Rapport Analyse de données

KSSIM Aymane - CARREZ Jérémie - HINNIGER-FORAY Nohé

Mars-Avril 2022

1 Introduction

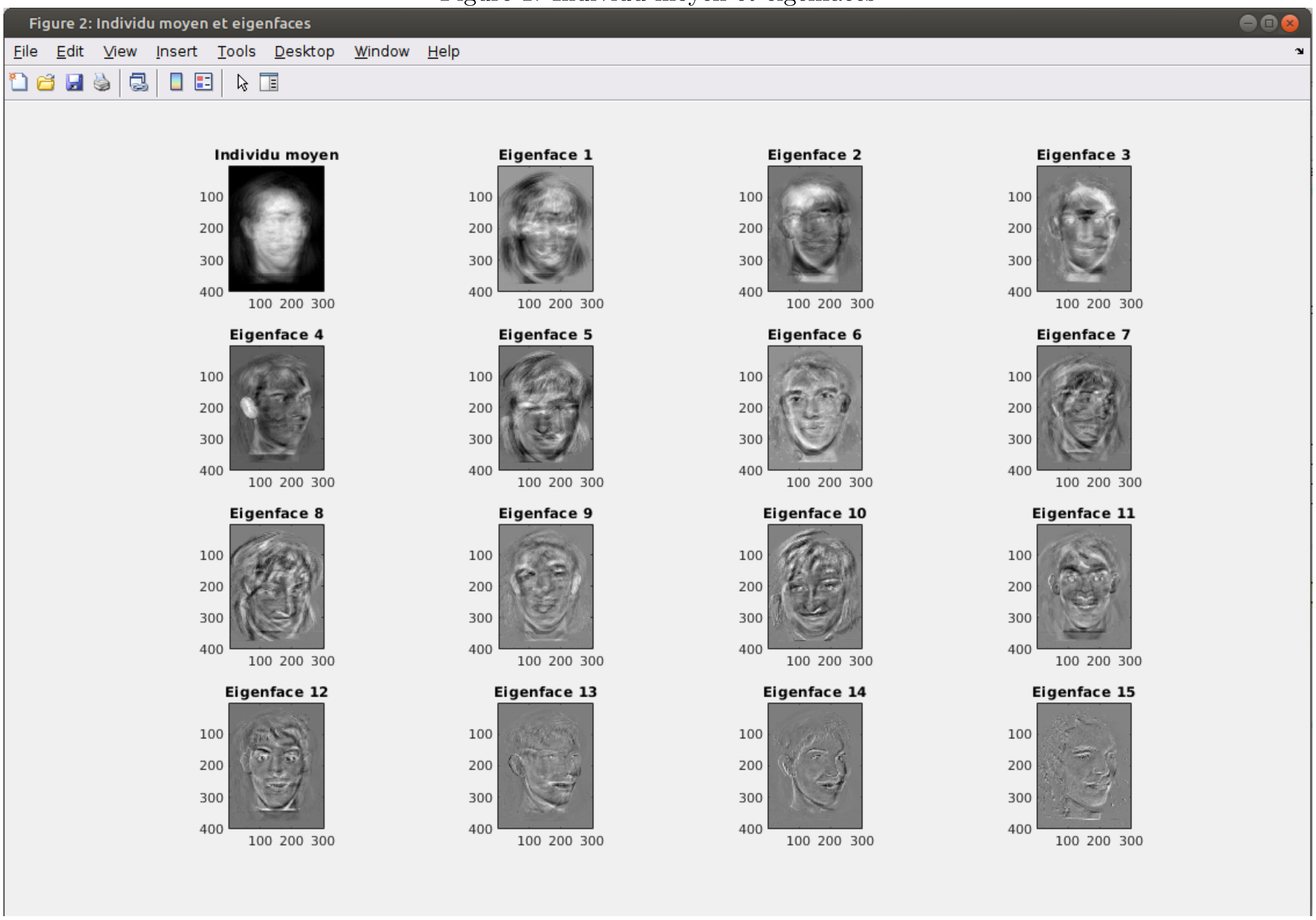
Dans ce projet, nous allons reconstituer la partie cachée par un masque de photos de plusieures personnes ("inpainting") en utilisant la technique des eigenfaces.

2 Eigenfaces

2.1 Question 1 :

Le résultat que nous obtenu correspond au résultat attendu :

Figure 1: Individu moyen et eigenfaces



2.2 Question 2 :

Le résultat que nous obtenu correspond au résultat attendu :

Figure 2: Utilisation des 15 premières composantes principales

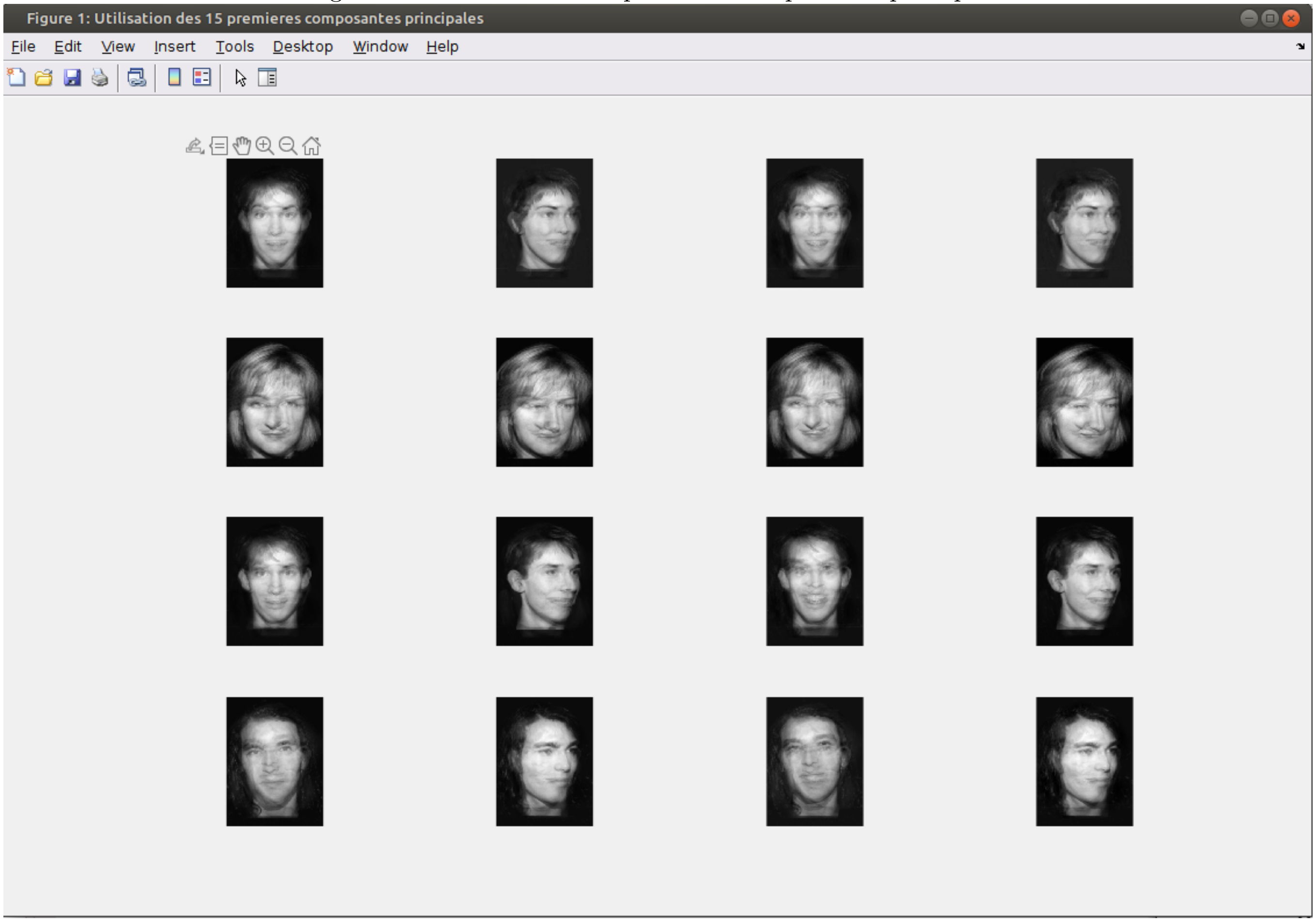
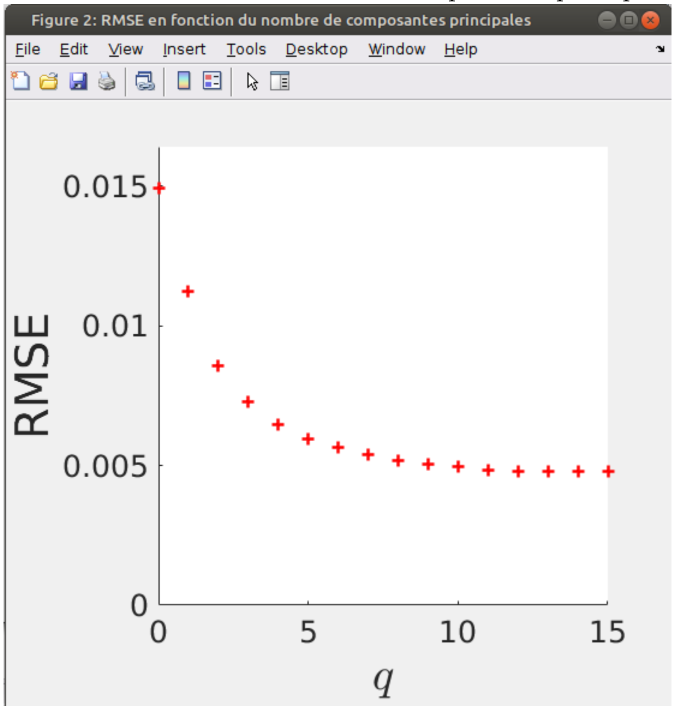


Figure 3: Erreur en fonction du nombre de composantes principales utilisées



2.3 Question 3 :

Le résultat que nous obtenu correspond au résultat attendu :

Figure 4: Individu moyen et eigenfaces avec masque

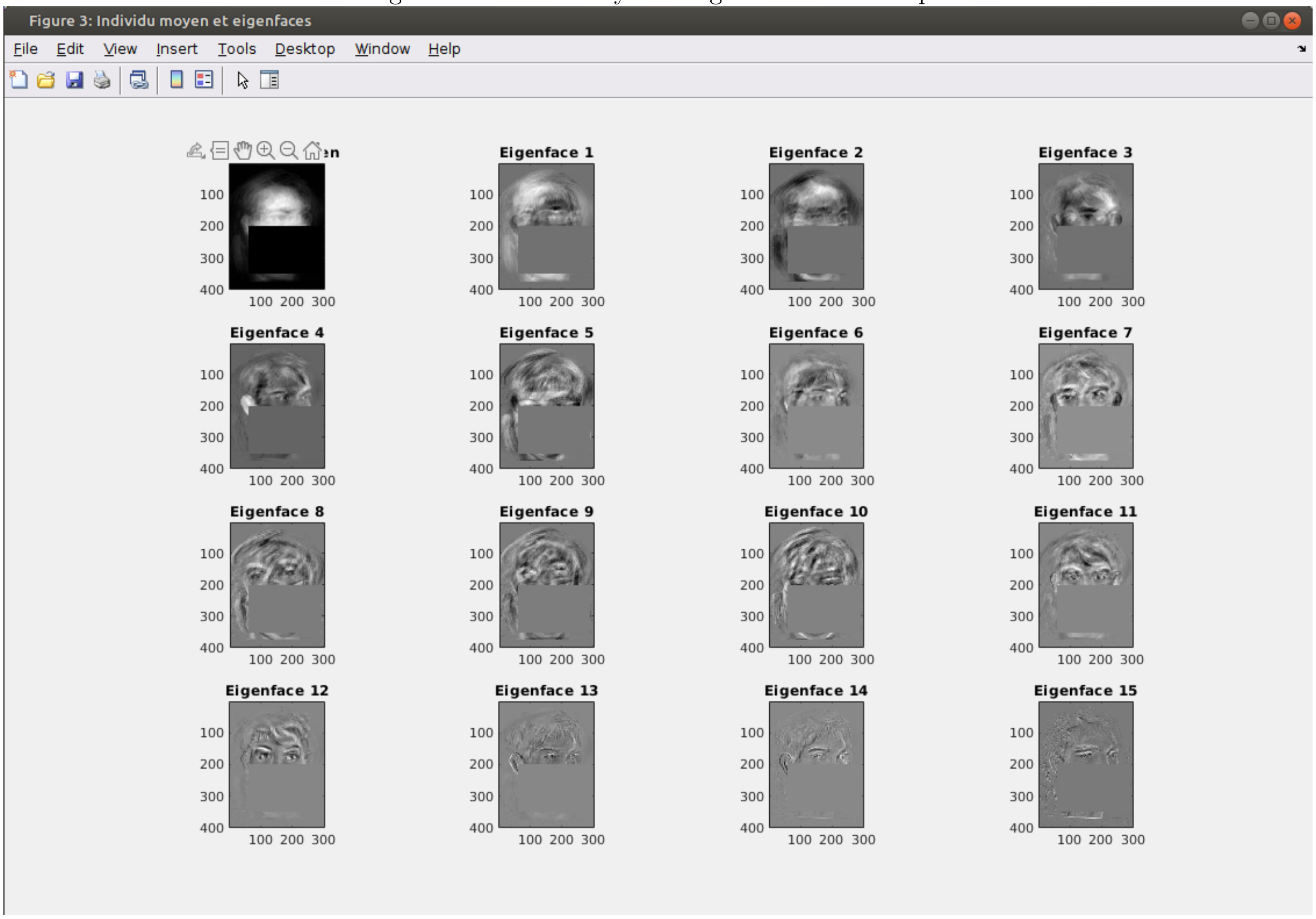


Figure 5: Utilisation des 15 premières composantes principales avec masque

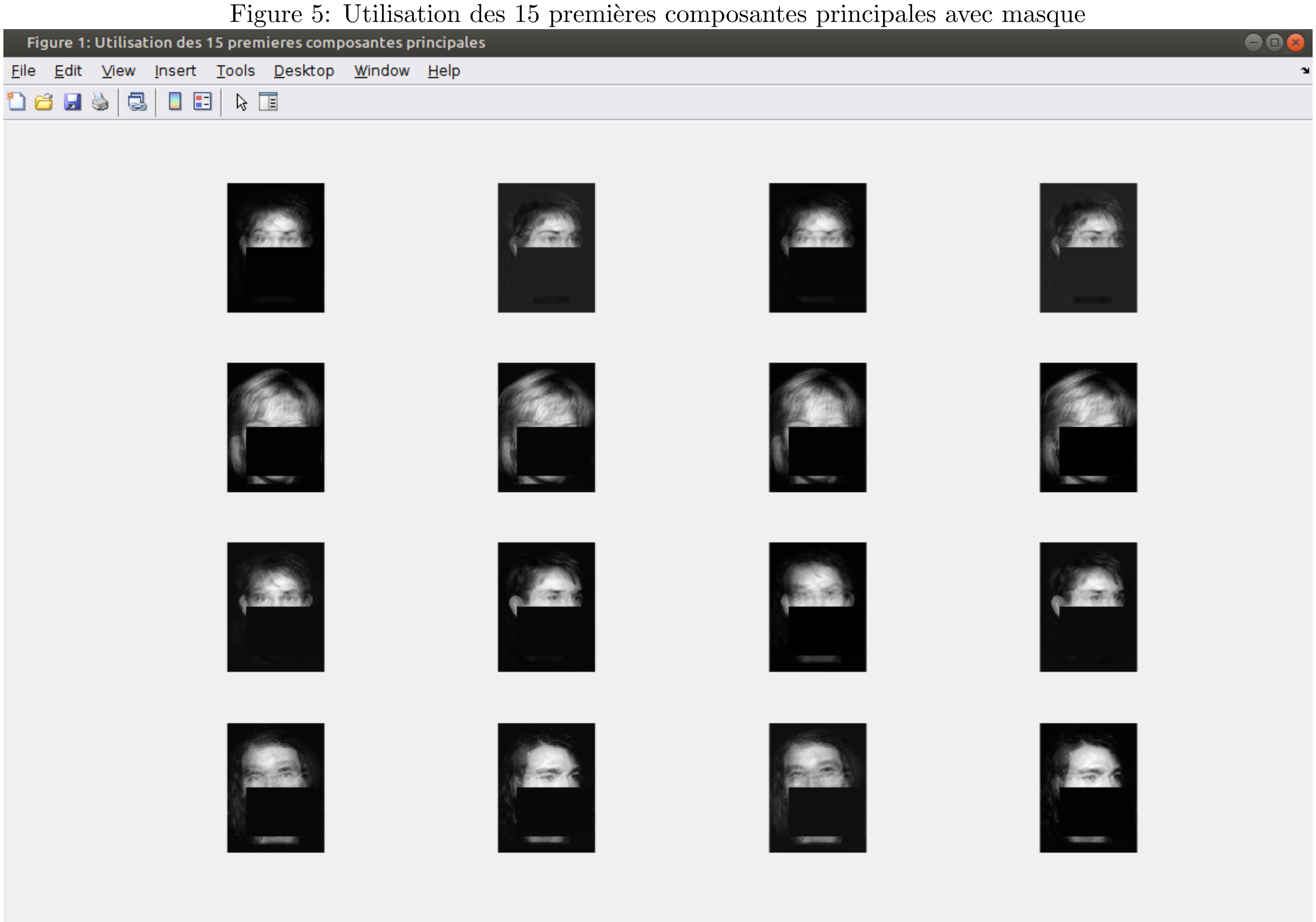
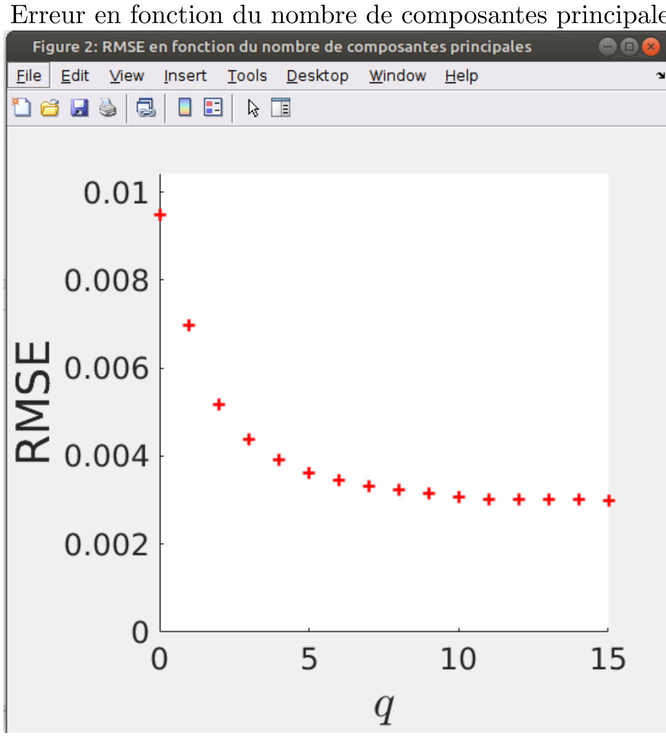


Figure 6: Erreur en fonction du nombre de composantes principales utilisées



2.4 Question 4 :

Soit $\mathbf{H} \in \mathbb{R}^{p \times n}$, $\lambda \in \mathbb{R}$ une valeur propre de $\mathbf{H}^\intercal\mathbf{H}$ et $\mathbf{X} \in \mathbb{R}^n$ un vecteur propre associé a λ alors :

$$\mathbf{H}^\intercal\mathbf{H}\mathbf{X} = \lambda\mathbf{X} \tag{1}$$

En multipliant les deux termes de l'équation par $\mathbf{H}$ à gauche, on obtient :

$$\mathbf{H}\mathbf{H}^\intercal\mathbf{H}\mathbf{X} = \lambda\mathbf{H}\mathbf{X} \tag{2}$$

En posant $\mathbf{Y} = \mathbf{H}\mathbf{X}$, on a :

$$\mathbf{H}\mathbf{H}^\intercal\mathbf{Y} = \lambda\mathbf{Y} \tag{3}$$

D'où λ est aussi une valeur propre de $\mathbf{H}\mathbf{H}^\intercal$ et $\mathbf{Y} = \mathbf{H}\mathbf{X}$ est un vecteur propre de $\mathbf{H}\mathbf{H}^\intercal$ pour la valeur propre λ

2.5 Question 5 :

Figure 7: Résultats du script puissance_iteree

```
Command Window
Erreur pour la methode avec la grande matrice = 9.887e-09
Erreur pour la methode avec la petite matrice = 9.787e-09
Ecart relatif entre les deux valeurs propres trouvees = 1.02e-02
Temps pour une ite avec la grande matrice = 4.132e-03
Temps pour une ite avec la petite matrice = 8.734e-05

Valeur propre dominante (methode avec la grande matrice) = 9.326e+04
Valeur propre dominante (methode avec la petite matrice) = 9.326e+04
Valeur propre dominante (fonction eig) = 9.326e+04
>>
```

2.6 Question 6 :

La fonction **eig** dans Matlab calcule toutes les valeurs propres d'une matrice alors que la méthode de la puissance itérée ne calcule que les valeurs propres distinctes de la matrice. Il est donc à priori plus intéressant quand on veut effectuer une ACP d'utiliser l'algorithme de la puissance itérée. Cependant, afin de précisément connaître la différence en temps de calcul des deux options il faut connaître l'algorithme utilise par la fonction **eig**.

2.7 Question 7 :

Puisque les deux matrices Σ et $\Sigma_2 = \mathbf{X}_c{}^\intercal\mathbf{X}_c/n$ ont les mêmes valeurs propres non nulles et qu'il est possible de retrouver les vecteurs propres de Σ à partir de ceux de Σ_2, alors afin de réduire le temps de calcul et la mémoire utilisée par le programme, il faut effectuer l'algorithme de la puissance itérée avec déflation sur Σ_2.
En effet pour un vecteur propre $\mathbf{V}_2$ de Σ_2 on peut retrouver $\mathbf{V} = \mathbf{X}_c\mathbf{V}_2$ qui est un vecteur propre de Σ par la même valeur propre λ commune des deux matrices. Cette propriété découle du fait que $\Sigma = \mathbf{X}_c\mathbf{X}_c{}^\intercal/n$ et que :

$$\Sigma_2\mathbf{V}_2 = \lambda\mathbf{V}_2$$

En multipliant par $\mathbf{X}_c$ à droite et en développant l'expression de Σ_2, on obtient :

$$\mathbf{X}\mathbf{X}^\intercal\mathbf{X}\mathbf{V}_2 = \lambda\mathbf{X}\mathbf{V}_2$$

On reconnaît l'expression de Σ et celle de $\mathbf{V}$, l'équation précédente devient donc :

$$\Sigma\mathbf{V} = \lambda\mathbf{V}$$