EE2703 : Applied Programming Lab
Assignment 3
Fitting Data to Models

Arun Krishna A M S
EE19B001

March 3, 2021

Aim

This assignment mainly focuses on

- Reading ”fitting.dat” file and parse data from it
- Analyze the generated data to extract information
- Linear fitting to the data and the effects of noise on it
- Plotting graphs for enhanced understanding

Introduction

The assignment starts with generating data as a linear combination of the **Bessel function.** and the input variable by running the `generate_data.py` file. To simulate real-life signals, a normally distributed noise with standard deviation sampled uniformly on a logarithmic scale has also been added. The data is stored in the `fitting.dat` file.

```
data = np.loadtxt("fitting.dat",dtype=float);
time = np.array(data[:,0]);
data = np.array(data[:,1:]);
```

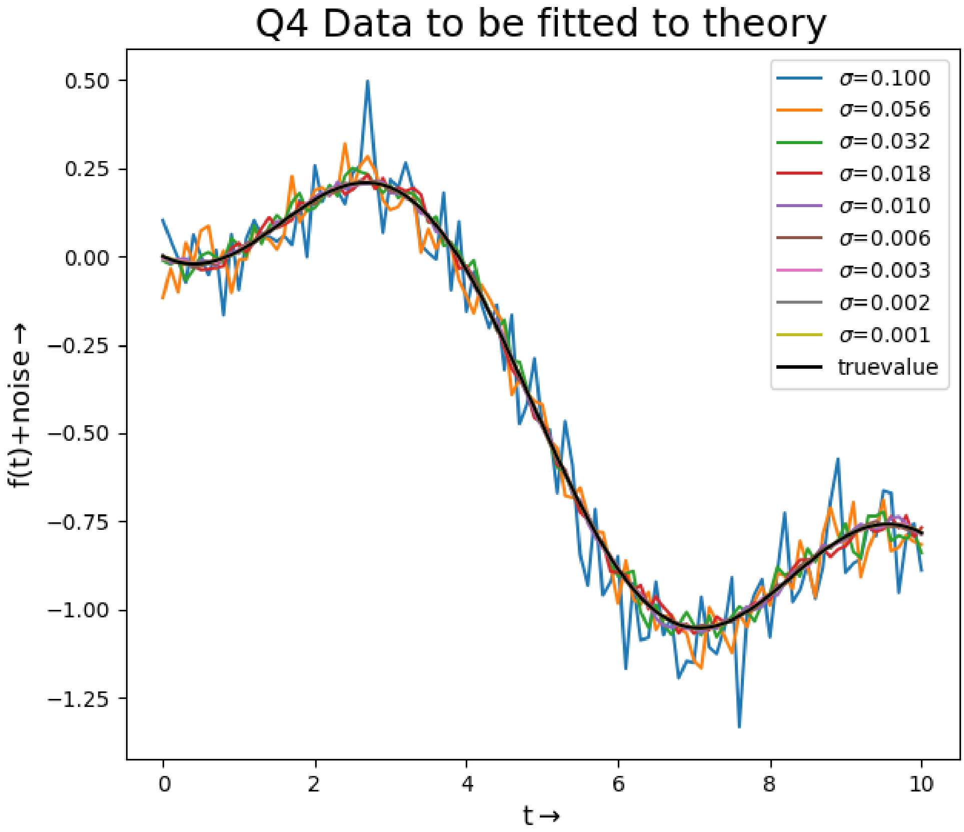
The `fitting.dat` file contains 10 columns. The first column is time, and the other 9 columns contain function:

$$f(t) = 1.05J_2(t) - 0.105t + n(t)$$

where $J_2(t)$ is the bessel function and
$n(t)$ is the normally distributed noise given by

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

where σ is generated using `sigma=logspace(-1,-3,9)`



Plot of all data columns

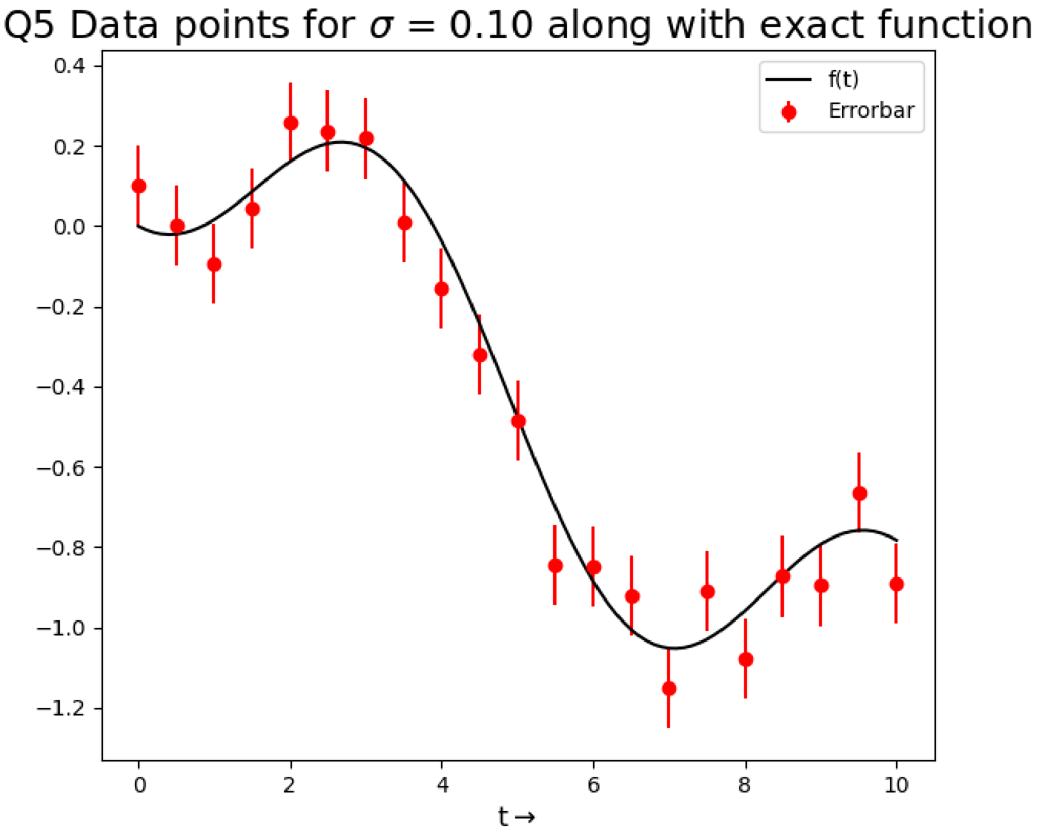
Errorbar Plot

Since the general shape of the given function is known beforehand, it is quite easy to fit the data to a model.

$$g(t; A, B) = AJ_2(t) + Bt$$

where A and B are the parameters that needs to be figured. In this assignment we take values of the parameters as $A = 1.05$ and $B = -0.105$.

Errorbar plot is an easy way of visualizing the noises present in the measurement. The first column of the data is plotted with errorbars. For readability, every 5^{th} item is plotted. The exact curve is also plotted to see how much the data diverges.



Errorbar Plot

Matrix Generation

Any experiment in reality, involve discreet variables. In this problem as well the variable `time` is discrete and known. Defining the matrix M as

$$M = \begin{pmatrix} J_2(t_1) & t_1 \\ \dots & \dots \\ J_2(t_m) & t_m \end{pmatrix}$$

and the vector p as

$$p = \begin{pmatrix} A \\ B \end{pmatrix}$$

The vector p contains the values of coefficients A and B and matrix M which contains the Bessel function values $J_2(t)$ and time t.

The function $g(t; A, B)$ can be written as

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

Mean Square Error Calculation

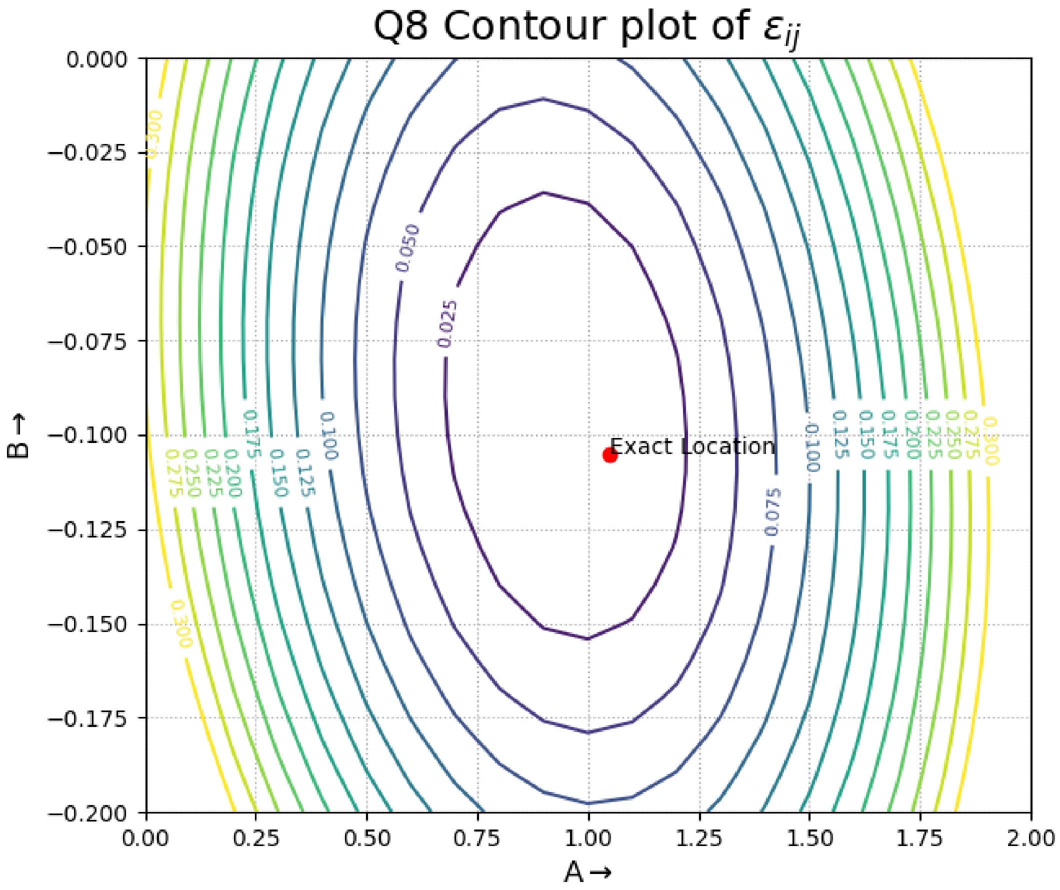
As we traverse through $A = 0, 0.1, ..., 2$ and $B = -0.2, -0.19, ...0$, the **mean squared error** is calculated between the data f_k and the assumed model using the formula

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

```
A = np.linspace(0,2,21);
B = np.linspace(-0.2,0,21);
epsilon = np.zeros((len(A),len(B)),float);

for Ai in range(len(A)):
    for Bi in range(len(B)):
        error = truevalue - g(time,A[Ai],B[Bi]);
        squareError = np.matmul(error,np.transpose(error));
        epsilon[Ai][Bi] = squareError/len(time);
```

The Mean squared error function values are plotted below

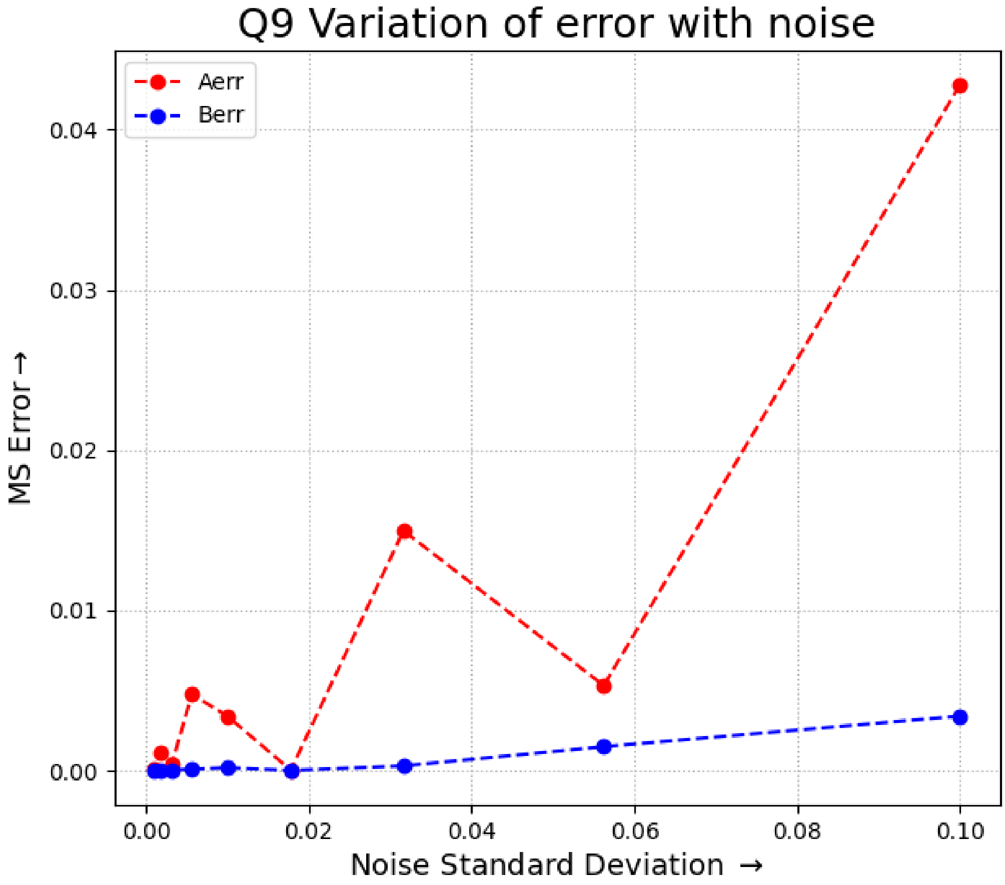


Mean Squared Error

The value of the paramters A_0 and B_0 corresponding to the least mean square error is the best estimate that can model the data is found using the python command:

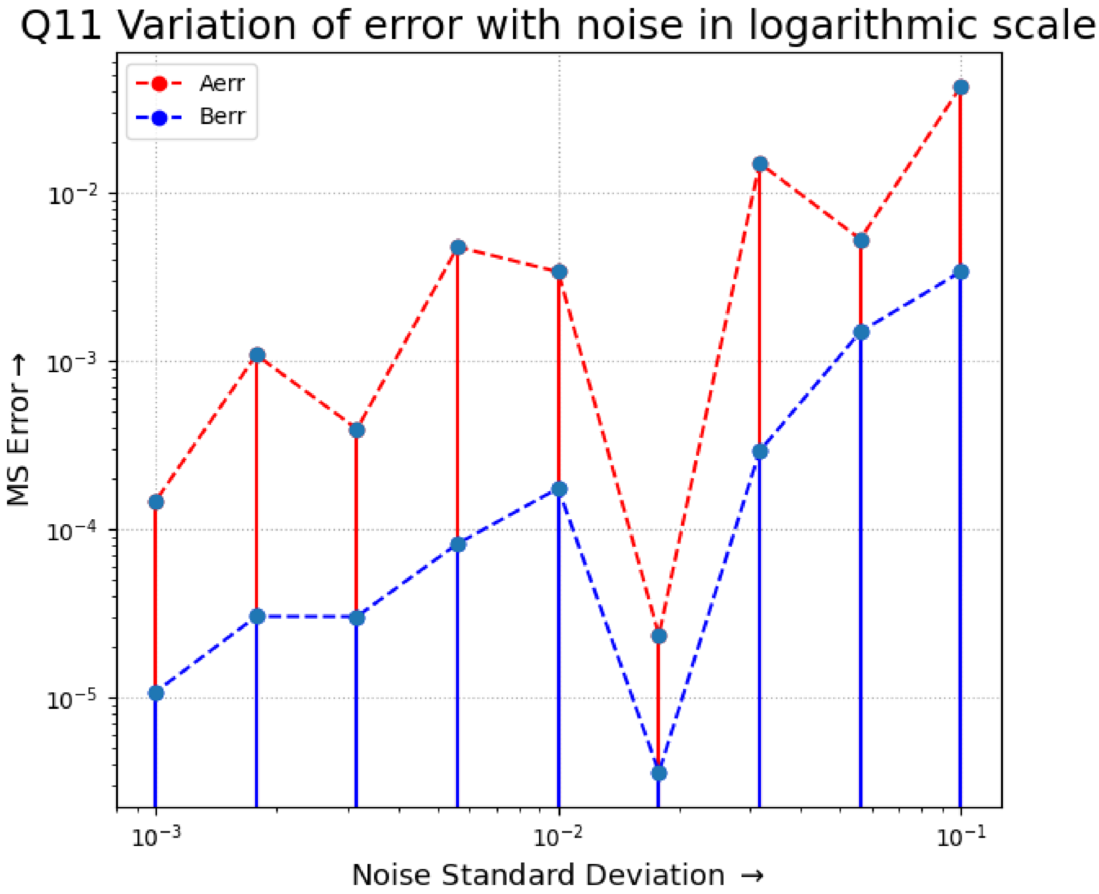
```
Est = [np.linalg.lstsq(M,data[:,i],rcond = None)[0] for i in range(9)];
Est = np.asarray(Est);
```

The `np.linalg.lstsq` function returns the least-squares solution to a linear matrix equation. Here $A_0 = 1.05$ and $B_0 = -0.105$. The error in the estimate of A and B for different values of σ is plotted. It can be observed that this error is in a non-linear fashion with respect to standard deviation of noise.



Error vs σ

On plotting on $log - log$ scale, it can be observed that the graph is approximately linear.



Error vs σ in logarithmic scale

Conclusion

A best possible fit is observed for the data obtained from the file with noises by minimizing the least mean square error. It is observed that the mean square error is directly proportional to the σ of the noise in the log scale