In The Name Of God

Shiraz University

دانشگاه شیراز

Estimation Theory

Dr. Somayeh Afrasiabi

HW 5

Alireza Garmsiri

40260337

Introduction

In this project, we focus on signal estimation and reconstruction using basis functions and regularized least squares techniques. The goal is to estimate the underlying clean signal from noisy data by fitting a model composed of cosine basis functions. We explore how the number of terms in the basis functions affects the quality of the reconstructed signal and the performance of the estimation in terms of error and model complexity. The analysis includes two main parts: one using regularized least squares estimation, and the other using recursive estimation with a growing number of basis functions. We also examine how the **Akaike Information Criterion (AIC)** helps in model selection by penalizing overly complex models with too many terms.

Implementation

Components of the Code

Signal Generation (generate_signal):

This function generates the clean signal based on a cosine function and adds Gaussian noise to simulate real-world noise. The noisy signal is then used for fitting models and estimation.

Basis Construction (construct_cosine_basis):

This function constructs a matrix of cosine basis functions. The number of terms used in the basis functions can be adjusted, which directly influences the complexity of the model and its ability to fit the noisy signal.

Regularized Least Squares Estimation (fit_signal_regularized):

This function fits the noisy signal using regularized least squares estimation. It solves for the coefficients that best fit the signal in the cosine basis, with regularization to prevent overfitting.

Recursive Estimation and Cost Function Calculation (recursive_estimation):

In the second part, this function iteratively adds more terms to the basis function and calculates the coefficients for the best fit at each step. The cost function at each step represents the sum of squared residuals, helping to monitor how well the model fits the noisy data.

Model Selection and AIC Calculation (compute_aic):

The AIC is computed to help evaluate the model and balance between fit and complexity. It penalizes models with more terms, encouraging simpler, more generalizable models.

Plotting Functions:

Several plotting functions are included to visualize the results of the estimation process, including the **Sum of Squared Error (SSE)** vs the number of cosine terms, the **AIC** vs the number of terms, and the comparison between the noisy signal and the reconstructed signal for different numbers of terms.

These plots help analyze how the fit improves as the number of terms increases and how the AIC helps in selecting the optimal model complexity.

Methodology

The methodology consists of two primary parts:

Signal Generation and Data Preparation:

We generate a clean signal that follows a cosine pattern and then introduce Gaussian noise to simulate real-world measurement errors. This noisy signal is used for both parts of the analysis.

n the first part, we construct a set of cosine basis functions and fit the signal using regularized least squares.

In the second part, we use a recursive approach where we progressively increase the number of cosine basis functions, reconstruct the signal, and monitor the cost function.

Regularized Least Squares Estimation:

The first part of the code applies regularized least squares estimation to fit a series of cosine basis functions to the noisy signal. This method is used to estimate the coefficients that best represent the noisy signal in terms of the chosen basis functions.

A regularization term is included to prevent overfitting by adding a penalty for using too many basis functions.

Recursive Estimation:

In the second part, we employ a recursive approach to progressively increase the number of basis functions used to fit the noisy signal. The coefficients for the best fit are estimated iteratively for different numbers of basis terms.

The cost function, representing the residual error between the actual noisy signal and the fitted signal, is computed at each iteration to help monitor the reconstruction error.

Model Selection using AIC:

For both methods, the performance of the model is assessed using the Akaike Information Criterion (AIC), which helps in choosing the best model by penalizing models with too many terms. The AIC balances model fit and complexity, ensuring that a more complex model doesn't simply overfit the data

Result

Regularized Least Squares Estimation:

As the number of cosine terms in the basis functions increases, the fit between the noisy signal and the estimated signal improves. However, when the number of terms is too large, the model may overfit the data, resulting in a larger error on the test data.

The plot of **Sum of Squared Error (SSE)** shows that as more cosine terms are added, the error initially decreases, but after a certain point, it begins to increase again due to overfitting.

The **Akaike Information Criterion (AIC)** plot indicates that the best model complexity is where the AIC is minimized. This corresponds to the point where the error is sufficiently low without introducing unnecessary complexity.

Recursive Estimation:

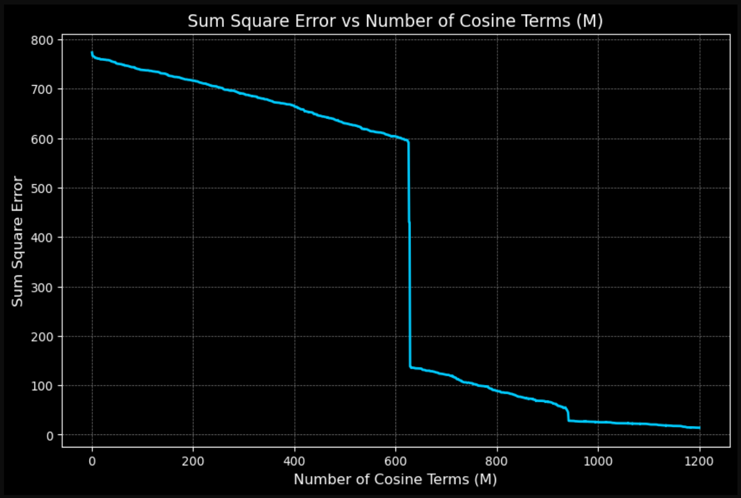
The recursive estimation process shows how the reconstruction of the signal improves as more basis terms are included. At each step, the cost function is calculated to monitor how well the signal is being reconstructed.

The reconstructed signal gets closer to the clean signal as more terms are used, and the error decreases. The cost function starts to flatten out as the model reaches a point where adding more terms doesn't significantly improve the reconstruction.

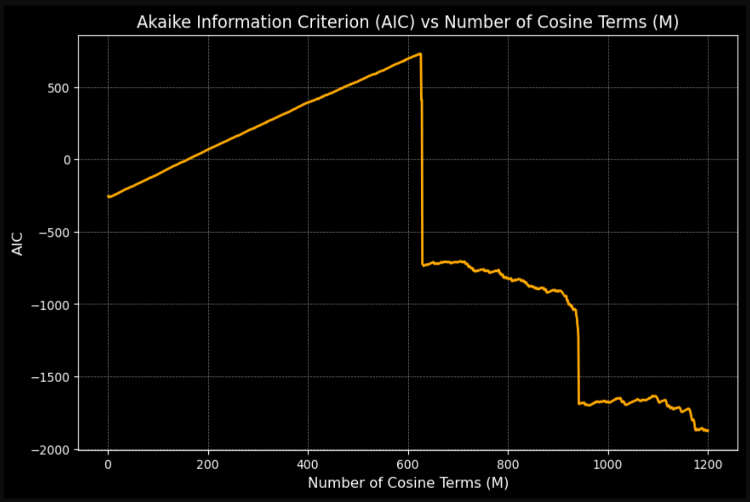
Model Selection:

The AIC plot clearly demonstrates the trade-off between model fit and complexity. Initially, as more terms are added, the fit improves, but after a certain number of terms, the complexity penalty of AIC outweighs the benefits of a better fit. The optimal model complexity is found by balancing these factors.

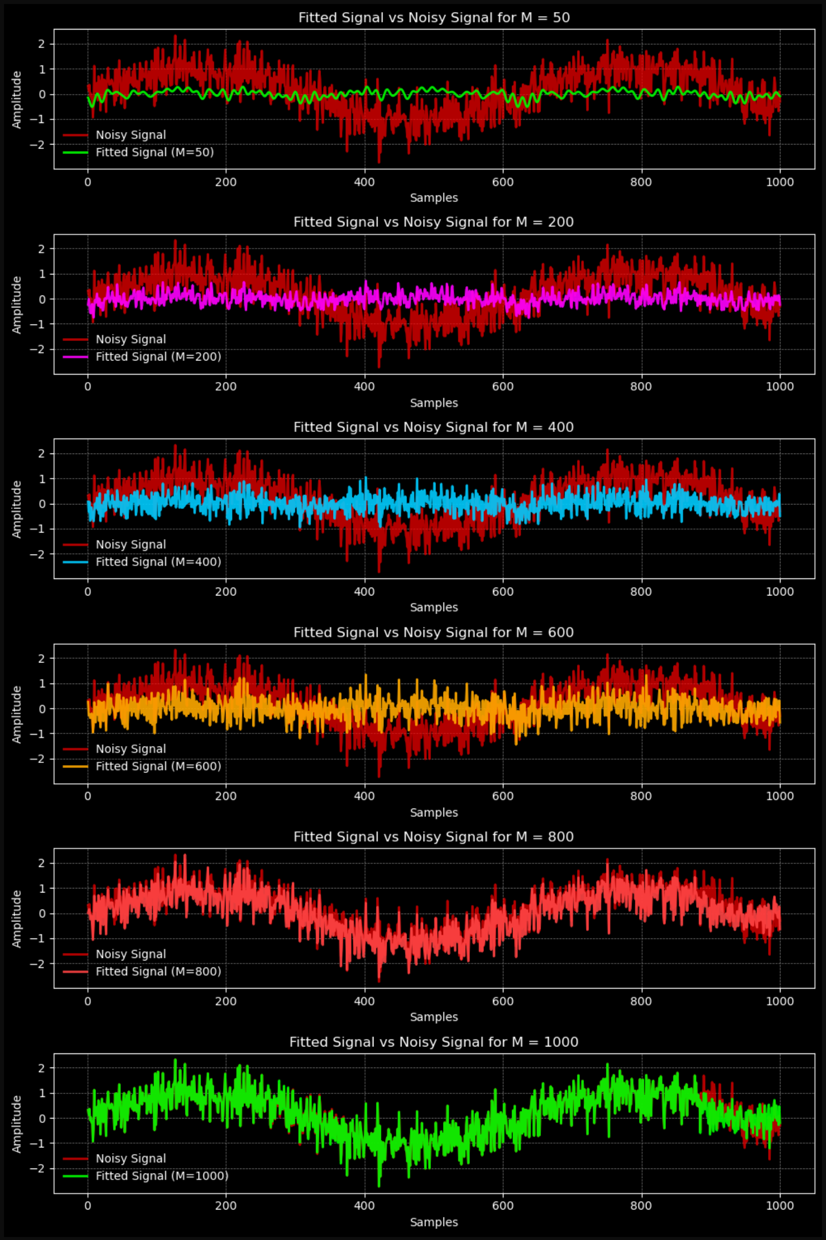
Plots



Sum Square Error vs Number of Cosine Terms (M)
Sum Square Error
Number of Cosine Terms (M)

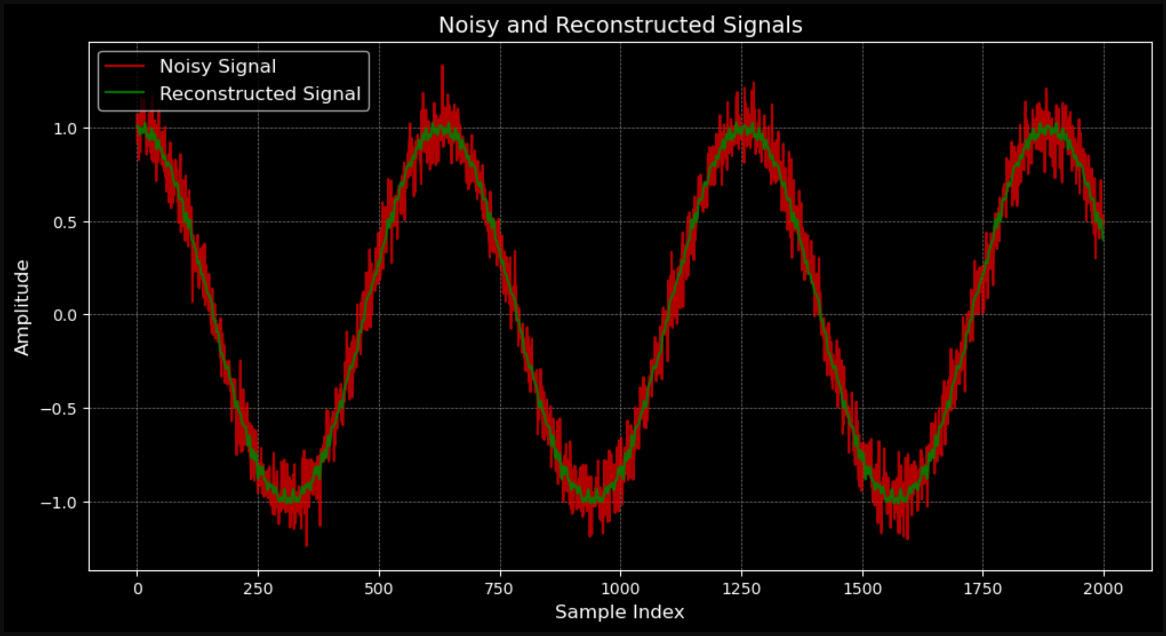


Akaike Information Criterion (AIC) vs Number of Cosine Terms (M)
AIC
Number of Cosine Terms (M)



Fitted Signal vs Noisy Signal for M = 50
Amplitude
Samples
Noisy Signal
Fitted Signal (M=50)
Fitted Signal vs Noisy Signal for M = 200
Noisy Signal
Fitted Signal (M=200)
Fitted Signal vs Noisy Signal for M = 400
Noisy Signal
Fitted Signal (M=400)
Fitted Signal vs Noisy Signal for M = 600
Noisy Signal
Fitted Signal (M=600)
Fitted Signal vs Noisy Signal for M = 800
Noisy Signal
Fitted Signal (M=800)
Fitted Signal vs Noisy Signal for M = 1000
Noisy Signal
Fitted Signal (M=1000)


Cost Function J_min vs Order
Cost Function
Order (k)
0
200
400
600
800
1000
20
40
60
80
100


Noisy and Reconstructed Signals
Noisy Signal
Reconstructed Signal
Amplitude
Sample Index
1.0
0.5
0.0
−0.5
−1.0
0
250
500
750
1000
1250
1500
1750
2000

Discussion

Impact of Model Complexity:

As the number of basis terms increases, the fit improves, but there is a point where additional terms lead to overfitting. The AIC effectively helps in identifying the point of optimal complexity by penalizing excessive terms.

The recursive estimation method provides a flexible way of progressively improving the model. By starting with a small number of terms and gradually increasing them, it helps avoid overfitting and allows for better control of the model complexity.

Role of Regularization:

Regularization helps prevent overfitting by penalizing overly complex models. By adding a small regularization term to the least squares solution, we ensure that the estimated coefficients don't grow too large, which could lead to unstable or overfit models.

Bias-Variance Trade-off:

The analysis shows the importance of the bias-variance trade-off in model selection. With too few basis terms, the model is too simple and underfits the data. With too many terms, the model overfits the data, leading to poor generalization to unseen data. The AIC provides a way to balance these competing factors.

Conclusion

Signal estimation and reconstruction using cosine basis functions and regularized least squares methods. The use of Akaike Information Criterion (AIC) helps in selecting the optimal model complexity by balancing model fit and complexity. The results show that increasing the number of basis functions improves the fit but can lead to overfitting if not controlled. The recursive estimation approach provides a systematic way of increasing model complexity, and the AIC ensures that the model remains as simple as possible while still fitting the data effectively.

This method is applicable in various fields, such as signal processing, time series analysis, and machine learning, where the goal is to estimate underlying patterns from noisy data. By understanding and applying these techniques, we can develop more robust models that generalize well to unseen data.