Assignment 4 – Solutions

1. A) Data provided '**steamdata.mat'.** The given data has *n* (28) variables and *N* (1000) samples. *IPCA* is performed to iteratively compute the error covariance matrix and the number of constraints (*m*) are determined if the number of eigenvalues close to 1 is equal to chosen *m* for the iterative procedure.
   - Since the error covariance matrix is assumed to be diagonal and *n* = 28, we need at least 7 constraints as $\frac{m(m+1)}{2} \geq n$. The initial *m* is chosen from *m* = *n-1* (i.e. 27) for the first *IPCA* and is progressively reduced till *m* = 7.
   - An identity matrix is chosen as the initial error covariance matrix and the values of the new error variances are updated to perform the *MLPCA* till convergence. The user defined function '**stdest.m'** is used for this procedure.

Table 1: Eigen values corresponding to the number of constraints

No of Constraints	M=7	M=10	M=11	M=17	M=21	M=25
1	576493.7	1438833	1456005.9	982522.5	2101955.2	1959.7
2	274845.5	868286.1	663480.2	316922.8	136433.7	90.3
3	161141.5	139240.5	123211.2	46936.4	31905.6	27.3
4	31026.8	77922.5	63942.4	17971	12862.7	5.83
5	18873.3	43918.2	44990	16105	35.3	4
6	11931	29940.4	27318.8	6853.1	20.6	3.78
7	7885.7	16987.5	15546.1	1701.8	11.5	2.43
8	5099.8	14961.9	13754.2	50.6	4	2.08
9	3528.6	8497.5	8636.2	21.2	3.2	1.45
10	2734.8	7730.5	5983.2	17.2	2.6	1.10
11	2252	6720.4	4107	12.5	2.03	0.986
12	1150.2	4323.9	3545	3.7	1.5	0.791
13	904.1	2946.7	2825.8	2.1	1.35	0.598
14	455	2781.1	2102.6	1.9	1.1	0.470
15	392.9	1853.3	1648.2	1.8	0.856	0.422
16	194.5	1497.1	1557.1	1.1	0.65	0.316
17	147.2	1338.1	1269.5	1.02	0.55	0.261
18	1.01	1.639	1.132	0.814	0.4	0.001
19	0.998	1.074	1.083	0.658	0.353	0.0002
20	0.976	1.06	1.049	0.555	0.0018	0.0002
21	0.732	1.031	1.044	0.003	0..001	0.0002
22	0.235	1.024	1.013	0.002	0.0007	0.0002
23	0.197	1.004	0.995	0.0019	0.0004	0.0001
24	0.110	0.987	0.973	0.0013	0.0003	0.0001
25	0.049	0.979	0.972	0.001	0.0003	0.0001

26	0.041	0.972	0.936	0.0007	0.0003	0.00007
27	0.0346	0.940	0.916	0.0006	0.0002	0.00006
28	0.0284	0.891	0.887	0.0006	0.0002	0.00003

- At PCs = 17, the number of eigenvalues ≈ 1 coincides with m = 11. The eigenvalues obtained for other constraint values are listed in Table 1.

The standard deviation of the variables are identified as (for m = 11):

Variable No:	Stand Error		Variable No:	Stand Error		Variable No:	Stand Error
1	0.1059		8	0.0984		15	0.1077
2	0.1005		9	0.0727		16	0.0854
3	0.1037		10	0.0961		17	0.0995
4	0.1053		11	0.1031		18	0.1036
5	0.0938		12	0.1046		19	0.0961
6	0.1098		13	0.1024		20	0.1033
7	0.1000		14	0.0811		21	0.0903

Variable No:	Stand Error
22	0.1065
23	0.1020
24	0.0969
25	0.1103
26	0.1050
27	0.0975
28	0.0876

This technique correctly identifies the model order and the number of unity Eigen values equals the number of constraints.

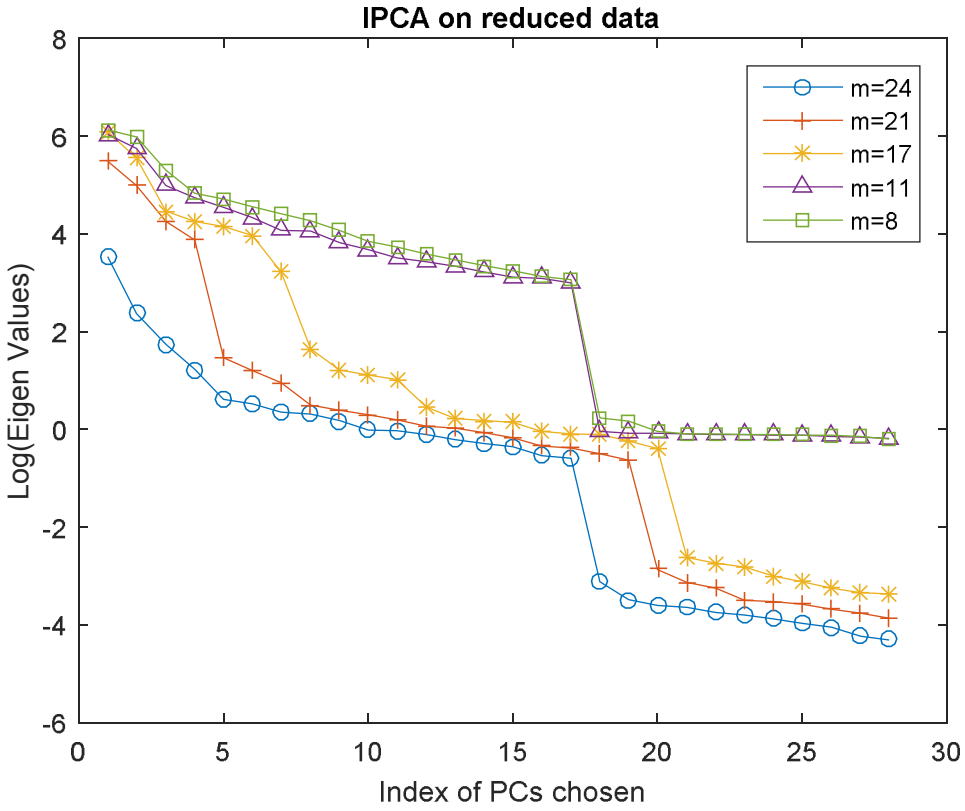
B) The regression matrix is given by $B = -(U_D)^{-1}(U_I)$ and the maximum abs difference $\max(|B_{est} - B_{true}|)$ is calculated as **0.0063.**

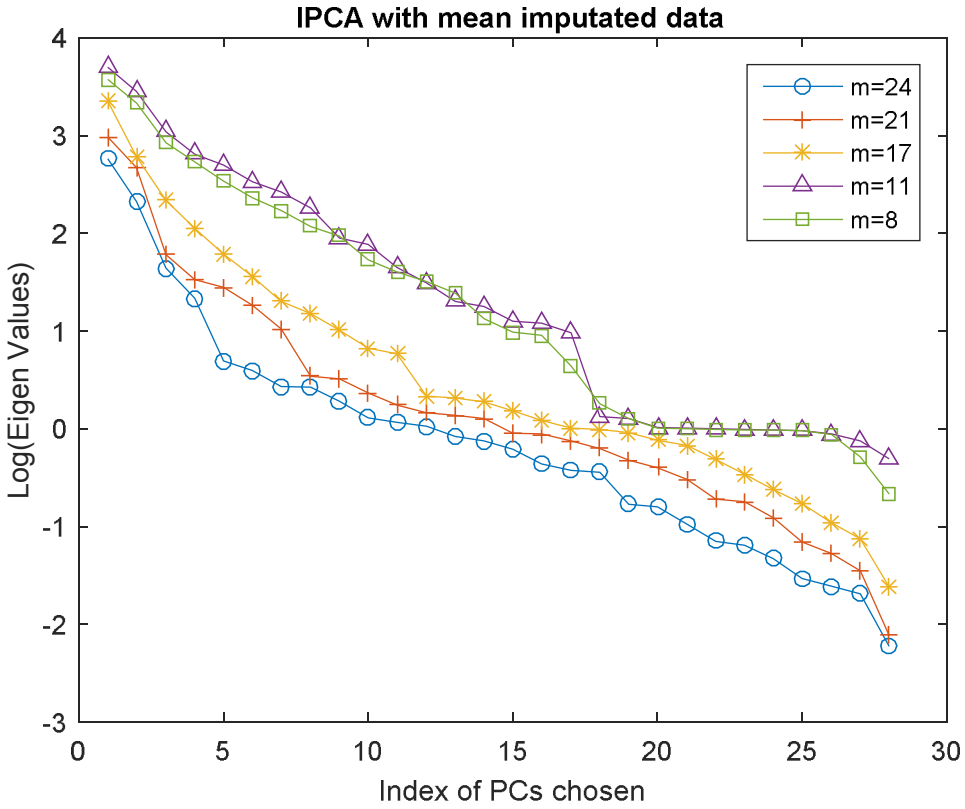
2. In this problem, some of the samples have *NaN* measurements and performance of the different strategies, employed using *PCA* and *IPCA* methods, are evaluated using the maximum abs difference between the estimated and the true regression matrix. The computed values are shown in Table 2 (for m = 11). Convergence criteria used is

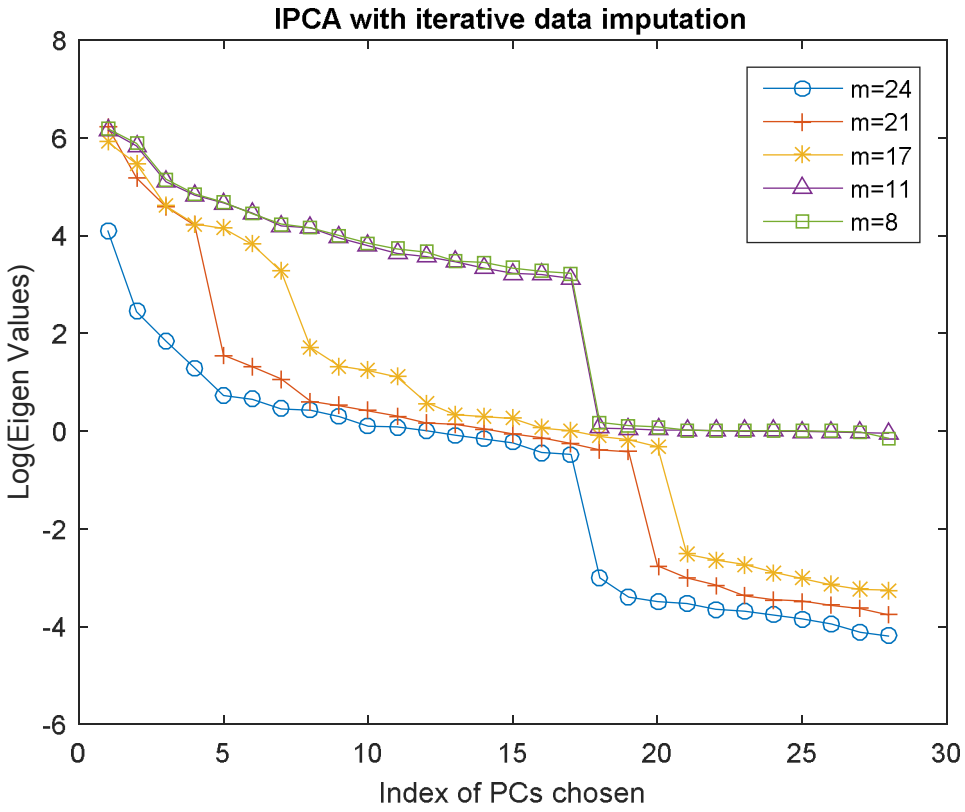
Table 2: Max absolute values corresponding to different methods

	PCA	IPCA
Reduced data	0.0082	0.0075
Mean imputed sampled data	0.1685	0.1758
Iterative imputation using PCA	0.0066	0.0060

The graphical results are as follows:




IPCA with mean imputated data
m=24
m=21
m=17
m=11
m=8
Log(Eigen Values)
Index of PCs chosen
4
3
2
1
0
-1
-2
-3
0
5
10
15
20
25
30


IPCA with iterative data imputation
m=24
m=21
m=17
m=11
m=8
Log(Eigen Values)
Index of PCs chosen
8
6
4
2
0
-2
-4
-6
0
5
10
15
20
25
30