**Diapositive 1 — 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** ,  
  et la **distance naturelle** est .

(mettre une image de la sortie graph)

**Diapositive 2 — 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 3 — Exemple : perception des couleurs**

**(Prendre un premier exemple cf le pdf)**

**Diapositive 4 — Distance, dissimilarité et similarité**

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

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

Étant donnée une matrice de dissimilarités , le MDS cherche à trouver  
des points  tels que :

* S’il existe une configuration exacte, on parle de **distance euclidienne**.
* Mais parfois, il n’existe **aucune configuration** qui reproduise exactement .  
  → On parle alors de **distance non euclidienne**.

**Diapositive 5 — 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 .

(rajouter photo explicative)

* Le MDS essaie malgré tout de trouver une **configuration approchée** minimisant l’écart entre  
   et .

**Diapositive 6 — MDS classique : idée général**

**But :** retrouver les coordonnées des points à partir d’une matrice de distances (ou dissimilarités) supposées **euclidiennes**.

* On cherche une configuration  telle que
* Plutôt que de calculer  directement, on travaille avec la **matrice de produits scalaires** :
* Si  est centrée (points autour du centre de gravité), on peut exprimer  à partir des distances :

où  est la matrice de **centrage**.

**Diapo 7 — étapes de calcul**

1. **Données d’entrée :** la matrice des distances .
2. **Centrage double :** calcul de
3. **Décomposition en valeurs propres :**
4. **Coordonnées finales :**

( mettre un schéma : “Distances → Matrice B → Décomposition → Coordonnées dans R²”)

**Diapo 8 — réduction des dimensions**

Souvent, on ne garde que les **p premières composantes** correspondant aux plus grandes valeurs propres :

où :

* est la **sous-matrice**  des **p premières valeurs propres** de ,
* contient les **p premières colonnes** de la matrice des vecteurs propres .

→ les colonnes de  sont les **axes principaux** (comme en ACP)

→ On obtient une **carte en 2D ou 3D** qui préserve au mieux les distances.

**Interprétation :**

* Les **axes** représentent les directions de variation principale.
* Les **distances entre points** sur la carte reflètent les dissimilarités d’origine.

( image : nuage de points avec axes 1 et 2)

**Diapo 9 — Résumé**

 Conserve les **distances euclidiennes** autant que possible.

 Produit des **coordonnées centrées** et ordonnées par importance (variance).

 Permet une **représentation graphique** claire des relations entre objets.

 Proche de l’**ACP**, mais part directement d’une matrice de distances.

**Diapo 16 — Exemples : MDS classique**

(voir les 3 exemples sur le pdf)

**Diapo 17— Distance scaling (ou mise à l’échelle des distances)**

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

c’est-à-dire que les **distances observées**  soient aussi proches que possible des **distances reconstruites** .

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

On assouplit la contrainte  du MDS classique en autorisant une **transformation monotone** des distances :

où  est une **fonction monotone croissante**.

**Types de MDS selon la nature des dissimilarités :**

* **MDS métrique** (*metric MDS*) : si les dissimilarités  sont **quantitatives** (valeurs numériques).
* **MDS non métrique** (*non-metric MDS*) : si les dissimilarités  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 18 — MDS métrique**

Le **MDS métrique (classique)**  
Étant donnée une dimension faible  et une fonction monotone ,  
le MDS métrique cherche à trouver une configuration optimale  telle que :

aussi proche que possible.

• La fonction  peut être prise comme une fonction monotone paramétrique,  
par exemple .

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

et le MDS métrique minimise  sur tous les  et .

• Le MDS métrique usuel est le cas particulier où  ;  
la solution du MDS métrique (par optimisation) **n’est pas égale** à celle du MDS classique.

**Diapo 19 — Cartographie de Sammon (Sammon Mapping)**

Sammon mapping est la généralusation du MDS metric.

**Fonction de stress de Sammon (pour etre minimiser) :**

Ce système de pondération normalise les erreurs quadratiques dans les

distances par paires en utilisant la distance dans l'espace d'origine.

En conséquence, le Sammon mapping préserve les petits dij , leur accordant

une plus grande importance dans la procédure d'ajustement que pour les valeurs plus élevées de dij

* Les **petites distances** ont plus de poids → meilleure préservation des **voisinages locaux**.

• La solution optimale est trouvée par calcul numérique (valeur initiale

par cMDS).

**Diapo 20— Comparaison : cMDS vs Sammon Mapping**

Prendre l’exemple du pdf

**Diapo 21 — 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 , le MDS non métrique cherche à trouver une **configuration optimale**  telle que :

aussi proche que possible.

• Contrairement au MDS métrique, ici la fonction  est **beaucoup plus générale** et **n’est définie qu’impliciteme**nt.  
• Les valeurs  sont appelées **disparités**, et elles ne préservent que **l’ordre** des dissimilarités, c’est-à-dire :

**Diapo 22 — MDS non métrique de Kruskal**

Kruskal a proposé de minimiser :

Les dissimilarités initiales ne servent qu’à comparer les **ordres** (pas les valeurs), dij <dkl <...<dmf.  
La fonction  agit comme une **courbe de régression monotone** entre dissimilarités et distances. (approximated dissimilarities dij as y , disparities dij \* as ˆ y , and the order of dissimilarities as explanatory)

Prendre la photo du pdf

**Diapo 23— Exemple : reconnaissance de lettres**

Prendre un exemple cf pdf

**Diapo 24 — Construction des dissimilarités**

**Comment déduire les dissimilarités à partir d’une matrice de similarité ?**  
À partir des similarités , on choisit une **valeur maximale de similarité** ,  
de sorte que :

• **Quelle méthode est la plus appropriée ?**  
Comme les dissimilarités  ont été **déduites des similarités**,  
leurs valeurs absolues dépendent du **choix arbitraire** de .  
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)**

Prendre un exemple cf pdf

**Diapo 34 — Résultats : valeurs propres**

Prendre un exemple cf pdf

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

Prendre un exemple cf pdf

**Diapo 37 — Résultats pour c = 210**

Prendre un exemple cf pdf

**Diapo 38 — Résumé : reconnaissance de lettres**

Prendre un exemple cf pdf

**Diapo 39 — Résumé (clusters de lettres)**

Prendre un exemple cf pdf