Take-Home

Name-Shreyas Kulkarni

EAS-502

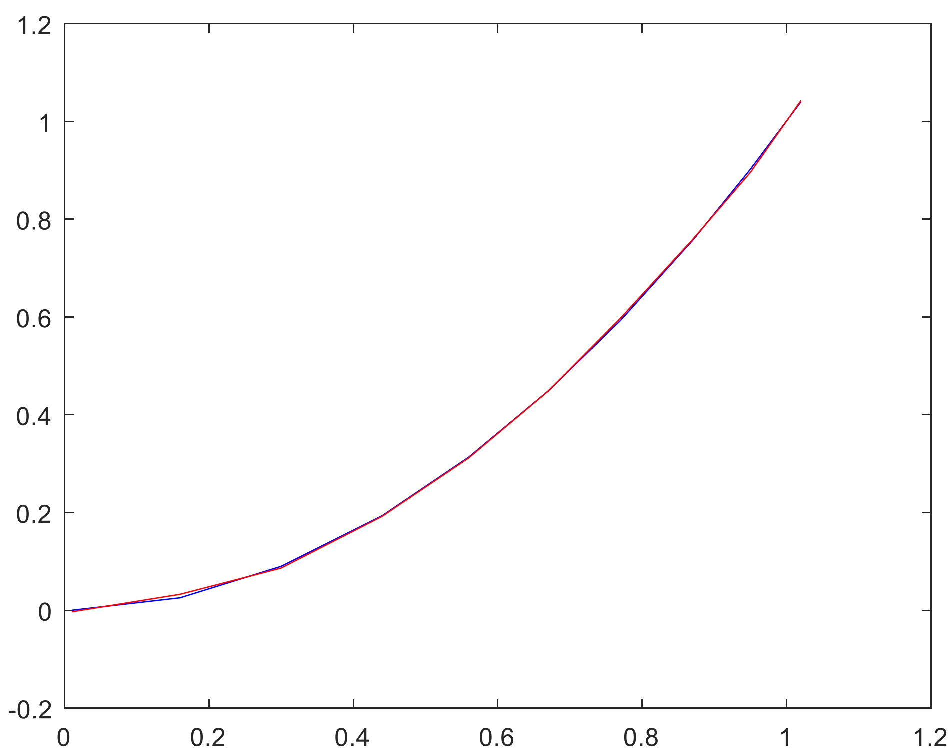
UBID- sk385

Q.1)

ay^2 + bxy + cx + dy + e = x^2

For this equation, we need to determine coefficients a, b, c, d, e for which ∑(a*y^2 + b*x*y + c*x + d*y + e - x^2) is minimum. i.e. difference between calculated value and actual value should be least. In matlab there is function which calculates coefficients for linear regression. For that we created a matrix of elements y^2, x*y, x, y and 1. We get 10*5 matrix X. Performed linear regression of X with Y(=x^2). So now we calculate output using these coefficients which gives calculated values.

We now compare theoretical values with calculated values. Graph below shows the comparison. Blue line represents actual value. Red value represents calculated value. We can infer that both lines almost overlap with each other.



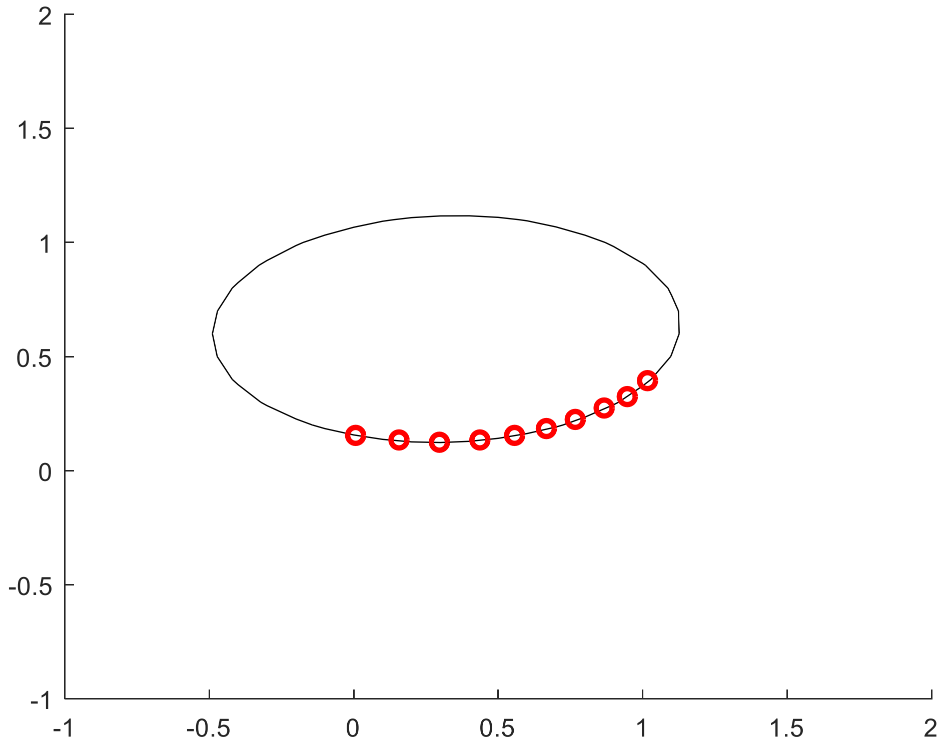


Figure above shows ellipse generated for given equation of orbital of planet. Red points represent section of ellipse(data points) which are calculated for the given equation.

Q.2)

i)

Rank of the matrix is maximum number of linearly independent columns or rows present in the given matrix.

For m*n matrix, if m > n, then maximum possible rank of the matrix is n

For m*n matrix, if m < n, then maximum possible rank of the matrix is m

Given matrix is,

A= 1/3 1/3 2/3

2/3 2/3 4/3

1/3 2/3 3/3

2/5 2/5 4/5

3/5	1/5	4/5

Now lets calculate echelon form of matrix A.

Perform C3-C1+C2

We get,

A=

1/3	1/3	0
2/3	2/3	0
1/3	2/3	0
2/5	2/5	0
3/5	1/5	0

Now C1 and C2 are independent i.e. we cant reduce any of these columns to 0 by performing some operation. So rank of the matrix is 2.

ii) Perform SVD on the matrix. We get three outputs.

[U,S,V]=svd(A)

Where, A=U*S*V'

1. U is orthogonal matrix of size m*m(5*5 in our case)

2. S is a rectangular diagonal matrix with non-negative real numbers of size m*n(5*3).

3. V is orthogonal matrix of size n*n(3*3).

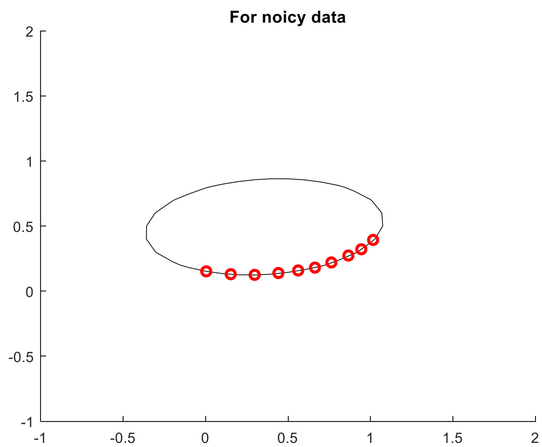
Rank of matrix A is no of non-zero diagonal elements present in the diagonal position of matrix S.

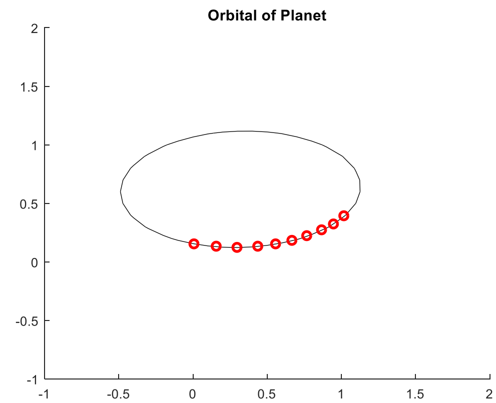
iii)

0.9^n is assigned to every (n,n) element of the matrix of size 2k*2k. Every value which is less than eps of every diagonal element * maximum dimension of the matrix(default tolerance) is considered as 0. The rank calculated with this tolerance is time consuming but most reliable.

B

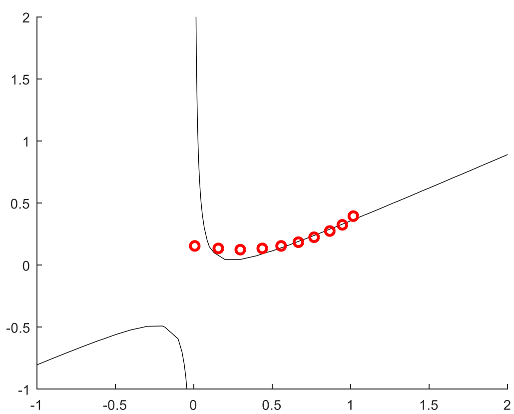
i)

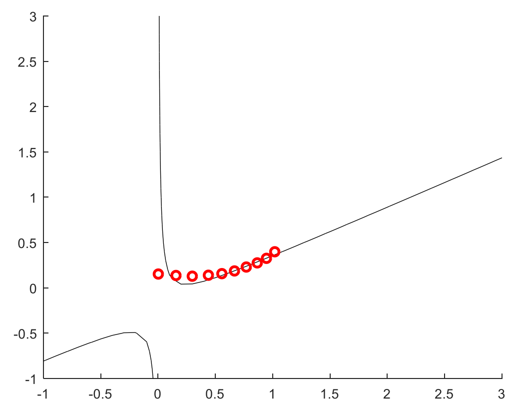




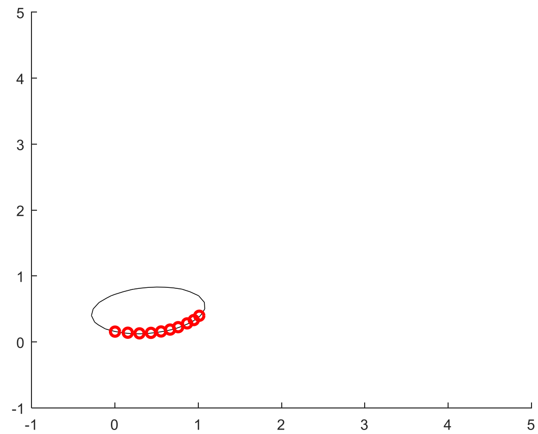
Comparison of normal and noicy data

When we add noice to the x & y co-ordinates of the points, the resultant diagram is also ellipse but is slightly tilted. Calculated data points are also on the ellipse.





This is the hyperbola we get for the tolerance = 10^-1 and rank 3. Data point are not over the plot. It also shows that low rank approximation doesn’t fit the data perfectly for the normal data(with noice and without noice)



For 10^-2 we get rank 4, low rank approximation fits the data for rank 4 also and forms an ellipse.

For other values, we get rank 5 and get normal ellipse for normal data and also for the noicy data, we get ellipse which is slightly tilted and bigger.

Q.3)

Dimension of S is same as the matrix. But contains only 100 nonzero elements at the diagonal position. Values are denser.

In the scatter plot of 1st column of U vs 2nd column of U, rotation is performed on the voters matrix

U1 and U2 represents first and second principle components of the multivariate function.

Analysis shows that it should decrease and it is decreasing. Most of the predictions tend to remain same in the future. Pattern is followed as for most of the representatives, fraction is close to 1.