# Analyse factorielle - Rappels et calculs matriciels

François Husson

Unité pédagogique de mathématiques appliquées - l'institut Agro

husson@agrocampus-ouest.fr

# Analyse factorielle - Rappels et calculs matriciels

1 Décomposition en valeurs singulières (SVD)

2 SVD et images

3 Lien SVD et ACP, AFC, ACM

1	2	3	4
2	4	6	8
5	10	15	20
-1	-2	-3	-4
-10	-20	-30	-40

	1	2	3	4
1	1	2	3	4
2	2	4	6	8
5	5	10	15	20
-1	-1	-2	-3	-4
-10	-10	-20	-30	-40

	1	2	3	4
1	1	2	3	4
2	2	4	6	8
5	5	10	15	20
-1	-1	-2	-3	-4
-10	-10	-20	-30	-40

1	2	3	4		1
2	4	6	8		2
5	10	15	20	=	5
-1	-2	-3	-4		-1
-10	-20	-30	-40		-10

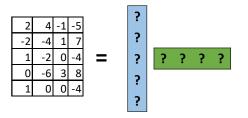


	1	2	3	4
1	1	2	3	4
2	2	4	6	8
5	5	10	15	20
-1	-1	-2	-3	-4
-10	-10	-20	-30	-40

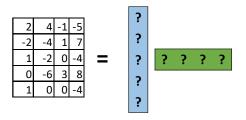
				_						
1	2	3	4		1					
2	4	6	8		2	ı				
5	10	15	20	=	5		1	2	3	4
-1	-2	-3	-4		-1					
-10	-20	-30	-40		-10					

$$5*4 = 20$$
 valeurs  $5+4 = 9$  valeurs

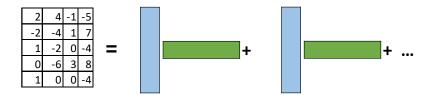
Toutes les matrices sont-elles de rang 1?



Toutes les matrices sont-elles de rang 1? Non, mais ...



Toutes les matrices sont-elles de rang 1? **Non**, mais ...elles peuvent toutes s'écrire comme une somme de matrices de rang 1



$$X = U \Lambda V'$$
 avec  $UU' = U'U = Id_n$  et  $VV' = V'V = Id_p$ 

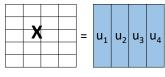


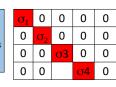






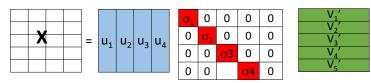
$$X = U\Lambda V'$$
 avec  $UU' = U'U = Id_n$  et  $VV' = V'V = Id_p$ 

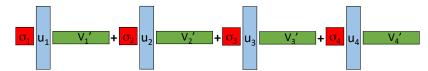




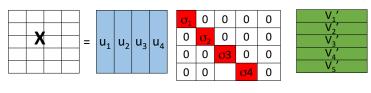


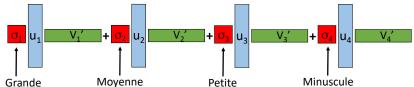
$$X = U \Lambda V'$$
 avec  $UU' = U'U = Id_n$  et  $VV' = V'V = Id_p$ 





$$X = U \Lambda V'$$
 avec  $UU' = U'U = Id_n$  et  $VV' = V'V = Id_p$ 





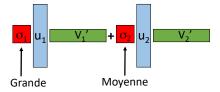
La SVD d'une matrice  $X_{(n,p)}$  donne les matrices  $U_{(n,r)}$ ,  $\Lambda_{(r,r)}$  et  $V_{(p,r)}$  telles que :

$$X = U\Lambda V'$$
 avec  $UU' = U'U = Id_n$  et  $VV' = V'V = Id_p$ 



$\sigma_1$	0	0	0	0
0	$\sigma_2$	0	0	0
0	0	σ3	0	0
0	0		σ4	0





Avec les 2 premières dimensions, on a une approximation de X de rang 2, avec  $2 \times (n+1+p)$  valeurs au lieu de  $n \times p$ 

2	4	-1	-5
-2	-4	1	7
1	-2	0	-4
0	-6	3	8
1	0	0	-4

2	4	-1	-5
-2	-4	1	7
1	-2	0	-4
0	-6	3	8
1	0	0	-4

=

-0,43	0,11	0,59	0,54	15,4			
0,54	0,09	-0,40	0,41		4,7		
-0,16	-0,80	-0,25	0,47			2.1	
0,66	-0,35	0,65	-0,06			۷, ۲	0.20
-0,23	-0,46	0,00	-0,56				0,36

	-0,49		
-0,26	0,81	-0,23	0,48
0,82	0,27	0,45	0,20
0,48	-0,19	-0,84	0,17

2	4	-1	-5
-2	-4	1	7
1	-2	0	-4
0	-6	3	8
1	0	0	-4

	-0,43	0,11	0,59	0,54	15,4			
	0,54	0,09	-0,40	0,41		4,7		
=	-0,16	-0,80	-0,25	0,47			2 1	
	0,66	-0,35	0,65	-0,06			۷, ۱	0.26
	-0,23	-0,46	0,00	-0,56				0,36

	-0,49		
-0,26	0,81	-0,23	0,48
0,82	0,27	0,45	0,20
0,48	-0,19	-0,84	0,17

-0,43

15,4 -0,16 -0,15 -0,49 0,19 0,84 0,66 -0,23

1.01	3.27	-1.28	-5.55
-1.26	-4.09	1.61	6.94
0.38	1.23	-0.48	-2.10
-1.55	-5.02	1.97	8.53
0.53	1.72	-0.67	-2.92

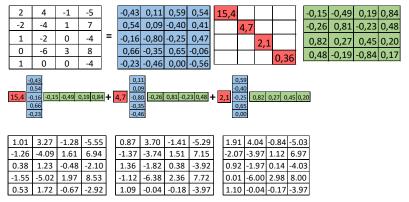
Approximation acceptable

2	4	-1	-5		-0,43	0,11	0,59	0,54	15,4				-0,15	-0,49	0,19	0,84
-2	-4	1	7		0,54	0,09	-0,40	0,41		4,7			-0,26	0,81	-0,23	0,48
1	-2	0	-4	=	-0,16	-0,80	-0,25	0,47		Ť	2,1		0,82	0.27	0.45	0,20
0	-6	3	8		0,66	-0,35	0,65	-0,06			-	0.26	0,48	-0,19	-	-
1	0	0	-4		-0,23	-0,46	0,00	-0,56				0,36	0,10	0,13	0,01	0,17
-0,43 0,54 15,4 0.16 0.15 0.49 0.19 0.84 + 4,7 0,66 0,23 0.48 0.35 0.48																

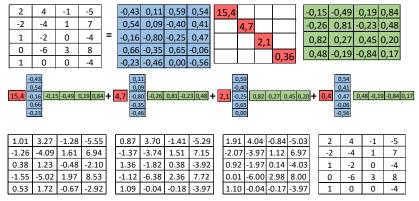
1.01	3.27	-1.28	-5.55	
-1.26	-4.09	1.61	6.94	
0.38	1.23	-0.48	-2.10	
-1.55	-5.02	1.97	8.53	
0.53	1.72	-0.67	-2.92	

0.87	3.70	-1.41	-5.29
-1.37	-3.74	1.51	7.15
1.36	-1.82	0.38	-3.92
-1.12	-6.38	2.36	7.72
1.09	-0.04	-0.18	-3.97

Approximation bonne



Approximation très bonne



On retrouve X

Soit  $X_{(n,p)}$  une matrice, comment faire la SVD de X, i.e. obtenir les matrices  $U_{(n,r)}$ ,  $\Lambda_{(r,r)}$  et  $V_{(p,r)}$  telles que :

$$X = U \Lambda V'$$
 avec  $UU' = U'U = Id_n$  et  $VV' = V'V = Id_p$ 

X'X est la matrice des covariances (au n près) XX' est la matrice des produits scalaires

Comment obtenir 
$$U_{(n,r)}$$
,  $\Lambda_{(r,r)}$  et  $V_{(p,r)}$  telles que  $X = U\Lambda V'$  avec  $UU' = U'U = Id_n$  et  $VV' = V'V = Id_p$ 

Comment obtenir 
$$U_{(n,r)}$$
,  $\Lambda_{(r,r)}$  et  $V_{(p,r)}$  telles que  $X = U\Lambda V'$  avec  $UU' = U'U = Id_n$  et  $VV' = V'V = Id_p$  
$$X'X = (U\Lambda V')'(U\Lambda V') = (V\Lambda U')(U\Lambda V') = V\Lambda I_d\Lambda V' = V\Lambda^2 V'$$

Comment obtenir 
$$U_{(n,r)}$$
,  $\Lambda_{(r,r)}$  et  $V_{(p,r)}$  telles que  $X = U\Lambda V'$  avec  $UU' = U'U = Id_n$  et  $VV' = V'V = Id_p$  
$$X'X = (U\Lambda V')'(U\Lambda V') = (V\Lambda U')(U\Lambda V') = V\Lambda I_d\Lambda V' = V\Lambda^2 V'$$
 
$$\Rightarrow X'XV = V\Lambda^2 V'V = V\Lambda^2$$

Comment obtenir 
$$U_{(n,r)}$$
,  $\Lambda_{(r,r)}$  et  $V_{(p,r)}$  telles que  $X = U \Lambda V'$  avec  $UU' = U'U = Id_n$  et  $VV' = V'V = Id_p$  
$$X'X = (U \Lambda V')'(U \Lambda V') = (V \Lambda U')(U \Lambda V') = V \Lambda I_d \Lambda V' = V \Lambda^2 V'$$
 
$$\Rightarrow X'XV = V \Lambda^2 V'V = V \Lambda^2$$

 $\Rightarrow \Lambda^2$  valeurs propres et V vecteurs propres de la matrice de covariance X'X

Comment obtenir 
$$U_{(n,r)}$$
,  $\Lambda_{(r,r)}$  et  $V_{(p,r)}$  telles que  $X = U\Lambda V'$  avec  $UU' = U'U = Id_n$  et  $VV' = V'V = Id_p$  
$$X'X = (U\Lambda V')'(U\Lambda V') = (V\Lambda U')(U\Lambda V') = V\Lambda I_d\Lambda V' = V\Lambda^2 V'$$
 
$$\Rightarrow X'XV = V\Lambda^2 V'V = V\Lambda^2$$

 $\Rightarrow \Lambda^2$  valeurs propres et V vecteurs propres de la matrice de covariance X'X

$$XX' = (U \wedge V')(U \wedge V')' = (U \wedge V')(V \wedge U') = U \wedge I_d \wedge U' = U \wedge^2 U'$$

Comment obtenir 
$$U_{(n,r)}$$
,  $\Lambda_{(r,r)}$  et  $V_{(p,r)}$  telles que  $X = U \Lambda V'$  avec  $UU' = U'U = Id_n$  et  $VV' = V'V = Id_p$  
$$X'X = (U \Lambda V')'(U \Lambda V') = (V \Lambda U')(U \Lambda V') = V \Lambda I_d \Lambda V' = V \Lambda^2 V'$$
 
$$\Rightarrow X'XV = V \Lambda^2 V'V = V \Lambda^2$$

 $\Rightarrow \Lambda^2$  valeurs propres et V vecteurs propres de la matrice de covariance X'X

$$XX' = (U \wedge V')(U \wedge V')' = (U \wedge V')(V \wedge U') = U \wedge I_d \wedge U' = U \wedge^2 U'$$
  
$$\Rightarrow XX'U = U \wedge^2 U'U = U \wedge^2$$

Comment obtenir 
$$U_{(n,r)}$$
,  $\Lambda_{(r,r)}$  et  $V_{(p,r)}$  telles que  $X = U\Lambda V'$  avec  $UU' = U'U = Id_n$  et  $VV' = V'V = Id_p$  
$$X'X = (U\Lambda V')'(U\Lambda V') = (V\Lambda U')(U\Lambda V') = V\Lambda I_d\Lambda V' = V\Lambda^2 V'$$
 
$$\Rightarrow X'XV = V\Lambda^2 V'V = V\Lambda^2$$

 $\Rightarrow \Lambda^2$  valeurs propres et V vecteurs propres de la matrice de covariance X'X

$$XX' = (U\Lambda V')(U\Lambda V')' = (U\Lambda V')(V\Lambda U') = U\Lambda I_d \Lambda U' = U\Lambda^2 U'$$
  
$$\Rightarrow XX'U = U\Lambda^2 U'U = U\Lambda^2$$

 $\Rightarrow \Lambda^2$  valeurs propres et U vecteurs propres de la matrice des produits scalaires XX'

Comment obtenir 
$$U_{(n,r)}$$
,  $\Lambda_{(r,r)}$  et  $V_{(p,r)}$  telles que  $X = U\Lambda V'$  avec  $UU' = U'U = Id_n$  et  $VV' = V'V = Id_p$  
$$X'X = (U\Lambda V')'(U\Lambda V') = (V\Lambda U')(U\Lambda V') = V\Lambda I_d\Lambda V' = V\Lambda^2 V'$$
 
$$\Rightarrow X'XV = V\Lambda^2 V'V = V\Lambda^2$$

 $\Rightarrow \Lambda^2$  valeurs propres et V vecteurs propres de la matrice de covariance X'X

$$XX' = (U \wedge V')(U \wedge V')' = (U \wedge V')(V \wedge U') = U \wedge I_d \wedge U' = U \wedge^2 U'$$
  
$$\Rightarrow XX'U = U \wedge^2 U'U = U \wedge^2$$

 $\Rightarrow \Lambda^2$  valeurs propres et U vecteurs propres de la matrice des produits scalaires XX'

Les valeurs singulières sont les racines carrées des valeurs propres de la matrice de covariance (= valeurs propres de la matrice des produits scalaires)

# Analyse factorielle - Rappels et calculs matriciels

1 Décomposition en valeurs singulières (SVD)

2 SVD et images

3 Lien SVD et ACP, AFC, ACM

# Exercice: SVD et compression d'image

1 Importer l'image (Léna) et la transformer en matrice



library(raster) # raster et rgdal à installer
photo <- raster("https://husson.github.io/img/Lena.png")
photo <- as.matrix(photo)
dim(photo)</pre>

- 2 Faire la SVD sur les données définissant cette image
- 3 Reconstruire l'image en utilisant la reconstruction de rang 5 (faire de même avec les rangs 20, 50 puis 100)
- 4 Pour chaque rang, donner le nombre de données qu'il est nécessaire de stocker (et le pourcentage de données par rapport au nombre de données de l'image originale)

# SVD pour la reconnaissance faciale : eigenfaces

400 images de visages sur photo  $64 \times 64$  pixels











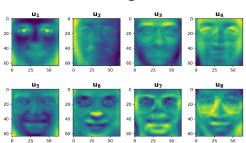
Les images sont disponibles ici :

https://github.com/lloydmeta/Olivetti-PNG/tree/master/images Transformation de chaque image en un vecteur de taille  $64 \times 64 = 4096$ 

Création d'une matrice  $4096 \times 400$  puis on fait la SVD

#### Reconnaissance faciale: eigenfaces

Voici les 8 premiers vecteurs propres mis sous 60
forme d'image (les valeurs ne sont pas comprises entre 0 et 1)



#### Reconnaissance faciale: eigenfaces

Voici les 8 premiers vectures propres mis sous forme d'image (les valeurs ne sont pas comprises entre 0 et 1)

20 20 -40 u<sub>8</sub> 20 40 25 Original image k = 10k = 15 k = 20k = 35

On peut reconstruire une image :

# SVD pour le débruitage d'images

On peut débruiter cette image



# SVD pour le débruitage d'images

On peut débruiter cette image

Reconstitution par SVD en ne conservant que les premières dimensions (ce qui revient à éliminer le bruit sur les dernières dimensions)

l'aime la science des données

r= 20 7 9 %

l'aime la science des données

r= 100 , 39.1 %

l'aime la science des données r= 60 , 23.5 %

l'aime la science des données

r= 200 , 78.2 %

l'aime la science des données

# Analyse factorielle - Rappels et calculs matriciels

1 Décomposition en valeurs singulières (SVD)

2 SVD et images

3 Lien SVD et ACP, AFC, ACM

#### Lien SVD et ACP, AFC, ACM

- ACP est une SVD sur données centrées ou centrées-réduites si l'ACP est normée Plus précisément, avec  $M=diag(\frac{1}{\sigma_1^2},\frac{1}{\sigma_2^2},...\frac{1}{\sigma_p^2})$  et N la matrice diagonale des poids des lignes (1/n), la SVD de  $N^{1/2}XM^{1/2}$  donnent les résultats de l'ACP normée (i.e. les valeurs propres et vecteurs propres de XMX'N et X'NXM)
- AFC est une SVD de la matrice  $S=D_r^{-1/2}(P-rc')D_c^{-1/2}$  avec P=X/n,  $D_r$  et  $D_c$  les matrices diagonales des marges lignes et colonnes de P Les coordonnées des lignes sont :  $F=D_r^{-1/2}U\Lambda$  Les coordonnées des colonnes sont :  $G=D_c^{-1/2}V\Lambda$
- ACM est une AFC sur le tableau disjonctif de X, et donc une SVD.

#### Fiche récapitulative de l'ACP

- Quels tableaux de données? Quels objectifs?
- Comment interpréter?
- Comment considérer des individus supplémentaires, variables qualitatives, variables quantitatives supplémentaires?
- Quelle différence entre ACP normée et non normée?
- Dans un tableau avec 1 variable qualitative, les axes de l'ACP obtenus sur le tableau individus × variables quantitatives sont-ils identiques à ceux obtenus à partir des moyennes par modalité, moyennes pondérées par l'effectif de la modalité?
   Donner un contre-exemple, expliciter les différences d'objectif OU démontrer l'égalité.

#### Fiche récapitulative sur l'AFC

- Quels types de tableaux de données? Quels jeux de données?
   Quels objectifs?
- Comment interpréter?
- Considérer le jeu de données Nobel avec le code suivant :

```
fichier <- "https://husson.github.io/MOOC_AnaDo/AnaDo_JeuDonnees_Nobel_avecMaths.csv"
Nobel <- read.table(fichier, header=TRUE, sep=";", row.names=1, check.names=FALSE)
Nobel <- Nobel[1:8,]</pre>
```

Comparer les objectifs et les résultats de l'ACP et ceux de l'AFC sur ce jeu de données. Bien expliciter la différence.

#### Fiche récapitulative sur l'ACM

- Quels tableaux de données? Quels objectifs? Comment interpréter?
- Prendre le jeu tea du package FactoMineR et faire l'ACM sur le tableau avec uniquement les variables 14 et 18. Puis faire l'AFC sur le tableau de contingence croisant ces 2 variables.

```
library(FactoMineR)
data(tea)
don <- tea[, c(14,18)] ; MCA(don)
TabCont <- table(don) ; CA(TabCont)
```

Comparer objectifs et résultats de l'ACM et l'AFC sur ce jeu de données. Expliciter les différences (dimensionalité et inertie notamment)