Université de Toulouse

Compte Rendu

TP2 : Analyse en composants principales (ACP) et analyse factorielle discriminante (AFD).

- Module : *Analyse statistique de données*
- Réalisé par :
  - KABOU Abdeldjalil
  - GHENAI Islem

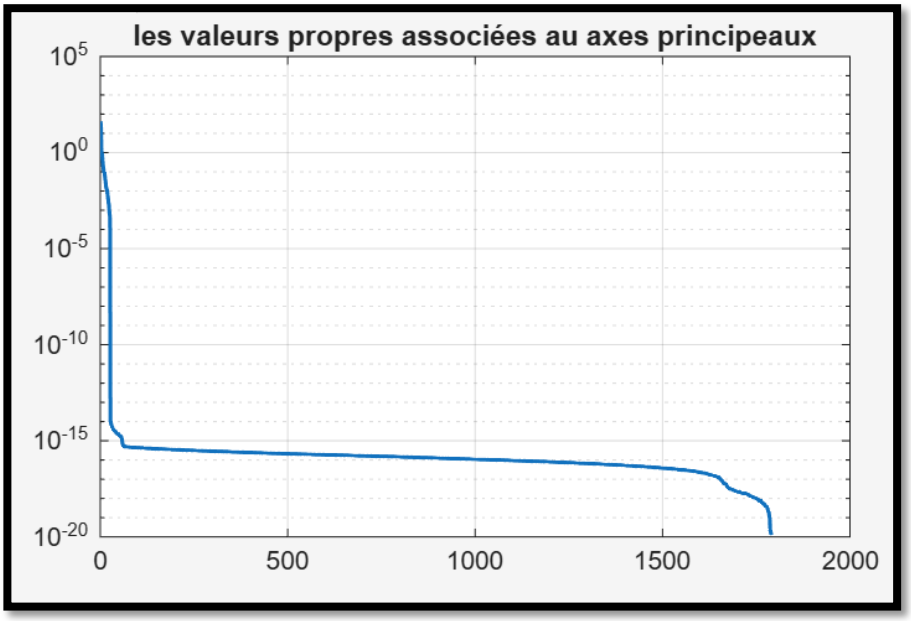
2025 / 2026

Sujet 2 : Analyse en composants principales (ACP) et analyse factorielle discriminante (AFD).

1) ACP : Première étude de la réflectance de matériaux

a. ACP avec métrique identité :

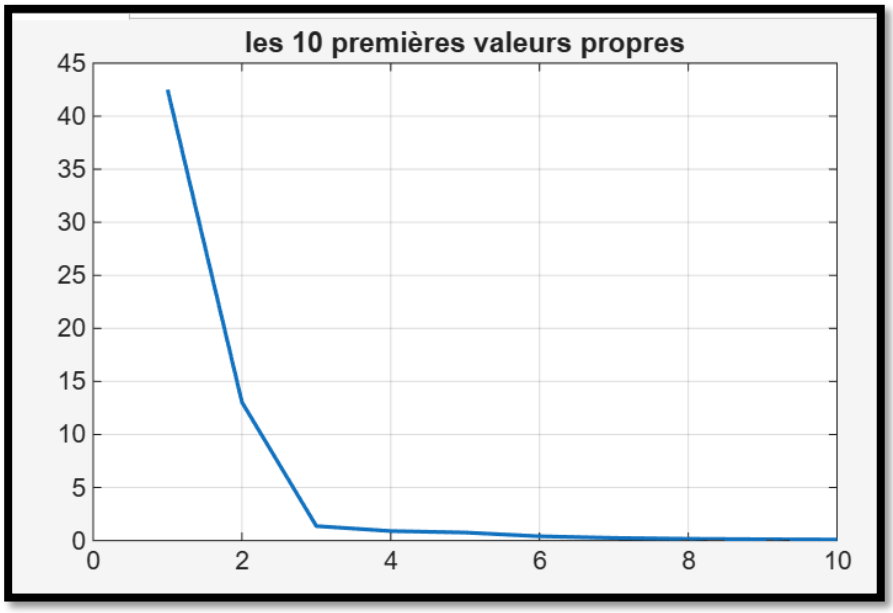
1) Représentation des valeurs propres associées au axes principaux en fonction des indices



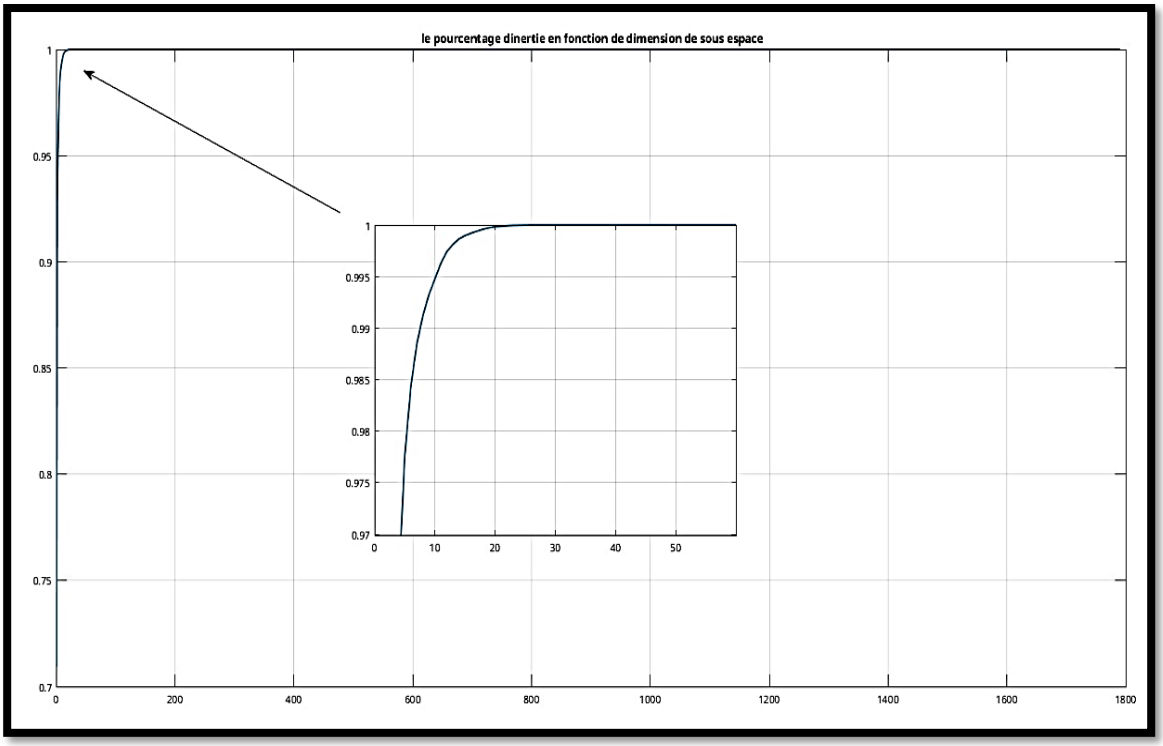
Commentaire : La décroissance rapide des valeurs propres indique que les données peuvent être bien représentées dans un sous-espace de faible dimension (quelques composantes principales suffisent), ce qui est typique dans les mesures spectrales (les longueurs d'onde sont fortement corrélées).

2) Tracer Les 10 premières valeurs propres en fonction des indices

Commentaire : La courbe montre une décroissance rapide : la 1ère composante principale explique une part importante de la variance. Après la 3eme ou 4eme valeur propre, les valeurs sont très faibles, ce qui indique que seules quelques composantes sont nécessaires pour représenter efficacement l'information contenue dans les données.

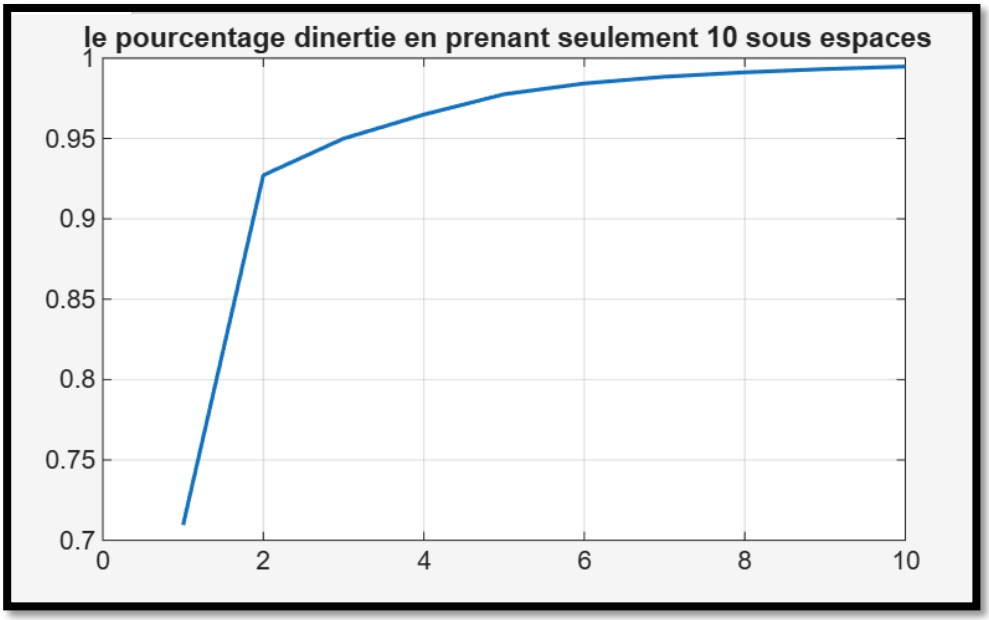


3) <u>Tracer Le pourcentage d'inertie expliquées en fonction de la dimension des sous espaces</u>



Commentaire : Le pourcentage d'inertie expliquée augmente rapidement avec les premières composantes principales. En utilisant un sous-espace de dimension réduite (≈ 5–10), on explique déjà plus de 99 % de la variance totale. Au-delà, l'apport supplémentaire est négligeable, ce qui confirme la forte redondance entre les variables et la possibilité de réduction de dimension efficace.

4) Tracer Le % d'inertie expliquées en prenant seulement les 10 premières sous espaces



On a trouvé le pourcentage d'inertie expliquée pour k = 1, 2, 3 :

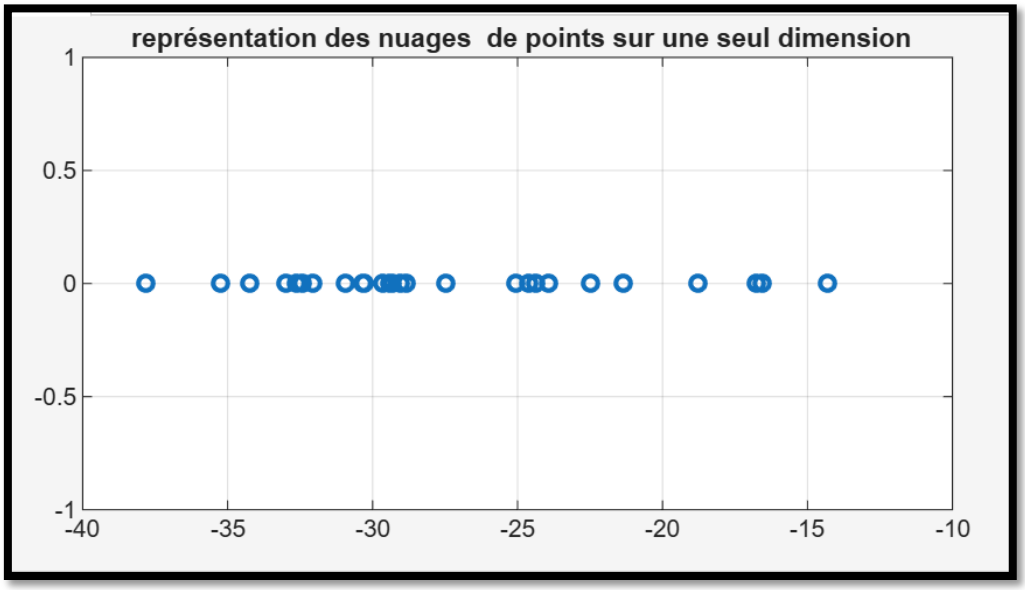
- k = 1 : ≈ 71 %
- k = 2 : ≈ 93 %
- k = 3 : ≈ 95 %

- il ne semble pas raisonnable de projeter sur un sous espace de dimension 1 , car en fait pour k = 1 on a un pourcentage d'inertie 71% et donc on perd presque 30 % de la variance. C'est trop pour une bonne représentation.
- *Mais si on prend un espace de dimension 2, c'est* très raisonnable : on garde >90 % de la variance et on obtient un plan facile à visualiser.
- en dimension 3 c'est légèrement mieux (~95 %), mais avec 3 dimensions la visualisation des données est plus complexe, le gain vs 2D est faible au regard de la complexité en plus (visualisation 3D moins pratique).
- *Et c'est pour cette raison on utilise la dimension 2 :*

- Excellent compromis variance conservée / lisibilité,
- Permet de la visualisation claire des individus/variables,

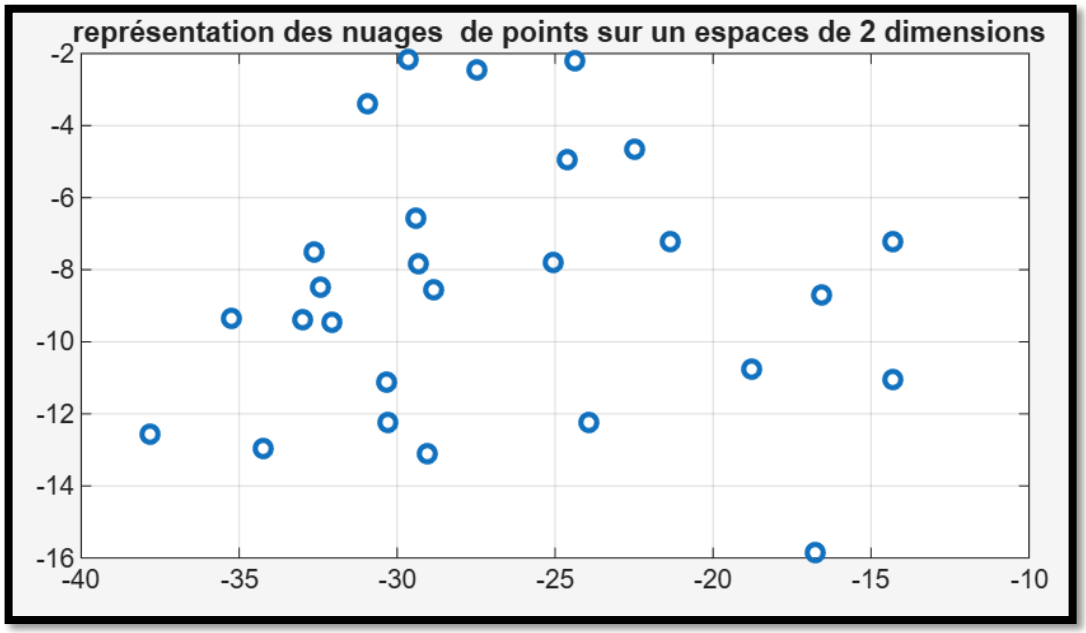
5) Représenter les nuages de points sur un sous ensemble de dimension 1 :

Commentaire : La projection en dimension 1 aligne les individus sur l'axe expliquant la plus grande variance. L'étalement des points le long de l'axe reflète l'inertie captée par la 1re composante. Ici, PC1 capture l'essentiel de l'information.



6) <u>Représenter les nuages de points sur un sous ensemble de dimension 2 :</u>

Commentaire : La projection sur les deux premières composantes conserve ~>90 % de la variance. La dispersion le long de PC1 est la plus forte ; PC2 apporte un complément d'information et permet une visualisation 2D lisible des individus.



7) <u>Les valeurs de la première composante principale :</u>

Sont données dans le fichier Matlab du code

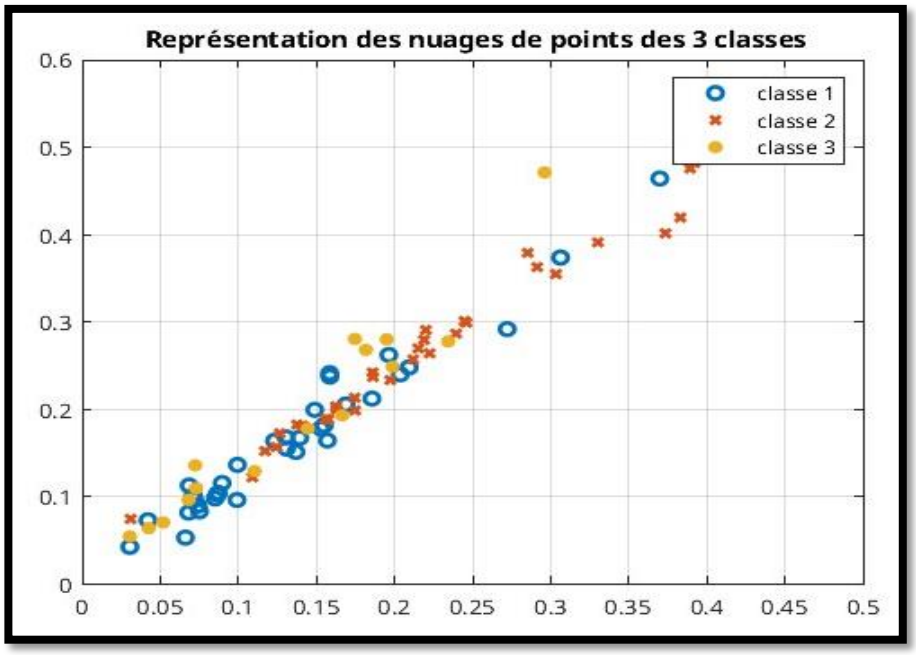
2) *ACP : Deuxième étude de la réflectance de matériaux*

a. Données considérées

Sont données dans le fichier Matlab du code

b. Représentation Partielles des données

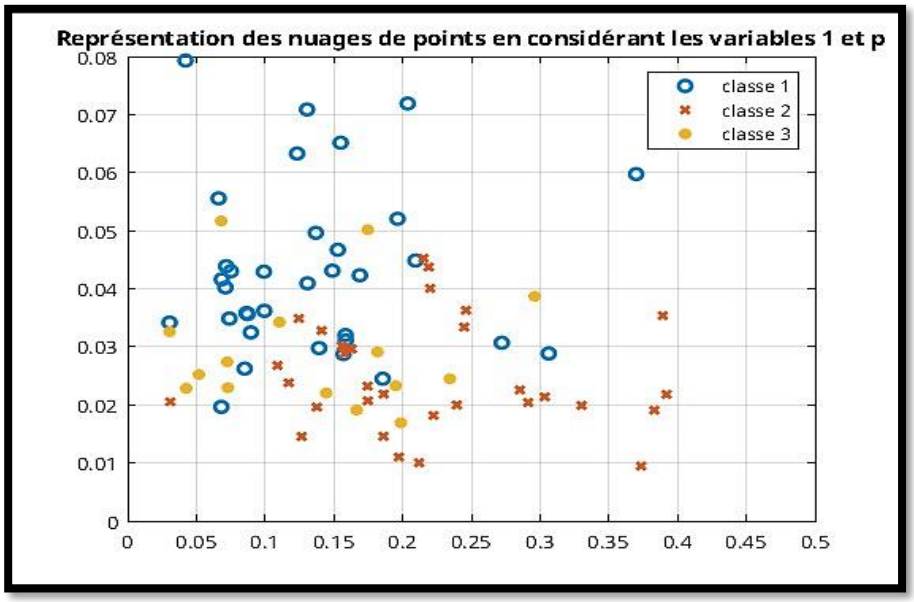
1. ***Représentation des nuages de points bidimensionnels en considérant les variables 1 et 2***



Commentaire : On observe une distribution globalement linéaire entre les variables. Les classes sont partiellement séparées : la classe 1 et la classe 2 occupent des zones légèrement différentes, tandis que la classe 3 est plus dispersée. Cette représentation donne une bonne vue d'ensemble mais ne permet pas une séparation nette des classes.

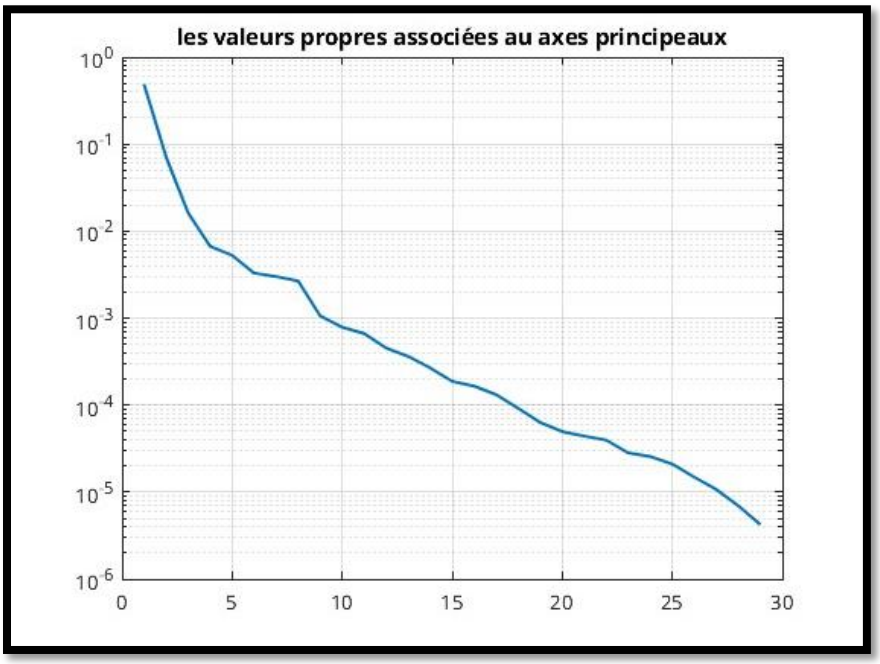
2. ***Représentation des nuages de points bidimensionnels en considérant les variables 1 et p :***

Commentaire : En considérant les variables d'indices 1 et p, les classes sont moins bien séparées et les points sont plus dispersés. Cela montre que la variable p est moins pertinent que la variable 2 pour représenter les données.



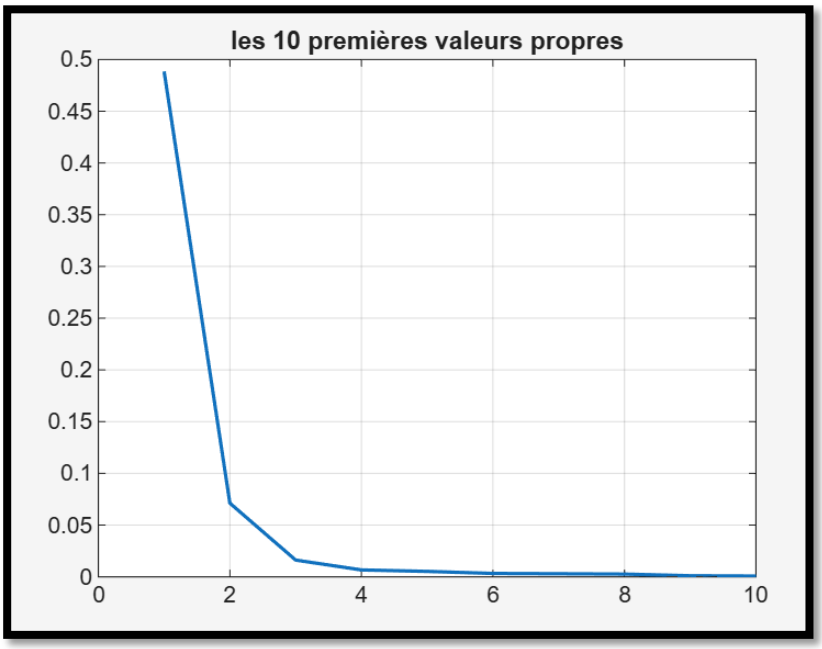
c. ACP avec métrique identité

1) Les valeurs propres associées au axes principaux



Commentaire : Les valeurs propres décroissent rapidement, ce qui montre que quelques axes principaux concentrent la majeure partie de l'information. Les dernières valeurs sont très faibles, indiquant qu'elles n'apportent que peu de variance supplémentaire. L'échelle logarithmique met bien en évidence cette décroissance.

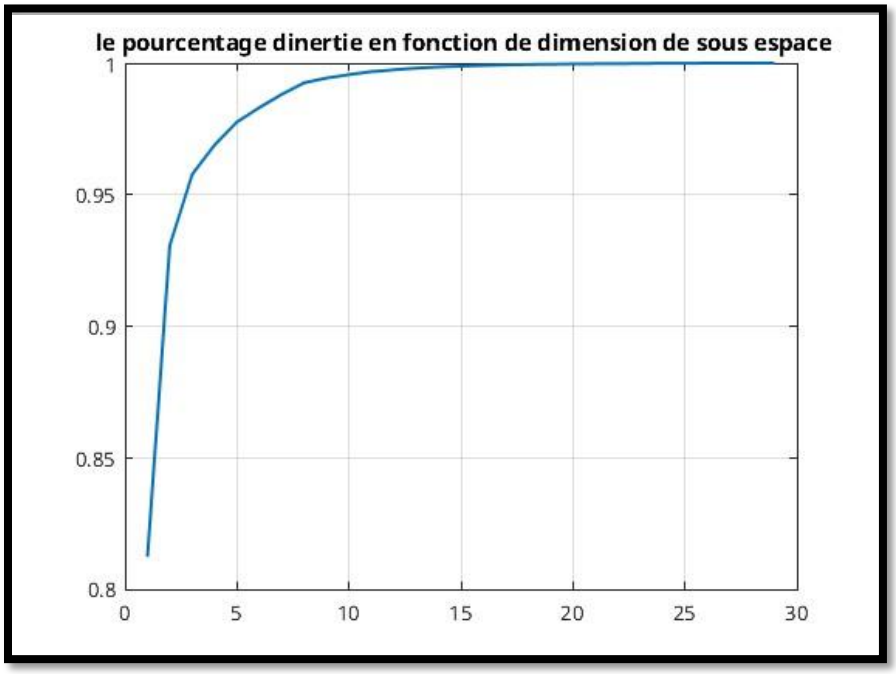
2) ***Représentation des 10 premières valeurs propres***



Commentaire : On observe que les deux premières valeurs propres sont largement dominantes, la première étant particulièrement élevée. À partir de la troisième, les valeurs deviennent très faibles. Cela indique que l'essentiel de l'inertie est concentré dans les deux premiers axes, ce qui justifie une réduction de dimension efficace avec très peu de perte d'information.

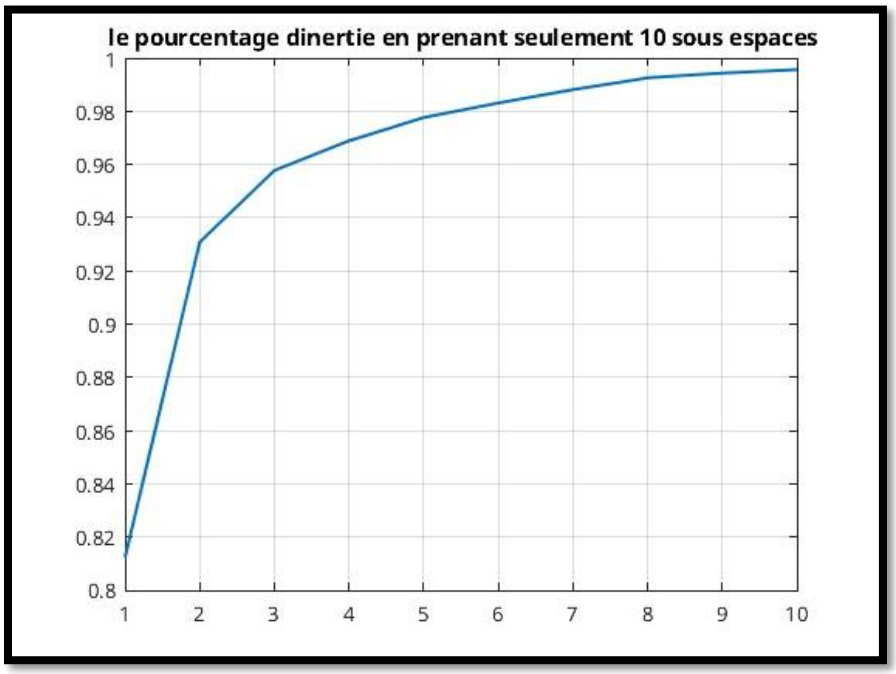
3) ***Pourcentage d'inertie en fonction de dimension de sous espace***

Commentaire : La courbe montre que le pourcentage d'inertie expliquée augmente rapidement au début, puis se stabilise progressivement. Cela indique que les premières composantes principales contiennent l'essentiel de la variance des données. À partir d'un certain nombre de dimensions (environ 8 à 10), l'ajout de nouvelles composantes n'apporte quasiment plus d'information supplémentaire.

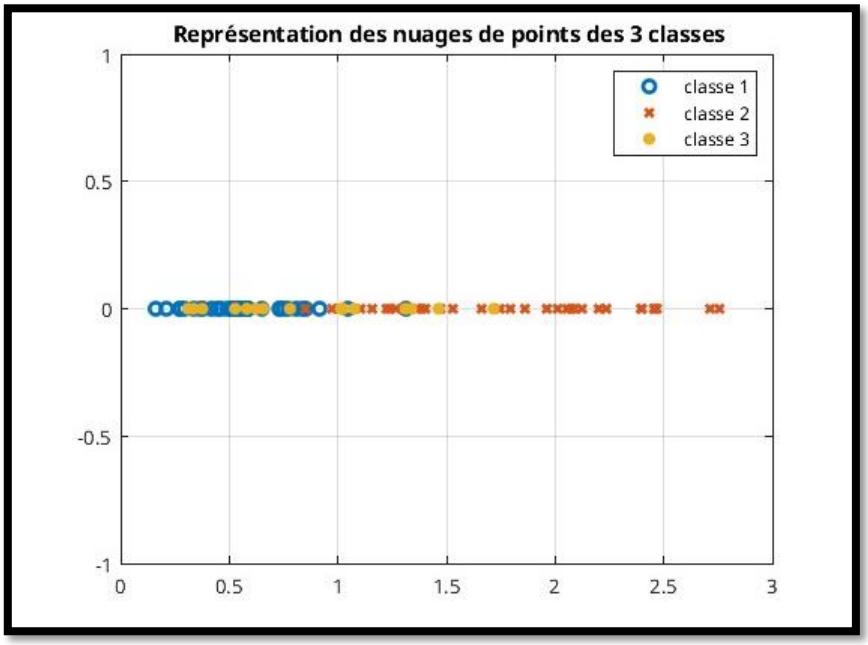


4) ***Pourcentage d'inertie expliquées en prenant seulement les 10 premières sous espaces***

Commentaire : Cette figure zoome sur les 10 premières composantes. On observe qu'avec seulement 2 ou 3 composantes, on explique déjà une très grande partie de la variance (>90 %). Cela confirme qu'il est possible de réduire fortement la dimension tout en conservant l'essentiel de l'information, ce qui est particulièrement utile pour la visualisation ou le prétraitement.



5) ***Représentation des nuages de points projetées sur un sous espace de dimension 1 :***



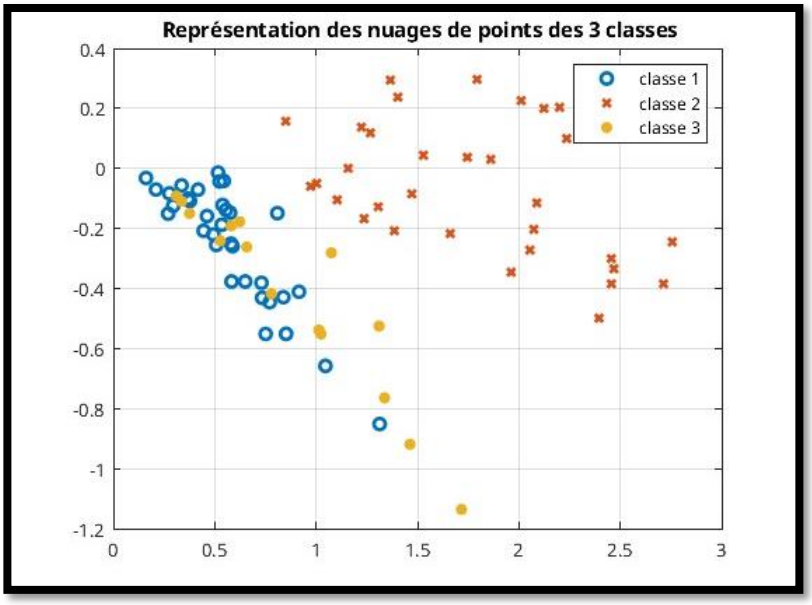
Commentaire : On observe que les points sont alignés sur une seule dimension (axe horizontal), ce qui est normal puisque la projection se fait sur une seule composante. On remarque que les trois classes sont relativement regroupées, mais il n'y a pas une séparation nette entre elles : les points des différentes classes se chevauchent partiellement. Cela indique que la première composante explique une grande partie de la variance, mais elle ne suffit pas à elle seule à séparer parfaitement les classes.

6) ***Représentation des nuages de points projetées sur un sous espace de dimension 2 :***

Commentaire : On observe que la dispersion des points est beaucoup plus informative qu'en 1D :

- La classe 1 (cercles bleus) est regroupée dans une zone précise (en bas à gauche).
- La classe 2 (croix rouges) occupe une zone bien distincte, plutôt vers la droite.
- La classe 3 (points jaunes) se situe principalement au centre/bas, avec une répartition moins nette.

La projection sur deux dimensions permet de mieux visualiser la structure globale des données et les relations entre classes. On distingue déjà une meilleure séparation entre la classe 2 et les deux autres, même si les classes 1 et 3 restent partiellement superposées.

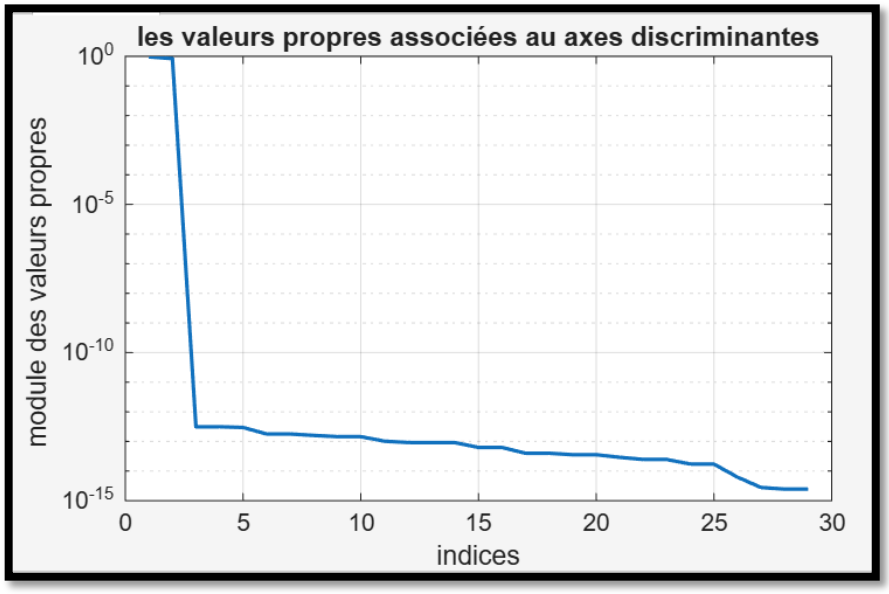


3) *AFD : Etude de la réflectance des matériaux :*

a. AFD

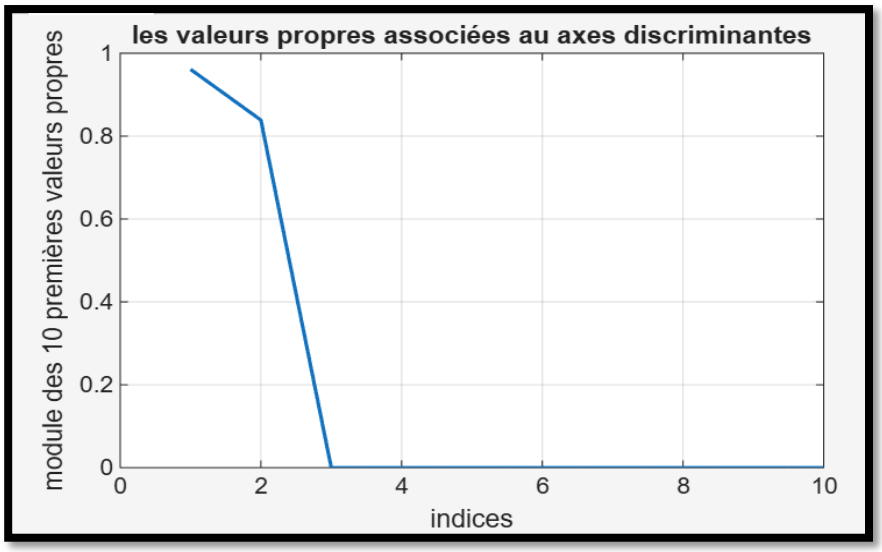
1) *Les valeurs propres associées aux axes discriminants*

Commentaire : La figure montre la décroissance rapide des valeurs propres associées aux axes discriminants. On observe que seules les premières valeurs sont significatives, tandis que les suivantes sont très proches de zéro (de l'ordre de 10^{-15}). Cela signifie que l'essentiel de l'information discriminante est contenu dans les premiers axes, et que les axes suivants n'apportent pratiquement aucune information supplémentaire utile à la discrimination entre classes.



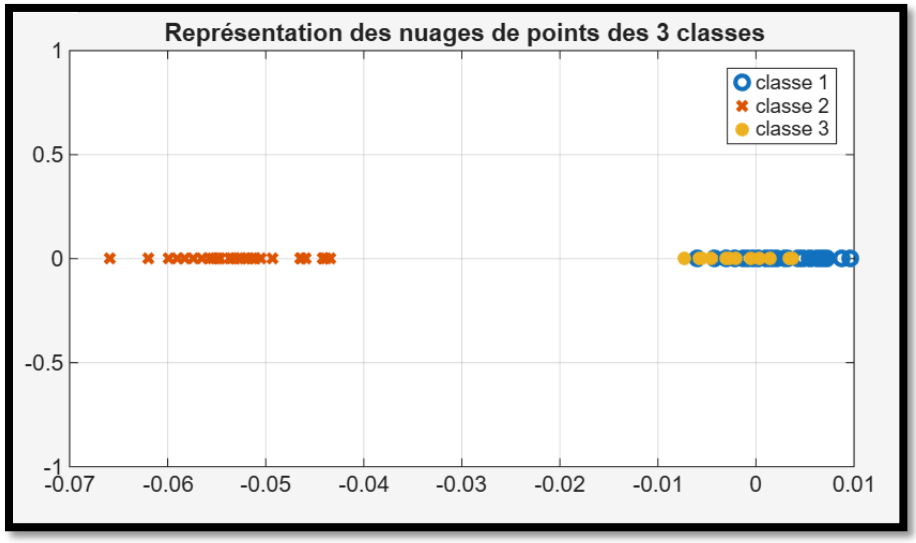
2) *Courbe des 10 premières valeurs propres discriminantes*

Commentaire : On observe que les deux premières valeurs propres sont nettement supérieures aux autres (≈ 0.95 et 0.85), cela indiquant que l'essentiel de l'information discriminante est concentré dans les deux premiers axes. À partir de la troisième valeur, les valeurs propres deviennent presque nulles, ce qui signifie que les axes suivants n'apportent quasiment aucune information discriminante supplémentaire. Cela suggère qu'une réduction de dimension à 2 axes discriminants serait suffisante pour bien séparer les classes, sans perte significative d'information.

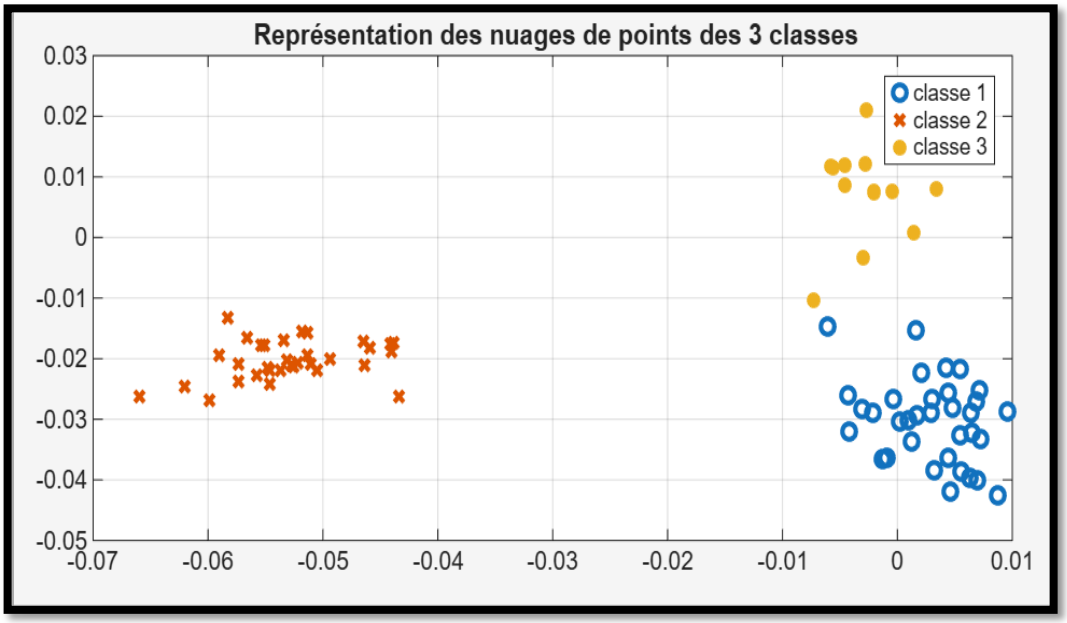


3) *Représentation du nuage non centré projeté sur un sous-espace de dimension 1*

Commentaire : La projection sur une dimension sépare bien la classe 2 des classes 1 et 3, mais ces deux dernières restent mélangées. Un seul axe ne suffit donc pas à bien distinguer toutes les classes.



4) ***Représentation du nuage non centré projeté sur un sous-espace de dimension 1***



Commentaire : On observe une séparation plus nette entre les trois classes lorsqu'on projette les données sur un sous-espace de dimension 2. Les groupes sont mieux individualisés qu'en dimension 1 : la classe 2 se distingue clairement, tandis que les classes 1 et 3 forment également des nuages séparés dans l'espace projeté. Cette représentation bidimensionnelle facilite ainsi la visualisation des structures discriminantes entre classes.