Algebra of Principal Component Analysis $\left(\mathbf{Y}_c, \mathbf{Q} = \mathbf{I}_p, \mathbf{D} = \frac{1}{n}\mathbf{I}_n\right)$

Data: $\mathbf{Y} = \begin{bmatrix} 2 & 1 \\ 3 & 4 \\ 5 & 0 \\ 7 & 6 \\ 9 & 2 \end{bmatrix}$ Centre each column on its mean: $\mathbf{Y}_c = [y - \bar{y}] = \begin{bmatrix} -3.2 & -1.6 \\ -2.2 & 1.4 \\ -0.2 & -2.6 \\ 1.8 & 3.4 \\ 3.8 & -0.6 \end{bmatrix}$

Where $\overline{\boldsymbol{y}} = [5.2 \quad 2.6]^t$

$\mathbf{S}_I = \mathbf{SQ} = \frac{1}{5}\mathbf{Y}_c^t\mathbf{Y}_c$ is the covariance matrix (2 variables): $\boldsymbol{S}_I = \begin{bmatrix} 6.56 & 1.28 \\ 1.28 & 4.64 \end{bmatrix}$

Equation for eigenvalues and eigenvectors of $\mathbf{S}_I$: $(\mathbf{S}_I - \lambda_k \mathbf{I} \quad)\boldsymbol{a}_k = 0$

- Axis $\boldsymbol{a}_k$
- Loadings (principal factors) $\boldsymbol{u}_k = \mathbf{Q}\boldsymbol{a}_k$ because $\mathbf{Q} = \mathbf{I}$

Eigenvalues: $\lambda_1 = 7.2$; $\lambda_2 = 4.0$ => $I_g = tr(\mathbf{S}_I) = \lambda_1 + \lambda_2 = 11.2$

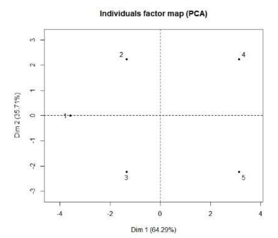
Matrix of eigenvalues : $\mathbf{\Lambda} = \begin{bmatrix} 7.2 & 0 \\ 0 & 4.0 \end{bmatrix}$

Matrix of eigenvectors: $\mathbf{U} = [\boldsymbol{u}_1 \quad \boldsymbol{u}_2] \approx \begin{bmatrix} 0.89 & -0.45 \\ 0.45 & 0.89 \end{bmatrix}$

Coord of the 5 objects in ordination space: $\mathbf{F} = \mathbf{Y}_c\mathbf{U}$ => Principal components

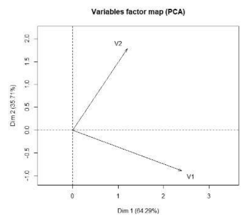
$$\mathbf{F} = [\boldsymbol{F}^1 \quad \boldsymbol{F}^2] \approx \begin{bmatrix} -3.58 & 0 \\ -1.34 & 2.24 \\ -1.34 & -2.24 \\ 3.13 & 2.24 \\ 3.13 & -2.24 \end{bmatrix}$$

with F'F=I



Coord of the 2 variables in ordination space: $\boldsymbol{G}^{\boldsymbol{k}} = \sqrt{\lambda_k}\boldsymbol{a}_k$

$$\mathbf{G} = [\boldsymbol{G}^1 \quad \boldsymbol{G}^2] \approx \begin{bmatrix} 2.4 & -0.89 \\ 1.2 & 1.79 \end{bmatrix}$$




```
# exemple à la main => sous R : résultats au signe près
# ACP (Yc,Ip,(1/n)In)
#
y<-c(2,1,3,4,5,0,7,6,9,2)
Y<-matrix(y,ncol=2,byrow = T) ; Y
Yc<-scale(Y,scale=FALSE);Yc
dim(Yc) ; n=nrow(Yc)
SI<-(1/n)*t(Yc)%*%Yc ; SI
eig=eigen(SI) ; eig
#cP
F=Yc%*%eig$vectors ; round(F,2)
plot(F)
#coord des variables
G<-matrix(0,2,2)
G[,1]=sqrt(eig$values[1])%*%eig$vectors[,1]
G[,2]=sqrt(eig$values[2])%*%eig$vectors[,2]
round(G,2)

library(FactoMineR)
PCA(as.data.frame(Yc),scale.unit=FALSE)

#resultat avec PCA au signe près!
```