

Rapport Tp1 Calcul Scientifique - Analyse de Données

Mdaa Saad // El Bennouri Abdeslam // Dahhoumi Mouad

Département Sciences du Numérique - 1A
2019-2020

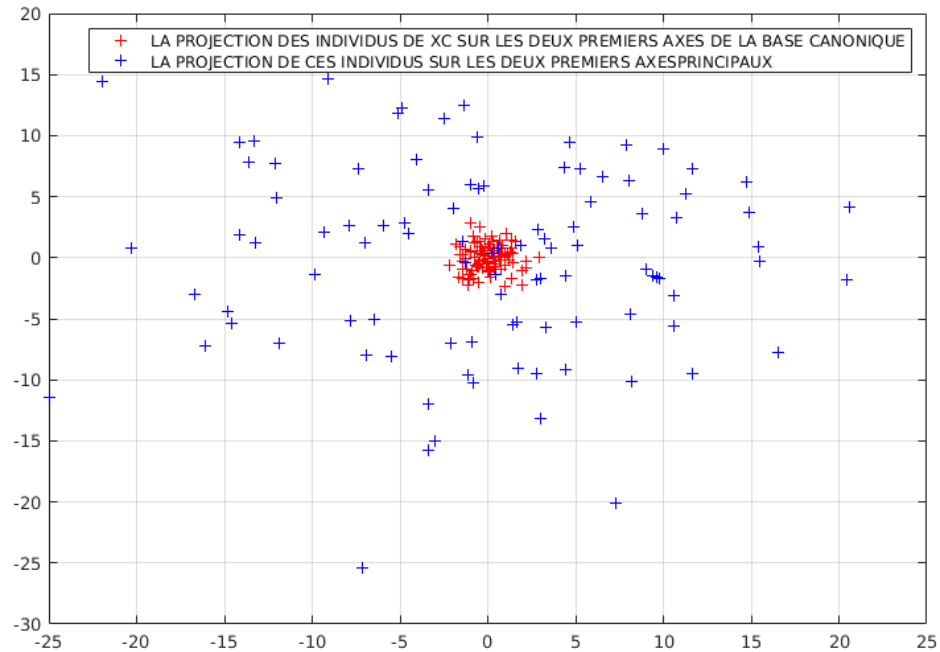
Question 1

Pendant le premier Tp analyse de Données on a eu $X = [R(:) \ V(:) \ B(:)]$. Donc X a 3 colonnes qui correspondent aux niveaux de vert, rouge, bleu par pixel. le tableau X est de taille $(71070, 3)$ puisque l'image est de taille $206 * 345 = 71070$ (ie l'image contient 71070 pixels)

Question 2

voir *visualisation.m*

Question 3



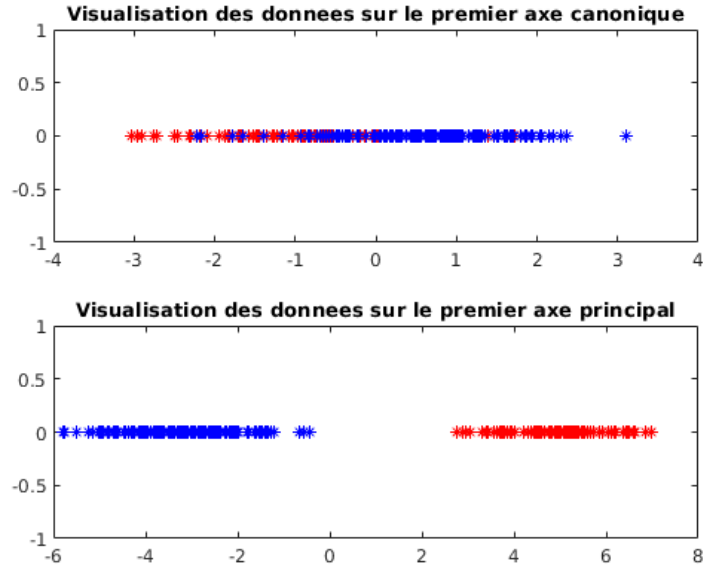
On remarque que la projection des données centrées sur l'axe principale nous permet de bien les distinguer (ie les données sont dispersées) contrairement à la projection des données centrées sur l'axe canonique qui est bien compacte donc moins claire.

Question 4

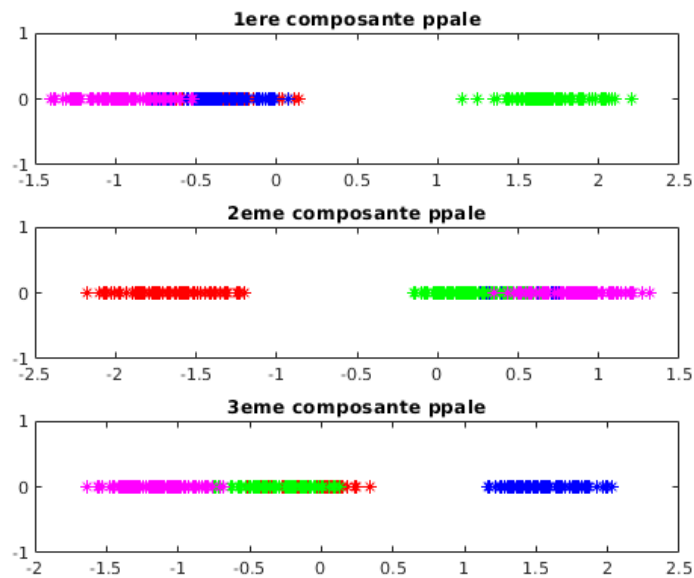
On peut quantifier l'information contenue dans les q premières composantes principales à partir de la matrice Σ en calculant la contraste conservé par les q Composantes Principales :

$$\frac{\sum_{j=1}^q \lambda_j}{\sum_{j=1}^p \lambda_j}$$

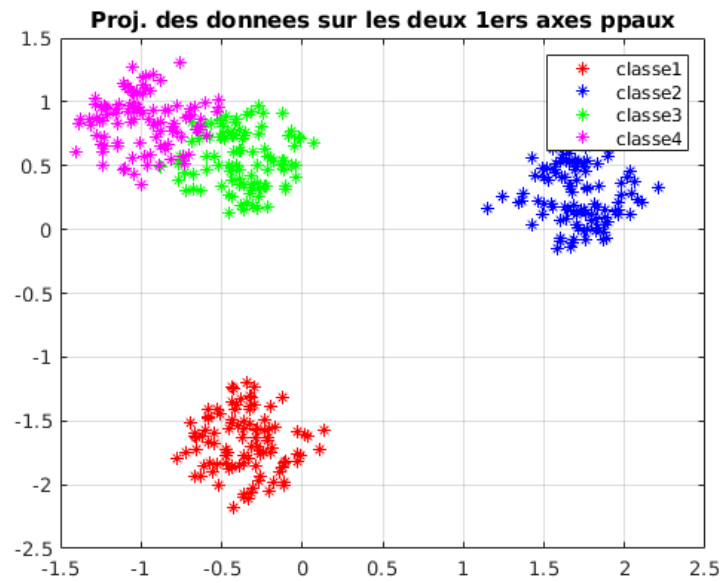
Question 5



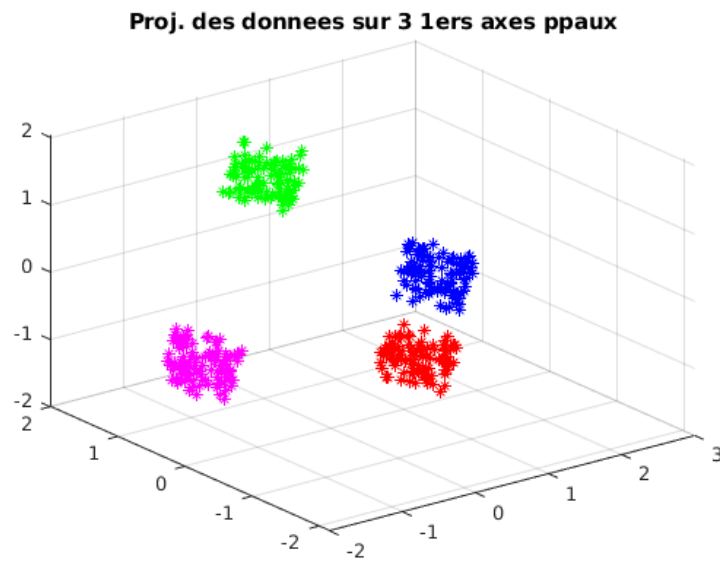
On remarque que la projection sur le premier axe principale nous permet de distinguer les 2 classes clairement contrairement à la projection sur l'axe canonique



Avec la 1er composante on est capable de constater 2 classes , avec la 2eme composante on est capable de constater 2 classes , avec la 3eme composante on est capable de constater 3 classes , et avec tous les trois on peut constater les 4 classes.

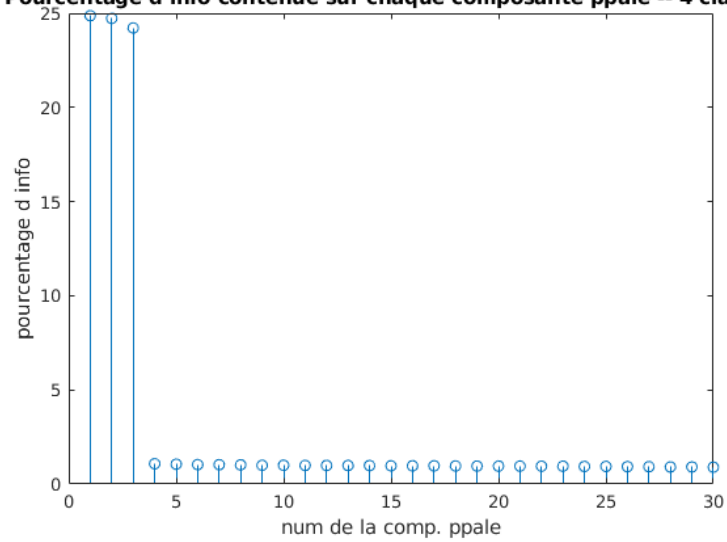


Dans le plan on peut distinguer les 4 classes mais la classe 3 et 4 sont un peu confondue



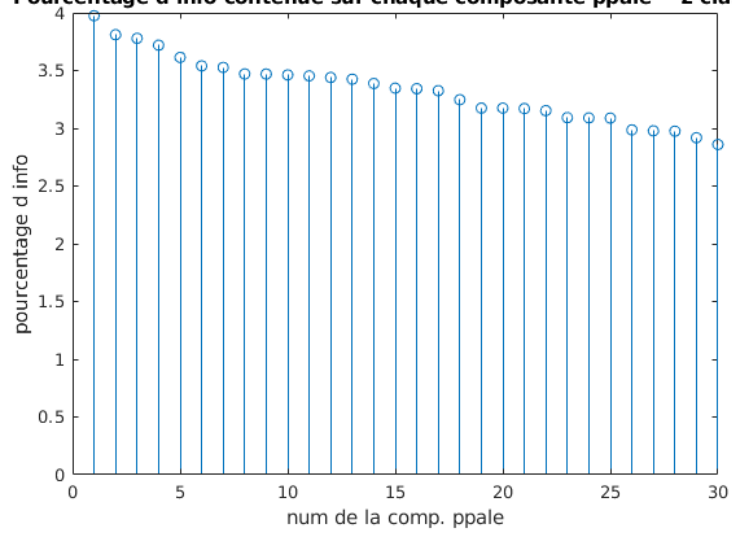
Dans l'espace on distingue clairement les 4 classes

Pourcentage d info contenue sur chaque composante ppale -- 4 classes

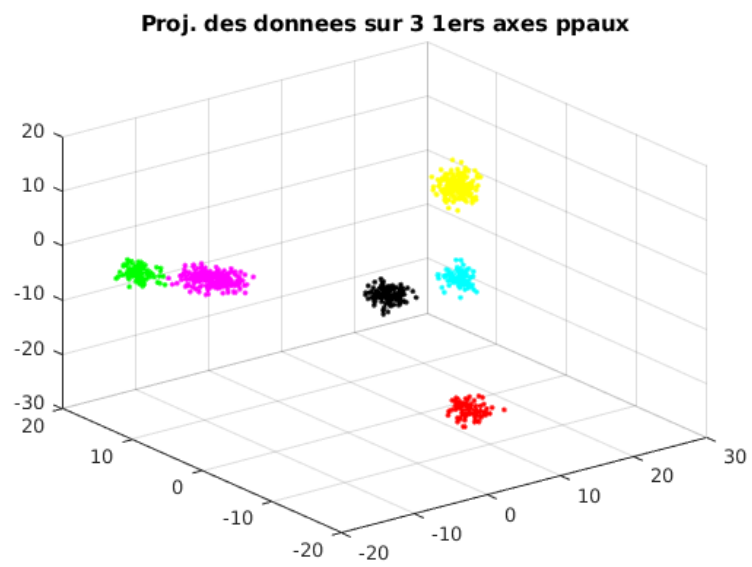
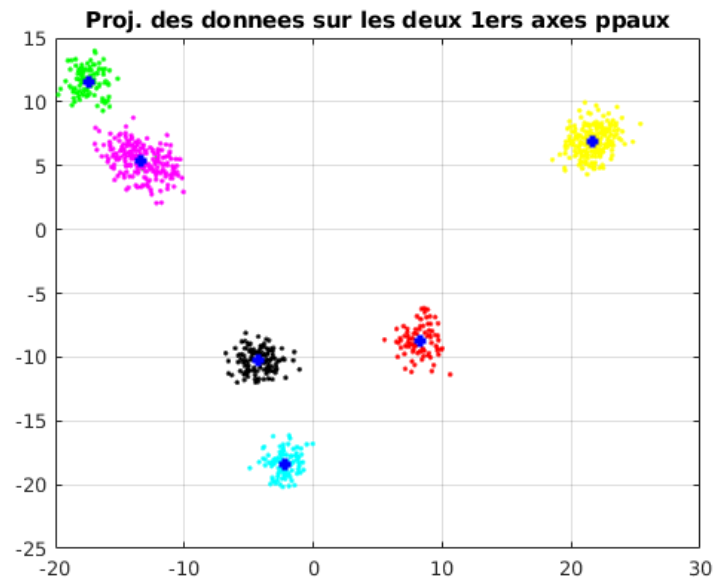


On constate que dans la donnée quatre classes la plus part de l'information est contenue dans les 3 premières composantes contrairement aux données deux classes

Pourcentage d info contenue sur chaque composante ppale -- 2 classes

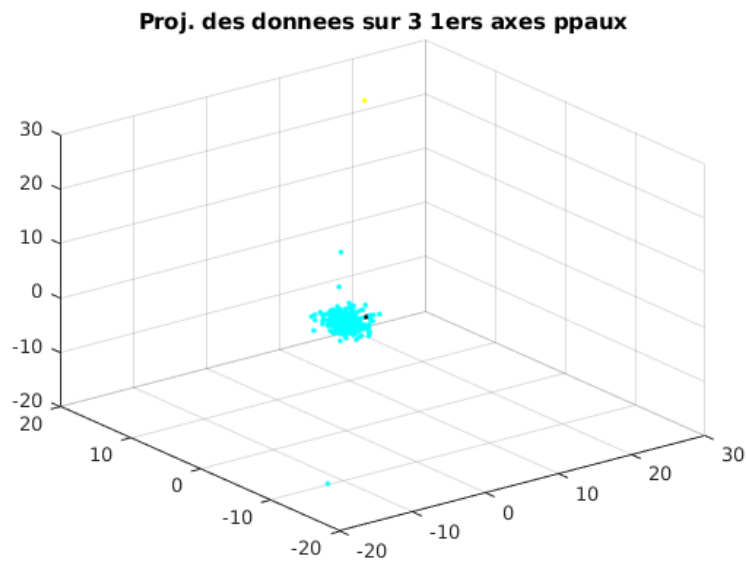
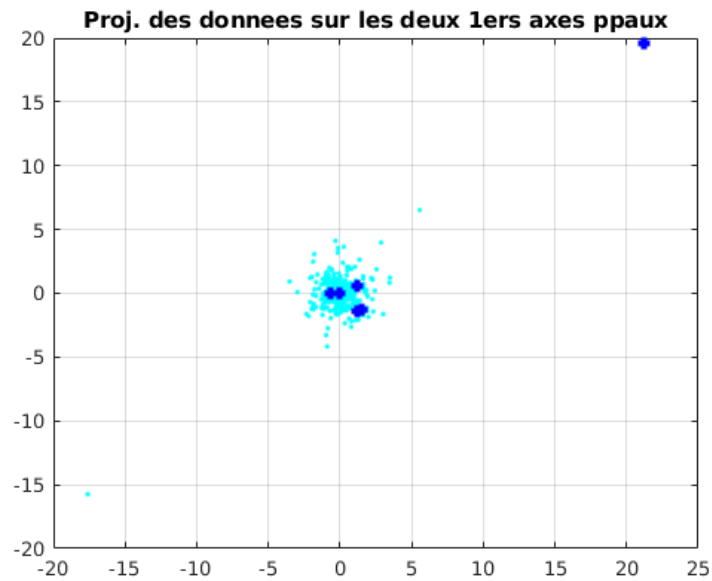


Question 6



Dans le jeu des données dataset.mat On peut distinguer 6 classes d'individus

Question 7



Dans le jeu des données dataset.mat On peut distinguer 1 classe de variables

Question 8

soit (λ, x) un couple propre de $H^T H$ donc : $H(H^T H).x = (HH^T)H.x = \lambda.H.x$
on pose $y = H.x$ donc

$$\boxed{HH^T.y = \lambda.y}$$

Donc connaître les éléments propres de $H^T H$ permet de connaître les éléments propres de HH^T

Question 9

```
Erreur relative pour la methode avec la grande matrice = 9.996e-09
Erreur relative pour la methode avec la petite matrice = 9.651e-09
Ecart relatif entre les deux valeurs propres trouvees = 5.47e-04
Temps pour une ite avec la grande matrice = 3.378e-03
Temps pour une ite avec la petite matrice = 1.274e-04
>>
```

Question 10

Lien avec l'ACP : à ce qui concerne l'ACP on cherche seulement à réduire les dimensions d'un espace, donc on a pas besoin de savoir toutes les valeurs propres mais seulement les q -premières v.p. selon l'ordre croissant.

Question 11

Si on choisit d'utiliser la méthode de la puissance itérée pour calculer les éléments propres de Σ on doit appliquer la méthode sur la matrice $Xc^T Xc$