EE2703 : Applied Programming Lab
Assignment 3
Fitting data to Models

Chagari Koushal Kumar Reddy
EE20B023

February 18, 2022

Aim

The aim of this assignment is to :

- Record data from a noisy environment and process it
- Fit the data using a given model
- Observe how fitting model paramters vary with noise

Theory

Run the python file "generate_data.py" to generate the file "fitting.dat" which has 10 columns, the first one being time and the rest being different amounts of noise added to the function. The data columns correspond to the following equation:

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

Here $n(t)$ is the noise added which follows normal distribution given by:

$$Pr(n(t)|\sigma) = \frac{1}{\sigma\sqrt{2\pi}}\exp(-\frac{n(t)^2}{2\sigma^2})$$

where σ is generated using the python function "logspace"

```
sigma = logspace(-1,-3,9)
```

The model function which is used to fit the data is:

$$g(t;A,B) = AJ_2(t) + Bt \tag{2}$$

with true values of A, B being

$$A = 1.05, B = -0.105$$

To solve this problem, we first create the matrix M whose second column is the time vector and the corresponding values of the Bessel function $J_2(t)$ is the first column. Another column vector p is created where the first element is A and second element is B. We obtain $g(t;A,B)$ as a column vector by multiplying matrices M and p.

$$g(t;A,B) = \begin{pmatrix} J_2(t_1) & t_1 \\ \dots & \dots \\ J_2(t_m) & t_m \end{pmatrix} \cdot \begin{pmatrix} A \\ B \end{pmatrix} = M.p \tag{3}$$

Now the mean square error for the data is found for $A = 0, 0.1, \ldots, 2$ and $B = -0.2, -0.19, \ldots, 0$ using the formula

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

where f_k is the first column of data. A contour plot of the mean square error with the values of A and B gives an idea of how error varies with A, B. By analyzing the graph we can find for which values of A, B minimum error occurs.

An estimate of the values of A and B to fit the given noisy data is found using the method of least squares. The required python command is

```
scipy.linalg.lstsq(M,Data_column_to_be_fitted)
```

The above command gives an estimate for A and B which minimizes the mean square error.

Codes

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

```python
def g(t,A,B):
    j2_t = sp.jv(2,t)
    return A*j2_t + B*t
```

The piece of python code used to calculate the mean square error for $A = 0, 0.1, \ldots, 2$ and $B = -0.2, -0.19, \ldots, 0$ is as follows:

```python
a = np.linspace(0, 2, 21)
b = np.linspace(-0.2, 0, 21)
MSE = np.zeros((len(a), len(b)))

for i in range(len(a)):
    for j in range(len(b)):
        diff = np.zeros(0)       # Empty array
        for k in range(0, len(time)):
            g_value = np.array([g(t, a[i], b[j]) for t in time])
            diff = np.append(diff, (f_with_noise[0][k] - g_value[k])**2)
# f_with_noise[0][k] is the first column of data
        MSE[i][j] = diff.mean()
```

Results and Plots

1. The noisy datasets from ”fitting.dat” are plotted here. The first column of ”fitting.dat” corresponds to time and the remaining 9 columns correspond to the function with noise.

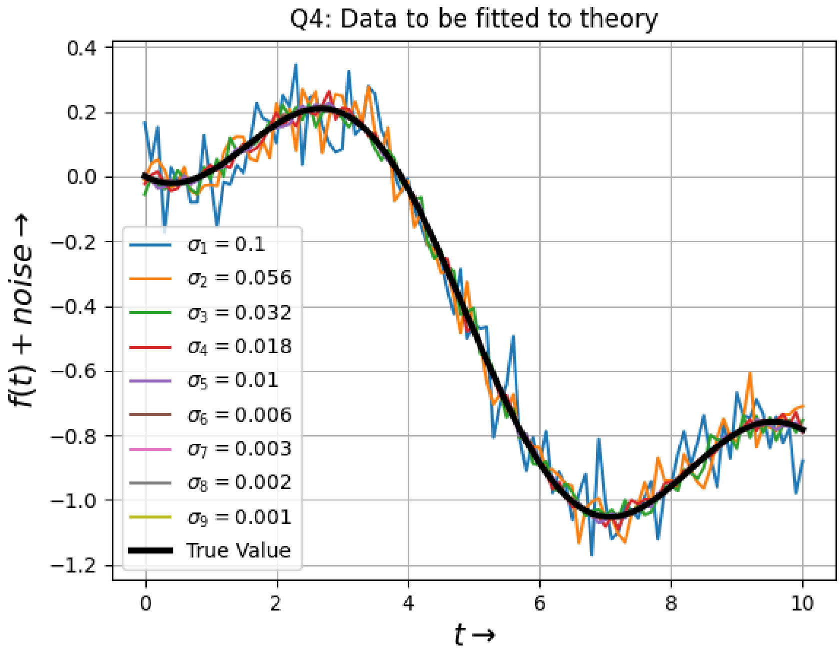

Figure 1: Data Plot

2. Error bars are used to show the deviation of noisy data(Here of first column) from the true value. The bright red dot in the center is the value of first column.

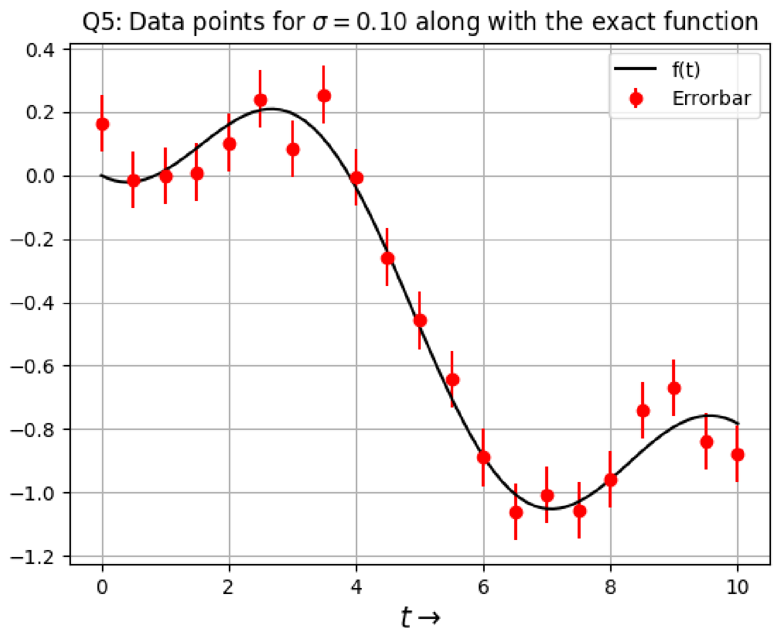

Figure 2: Error Bars

3. Contour plots of Mean squared error are plotted for various values of A and B.

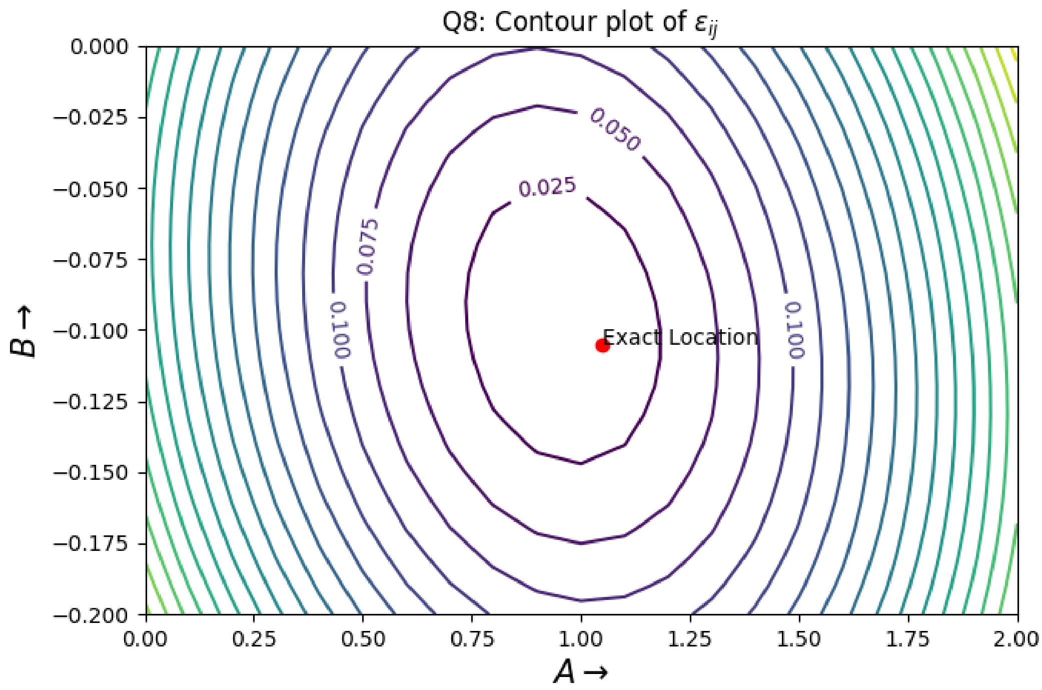

Figure 3: Contour Plot

4. This is the plot of deviation of parameters A, B from their true values with respect to the standard deviation of the noise present in the data (Linear scale).

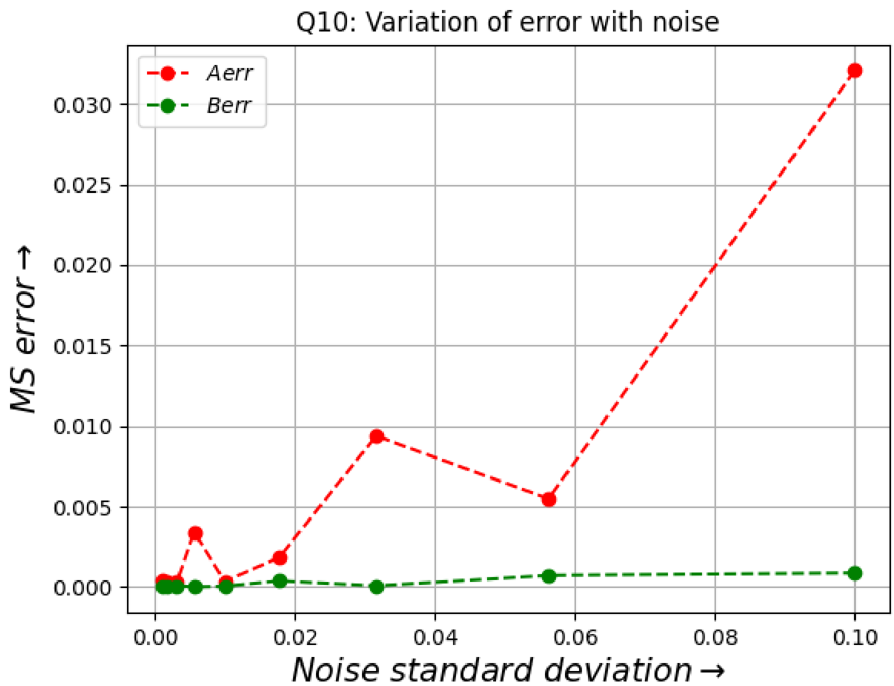

Figure 4: Linear scale Error plot

5. This is the plot of deviation of parameters A, B from their true values with respect to the standard deviation of the noise present in the data (Logarithmic scale).

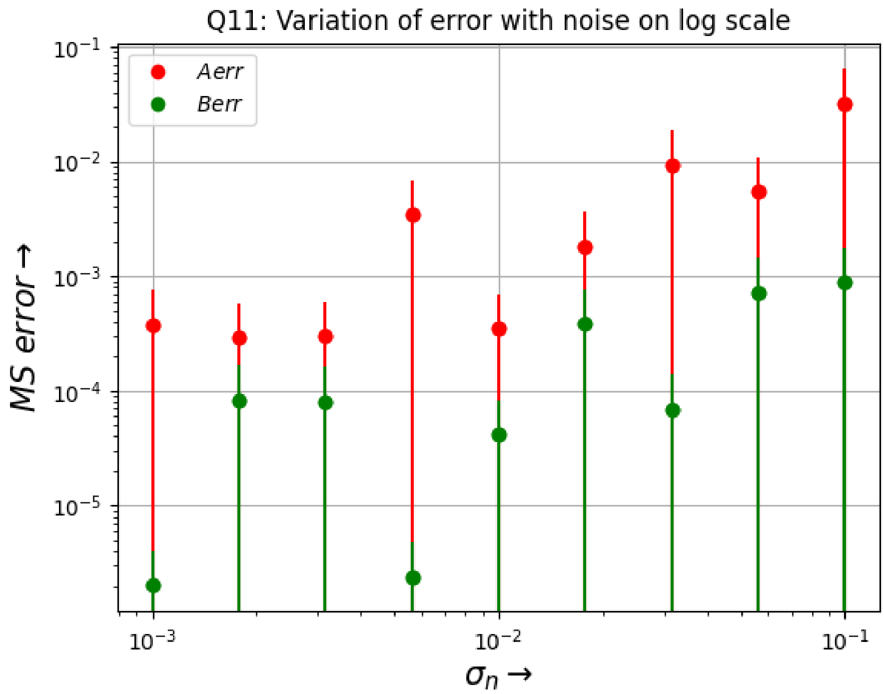

Figure 5: Logarithmic scale Error plot

The logarithmic errors of A, B show somewhat linear behaviour with few deviations.

Observations and Conclusions

- From the error bar plot we observe the estimated parameters of A, B provide a very close fit to the actual values of the function with noise of standard deviation 0.1.
- From the contour plot it is observed that the mean squared error converges to a minimum value as A, B approach their true values which are 1.05, -0.105 respectively. For the noisy data there is a single minima which is obtained using the least square solution.
- The errors in the estimate of A, B are not varying in a linear fashion with respect to time. Also the error in A is more suspectible to noise than B. This implies that the Bessel's function values accounts more to the noise than the time values.
- Logarithmic errors of A, B are linearly varying with respect to the logarithm of noise with slight deviations. This linear relation means that the errors in estimate of A, B vary exponentially with standard deviation of noise as seen in Figure 4.