DSAA ASSIGNMENT-3 REPORT

Name: Aayush Goel
Roll No: 20171188

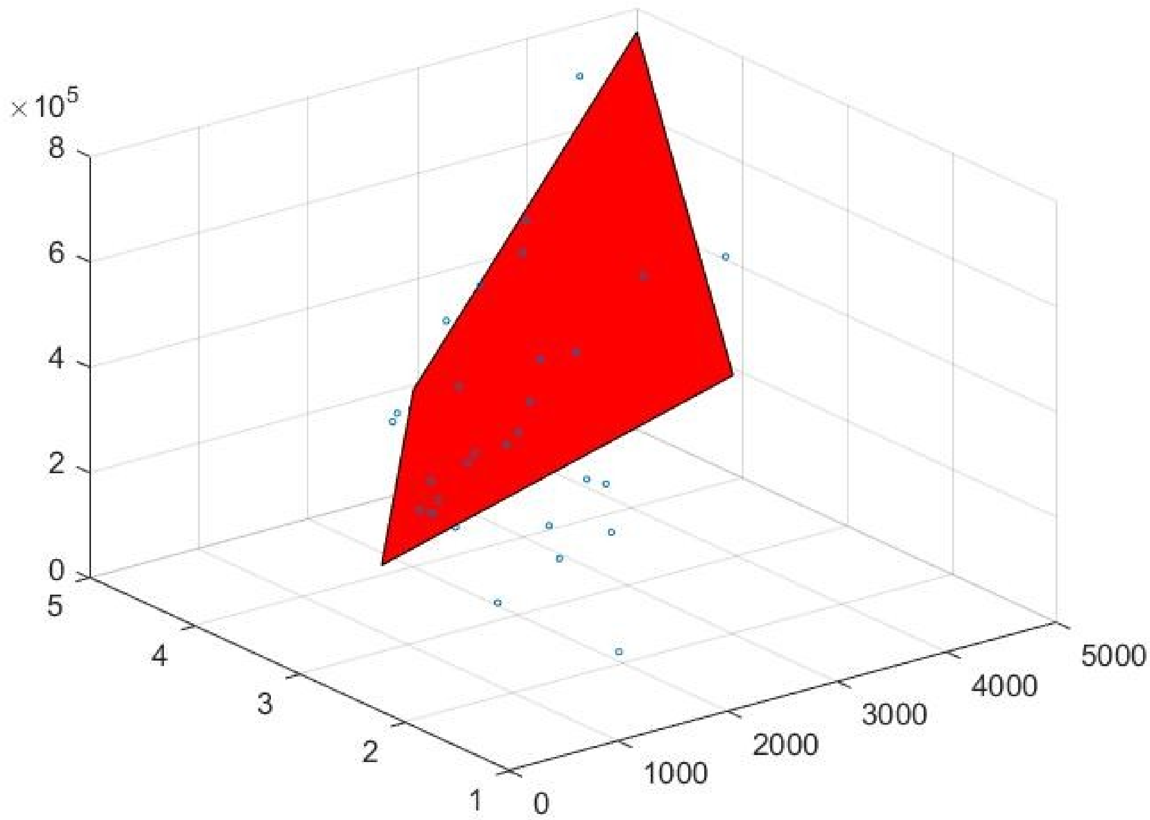
Problem 1)

Part a:

We use solution of multivariate linear regression ie,

(X'*X)-1*(X'*Price)

To calculate the various coefficients of the variables.



Coefficients are : 77731,141,-6362

Normalised Coefficients are: 213200,493300,-25450

Price of house with x1 = 1400 and x2 = 4 is 250851.5

L2 Norm: 125570

Part b:
No normalizing did not help in any way. The L2 norm without normaliztion is 125572.1203 and with normaliztion is 125572.1203. As L2 norm is equal to L1 norm so normalization does not help. But normalisation helps us to deal only with small values.

Part c:

Theoretically the mean values must satisfy the regression equation.

This is in fact the case.

We get a value of 5.82076609134674e-11 which can be easily be approximated to 0.

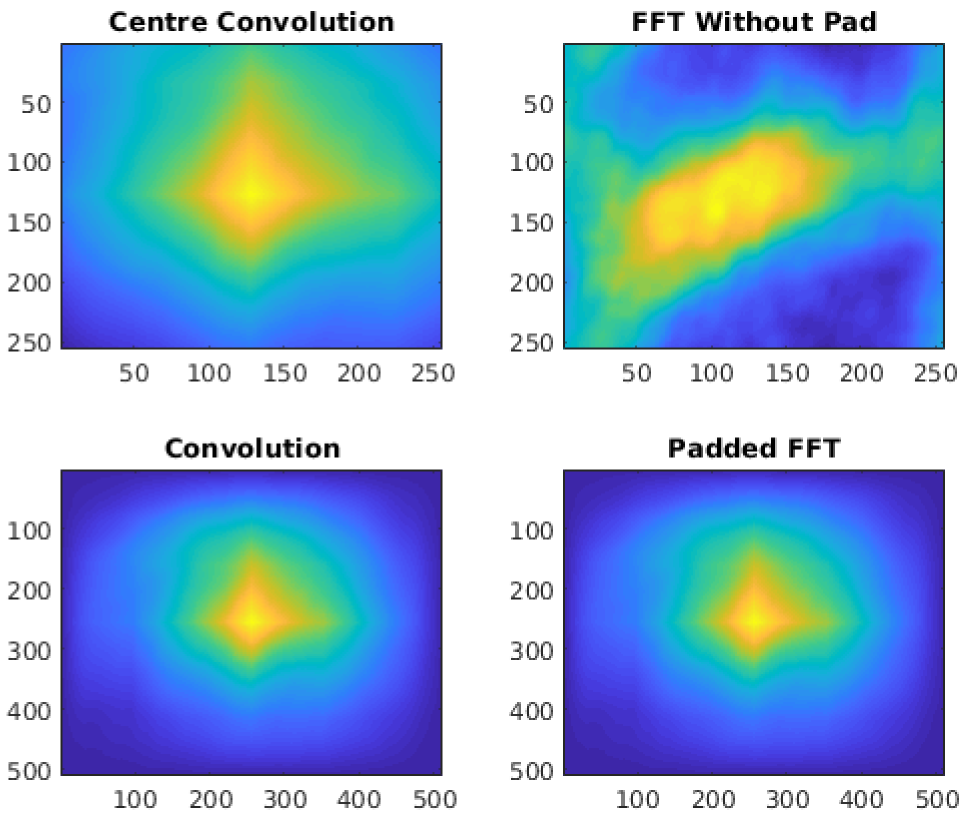
Part d:

No, this method is inefficient for large values of N as the complexity of computing the inverse is O(nlog27) where n ~ 10^6. Hence it not computationally efficient.

Problem 2)
The two images that are used is ‘sample.jpeg’ and ‘cameraman.tif’
1. iDFT[FH] is not equal to f*h as it can be seen by there average squared difference(which is calculated below and is very high).

2. The average squared difference between pixel values in iDFT[FH] and the central 256 X 256 portion of f * h is 1.1108×10^{11}.

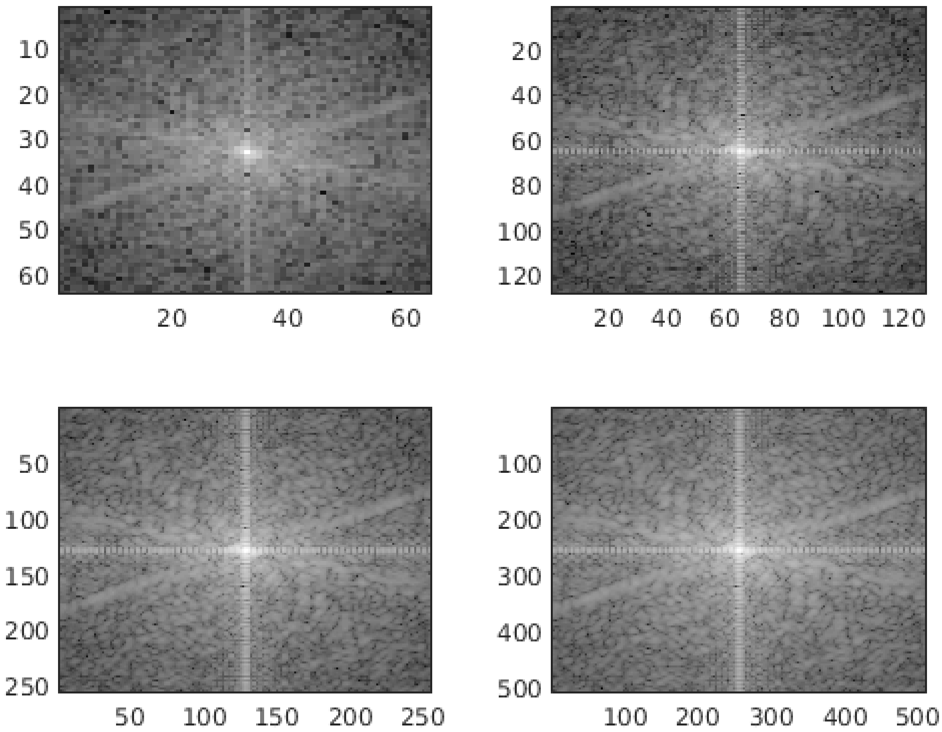
3. After padding the fft with zeroes at the end we get the both equal with minimal error of range 10 power -11 which can be neglected.



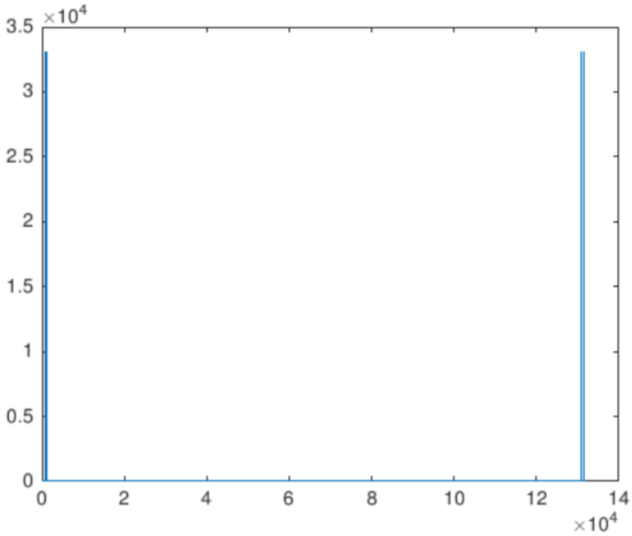
Problem 3)

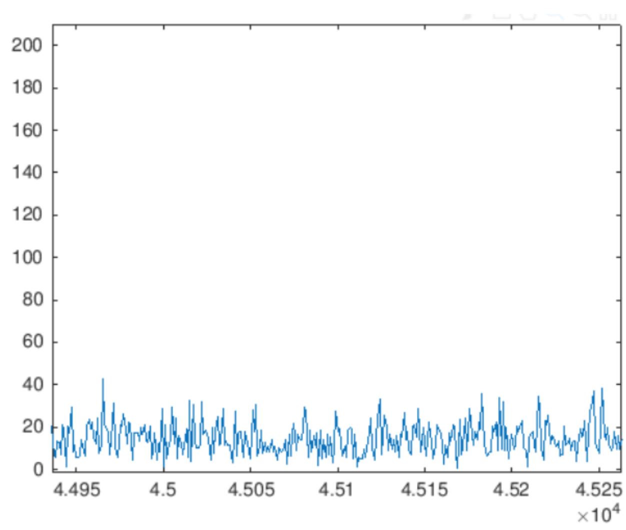
We observe that the central portion remains same even after padding. The only difference is that we see horizontal and vertical disturbances in the signal.

These occur as we introduce discontinuity when we pad with zeros. The effect is seen as sic function in the Fourier Domain.



Problem 4)
First we take FFT of the X and plot it. Here two peak frequency and noise of very little amplitude throughtout the signal, as shown.

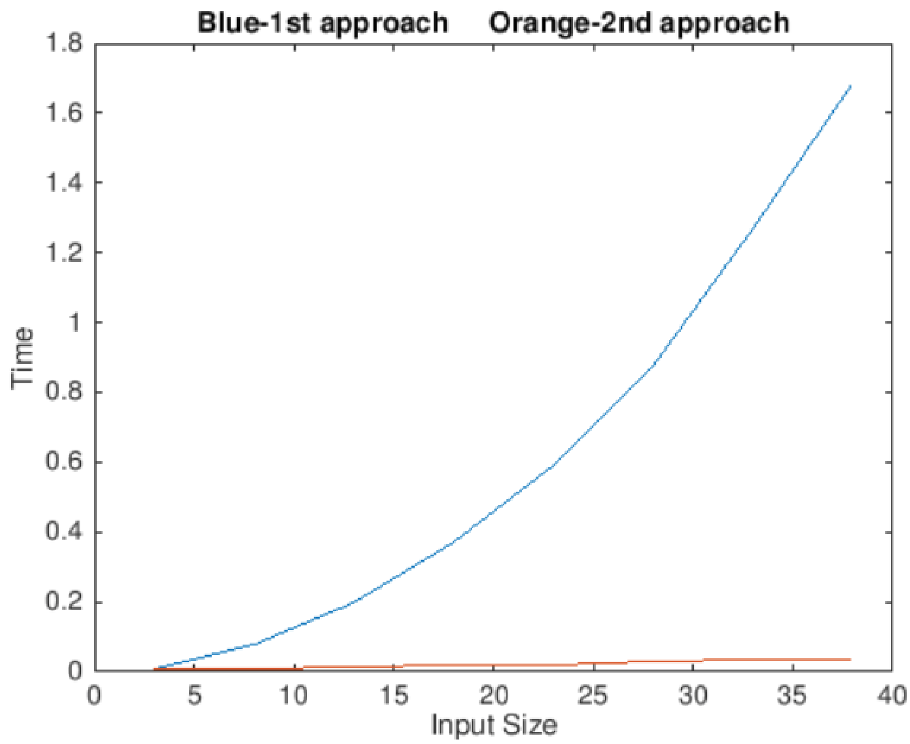




Now as it is clear from the FFT diagram that we have high frequency noise so we use low pass filter to remove these noise. As we observe that there are two frequency of value 882 and 1321. Thus any filter that inhabits frequecy higher thatn 1321 will do. I have used lowpass filter with cutoff frequency of 2000.

Problem 5)
It is given that there is an overlap of 3 to 5 seconds. So first we take 5 seconds(maximum overlap) of segment at starting and ending of all the sounds. Then to compare them we take xcorr of each and every starting and ending segment and take out top 4 values from it and store their sequence and thus in this way we get the required sequence which is **3,5,1,2,4**. For denoise I have used a mixture of median , sgolay and gaussian filter.

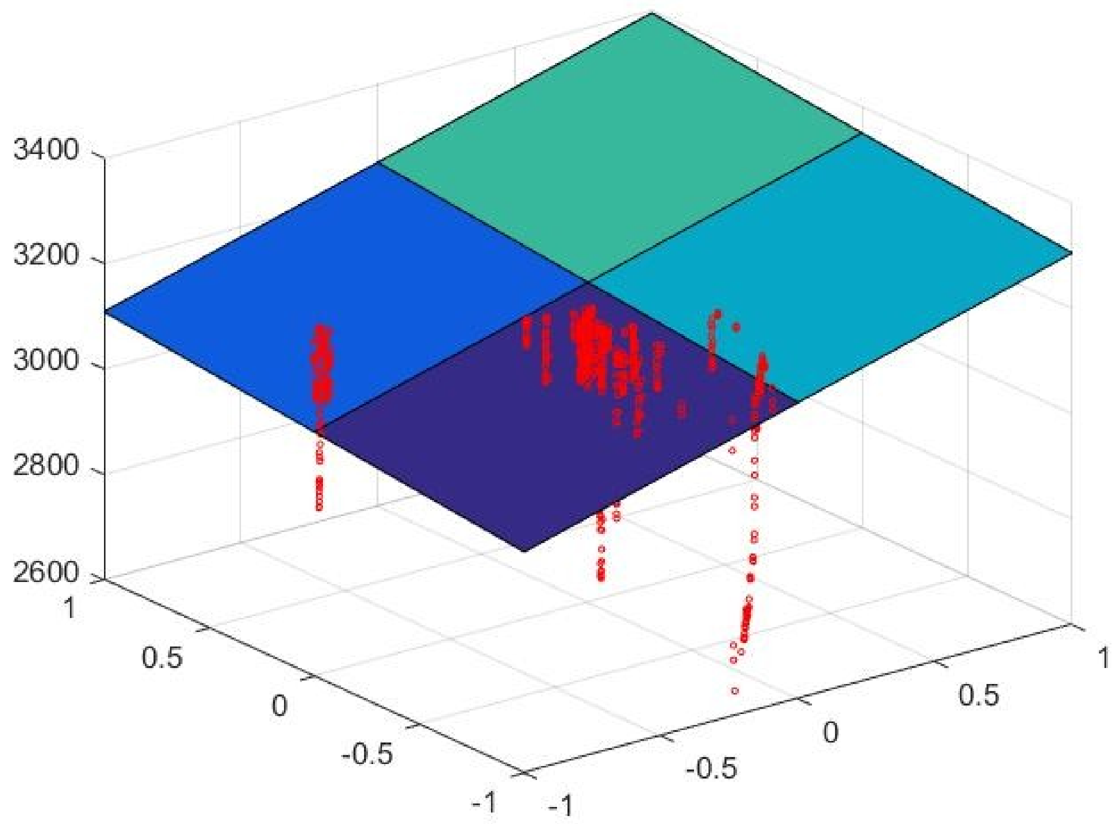
Problem 6)
If you have two numbers a and b such that both are not equal then their average will always come between them. Which means that the average will be less than the higher value and thus the higher value gets diminished while the lower value increases and thus in a way implementing low pass filter. When we use adjacent pixels to calculate average we observe that the computational time decreases significantly as shown in the figure below.



1st approach is by recalculating the average everytime.
2nd approach is by using adjacent pixel for calculating average.

Result:

Problem 7)



Gradient Descent is used to calculate the linear coeffecients for the inputs. The 90% data is used to calculate the values and remaining 10% as test data.

1000 iterations are used to calculate theta in each iteration to reduce the error.

The theta values obtained for batch are 3019.3 , 44.5 , 142.7

The theta values obtained for stochastic are 1975.4,1020.5,977.5

The input has been normalised so that range is between 0 and 1.

Batch l2 norm of difference between predicted and actual output is 1195000.

Stochastic l2 norm of difference between predicted and actual output is 1182000.

The results show Batch is worse than Stochastic possibly because of the updates by each entry in the table.