ME644-Assignment_5

Introduction

System identification is a method used in engineering and science to build mathematical models of dynamic systems from measured data. This document outlines the steps taken to discover the governing equation of a simple pendulum from experimental data provided in a CSV file. The data consists of angular displacement (θ), angular velocity ($\dot{\theta}$), and angular acceleration ($\ddot{\theta}$).

Methodology

1. Data Preparation

The first step involves loading the experimental data from a CSV file into a pandas Data Frame. The data contains three columns:

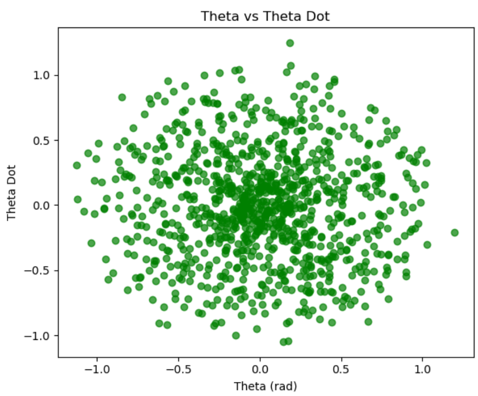
- **Theta** (θ): Angular displacement in radians.
- **Theta dot** ($\dot{\theta}$): Angular velocity in radians per second.
- **Theta double dot** ($\ddot{\theta}$): Angular acceleration in radians per second squared.

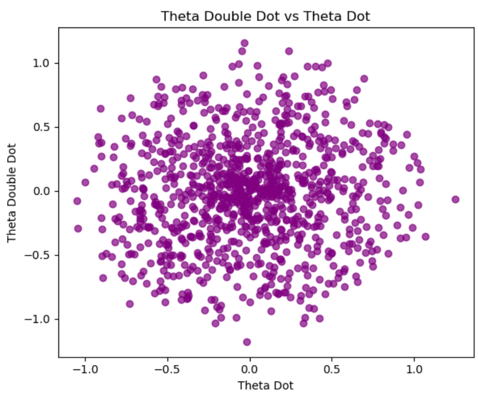
After loading the data, we convert the columns to NumPy arrays for easier manipulation.

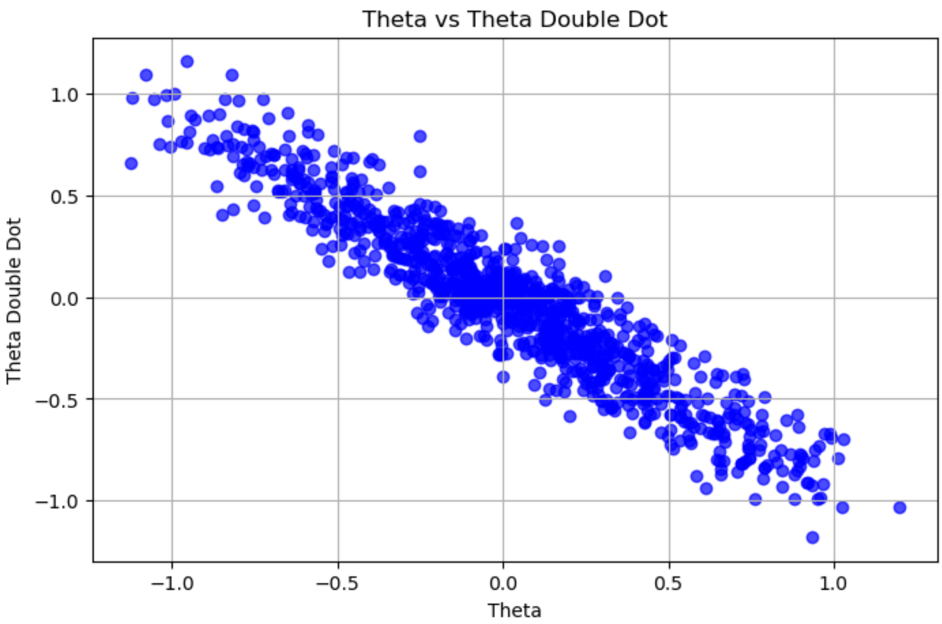
b. Scatter Plots

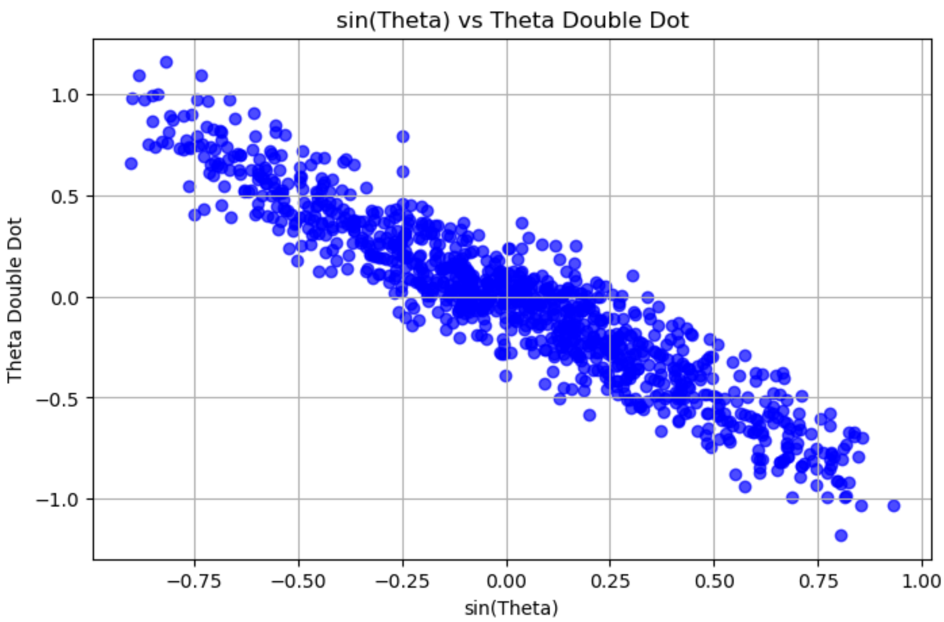
We create scatter plots to visually inspect the relationships between the variables:

- **Theta vs. Theta dot**: This plot shows how angular displacement relates to angular velocity.
- **Theta dot vs. Theta double dot**: This plot illustrates the relationship between angular velocity and angular acceleration.
- **Sin (Theta) vs. Theta double dot**: This plot is useful for observing how the sine of angular displacement correlates with angular acceleration.
- **Theta vs. Theta double dot:** This plot is useful for observing how the angular displacement correlates with angular acceleration.









Since there is no strong correlation in angular acceleration and angular velocity and correlation coefficient is close to 0 so I used a linear ridge model with angular acceleration as Y and parameters are angular displacement and angular velocity.

3. Model Development

a. Feature Matrix

We create a feature matrix X containing the predictors θ and sin(θ), and we set y as the target variable $\ddot{\theta}$. This allows us to build a model that predicts angular acceleration based on angular displacement and its sine.

b. Ridge Regression

To discover the governing equation of the pendulum, we use Ridge Regression, a regularized linear regression method that can prevent overfitting. We implement the Ridge Regression model from scratch, incorporating methods for fitting the model and making predictions.

- **Fitting the Model**: The model uses a closed-form solution to calculate the coefficients. A bias term (intercept) is added to the feature matrix during fitting.
- **Predictions**: After fitting, the model can predict angular acceleration based on new input data.
- **Cross-Validation**: We implement k-fold cross-validation to evaluate the model's performance and ensure that it generalizes well to unseen data. The Mean Squared Error (MSE) is used as the performance metric.

4. Results

After fitting the model, we print the best alpha value (the regularization strength) and the coefficients obtained from the Ridge Regression. We also calculate and display the mean cross-validation score to assess model performance.

5. Conclusion

This document outlines the process of system identification for a simple pendulum using experimental data. By analyzing the relationships between the variables, preparing the data for regression, and implementing Ridge Regression from scratch, we can discover the governing equations that describe the dynamics of the pendulum. The combination of data analysis, model fitting, and cross-validation provides a robust approach for system identification in engineering applications.