EE2703 : Applied Programming Lab
Assignment 3
Fitting Data to Models

V. Ruban Vishnu Pandian
EE19B138

March 6, 2021

Aim

The aim of this assignment is to:

- Record data from a noisy environment
- Fit the data using a given model
- Observe how the fitting model parameters are affected by the noise

Procedure

Run the python file "generate_data.py" to generate a set of data following the equation:

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

With n(t) being the noise. The noise in each data set follows the normal disribution,

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

where σ is generated using python function "logspace"

```
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 and B being

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

The different values of t are known. Also, t is treated as a column vector with different values of time. With this understanding, matrices named M and p are created, where M is a 2-column matrix with the following columns:

First column: Contains the Bessel function $J_2(t)$ values for different values of t
Second column: Is the t vector itself

p matrix is a single 2 element column vector where the first element is A and second element is B

We have:

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

Now the mean squared error of the data is found for $A = 0, 0.1, ..., 2$ and $B = -0.2, -0.19, ...0$ 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 reaches its minimum

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_to_be_fitted)
```

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

Results and plots

1. The plots of the different noisy datasets to be fitted:

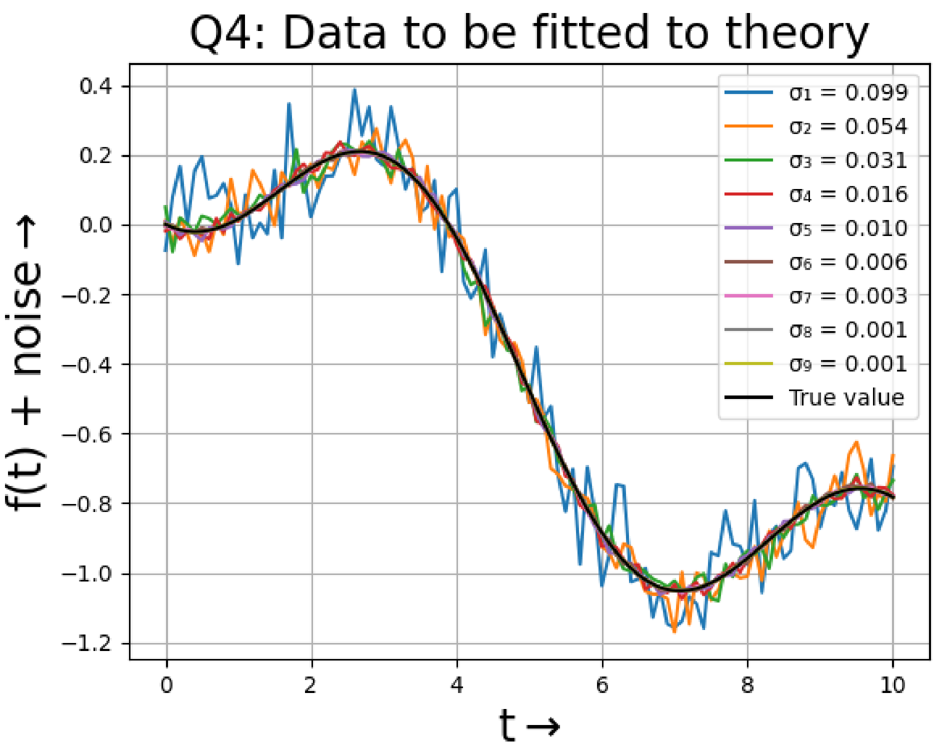

Figure 1: Data Plot

2. Errorbars are used to show the deviation of the noisy data (with standard deviation 0.1) from the true value:

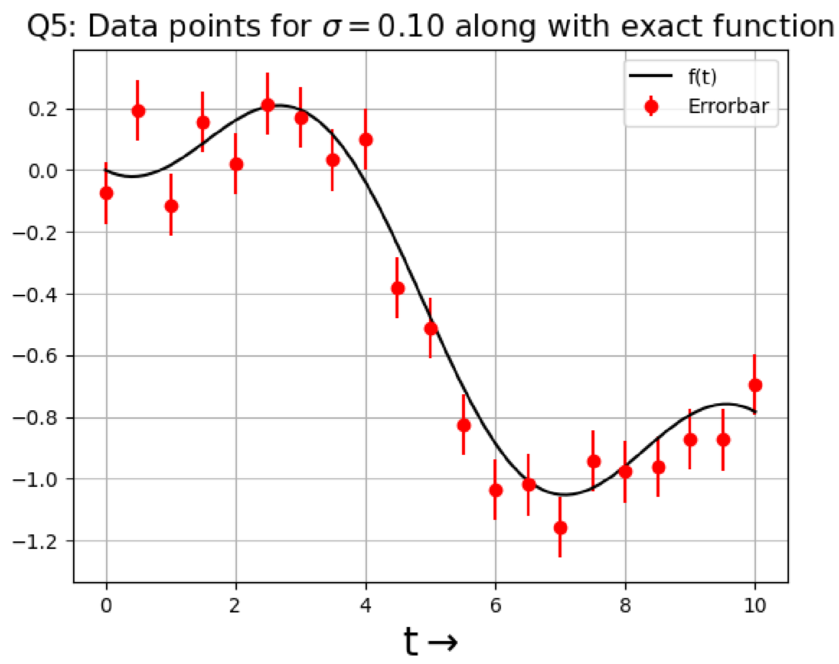

Figure 2: Error Bars

3. Contour plot of the mean squared error with various combinations of A and B:

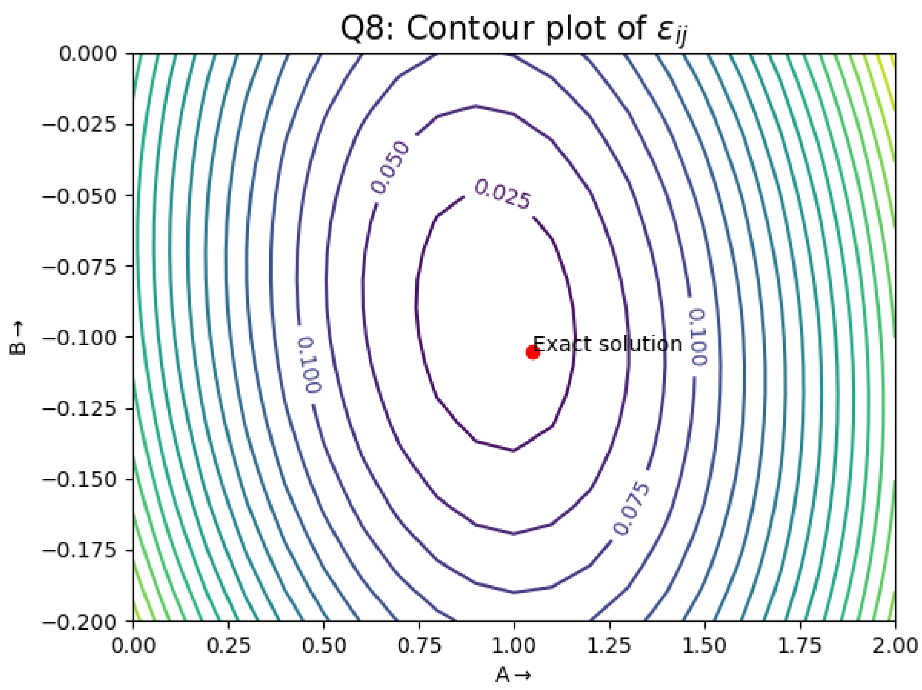

Figure 3: Contour Plot

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

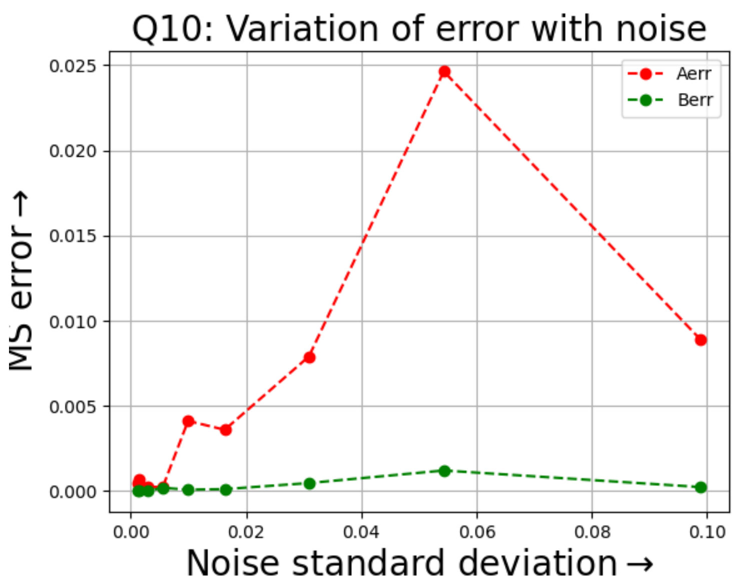

Figure 4: Linear scale error plot

The deviation of A is more susceptible to noise more than the deviation of B.

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

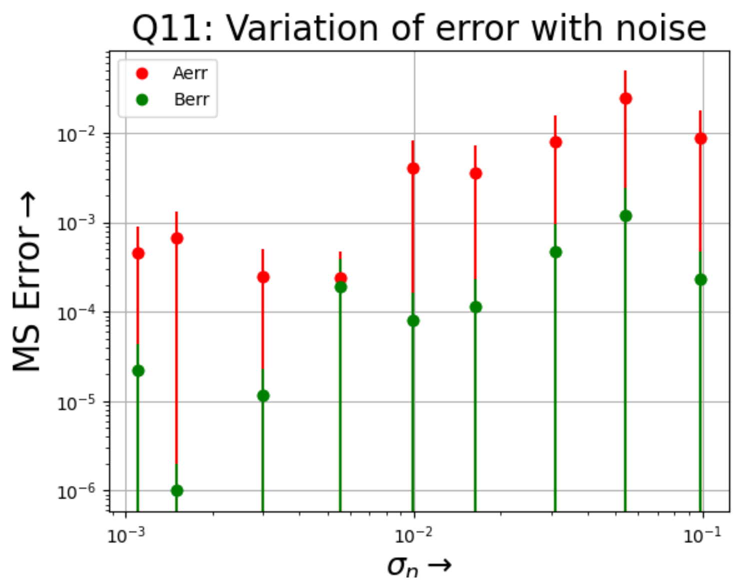

Figure 5: Logarithmic scale error plot

The logarithmic errors of A, B show somewhat linear behaviour with few deviations from the linear fashion

Observations and Conclusions

My conclusions are:

1. From the contour plot, it is observed that the mean squared error converges to a minimum value as A and B approach their true values which are 1.05 and -0.105. And also, there are no multiple minima. There's only a single minimum and that is the least square solution of A, B.

2. The errors in estimate of A, B are not varying in a linear fashion with respect to the noise. Also, error in A is more susceptible to noise, implying the Bessel's function values provide more contribution in compensating the noise than the time values do.

3. Logarithmic errors of A and B are linearly varying with respect to the logarithm of noise with some deviations. This linear variation means that the errors in estimate of A, B are exponents of standard deviation of the noise which is also evident from Figure 4.

   (Let ϵ_A = Error in A, σ = Standard deviation. Then we have: $\log(\epsilon_A) = k{\cdot}\log(\sigma){+}c$. Hence, $\epsilon_A = 10^c \cdot \sigma^k$, where k and c are coefficients in the linear relation. Similar relation follows for ϵ_B (Error in B))