Alex Ho
ID: 1001294712
CSE 4334-001

Assignment 1 Report

Problem 1:

Part 1:

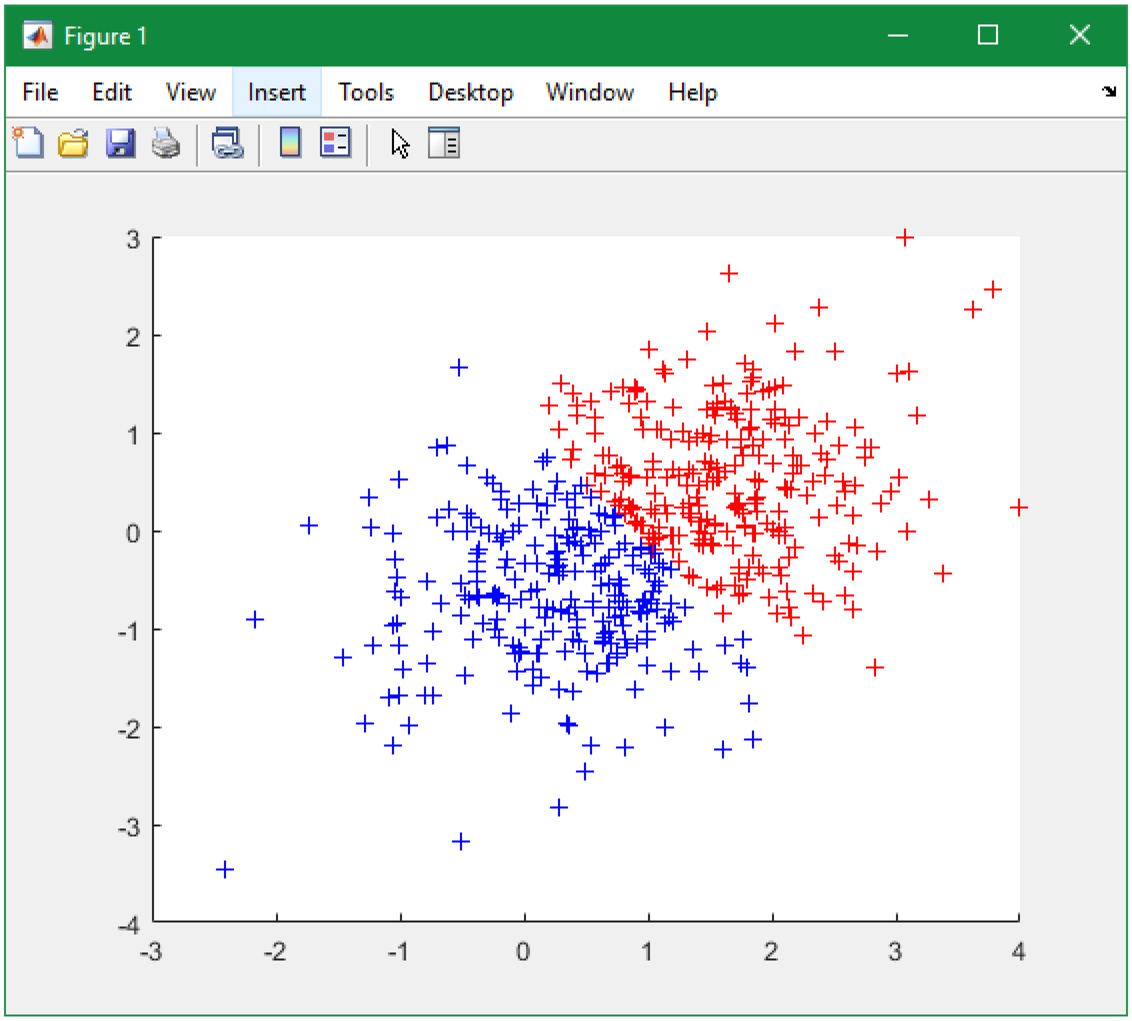
Mykmeans.m was relatively simple to implement. It completes k means and returns an array of the data point's clusters. However, since it's only returning the centers they belong to, I wasn't sure how to plot the new centers outside of the method. I do have debug code within the method that will plot it, but I left it commented.

The biggest problem I had when solving the problem was figuring out how to converge the centers, but I was able to do it by getting the average of all the data points and setting the new center to it. My function also has two built in functions, one that finds the L2 from a point to a given center vector, and one that takes two vectors and finds the distances for each corresponding row. The first one is used when iterating through each data point to find its distance to every center. The second one was used to simply find the distance from all old centers to all new centers.

Part 2:

Applying the code to the data given with the specified parameters in part 2, the run in Figure 1 below took 11 iterations. The centers found are at (1.6737, 0.5212) and (0.2181, -0.6804). I did not plot the centers in the figure, as it would require plotting in the method itself.

Figure 1



<u>Part 3:</u>

Doing the same for part 3 as I did in part 2, it took 9 iterations. The centers were at (.8642, 2.4646) , (-1.1760, 0.3076), (0.3224, 0.9326) and (-0.6405, 1.7524). The result is in Figure 2

Figure 2

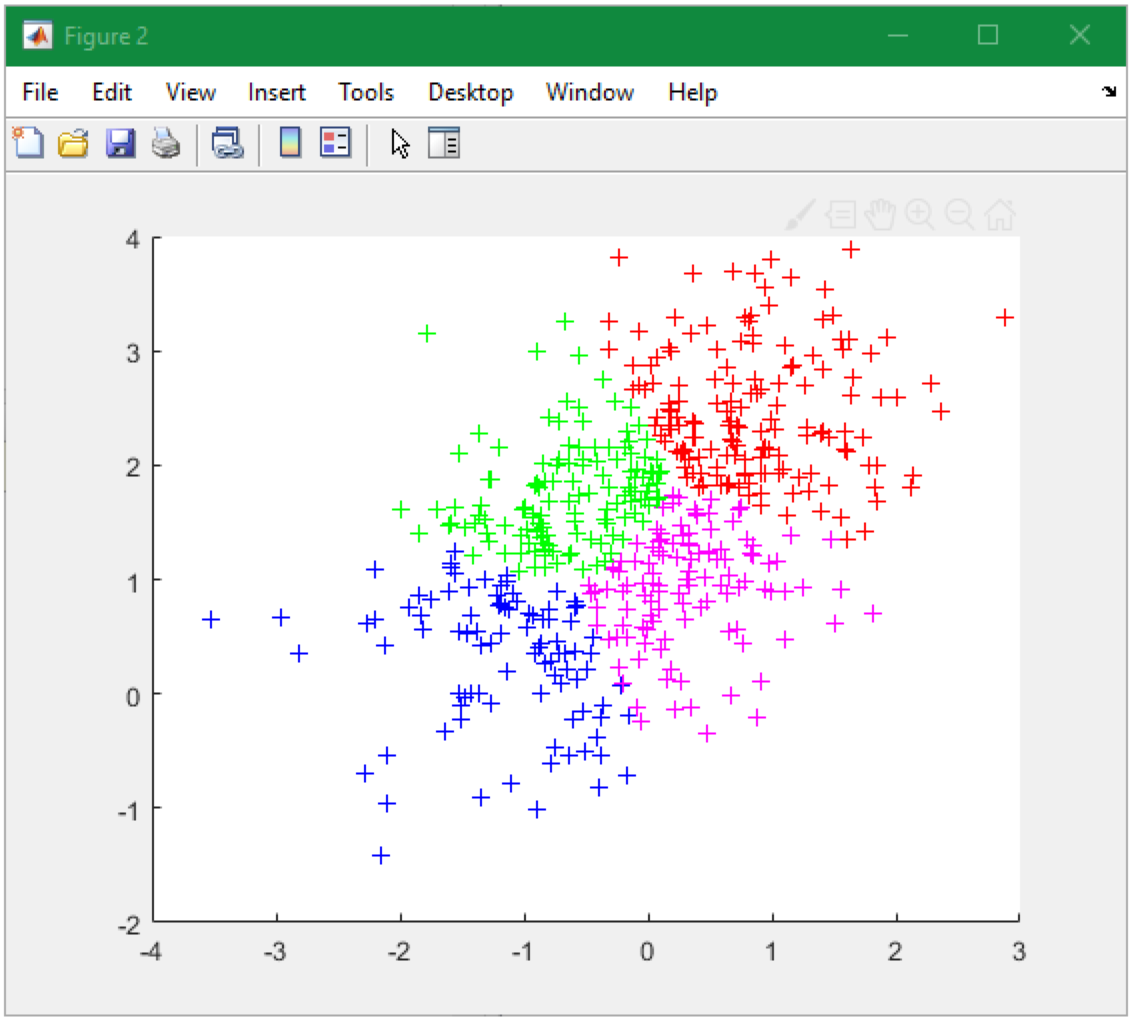


Figure 2
File
Edit
View
Insert
Tools
Desktop
Window
Help
4
3
2
1
0
-1
-2
-4
-3
-2
-1
0
1
2
3

Problem 2:

Problem 2 was much harder. I was confused on how to implement the kernel density estimation formula, but I was able to figure it out in the end. However, there are a couple of potential issues with the code:

1: I wasn't entirely sure what kernel function I should use. I decided to go with the Parzen Window function in the slides, but I also included a Gaussian kernel function commented out within the method.

2: I didn't understand what data Part 3 was asking for. It listed two datasets, but the first dataset was the same as Part 1. In the end, I decided to omit the duplicate data set and only use the new new mu and sigma (0 and 0.2).

3: When I compared mykde to the built in kde function, I noticed that their line reached all the way to the bottom. This goes past the given domain, but I wasn't sure how to implement it or if I was supposed to. Therefore, my kde line stops at the end of the domain.

4: I had trouble understanding how to implement a the multivariate kernel density estimation for a matrix. I ended up using the built in mvksdensity function to compute the probability.

Beyond these problems, the function should work as intended. The figures provided below corrispond these parts:

Figure 1-4 = kde of Part 2 at bandwidths h = {.1, 1, 5, 10}. Data is N = 1000 gaussian random data with $\mu = 5$ and $\sigma 1 = 1$

Figure 5-8 = kde of Part 3 at bandwidths h = {.1, 1, 5, 10}. Data is N = 1000 gaussian random data with $\mu = 0$ and $\sigma 1 = 0.2$

Figure 9-12 = kde of Part 4, N_1 at bandwidths h = {.1, 1, 5, 10}. Data is N1 = 500 gaussian random data with $\mu = [1, 0]$ and $\Sigma = [0.9, 0.4 ; 0.4 , 0.9]$.

Figure 13-16 = kde of Part 4, N_2 at bandwidths h = {.1, 1, 5, 10}. Data is N2 = 500 gaussian random data with $\mu = [0, 1.5]$ and $\Sigma = [0.9, 0.4 ; 0.4 , 0.9]$.

Figure 1-4: Part 2

Figure 1

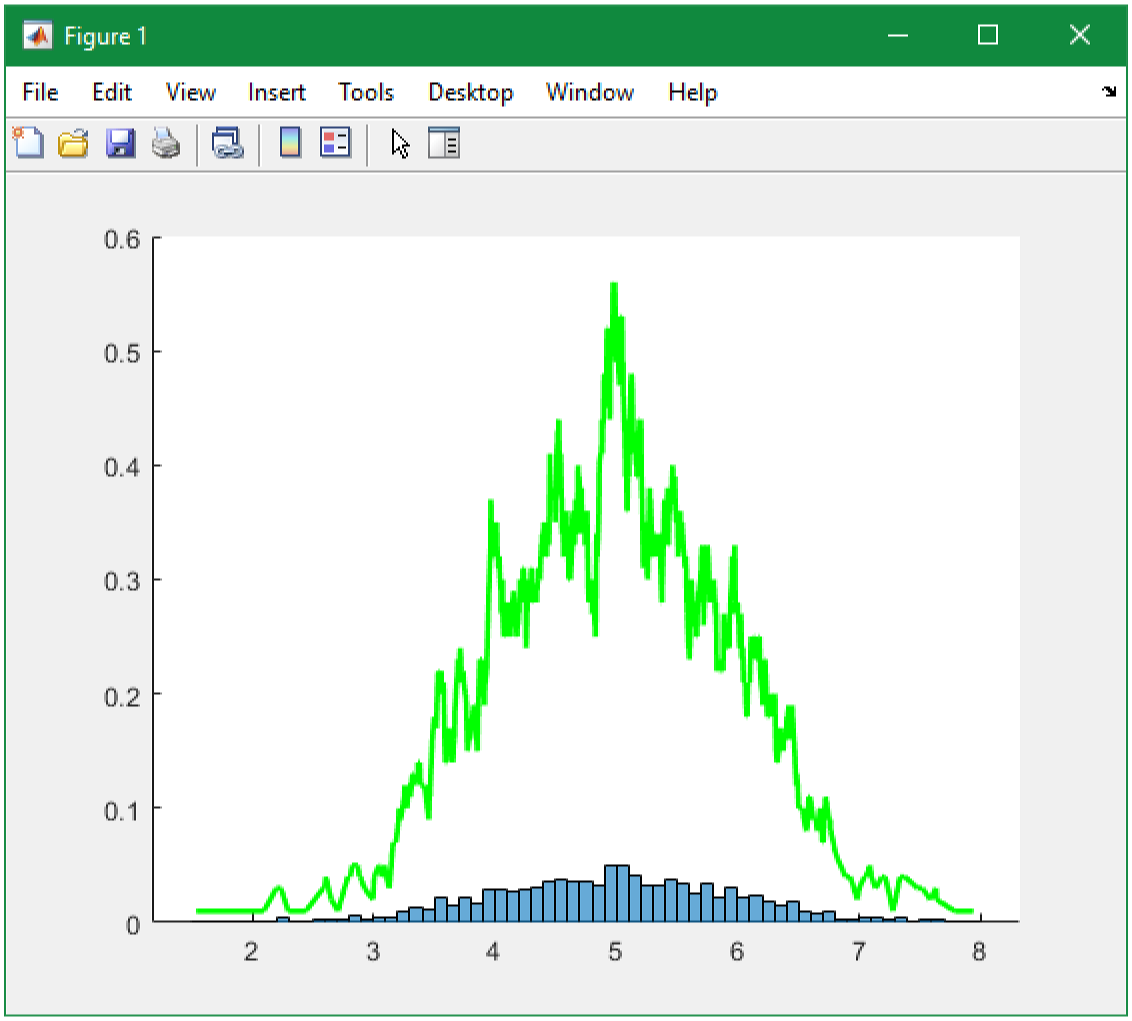


Figure 2

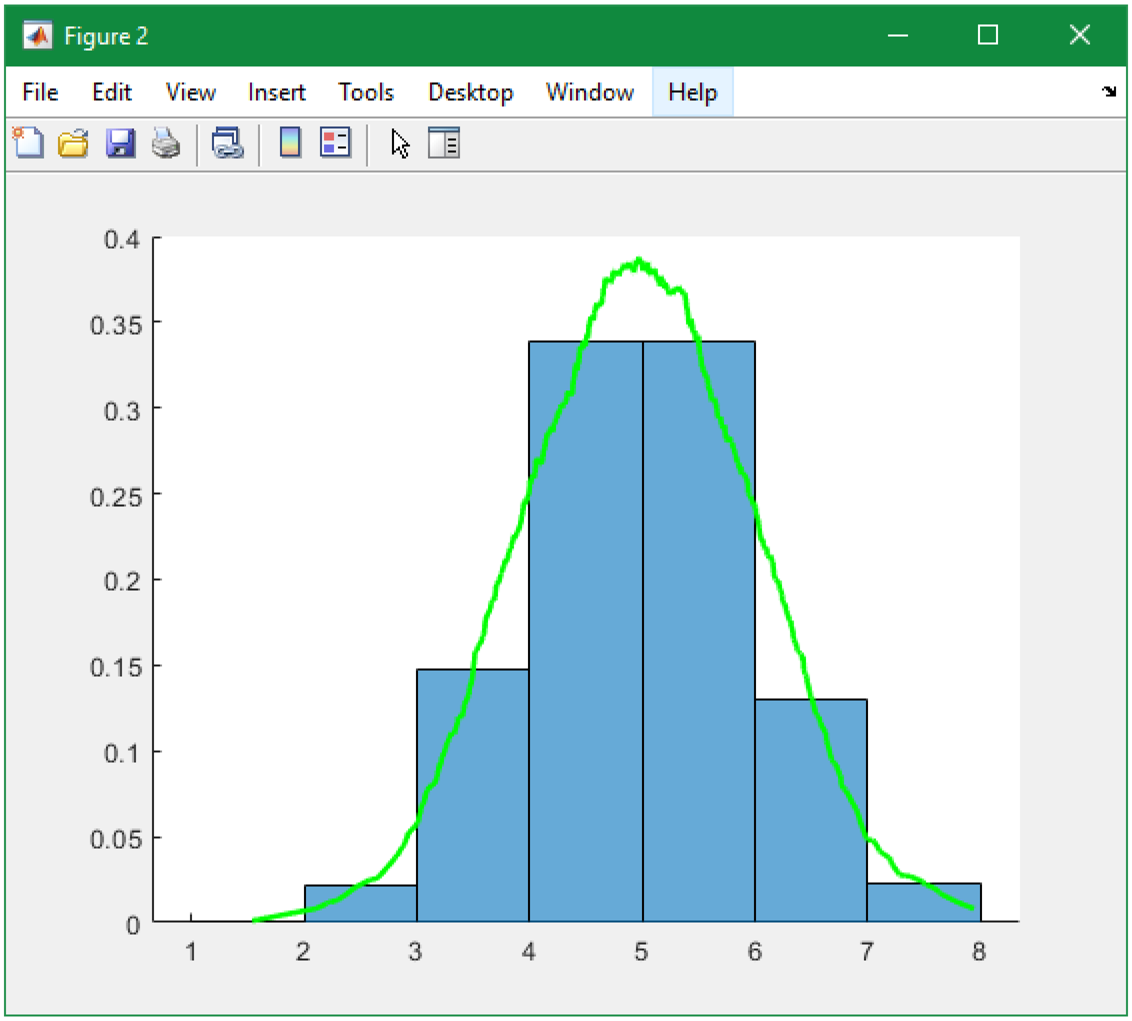


Figure 2
File
Edit
View
Insert
Tools
Desktop
Window
Help
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
1
2
3
4
5
6
7
8

Figure 3

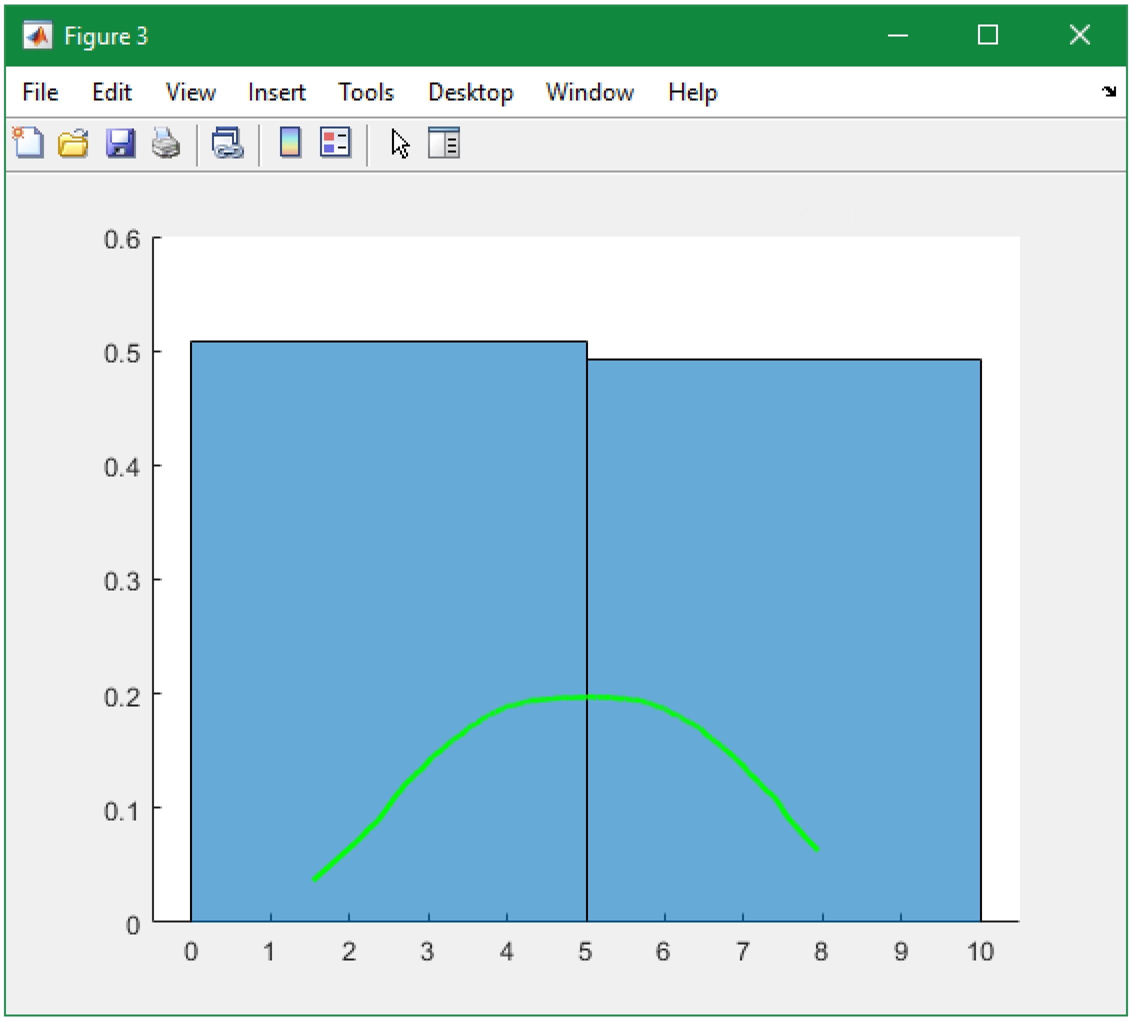


Figure 3
File Edit View Insert Tools Desktop Window Help
0.6
0.5
0.4
0.3
0.2
0.1
0
0 1 2 3 4 5 6 7 8 9 10

Figure 4

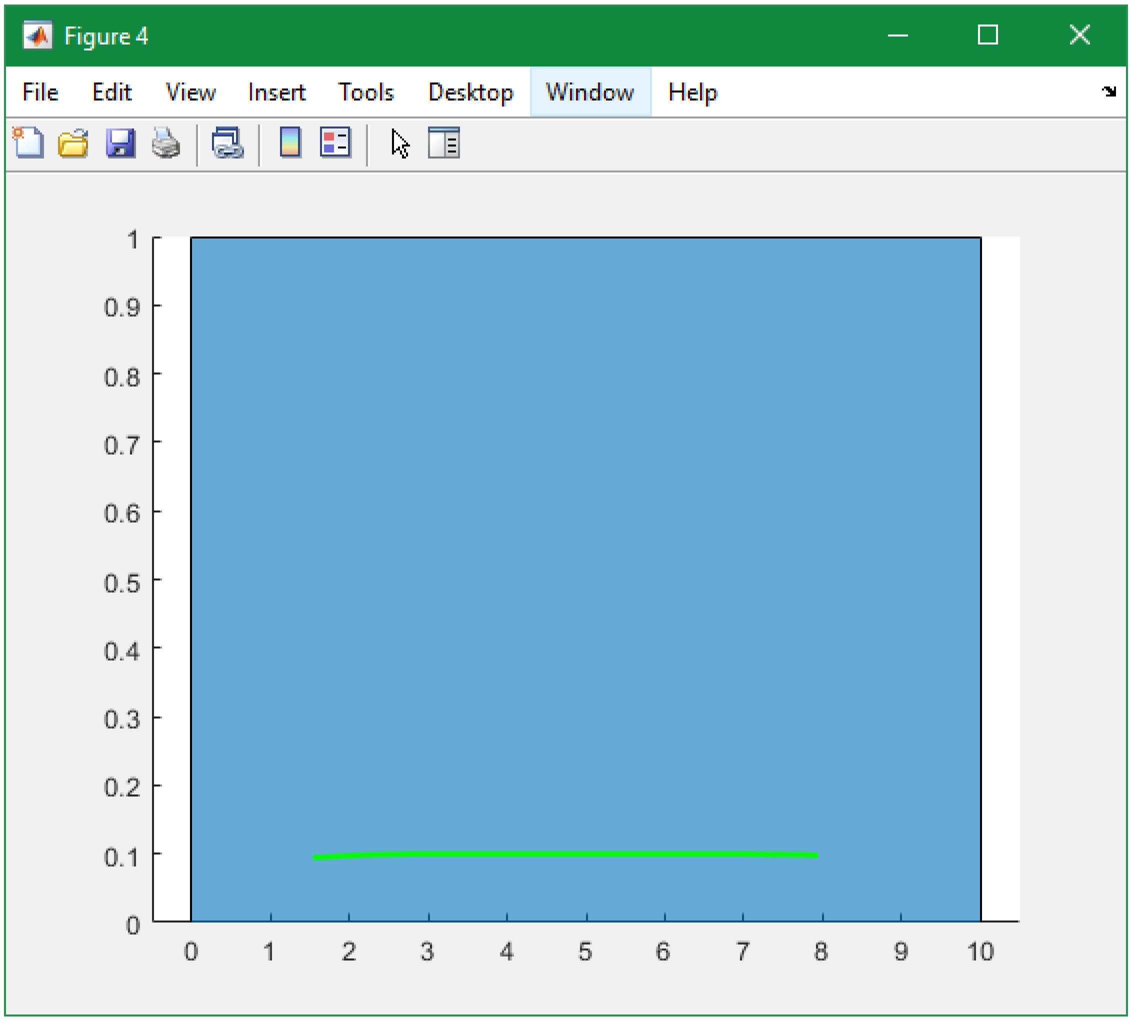


Figure 4
File
Edit
View
Insert
Tools
Desktop
Window
Help
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
1
2
3
4
5
6
7
8
9
10

Figure 5-8: Part 3

Figure 5

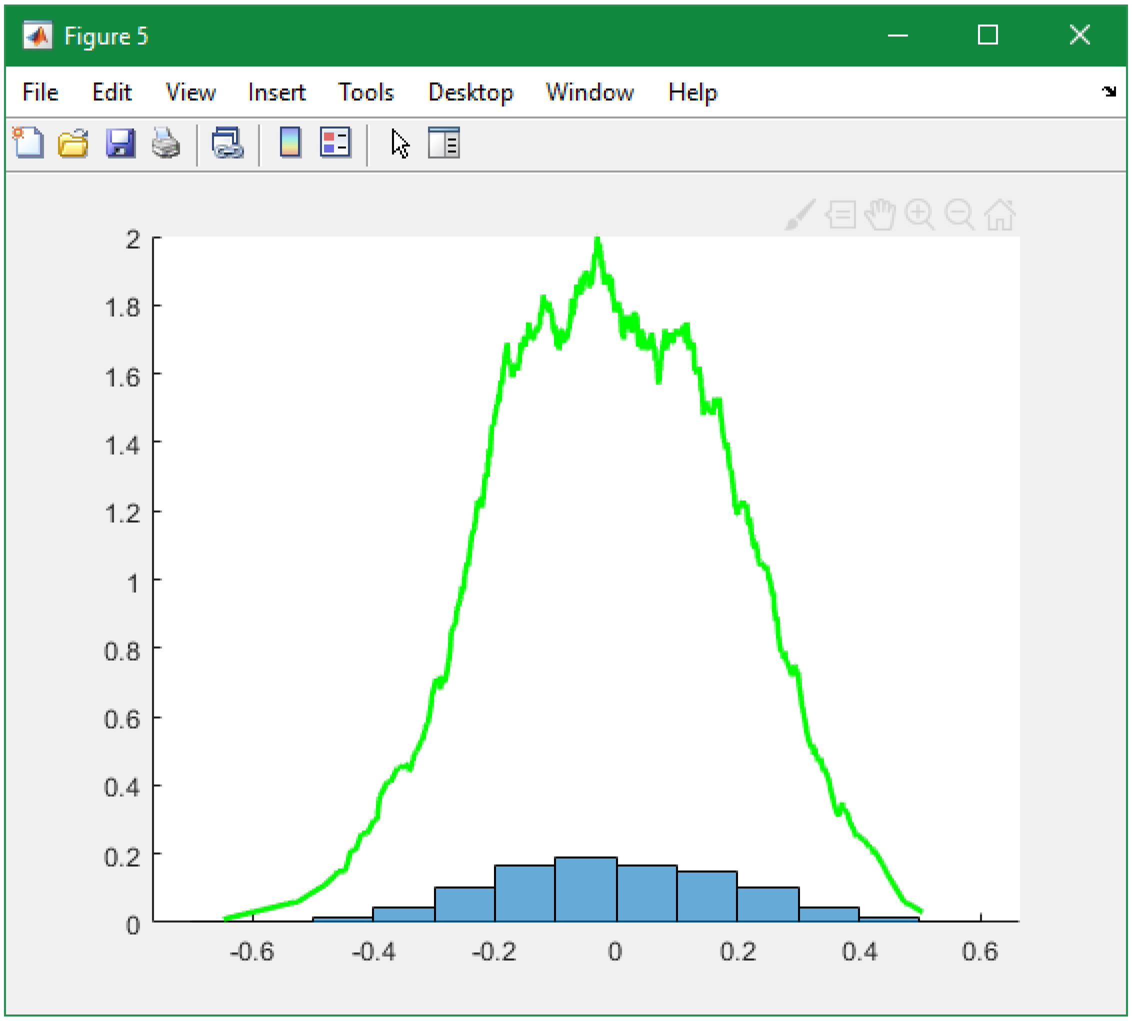


Figure 6

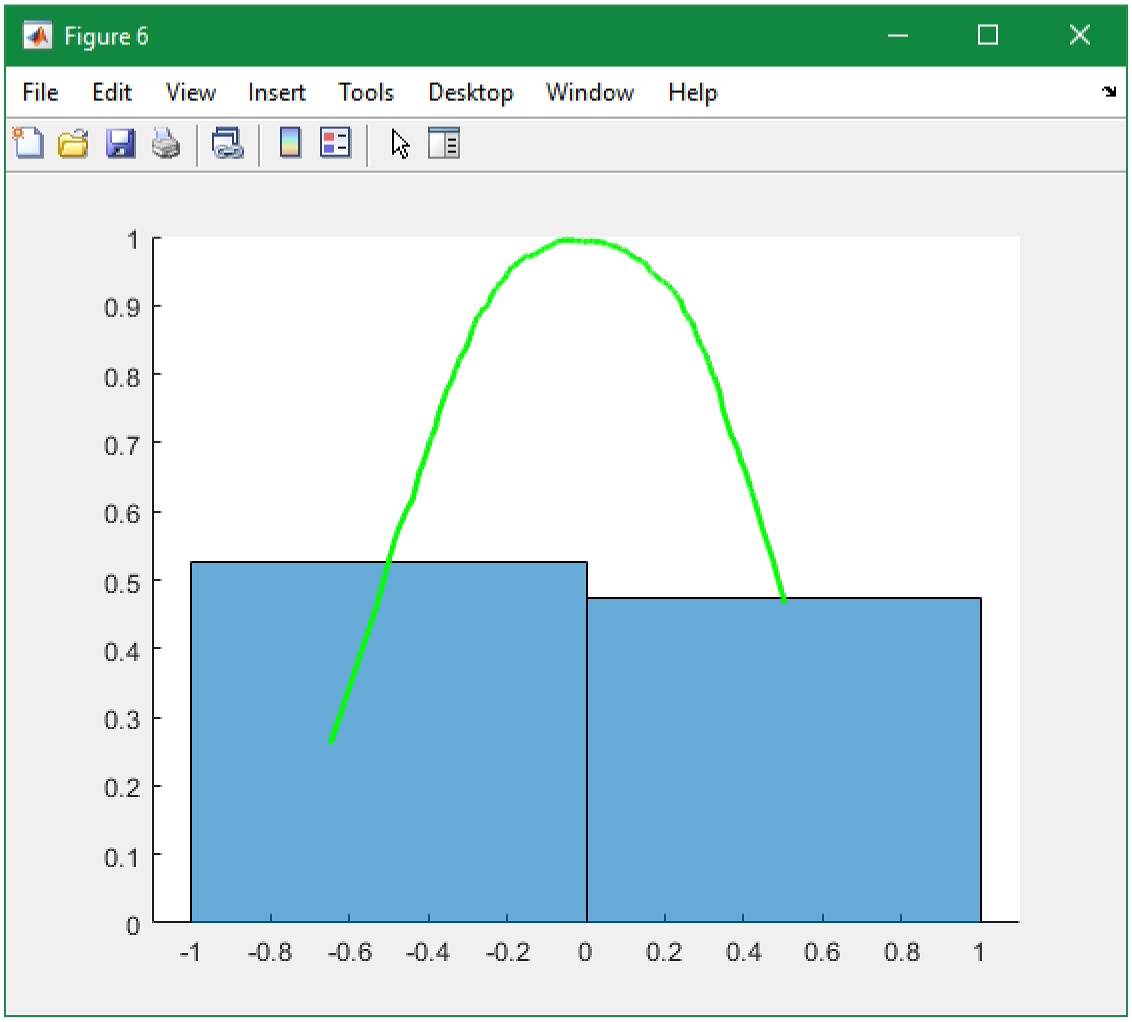


Figure 6
File
Edit
View
Insert
Tools
Desktop
Window
Help
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-1
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1

Figure 7

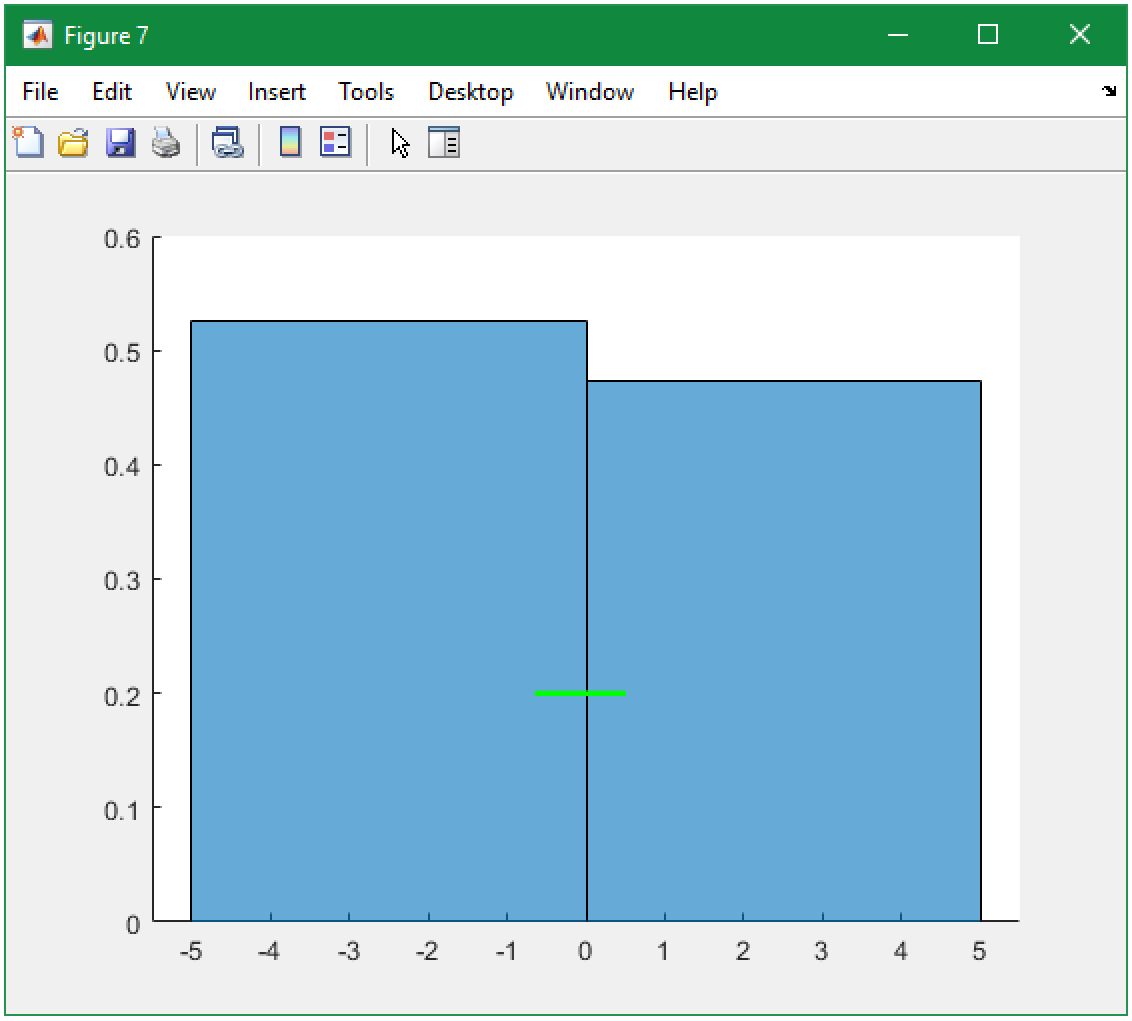


Figure 7
File
Edit
View
Insert
Tools
Desktop
Window
Help
0.6
0.5
0.4
0.3
0.2
0.1
0
-5
-4
-3
-2
-1
0
1
2
3
4
5

Figure 8

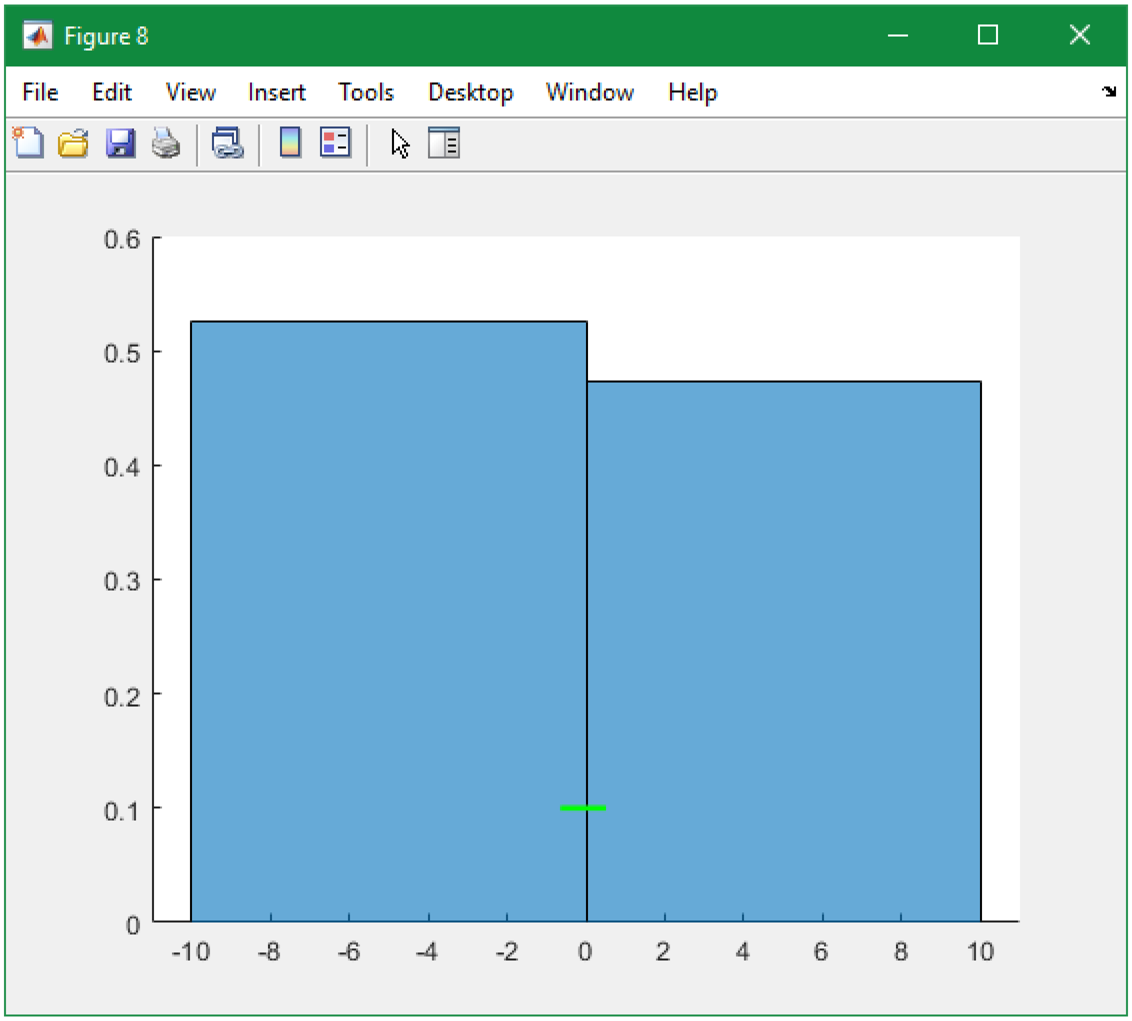


Figure 8
File
Edit
View
Insert
Tools
Desktop
Window
Help
0.6
0.5
0.4
0.3
0.2
0.1
0
-10
-8
-6
-4
-2
0
2
4
6
8
10

Figure 9-12: Part 4 - N1

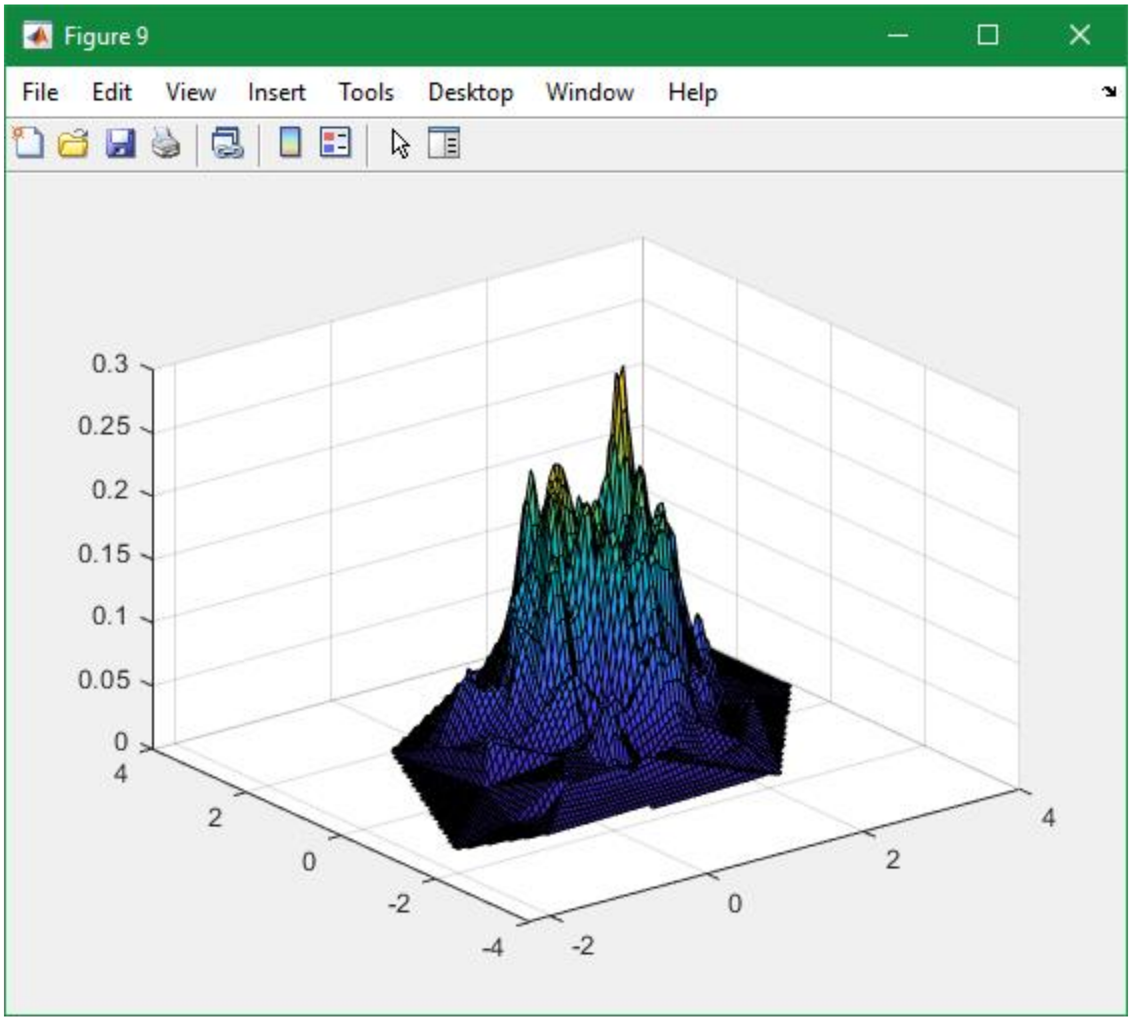


Figure 10

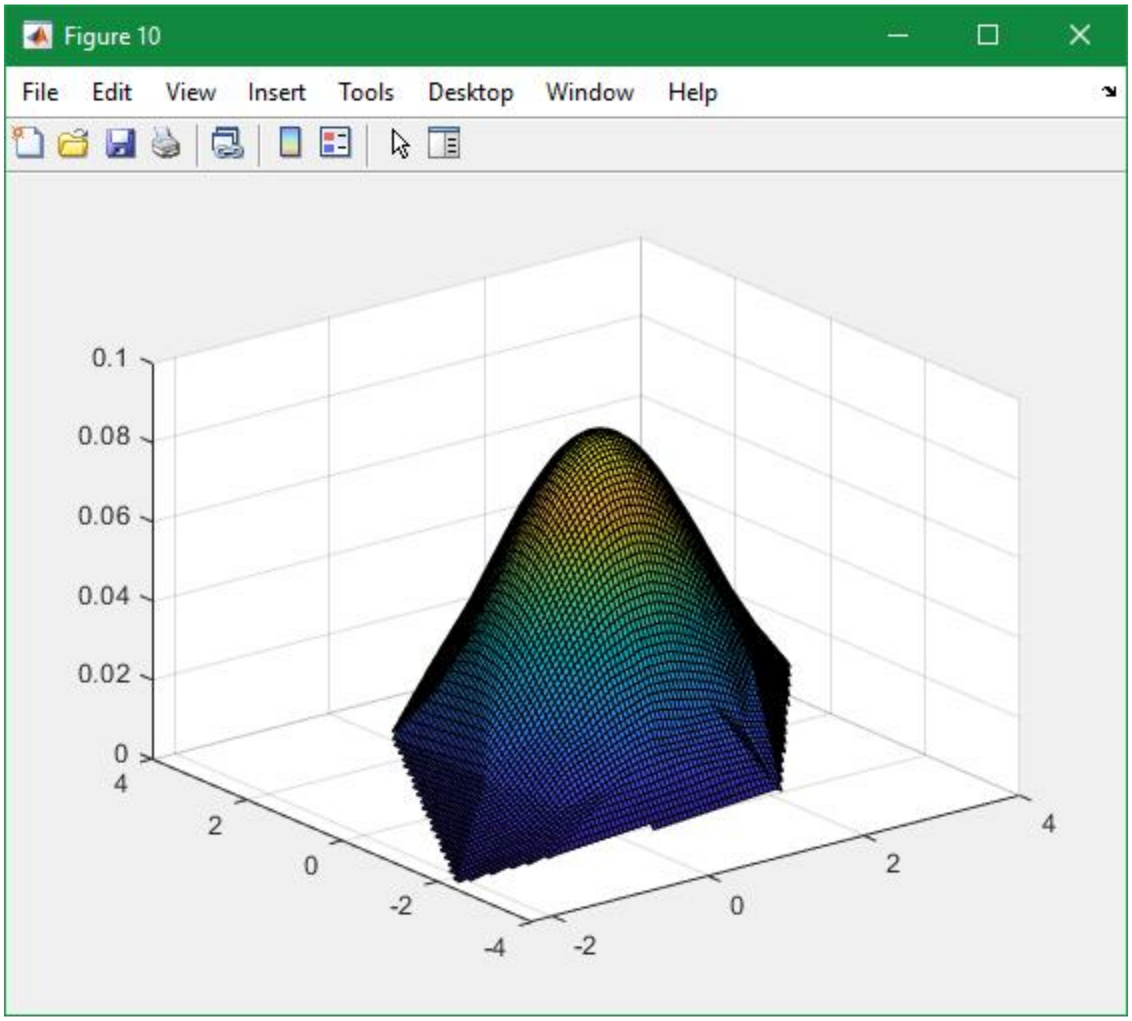


Figure 10
File Edit View Insert Tools Desktop Window Help
0.1
0.08
0.06
0.04
0.02
0
4
2
0
-2
-4
-2
0
2
4

Figure 11

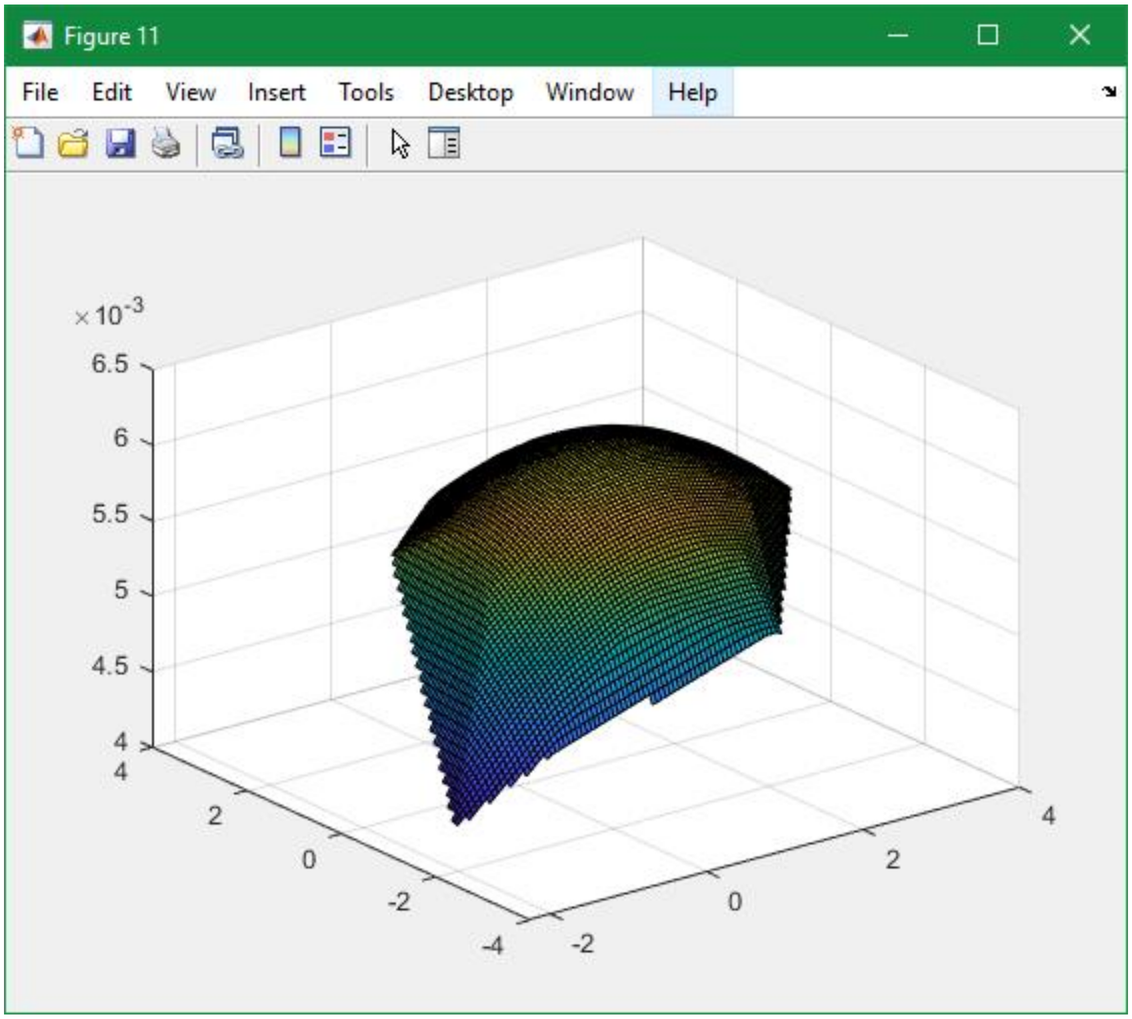


Figure 11
File
Edit
View
Insert
Tools
Desktop
Window
Help
× 10^-3
6.5
6
5.5
5
4.5
4
4
2
0
-2
-4
-2
0
2
4

Figure 12

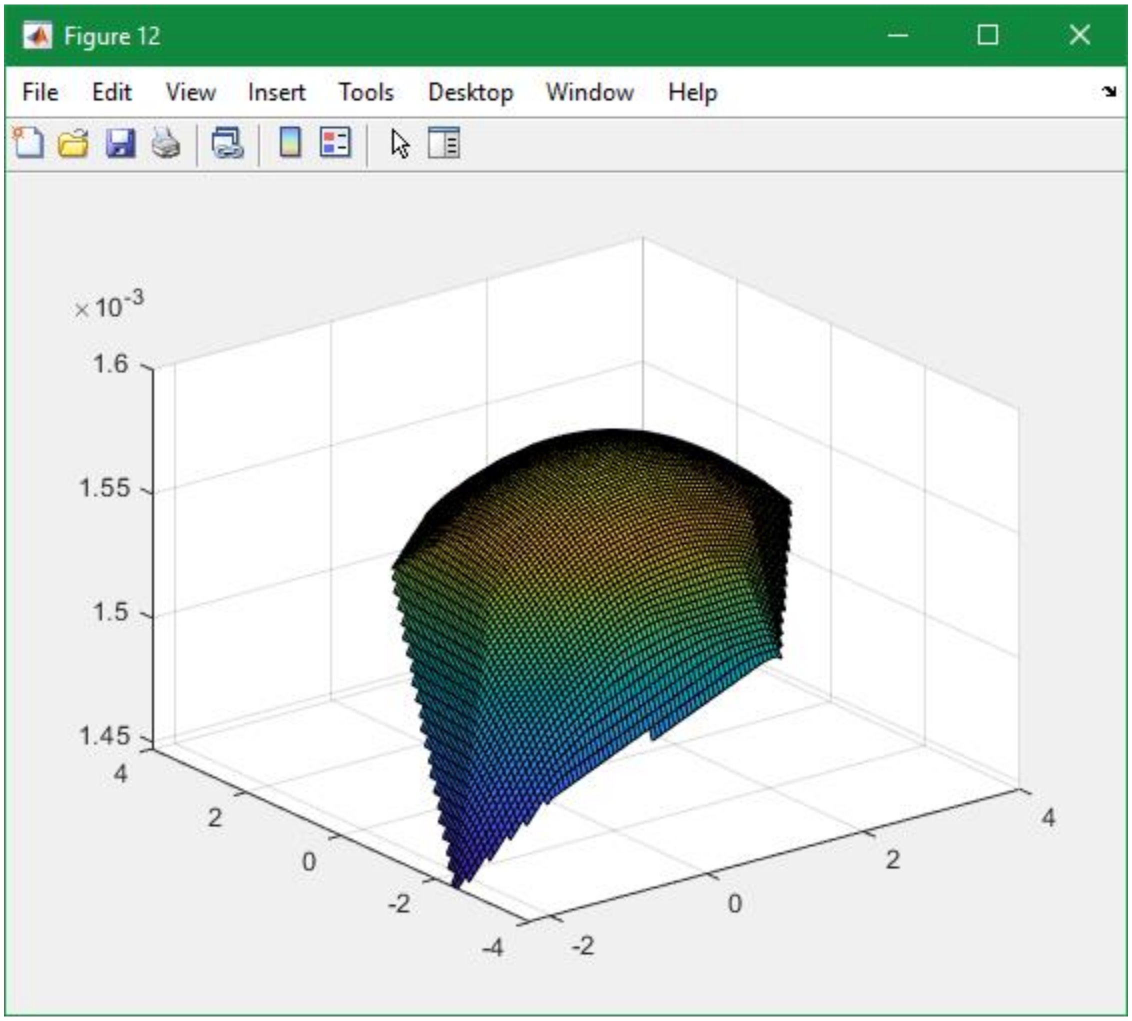


Figure 12
File
Edit
View
Insert
Tools
Desktop
Window
Help
×10^-3
1.6
1.55
1.5
1.45
4
2
0
-2
-4
-2
0
2
4

<u>Figure 13 – 16: Part 4 – N2</u>

Figure 13

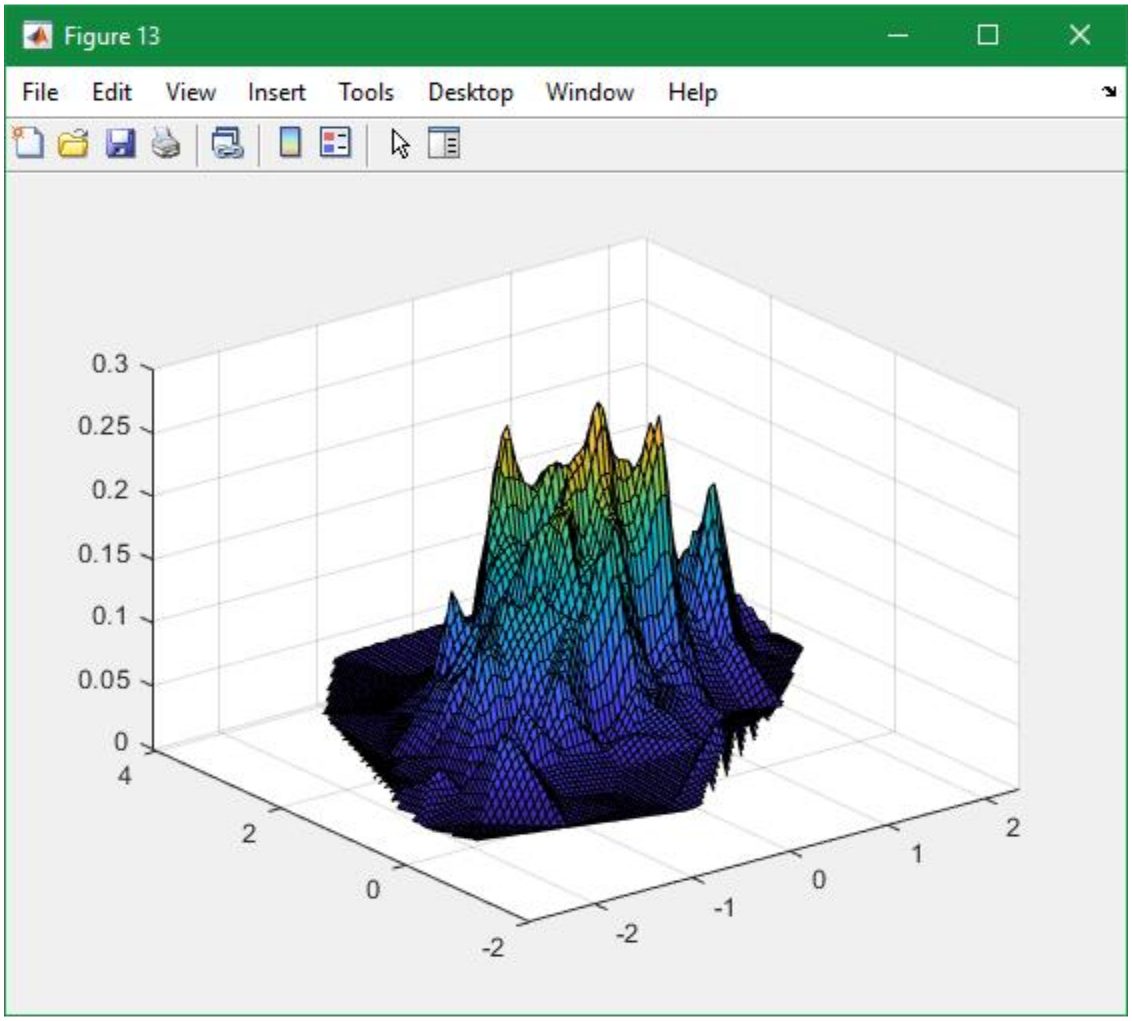


Figure 14

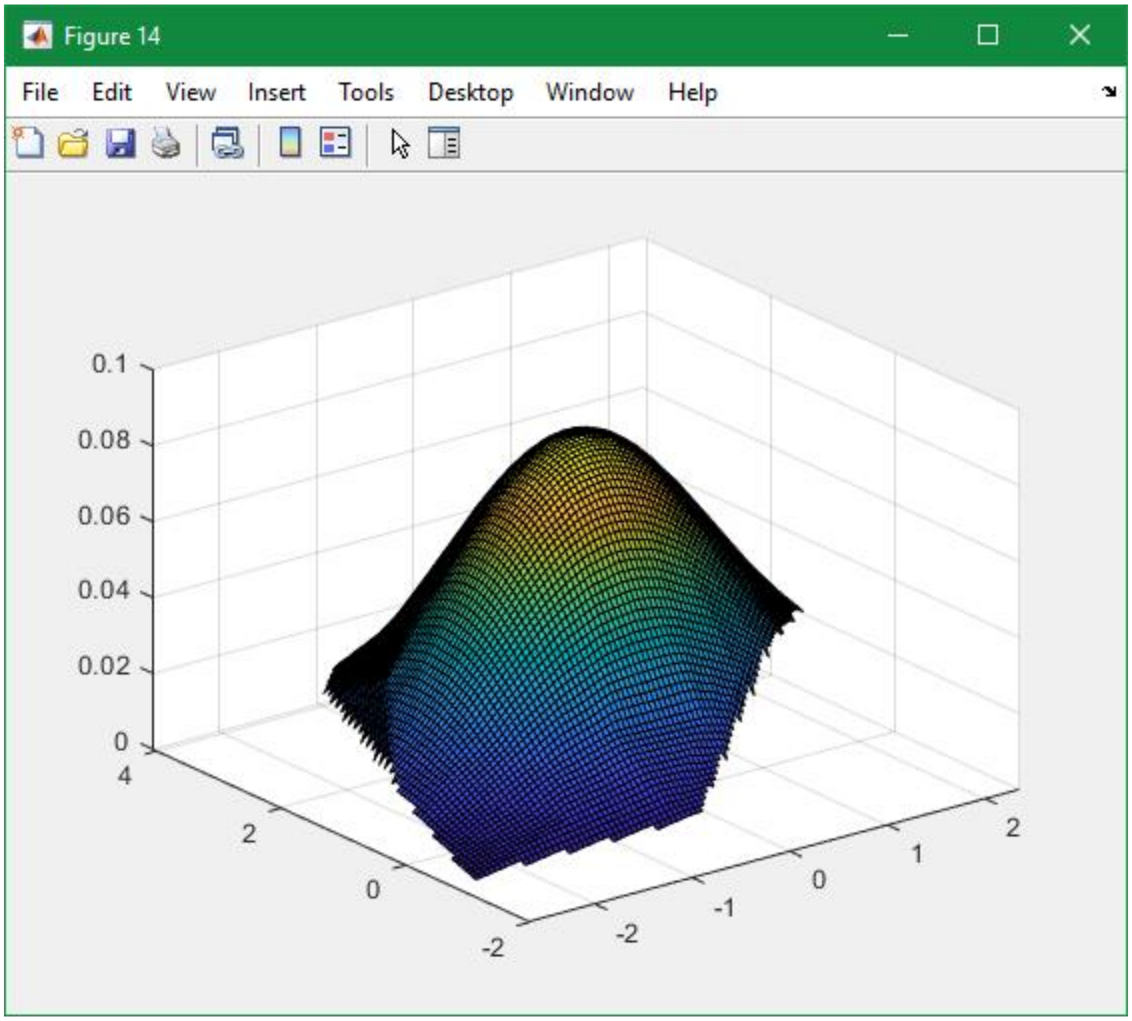


Figure 14
File
Edit
View
Insert
Tools
Desktop
Window
Help
0.1
0.08
0.06
0.04
0.02
0
4
2
0
-2
-2
-1
0
1
2

Figure 15

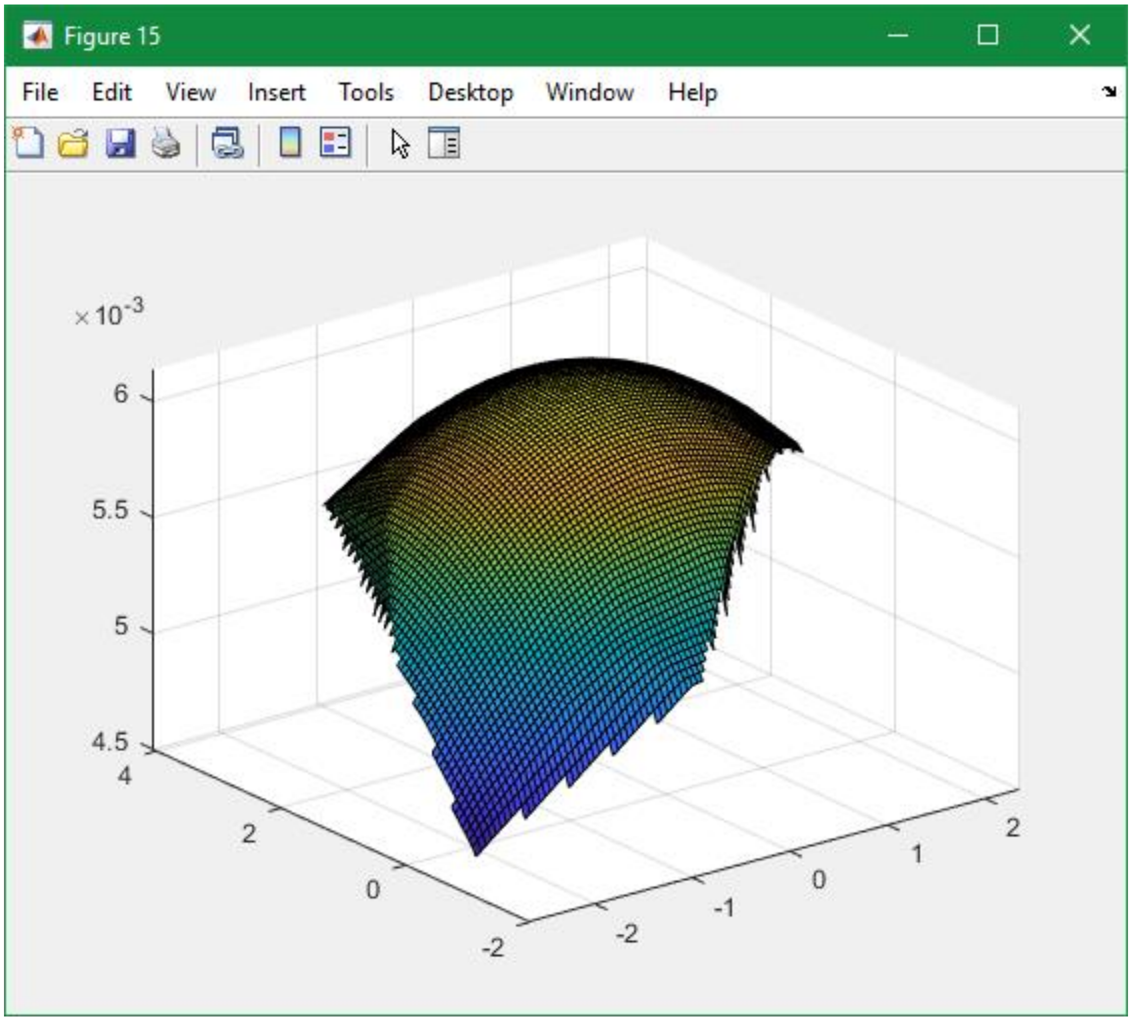


Figure 15
File
Edit
View
Insert
Tools
Desktop
Window
Help
× 10^-3
6
5.5
5
4.5
4
2
0
-2
-2
-1
0
1
2

Figure 16

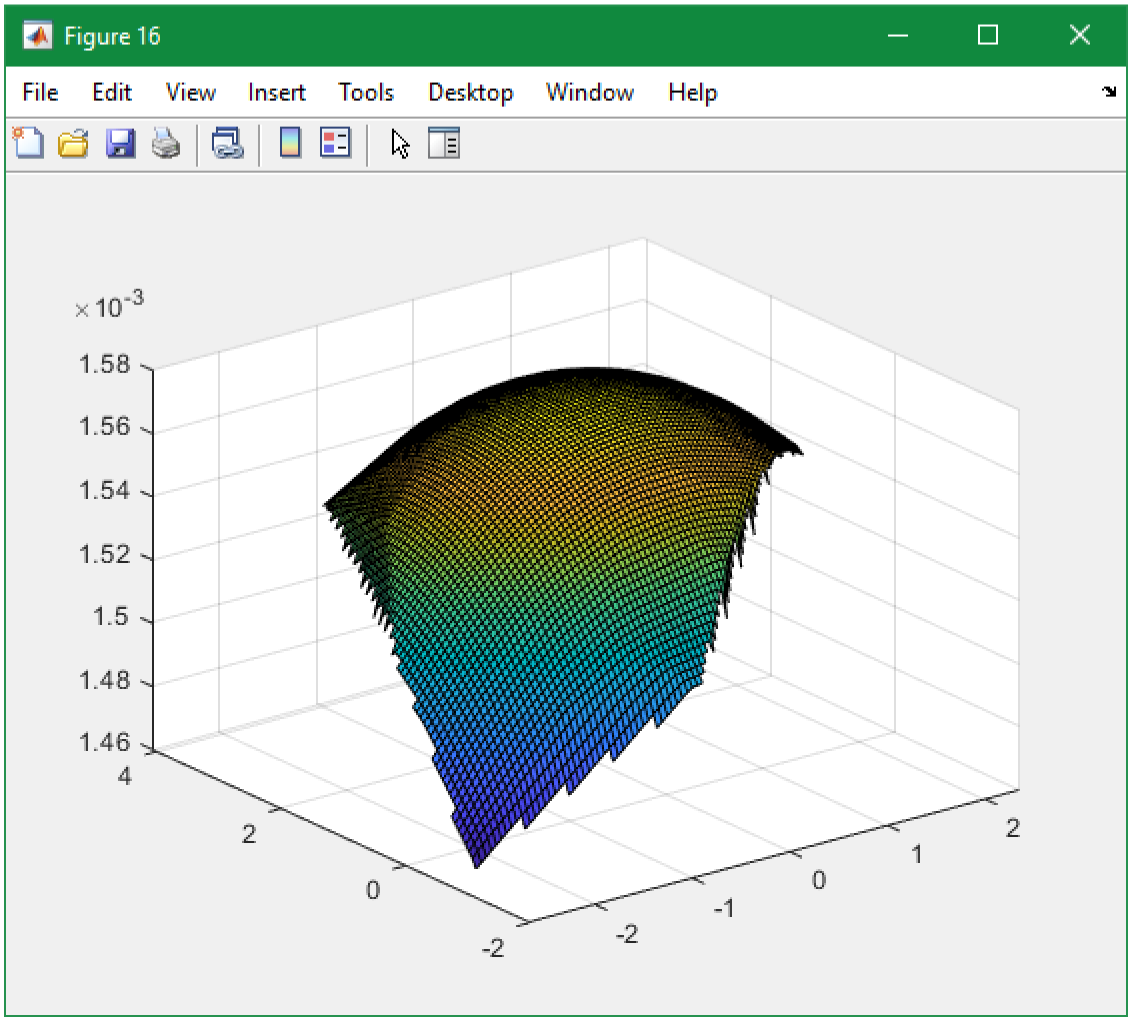


Figure 16
File
Edit
View
Insert
Tools
Desktop
Window
Help
× 10-3
1.58
1.56
1.54
1.52
1.5
1.48
1.46
4
2
0
-2
-2
-1
0
1
2