Assignment 3 : Fitting Data To Models

Tadigoppala Yeswanth[EE19B122]

March 2021

1 Abstract

This Assignment mainly focuses on fitting data to models.
The main contents of are:

- Analysing the data to extract information
- Study the effect of noise in fitting process
- Plotting graphs

2 Loading *generate_data.py* and plotting it

First run the python code ***generate_data.py*** then it saves a file ***generatedata.py*** into the directory in which our ***fitting.dat*** is there. For loading this fitting.dat flie into the code we have to give the following command

fitdata = loadtxt("fitting.dat")

After this we will extract time from fitdata and we will use it to generate $g(t, 1.05, -0.015)$. Then we will plot the fitdata along different added noises with the following python code.

```
figure(0)
for i in range(1,10):
plot(t,fitdata[:,i],label="stdev= sigma[i-1])
plot(t,y,label="True Value",color='green',linewidth=3)
title("Data to be fitted to theory",size = 25)
xlabel('t',size = 20)
ylabel('f(t)+n',size = 20)
```

grid(True)
legend()

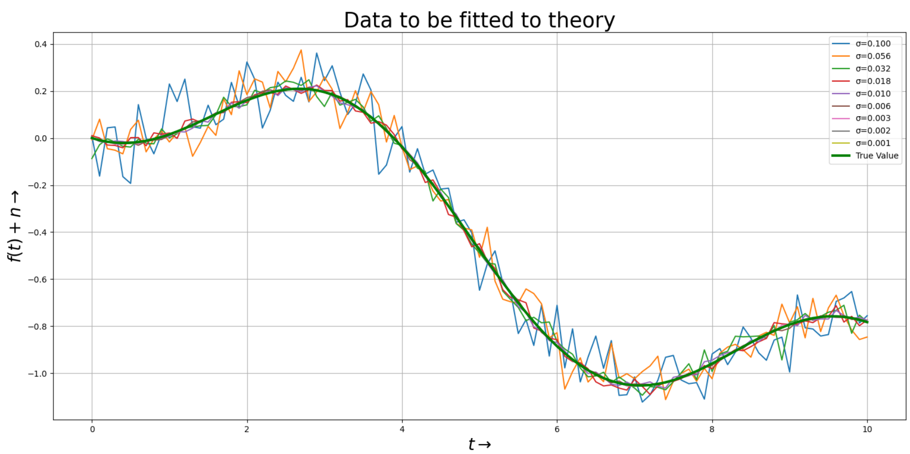

Figure 1: plot of True value along with added noise

This fitdata consists of 10 columns and 1st column corresponds to time and the remaining columns corresponds to data. Each column will have its own standard deviation and it is shown in the python code given below.
sigma = logspace(-1,-3,9)

3 Plotting Error Bar

Error bars are useful to problem solvers because error bars show the confidence or precision in a set of measurements or calculated values. In this we have to plot error bars of 1st column by taking every 5th element by using **errorbar()** function. Python code to plot is as shown below :

plot(t,y,label="True Value",color='blue',linewidth=3)
errorbar(t[::5],d[::5],sigma[0],fmt='ro',label="Noise")

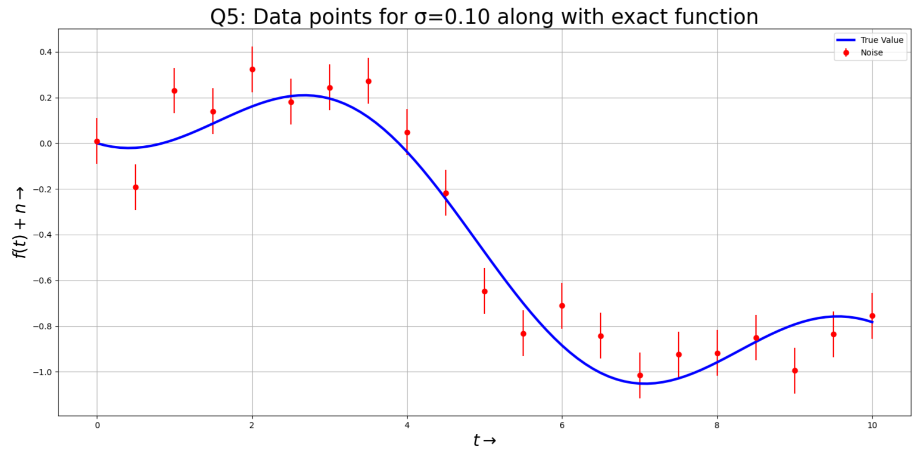

Figure 2: Errorbar plot of 1st column in data

4 Checking Matrix Equation

The matrix equation is

$$M.P = g(t, A, B)$$

.

we have to verify whether this equation holds or not. So first we will generate the M matrix. Matrix P will be equal to [A,B] where A=1.05 and B=-0.105. To compare two matrices we will use **array_equal()** function. The python code snippet for this is:

M = empty((a,2))
for i in range(a):
M[i] = (jv(2,t[i]),t[i])
p = array([1.05,-0.105])
out = dot(M,p)
if array_equal(out,g(t,1.05,-0.105)) : print("solution for Q4 and Q6 are equal")
else : print ("solution for Q4 and Q6 is not same")

5 Mean Square Error

The mean square error is the error between noisy data and True value. It is calculated as follows:

$$\varepsilon_{i,j} = (1/101) \sum_{k=0}^{k=101} (f_k - g(t_k, A_i, B_j))^2$$

This error is calculated for various values of A,B and a contour is plotted. The python code snippet to calculate mean square error is as follow:// **figure(2)**
X,Y = meshgrid(A,B)
Contour=contour(X,Y,E,20)
clabel(Contour,Contour.levels[:5],inline=1)
title("Q8:Contour Plot for Eij",size=20)
xlabel('A',size=20)
ylabel('B',size=20)
grid(True)
The contour plot for various values of A and B is :

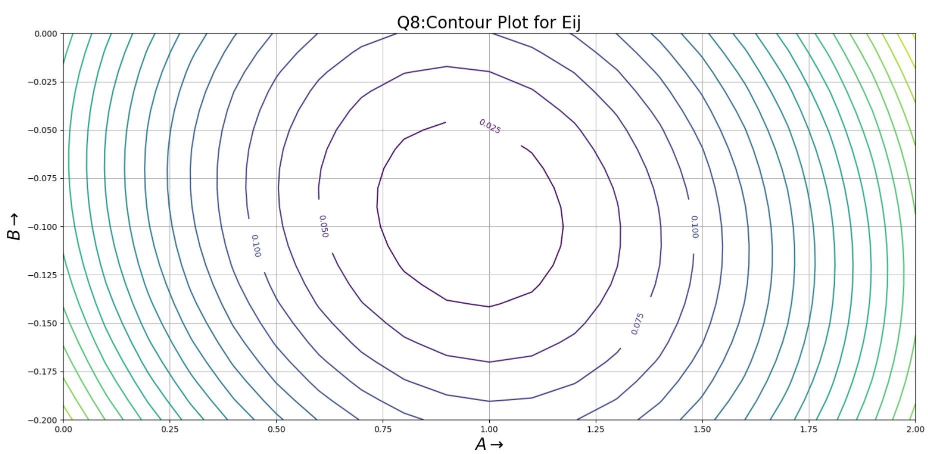

Figure 3: Contour plot of ε

5.1 Conclusion

From the above plot we can conclude that there is only one minimum value for ε

6 Error In Estimation of A and B

A and B are estimated using **lstsq()** function. A and B are estimated like this for all the 9 data columns and the estimated values are stored in arrays E_a and E_b respectively. Actual values of A(=1.05) and B(=-0.105) are previously stored in P matrix. So the error in estimation will be absolute value of difference between computed value and actual value. The python code snippet is as follows:

```
Ea = empty((9,1))
Eb = empty((9,1))
for j in range(9):
AB = linalg.lstsq(M,fitdata[:,j+1],rcond=None)
Ea[j] = abs(AB[0][0]-p[0])
Eb[j] = abs(AB[0][1]-p[1])
```

The plot of error in A,B versus standard deviation:

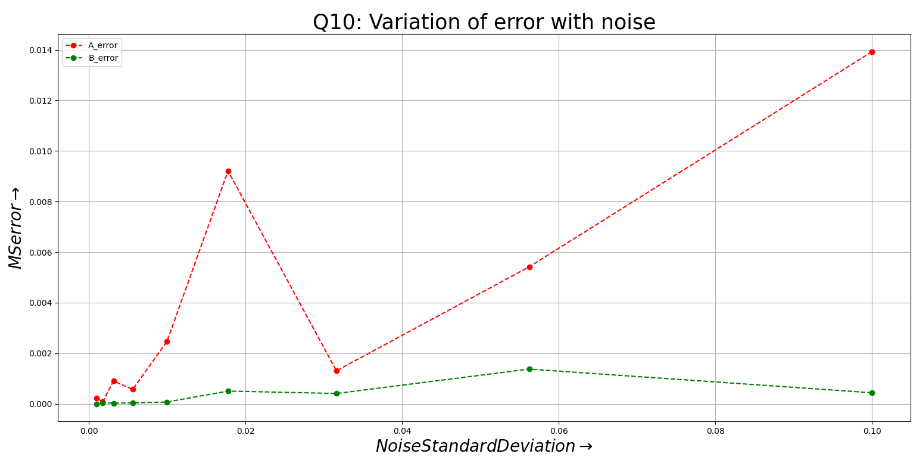

Figure 4: Error Vs Standard deviation

6.1 Conclusion

From the plot we can clearly see that the error is not varying linearly with noise

7 Error computation plot in loglog.

The python snippet code for plotting Error in loglog. is as follows:

```
figure(4)
loglog(sigma,Ea,'ro',label=" A_error")
errorbar(logspace(-1,-3,9),Ea,std(Ea),fmt='ro')
loglog(sigma,Eb,'go',label=" B_error")
errorbar(logspace(-1,-3,9),Eb,std(Eb),fmt='go')
title("Q11: Variation of error with noise",size=25)
xlabel('Noise standard deviation',size=20)
ylabel('MSerror',size=20)
grid(True)
legend()
```

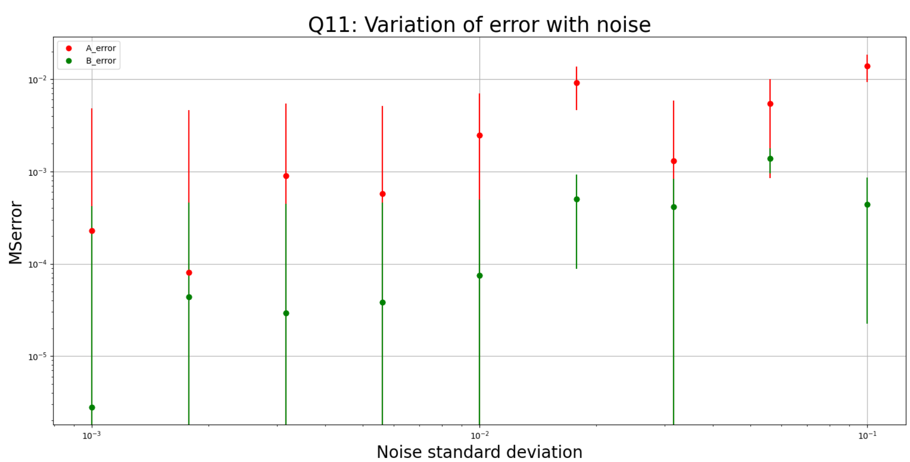

Figure 5: loglog. plot of Error Vs Standard deviation

7.1 Conclusion

From the plot we can clearly see that log(error) is not linearly varying with log(noise).

Inference

The given noisy data was extracted and the best possible estimate for the underlying model parameters were found by minimizing the mean squared error. This is one of the most general engineering use of a computer, modelling of real data.