Homework 9

November 13, 2006

Problem1

1. Let us first derive an analytical espression for $S_{model}(\lambda, t, \theta)$ using both the models mentioned in the problem statement.

   - *Model 1: Sequential Reaction Model*

     According this model C is formed from B through an intermediate species X, we can write the rate of formation of B, X and C as shown in Equation(1). In writing the rate rules in Equation(1) we have made use of the fact that the volume is constant.

$$\begin{aligned} \frac{d[B]}{dt} &= -k_1[B] \\ \frac{d[X]}{dt} &= k_1[B] - k_2[X] \\ \frac{d[C]}{dt} &= k_2[X] \end{aligned} \tag{1}$$

     We can solve the equations simultaneously to get the expressions for the concentrations of $[B]$, $[X]$ and $[C]$ shown in Equation(2).

$$\begin{aligned} [B] &= [B]_0 e^{-k_1 t} \\ [X] &= [B]_0 k_1 \left(\frac{e^{-k_1 t}}{k_2 - k_1} + \frac{e^{-k_2 t}}{k_1 - k_2} \right) \\ [C] &= [B]_0 \left(1 + \frac{k_2 e^{-k_1 t}}{k_1 - k_2} + \frac{k_1 e^{-k_2 t}}{k_2 - k_1} \right) \end{aligned} \tag{2}$$

     If we look at the absoption spectra of B and C in Figure 1 as given by SVD we realise that B has two distinct peaks and C has

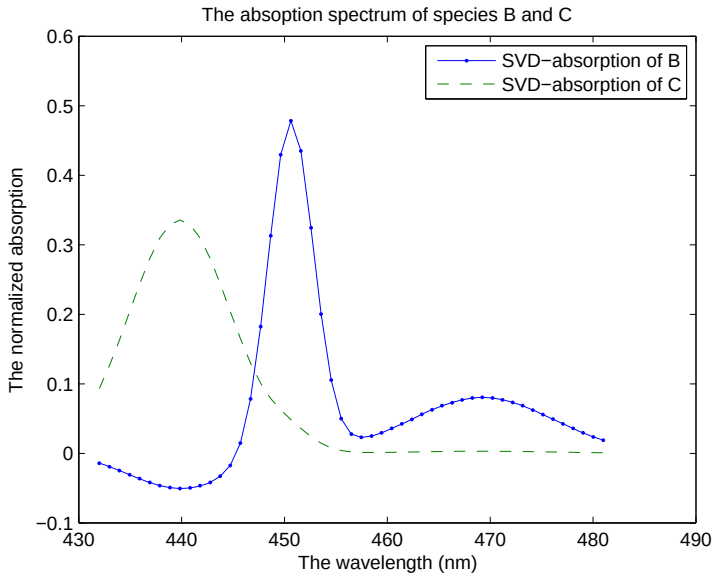

Figure 1: Normalized Absoption Spectrum of B and C obtained from SVD

only one peak, which suggests that B can be written as a sum of two different gaussians and C can be written as a sum of only one gaussian as shown in Equation(3).

$$\begin{aligned} A_B(\lambda) &= A_{B,1}\exp\left(-\frac{(\lambda-\lambda_{B,1})^2}{w_{B,1}^2}\right) + A_{B,2}\exp\left(-\frac{(\lambda-\lambda_{B,2})^2}{w_{B,2}^2}\right) \\ A_C(\lambda) &= A_{C,1}\exp\left(-\frac{(\lambda-\lambda_{C,1})^2}{w_{C,1}^2}\right) \end{aligned} \tag{3}$$

Putting everything together, we can get an analytical expression for $S_{model}(\lambda, t, \theta)$ as shown in Equation(4).

$$\begin{aligned} S_{model}(\lambda,t,\theta) &= [B]_0 e^{-k_1 t}\left(A_{B,1}\exp\left(-\frac{(\lambda-\lambda_{B,1})^2}{w_{B,1}^2}\right) + A_{B,2}\exp\left(-\frac{(\lambda-\lambda_{B,2})^2}{w_{B,2}^2}\right)\right) \\ &\quad + [B]_0\left(1+\frac{k_2 e^{-k_1 t}}{k_1-k_2}+\frac{k_1 e^{-k_2 t}}{k_2-k_1}\right)\left(A_{C,1}\exp\left(-\frac{(\lambda-\lambda_{C,1})^2}{w_{C,1}^2}\right)\right) \end{aligned} \tag{4}$$

As is mentioned in the problem statement we cannnot determine $[B]_0$ or the magnitude of absoptions independently and we have

to club some of the parameters together. If we do that we get Equation(5)

$$\begin{aligned} S_{model}(\lambda,t,\theta) &= e^{-k_1 t}\left(K_{B,1}\exp\left(-\frac{(\lambda-\lambda_{B,1})^2}{w_{B,1}^2}\right)+K_{B,2}\exp\left(-\frac{(\lambda-\lambda_{B,2})^2}{w_{B,2}^2}\right)\right) \\ &\quad+\left(1+\frac{k_2e^{-k_1t}}{k_1-k_2}+\frac{k_1e^{-k_2t}}{k_2-k_1}\right)\left(K_{C,1}\exp\left(-\frac{(\lambda-\lambda_{C,1})^2}{w_{C,1}^2}\right)\right) \end{aligned} \quad (5)$$

The parameters that we need to determined in the above model are k_1, k_2, $K_{B,1}$, $K_{B,2}$, $K_{C,1}$, $\lambda_{B,1}$, $\lambda_{B,2}$, $\lambda_{C,1}$, $w_{B,1}$, $w_{B,2}$ and $w_{C,1}$.

- *Model 2: Independent Reaction Model*
  In this reaction model the rate of change of B and C are independent of each other. The final expressions for the concentrations of B and C at any given time t is given in Equation(6).

$$\begin{aligned} [B] &= [B]_0 e^{-k_1 t} \\ [C] &= [Y]_0\left(1-e^{-k_3 t}\right) \end{aligned} \quad (6)$$

The absoption spectrum will be assumed to be similar to the one provided in Equation(3) and the analytical expression for $S_{model}(\lambda,t,\theta)$ can be written as shown in Equation(7).

$$\begin{aligned} S_{model}(\lambda,t,\theta) &= [B]_0 e^{-k_1 t}\left(A_{B,1}\exp\left(-\frac{(\lambda-\lambda_{B,1})^2}{w_{B,1}^2}\right)+A_{B,2}\exp\left(-\frac{(\lambda-\lambda_{B,2})^2}{w_{B,2}^2}\right)\right) \\ &\quad+[Y]_0\left(1-e^{-k_3 t}\right)\left(A_{C,1}\exp\left(-\frac{(\lambda-\lambda_{C,1})^2}{w_{C,1}^2}\right)\right) \end{aligned} \quad (7)$$

Again we can rewrite the model above in terms of the least number of determinable parameters as shown in Equation(8) and the parameters that we need to determine are k_1, k_3, $K_{B,1}$, $K_{B,2}$, $K_{C,1}$, $\lambda_{B,1}$, $\lambda_{B,2}$, $\lambda_{C,1}$, $w_{B,1}$, $w_{B,2}$ and $w_{C,1}$.

$$\begin{aligned} S_{model}(\lambda,t,\theta) &= e^{-k_1 t}\left(K_{B,1}\exp\left(-\frac{(\lambda-\lambda_{B,1})^2}{w_{B,1}^2}\right)+K_{B,2}\exp\left(-\frac{(\lambda-\lambda_{B,2})^2}{w_{B,2}^2}\right)\right) \\ &\quad+\left(1-e^{-k_3 t}\right)\left(K_{C,1}\exp\left(-\frac{(\lambda-\lambda_{C,1})^2}{w_{C,1}^2}\right)\right) \end{aligned} \quad (8)$$

2. The estimate of noise level given by SVD is 4.857×10^{-4}. This noise level is calculated by summing up the signal corresponding to all the small eigenvalues. Thus the noise matrix is given as

$$S=\sum_{i=3}\sigma_i x_i a_i^T$$

where x_i and a_i are respectively the concentration and absoptions of the species. We are told that the noise at each wavelength and each time comes from the same normal distribution. Thus we can calculate the standard deviation of the noise using the values obtained from the noise matix.

Since we dont know anything about the reactions taking place and the abosorption spectrum of the species we will just assume that all the parameters are uniformly distributed in some arbitrarily large range of values chosen. For example the values of the rate constant will be positive and have values less than 10^{15} which is the range of the rate constant for a unimolecular reaction. Also since we assume that all the parameters are independent of each other

$$p(\underline{\Theta}) = \prod_{i=1}^{n} p_i(\theta_i)$$

.

3. • *Model 1: Sequential Reaction Model*

   At t=0 we are told that the concentration of C is 0. Thus all the absoption signal that we observe is due to species B. Using Equation(5) we realise that the absoption at t=0 can be used to determine the value of $K_{B,1}$, $K_{B,2}$, $\lambda_{B,1}$, $\lambda_{B,2}$, $w_{B,1}$ and $w_{B,2}$. The matlab code to do the above fitting is provided in file `fitAbsForB.m` and the parameters that it calculates are $K_{B,1} = 0.3000$, $K_{B,2} = 0.0501$, $\lambda_{B,1} = 450.6204$, $\lambda_{B,1} = 469.2274$,$w_{B,1} = 3.1041$ and $w_{B,2} = 9.7590$. The comparison of the Best Fit Absorption and the Signal Strength at t=0 are provided in Figure 2.

   Also if compare the absoption spectrum of B and C using Figure 1, we realise that the absorption signal at wavelength 470 nm is solely due to B. Thus we can calculate the rate constant k_1 using the absoption signal at 470 nm and using the parameters $K_{B,1}$, $K_{B,2}$, $\lambda_{B,1}$, $\lambda_{B,2}$, $w_{B,1}$ and $w_{B,2}$ which we have calculated previously. The matlab code `bestfitk1.m` is used to get a best fit value of k_1. The best fit value of $k_1 = 0.0497$ and the graphical comparison of the model vs data is shown in Figure 3.

   From Figure 3 it is apparent that the concentration of B is close to zero at time t=100 s. Thus the at t=100, all the absoption is due to species C. We dont know the concentration of C and so we will be able to calculate the absoption profile of C to within a constant factor. This fitting is performed using the matlab file `fitAbsForC.m`.

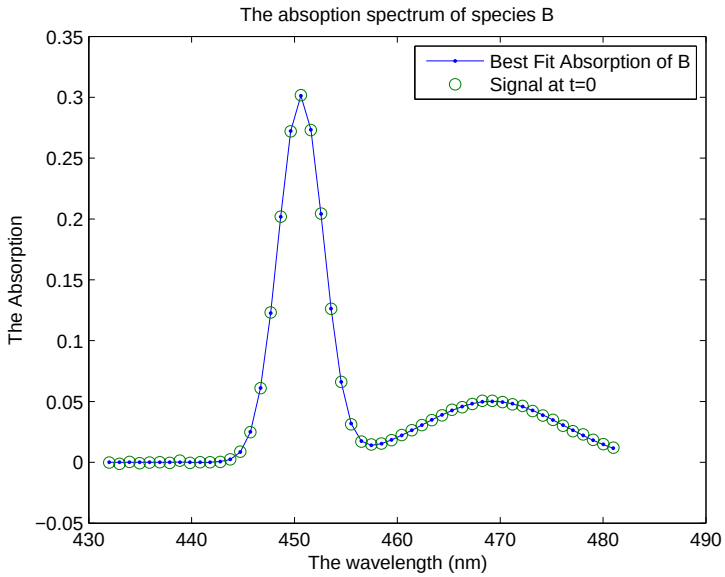

Figure 2: BestFit Absoptions of B compared to the Signal Strength at t=0

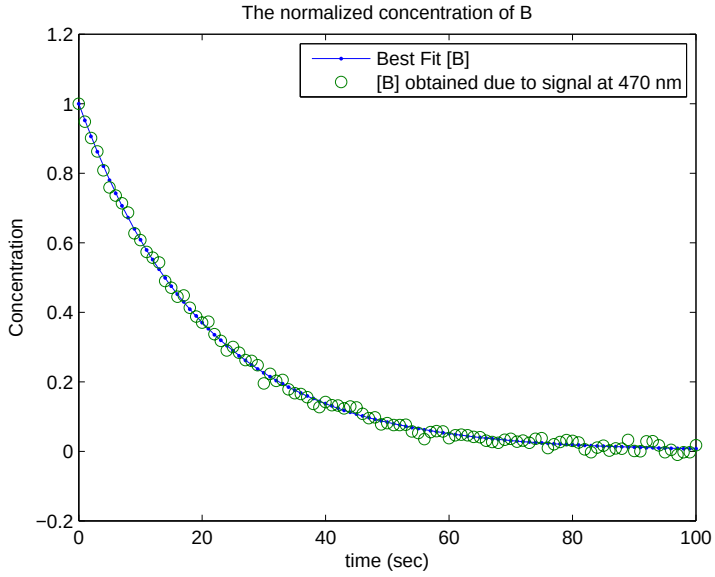

Figure 3: Normalized Absoption Spectrum of B

The parameter values are $K_{C,1} = 0.7339$, $\lambda_{C,1} = 439.844$ and $w_{C,1} = 6.94$. The comaprison of fitted value of absoption of C and signal strength at t=100 s is given in Figure 4.

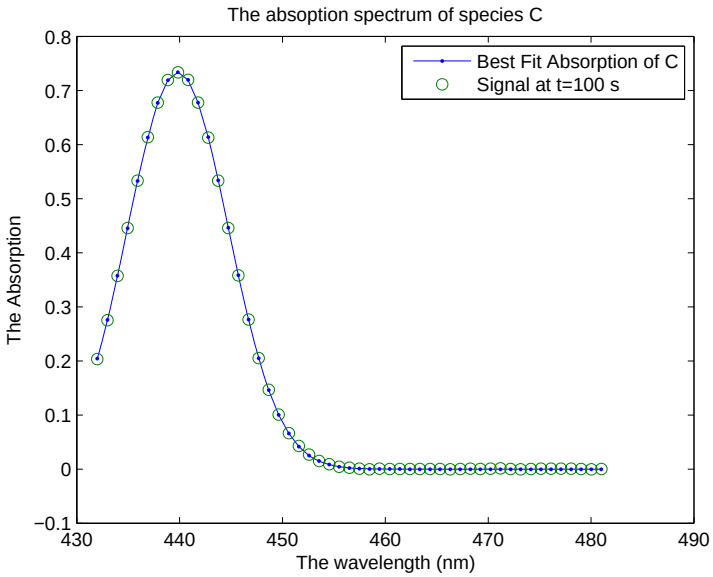

Figure 4: BestFit Absoptions of C compared to the Signal Strength at t=101

Finally we see that C has a absorption maximum at 439.844 nm and B does not absorb at that concentration. Thus we can use the signal strength at 439.844 nm to calculate the best fit value of k_2. This is done in matalb file `bestfitk2_model1.m`. The best fit value of $k_2 = 0.0889$ and the graphical comparison of normalized concentration of B with the signal at 440 nm is given in Figure 5.

We can now use all the values, estimated so far, as the initial guess for a routine which tries to fit the entire matrix of signal. The matlab function that accomplishes this task is `p3_model1_error_bestfit.m`. The deviation between model prediction and the data and the error is plotted in Figure 6. The best fit values of the parameters is given in Table 1.

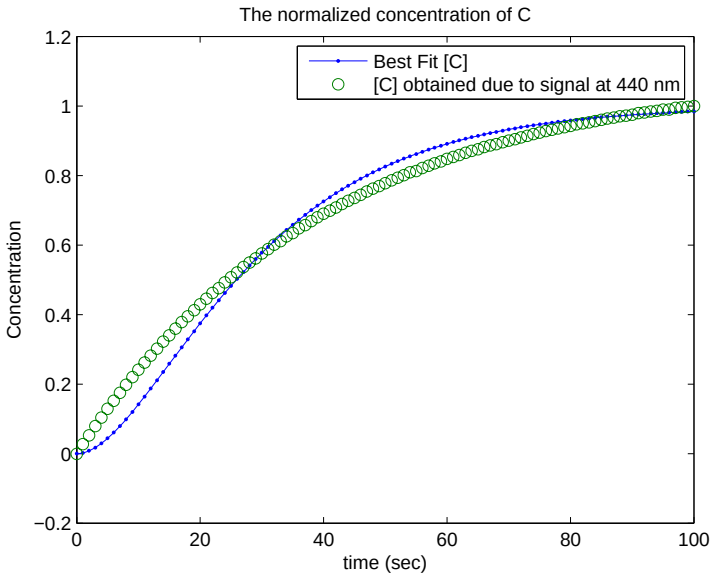

Figure 5: Normalized Absoption Spectrum of C

Table 1: Best Fit Values for Model 1

k_1	0.0286
k_2	0.8123
$K_{B,1}$	0.2259
$K_{B,2}$	0.0369
$K_{C,1}$	0.7733
$\lambda_{B,1}$	450.53
$\lambda_{B,2}$	469.23
$\lambda_{C,1}$	439.79
$w_{B,1}$	3.2342
$w_{B,2}$	9.7594
$w_{C,1}$	6.7872

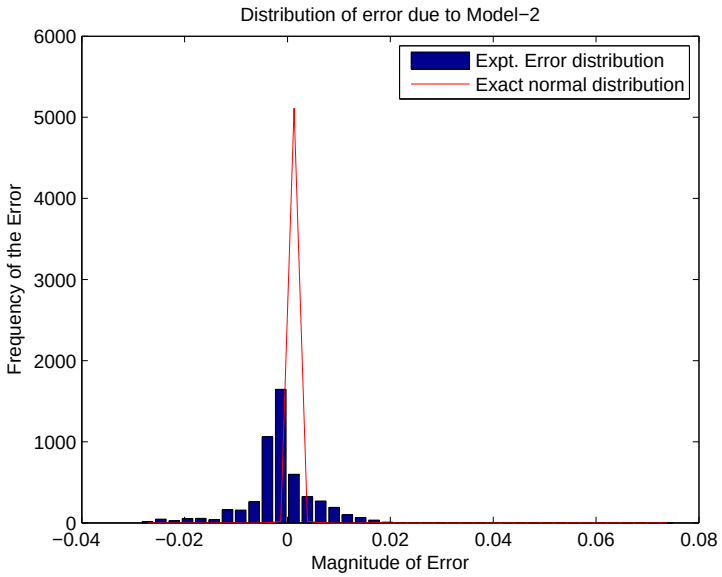

Figure 6: Error Distribution due to Model 1

- *Model 2: Independent Reaction Model*

  For this model the initial guesses for all the parameters besides k_3 is calculated in exactly the same way as the previous Model. For calculating the value of k_3, we take the wavelength 440 nm and fit the concentration of C to the analytical formula given in Equation(6). This fit is performed using the matlab function given in command `bestfitk2_model2.m`. The comparison of the fit with the Signal strength at 440 nm is given in Figure 7 and the value of $k_3 = 0.0303$. Again we use the values, calculated so far, as the initial guesses and perform a full fledged optimization to calculate the values of all the parameters using the matlab function `p3_model2_error_bestfit.m`. The values of the parameters calculated are given in Table 2. The distribution of the error is shown in Figure 8.

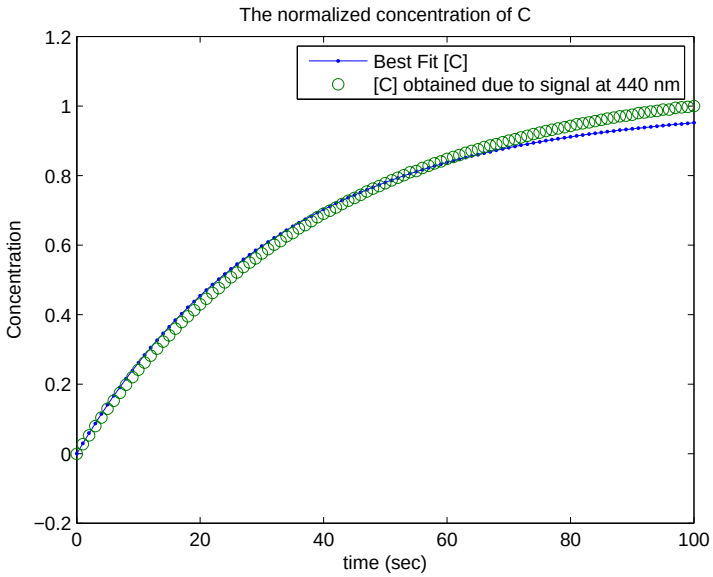

Figure 7: Normalized Absoption Spectrum of C

Table 2: Best Fit Values for Model 2

k_1	0.0500
k_3	0.0250
$K_{B,1}$	0.3000
$K_{B,2}$	0.0501
$K_{C,1}$	0.7999
$\lambda_{B,1}$	450.62
$\lambda_{B,2}$	469.24
$\lambda_{C,1}$	439.84
$w_{B,1}$	3.0993
$w_{B,2}$	9.7915
$w_{C,1}$	6.9303

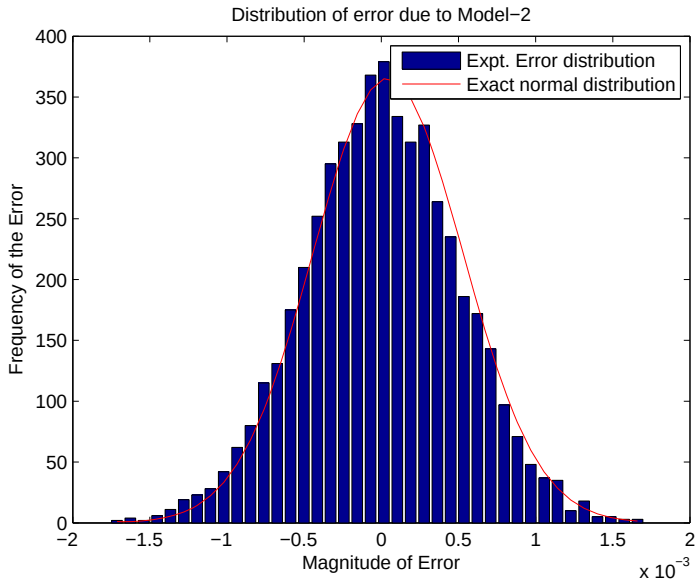

Figure 8: Error Distribution due to Model 2

4. To apply the χ^2 test we assume that the underlying distribution of errors is a normal distribution. Then we calculate the χ^2 value of the distribution using the Equation(9)

$$\chi^2 = \sum_i \frac{\delta_i^2}{\sigma} \tag{9}$$

In the above equation δ_i are the individual errors and σ is the standard deviation of the error around 0 that we have calculated in part 2 of the problem. The value χ^2 will be distributed according to the χ^2-distribution if the number of sample points in large. In our case the number of sample points are large (5151 points). The number of fitted parameters are 11 and thus the degree of freedom of the χ^2-distribution is 5140 (=5151-11).

- *Model 1: Sequential Reaction Model*
  The χ^2 value of model 1 $= 1.3 \times 10^6$. The p-value is 0. Thus there is 100% probability that the model is incorrect.
- *Model 1: Sequential Reaction Model*
  The χ^2 value of model 1 $= 5.480 \times 10^3$. The p-value is 0.0005. Thus there is 99.95% probability that the model is incorrect.

The values of χ^2 of model 1 ad model 2 are calculated using the functions `p3_model1_error_bestfit.m` and `p3_model2_error_bestfit.m` respectively.

5. For Model 2 we have made an array of k_1 and k_3 and calculated χ^2 values for all the combinations of k_1 and k_3. The graph is plotted using the matlab function `p5_contourplots.m` and the graph is shown in Figure 9.

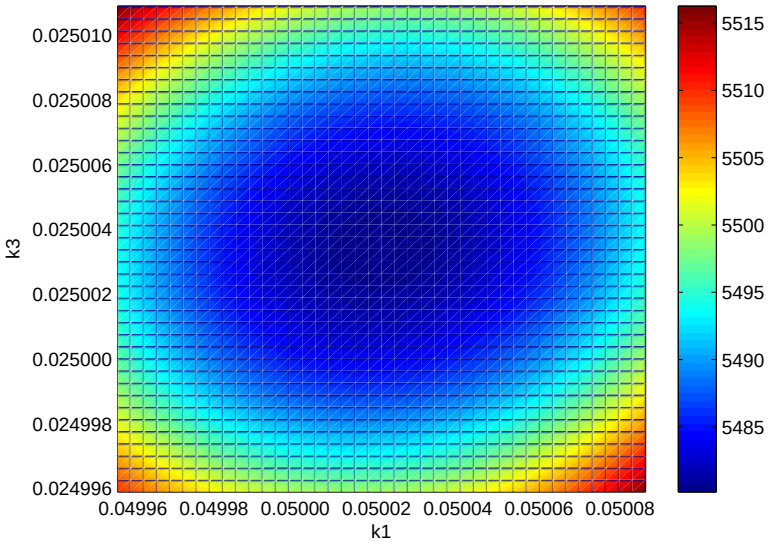

Figure 9: χ^2 with different values of k_f and k_r

To convert the χ^2 plot to a plot of probability we just calculate the value of $\Delta\chi^2 = \chi^2(k_1, k_3) - \chi^2_{min}$ and this value of $\Delta\chi^2$ is distributed with a χ^2 distribution of 2 degrees of freedom. The contour plot of 95% and 90% confidence interval of k_1 and k_2 are plotted in Figure 10.

From Figure 10 we can read off the 95% confidence interval for k_1 and k_2 as $0.04998 \leq k_1 \leq 0.050065$ and $0.024995 \leq k_3 \leq 0.025008$. Both the figures are generated using the matlab code `p5_countourplots.m`.

We can also use the built-in matlab functions `nlinfit` and `nlparci` to calculate the confidence intervals on k_1 and k_2 and the confidence intervals given by these functions are $0.049968 \leq k_1 \leq 0.050074$ and $0.024993 \leq k_3 \leq 0.025014$. The matlab function which generates these confidence inervals is `p5_model2_param_est.m`. The confidence intervals provided by the matlab function are very close to the values given

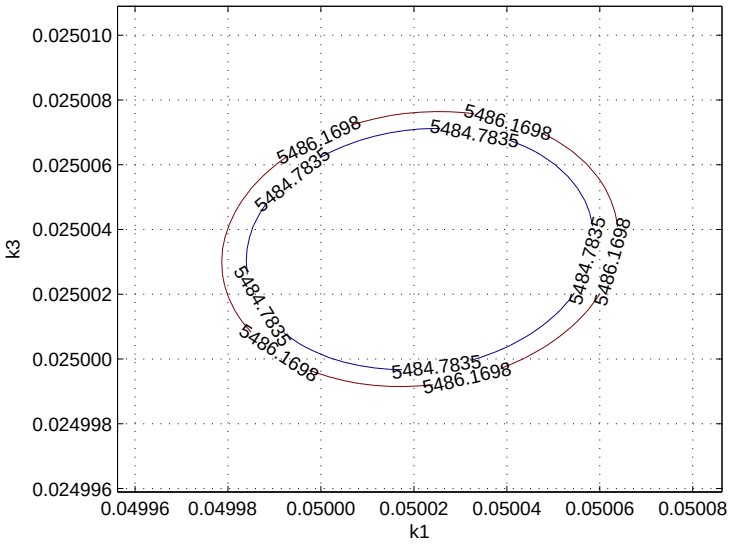

Figure 10: 90% and 95% confidence contours

by plotting the graph. Due to the large number of data points (5151 data points) and relatively fewer fitted parameters, the confidence intervals of all the parameters are small. The matlab functions approximate the bottom of the well of the χ^2 distribution as a parabola. If the confidence intervals are small then the parabola approximation is usually very good.

6. In part 2 of the problem we assumed that the distribution of θ was uniform. Specifically for k_1 and k_2 we assumed the probability distribution was uniform between 0 and 10^{14}. If one were to normalize the probability distribution the constant multiplying factor of $p_{prior}(k_1, k_2)$ will get eliminated. Thus for a uniform prior probability distribution the answer to this part will look exactly like Figure 10. The confidence intervals calculated will also remain unchanged. The values of k_1 and k_2 will be effected only if we have specific prior information about the probability distribution of k_1 and k_2.