Lecture 8: Multidimensional scaling

Advanced Applied Multivariate Analysis

STAT 2221, Fall 2013

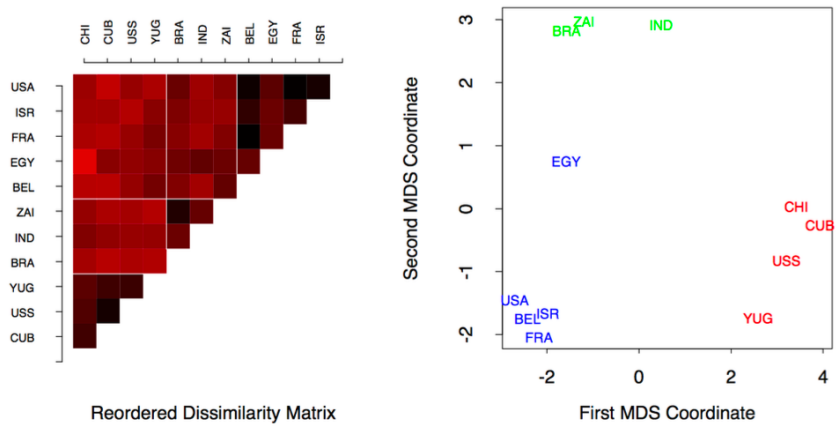
Sungkyu Jung
Department of Statistics
University of Pittsburgh

E-mail: sungkyu@pitt.edu
http://www.stat.pitt.edu/sungkyu/AAMA/

Diapositive 2 — Objectif du MDS

Objectif du MDS (Multidimensional Scaling)
À partir de mesures de dissimilarité entre paires d'objets, on cherche à reconstruire une **carte** qui **préserve les distances** entre les points.

- On peut partir de **toute mesure de dissimilarité** (pas forcément une distance métrique).
- La carte reconstruite fournit des **coordonnées** $x_i = (x_{i1}, x_{i2})$, et la **distance naturelle** est $\| x_i - x_j \|_2$.



Diapositive 3 — Famille de méthodes MDS

Le MDS n'est pas une seule méthode, mais une **famille d'algorithmes** visant à trouver une configuration optimale dans un espace de faible dimension (souvent **p = 2 ou 3**).

Les principales méthodes MDS sont :

1. **MDS classique** (*Classical MDS*)
2. **MDS métrique** (*Metric MDS*)
3. **MDS non métrique** (*Non-metric MDS*)

Diapositive 4 — Exemple : perception des couleurs

Étude de la **perception des couleurs par la vision humaine** (Ekman, 1954 ; Izenman §13.2.1).

- 14 couleurs, différant uniquement par leur **teinte** (longueurs d'onde de 434 à 674 μm).
- 31 personnes évaluent chaque paire de couleurs sur une **échelle de 0 à 4** :
  - 0 = aucune similarité,
  - 4 = identiques.
- On obtient donc $\binom{14}{2}$ paires.
- On fait la **moyenne** des 31 notations pour chaque paire → cela donne une **mesure de similarité**.
- On la convertit ensuite en **dissimilarité** :
  dissimilarité $= 1 -$ similarité.

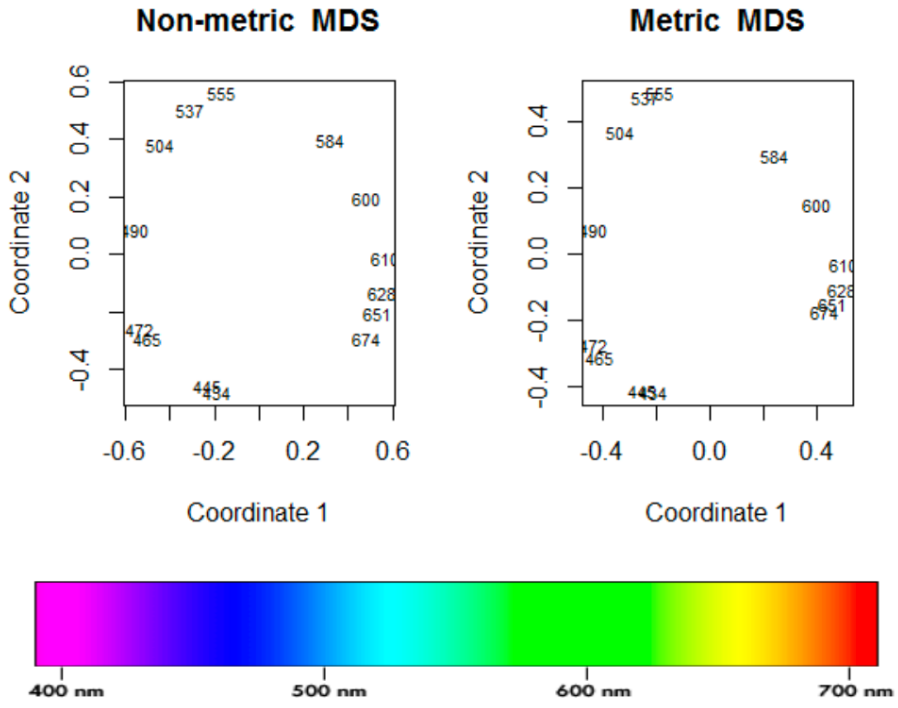
Diapositive 5 — Matrice de dissimilarité

On obtient une **matrice 14 × 14** de dissimilarités, symétrique, avec des **zéros sur la diagonale**.
Le MDS cherche une configuration **à deux dimensions** représentant ces couleurs.

```
     434  445  465  472  490  504  537  555  584  600  610  628  651
445 0.14
465 0.58 0.50
472 0.58 0.56 0.19
490 0.82 0.78 0.53 0.46
504 0.94 0.91 0.83 0.75 0.39
537 0.93 0.93 0.90 0.90 0.69 0.38
555 0.96 0.93 0.92 0.91 0.74 0.55 0.27
584 0.98 0.98 0.98 0.98 0.93 0.86 0.78 0.67
600 0.93 0.96 0.99 0.99 0.98 0.92 0.86 0.81 0.42
610 0.91 0.93 0.98 1.00 0.98 0.98 0.95 0.96 0.63 0.26
628 0.88 0.89 0.99 0.99 0.99 0.98 0.98 0.97 0.73 0.50 0.24
651 0.87 0.87 0.95 0.98 0.98 0.98 0.98 0.98 0.80 0.59 0.38 0.15
674 0.84 0.86 0.97 0.96 1.00 0.99 1.00 0.98 0.77 0.72 0.45 0.32 0.24
```

Diapositive 6 — Résultat sur la perception des couleurs

Le MDS reconstruit le célèbre **cercle des couleurs en deux dimensions**,
montrant que la **perception humaine de la teinte** est naturellement circulaire.



Diapositive 7 — Distance, dissimilarité et similarité

Les notions de **distance**, **dissimilarité** et **similarité (ou proximité)** sont définies pour toute paire d'objets.

En mathématiques, une fonction de distance (ou **métrique**) satisfait :

1. $d(x, y) \geq 0$
2. $d(x, y) = 0$ si et seulement si $x = y$
3. $d(x, y) = d(y, x)$
4. $d(x, z) \leq d(x, y) + d(y, z)$ (inégalité triangulaire)

On peut alors se demander si les dissimilarités données sont **vraiment des distances**, et si elles peuvent être **interprétées comme des distances euclidiennes**.

Diapositive 8 — Distance euclidienne et non euclidienne

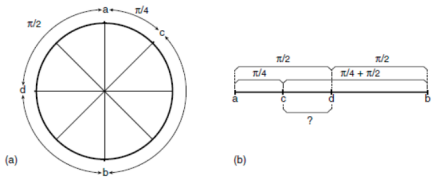
Étant donnée une matrice de dissimilarités $D = (d_{ij})$, le MDS cherche à trouver des points $x_1, \ldots, x_n \in \mathbb{R}^p$ tels que :

$$d_{ij} \approx \| x_i - x_j \|_2$$

- S'il existe une configuration exacte, on parle de **distance euclidienne**.
- Mais parfois, il n'existe **aucune configuration** qui reproduise exactement d_{ij}.
  → On parle alors de **distance non euclidienne**.

Diapositive 9 — Exemple de distance non euclidienne

- La **distance radiale sur un cercle** (la longueur de l'arc entre deux points) est bien une **métrique**,
  mais **ne peut pas être représentée exactement** dans un espace euclidien $\mathbb{R}^p$.



Point	a	b	c	d
a	0.0000	3.1416	0.7854	1.5708
b	3.1416	0.0000	2.3562	1.5708
c	0.7854	2.3562	0.0000	2.3562
d	1.5708	1.5708	2.3562	0.0000

- Le MDS essaie malgré tout de trouver une **configuration approchée** minimisant l'écart entre
  d_{ij} et $\| x_i - x_j \|_2$.

Diapositive 10 — MDS classique : théorie

On suppose que la matrice de distances $D = (d_{ij})$ est **euclidienne**.

L'objectif du **MDS classique (cMDS)** est de trouver une matrice de coordonnées $X = [x_1, \dots, x_n]$ telle que :

$$\| x_i - x_j \| = d_{ij}$$

Cette solution **n'est pas unique**, car un déplacement global $X^* = X + c$, $c \in \mathbb{R}^q$, donne les mêmes distances.

On impose donc une **centrage** des coordonnées :

$$\sum_{i=1}^{n} x_{ik} = 0 \forall k$$

afin de stabiliser la solution et faciliter la réduction de dimension.

Diapo 11 — MDS classique : théorie (suite)

En résumé, le **MDS classique (cMDS)** cherche une configuration **centrée** $x_1, \dots, x_n \in \mathbb{R}^q$ (pour un certain $q \geq n - 1$)
telle que leurs distances mutuelles correspondent à celles de D.

On calcule plutôt la **matrice de Gram** $B = X'X$ plutôt que X directement.
C'est une **matrice de produits scalaires** (puisque X est centrée).

On a la relation :

$$d_{ij}^2 = b_{ii} + b_{jj} - 2b_{ij}$$

provenant de :

$$\| x_i - x_j \|^2 = x_i' x_i + x_j' x_j - 2x_i' x_j$$

Diapo 12 — MDS classique : théorie (suite 2)

La contrainte de centrage ($\sum_i \quad x_{ik} = 0$) implique :

$$\sum_{i=1}^{n} b_{ij} = \sum_{i=1}^{n}\sum_{k=1}^{q} x_{ik}x_{jk} = \sum_{k=1}^{q} x_{jk}\sum_{i=1}^{n} x_{ik} = 0,$$

pour tout j.

En notant $T = \text{trace}(B) = \sum_i \quad b_{ii}$, on obtient les relations :

$$\sum_{i=1}^{n} d_{ij}^2 = T + nb_{jj},\ \sum_{j=1}^{n} d_{ij}^2 = T + nb_{ii},\ \sum_{j=1}^{n}\sum_{i=1}^{n} d_{ij}^2 = 2nT. \quad (3)$$

Diapo 13 — MDS classique : solution analytique

En combinant les équations précédentes, on obtient la solution unique :

$$b_{ij} = -\frac{1}{2}(d_{ij}^2 - d_{\cdot j}^2 - d_{i\cdot}^2 + d_{\cdot\cdot}^2)$$

ou encore sous forme matricielle :

$$B = -\frac{1}{2} C D^2 C$$

où C est la **matrice de centrage**.

La solution X est alors obtenue par **décomposition en valeurs propres** :

$$B = V\Lambda V'$$
$$X = \Lambda^{1/2} V'$$

Diapo 14 — Interprétation géométrique

L'espace dans lequel se trouve **X** est **l'espace propre** (*eigenspace*),
où la **première coordonnée** correspond à la **plus grande variation** —
cet espace est noté $\mathbb{R}^q$.

Si l'on souhaite **réduire la dimension** à $p \leq q$,
alors les **p premières lignes** de $X^{(p)}$ conservent **le mieux possible**
les distances d_{ij}, parmi toutes les réductions linéaires possibles de X en dimension p.

Ainsi :

$$X^{(p)} = \Lambda_p^{1/2} V_p'$$

où :

- Λ_p est la **sous-matrice** $p \times p$ des **p premières valeurs propres** de Λ,
- V_p contient les **p premières colonnes** de la matrice des vecteurs propres V.

Diapo 15 — Récapitulatif : MDS classique

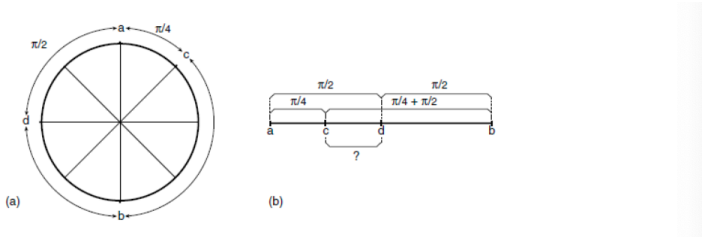
Le **cMDS** permet de :

- Donner des configurations $X^{(p)}$ dans $\mathbb{R}^p$ pour tout $p = 1,2,\ldots,q$
- Avoir des coordonnées **centrées**
- Ordonner les axes selon la **variance décroissante**
- Réduire la dimension (comme l'ACP)
- Obtenir une **solution exacte** si les distances sont euclidiennes
- Être utilisé même si les distances **ne le sont pas strictement**

Diapo 16 — Exemples : MDS classique

On considère plusieurs exemples :

1. Un **tétraèdre** de géométrie euclidienne (arêtes = 1)
2. Une **géométrie circulaire** (distance sur un cercle)



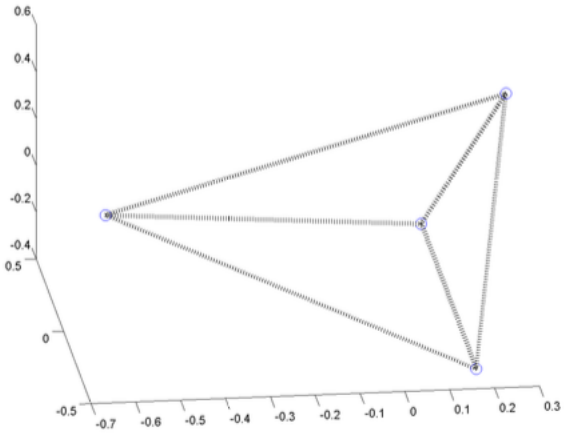
3. L'exemple des **distances aériennes** entre villes (Izenman §13.1.1)

Diapo 17 — Exemple : tétraèdre

Matrice de distances entre les 4 sommets d'un tétraèdre (toutes égales à 1) :

$$D = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}$$

Le calcul donne une **matrice de Gram** dont les valeurs propres sont : (0.5, 0.5, 0.5, 0).
→ En utilisant $p = 3$, on retrouve **parfaitement le tétraèdre**.



Diapo 18 — Exemple : distances circulaires

Matrice des distances par paires :

Point	a	b	c	d
a	0.0000	3.1416	0.7854	1.5708
b	3.1416	0.0000	2.3562	1.5708
c	0.7854	2.3562	0.0000	2.3562
d	1.5708	1.5708	2.3562	0.0000

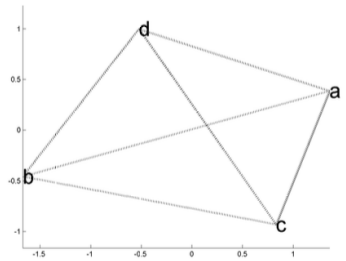
Matrice de distances correspondant à des points sur un cercle.
Les valeurs propres de la matrice de Gram $B_{(4\times4)}$ sont :

$$(5.6117, -1.2039, -0.0000, 2.2234)$$

On ne peut pas prendre la racine carrée des valeurs **négatives**, donc on garde uniquement les **valeurs propres positives**→ approximation de la géométrie circulaire par une géométrie euclidienne.

Diapo 19 — Exemple : distances circulaires (suite)

En utilisant $p = 2$, on obtient une configuration 2D $X^{(2)}$.



On compare la matrice des distances réelles D et celle reconstituée $\widehat{D} = \| x_i - x_j \|^2$.
Les valeurs sont très proches → la reconstruction est fidèle.

$$\begin{pmatrix} 0 & 3.1416 & 0.7854 & 1.5708 \\ 3.1416 & 0 & 2.3562 & 1.5708 \\ 0.7854 & 2.3562 & 0 & 2.3562 \\ 1.5708 & 1.5708 & 2.3562 & 0 \end{pmatrix}, \quad \hat{D} = \begin{pmatrix} 0 & 3.1489 & 1.4218 & 1.9784 \\ 3.1489 & 0 & 2.5482 & 1.8557 \\ 1.4218 & 2.5482 & 0 & 2.3563 \\ 1.9784 & 1.8557 & 2.3563 & 0 \end{pmatrix}$$

Diapo 20 — Exemple : distances aériennes

Exemple d'application du cMDS aux **distances aériennes** entre grandes villes américaines.

TABLE 13.2. *Airline distances (km) between 18 cities. Source: Atlas of the World, Revised 6th Edition, National Geographic Society, 1995, p. 131.*

	Beijing	Cape Town	Hong Kong	Honolulu	London	Melbourne
Cape Town	12947					
Hong Kong	1972	11867				
Honolulu	8171	18562	8945			
London	8160	9635	9646	11653		
Melbourne	9093	10338	7392	8862	16902	
Mexico	12478	13703	14155	6098	8947	13557
Montreal	10490	12744	12462	7915	5240	16730
Moscow	5809	10101	7158	11342	2506	14418
New Delhi	3788	9284	3770	11930	6724	10192
New York	11012	12551	12984	7996	5586	16671
Paris	8236	9307	9650	11988	341	16793
Rio de Janeiro	17325	6075	17710	13343	9254	13227
Rome	8144	8417	9300	12936	1434	15987
San Francisco	9524	16487	11121	3857	8640	12644
Singapore	4465	9671	2575	10824	10860	6050
Stockholm	6725	10334	8243	11059	1436	15593
Tokyo	2104	14737	2893	6208	9585	8159

	Mexico	Montreal	Moscow	New Delhi	New York	Paris
Montreal	3728					
Moscow	10740	7077				
New Delhi	14679	11286	4349			
New York	3362	533	7530	11779		
Paris	9213	5522	2492	6601	5851	

Diapo 21 — Exemple : distances aériennes (suite)

TABLE 13.6. *Eigenvalues of* B *and the eigenvectors corresponding to the first three largest eigenvalues (in red) for the airline distances example.*

	Eigenvalues	Eigenvectors		
1	471582511	0.245	−0.072	0.183
2	316824787	0.003	0.502	-0.347
3	253943687	0.323	−0.017	0.103
4	−98466163	0.044	−0.487	-0.080
5	−74912121	−0.145	0.144	0.205
6	−47505097	0.366	−0.128	−0.569
7	31736348	−0.281	−0.275	−0.174
8	−7508328	−0.272	−0.115	0.094
9	4338497	−0.010	0.134	0.202
10	1747583	0.209	0.195	0.110
11	−1498641	−0.292	−0.117	0.061
12	145113	−0.141	0.163	0.196
13	−102966	−0.364	0.172	−0.473
14	60477	−0.104	0.220	0.163
15	−6334	−0.140	−0.356	−0.009
16	−1362	0.375	0.139	−0.054
17	100	−0.074	0.112	0.215
18	0	0.260	−0.214	0.173

Les distances aériennes ne sont **pas strictement euclidiennes**.
On retient les **3 plus grandes valeurs propres** (à l'aide du *scree plot*).
→ Représentation tridimensionnelle approximative.

Diapo 22-23 — Exemple : distances aériennes (visualisation)

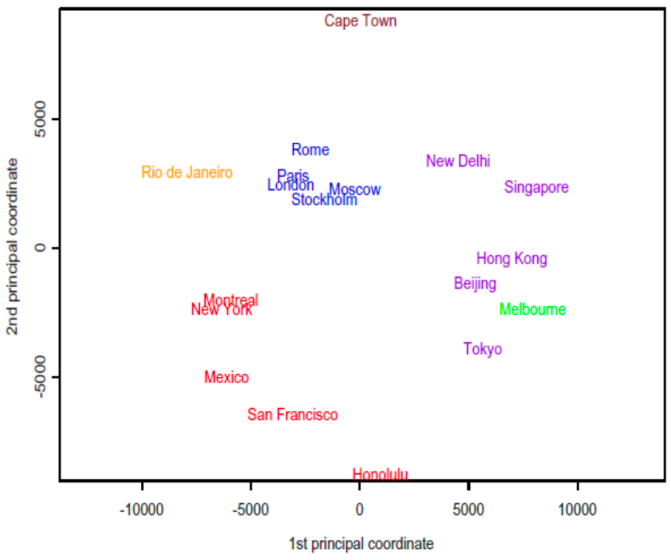


FIGURE 13.1. *Two-dimensional map of 18 world cities using the classical scaling algorithm on airline distances between those cities. The colors*

Visualisation des villes dans un plan ou espace 3D :
la carte reconstituée ressemble à une **carte géographique déformée** des États-Unis.
Les distances sont respectées dans la mesure du possible.

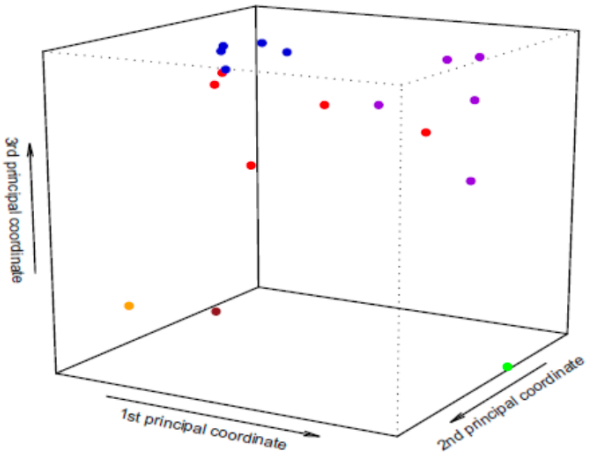


FIGURE 13.2. *Three-dimensional map of 18 world cities using the classical scaling algorithm on airline distances between those cities. The colors reflect the different continents: Asia (purple), North America (red), South America (yellow), Europe (blue), Africa (brown), and Australasia (green).*

Diapo 24 — Distance scaling (ou mise à l'échelle des distances)

Le **MDS classique (classical MDS)** cherche à trouver une **configuration optimale** des points x_i telle que :

$$d_{ij} \approx \hat{d}_{ij} = \| x_i - x_j \|_2$$

c'est-à-dire que les **distances observées** d_{ij} soient aussi proches que possible des **distances reconstruites** $\hat{d}_{ij}$.

Mise à l'échelle des distances (*Distance Scaling*)

On assouplit la contrainte $d_{ij} \approx \hat{d}_{ij}$ du MDS classique en autorisant une **transformation monotone** des distances :

$$\hat{d}_{ij} \approx f(d_{ij})$$

où f est une **fonction monotone croissante**.

Types de MDS selon la nature des dissimilarités :

- **MDS métrique** (*metric MDS*) : si les dissimilarités d_{ij} sont **quantitatives** (valeurs numériques).

- **MDS non métrique** (*non-metric MDS*) : si les dissimilarités d_{ij} sont **qualitatives ou ordinales** (par exemple : classement des similarités).

Différence avec le MDS classique :

Contrairement au **cMDS**, la **mise à l'échelle des distances** est un **processus d'optimisation** : on cherche à **minimiser une fonction de stress** (mesurant la différence entre distances réelles et reconstruites),
et la solution est obtenue par des **algorithmes itératifs** (numériques).

Diapo 25 — MDS métrique

Le **MDS métrique (classique)**
Étant donnée une dimension faible p et une fonction monotone f,
le MDS métrique cherche à trouver une configuration optimale $X \subset \mathbb{R}^p$ telle que :

$$f(d_{ij}) \approx \hat{d}_{ij} = \| x_i - x_j \|_2$$

aussi proche que possible.

• La fonction f peut être prise comme une fonction monotone paramétrique,
par exemple $f(d_{ij}) = \alpha + \beta d_{ij}$.

• « Aussi proche que possible » est maintenant défini explicitement par la **perte quadratique** :

$$\text{stress} = L(\hat{d}_{ij}) = \left(\frac{\sum_{i<j} (\hat{d}_{ij} - f(d_{ij}))^2}{\sum d_{ij}^2}\right)^{1/2}$$

et le MDS métrique minimise $L(\hat{d}_{ij})$ sur tous les $\hat{d}_{ij}$ et α, β.

• Le MDS métrique usuel est le cas particulier où $f(d_{ij}) = d_{ij}$;
la solution du MDS métrique (par optimisation) **n'est pas égale** à celle du MDS classique.

Diapo 26 — Cartographie de Sammon (Sammon Mapping)

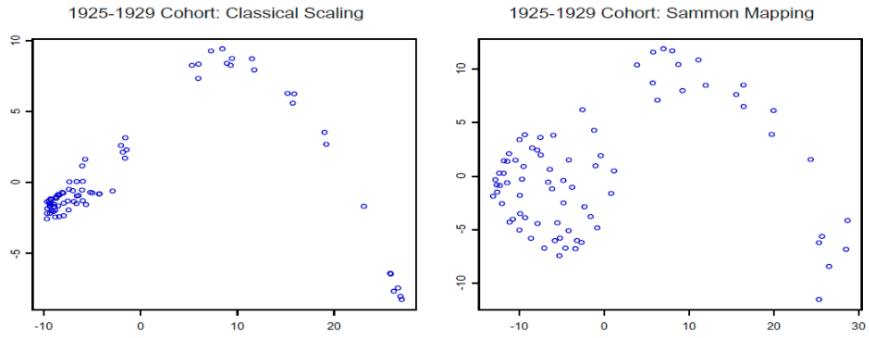
Une variante du MDS métrique.

Fonction de stress de Sammon :

$$\text{stress}_{\text{Sammon}} = \frac{1}{\sum_{l<k} d_{lk}} \sum_{i<j} \frac{(\hat{d}_{ij} - d_{ij})^2}{d_{ij}}$$

- Les **petites distances** ont plus de poids → meilleure préservation des **voisinages locaux**.
- La solution est trouvée **numériquement**, souvent en partant de la configuration cMDS.

Diapo 27 — Comparaison : cMDS vs Sammon Mapping



Résultats (Izenman, Fig. 13.9) :

- Le **cMDS** conserve bien les grandes distances.
- Le **Sammon Mapping** préserve mieux les **petites distances** (les objets proches restent proches).
  → C'est donc une méthode plus adaptée pour les **structures locales**.

Diapo 28 — MDS non métrique

Dans de nombreuses applications du MDS, les dissimilarités ne sont connues qu'à travers **leur ordre de classement**, et l'**écart** entre deux dissimilarités successives n'a **aucune importance** ou n'est **pas disponible**.

MDS non métrique

Étant donnée une dimension faible p, le MDS non métrique cherche à trouver une **configuration optimale** $X \subset \mathbb{R}^p$ telle que :

$$f(d_{ij}) \approx \hat{d}_{ij} = \| x_i - x_j \|_2$$

aussi proche que possible.

• Contrairement au MDS métrique, ici la fonction f est **beaucoup plus générale** et **n'est définie qu'implicitement**.
• Les valeurs $f(d_{ij}) = d^*_{ij}$ sont appelées **disparités**, et elles ne préservent que **l'ordre** des dissimilarités, c'est-à-dire :

$$d_{ij} < d_{k\ell} \quad \Leftrightarrow \quad f(d_{ij}) \leq f(d_{k\ell}) \quad \Leftrightarrow \quad d^*_{ij} \leq d^*_{k\ell}$$

Diapo 29 — MDS non métrique de Kruskal

Kruskal a proposé de minimiser :

$$\text{stress-1}(\hat{d}_{ij}, d^*_{ij}) = \left(\frac{\sum_{i<j} (\hat{d}_{ij} - d^*_{ij})^2}{\sum \hat{d}^2_{ij}}\right)^{1/2}$$

Les dissimilarités initiales ne servent qu'à comparer les **ordres** (pas les valeurs), $d_{ij} < d_{kl} < \ldots < d_{mf}$.
La fonction f agit comme une **courbe de régression monotone** entre dissimilarités et distances. (approximated dissimilarities d_{ij} as y , disparities d_{ij}^* as $\hat{}$ y , and the order of dissimilarities as explanatory)

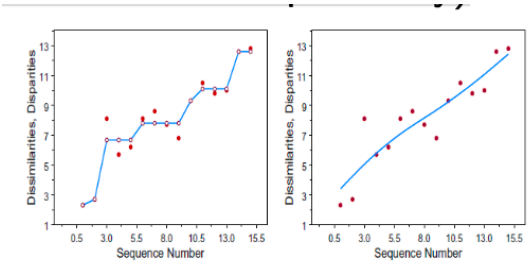


FIGURE 13.10. *Shepard diagram for the artificial example. Left panel: Isotonic regression. Right panel: Monotone spline. Horizontal axis is rank order. For the red points, the vertical axis is the dissimilarity d_{ij}, whereas for the fitted blue points, the vertical axis is the disparity $\hat{d}_{ij}$.*

Diapo 30 — Exemple : reconnaissance de lettres

Wolford et Hollingsworth (1974) se sont intéressés aux **erreurs de reconnaissance** commises lorsqu'une personne tente d'**identifier des lettres de l'alphabet** présentées pendant seulement quelques millisecondes.

Une **matrice de confusion** a été construite, indiquant la **fréquence** à laquelle chaque lettre présentée (**stimulus**) a été **confondue avec une autre**.

Une partie de cette matrice est présentée dans le tableau ci-dessous.

Letter	C	D	G	H	M	N	Q	W
C	–							
D	5	–						
G	12	2	–					
H	2	4	3	–				
M	2	3	2	19	–			
N	2	4	1	18	16	–		
Q	9	20	9	1	2	8	–	
W	1	5	2	5	18	13	4	–

Question : est-ce une matrice de dissimilarité ?

Diapo 31 — Construction des dissimilarités

Comment déduire les dissimilarités à partir d'une matrice de similarité ?
À partir des similarités δ_{ij}, on choisit une **valeur maximale de similarité** $c \geq \max(\delta_{ij})$, de sorte que :

$$d_{ij} = c - \delta_{ij} \text{si } i \neq j, \text{et } \; d_{ii} = 0.$$

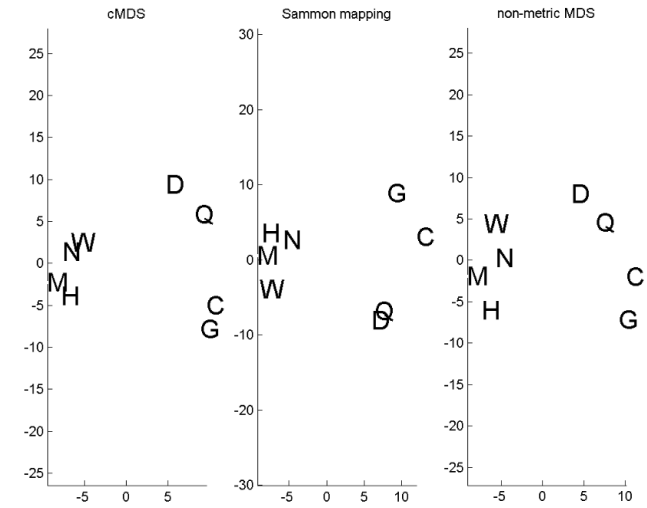
• **Quelle méthode est la plus appropriée ?**
Comme les dissimilarités d_{ij} ont été **déduites des similarités**,
leurs valeurs absolues dépendent du **choix arbitraire** de c.
C'est donc un cas où le **MDS non métrique** est le plus logique à utiliser.
Cependant, on verra que les **méthodes métriques** (MDS classique et *Sammon mapping*) peuvent également donner de bons résultats.

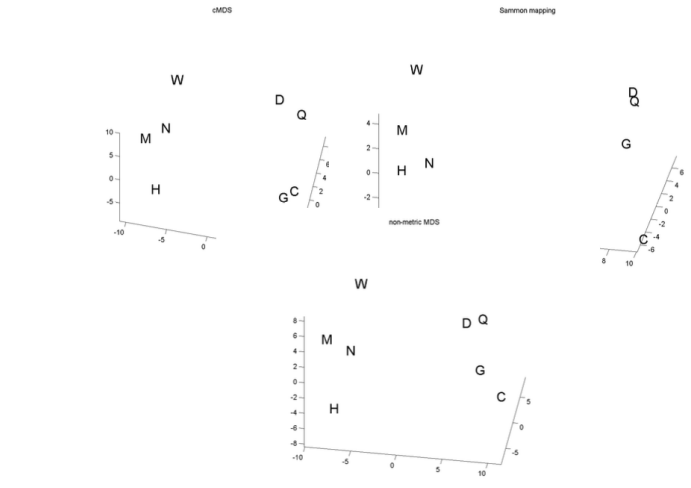
• **Combien de dimensions choisir ?**
En observant les **valeurs propres** obtenues à partir de la solution du **MDS classique (cMDS)**.

Diapo 32–33 — Exemple : lettres (c = 21)

On choisit $c = 21 = \max(\delta_{ij}) + 1$.
Comparaison des résultats du MDS (dimension 2 ou 3) avec :

- MDS classique (cMDS)
- Sammon Mapping
- MDS non métrique (stress-1)



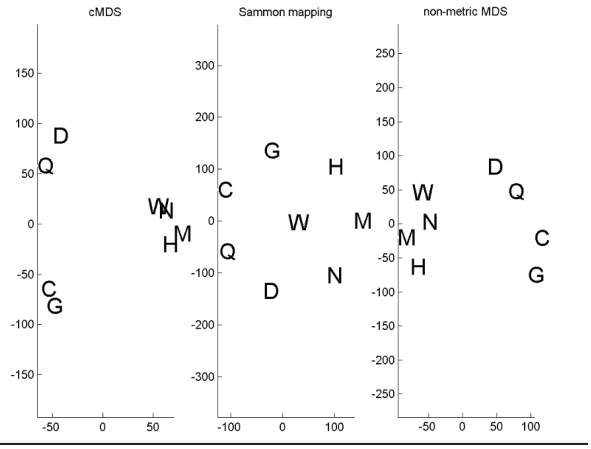


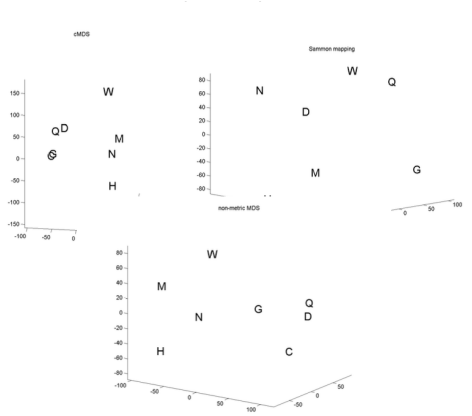
Diapo 34 — Résultats : valeurs propres

Pour $c = 21 = \max(\delta_{ij}) + 1.$, les valeurs propres le la gram matrice B dans le calcul du cMDS sont :
508.6, 236.1, 124.8, 56.1, 39.7, -0.0, -35.5, -97.2
→ Choisir $p = 2$ ou 3 semble raisonnable.

Diapo 35–36 — Exemple : lettres (c = 210)

Deuxième choix : $c = 210 = \max(\delta_{ij}) + 190$.
Même comparaison que précédemment (p = 2 et p = 3).





Diapo 37 — Résultats pour c = 210

Valeurs propres du cMDS (en ×10⁴) :
2.7210, 2.2978, 2.1084, 1.9623, 1.9133, 1.7696, 1.6842, 0.0000
→ Plus de dimensions nécessaires (p > 3).

Diapo 38 — Résumé : reconnaissance de lettres

- Données adaptées au **MDS non métrique**
- **Kruskal non-metric scaling** :
    1. Convient si seuls les **ordres** des dissimilarités sont fiables
    2. Sensible aux **minima locaux** (plusieurs solutions possibles)
    3. **Long à calculer**
- **cMDS** : rapide et globalement performant
- **Sammon Mapping** : échoue quand $c = 210$

Diapo 39 — Résumé (clusters de lettres)

Des **groupes de lettres** apparaissent :

- (C, G)
- (D, Q)
- (H, M, N, W)
  Confirmés par une **analyse de clustering hiérarchique** (liaison moyenne).