

**Rapport du Projet TP1 de Calcul
Scientifique/Analyse de données : Se familiariser
avec l'Analyse en Composantes Principales (ACP)**

Théo PETIT
Thomas SADURNI
Thibault ROUX

Département Sciences du Numérique - Première année
2019-2020

Table des matières

1	Visualiser les données	3
2	L'Analyse en Composantes Principales	5
3	L'ACP et la classification de données	5
4	L'ACP et la méthode de la puissance itérée	9

Table des figures

1	Représentation d'éléments de \mathbb{R}	3
2	Représentation d'éléments de \mathbb{R}^2	4
3	Représentation d'éléments de \mathbb{R}^3	4
4	Le même nuage de point avant et après ACP.	5
5	ACP sur une population séparée en 2 classes.	6
6	Projection sur les 3 premières composantes principales en 1 dimension.	6
7	Projection sur les 3 premières composantes principales en 3 dimensions.	7
8	Projection sur les 3 premières composantes principales.	7
9	Projection des données sur différentes CP	8
10	Projection sur les CP 1, 4 et 6.	9

1 Visualiser les données

Question 1 :

Les données sur lesquelles on a appliqué l'ACP dans le TP1 sont les 3 vecteurs de couleur RVB. Le tableau de données X correspond la matrice des données. Chaque colonne contient les données de chaque canal de couleur centrées.

Question 2 :

Chaque image étant de 264×400 , les données sont de $3 \times 264 \times 400$ donc de 3×105600 ($=316800$). Les figures 1, 2 et 3 montrent des représentations de nuages de points dans les dimensions 1, 2 et 3.

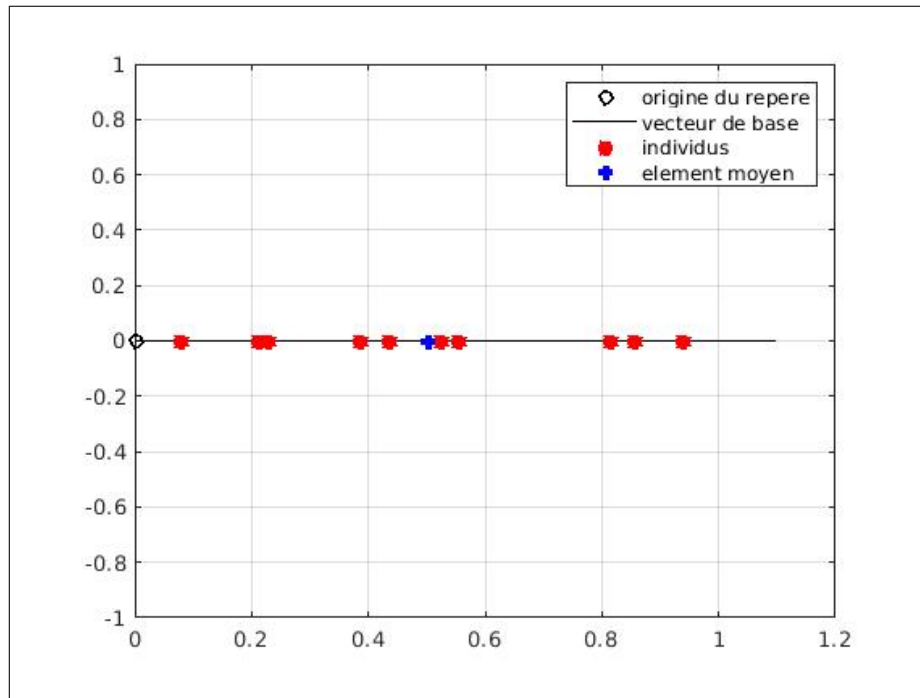


FIGURE 1 – Représentation d'éléments de \mathbb{R} .

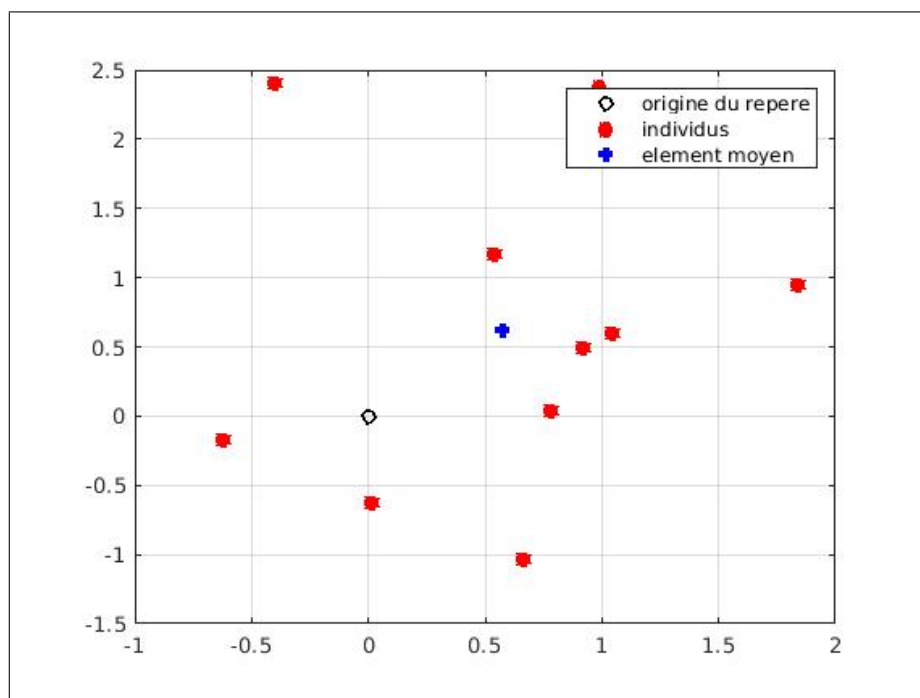


FIGURE 2 – Représentation d'éléments de \mathbb{R}^2 .

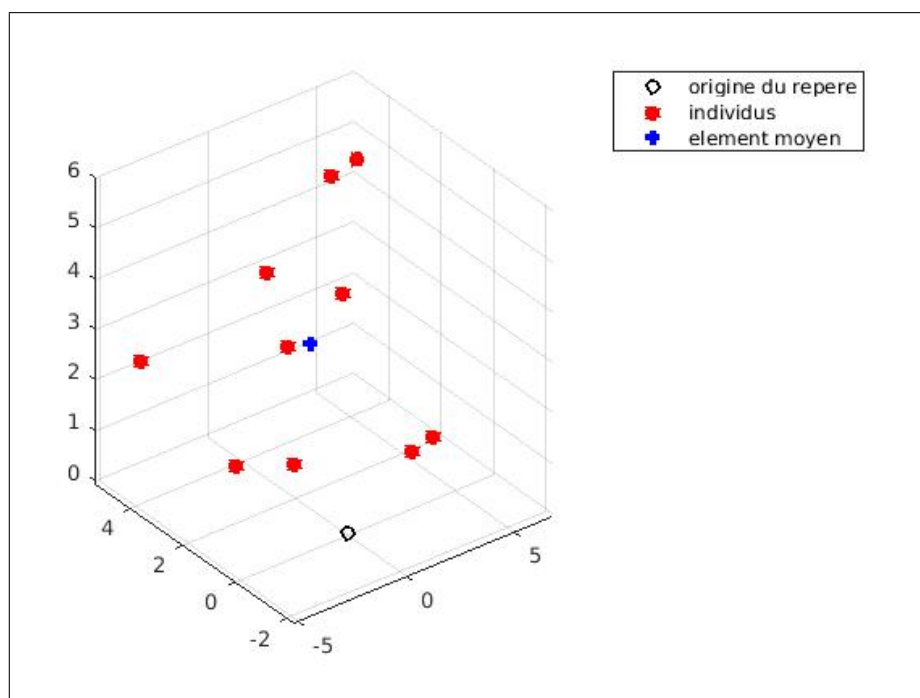


FIGURE 3 – Représentation d'éléments de \mathbb{R}^3 .

2 L'Analyse en Composantes Principales

Question 3 :

On observe en figure 4 deux nuages de points. Le nuage des points initiaux est bien plus dense, les points sont moins espacés, ce qui rend plus difficile leur regroupement en différentes catégories. Les points projetés sur les composantes principales sont bien plus espacés.

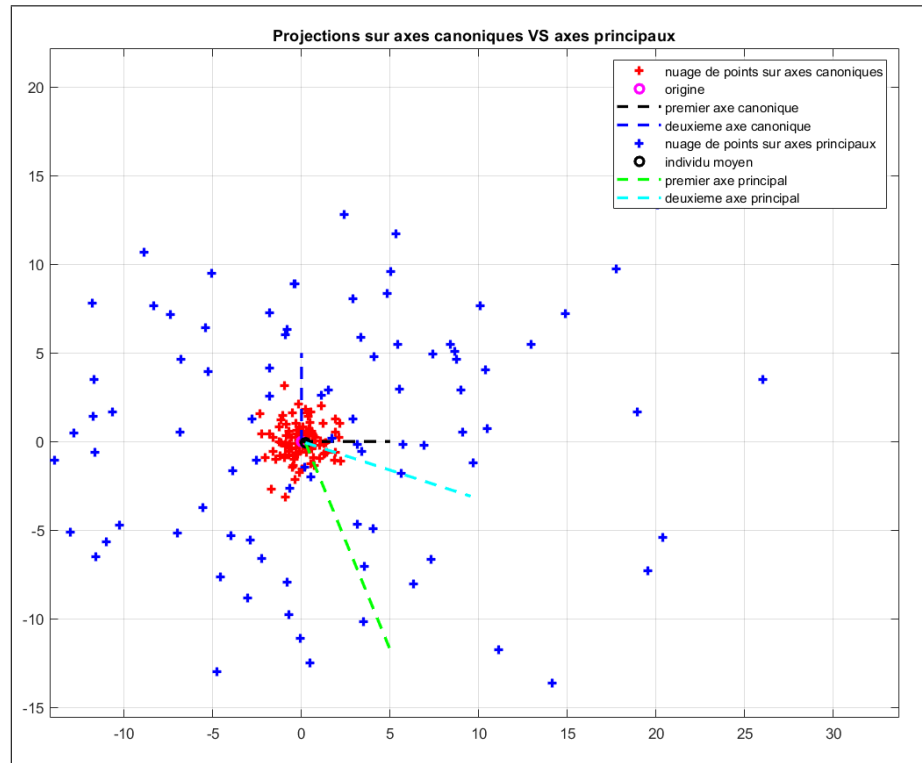


FIGURE 4 – Le même nuage de point avant et après ACP.

Question 4 :

Plus le coefficient sur la diagonale de Σ est important, plus il y a d'informations sur cette composante.

3 L'ACP et la classification de données

Question 5 :

Sur la figure 5, on peut observer deux classes distinctes en projetant sur la première composante principale. On peut donc en conclure que cette dernière concentre beaucoup d'informations.

En revanche, comme on remarque sur la figure 6, la première composante principale n'est pas toujours suffisante pour conclure. La figure 7 nous renseigne sur le fait que les 3 premières composantes principales contiennent la majorité de l'information (environ 25% chacune). Ces trois composantes principales sont suffisantes pour conclure, comme nous l'indique la figure 9.

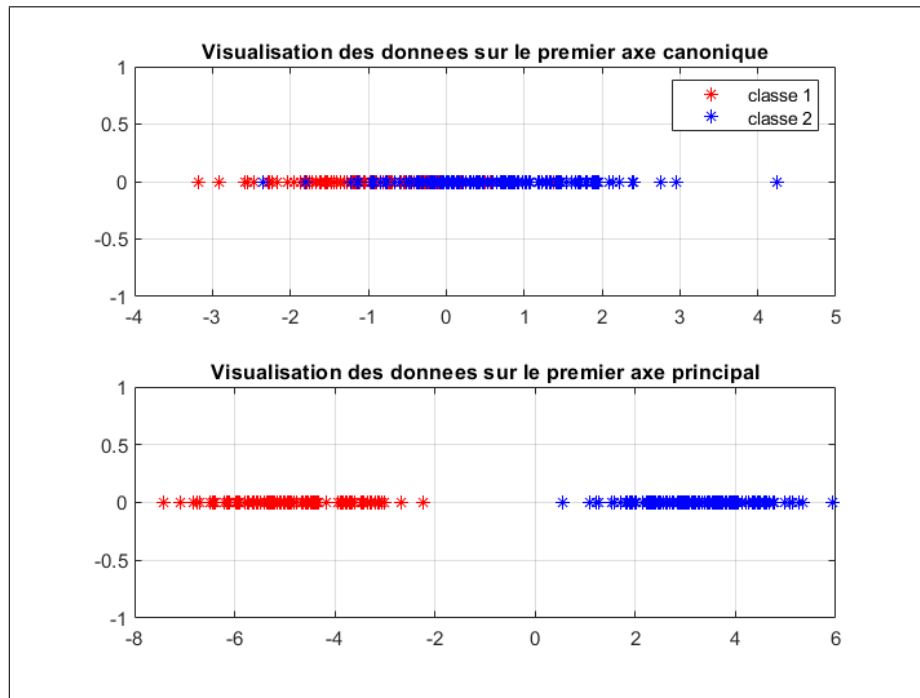


FIGURE 5 – ACP sur une population séparée en 2 classes.

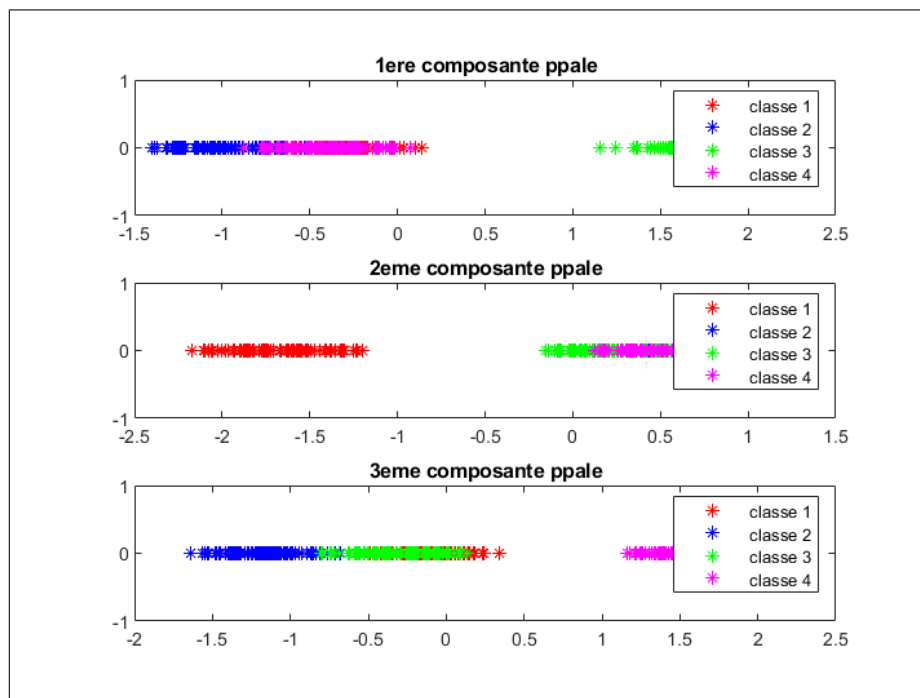


FIGURE 6 – Projection sur les 3 premières composantes principales en 1 dimension.

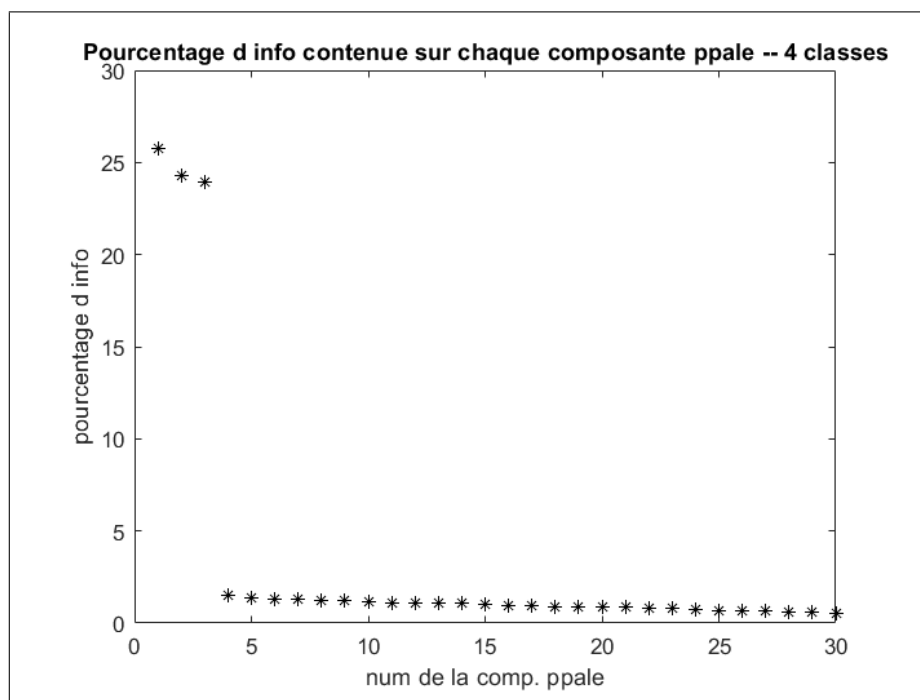


FIGURE 7 – Projection sur les 3 premières composantes principales en 3 dimensions.

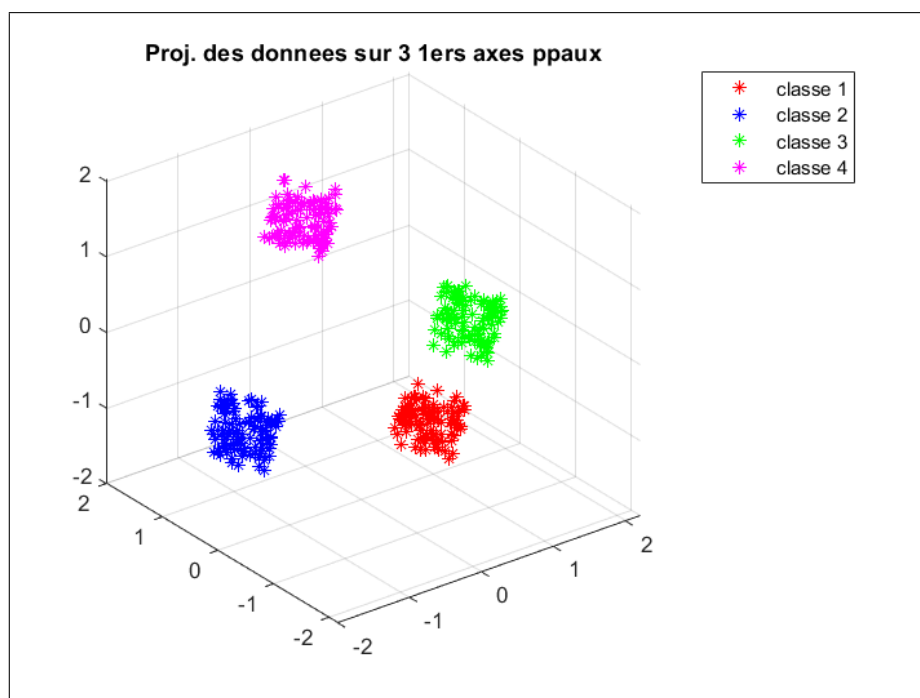


FIGURE 8 – Projection sur les 3 premières composantes principales.

Question 6 :

Avec une composante principale on observe 4 classes, en revanche, avec deux ou trois composantes principales on en observe entre 5 et 8. La figure 9 en représente quelques unes. La figure 10 représente la projection sur 3 des composantes principales regroupant de l'information. De plus, voici un tableau qui récapitule le nombre de classes observées suivant les 3 composantes principales utilisées lors de la projection :

CPs	1,2,3	1,2,4	1,2,5	1,2,6	1,3,4	1,3,5	1,3,6	1,4,5	1,4,6	1,5,6
Nb classes	6	7	7	7	7	7	6	7	8	7
	2,3,4	2,3,5	2,3,6	2,4,5	2,4,6	2,5,6	3,4,5	3,4,6	3,5,6	4,5,6
	6	5	7	6	7	7	7	8	6	7

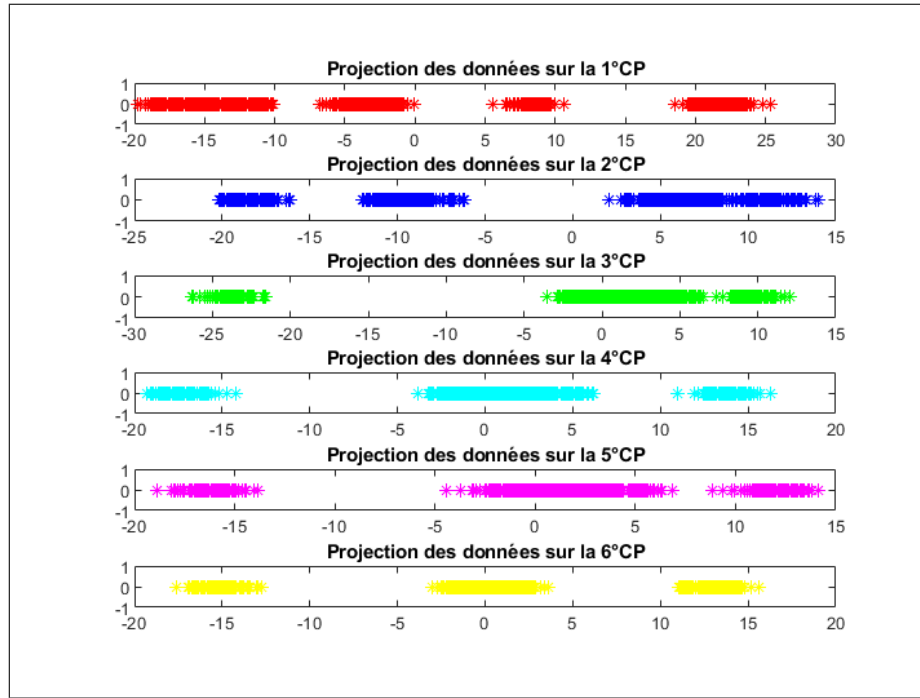


FIGURE 9 – Projection des données sur différentes CP

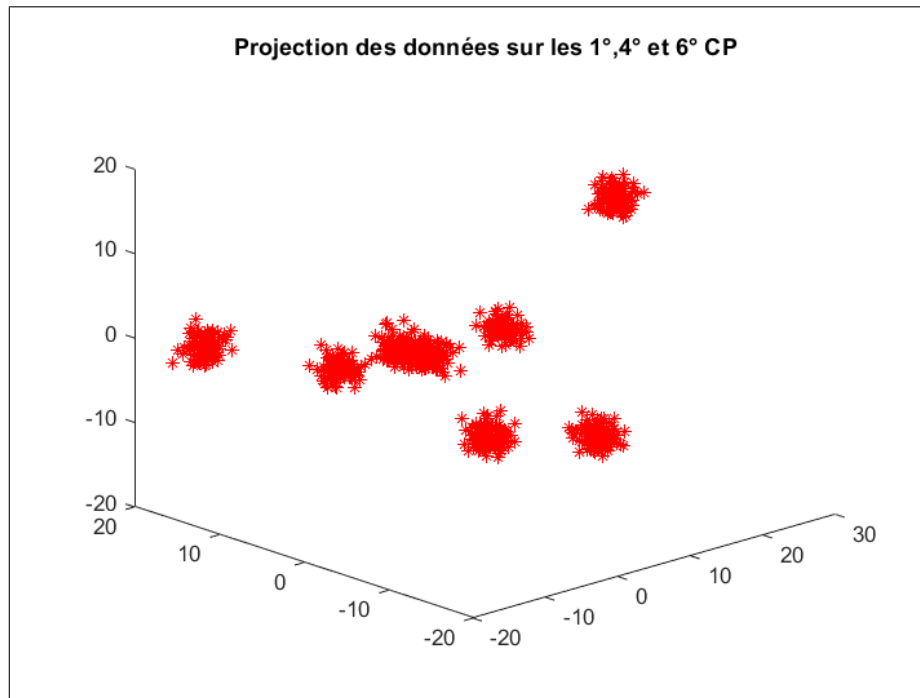


FIGURE 10 – Projection sur les CP 1, 4 et 6.

4 L'ACP et la méthode de la puissance itérée

Question 8 :

Soit $\lambda \in Sp(H^T H)$ et $X \vec{v}_p$

Alors $(H^T H X) = \lambda X$

$\Rightarrow (H H^T) H X = \lambda H X$

$\Rightarrow \lambda \in Sp(H H^T)$ et $H X$ est le \vec{v}_p associé.

Mêmes valeurs propres et on peut avoir les vecteurs propres en multipliant soit par H soit par H^T .

Question 9 :

Cf code Matlab.

Question 10 :

Il est plus utile et moins coûteux d'utiliser la méthode de la puissance itérée que la fonction eig. En effet, il est possible avec cette méthode de s'arrêter après avoir trouvé les premières plus grandes valeurs propres, seulement celles dont nous avons besoin pour réaliser l'ACP. Alors que la fonction eig va calculer toutes les valeurs propres, même celles qui ne nous seront pas utiles.

Question 11 :

Il faut d'abord comparer les dimensions de A (n et p), si $p > n$, il est préférable d'utiliser la méthode sur $A A^T$ puisqu'elle est carrée de taille n , sinon, il faudrait utiliser la méthode sur $A^T A$, carrée de taille p .