<u>Assignment 2</u> <u>EOSC 410, Angelene Leow</u>

<u>PROBLEM 1</u>

(a) Time series for each variable

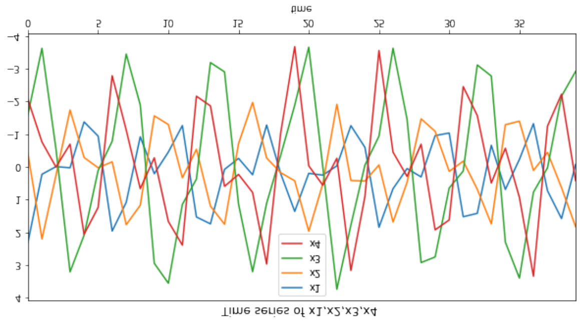

Figure 1

(b) Perform PCA

Eigenvectors obtained from each mode are:

[[-0.22466149 -0.33931661 0.8266404   0.38866064]
 [0.42848482 -0.35586436 0.32893107 -0.76260453]
 [0.78244359 -0.31022927 -0.15751461 0.51645811]
 [0.39205867 0.81362313 0.42855471 0.025461]]

where each column represents 1 eigenvector corresponding to each mode.

percentage of each mode are:

[68.64151021 29.8159161   1.10174125 0.44083245]

whereby mode 1 explains 68.64% of the variance, mode 2 explains 29.82%, mode 3 explains 1.1% and mode4 explains 0.44%.

(c) Determine PCA modes to keep

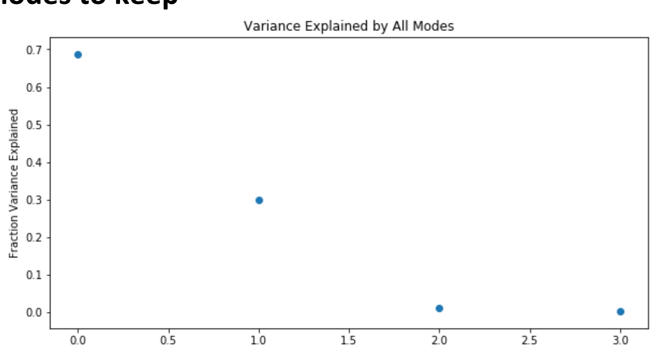

Figure 2

Figure 2 shows the fraction of variance explained with relation to the mode number.

Since mode 1 and mode 2 sums up to 98.46% of the variance explained (>95%), we can say that the first two modes should be kept to reconstruct data.

The corresponding eigenvectors are:

```
e1 = [-0.22466149  0.42848482  0.78244359  0.39205867]
e2 = [-0.33931661 -0.35586436 -0.31022927  0.81362313]
```

(d) PC's of significant modes

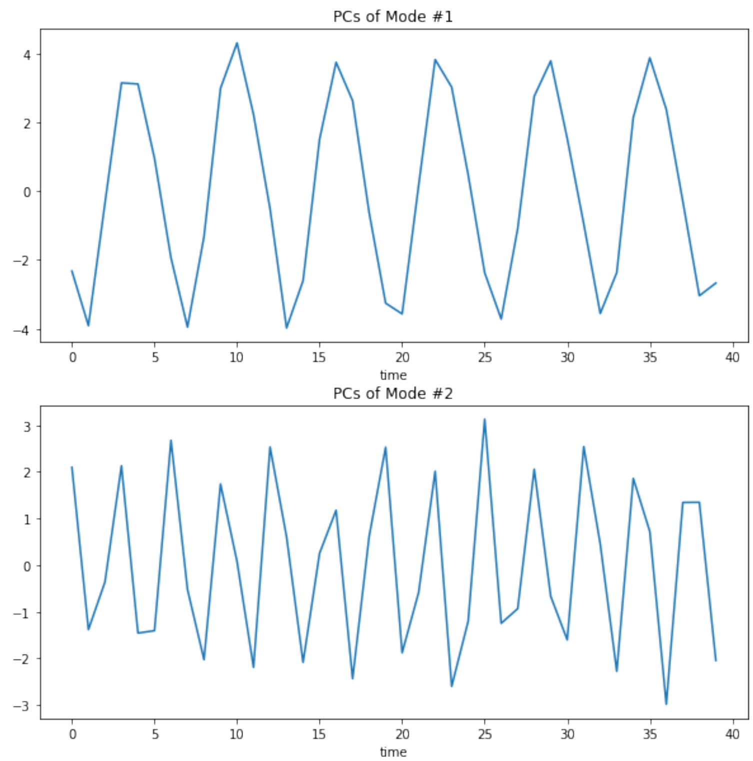

Figure 3

PC1 PC1 and PC2 both show periodic oscillation. PC1 oscillates with a period T while PC2 oscillates with a period of ½ T. The PCs show how strongly the influence each point in time.

(e) PC1 vs PC2 plot

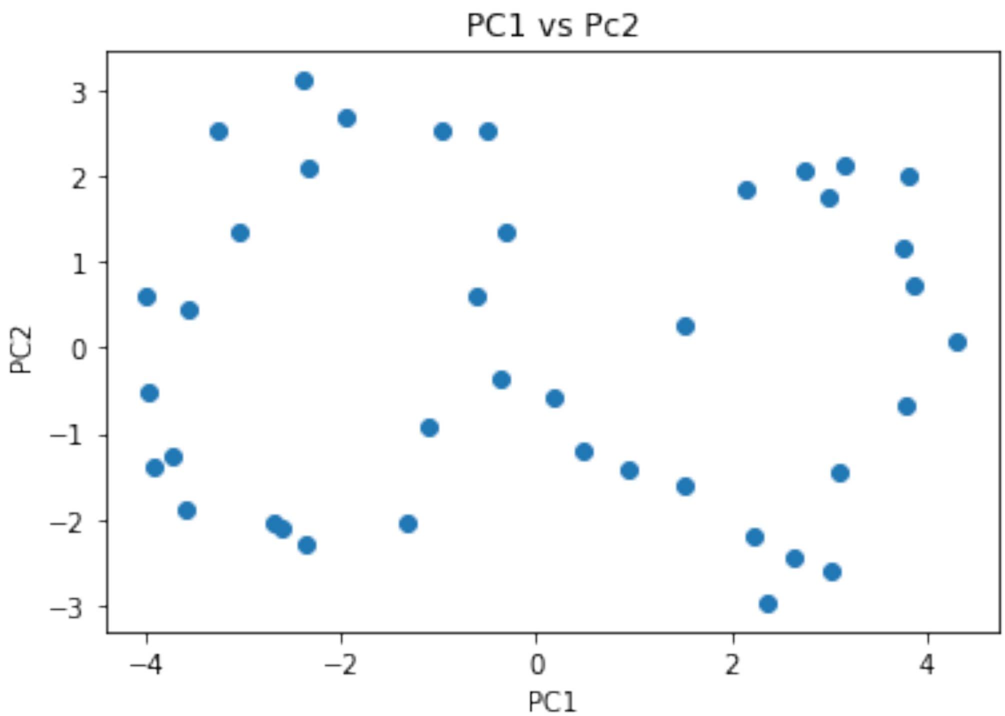

Figure 4

The 2 PCs seems to map out to look like an infinity loop. However after prolonged thoughts over sleepless nights, there does not seem to be any correlation between PC1 and PC2.

PROBLEM 2:

(a)

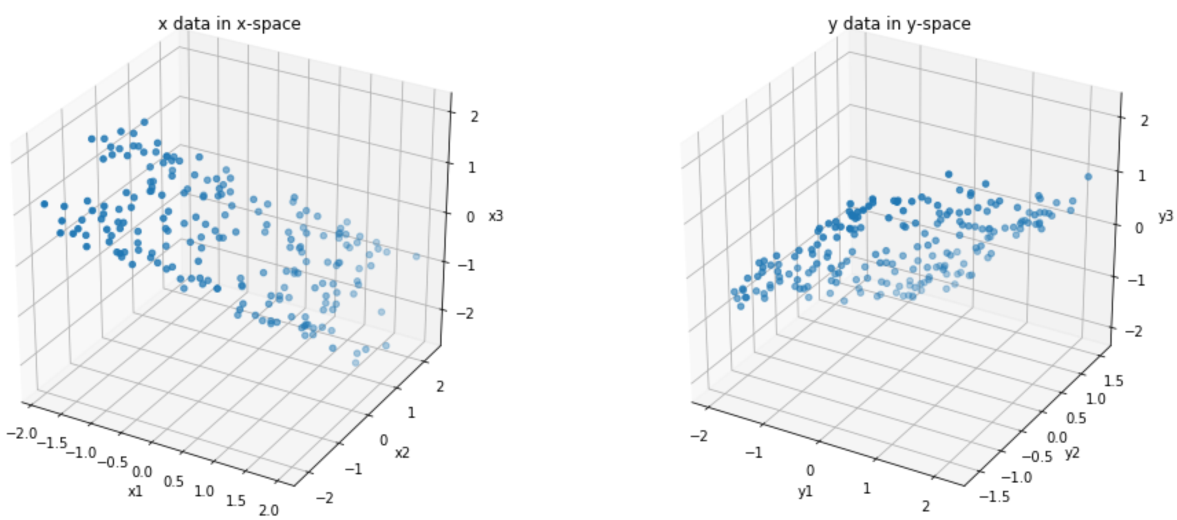

Figure 5

(b) Performing CCA

Shape of x: 200 rows, 3 columns
Shape of y: 200 rows, 3 columns
r = 0.9869 ,0.9199, 0.1178 for each mode.

r value shows the correlation coefficient of data between x and y for each 3 modes.
r1 and r2 shows high correlation.

```
>> A
[[ 0.02347783  0.92506524 -0.37908191]
 [ 0.6981136   0.25625102  0.66856026]
 [ 0.71560198 -0.28033858 -0.6397844 ]]

>> B
[[-0.47294885  0.61636613 -0.62961272]
 [-0.74645976  0.09934212  0.65797338]
 [ 0.46809957  0.78116831  0.41310878]]
```

where the first column of **A** is the first CCA mode vector **F1** in the x-space, and the first column of **B** corresponds to the first CCA mode vector **G1** in the y- space, similarly with the second column corresponding to mode 2 and so on.

After computing the covariance matrices C_{xx} and C_{yy},

$F = C_{xx} A$

$G = C_{yy} B$

```
>> F
[[ 0.00446775  1.38448709  0.74085775]
 [ 0.36774811  1.18278841  0.65839886]
 [ 0.35360286 -1.20036722 -0.64154713]]

>> G
[[-0.60781613  1.04238248 -0.11125062]
 [-1.21694479 -0.0053953   0.14002947]
 [ 0.62447444  1.02810331 -0.23102376]]
```

where the first column of **F** gives the direction of the CCA mode 1 in the x-space, and the first column of **G** gives the direction of the CCA mode 1 in the y-space, similarly with the second column corresponding to mode 2 and so on.

(c) We chose the first 2 modes to be plotted since they have high correlation(r values). Figure 6 below shows F1 and F2 in the x-space, G1 and G2 in the y-space.

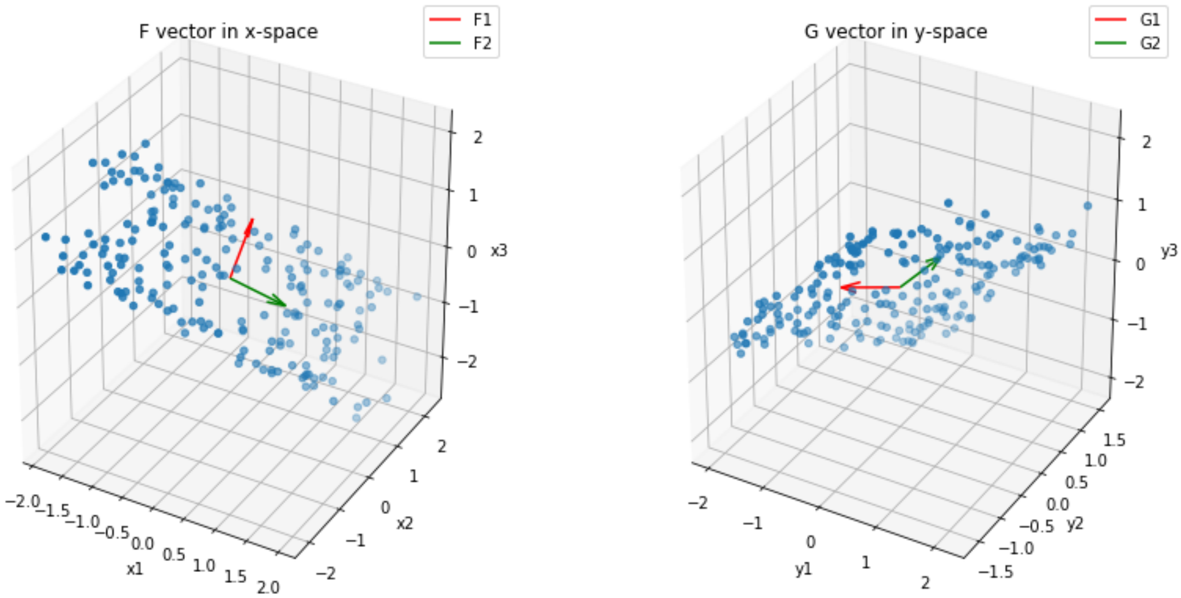

Figure 6

(d) Plot of V(t) versus U(t) for all modes

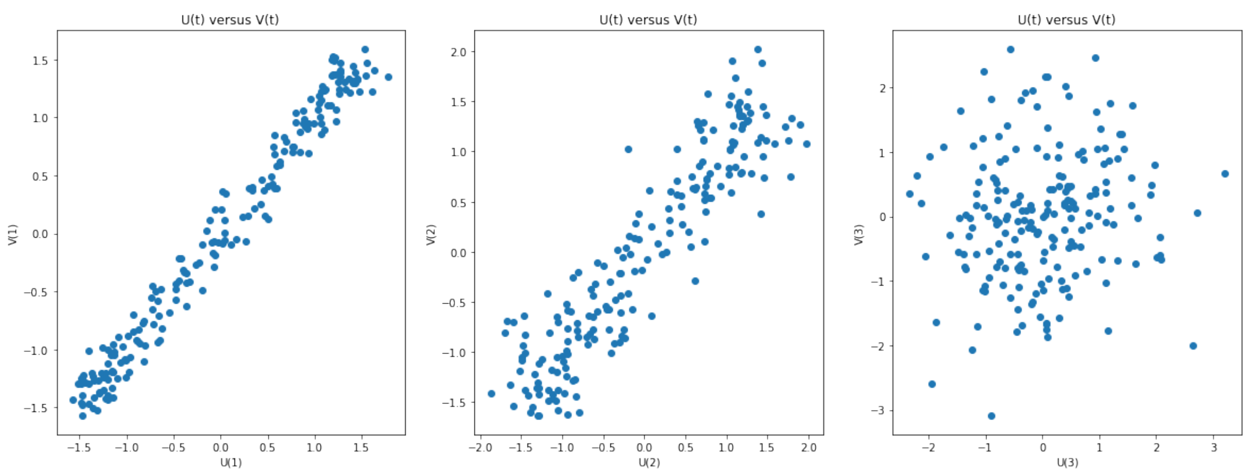

Figure 7

The first two subplots show high correlation for both mode 1 and mode 2, and this result aligns with our correlation coefficient, r , that mode 1 and mode 2 are significant when comparing between x and y datasets, while mode 3 shows low correlation between the two datasets (r = 0.1178).

(e) Plotting F1, F2 and G1, G2 vectors with their corresponding eigenvectors.

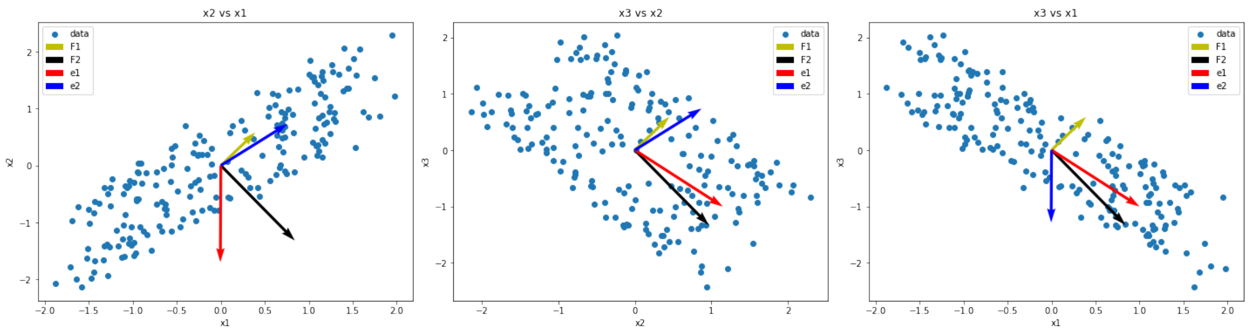

Figure 8

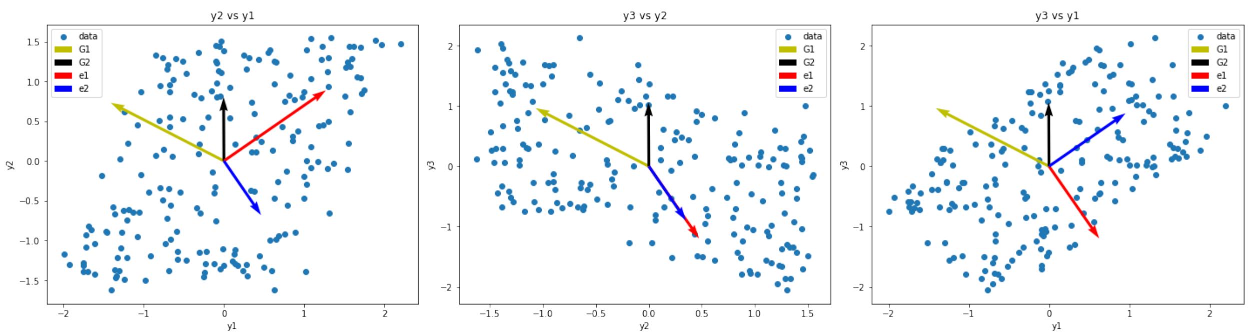

Figure 9

We are only plotting vectors from mode 1 and 2 since they are significant modes for the comparison of x and y data.
From Figure 8, 1 and e2 behave similarly (points approximately in the same direction) when plotted in the x1 and x2 space.
In the second subplot, F1 and e2 behave similarly, as well as F2 and e1 when plotted in the x2 and x3 space.
F2 and e1 also behave similarly when plotted in the x1 and x3 space.

From Figure 9, none of the CCA directional modes (G1 and G2) behave in a similar way corresponding to the eigenvectors in y1, y2 and y3 spaces.

Hence we can conclude that the CCA modes fits the x dataset better. This indicates that the CCA modes in the x dataset correspond to the modes of largest variability in the x-data.