S7001E Lab 3

Calle Rautio
calrau-1@student.ltu.se

Jim Agnevik
jimagn-8@student.ltu.se

Contents

1 Introduction

This lab investigates prediction and system identification using MATLAB. For prediction, it explores the theory and implementation of a linear minimum mean square error (LMMSE) predictor on a moving average (MA) process of order three. The task is to theoretically determine predictor parameters and then simulate and analyze the predictive accuracy on a generated random sequence, observing any delay or startup effects.

The system identification section focuses on estimating an unknown system's impulse response by analyzing its input and output signals. Using auto-correlation and cross-correlation functions, participants compute and interpret power spectral densities and, through an LMMSE approach, estimate the system parameters. They evaluate mean-square errors across different model orders to identify an optimal model configuration. The goal is to develop practical skills in system modeling and analysis using signal processing techniques in MATLAB.

2 Prediction

2.1 Assignment 1

To solve the problem presented in the lab report, which requires finding the theoretically optimal Linear Minimum Mean Square Error (LMMSE) predictor for a Moving Average (MA) process of order 3, we will use the Yule-Walker equations. We will first derive these equations before applying them.

The goal is to predict $(X[n+1])$ using the previous three values of the process: $(X[n], X[n-1], X[n-2])$. To achieve this, we need to determine the coefficients (a_0, a_1, a_2) such that the predictor has the following form:

$$\hat{X}[n+1] = a_0 X[n] + a_1 X[n-1] + a_2 X[n-2] \tag{1}$$

The MA(3) process is given in terms of the coefficients (d_0, d_1, d_2, d_3) as follows:

$$X[n] = 2W[n] + W[n-1] + 2W[n-2] + W[n-3] \tag{2}$$

The LMMSE predictor is derived by minimizing the mean squared error between $(X[n+1])$ and the predicted value $(\hat{X}[n+1])$. To accomplish this, we calculate the autocorrelation function of the process and use it to solve the Yule-Walker equations.

The autocorrelation function of the MA process can be computed using the coefficients (d_0, d_1, d_2, d_3), and the autocorrelation of the white noise sequence $(W[n])$, which is zero for all lags except at lag 0, where it equals the variance of $(W[n])$. The variance of $(W[n])$ is given as 3 in the lab report, $(\mathrm{Var}(W[n]) = 3)$, from $(W[n] \sim N(0,3))$.

Now, we calculate the autocorrelation $(R_{X[k]})$ of the process at different lags (k):

$$R_X[0] = 3(2^2 + 1^2 + 2^2 + 1^2) = 3(4+1+4+1) = 30 \tag{3}$$

$$R_X[1] = 3(2 \cdot 1 + 1 \cdot 2 + 2 \cdot 1) = 3(2+2+2) = 18 \tag{4}$$

$$R_X[2] = 3(2 \cdot 2 + 1 \cdot 1) = 3(4+1) = 15 \tag{5}$$

$$R_X[3] = 3(2 \cdot 1) = 6 \tag{6}$$

$$R_X[k] = 0 \quad \text{for} \quad k > 3 \tag{7}$$

We can now set up the Yule-Walker equations to determine the coefficients (a_0, a_1, a_2):

$$\begin{aligned} R_X[1] &= a_0 R_X[0] + a_1 R_X[1] + a_2 R_X[2] \\ R_X[2] &= a_0 R_X[1] + a_1 R_X[0] + a_2 R_X[1] \\ R_X[3] &= a_0 R_X[2] + a_1 R_X[1] + a_2 R_X[0] \end{aligned} \tag{8}$$

Substituting the previously calculated values of $(R_{X[k]})$:

$$\begin{aligned} 18 &= a_0 \cdot 30 + a_1 \cdot 18 + a_2 \cdot 15 \\ 15 &= a_0 \cdot 18 + a_1 \cdot 30 + a_2 \cdot 18 \\ 6 &= a_0 \cdot 15 + a_1 \cdot 18 + a_2 \cdot 30 \end{aligned} \tag{9}$$

This system of linear equations in (a_0, a_1, a_2) can be solved to find the optimal predictor coefficients. Using a calculator, the solutions are as follows (which match the results obtained from MATLAB):

$$a_0 = 0.5282, \quad a_1 = 0.3462, \quad a_2 = -0.2718 \tag{10}$$

Therefore, the LMMSE predictor is given by:

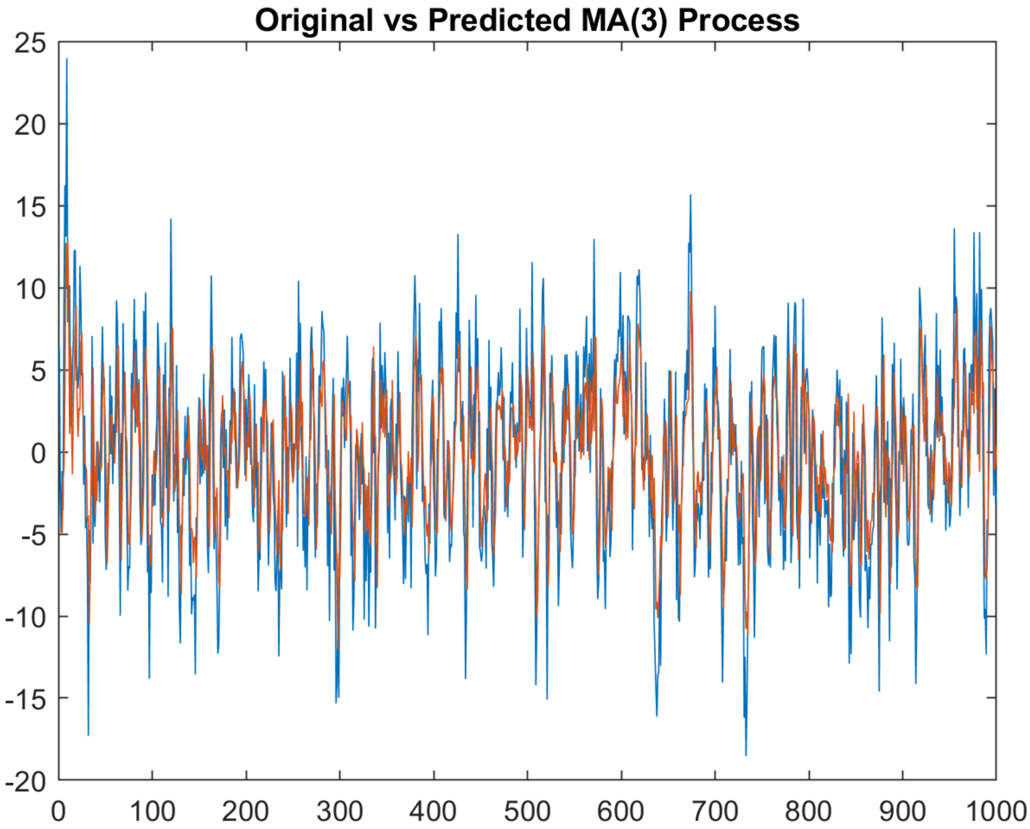
$$\hat{X}[n+1] = 0.5282X[n] + 0.3462X[n-1] + 0.2718X[n-2] \tag{11}$$

Image from MATLAB, the code will be appended together with the rest of the code in Prediction.

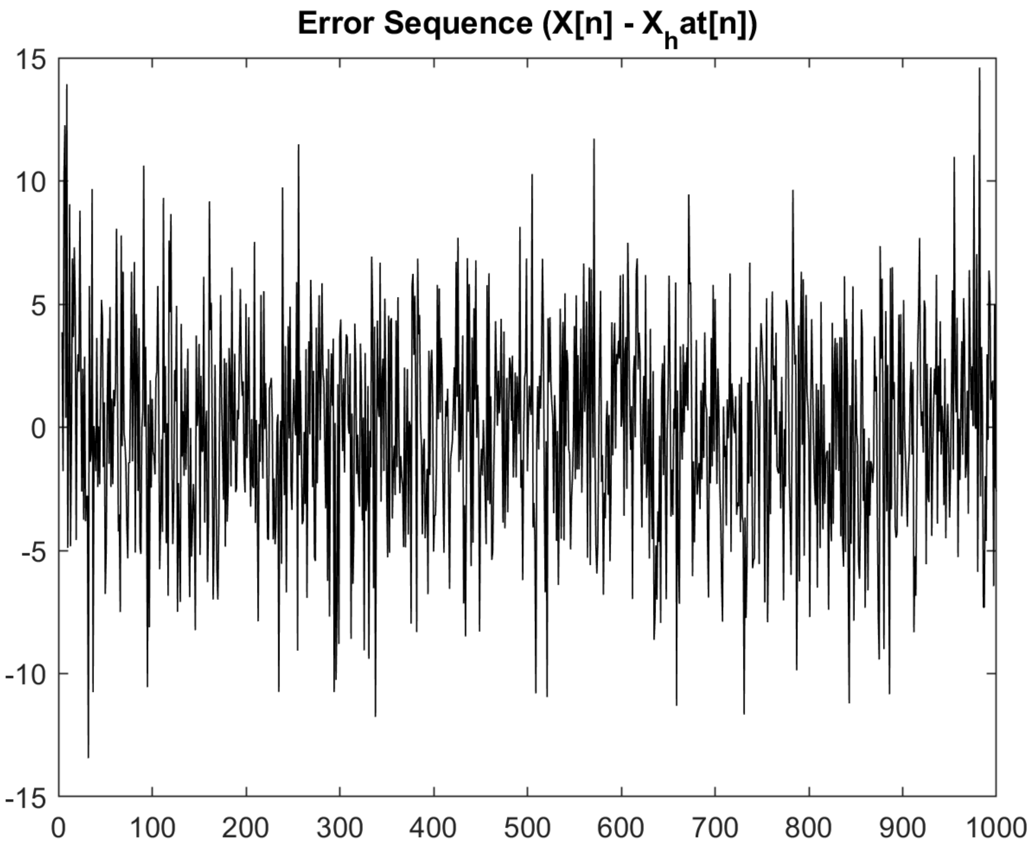
```
The calculated coefficients are:
a_0 = 0.5282
a_1 = 0.3462
a_2 = -0.2718
```

2.2 Assignment 2

Original vs Predicted MA(3) Process image from MATLAB.



Error Sequence from Matlab.



Conclusion:

The predictor captures the general trend of the original MA(3) process well in our opinion, though there are some discrepancies, particularly in high-frequency variations throughout the entire signal. Overall, the predictor performs effectively, with minor deviations arising from noise and model limitations.

When it comes to the error sequence it fluctuates around zero, indicating the predictor is unbiased, though there are occasional significant deviations, suggesting the predictor occasionally struggles with randomness in the process. The error sequence is stationary, with no visible delay or startup issues, implying the predictor works effectively from the beginning.

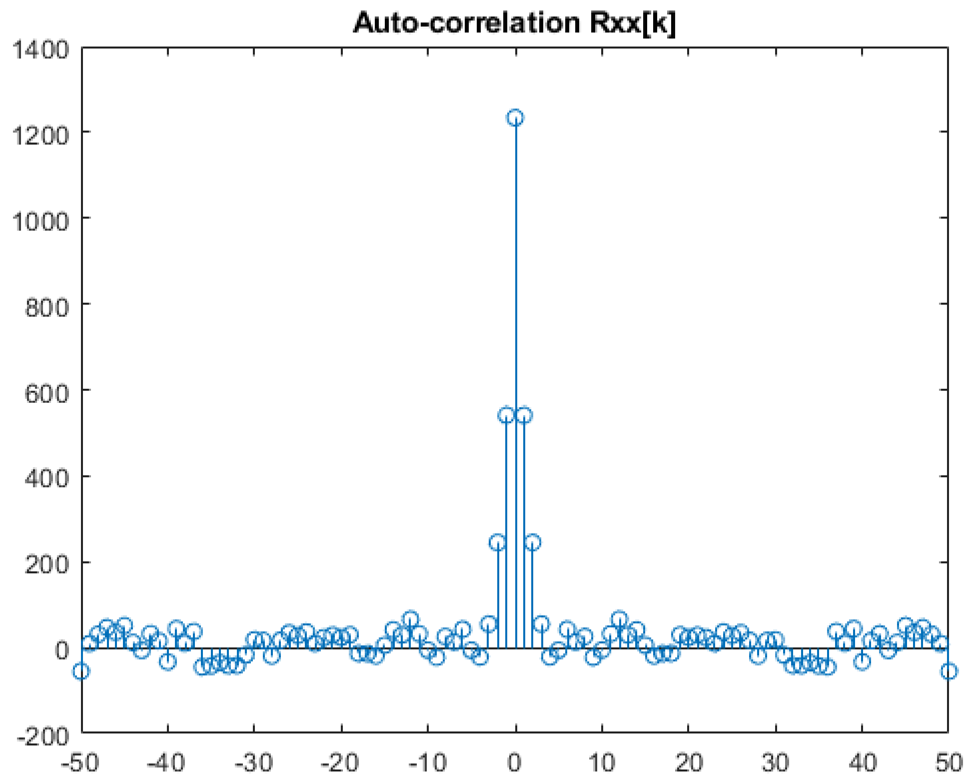
In conclusion, the predictor approximates the MA(3) process well, with occasional significant errors, but without any startup delay.

3 System Identification

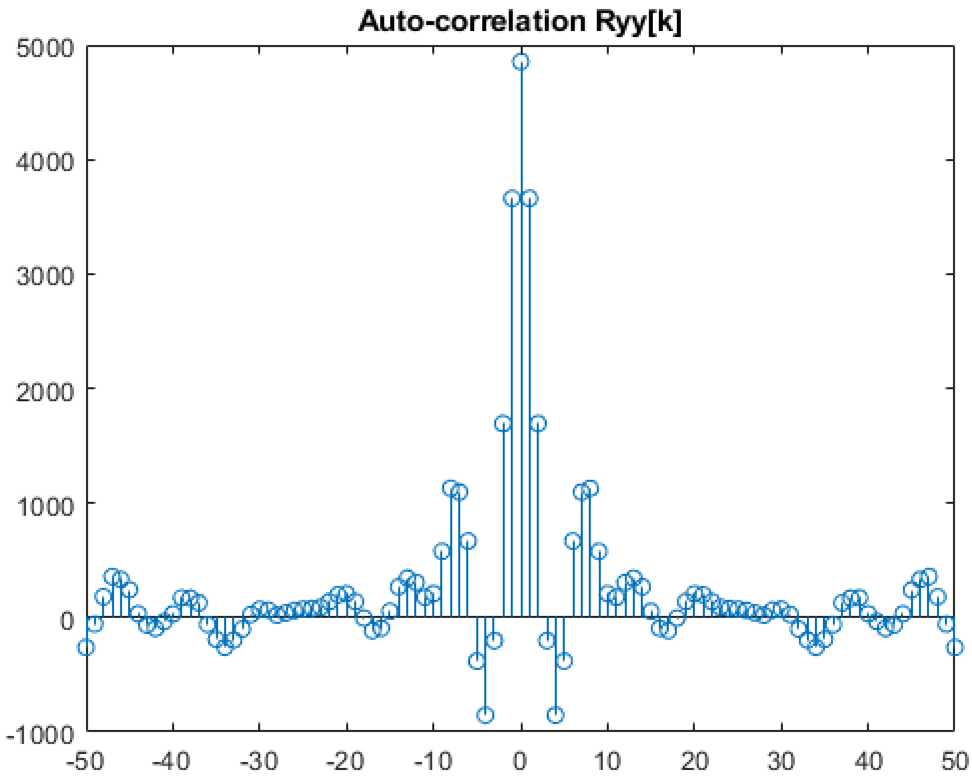
The objective of this part was to to find the impulse response $h[n]$ using the input $x[n]$ and output $y[n]$ through system identification. This is possible through Auto-correlation and cross-correlation for the input $x[n]$ and output $y[n]$.

3.1 Auto-correlation and cross-correlation

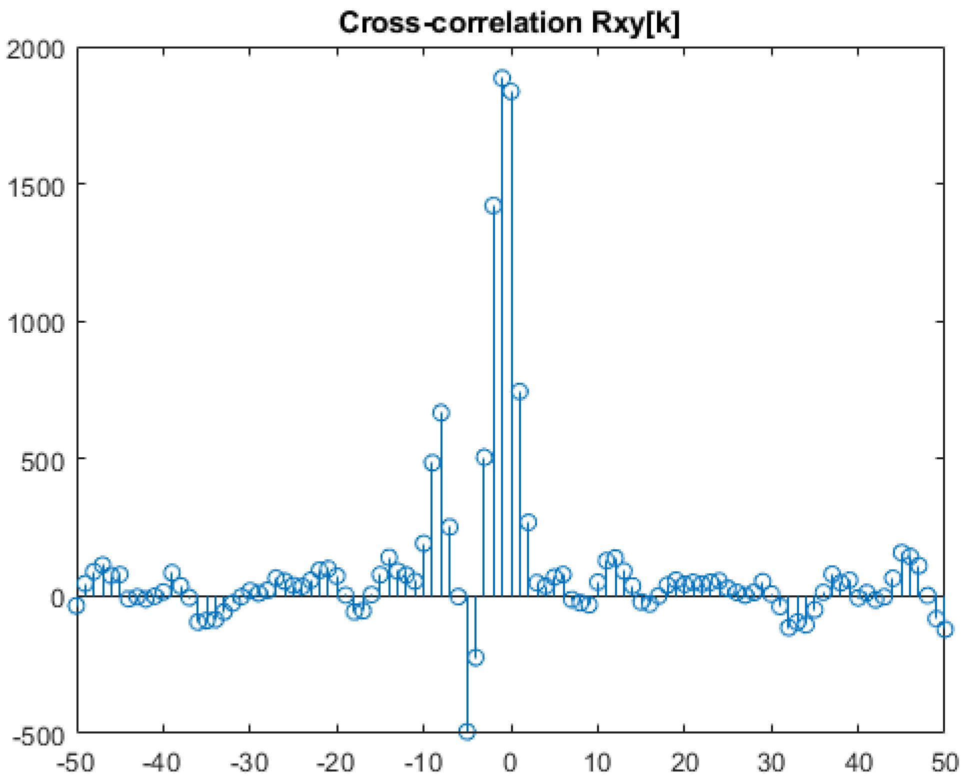
Calculating the auto-correlation and cross-correlation gives the following results.



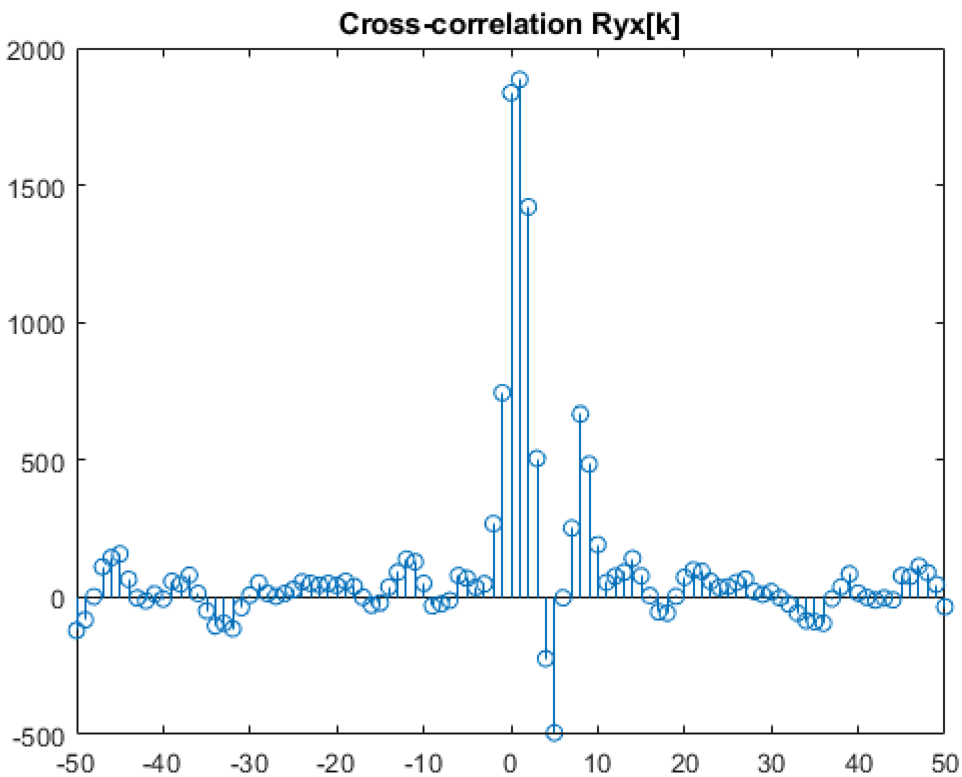
Auto-correlation for input x with lag 50.



Auto-correlation for system output y with lag 50.



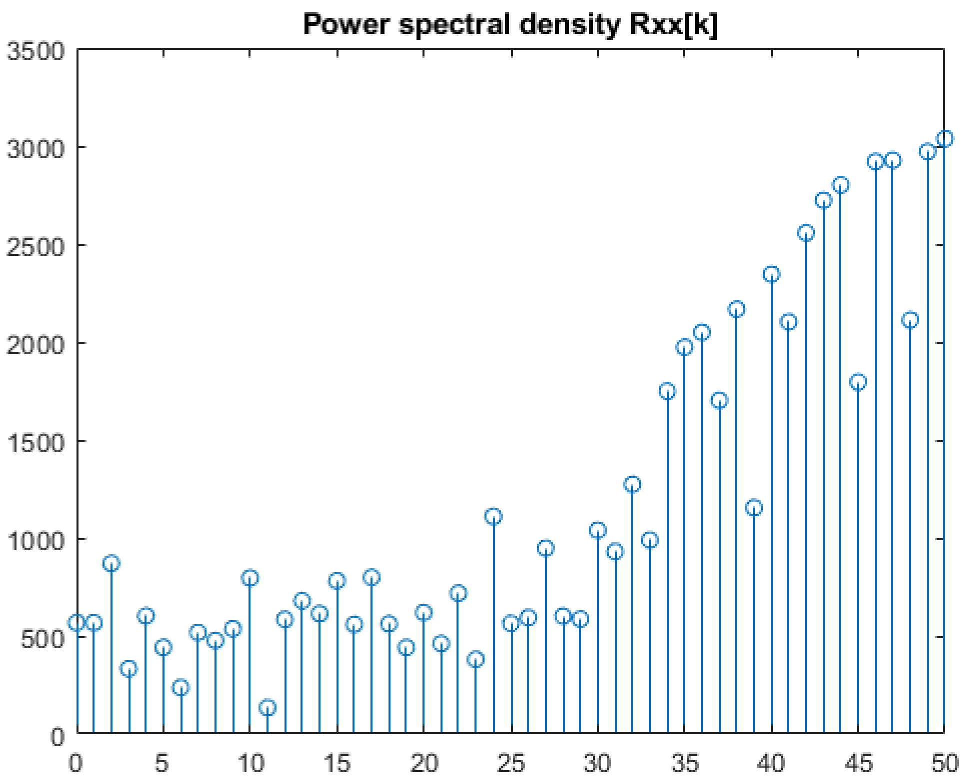
Cross-correlation between the input and output of the system with y lag 50.

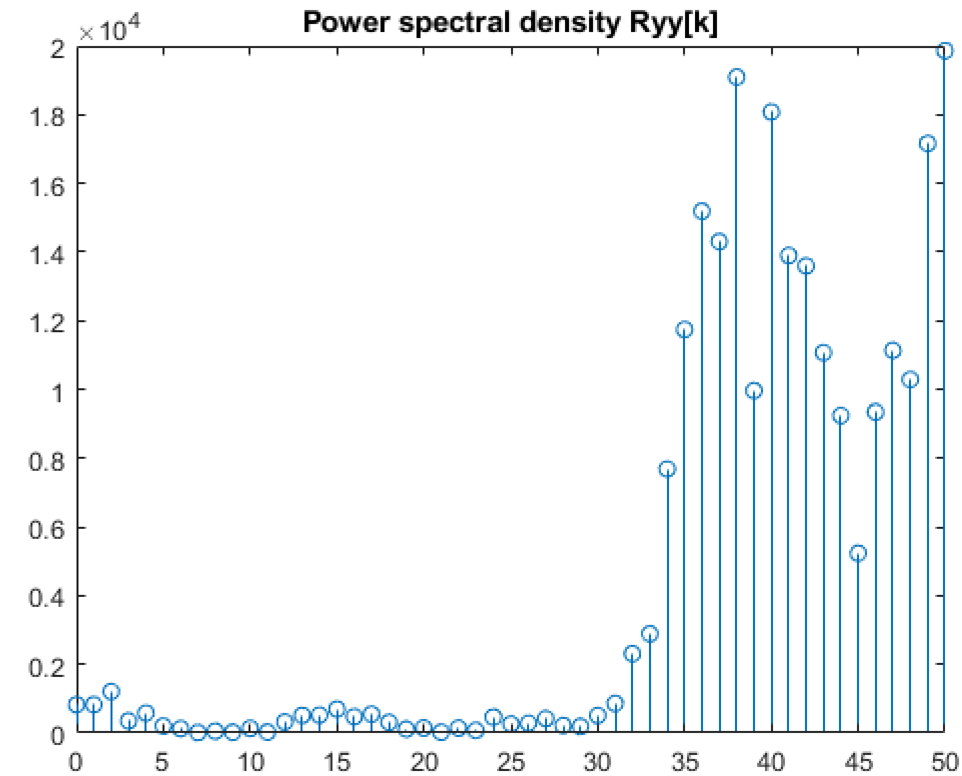


Cross-correlation between the input and output of the system with x lag 50. This is the one that will be used in the system identification.

3.2 Power spectral densities

Doing Fourier transform on the cross-correlation gives the power spectral for the input and output sequences.





1. Input Power Spectral Density $R_{\{xx\}}[k]$:
   - The PSD $R_{\{xx\}}[k]$ shows a steadily increasing trend with a series of peaks, which suggests that the input signal might contain components across a wide range of frequencies, with higher power concentrated at higher frequencies.
2. Output Power Spectral Density $R_{\{yy\}}[k]$:
   - The PSD $R_{\{yy\}}[k]$ shows a prominent set of peaks, particularly in the higher frequency range (around (k = 40) to (k = 50)), with much larger magnitudes compared to the input.
   - This indicates that the system amplifies higher frequency components significantly, implying it may have characteristics similar to a high-pass filter or an amplifier that boosts certain frequencies.

Conclusion:

The plots suggest that the system has a tendency to enhance higher-frequency components of the input signal, possibly due to resonance at certain frequencies or inherent filtering characteristics that attenuate low frequencies while amplifying high ones.

3.3 Estimated system output

Assuming the unknown system is a linear time (or shift) invariant system of the form.

$$y[n] = \sum_{k=0}^{M} h[k]x[n-k] \tag{12}$$

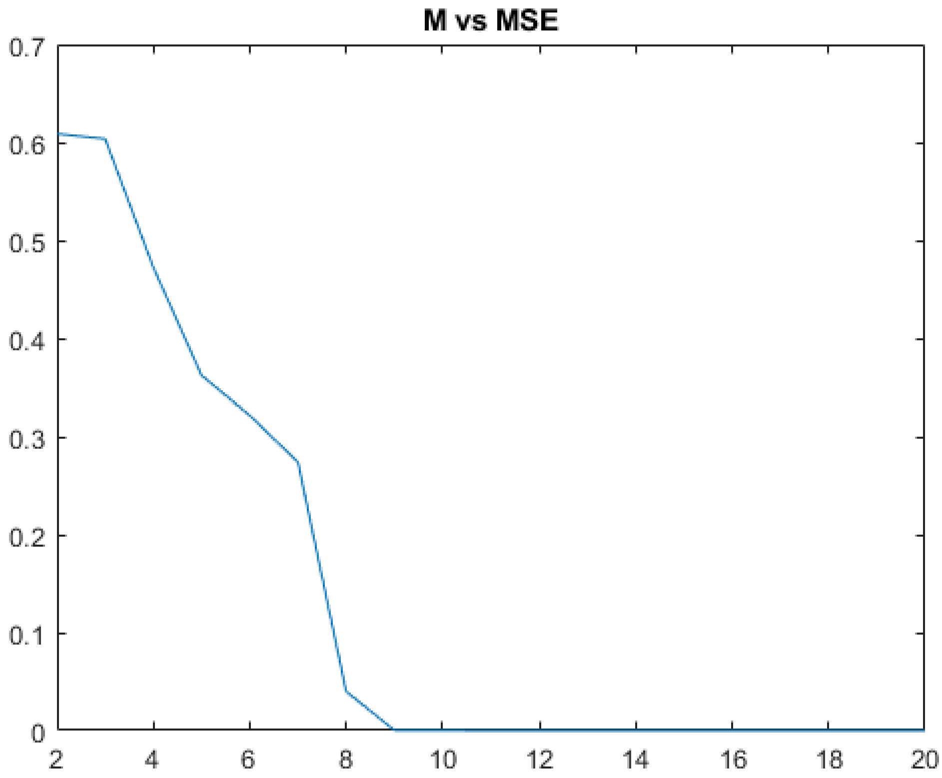
Using the estimated correlations, to compute an estimated system parametrization $\hat{h}[n]$. The estimated system's output is then computed as $\hat{y}[n]$.

$$\hat{y}[n] = \sum_{k=0}^{M} \hat{h}[k]x[n-k] \tag{13}$$

The method that was used followed the LMMSE criterion and made use of the orthogonality principle.

$$E\{(y[n] - \sum_{k=0}^{M} \hat{h}[k]x[n-k])x[m]\} = 0 \qquad form = n, n-1, ..., n-M \tag{14}$$

This was converted to vector-matrix form and calculated in Matlab. Estimating the first $M = 2, ...20$ and calculating the MSE as seen in graph.



As M increases the MSE decreases until a certain point were the estimated system becomes to complex. The initial decrease of MSE with higher M is because more of the systems gets captured in the estimator.
From the Graph the M value that gives the lowest MSE was determined to be $M = 11$ before it starts to increase once again. Giving us model order $M = 11$.