EE2703: Applied Programming Lab
Week3: Fitting Data to Models

Niyas Mon P — EE20B094

February 18, 2022

1 Objectives

- Make the best model/estimate for the data given
- Observe on how error is varied with respect to the noise in the data.

2 Theory

The file "generate data.py" will generate a set of data with different levels of noise in it as follows:

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

where n(t) is the amounts of noise according to the following distribution:

$$P(n(t)|\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{n(t)^2}{2\sigma^2}} \tag{2}$$

Now we have to fit the datas in "fitting.dat" using the following model.Here, we have to find the values of A and B which would best fit using least square error

$$g(t; A, B) = AJ_2(t) + Bt \tag{3}$$

while the true values of A and B are:

$$A = 1.05, B = -0.105 \tag{4}$$

To solve this first we make a matrix M whose columns are Bassel function *J(t)* and time t and vector P for the values of A and B

$$g(t;A,B) = \begin{pmatrix} J_2(t_1) & t_1 \\ J_2(t_2) & t_2 \\ ... & ... \\ J_2(t_m) & t_m \end{pmatrix} \cdot \begin{pmatrix} A \\ B \end{pmatrix} = M.p \tag{5}$$

We are also doing contour plot for different values of A and B for the first data. For A *=0, 0.1, ..., 2* and B *=0.2,0.19,...0* we calculate:

$$\epsilon_{ij} = \frac{1}{101}\sum_{k=0}^{101}(f(t_k) - g(t_k, A_i, B_j))^2 \tag{6}$$

And finally we are finally plotting errors in values of A and B in normal and log-log scale

3 Plots

3.1 Extraction and Plotting of noisy data

The data in the fitting.dat file is plotted:

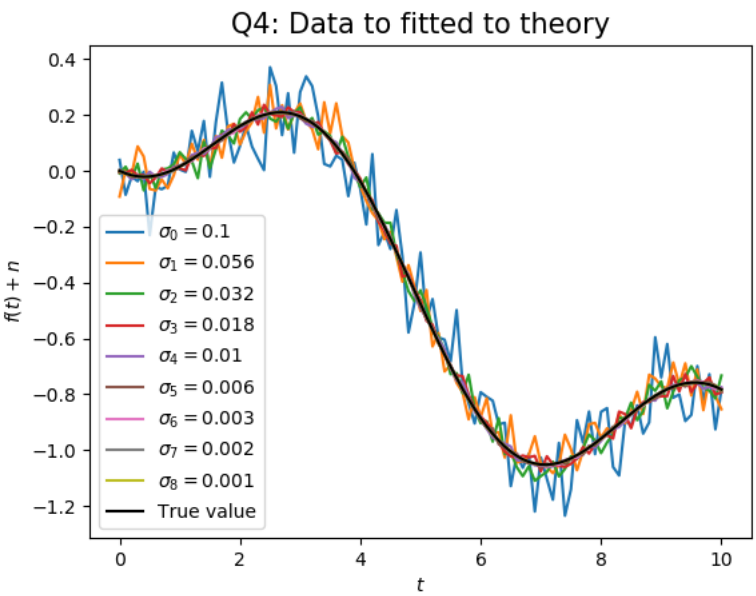

Figure 1: Noisy Data and True Data

3.2 Plotting Errorbars

Now we are plotting first data in "fitting.dat" file with error bars with every fifth data in it along with true values to compare

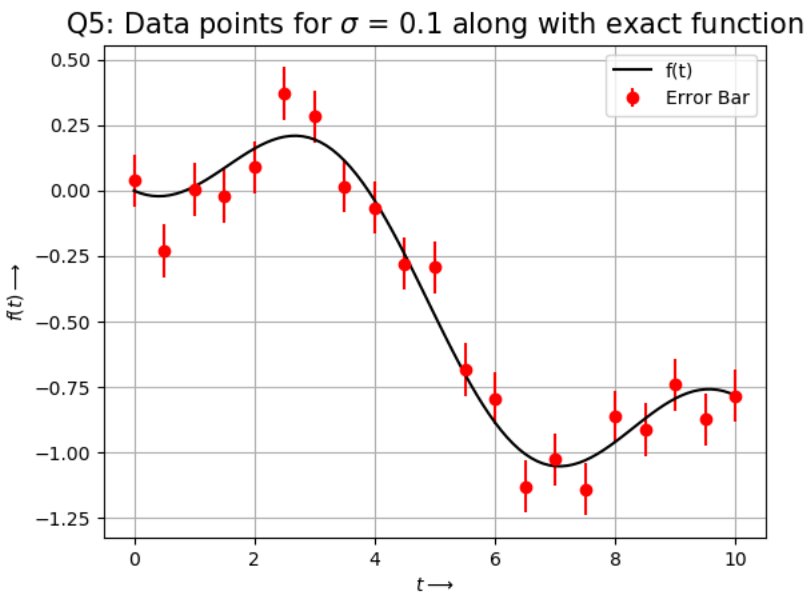

Figure 2: Noisy Data with Errorbar

3.3 Finding the best fit for the data

From the data, we can say that the data can be fitted into a function of the form:

$$g(t, A, B) = AJ_2(t) + Bt \tag{7}$$

where the coefficients A and B are to be found.

To find the coefficients A and B, we are finding the mean square error between the function and the data for a range of values of A and B,given by:

$$\epsilon_{ij} = \frac{1}{101}\sum_{k=0}^{101}(f(t_k) - g(t_k, A_i, B_j))^2 \tag{8}$$

and plotting the contour line and finding the minimum for the best fit as follows:

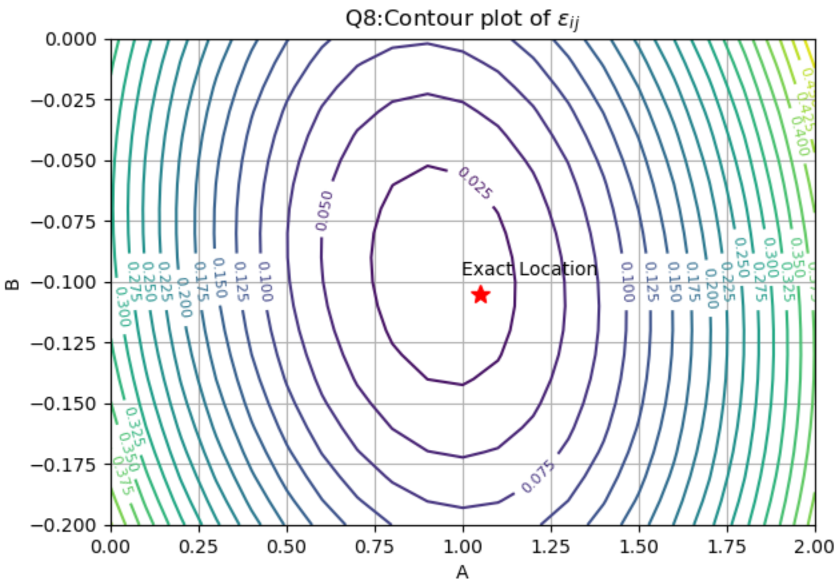

Figure 3: Contour Plot of ϵ_{ij}

The minima occur at **A = 1.05 and B = -0.105**

3.4 Error plots: Variation of error with σ

The plot of error in our approximation of A and B using the `lstsq` function with the standard deviation of the noise in the data

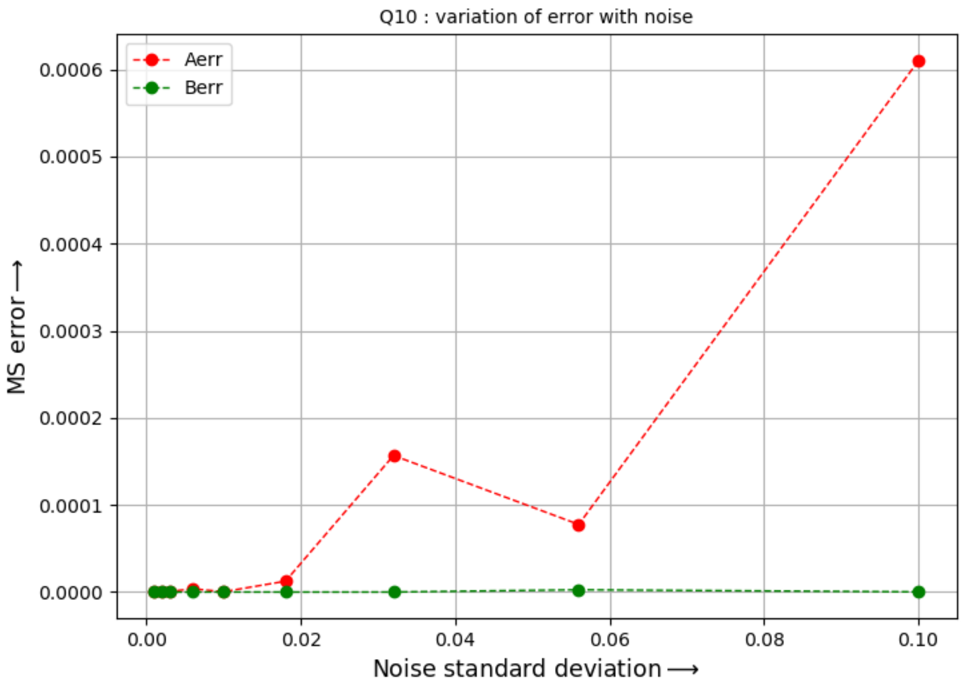

Figure 4: Mean Squared Error vs Standard Deviation

Now we are plotting same info in log-log scale

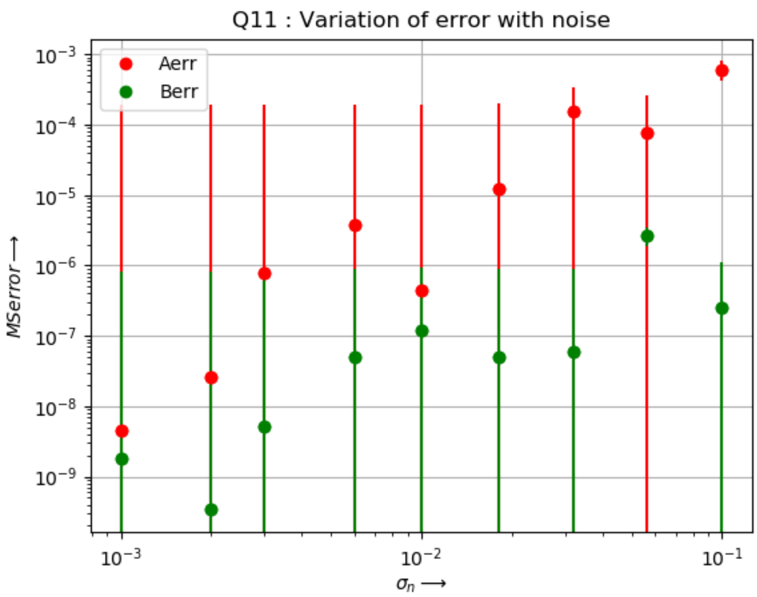

Figure 5: MS Error vs Standard Deviation (`loglog`) Plot

We can see an approximately linear relation between σ_n and ϵ.

4 Conclusions

- In the contour plot, we can see that the MS error of the data converges to the true values of A and B and minimizing it using the least squares method we obtain a good estimation
- The value of B parameter changes very slowly compared to value of A in the normal scale but both A and b changes almost linearly with standard deviation in the logarithmic scale.