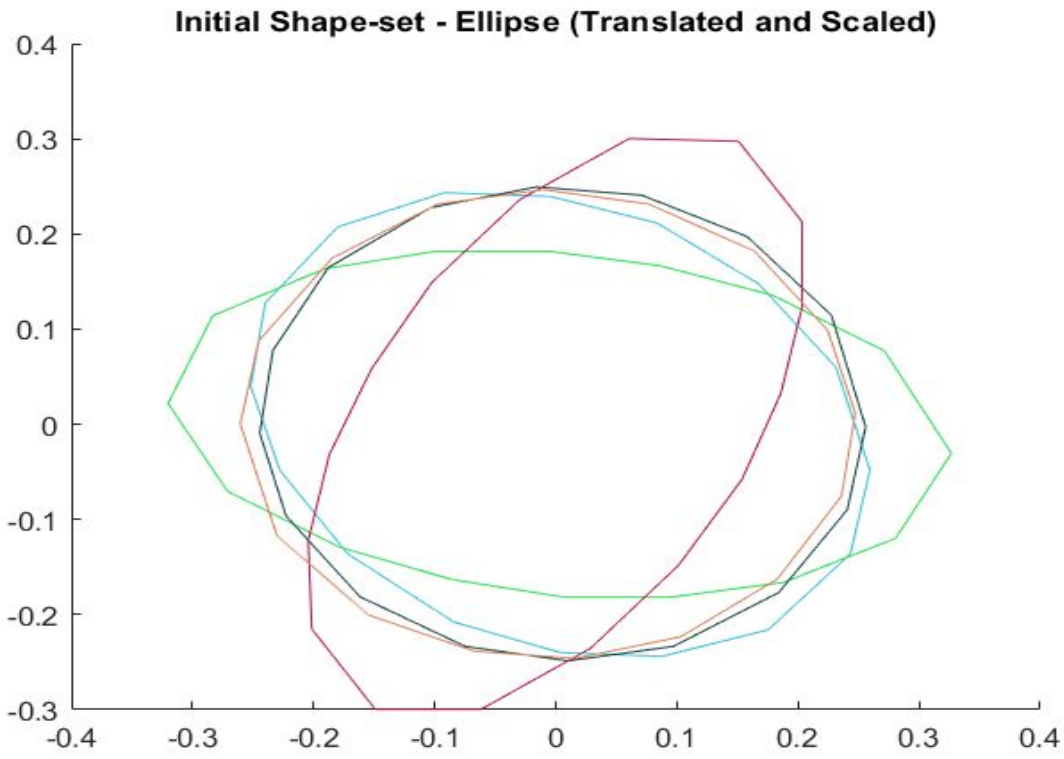
Assignment 1 - Shape Analysis
Report (Question 1)

Members: Bavish (170040106) and Manan (170040067)

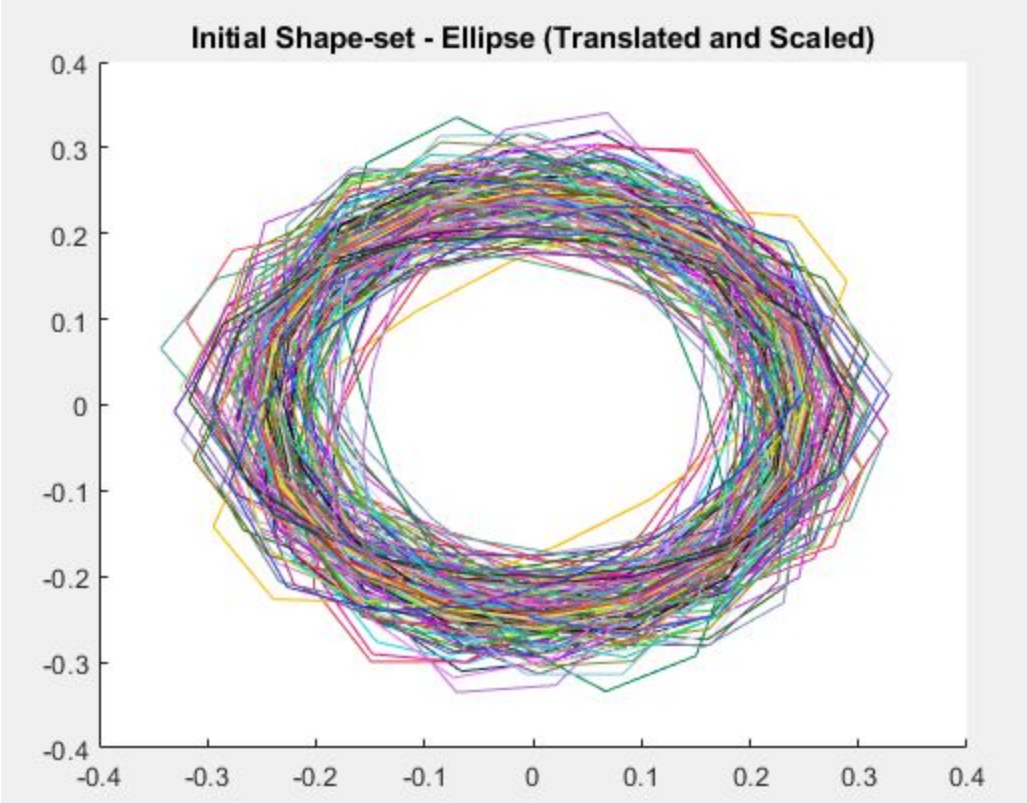
Data Collection (refer getPointset.m file):

For ellipse dataset, the data was obtained in an automatic manner. The image was binarized (i.e. all pixel intensities above a particular threshold, here 0.1, were set to 1, else, 0). Matlab's Image Processing Toolbox™ contains special functions for tracing boundaries for images. For the binary images, 'bwboundaries' function which returns a matrix with all the boundary points collected in order, was used.

For Procrustes' shape analysis, two captured pointsets must have the corresponding points at same indices. But 'bwboundaries' captures points in an arbitrary manner; it picks up the starting point from a fixed direction in an image and starts tracing from it. Since the ellipse dataset has ellipses with major and minor axis aligned randomly with respect to the image coordinates, the above condition gets violated. For remedying this, once all boundary points are captured, a best fit line is plotted. Now for an ellipse, the best fit line will overlap with the major axis. The boundary point which is nearest to the best fit line serves as the corner point of ellipse, and serves as the first index for the pointset. The remaining points are picked from the list after equal index interval for obtaining a symmetrical pointset.
Following image shows 5 pointsets sampled this way.

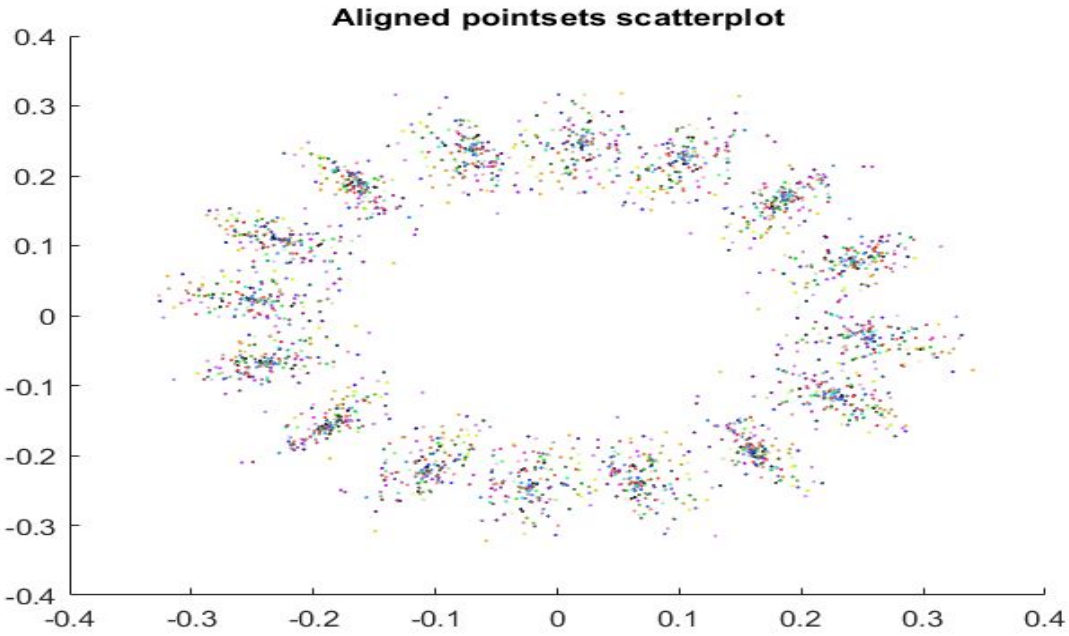


Plot of all pointsets after translating and scaling:

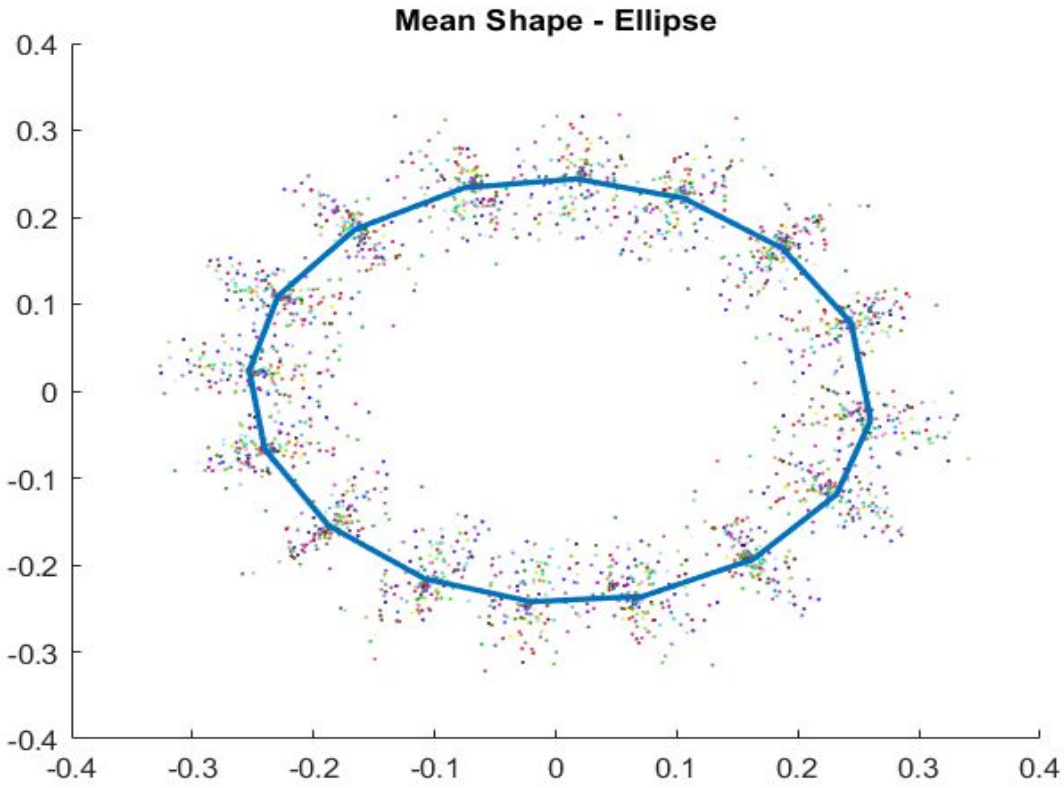


<u>Pointset Alignment</u>:

Once all the pointsets were captured, each pointset was translated to origin and scaled down by its norm. The scatterplot looked like (note only translation and scaling is done yet):



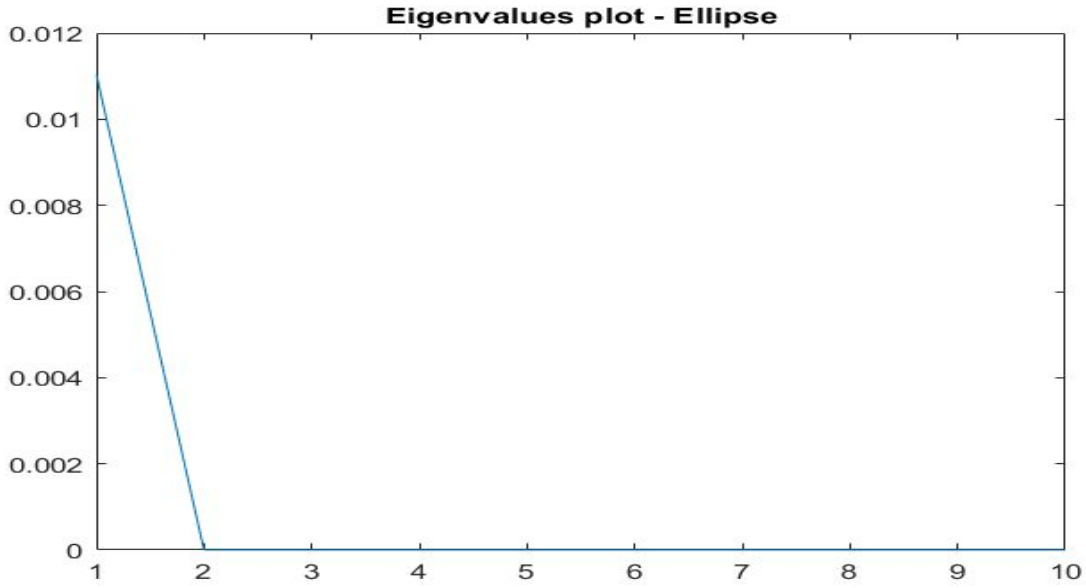
The mean was calculated by plotting together the mean of each individual pointclouds. The resulting plot was:

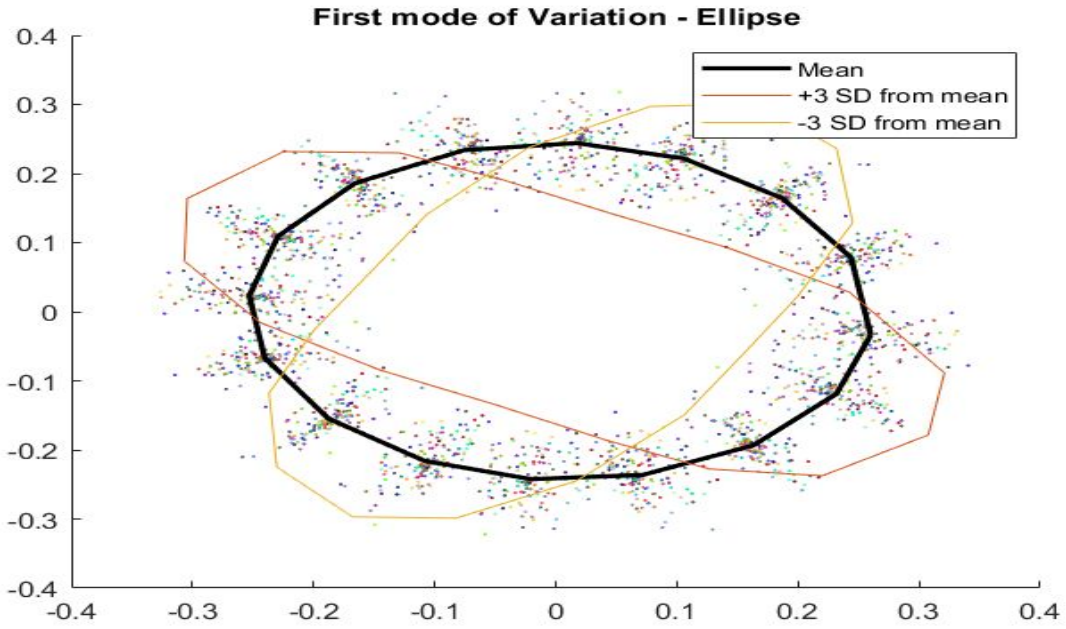


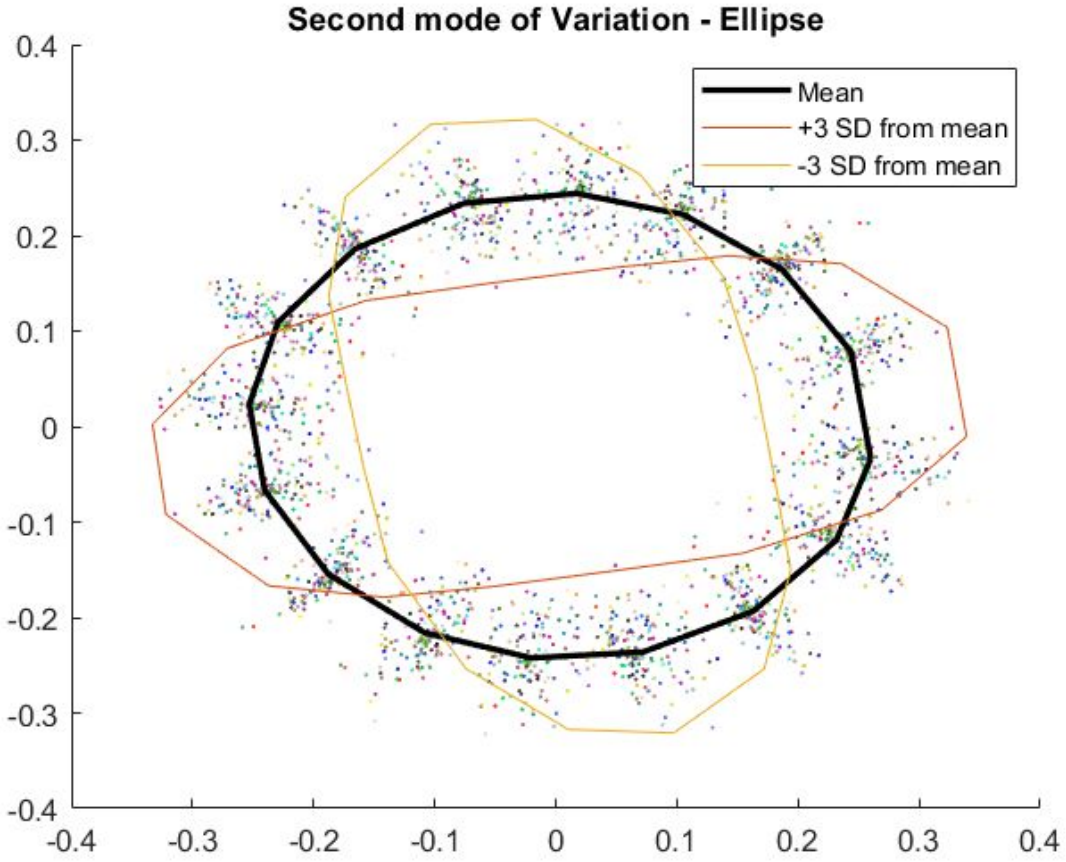
Once the mean is obtained, each pointset is aligned (rotated; since all are already translated and scaled) with respect to the mean shape.

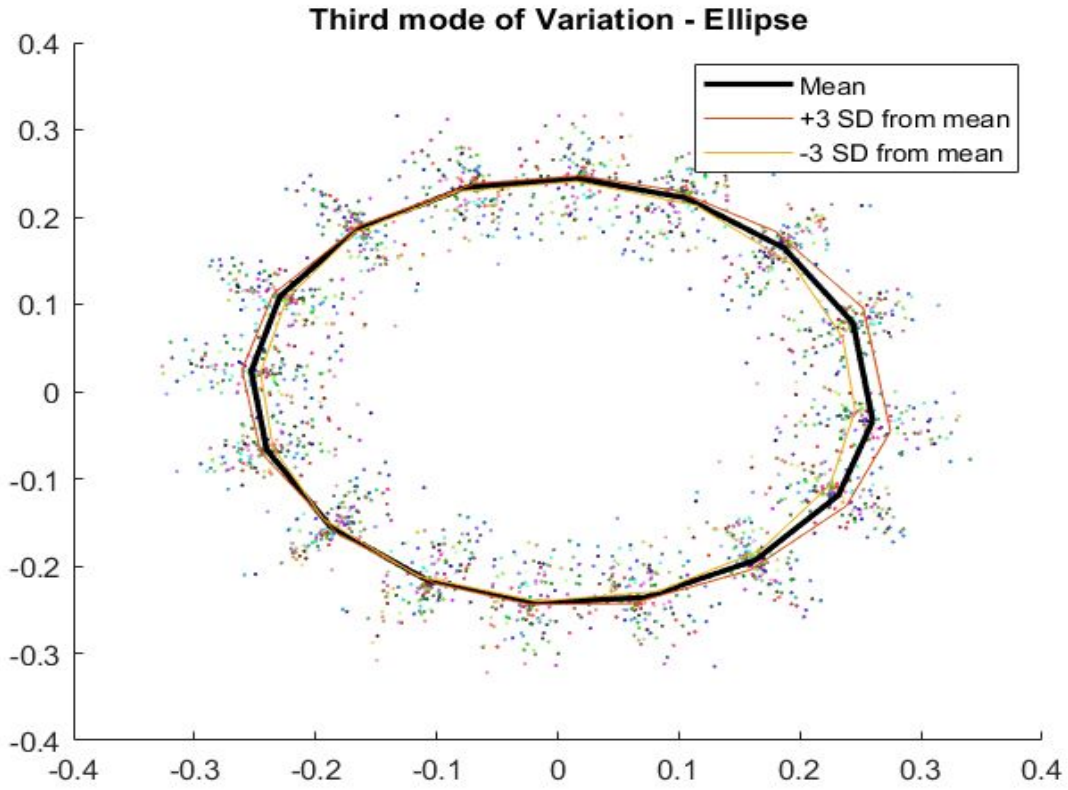
Statistical Analysis:

Instead of investigating isolated point variance, the set of shapes was stretched into a single vector of the form [x_1, y_1, x_2, y_2,.....] and PCA shape decomposition was carried out. This results in an ordered basis where each component is ranked after variance. The PCA shape decomposition is able to represent much of the variance, just using the three parameters. Following are the plots of the sorted(first 10) eigenvalues and first three modes of variation:

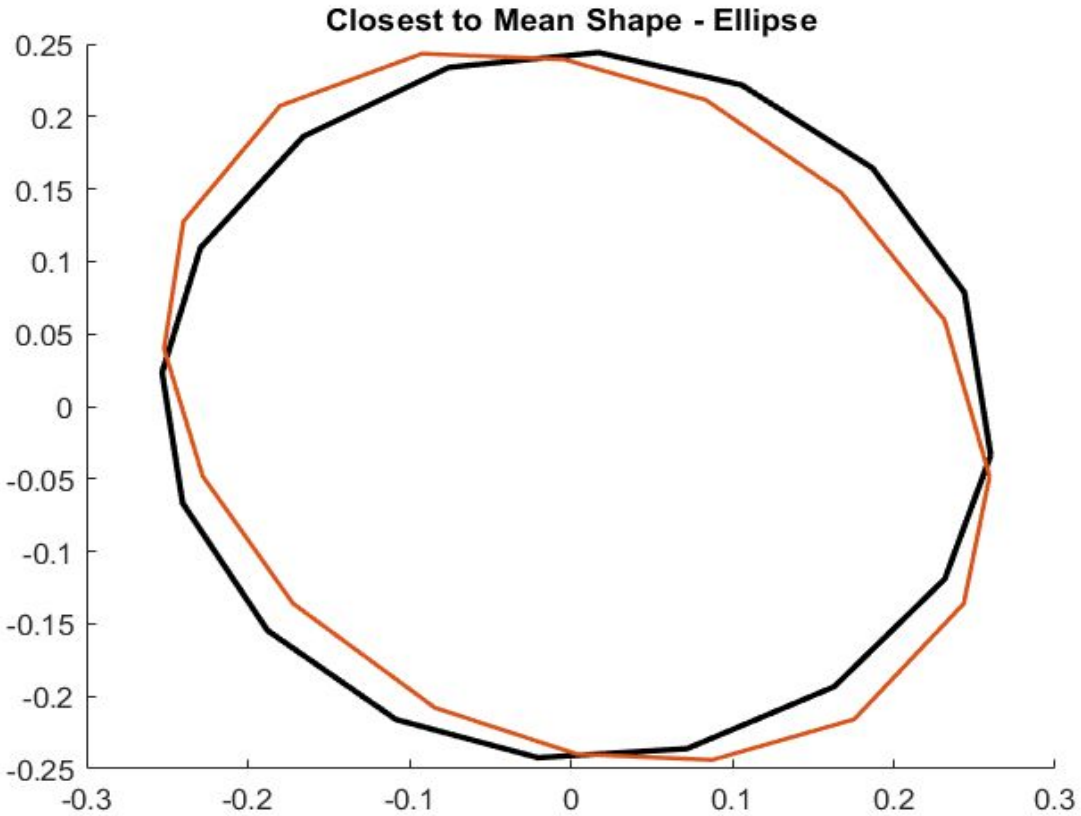


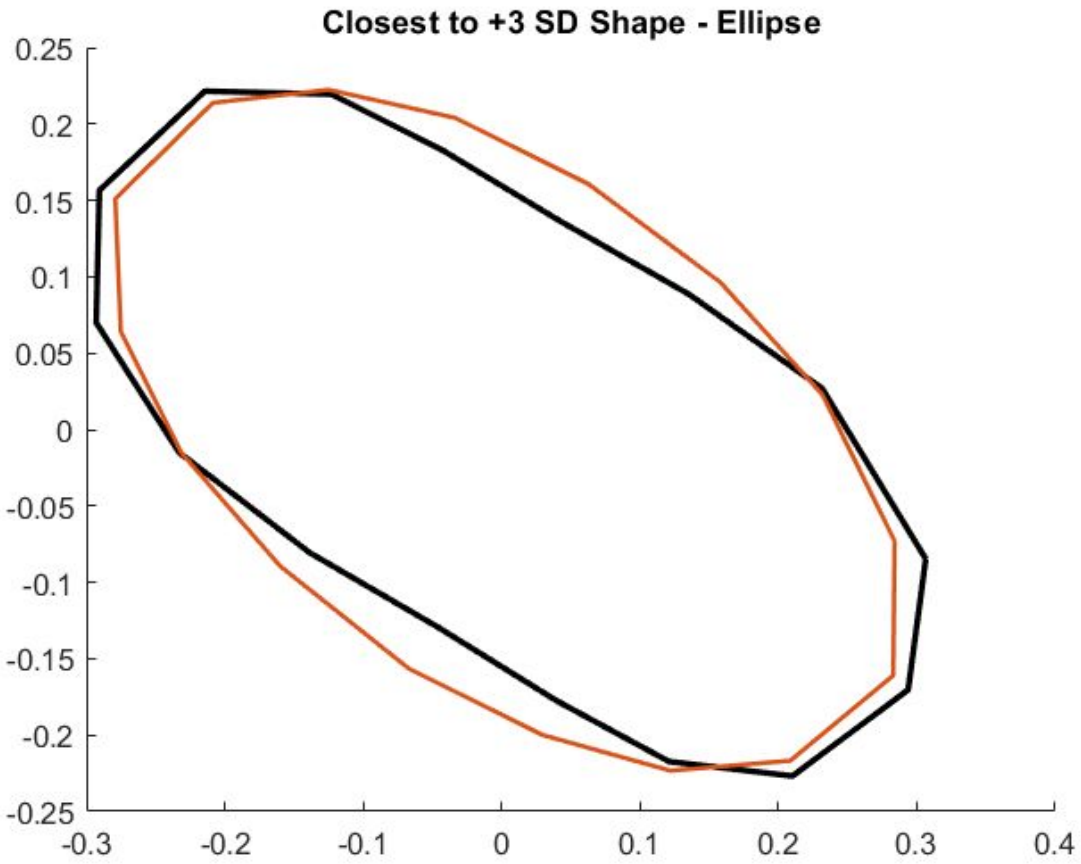

First mode of Variation - Ellipse
Mean
+3 SD from mean
-3 SD from mean
0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
-0.4
-0.3
-0.2
-0.1
0
0.1
0.2
0.3
0.4

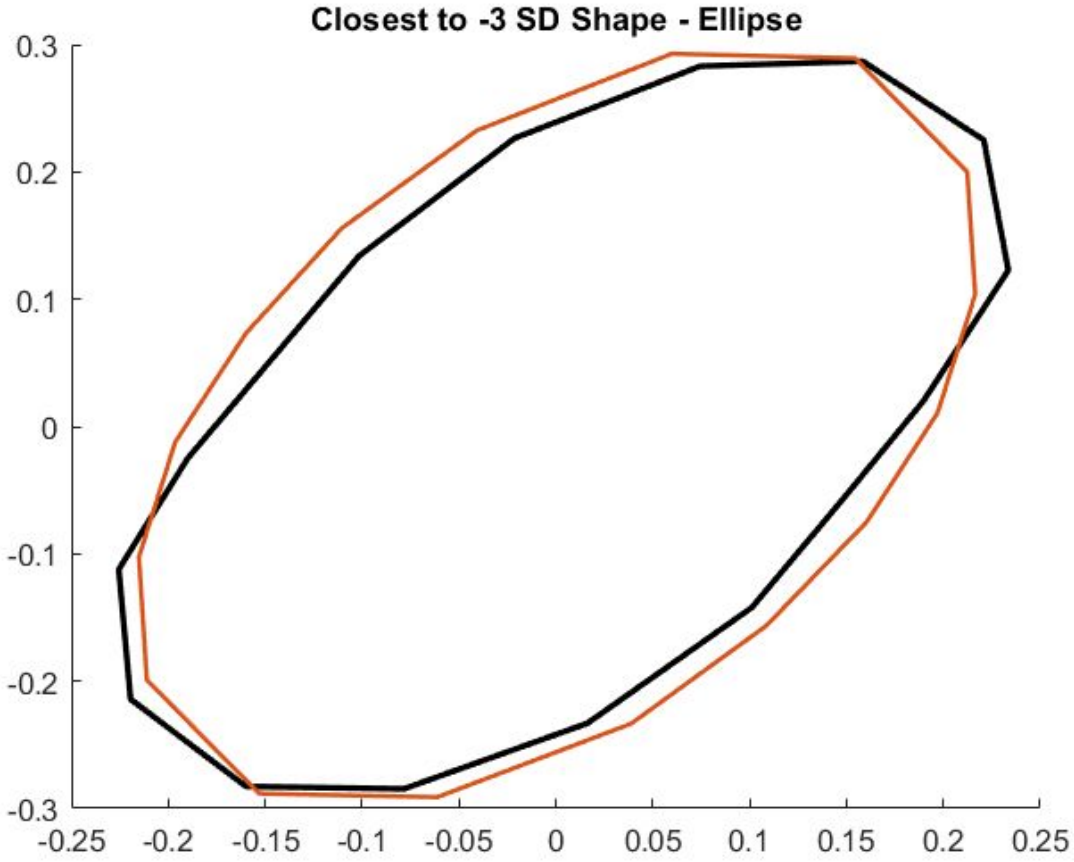

Second mode of Variation - Ellipse
Mean
+3 SD from mean
-3 SD from mean
0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
-0.4
-0.3
-0.2
-0.1
0
0.1
0.2
0.3
0.4



The following plots are the pointsets which were most similar(ranked against L2 norm) with the mean, +3 SD away from the top mode of variation and -3 SD away from the top mode of variation respectively (black curve is the reference curve):




Closest to +3 SD Shape - Ellipse
0.25
0.2
0.15
0.1
0.05
0
-0.05
-0.1
-0.15
-0.2
-0.25
-0.3
-0.2
-0.1
0
0.1
0.2
0.3
0.4


Closest to -3 SD Shape - Ellipse
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.25
-0.2
-0.15
-0.1
-0.05
0
0.05
0.1
0.15
0.2
0.25