) = 0
  For k = 1 To datos
    For i = 1 To datos
      For j = 1 To q
        norm(q) = norm(q) + (Y(i, j, k) - X(i, j)) ^ 2
      Next j
    Next i
  Next k
Next q

For q = 1 To datos - 1
  singles(q) = 1 - num2(q) / norm(q)
  disp_modelo(q) = 1 - num1(q) / norm(q)
Next q

Open "ajuste.dat" For Output As #1
  Print #1, "dimension", "    modelo", "    residuos "
  Print #1, "*****", "    *****", "    *****"
  For q = 1 To datos - 1

    Print #1, q, disp_modelo(q), singles(q)
    Print #1, ""
```

```
    Next q  
  Close 1
```

```
    Command1.Visible = True  
  End Sub
```

```
Private Sub Command1_Click()  
  End  
End Sub
```

## Bibliografía

- [1] Carroll, J.D. & Chang, J.J. (1970). *Analysis of individual differences in multidimensional scaling via an n-way generalization of "Eckart-Young" decomposition*. PSYCHOMETRIKA 35, 283-319.
- [2] Cliff, N. (1966). *Orthogonal rotation to congruence*. PSYCHOMETRIKA 31, 33-42.
- [3] Commandeur, J.F. (1991). *Matching configurations*. M& T SERIES, 19. DSWO. Press.
- [4] De Leeuw, J. & Heiser, W. (1980). *Multidimensional scaling with restrictions on the configuration*. En Krishnaiah, P.R. (ed.), "MULTIVARIATE ANALYSIS V", AMSTERDAM: NORTH HOLLAND.
- [5] De Leeuw, J. & Meulman J. (1986). *A special Jackknife for MDS*. JOURNAL OF CLASSIFICATION 3, 97-112.
- [6] Dinkelbach, W. (1967). *On nonlinear fractional programming*. MANAGEMENT SCIENCE, 13, 492-498.
- [7] Everitt, B.S. & Gower, J.C. (1981). *Plotting the optimum positions of an array of cortical electrical phosphenes*. En Barnett (Ed.) "INTERPRETING MULTIVARIATE DATA". CHICHESTER: WILEY.
- [8] Gower, J.C. (1975). *Generalized Procrustes analysis*. PSYCHOMETRIKA 40, 33-51.
- [9] Gower, J.C. (1984). *Distance matrices and their euclidean approximation*. DATA ANALYSIS AND INFORMATICS III. DIDAY
- [10] Gower, J.C. (1986). *Euclidean distance matrices*. MULTIDIMENSIONAL DATA ANALYSIS. DSWO PRESS, LEIDEN.

- [11] Green, B.F. (1952). *The orthogonal approximation of an oblique structure in factor analysis*. PSYCHOMETRIKA 17, 429-440.
- [12] Groenen, P.J.F. (1990). *The tunneling method applied to multidimensional scaling*. PROGRESS REPORT 1. RESEARCH REPORT RR-90-03. LEIDEN: DEPARTMENT OF DATA THEORY.
- [13] Groenen, P.J.F. (1993). *The majorization approach to multidimensional scaling: some problems and extentions*. M& T SERIES, 26. DSWO. PRESS.
- [14] Groenen, P.J.F. & Heiser, W.J. (1991). *An improved tunneling function for finding a decreasing series of local minima*. (TECH. REP. NO. RR-91-06). LEIDEN, THE NETHERLANDS: DEPARTMENT OF DATA THEORY, LEIDEN UNIVERSITY.
- [15] Groenen, P.J.F. & Heiser, W.J. (1996). *The tunneling method for global optimization in Multidimensional Scaling*. PSYCHOMETRIKA 61, 529-550.
- [16] Guttman, L. (1968). *A general nonmetric technique for finding the smallest coordinate space for a configuration of points*. PSYCHOMETRIKA 33, 469-506.
- [17] Heiser, W. (1995). *Nested Solutions in Least Squares Multidimensional Scaling*. DEPARTMENT OF DATA THEORY, FACULTY OF SOCIAL SCIENCES, UNIVERSITY OF LEIDEN, LEIDEN, THE NETHERLANDS.
- [18] Heiser, W.J. & Groenen, P.J.F. (1997). *Cluster differences scaling with a within-clusters loss component and a fuzzy successive approximation strategy to avoid local minima*. PSYCHOMETRIKA 62, 63-83.
- [19] Horst, R. & Tuy, H. (1990). *Global optimization: Deterministic approaches*. BERLIN: SPRINGER-VERLAG.
- [20] Hubert, L.J. & Arabie, P. (1986). *Unidimensional scaling and combinatorial optimization*. En: J. De Leeuw, W.J. Heiser, J. Meulman & F. Critchley (Eds.), "MULTIDIMENSIONAL DATA ANALYSIS", 181-196. LEIDEN: DSWO PRESS, UNIVERSITY OF LEIDEN.
- [21] Kristof, W. & Wingersky, B. (1971). *Generalization of the orthogonal Procrustes rotation procedure to more than two matrices*. PROCEEDINGS, 79TH ANNUAL CONVENTION OF THE AMERICAN PSYCHOLOGICAL ASSOCIATION, PP 81-90.

- [22] Kruskal, J.B. (1964a). *Multidimensional scaling by optimizing goodness-of-fit to a nonmetric hypothesis*. PSYCHOMETRIKA 29, 1-27.
- [23] Kruskal, J.B. (1964b). *Nonmetric multidimensional scaling: a numerical method*. PSYCHOMETRIKA 29, 115-129.
- [24] Laurence G. Grim & Paur R. Yarnold (1995). *Reading and understanding Multivariate Statistics*. AMERICAN PSYCHOLOGICAL ASSOCIATION.
- [25] McFarlane, M. & Young, F.W. (1994). *Graphical Sensitivity Analysis for Multidimensional Scaling*. JOURNAL OF COMPUTATIONAL AND GRAPHICAL STATISTICS, 3, NUMBER 1, 23-33.
- [26] McGee, V.E. (1968). *Multidimensional scaling of N sets of similarity measures: A nonmetric individual differences approach*. MULTIVARIATE BEHAVIORAL RESEARCH 3, 233-248.
- [27] Peay, E.R. (1988). *Multidimensional rotation and scaling of configurations to optimal agreement*. PSYCHOMETRIKA 53, 199-208.
- [28] Ramsay, J.O. (1977). *Maximum likelihood estimation in multidimensional scaling*. PSYCHOMETRIKA 42, 241-266.
- [29] Ramsay, J.O. (1978a). *Confidence regions for multidimensional scaling analysis*. PSYCHOMETRIKA 43, 145-160.
- [30] Ramsay, J.O. (1980). *Some small sample results for maximum likelihood estimation in multidimensional scaling*. PSYCHOMETRIKA 45, 141-146.
- [31] Ramsay, J.O. (1982). *Some statistical approaches to multidimensional scaling data*. J. R. STATIS. SOC. A 145, 285-312.
- [32] Román, Y., Vera, J.F. & González, A. (1995). *Diseños Anidados con SSTRESS*. XXII CONGRESO NACIONAL DE ESTADÍSTICA E INVESTIGACIÓN OPERATIVA (Sevilla).
- [33] Román, Y., Vera, J.F. & González, A. (1997). *Soluciones anidadas en ALSCAL para datos de proximidad con ponderaciones*. XXIII CONGRESO NACIONAL DE ESTADÍSTICA E INVESTIGACIÓN OPERATIVA (Valencia).

- [34] Román, Y., Vera, J.F. & González, A. (1998). *Análisis de las soluciones anidadas en MDS*. XXIV CONGRESO NACIONAL DE ESTADÍSTICA E INVESTIGACIÓN OPERATIVA (Almería).
- [35] Román, Y., González, A., & Vera, J.F. (1999). *Anidamiento en MDS mínimo cuadrático no métrico*. ESTADÍSTICA ESPAÑOLA. Vol. 41, Número 144.
- [36] Schönemann, P.H. (1966). *A generalized solution of the orthogonal procrustes problem*. PSYCHOMETRIKA 31, 1-10.
- [37] Schönemann, P.H. (1972). *An algebraic solution for a class of subjective metric models*. PSYCHOMETRIKA 37, 441-451.
- [38] Sherman, C.R. (1972). *Nonmetric MDS: A Monte Carlo study of the basic parameters*. PSYCHOMETRIKA 37, 323-355.
- [39] Shepard, R.N. (1962a). *The analysis of proximities: multidimensional scaling with an unknown distance function I*. PSYCHOMETRIKA 27, 125-140.
- [40] Shepard, R.N. (1962b). *The analysis of proximities: multidimensional scaling with an unknown distance function II*. PSYCHOMETRIKA 27, 219-246.
- [41] Schönemann, P.H. (1966). *A generalized solution of the orthogonal Procrustes problem*. PSYCHOMETRIKA 31, 1-10.
- [42] Schönemann, P.H. & Carroll, R.M. (1970). *Fitting one matrix to another under choice of a central dilation and a rigid motion*. PSYCHOMETRIKA 35, 245-256.
- [43] Takane, Y. (1981). *Multidimensional successive categories scaling: a maximum likelihood method*. PSYCHOMETRIKA 46, 9-28.
- [44] Takane, Y., Young, F.W. & de Leeuw, J. (1977). *Nonmetric individual differences multidimensional scaling: an alternating least squares method with optimal scaling features*. PSYCHOMETRIKA 42, 7-67.
- [45] Ten Berge, J.M.F. (1977). *Orthogonal Procrustes rotation for two or more matrices*. PSYCHOMETRIKA, 42, 267-276.
- [46] Ten Berge, J.M.F. & Knol, D.L. (1984). *Orthogonal rotations to maximal agreement for two or more matrices of different column orders*. PSYCHOMETRIKA 49, 49-55.

- [47] Törn, A.A. & Zilinskas, A. (1989). *Global optimization*. En: G.Goos & J. Hartmanis (Eds.). "LECTURE NOTES IN COMPUTER SCIENCE", VOL.350. BERLIN: SPRINGER-VERLAG.
- [48] Young, G. & Householder, A.S. (1938). *Discussion of a set of points in terms of their mutual distances*. PSYCHOMETRIKA 3, 19-22
- [49] Young, F.W., Takane, Y. & Lewyckyj, R. (1978). *Three notes on ALSCAL*. PSYCHOMETRIKA, 43, 433-435.
- [50] Torgerson, W.S. (1952). *Multidimensional Scaling: 1. Theory and Method*. PSYCHOMETRIKA 17, 401-419.
- [51] Tucker, L.R. & Messick, S. (1963). *An individual differences model for multidimensional scaling*. PSYCHOMETRIKA 28, 333-367.
- [52] Young, F.W. & Rheingans, P. (1991). *Visualizing Structure in High-Dimensional Multivariate Data*. THE IBM JOURNAL OF RESEARCH AND DEVELOPMENT, VOL. 35, NUMBER 1/2.
- [53] Zinnes, J.L. & MacKay, D.B. (1983). *Probabilistic multidimensional scaling: complete and incomplete data*. PSYCHOMETRIKA 48, 27-48.