Rapport du projet AD/CS : Partie 1
Application de l'ACP : les "Eigenfaces"
Méthode de la Puissance Itérée

EL ALOUT Ismail
KARMAOUI Oussama
YOUNES Yahya

Département Sciences du Numérique - Première année
2021-2022

Table des matières

1 Introduction

Ce document présente un rapport de la première partie du projet de calcul scientifique et analyse de données. Dans cette première partie on s'est intéressé à l'étude de l'outil d'analyse en composante principale (ACP).

L'ACP est une technique descriptive permettant d'étudier les relations qui existent entre des variables quantitatives, notamment sur des échantillons de grande taille. Les domaines d'application de cette méthode sont très variés. L'ACP permet également le partitionnement des données.

2 Les "Eigenfaces"

La base de données sur laquelle ce projet est appliqué est une collection d'images utiles pour mener des expériences en psychologie (Psychological Image Collection at Stirling (PICS) 1). Elle comporte 32 individus (16 hommes et 16 femmes) avec 6 postures/expressions faciales (face, trois quart face et trois émotions différentes par posture). La Figure 1 montre une base de n = 16 visages présentant 4 personnes masquées et non masquées dans 4 positions/émotions différentes relativement à la caméra. Pour cette partie du projet, nous conisdérons seulement 4 personnes (2 femmes et 2 hommes) et 4 postures pour faire partie de la base d'apprentissage (Figure 1).

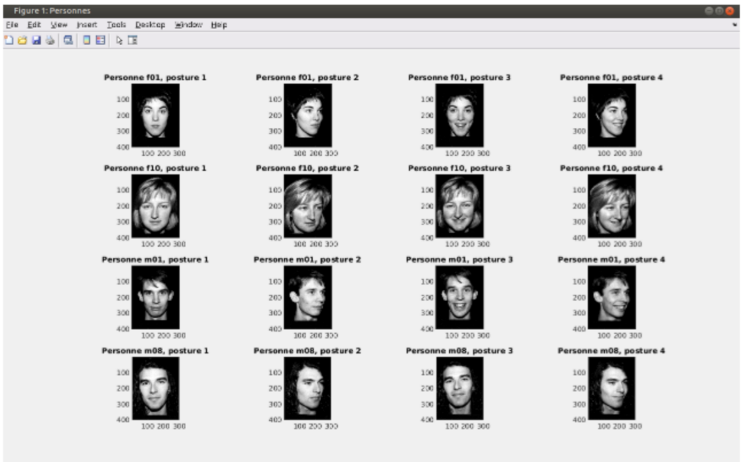

FIGURE 1 – Une base d'apprentissage

2.1 Question 1 : Analyse en Composantes Principales

On visualise sur la figure 2 les résultats de l'exécution du script eigenfaces.m qui vise à calculer les axes principaux des images d'apprentissage à partir des vecteurs propres associés aux n - 1 valeurs propres non nulles de la matrice de variance/covariance des données.

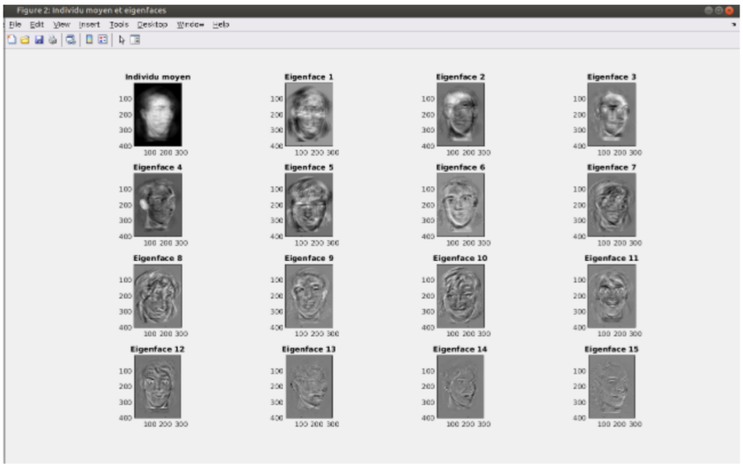

FIGURE 2 – Les "Eigenfaces"

2.2 Question 2 : Projection des images sur les eigenfaces

On visualise les images d'apprentissage reconstruites à l'aide des q premières Eigenfaces et des q premières composantes principales. On remarque que plus le nombre de composantes principales utilisées dans la reconstruction des images augmente, plus les images d'apprentissage reconstruites sont claires.

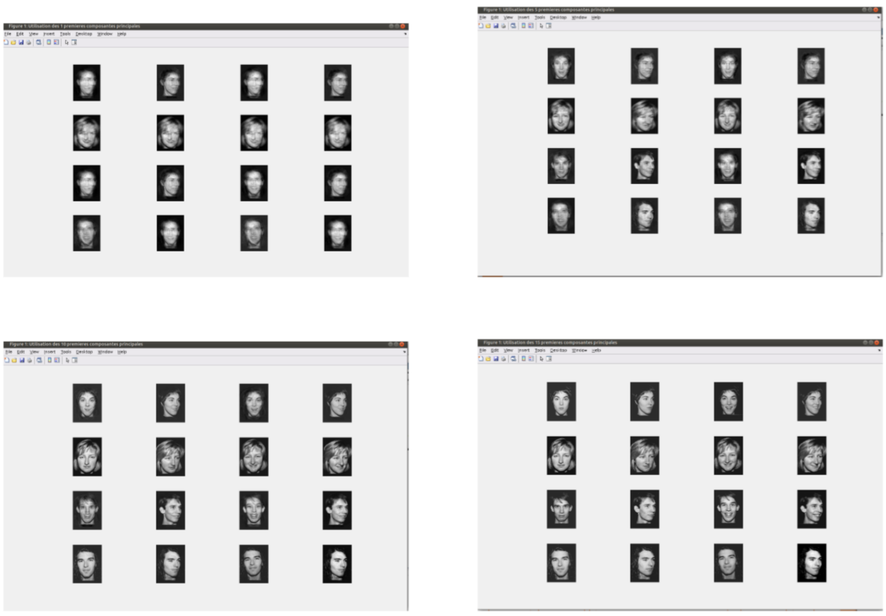

FIGURE 3 – les images d'apprentissage reconstruites

On visualise sur la figure 4 l'évolution de la racine carrée de l'erreur quadratique moyenne(RMSE) entre les images originales et les images reconstruites en fonction de q. On voit bien que "RMSE" est décroissante et s'annule pour q=14 et q=15.

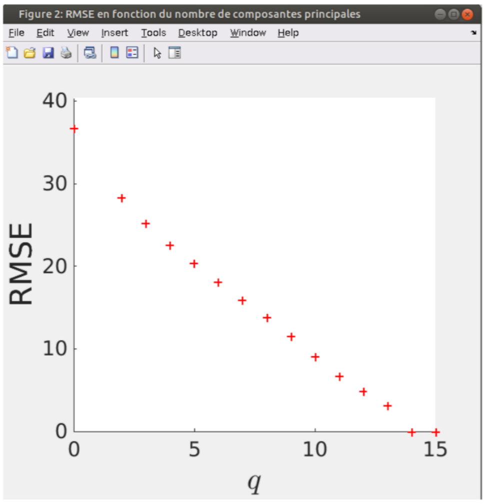

FIGURE 4 – Root Mean Square Error

2.3 Question 3 : Travail sur les visages masqués

On refait le même travail avec les images masquées et on visualise les résultats.

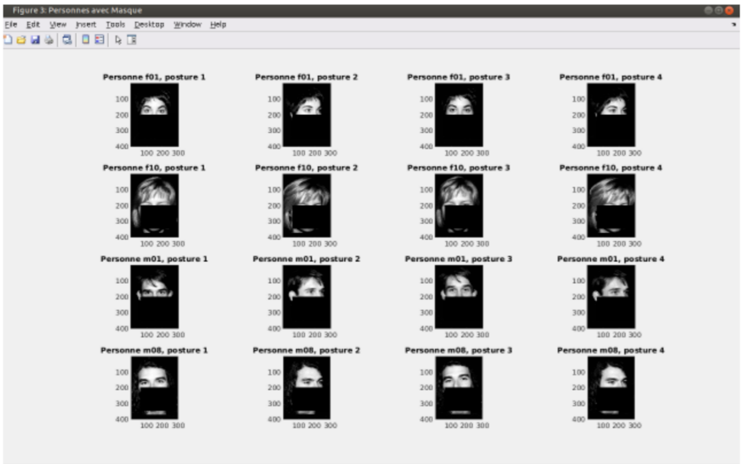

Figure 5 – Les visages masqués

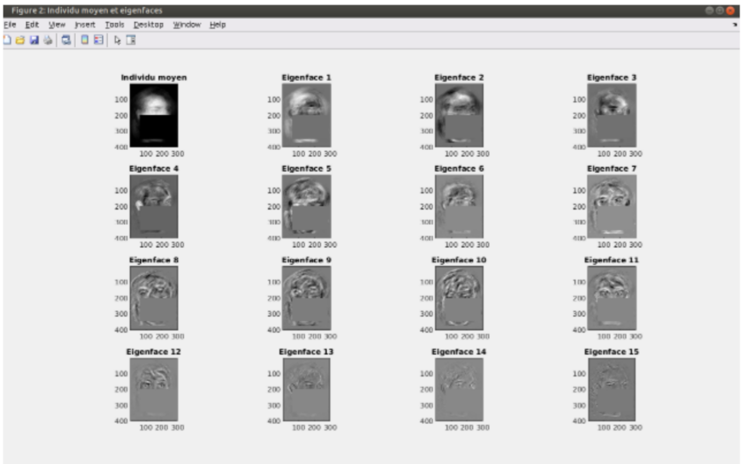

Figure 6 – Les "eigenfaces" masqués

On visualise grâce au script de "projection_masque.m" les images masquées reconstruites à l'aide des q premières eigenfaces masqués et des q premières composantes principales.

FIGURE 7 – les images masquées reconstruites

3 L'ACP et la méthode de la puissance itérée

3.1 Question 4

Montrons que connaître les valeurs propres et les vecteurs propres de ${}^tH$H permet de connaître les éléments propres de H ${}^tH$.

soit λ une valeur propre non nulle de ${}^tH$H.

Donc, $\exists$ x vecteur propre non nulle tel que ${}^tH$H.x = λ.x. Alors H. ${}^tH$H.x = λ.H.x.
On pose y = H.x, y $\neq 0$ car sinon ${}^tH$H.x = ${}^tH$y = 0 = λ.x et x $\neq 0$ et $\lambda \neq 0$
donc H ${}^tH$.y = λ.y. Alors λ est une valeur propre de H ${}^tH$.

D'ou ${}^tH$H et H ${}^tH$ ont les mêmes valeurs propres non nulles.

3.2 Question 5

On complète le script puissance_iteree.m. Les résultats de l'exécution sont donnés ci-dessous :

```
Erreur pour la methode avec la grande matrice = 9.992e-09
Erreur pour la methode avec la petite matrice = 9.964e-09
Ecart relatif entre les deux valeurs propres trouvees = 0.00e+00
Temps pour une ite avec la grande matrice = 1.608e-03
Temps pour une ite avec la petite matrice = 1.306e-04

Valeur propre dominante (methode avec la grande matrice) = 9.227e+04
Valeur propre dominante (methode avec la petite matrice) = 9.227e+04
Valeur propre dominante (fonction eig) = 9.227e+04
```

FIGURE 8 – Exécution de puissance-iteree.m

3.3 Question 6

Il est plus utile d'utiliser la méthode de la puissance itérée pour calculer les éléments propres de la matrice de variance/covariance puisqu'elle trie les éléments propres par ordre décroissant au fur et à mesure contrairement à la méthode eig de matlab qui les renvoie dans leur ordre initial. La seule contrainte de cette méthode est que parfois on peut avoir une boucle infinie dans certains cas, donc dans ces cas il vaut mieux faire appel à la fonction eig de matlab.

3.4 Question 7

Afin de minimiser le temps de calcul et la mémoire utilisée, on doit appliquer la méthode de la puissance itérée avec déflation à la petite matrice, car elle met moins de temps lors d'une itération que la grande matrice. Ceci paraît logique puisque la petite matrice occupe moins de mémoire que la grande.