Biophysical Chemistry – CH 4404 01
Assignment 9B – Computational Lab

Due Tuesday, December 3 (at the start of class)

Introduction

During the last three lectures, we have discussed chemical kinetics and reaction mechanisms. Mathematically, we have used differential equations to derive the anticipated time course of a reaction given a specified mechanism. In many of our explorations, we have arrived at rather complex systems of coupled differential equations: even seemingly trivial reaction mechanisms can often lead to equations that are impossible to solve analytically.

When no analytical solution exists, scientists often resort to numerical methods for solving differential equations. These methods are collectively referred to as *numerical integration*, because the equations are integrated computationally to give the final integrated result. Assuming a set of starting conditions, the calculation proceeds by determining how the system changes during a small time step; eventually, after enough time steps have been calculated, the final state is determined.

Additionally, throughout the course we have examined how different mathematical models can be used to describe our observations of chemical systems. These models have parameters that need to be *fit* (or adjusted) to optimize the agreement between the model and the observed data. Thus, in addition to allowing us to solve complex differential, computers have made it possible to determine the best fit parameters in a straightforward way.

An in-depth discussion of either of these topics – numerical integration or model fitting – is well beyond the scope of this course; nevertheless, a simple Excel worksheet can provide some insight on what is going on when scientists solve differential equations computationally. In this assignment, we will examine numerical solutions to some chemical kinetics problems, including several mechanisms we have already worked with in our discussions so far. Then, we will look at some actual data and see how we can adjust model parameters manually to obtain good agreement with our experimental results.

This assignment contains both written and computational portions. You will be expected to submit several Excel worksheets via email to the instructor at nfitzkee@chemistry.msstate.edu. If you wish, you may submit the written portions electronically as well, although it's probably easiest to submit them in person. Either way, all emails and answers must be received by the specified deadline to receive full credit.

Numerical Integration: The Euler Method

For a first-order chemical mechanism (A → B), one arrives at a simple differential equation to describe the dependence of A vs. time.

$$\frac{dA}{dt} = -kA$$

This equation says that the rate of change of A (the derivative) is proportional to A itself; when A is small, the rate of change in A is correspondingly small. When A is large, it will change faster. In class, we showed that this equation can be solved analytically using separation of variables. But what if we didn't know how to do that? Could we still solve the equation if we were willing to make enough calculations?

The answer, as you may have guessed, is yes, and the simplest procedure is as follows: Suppose we know the original concentration of A at $t = 0$, and that $A(0) = A_0$. We'd like to calculate the concentration of A at a short time later, $t = \Delta t$. If Δt is small enough, we can write:

$$\frac{\Delta A}{\Delta t} = -kA \tag{1}$$

$$\Delta A = -kA\Delta t \tag{2}$$

So, the change in A after a small Δt is going to be the kinetic constant k, multiplied by the current value of A, multiplied by the time step Δt. So, we can *approximate* $A(\Delta t)$ using the following formula:

$$A(0 + \Delta t) = A(0) + \Delta A \tag{3}$$

$$A(\Delta t) = A(0) - kA(0)\Delta t \tag{4}$$

As Δt becomes smaller, the approximation will be more accurate. The key point here is that we have used the value of A at time 0 to determine the value of A at time Δt. A more general form of this equation can be written for later time points:

$$A(t_i + \Delta t) = A(t_i) + \frac{dA}{dt}\Delta t \tag{5}$$

$$A(t_i + \Delta t) = A(t_i) - kA(t_i)\Delta t \tag{6}$$

$$A(t_{i+1}) = A(t_i) - kA(t_i)\Delta t \tag{7}$$

All the terms on the right hand side of the equation are known, whereas the next value for A is not known. Thus, *we can calculate future values of A using values of A from the past*. This is the fundamental idea of Euler integration (and more sophisticated methods as well). It makes good physical sense, too. Similar equations can be derived for how B responds to time, and the method can be generalized to solve any system of differential equations.

As an example of this, download `first_order.xlsx` from the course web page. This is an Excel spreadsheet that implements this idea for a first order kinetic scheme. There are two worksheets in the workbook: One is called "Simulation Data" and contains the parameters as well as values for the plot. The other worksheet is called "Simulation Plot," and this sheet contains the plot of the simulation. Since we have an analytical solution for this case, I have included the analytical curves on the plot for comparison.

Looking at the simulation data, you can see that many time points are implemented (about 15,000). Depending on the complexity of the reaction, you will need more or fewer time points to simulate the kinetics accurately. After the first time point, the rest are generated automatically using Excel's formula capability. Similarly, the initial values for A and B are placed into the first time point row; future time points must be calculated from these values. Looking carefully at the formulas for $\frac{dA}{dt}$ (cell C19) and $\frac{dB}{dt}$ (cell E19), you should see equation (1): we calculate the derivatives using the known values of A and B. Moving to the first calculated time point (the second row of data, row 20), we now calculate the concentrations of A and B using equation (5) (see cells B20 and D20). The process is repeated 15,000 times, and eventually the concentration dependence of A and B is determined. You can change the values for the time step, the kinetic rate constant *k*, as well as the initial concentrations of A and B to see how the simulation changes when compared to experiment.

One thing is very important to remember: **by applying the Euler method, we are generating an *approximation* of the analytical solution**. As you will see below, our numerical solution can be made more or less accurate by changing the size of the time step. This is a fundamental characteristic of all numerical methods: a smaller time step will produce more accurate results, but it will require more calculations for the same amount of simulation time. For example, suppose you wanted to calculate the concentration of A at 5 seconds. If you used a time step of 0.01 s, this would require 500 integration steps. On the other hand, if you used a time step of 0.0001, this would require 50,000 steps. The longer time step would require fewer integration steps to get to 5 seconds, but it would be less accurate than the shorter time step for the concentration of A. And both values for A would still differ slightly from the analytical solution for A vs. t if such a solution exists. Thus, the Euler method only allows us to calculate an *approximate* answer for A, but that approximation can be made more and more accurate by using a smaller time step.

A complete discussion of the error introduced by the Euler method beyond the scope of our course, but it can be shown that the integration error for a single step from t to $t + \Delta t$ is proportional to $(\Delta t)^2$. Compared to other methods for numerical integration, this is quite poor, and algorithms are frequently used with an error proportional to $(\Delta t)^4$ or better (compare $(\Delta t)^2$ vs. $(\Delta t)^4$ when $\Delta t = 0.01$ s). These more advanced algorithms allow the calculation of accurate results with much fewer calculations, giving scientists an accurate answer very quickly. This is important because the integration error for a single step will tend to accumulate over many steps. It is also important because, in many advanced scientific applications, no analytical solution exists. It's therefore necessary to make our numerical solution as accurate as possible. An example of an algorithm which outperforms the Euler method is the 4th-order Runge-Kutta method.

At this point, you should spend some time looking at the spreadsheet, changing the parameters, examining the formulas, etc. Below, you will answer some questions about your observations. You should also take a look at the equations themselves. Can you see how the Euler equations relate to the formulas in the Excel sheet?

Depending on your familiarity with Excel, you may not recognize the dollar signs present in some of the formula. In this class, I will assume that you can write formulas in Excel and use the drag feature to copy those formulas down to other cells. When you do this, Excel normally tries to adjust your formulas so that the references will update with each new row or column you occupy. However, sometimes you want a reference to stay in the same place. The dollar signs tell Excel that you want that row or column reference to stay constant, even if you copy the formula somewhere else. For example, if a cell contains the formula “=A1*5”, you can drag that cell down a column to fill that formula throughout the column. Normally, Excel would adjust that cell to “=B1*5,” “=C1*5” etc. But if you entered “=A1*5”, that value would stay constant, no matter where you moved the cell. This can be very useful if you want a cell to contain a constant that can be used in other equations, and we will use it here for our rate constants, initial concentrations, etc. Obviously, if one cell contains a constant, it shouldn’t move around as you use it in formulas.

Examining the simulation data (re-download the file if you changed the original values), you can observe that the calculated values (solid curves) overlay almost perfectly with the analytical solution. In this assignment, we will experiment with this behavior: under what conditions do we get good agreement? We will work up to the point where you will be able to set up your own simulation for a fairly complex mechanism that cannot be solved analytically.

Questions

1. Consider the `first_order.xlsx` spreadsheet.

    a. Looking at the spreadsheet, what is the Excel formula (in Excel syntax) for calculating *A* for the second time point (cell B20)? Explain how this formula is equivalent to equation (4) above. (4 points)

    b. Vary the time step for the simulation. At what time step do you begin to see the calculated curve deviate from the analytical solution? (2 points)

    c. Start at the time step from (b) and gradually increase the time step to 0.1 s. Describe the behavior you observe. Why do yout think the simulation behave this way? Frame your answer in terms of the differential equation used to describe this system. (3 points)

    d. In this experiment, we always calculate 15,000 steps. If we wanted to perform fewer calculations, we could stop the experiment after A had decayed to 1% of its original value (for example). Using the original time step of 0.0001, when does this occur? (1 points)

    e. Suppose that it took several seconds to calculate the value of each cell instead of less than a millisecond. Imagine that you wanted to stop the simulation before 15,000 steps, when the reaction had finished. What advantage would there be to using a larger time step? (3 points)

2. Download the `equilibrium_blank.xlsx` spreadsheet from the course web page. This spreadsheet contains a similar situation except that A and B are now allowed to be in equilibrium. Carefully study the equations used to generate the values of A and B in the simulation and answer the questions below.

a. Given initial concentrations A_0 and B_0, as well as the kinetic rates constants k_1 and k_{-1}, use the methods you learned in chapter 4 to write a general expression or calculating the equilibrium concentration of A and B (A_{eq} and B_{eq}). Submit your derivation separately, and fill in your formulas in to cells F11 and F12. *Hint:* There is really no "kinetics" involved in this problem: it's simply a matter of finding equilibrium concentrations given initial conditions, like we did with a table in chapter 4. (4 points)

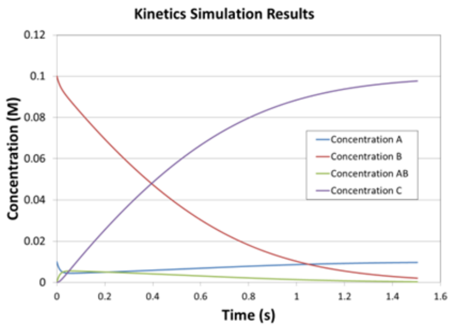
b. The analytical solution to this kinetic scheme is found on p. 350 in your book. Convert this equation into Excel formulas for the concentration of A and B vs. time using the given rates and starting conditions. Then, fill in the appropriate columns for A and B in terms of the parameters given in the Excel spreadsheet (the first cells of your calculation are G20 and H20, then you should fill those formulas down to the bottom of the worksheet (G15053 and H15053). Using the given time step and rate constants, you should see excellent agreement with the simulated concentrations. Email your final spreadsheet to the instructor as `<lastname>_equilibrium.xlsx`. (5 points)

c. Gradually increase the time step of this simulation. At what point to you begin to see deviation from the analytical result? (2 points)

d. Given your results from part 1(c) and 2(c), can you come up with a rule of thumb for determining the maximum "safe" time step to use? *Hint:* Compare the maximum time step to the rate constant, and keep in mind that, if the rate constant is k, the "characteristic time" for that rate constant is $\tau = 1/k$. (4 points)

e. Using the given kinetic constants, set the time step to 0.025 s. This value should be above the maximum you determined in part (d). Describe the behavior of the simulation for this time step. Why do you think this is happening? Describe your answer in terms of the differential equations for this mechanism. (3 points)

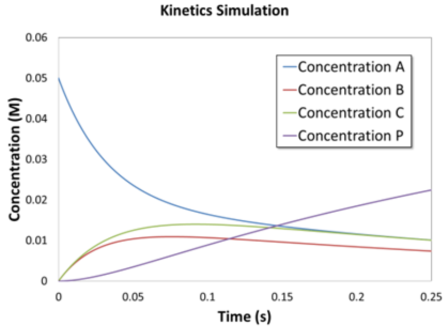
3. Now, we're going to get some practice actually integrating a differential equation numerically. Download the `enzyme_blank.xlsx` spreadsheet. In this spreadsheet you will the mechanism for a simple enzyme: the enzyme (A) binds to the substrate (B) to form a complex (AB). Then, the enzyme can catalyze the formation of product (C) from B, leaving the enzyme unchanged. For this scheme, write and submit the complete set of differential equations describing the behavior of A, B, AB, and C. Then, fill in the appropriate cells to calculate these concentrations over time using the Euler method. Submit your final spreadsheet to the instructor as `<lastname>_enzyme.xlsx`. As a hint, your final graph should look like this: (10 points)



Now, we will combine the ability to simulate differential equations with the functionality to adjust the rate constants to optimize the best fit to some experimental data. At this point you should read the supplemental handout on models and model fitting. Then, complete the problem below.

4. Download the `triangle_blank.xlsx` spreadsheet from the course web page. This spreadsheet simulates the triangular mechanism from last week's assignment. In this question, you will complete the simulation for this mechanism, and then you will optimize the fit to the included experimental data.

   a. In the excel spreadsheet, the column for $\frac{dB}{dt}$ has been left empty. Fill in this column with the appropriate formula. If you're stuck, look at the solutions to last week's assignment. Do not manipulate the time step or rate constants yet. When you are finished, the simulation plot will look like this: (5 points)



   b. Now click on the "Fitting" worksheet. From this worksheet you can compare how your simulated curves agree with the experimentally determined concentrations of A, B, C, and P. On this worksheet, you should manipulate the rate constants to optimize the agreement between the model and the data, as described in the models handout. *Do not adjust the rate constants on any of the other worksheets except for the Fitting worksheet!* You can examine how the SSR and residuals plots change as you manipulate the rates. When you are satisfied with your fit, you should submit this worksheet as `<lastname>_triangle.xlsx`. (10 points)

   A strategy for finding a good fit would be manipulate the rate constants in medium-size increments at first (10 s^{-1}), trying to get a rough set of parameters that fits the data. It would be easiest to adjust the scale on Excel so that you can see all the curves on your monitor at once. That way, if one curve gets better while another gets worse, you can adjust to avoid getting on the wrong track. Once you have a set of rough values, then you can start making smaller changes to hone in on the right solution. This solution is intrinsically non-linear, and therefore Excel's solver may have trouble finding an answer; however, motivated students are welcome to give it a try.

As an incentive for finding a good fit, the best SSR will receive **five bonus points** on this assignment. The second best SSR will receive three bonus points.

c. After you have finished optimizing your answer, use your simulation to determine the equilibrium concentrations of A, B, and C if k_4 is zero? This is not hard – simply set k_4 to zero and read the final concentrations off of the graph. (3 points)

At this point, hopefully you appreciate the sophistication of what you've done in this assignment: not only have you developed a method for solving *any* differential equation numerically, you've been able to optimize the solution to a differential equation so that it fits to experimental data. In the "real world" this would be done using more efficient algorithms with computer-aided optimization, but at a fundamental level those computer programs are simply implementing what you've been able to do here manually. And while there hasn't been any programming in this assignment, the tools you have implemented are similar to those that software developers would write to solve these problems in a more advanced way.

Authorship credit: Parts of this assignment were implemented by Will Buchanan as his undergraduate research project during the Spring of 2011.