Numerical Methods and MATLAB Refresher

For this lab we will be using MATLAB's ODE simulation and optimization tools to characterize bacterial growth data. For your submission, include all MATLAB code, figures generated, and answers to included questions.

Problem 1: Conversion between optical density and cell mass data.

The table below contains data of measured optical density (nm) and determined colony forming units, i.e. the number of cells/10^9.

OD600 (nm)	CFU (cells/10^9)
0	0
0.18	0.41
0.37	0.72
0.68	1.48
1.29	2.7

Use MATLAB's polyfit function to perform a linear regression on the dataset above to obtain a conversion factor to convert optical density to cell count/10^9. Plot your linear regression predictions and the experimental datapoints on the same axis.

Problem 2: Logistic growth and parameter estimation.

The table below contains optical density v. time data for a growing colony of *E. coli*.

Time (Hours)	OD600 (nm)
0	0.10
0.1	0.11
0.5	0.19
1	0.36
1.5	0.61
2	0.94
2.5	1.29
3	1.58
4	1.88
5	1.97
6	1.99

a) Use the linear regression parameters from Problem 1 to convert the above optical density measurements to cell counts with MATLAB's polyval function.
b) Using a model of logistic growth $\left(\frac{dN}{dt} = \gamma N\left(1 - {}^{N}/_{K}\right)\right)$ estimate the carrying capacity (K, $\frac{cells}{10^9}$) and growth rate ($\gamma, hours^{-1}$) of the growing colony by fitting the model to the time-course data above. Use the determined cell count value at time zero as your initial condition. To do this, use MATLAB's fminsearch function to minimize the sum or squared error between experimental measurements and model predictions ($error = \sum(prediction - measurement)^2$. Use initial parameter estimates of $K = 4\frac{cells}{10^9}, \gamma = 2\ hours^{-1}$.

c) Using the same model as in part a), re-estimate the model parameters by only fitting to the first five datapoints. Are your estimates of the parameters significantly different from part a)? What does this imply about the process of parameter estimation via fit to data?

d) The data used in parts a)-c) was easily fit because the values did not contain noise/uncertainty. Consider the new dataset below where some of the values from the previous datasets were taken and noise was added corresponding to +/- 25% of its magnitude. Re-fit the model to the dataset below with the same steps as in parts b) and c). How does your fit and estimated parameters compare to the previous steps? What does this imply about the role of noise in modeling biological systems?

Time (Hours)	Cell Count ($\frac{cells}{10^9}$)
0	0.22
0.1	0.26
0.5	0.36
1	0.77
4	4.72
5	4.92
6	3.39

e) Now repeat the same steps you took in part d), except rather than fitting data by minimizing sum of squared error, minimize the sum of squared error normalized by the magnitude of the datapoint ($error = \sum(prediction - measurement)^2/measurement^2$). Are the parameter values significantly different from those obtained in part d)?

f) Using the parameter values obtained in parts d) and e), predict the cell counts at each timepoint from the table above. Then generate the following two plots: 1) Plot $error = (prediction - measurement)^2$ on the y axis and time on the x axis. 2) Plot $error = (prediction - measurement)^2/measurement^2$ on the y axis and time on the x axis. Do you notice any significant differences for the predictions made with each parameter set? What do these differences mean about the use for each of these error functions?