Une image contenant texte, capture d’écran, carte de visite, Police

Le contenu généré par l’IA peut être incorrect.

**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** ,  
  et la **distance naturelle** est .

Une image contenant capture d’écran, texte, diagramme, Tracé

Le contenu généré par l’IA peut être incorrect.

**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  paires.
* On fait la **moyenne** des 31 notations pour chaque paire → cela donne une **mesure de similarité**.
* On la convertit ensuite en **dissimilarité** :  
  .

**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.

Une image contenant texte, capture d’écran, Police, document

Le contenu généré par l’IA peut être incorrect.

**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.

Une image contenant texte, diagramme, capture d’écran, nombre

Le contenu généré par l’IA peut être incorrect.

**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. si et seulement si
2. (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 , 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 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 .

Une image contenant diagramme, texte, Police, cercle

Le contenu généré par l’IA peut être incorrect.

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

**Diapositive 10 — MDS classique : théorie**

On suppose que la matrice de distances  est **euclidienne**.

L’objectif du **MDS classique (cMDS)** est de trouver une matrice de coordonnées  
 telle que :

Cette solution **n’est pas unique**, car un déplacement global , ,  
donne les mêmes distances.

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

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**  
 (pour un certain )  
telle que leurs distances mutuelles correspondent à celles de .

On calcule plutôt la **matrice de Gram**  plutôt que  directement.  
C’est une **matrice de produits scalaires** (puisque  est centrée).

On a la relation :

provenant de :

**Diapo 12 — MDS classique : théorie (suite 2)**

La contrainte de centrage () implique :

*Une image contenant texte, Police, blanc, typographie

Le contenu généré par l’IA peut être incorrect.*

pour tout .

En notant , on obtient les relations :

*Une image contenant texte, Police, blanc, écriture manuscrite

Le contenu généré par l’IA peut être incorrect.*

**Diapo 13 — MDS classique : solution analytique**

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

ou encore sous forme matricielle :

où  est la **matrice de centrage**.

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

**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 .

Si l’on souhaite **réduire la dimension** à ,  
alors les **p premières lignes** de  conservent **le mieux possible**  
les distances , parmi toutes les réductions linéaires possibles de  en dimension .

Ainsi :

où :

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

**Diapo 15 — Récapitulatif : MDS classique**

Le **cMDS** permet de :

* Donner des configurations  dans  pour tout
* 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)

Une image contenant diagramme, croquis, blanc, ligne

Le contenu généré par l’IA peut être incorrect.

1. 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) :

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

Une image contenant ligne, diagramme, triangle, pente

Le contenu généré par l’IA peut être incorrect.

**Diapo 18 — Exemple : distances circulaires**

Matrice des distances par paires :

Une image contenant texte, Police, capture d’écran, ligne

Le contenu généré par l’IA peut être incorrect.

Matrice de distances correspondant à des points sur un cercle.  
Les valeurs propres de la matrice de Gram B(4×4) sont :

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 , on obtient une configuration 2D .

Une image contenant ligne, diagramme, triangle

Le contenu généré par l’IA peut être incorrect.  
On compare la matrice des distances réelles  et celle reconstituée .  
Les valeurs sont très proches → la reconstruction est fidèle.

Une image contenant texte, Police, blanc

Le contenu généré par l’IA peut être incorrect.

**Diapo 20 — Exemple : distances aériennes**

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

Une image contenant texte, capture d’écran, Police, nombre

Le contenu généré par l’IA peut être incorrect.

**Diapo 21 — Exemple : distances aériennes (suite)**

Une image contenant texte, capture d’écran, menu, nombre

Le contenu généré par l’IA peut être incorrect.

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)**

**Une image contenant texte, capture d’écran, diagramme

Le contenu généré par l’IA peut être incorrect.**

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.

Une image contenant texte, diagramme, capture d’écran, ligne

Le contenu généré par l’IA peut être incorrect.

**Diapo 24 — 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 25 — 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 26 — Cartographie de Sammon (Sammon Mapping)**

Une variante du MDS métrique.

**Fonction de stress de Sammon :**

* 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**

**Une image contenant capture d’écran, ligne, diagramme

Le contenu généré par l’IA peut être incorrect.**

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 , 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 29 — 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)

Une image contenant texte, capture d’écran, diagramme, ligne

Le contenu généré par l’IA peut être incorrect.

**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.

Une image contenant texte, capture d’écran, nombre, Police

Le contenu généré par l’IA peut être incorrect.  
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 , 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)**

On choisit .  
Comparaison des résultats du MDS (dimension 2 ou 3) avec :

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

Une image contenant texte, diagramme, nombre

Le contenu généré par l’IA peut être incorrect.

Une image contenant diagramme, croquis, dessin, texte

Le contenu généré par l’IA peut être incorrect.

**Diapo 34 — Résultats : valeurs propres**

Pour ., 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  ou  semble raisonnable.

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

Deuxième choix : .  
Même comparaison que précédemment (p = 2 et p = 3).

Une image contenant texte, capture d’écran, diagramme, nombre

Le contenu généré par l’IA peut être incorrect.

Une image contenant diagramme, croquis, dessin, texte

Le contenu généré par l’IA peut être incorrect.

**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

**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).

Une image contenant diagramme, ligne, Plan, Dessin technique

Le contenu généré par l’IA peut être incorrect.