Séance-Projet 1

Application de l'ACP :

Les «Eigenfaces» – Méthode de la Puissance Itérée

Nihal Belmakhfi, Augustin Boulard, Héloïse Lafargue

Calcul Scientifique et Analyse de Données

Département Sciences du Numérique - Première année

2021-2021

Table des matières

Table des figures

1 Introduction

Ce document constitue la première partie du projet de calcul scientifique et analyse de données. Cette première partie s'intéresse à l'étude de l'ACP (analyse en composante principale).
L'ACP permet de créer un système de reconnaissance grâce aux composantes principales pour retrouver le visage entier dans la base d'apprentissage le plus similaire au visage masqué et de permettre une reconstruction de la zone du masque.

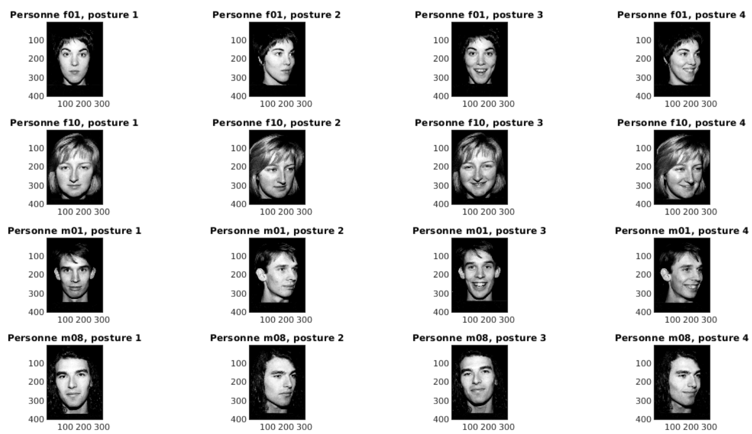

FIGURE 1 – Une base de visages

2 Les «Eigenfaces»

2.1 Exercice 1 : Analyse en Composantes Principales

Question 1 : En complétant le script eigenfaces.m, on calcule les axes principaux des images d'apprentissage, appelés les eigenfaces. On obtient la figure 2 qui représente l'individu moyen ou eigenfaces.

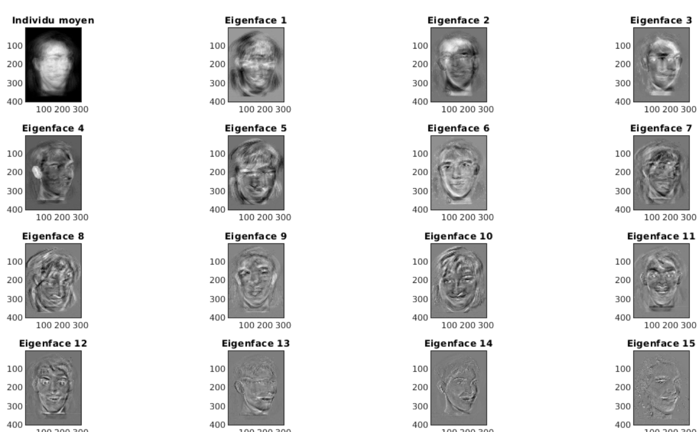

FIGURE 2 – Les "eigenfaces"

2.2 Exercice 2 : projection des images sur les eigenfaces

Question 2 : En complétant le script projection.m, on affiche les images d'apprentissage reconstruites à l'aide des q premières eigenfaces et des q premières composantes principales, pour $q \in [0, n-1]$ (q = 0 correspond à l'individu moyen). On obtient les images sur la figure 3 et la RMSE sur la figure 4.

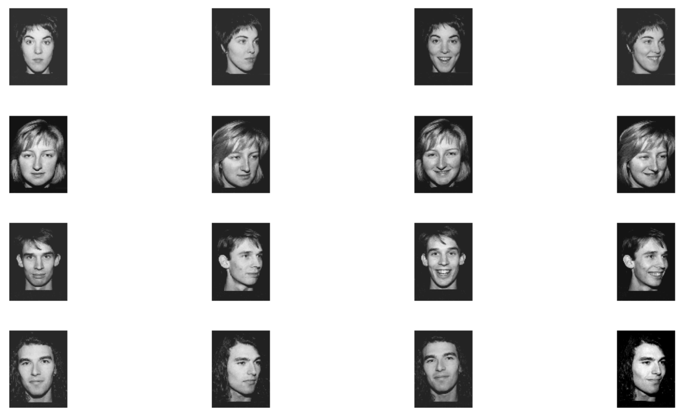

FIGURE 3 – Projection des images sur les eigenfaces

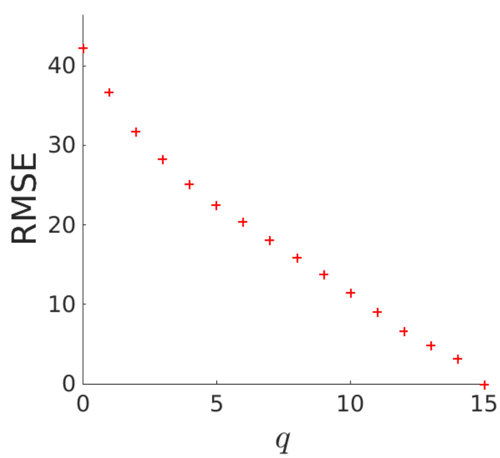

FIGURE 4 – Calcul de la RMSE

2.3 Exercice 3 : travail sur les visages masqués

Question 3 : On réalise les eigenfaces sur les images masquées et la projection des images sur ces eigenfaces. On obtient les eigenfaces masquées sur la figure 5, la projection sur la figure 6 et la RMSE sur figure 7.

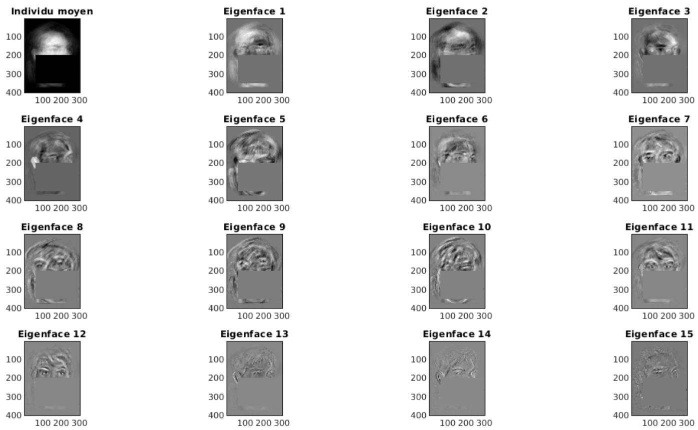

FIGURE 5 – Les eigenfaces masquées

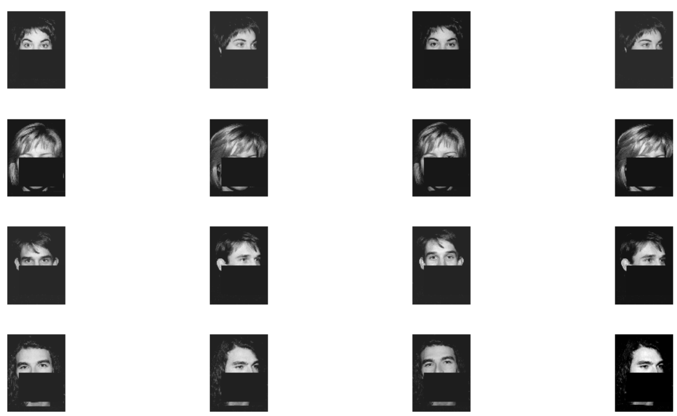

FIGURE 6 – Projection des images sur les eigenfaces

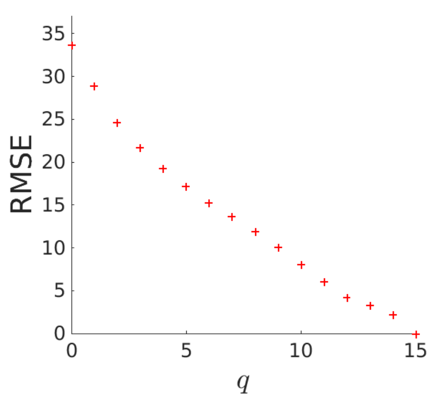

FIGURE 7 – Calcul de la RMSE avec les masques

3 L'ACP et la méthode de la puissance itérée

Question 4 : Soit une matrice rectangulaire H $\in \mathbb{R}^{p \times n}$. Expliquer pourquoi connaître les éléments propres – i.e. les valeurs propres et les vecteurs propres – de ${}^tHH$ permet de connaître les éléments propres de $H\,{}^tH$.

Si on pose $H = UD\,{}^tV$ la SVD de H, avec v_i un vecteur propre de ${}^tHH$ associé à la valeur propre λ_i, alors u_i est un vecteur propre de H${}^tH$ tel que $u_i = \frac{1}{\sqrt{\lambda_i}}Hv_i$ comme nous avons vu en cours. Les deux matrices ont donc les mêmes éléments propres.
(démo : Soit v_i un vecteur propre de ${}^tHH$ associé à la valeur propre λ_i. On pose $u_i = \frac{1}{\sqrt{\lambda_i}}Hv_i$. Alors $HH\,{}^tu_i = \frac{1}{\sqrt{\lambda_i}}HH\,{}^tHv_i = \frac{1}{\sqrt{\lambda_i}}H\lambda_i v_i = \sqrt{\lambda_i}Hv_i = \lambda_i u_i$.)

Question 5 :
On complète le script puissance-iteree.m qui contient réalise la méthode de la puissance itérée pour ${}^tAA$ et pour $A\,{}^tA$. Cette méthode permet de trouver le couple propre dominant d'une matrice.

On observe que l'algorithme de la puissance itérée est plus efficace pour une petite matrice que pour une grande matrice (cf temps d'exécution pour une itération - figure 8).

```
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 = 3.893e-03
Temps pour une ite avec la petite matrice = 1.597e-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 – Méthode de la puisance itérée

Question 6 : Lien avec l'ACP : est-il plus utile en théorie d'utiliser une fonction telle que eig ou la méthode de la puissance itérée pour calculer les éléments propres de Σ si le but est d'effectuer une ACP pour réduire les dimensions d'un espace ? Justifier.

En théorie, pour effectuer une ACP, il est mieux d'utiliser la méthode de la puissance itérée pour calculer les éléments propres de Σ car elle trie les éléments propres par ordre décroissant au fur et à mesure. Alors que eig calcule tous les couples propres et non ordonnés ce qui constitue une perte de temps et de mémoire.

Question 7 : Si l'on choisit d'utiliser la méthode de la puissance itérée avec déflation pour calculer les éléments propres de Σ, sur quelle matrice doit-on appliquer la méthode pour minimiser le temps de calcul et la mémoire utilisée ?

On peut écrire Σ sous la forme : $\Sigma = {}^tX_cX_c/n$, et d'après la question 4 , les deux matrices ${}^tX_cX_c$ et $X_c\,{}^tX_c$ ont les mêmes éléments propres. Donc il suffit de choisir entre ces matrices celle de plus faible taille. En effet, d'après Q5, pour minimiser le temps de calcul et la mémoire utilisée, il faut choisir une matrice de petite taille.
Si A $\in M_{n,p}(\mathbb{R})$, avec $n > p$, alors on choisit ${}^tAA$ pour le calcul des éléments propres.

4 Conclusion

Cette première partie du projet nous a permis d'obtenir les eigenfaces qui permettent de restaurer la zone dégradée et de comprendre l'ACP, notamment avec la méthode de la puissance itérée.