EE2703 : Applied Programming Lab
Assignment 3
Fitting Data to Models

Krishna Somasundaram
EE19B147

April 17, 2021

Aim

This assignment aims to

- Take data from a noisy environment and process it
- Study how to fit the data into a specified model
- Study how noise affects the fitting
- Plotting graphs to enhance our understanding
- Learning latex

Introduction

First, we run the generate_data.py file to generate the fitting.dat file. This fitting.dat file contains 10 columns, the first of which represents time while the next 9 contain the function:

$$f(t) = 1.05 J_2(t) - 0.105t + n(t) \tag{1}$$

With n(t) being the normal distribution of noise.

$$\mathrm{P}(n(t)|\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{n(t)^2}{2\sigma^2}\right)$$

where σ is generated using python function "logspace()"

This data is to be fitted into the function,

$$g(t; A, B) = A * J_2(t) + B * t \tag{2}$$

with true values of A = 1.05, B = -0.105

In this problem, the values of t are known. So we generate matrices M which contains the values of the Bessel function $J_2(t)$ and p, which contains the values of coefficients A and B to generate the column vector representing the function $g(t; A, B)$ by taking the product of the matrices.

$$g(t; A, B) = M \cdot p \tag{3}$$

Next the mean squared error of the data is taken with $A = 0, 0.1, 0.2, .., 2$ and $B = -0.2, -0.19, -0.18, ..0$ using the formula

$$\epsilon_{ij} = \frac{1}{101}\sum_{k=0}^{101}(f_k - g(t;A,B))^2 \tag{4}$$

A contour plot of the mean squared error with the values of A and B gives an estimate on the values of A and B where the error approaches 0

An estimate of the value of A and B to fit the given data is found using the method of least squares. This is done using the python command

This gives an estimate for A and B which minimizes the mean squared error.

0.1 Results:

0.1.1 Extracting and Visualising Data:

1. The data was obtained by running the given “generate_data.py”. A file named “fitting.dat” is produced as the output.

2. The fitting.dat file contains 10 columns of data: first column corresponding to the time stamps and the other 9 columns correspond to the function along with the noise.

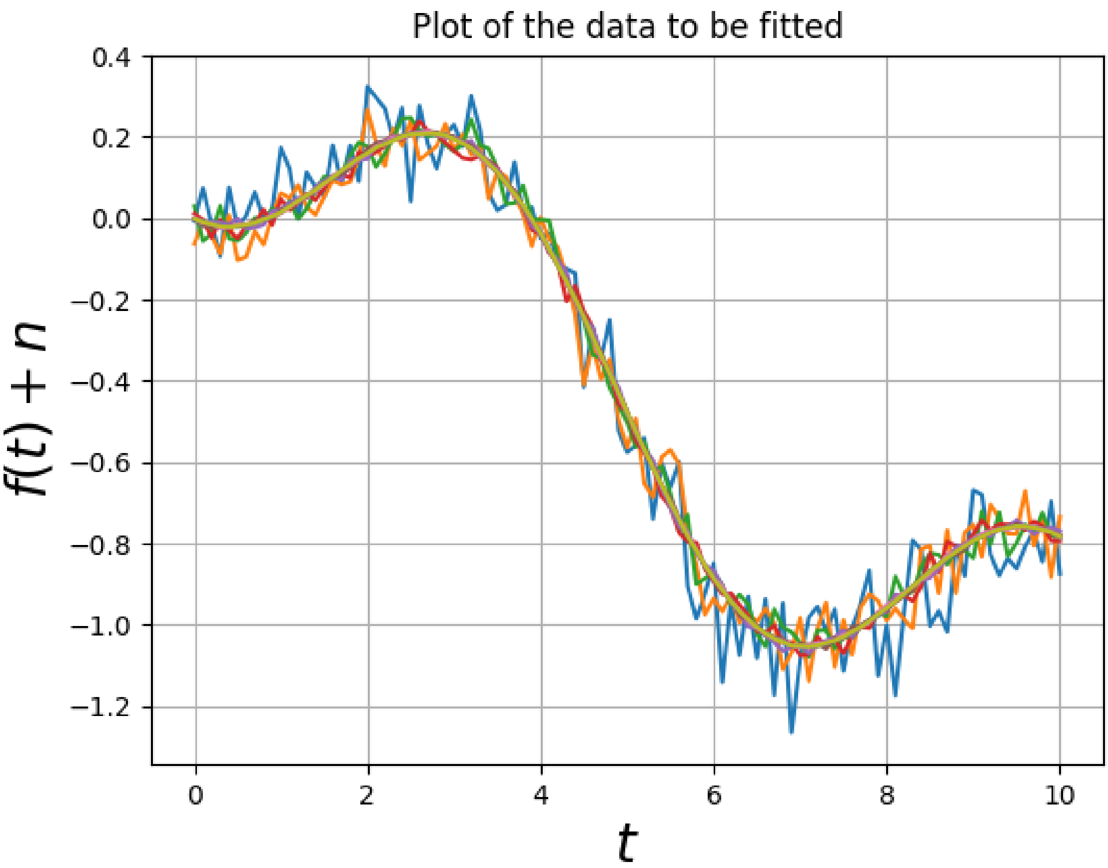

Figure 1: Given data

0.1.2 Obtaining the Function:

3. The python function to compute $g(t; A, B) = AJ_2(t) + Bt$ is as follows:

4. On fitting $g(t; A, B) = AJ_2(t) + Bt$ with $A = 1.05$ and $B = -0.105$, we obtain the following graph:

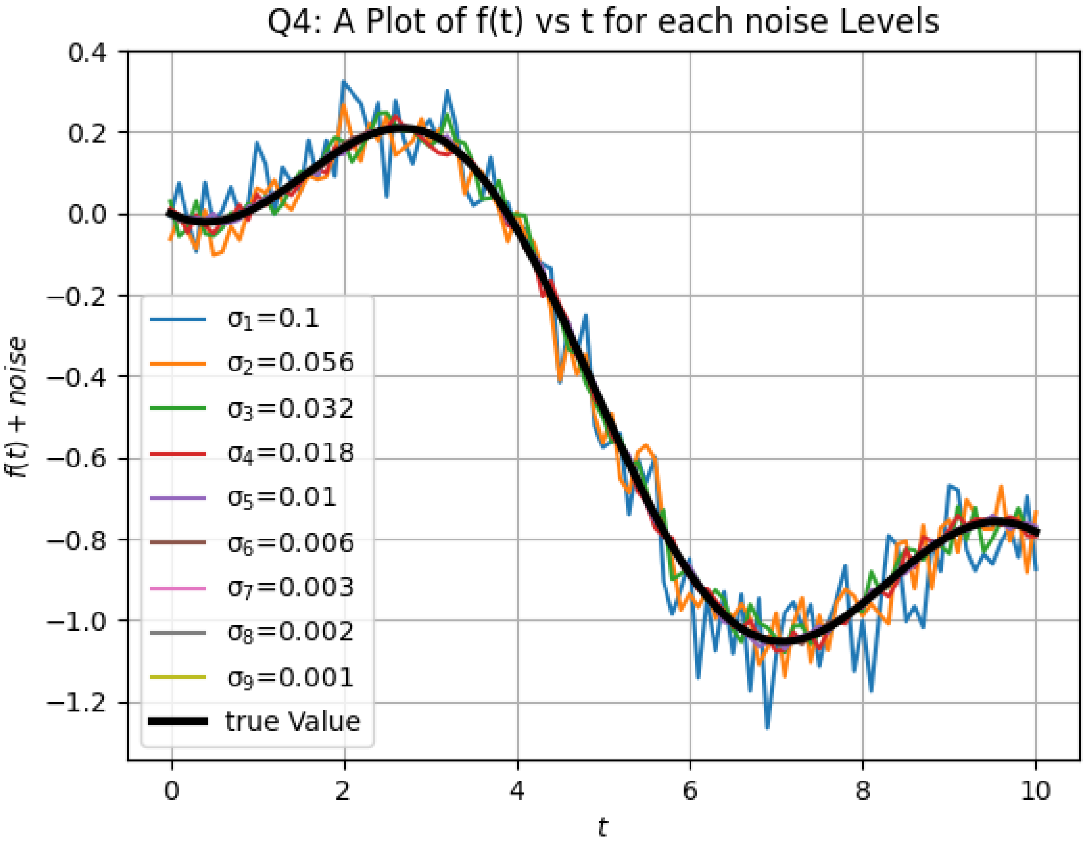

Figure 2: Obtaining the function

0.1.3 Errorbar Plot:

5. It is a convenient way to visualize the Measurements along with their errors. The errorbar plot of the first column is plotted below:

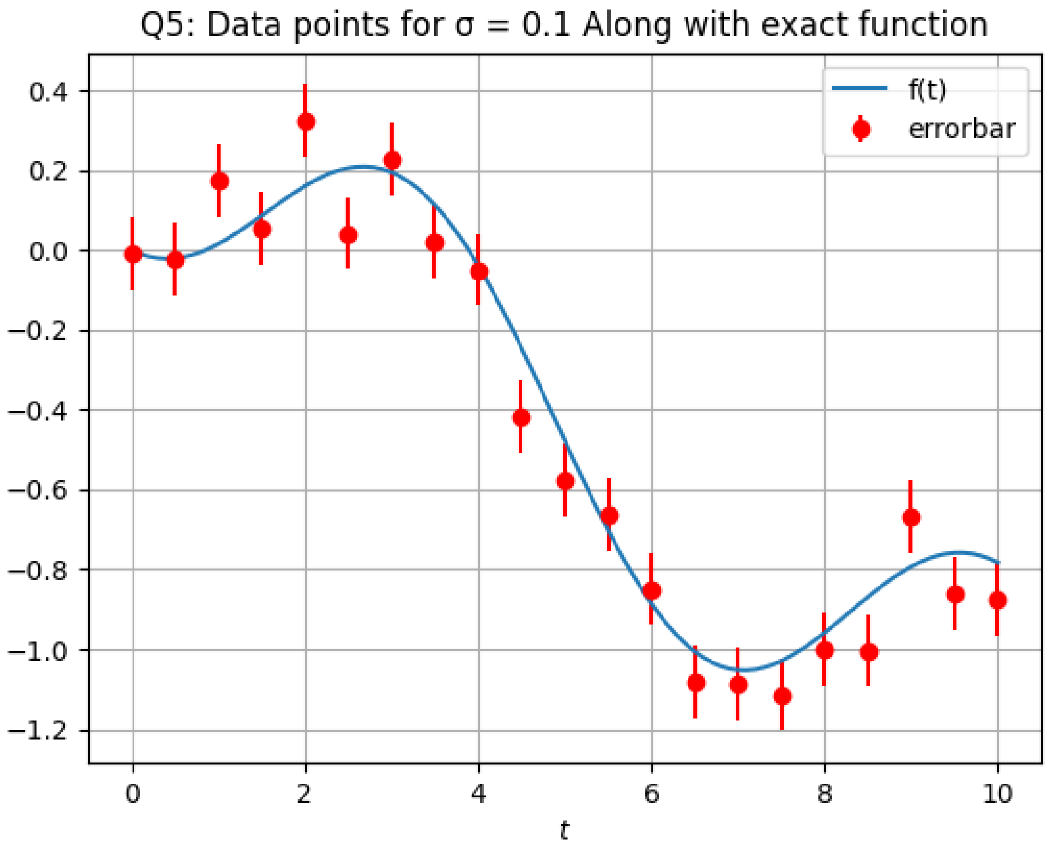

Figure 3: Errorbar plot

0.1.4 Predicting the Mean Square Error:

The Mean Squared Error can be computed using Python as follows:

$$\epsilon_{ij} = \frac{1}{101}\sum_{k=0}^{101}(f_k - g(t; A_i, B_j))^2 \tag{5}$$

0.1.5 Contour Plot:

6. Contour plot of the mean squared error with various values of A and B

   From the graph, the contours seem to have only one minimum and seems converge near the exact value at $A = 1.05$ and $B = -0.105$.

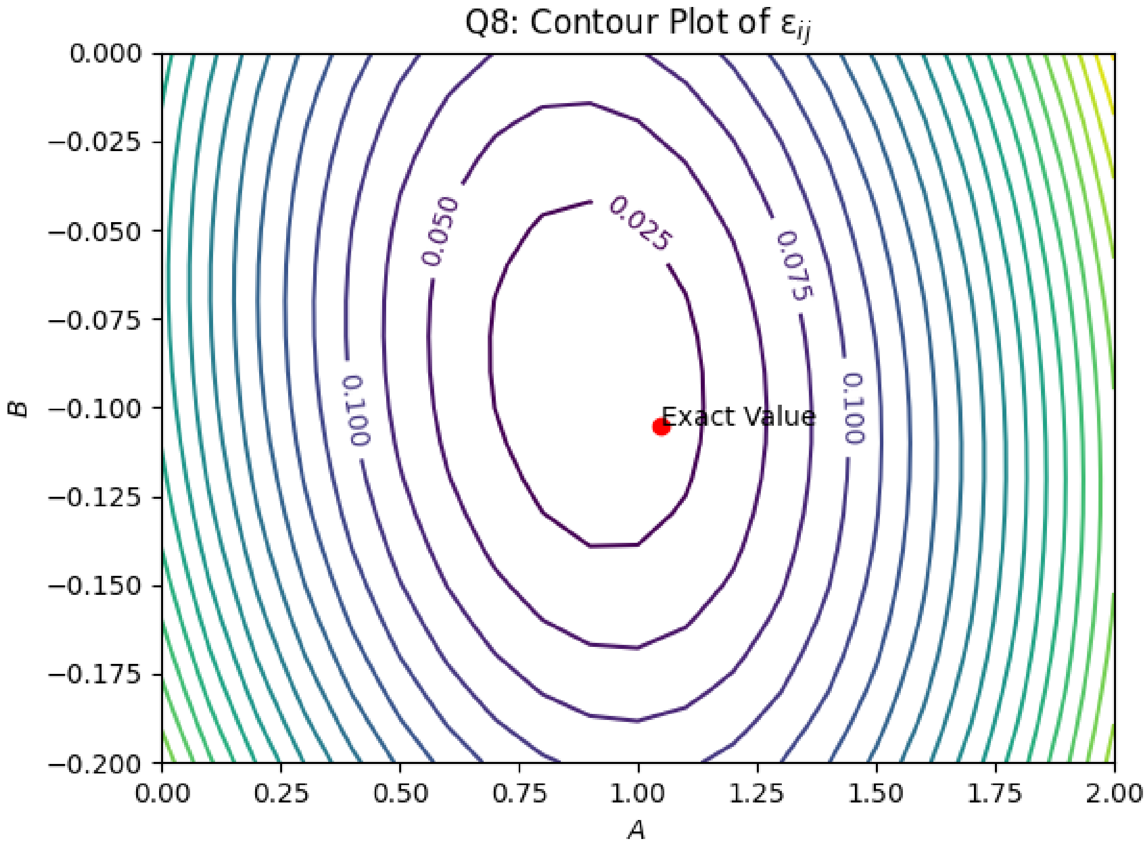

Figure 4: Contour plot

0.1.6 Parameter Estimation and Errors:

7. The plot of variation of error in approximation with the standard deviation of the noise in the data is as follows:

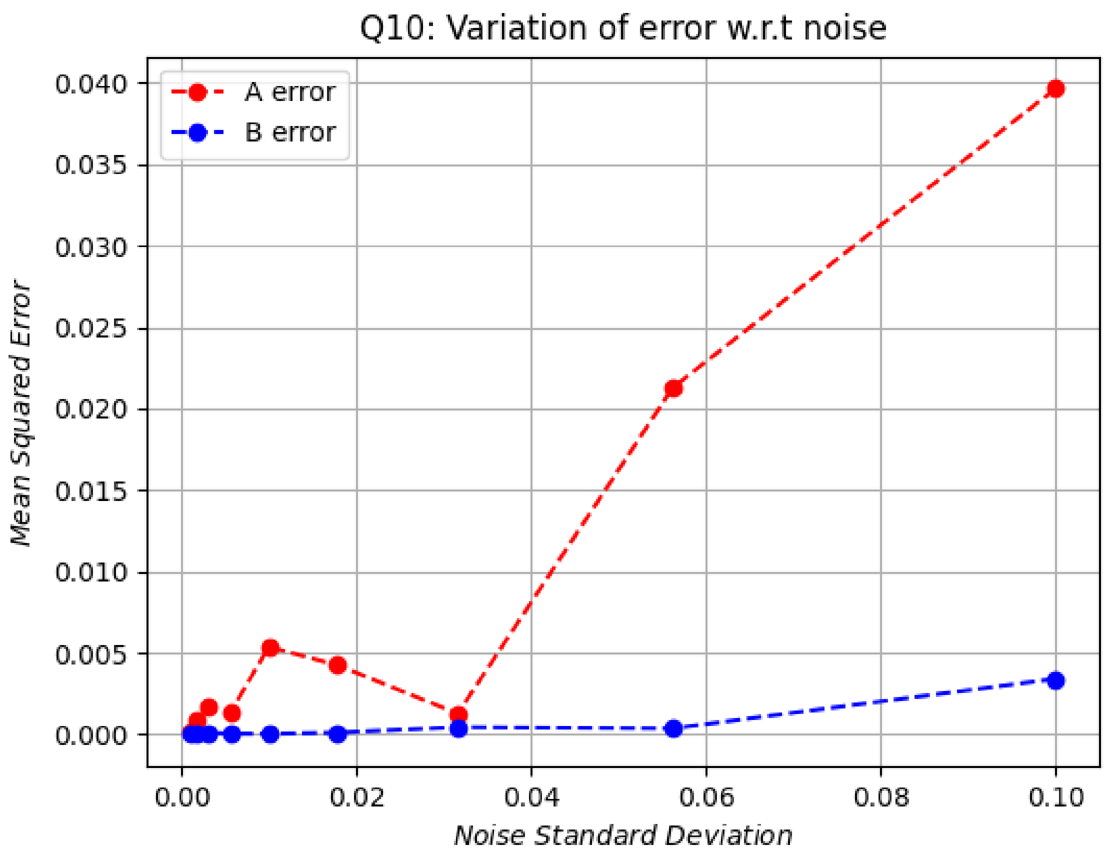

Figure 5: Variation of error w.r.t noise

The error estimate of A and B is non-linear with respect to the noise.

8. The plot of variation of error in approximation with the standard deviation of the noise in the data in logarithmic scale:

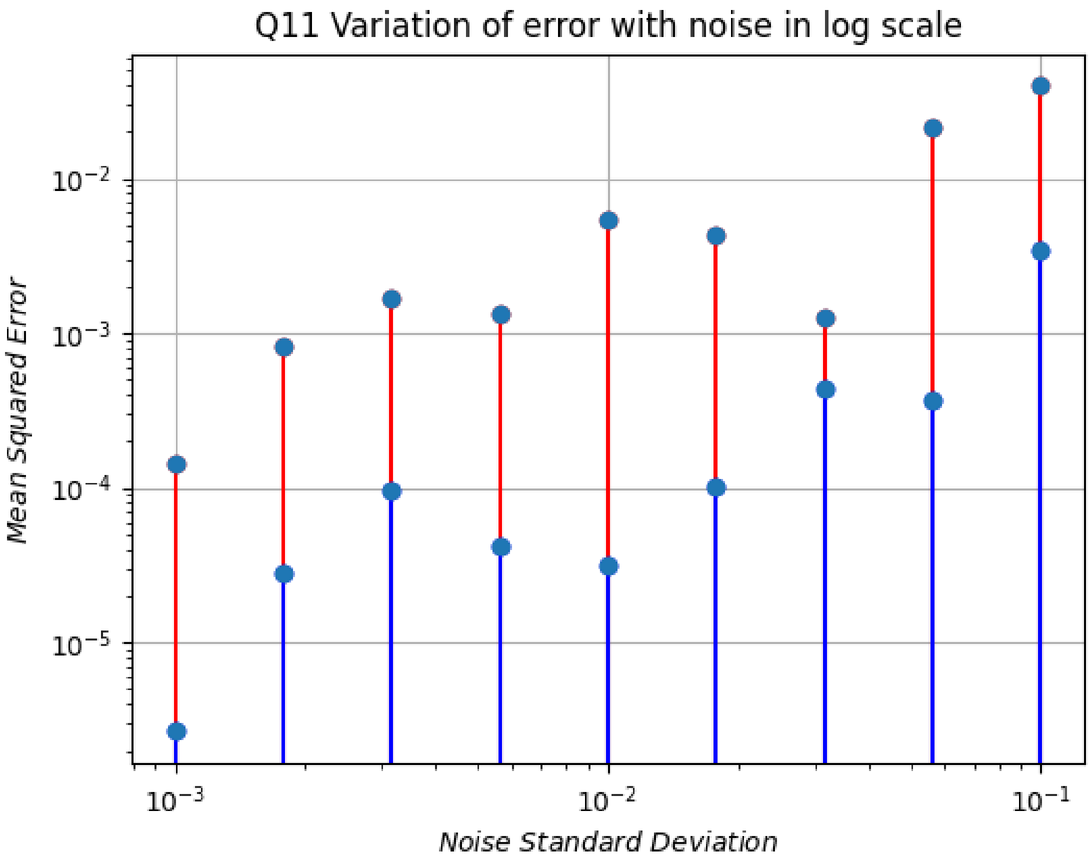

Figure 6: Variation of error w.r.t noise

On plotting both axes in the *log* scale, the graph becomes approximately linear, as the σ values are equally spaced in the *log* scale.

0.1.7 Conclusions

We can see that the mean squared error of the data converges close to the true value and minimizing it using the least squares method gives an estimation with less than 5% with the data of standard deviation close to 0.1.

It can also be seen that the error in A and B approximately varies almost linearly with the standard deviation of noise in the logarithmic scale.