# TRIED RNRF TP EX11 Report

## Carl Robinson 11th Oct 2017



# A – ACP sur données centrées

## a) Centrage des données

Raw data:

| Student | Maths | Phys | Fran | Latin | Dessin |
|---|---|---|---|---|---|
| A | 6.0 | 6.0 | 5.0 | 5.5 | 8.0 |
| B | 8.0 | 8.0 | 8.0 | 8.0 | 9.0 |
| C | 6.0 | 7.0 | 11.0 | 9.5 | 11.0 |
| D | 14.5 | 14.5 | 15.5 | 15.0 | 8.0 |
| E | 14.0 | 14.0 | 12.0 | 12.5 | 10.0 |
| F | 11.0 | 10.0 | 5.5 | 7.0 | 13.0 |
| G | 5.5 | 7.0 | 14.0 | 11.5 | 10.0 |
| H | 13.0 | 12.5 | 8.5 | 9.5 | 12.0 |
| I | 9.0 | 9.5 | 12.5 | 12.0 | 18.0 |

Centered data (mean normalised):

| Student | Maths | Phys | Fran | Latin | Dessin |
|---|---|---|---|---|---|
| A | -3.667 | -3.833 | -5.222 | -4.556 | -3.0 |
| B | -1.667 | -1.833 | -2.222 | -2.056 | -2.0 |
| C | -3.667 | -2.833 | 0.778 | -0.556 | 0.0 |
| D | 4.833 | 4.667 | 5.278 | 4.944 | -3.0 |
| E | 4.333 | 4.167 | 1.778 | 2.444 | -1.0 |
| F | 1.333 | 0.167 | -4.722 | -3.056 | 2.0 |
| G | -4.167 | -2.833 | 3.778 | 1.444 | -1.0 |
| H | 3.333 | 2.667 | -1.722 | -0.556 | 1.0 |
| I | -0.667 | -0.333 | 2.278 | 1.944 | 7.0 |

- In the centered data, the mean value of all students scores in each subject is 0.
- Knowing this, the extreme values become more apparent e.g. the high value of student I for dessin
- By using centered values in the PCA plot, the centre of gravity will be in the middle of the cloud of points. This results in an even distribution of points, making the plot easier to interpret.

## b) Ajustement par un sous espace de $R^p$

Eigenvectors U of the matrix X'X:

| **Eigenvector ID** | **Dim 1** | **Dim 2** | **Dim 3** | **Dim 4** | **Dim 5** |
|---|---|---|---|---|---|
| 1 | -0.515 | -0.567 | 0.051 | -0.289 | 0.573 |
| 2 | -0.507 | -0.372 | 0.014 | 0.553 | -0.546 |
| 3 | -0.492 | 0.65 | -0.108 | 0.394 | 0.41 |
| 4 | -0.485 | 0.323 | -0.023 | -0.674 | -0.453 |
| 5 | -0.031 | 0.113 | 0.992 | 0.034 | 0.013 |

- The eigenvectors are the directions of the new axes.

Eigenvalues λ of the matrix X'X:

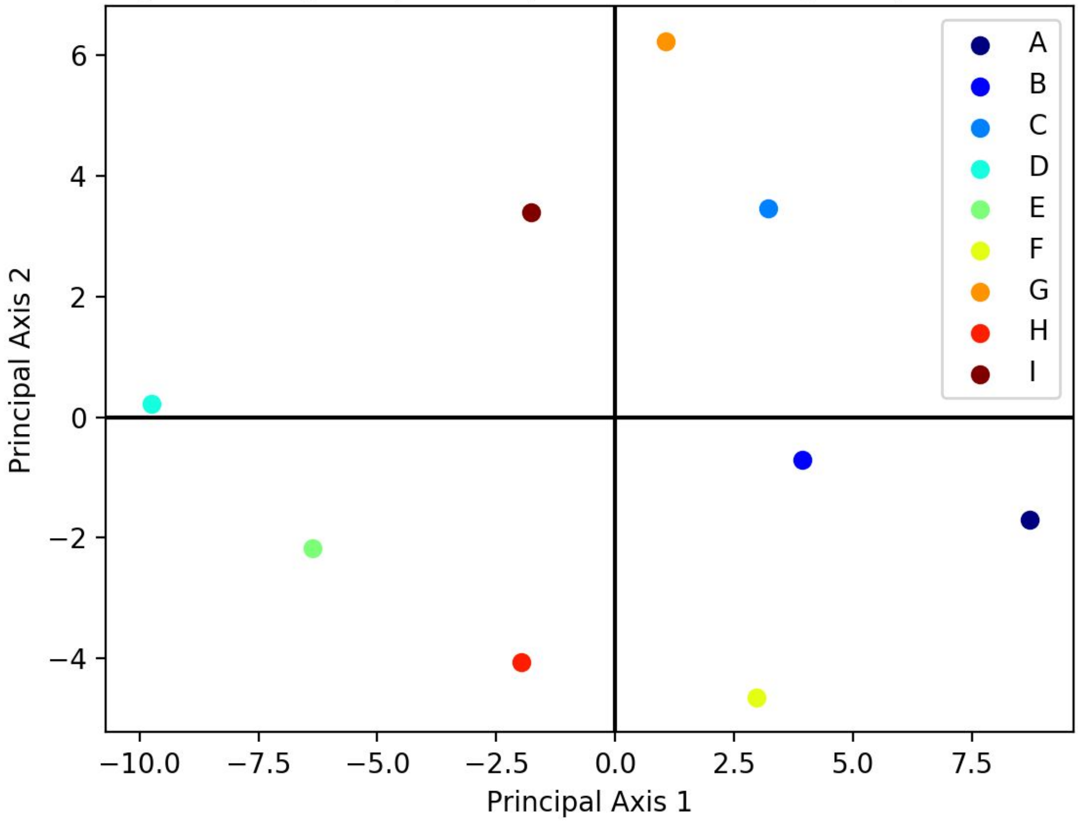
| **Eigenvalue ID** | **Value** |
|---|---|
| 1 | 254.279 |
| 2 | 108.673 |
| 3 | 77.542 |
| 4 | 0.196 |
| 5 | 0.089 |

- The eigenvalues are the multiplicative factors of each eigenvector.
- Ranking in decreasing order of size allows for the identification of the eigenvectors which captures the largest variance in the data.

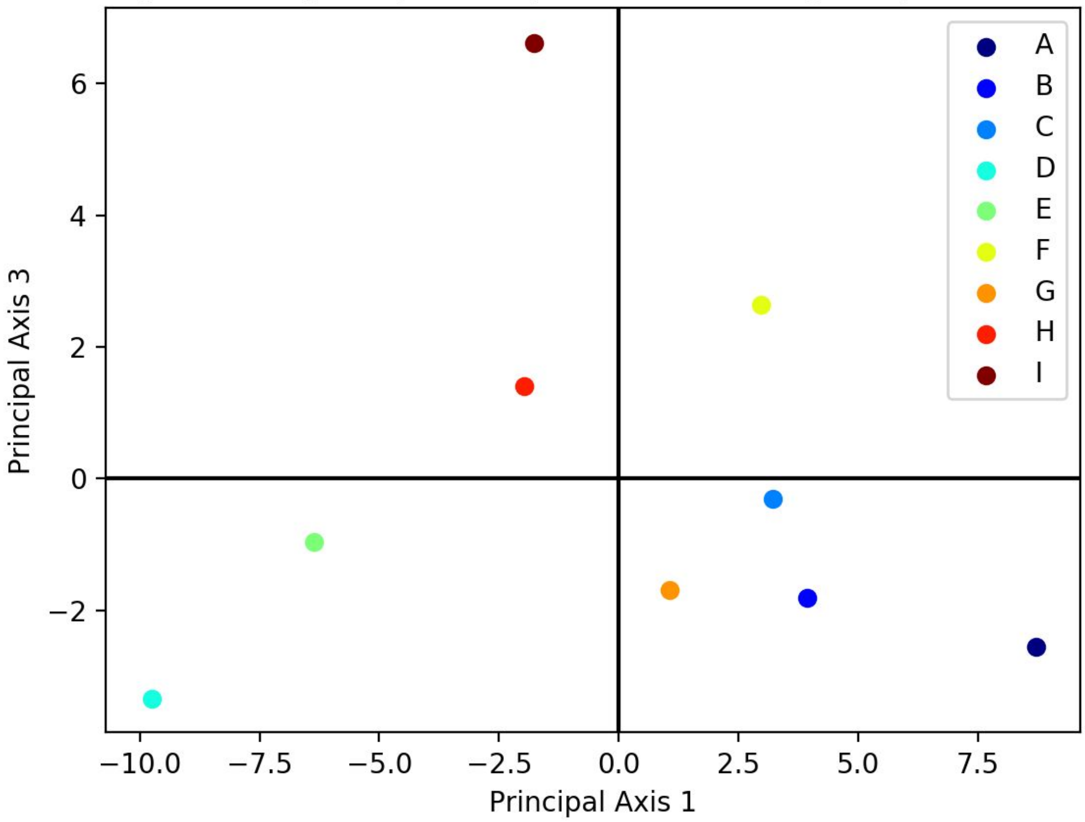
Les nouvelles coordonnées XU en projection sur les axes principaux:

| **Student** | **Axis 1** | **Axis 2** | **Axis 3** | **Axis 4** | **Axis 5** |
|---|---|---|---|---|---|
| A | 8.701 | -1.703 | -2.554 | -0.149 | -0.117 |
| B | 3.939 | -0.709 | -1.81 | -0.091 | 0.043 |
| C | 3.209 | 3.459 | -0.301 | 0.173 | 0.019 |
| D | -9.756 | 0.216 | -3.344 | -0.173 | 0.1 |
| E | -6.371 | -2.173 | -0.957 | 0.071 | -0.188 |
| F | 2.974 | -4.651 | 2.635 | -0.023 | 0.148 |
| G | 1.051 | 6.227 | -1.688 | 0.115 | 0.043 |
| H | -1.981 | -4.069 | 1.401 | 0.243 | 0.01 |
| I | -1.766 | 3.402 | 6.618 | -0.165 | -0.059 |

PCA Projection of principal components Xu onto principal axes 1 and 2




PCA Projection of principal components Xu onto principal axes 1 and 3



## c) Ajustement par un sous espace de $R^n$

Eigenvectors U of the matrix XX':

| Eigenvector ID | Dim 1 | Dim 2 | Dim 3 | Dim 4 | Dim 5 | Dim 6 | Dim 7 | Dim 8 | Dim 9 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | -0.546 | -0.163 | 0.29 | -0.338 | 0.394 | -0.256 | -0.406 | 0.078 | 0.227 |
| 2 | -0.247 | -0.068 | 0.206 | -0.205 | -0.146 | 0.795 | -0.287 | 0.043 | -0.039 |
| 3 | -0.201 | 0.332 | 0.034 | 0.39 | -0.065 | 0.299 | -0.279 | -0.277 | 0.038 |
| 4 | 0.612 | 0.021 | 0.38 | -0.392 | -0.337 | -0.039 | -0.393 | -0.05 | 0.122 |
| 5 | 0.4 | -0.208 | 0.109 | 0.16 | 0.631 | 0.406 | -0.018 | 0.452 | 0.485 |
| 6 | -0.187 | -0.446 | -0.299 | -0.052 | -0.497 | 0.136 | -0.01 | 0.537 | 0.615 |
| 7 | -0.066 | 0.597 | 0.192 | 0.261 | -0.144 | -0.037 | -0.179 | 0.634 | 0.546 |
| 8 | 0.124 | -0.39 | -0.159 | 0.55 | -0.035 | -0.101 | -0.596 | -0.126 | -0.074 |
| 9 | 0.111 | 0.326 | -0.752 | -0.373 | 0.199 | 0.129 | -0.364 | 0.044 | 0.121 |

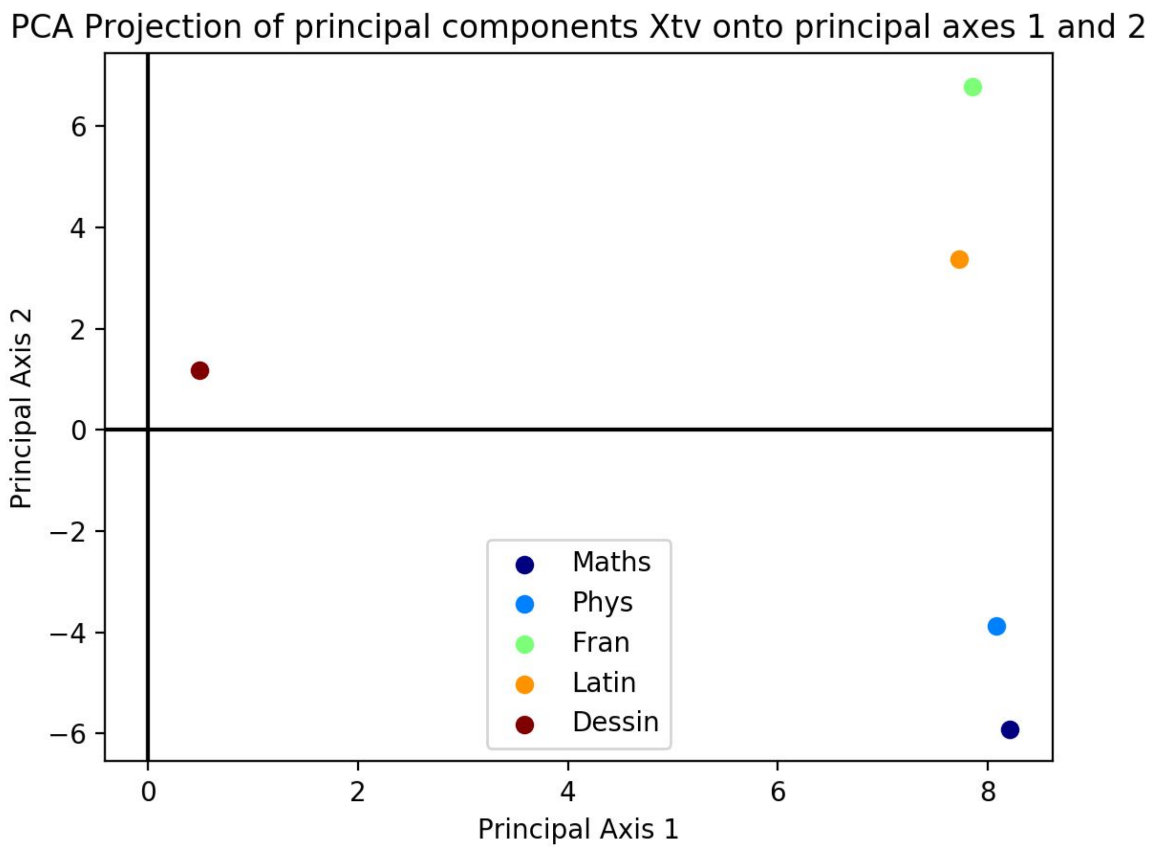
Eigenvalues λ of the matrix XX':

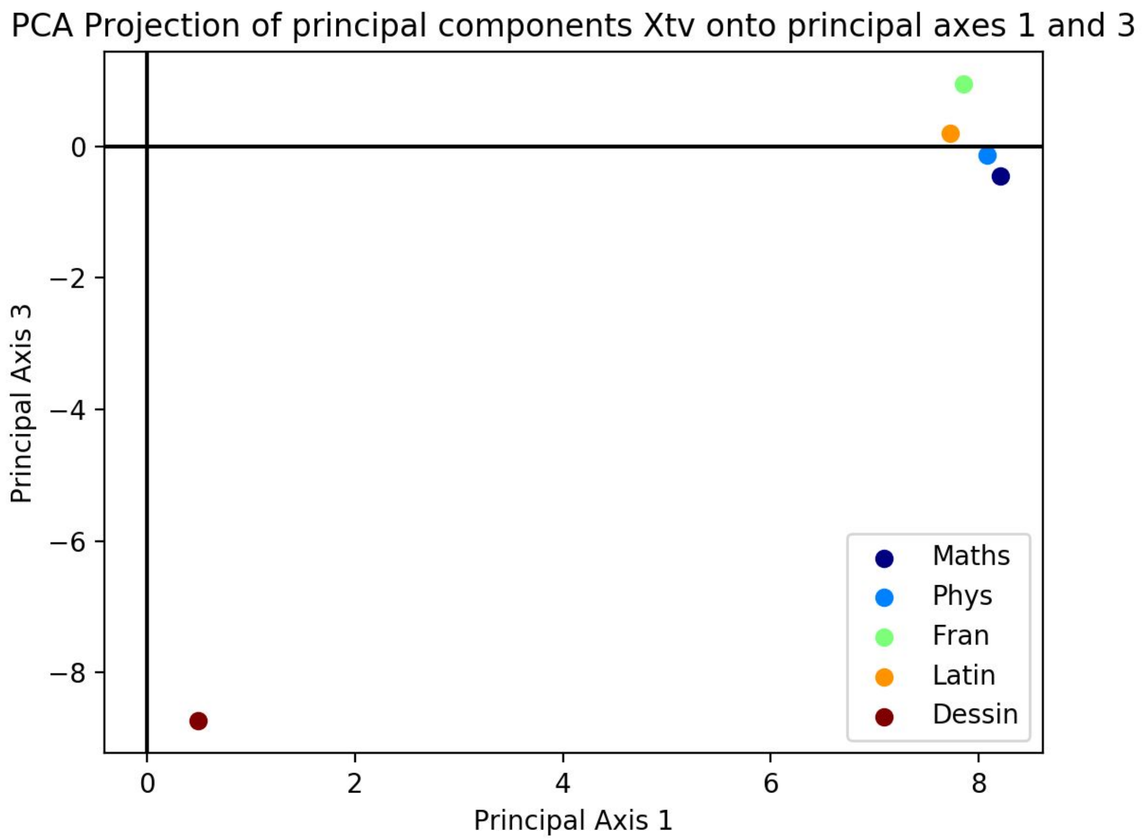
| Eigenvalue ID | Value |
|---|---|
| 1 | 254.279 |
| 2 | 108.673 |
| 3 | 77.542 |
| 4 | 0.196 |
| 5 | 0.089 |
| 6 | 0.0 |
| 7 | 0.0 |
| 8 | -0.0 |
| 9 | -0.0 |

- The first 5 eigenvalues are identical to the eigenvalues of X'X

Les nouvelles coordonnées X'V en projection sur les axes principaux:

| Subject | Axis 1 | Axis 2 | Axis 3 | Axis 4 | Axis 5 |
|---|---|---|---|---|---|
| Maths | 8.205 | -5.91 | -0.452 | -0.128 | -0.171 |
| Phys | 8.085 | -3.878 | -0.127 | 0.245 | 0.163 |
| Fran | 7.851 | 6.78 | 0.952 | 0.174 | -0.122 |
| Latin | 7.728 | 3.37 | 0.199 | -0.298 | 0.135 |
| Dessin | 0.488 | 1.177 | -8.739 | 0.015 | -0.004 |

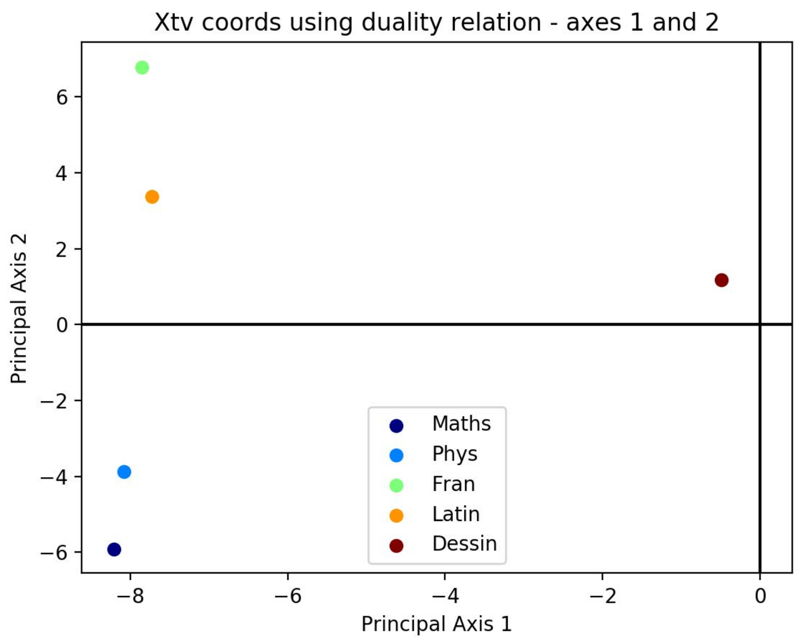

PCA Projection of principal components Xtv onto principal axes 1 and 2
Principal Axis 1
Principal Axis 2
Maths
Phys
Fran
Latin
Dessin

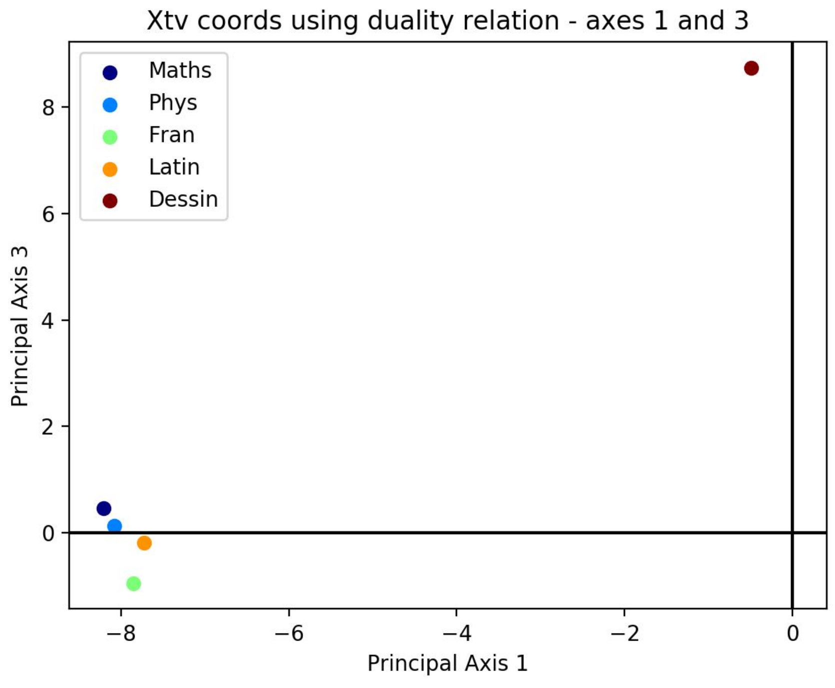

PCA Projection of principal components Xtv onto principal axes 1 and 3
Principal Axis 1
Principal Axis 3
Maths
Phys
Fran
Latin
Dessin

## d) Relation duales

Coordinates X'V of the variables in the base of principal components, calculated as $U_i * \mathrm{sqrt}(\lambda_i)$:

| Subject | Axis 1 | Axis 2 | Axis 3 | Axis 4 | Axis 5 |
|---|---|---|---|---|---|
| Maths | -8.205 | -5.91 | 0.452 | -0.128 | 0.171 |
| Phys | -8.085 | -3.878 | 0.127 | 0.245 | -0.163 |
| Fran | -7.851 | 6.78 | -0.952 | 0.174 | 0.122 |
| Latin | -7.728 | 3.37 | -0.199 | -0.298 | -0.135 |
| Dessin | -0.488 | 1.177 | 8.739 | 0.015 | 0.004 |




- They are identical to the previous two graphs, except the axis are flipped.

Eigenvectors V, recalculated from $XU_i$ / sqrt($\lambda_i$):

| Eigenvector ID | Dim 1 | Dim 2 | Dim 3 | Dim 4 | Dim 5 |
|---|---|---|---|---|---|
| 1 | 0.546 | -0.163 | -0.29 | -0.338 | -0.394 |
| 2 | 0.247 | -0.068 | -0.206 | -0.205 | 0.146 |
| 3 | 0.201 | 0.332 | -0.034 | 0.39 | 0.065 |
| 4 | -0.612 | 0.021 | -0.38 | -0.392 | 0.337 |
| 5 | -0.4 | -0.208 | -0.109 | 0.16 | -0.631 |
| 6 | 0.187 | -0.446 | 0.299 | -0.052 | 0.497 |
| 7 | 0.066 | 0.597 | -0.192 | 0.261 | 0.144 |
| 8 | -0.124 | -0.39 | 0.159 | 0.55 | 0.035 |
| 9 | -0.111 | 0.326 | 0.752 | -0.373 | -0.199 |

Coordinates X'V of the variables in the base of principal components, calculated as X' * V:

| Subject | Axis 1 | Axis 2 | Axis 3 | Axis 4 | Axis 5 |
|---|---|---|---|---|---|
| Maths | -8.205 | -5.91 | 0.452 | -0.128 | 0.171 |
| Phys | -8.085 | -3.878 | 0.127 | 0.245 | -0.163 |
| Fran | -7.851 | 6.78 | -0.952 | 0.174 | 0.122 |
| Latin | -7.728 | 3.37 | -0.199 | -0.298 | -0.135 |
| Dessin | -0.488 | 1.177 | 8.739 | 0.015 | 0.004 |

- The values are indeed identical to the X'V coordinates, calculated as $U_i$ * sqrt($\lambda_i$). The equivalence of the duality formulae is verified.

## e) Formule de reconstitution

Original centered X values

| Student | Maths | Phys | Fran | Latin | Dessin |
|---|---|---|---|---|---|
| A | -3.6666666667 | -3.8333333333 | -5.2222222222 | -4.5555555556 | -3.0 |
| B | -1.6666666667 | -1.8333333333 | -2.2222222222 | -2.0555555556 | -2.0 |
| C | -3.6666666667 | -2.8333333333 | 0.7777777778 | -0.5555555556 | 0.0 |
| D | 4.8333333333 | 4.6666666667 | 5.2777777778 | 4.9444444444 | -3.0 |
| E | 4.3333333333 | 4.1666666667 | 1.7777777778 | 2.4444444444 | -1.0 |
| F | 1.3333333333 | 0.1666666667 | -4.7222222222 | -3.0555555556 | 2.0 |
| G | -4.1666666667 | -2.8333333333 | 3.7777777778 | 1.4444444444 | -1.0 |
| H | 3.3333333333 | 2.6666666667 | -1.7222222222 | -0.5555555556 | 1.0 |
| I | -0.6666666667 | -0.3333333333 | 2.2777777778 | 1.9444444444 | 7.0 |

Reconstituted centered X values, using X = V[:, i] * sqrt($\lambda$[i]) * U[:, i]$^t$ - these are identical in all cases.

| Student | Maths | Phys | Fran | Latin | Dessin |
|---|---|---|---|---|---|
| A | -3.6666666667 | -3.8333333333 | -5.2222222222 | -4.5555555556 | -3.0 |
| B | -1.6666666667 | -1.8333333333 | -2.2222222222 | -2.0555555556 | -2.0 |
| C | -3.6666666667 | -2.8333333333 | 0.7777777778 | -0.5555555556 | 0.0 |
| D | 4.8333333333 | 4.6666666667 | 5.2777777778 | 4.9444444444 | -3.0 |
| E | 4.3333333333 | 4.1666666667 | 1.7777777778 | 2.4444444444 | -1.0 |
| F | 1.3333333333 | 0.1666666667 | -4.7222222222 | -3.0555555556 | 2.0 |
| G | -4.1666666667 | -2.8333333333 | 3.7777777778 | 1.4444444444 | -1.0 |
| H | 3.3333333333 | 2.6666666667 | -1.7222222222 | -0.5555555556 | 1.0 |
| I | -0.6666666667 | -0.3333333333 | 2.2777777778 | 1.9444444444 | 7.0 |

Reconstituted centered X values, using first three principal axes only:

| Student | Maths | Phys | Fran | Latin | Dessin |
|---|---|---|---|---|---|
| A | -3.6426521601 | -3.8147727497 | -5.115302948 | -4.7095120691 | -2.9933671093 |
| B | -1.7177569393 | -1.7594140947 | -2.2043419612 | -2.0969700871 | -1.9974282208 |
| C | -3.6278851452 | -2.9182244193 | 0.7019400785 | -0.4304847284 | -0.0061860621 |
| D | 4.7257514887 | 4.8174679092 | 5.3049315245 | 4.8730224588 | -2.9952983278 |
| E | 4.4613718438 | 4.0248761368 | 1.8269912837 | 2.4068422494 | -1.0000518172 |
| F | 1.2418386793 | 0.2604172246 | -4.7737692223 | -3.0040538819 | 1.9989232504 |
| G | -4.1578872428 | -2.8737078686 | 3.714838076 | 1.5415889604 | -1.0045127569 |
| H | 3.3976074044 | 2.5378373644 | -1.8222440525 | -0.3868685907 | 0.9914928892 |
| I | -0.6803879289 | -0.2744795027 | 2.3669572215 | 1.8064356885 | 7.0064281545 |

Residuals between reconstituted centered X values using first three principal axes only, and original centered X values:

| Student | Maths | Phys | Fran | Latin | Dessin |
|---|---|---|---|---|---|
| A | -0.024 | -0.019 | -0.107 | 0.154 | -0.007 |
| B | 0.051 | -0.074 | -0.018 | 0.041 | -0.003 |
| C | -0.039 | 0.085 | 0.076 | -0.125 | 0.006 |
| D | 0.108 | -0.151 | -0.027 | 0.071 | -0.005 |
| E | -0.128 | 0.142 | -0.049 | 0.038 | 0.0 |
| F | 0.091 | -0.094 | 0.052 | -0.052 | 0.001 |
| G | -0.009 | 0.04 | 0.063 | -0.097 | 0.005 |
| H | -0.064 | 0.129 | 0.1 | -0.169 | 0.009 |
| I | 0.014 | -0.059 | -0.089 | 0.138 | -0.006 |

# B – ACP NORMEE et cercle des corrélations

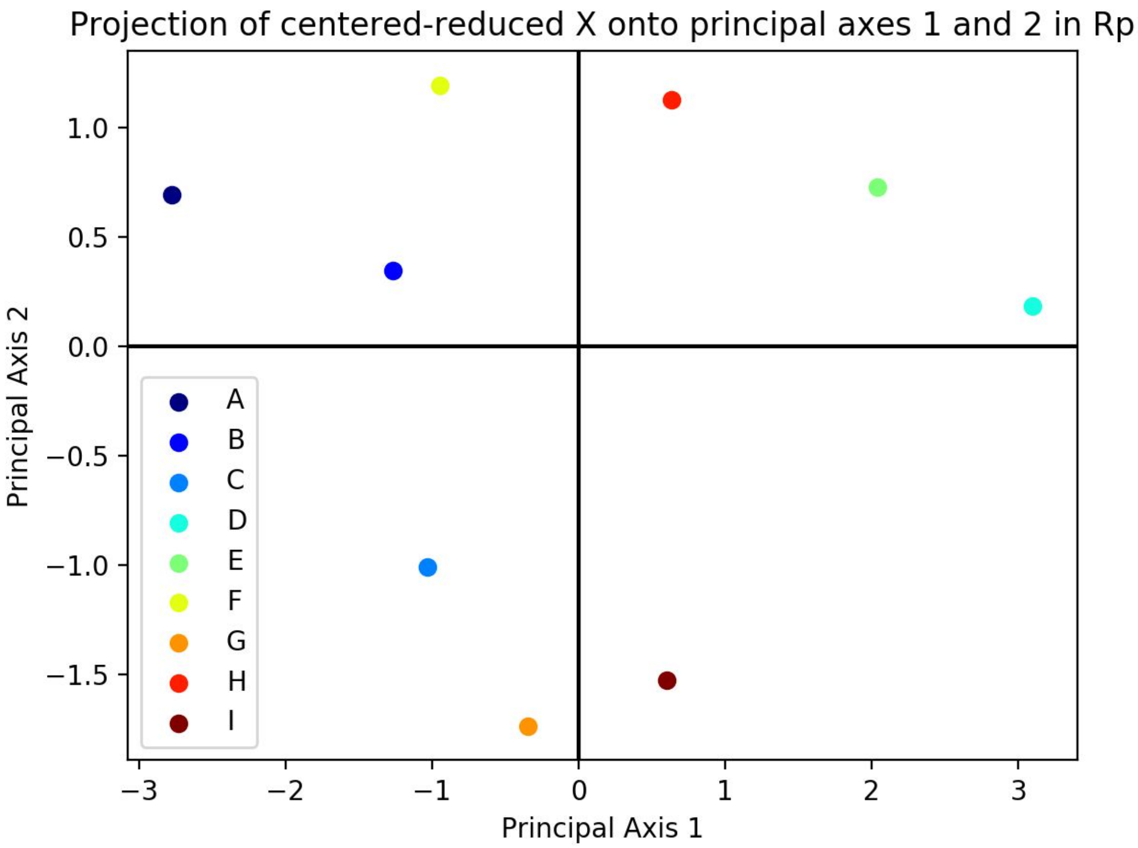
## a) Codage des données

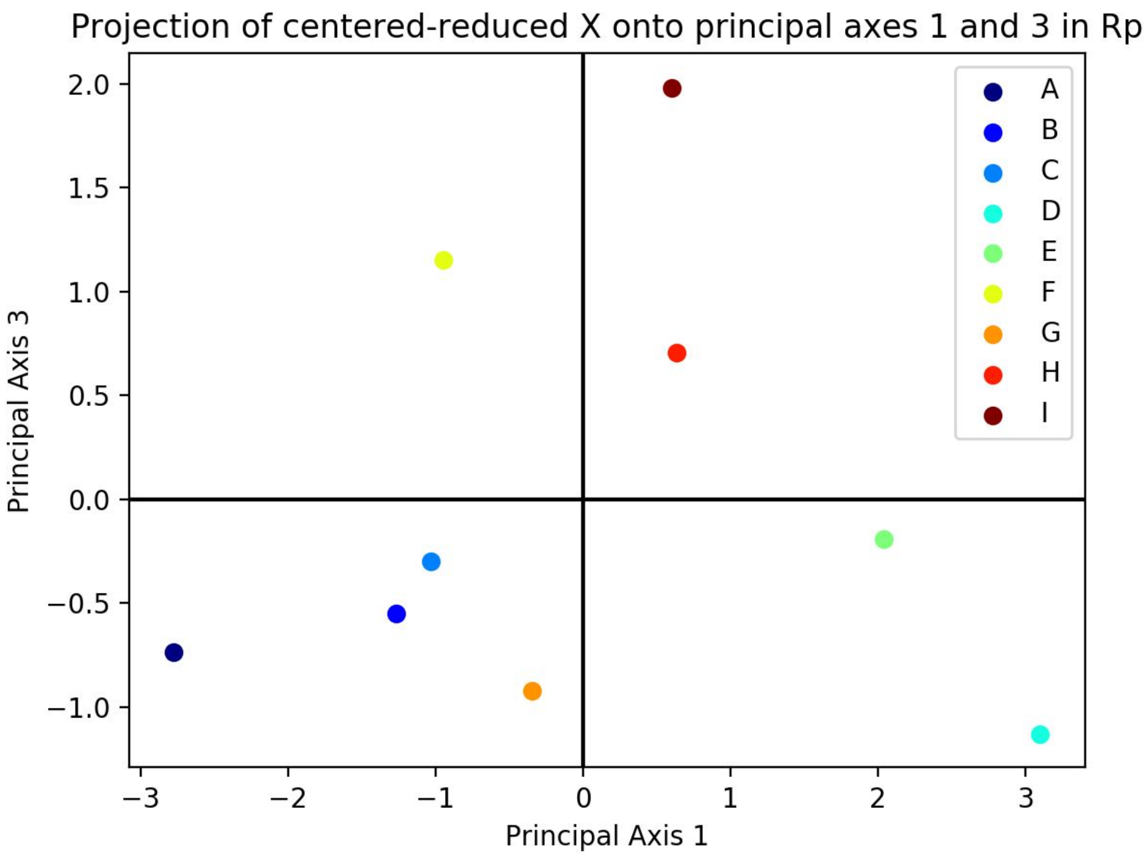
- Normed analysis is
- It is used when
- You mean normalise the data by subtracting the mean of the feature from each of its values. Then you scale the values by dividing by the variance of the feature, which causes each feature variable to have unit variance. This is useful when feature variables are on very different scales, where if it were not for feature scaling, one variable would heavily dominate the calculations and prevent an accurate comparison. Dividing the centered data by the standard deviation for each variable produces a transformed data set whose covariance matrix is the correlation matrix of the original centered data.

Correlation matrix of X:

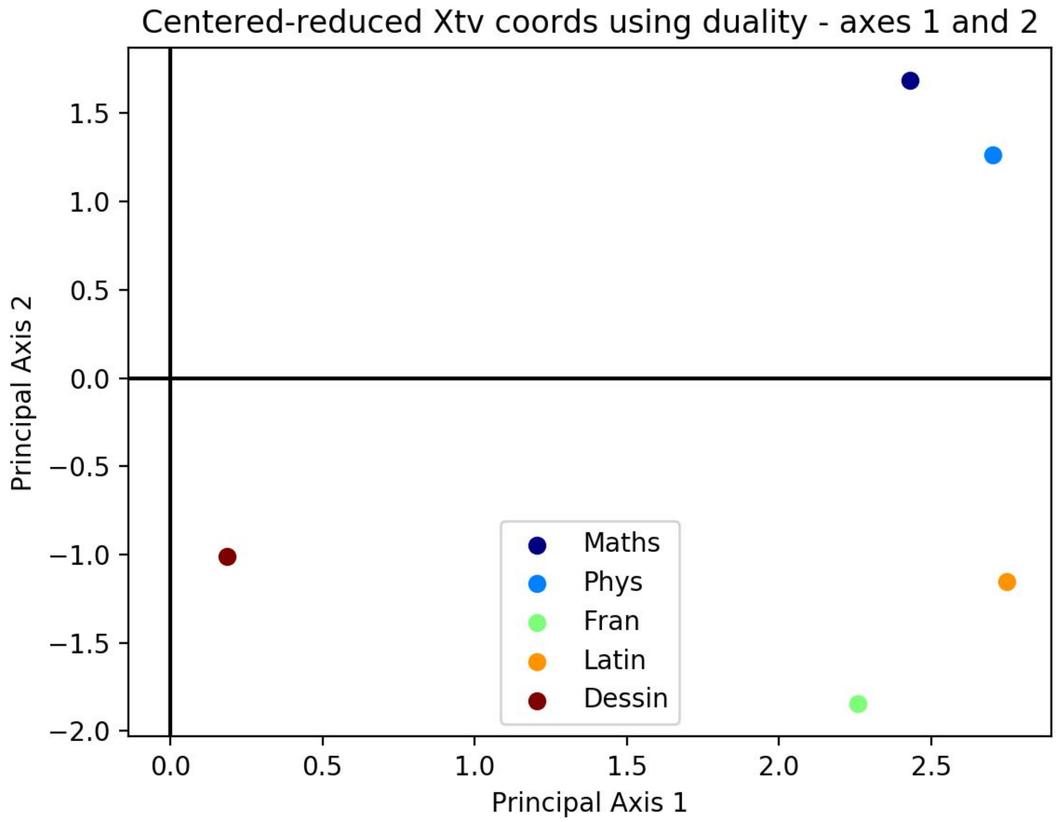
| Variable | Maths | Phys | Fran | Latin | Dessin |
|---|---|---|---|---|---|
| Maths | 1.0 | 0.983 | 0.227 | 0.508 | 0.011 |
| Phys | 0.983 | 1.0 | 0.397 | 0.652 | 0.006 |
| Fran | 0.227 | 0.397 | 1.0 | 0.951 | 0.038 |
| Latin | 0.508 | 0.652 | 0.951 | 1.0 | 0.081 |
| Dessin | 0.011 | 0.006 | 0.038 | 0.081 | 1.0 |

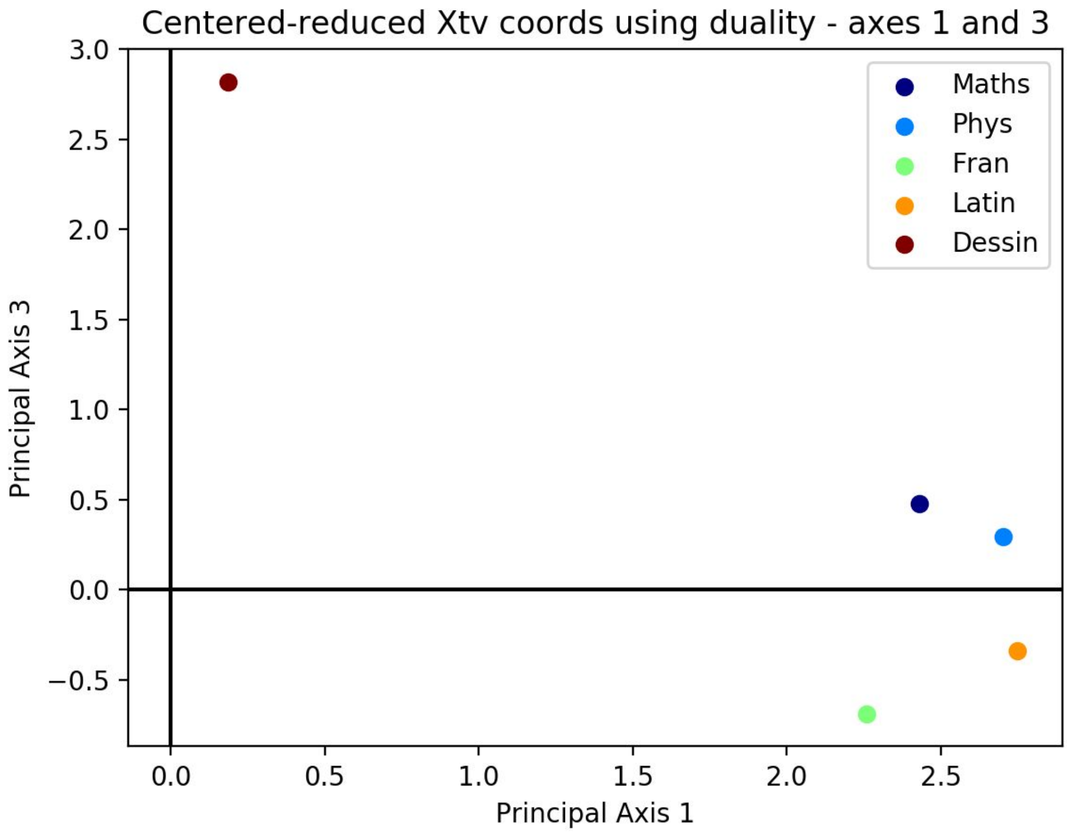
**Ajustement par un sous espace de $R^p$**



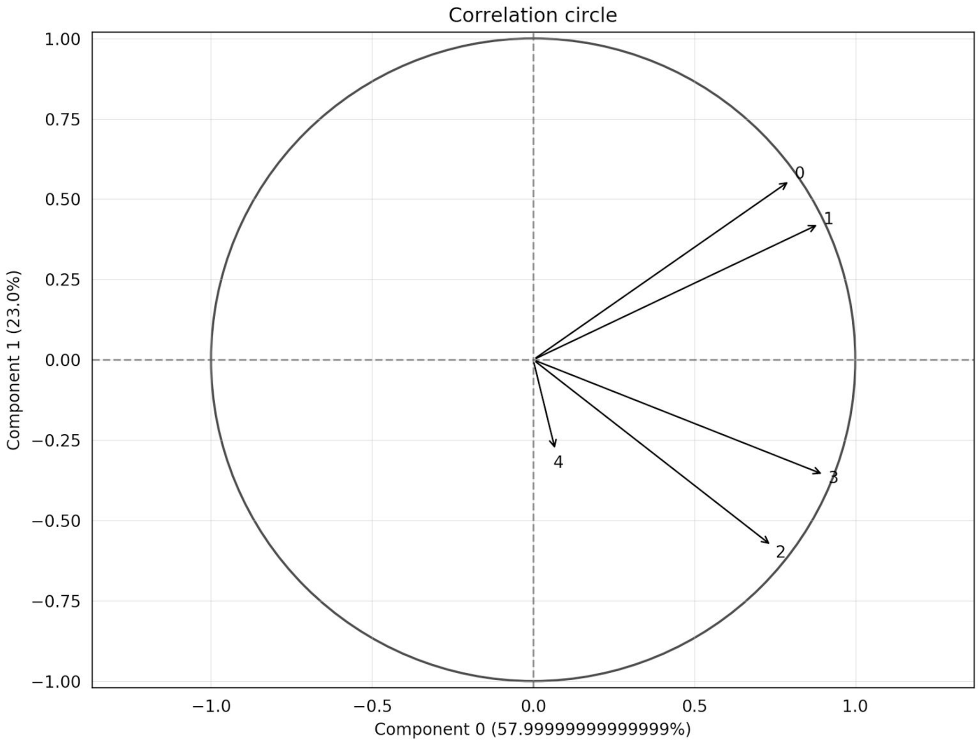
**c) Relations duales**



## d) Cercle des corrélations

Projection onto axes 1 and 2:




Projection onto axes 1 and 3: