Computer Laboratory Manual: Calculus for the Life Sciences

MATLAB and Maple Edition

Joseph M. Mahaffy

March 2014

Contents

1 Introduction **1**

2 Graphing in MATLAB and Maple **3**
1 Basic Graphing in MATLAB . . . 3
2 Basic Functions and Graphing in Maple . . . 7
3 Problems . . . 8

3 Function Models **13**
1 Error Analysis . . . 13
2 Linear Models . . . 15
3 Data and MATLAB . . . 18
4 Exponential and Logarithmic Functions . . . 28
5 Allometric Modeling . . . 30
6 Other Nonlinear Functions . . . 42
6.1 Weak Acid Chemistry . . . 42
6.2 Rational Functions . . . 44
6.3 Michaelis-Menten Enzyme Kinetics . . . 48

4 Discrete Dynamical Models **53**
1 Discrete Malthusian Growth . . . 53
2 Other Growth Models: Nonautonomous, Logistic, and Immigration . . . 64
3 Breathing Model . . . 78

A List of Problems **83**

CHAPTER 1:
INTRODUCTION

This document includes a LaTeXversion of many WeBWorK Laboratory problems created in the past. The problems are divided according to the mathematics that they use. The five categories are functions, discrete dynamical models, differentiation and optimization, differential equations, and integration. This manual begins with a brief introduction to basic graphing in MATLAB and Maple of simple functions and commands to use functions. Most computer applications in this lab manual are introduced with an example, then problems are presented to practice the computer techniques on interesting applications.

MATLAB is a short for Matrix Laboratory. This program was originally developed by Cleve Moler at the University of New Mexico with funding from NSF (National Science Foundation) in the 1970s. Its original design was to create an easy computer program to manage vectors and matrices, and its core language is C. It has been adopted widely for science, engineering, research, and teaching, and ranks very high amongst programs used worldwide. It is considered critical to know for any engineer and is often the favorite programming language of applied mathematicians and other scientists. The program is continually expanding through the packages that MathWorks develops, including recent expansions into bioinformatics and robotics. Its power is primarily from the routines that have been developed for advanced number crunching (numerical analysis) and graphing abilities. This manual is meant only as an introduction to MATLAB and will be fairly limited in the scope of MATLAB commands introduced. However, a student working through this manual will become fairly competent in using the language.

Maple is a Symbolic Algebra package. Maple was developed in 1980 at the University of Waterloo to become a more efficient program to its predecessor Macsyma. Most science, mathematics, and engineering students need to know some symbolic algebra program to handle complex modeling problems in the science of today. Maple's chief competitor is Mathematica, but at the level of this manual either language could be used with only minor variations in the syntax of the commands. Maple is continually upgraded and can manage a wide range of mathematics. Students of science and engineering today need both the numerical power of MATLAB and the symbolic algebra abilities of Maple. As with MATLAB, this manual only provides some basics to introduce students to the program.

The Problems in this Lab Manual were originally developed for a computer laboratory component of a Calculus for the Life Sciences course. Students can see a more complete development of many of the Calculus and modeling ideas in this Lab Manual in the texts *Calculus: A Modeling Approach for the Life Sciences*[9, 8]. The original problems were designed to introduce students to modeling applications of Calculus and teach basic programming techniques. The primary computer applications taught were Excel, Maple, and Word. To provide additional variation in the problems, they were adapted to WeBWorK, using its ability to individualize the problems with random changes to the parameters. (The problems in this text include the WeBWorK file name for easy reference, though some modifications are made.) This text has one version of the WeBWorK problems including most of the blanks that the students input to WeBWorK. The problems also

include parts that are meant to be written into a lab report. The lab report is designed to have students write complete paragraphs, explaining what was done, and to demonstrate their ability to create good graphs. The lab report portion of the problem will say, “In the Lab Report...”

CHAPTER 2:
GRAPHING IN MATLAB AND MAPLE

Keywords: Graphing, MATLAB, Maple

This chapter shows the basic commands needed to produce a graph in MATLAB and Maple. An example for graphing a line and a quadratic is shown for both programs. As the manual progresses, the reader will learn to build on these basic commands to produce more intricate graphs. There are tutorials for both MATLAB and Maple available through the program help page, which the novice should consider viewing. This manual will present a number of good references for additional reading.

1 BASIC GRAPHING IN MATLAB

This section provides the basic MATLAB commands needed to produce a good graph for a linear and a quadratic function. The author has designed a WeBWorK problem to examine the linear and quadratic functions:

$$f(x) = 2\,x - 2.5 \qquad \text{and} \qquad g(x) = 5 - 3\,x - x^2. \tag{2.1}$$

MATLAB does have a Symbolic Math Toolbox, which is based on an old version of Maple. This symbolic package does find solutions to basic equations. For example, if we wanted to find the roots of the function $g(x)$, then the following MATLAB commands produce the answer:

```
syms x     % This invokes the Symbolic Math Toolbox and defines the variable x.
solve(5 - 3*x - x^2 == 0)   % This is the quadratic equation to be solved.
```

Any note following the % symbol is simply a comment in MATLAB, which the user may want to include for understanding the program. MATLAB ignores all comments in its execution. The answer that MATLAB produces is assigned to the variable 'ans.' For this quadratic equation, MATLAB gives the solution:

```
ans =
   29^(1/2)/2 - 3/2
 - 29^(1/2)/2 - 3/2
```

The reader can easily check with the quadratic formula that the answer to $g(x) = 0$ is

$$x = \frac{-3 \pm \sqrt{29}}{2}.$$

To find the point of intersection the following MATLAB commands are used:

```
syms x y
[x_soln,y_soln] = solve(y == 2*x - 2.5, y == 5 - 3*x - x^2)
```

MATLAB creates the vectors 'x_soln' and 'y_soln,' which have the solutions, so the points of intersection are found to be:

$$(x_1, y_1) = \left(\frac{-5 - \sqrt{55}}{2}, -\frac{15}{2} - \sqrt{55} \right) \quad \text{and} \quad (x_2, y_2) = \left(\frac{-5 + \sqrt{55}}{2}, -\frac{15}{2} + \sqrt{55} \right).$$

Note that these x-values are used to determine the domain of interest for the graph.

Graphing M-file

The main purpose of this section is to introduce how to create a graph in MATLAB, which can be used in a Lab Report. Before progressing to the actual MATLAB script, there are a couple of file management issues with MATLAB. The line above the *Command Window* shows the directory where MATLAB is looking for files and is where it creates and stores files. It is very important to have this line pointing to the folder where you are keeping your working files and data before proceeding. Often it is valuable to have in this folder a copy of the graphing M-file template (below) that can be copied to create new graphs. MATLAB searches this folder first, then goes to its root folders to execute commands.

MATLAB uses M-files as programs. These can be created from the *Home* folder by clicking on the *New Script* icon. This puts the user into the *Edit* mode to create new programs or functions. Once *saved* as an *M-file*, a new script can be executed by clicking on the *Run* icon in the editor mode or by typing the file name in the Command Window. It is easiest to copy a working *M-file*, then edit the file to produce the desired result.

Below is the script for creating a graph in MATLAB. This file is named *line_quad.m*. There are extensive comments, but a detailed description is presented after the script along with the output.

```
clear               % Clear previous definitions
figure(1)           % Assign figure number
clf                 % Clear previous figures
hold off            % Start with fresh graph

mytitle = 'Line and Quadratic';  % Title
xlab = '$$x$$';                % X-label
ylab = '$$y$$';                % Y-label

x1 = -7;       % Left x-value
x2 = 2;        % Right x-value
Npts = 500;    % Number of points plotted between x1 and x2
xx = linspace(x1,x2,Npts); % Creates vector of x-values

yy1 = 2*xx - 2.5;        % Equation of Line
yy2 = 5 - 3*xx - xx.^2;  % Equation of Quadratic
xpt = [-6.21,1.21];      % x points of intersection
ypt = [-14.9,-0.084];    % y points of intersection
```

```
plot(xx,yy1,'r-','LineWidth',1.5);  % Graph of Line
hold on                             % Plots Multiple graphs
plot(xx,yy2,'b-','LineWidth',1.5);  % Graph of Quadratic
plot(xpt,ypt,'ko','MarkerSize',10); % Intersection Points
grid                                % Adds Gridlines
legend('$$f(x)$$','$$g(x)$$',4);    % Inserts Legend in lower right
text(-6.2,-16,'(-6.21, -14.9)','FontSize',12); % Adds Text - intersection pt
text(-0.4,0.3,'(1.21, -0.084)','FontSize',12); % Adds Text - intersection pt

y1 = -25;             % Lower y-value
y2 = 10;              % Upper y-value
axis([x1 x2 y1 y2]);  % Defines limits of graph

fontlabs = 'Times New Roman';   % Font type used in labels
xlabel(xlab,'FontSize',14,'FontName',fontlabs);  % x-Label size and font
ylabel(ylab,'FontSize',14,'FontName',fontlabs);  % y-Label size and font
title(mytitle,'FontSize',16,'FontName','Times New Roman'); % Title size/font
set(gca,'FontSize',12);         % Axis tick font size

print -depsc line_quad.eps        % Create figure as EPS file
% print -djpeg line_quad.jpg      % Create figure as JPEG file
```

The *M-file* above creates a graph of the Eqns. (2.1). This file can be readily modified to create other graphs. The file is divided into several parts. The first four lines are for safety to clear variables from memory in MATLAB to start new. The next three lines provide the title and labels for the x and y-axes. These can be changed for a different graph. The labels with the $$ use a special Math font. **Note**: On some versions of MATLAB, the action to copy and paste the PDF text from this Lab Manual into the MATLAB Editor results in errors at the apostrophes throughout the script. The apostrophes may be red and underlined in the copied script. This is easily repaired in the MATLAB Editor by finding the *Find & Replace* window in the Editor. (It is under *Navigate*, then *Find* in MATLAB R2012b.) In this window, select one of these red apostrophes by highlighting and copying this into the *Find what* window. In the *Replace with* window, type an apostrophe. Choose *Replace All*. Save this script or function, and it will *Run* normally in MATLAB.

The third set of lines defines the domain of the graph. The user adjusts the x1 and x2 values to set the endpoints for the domain of the graph. The script file evenly divides this interval into Npts (set at 500) above. The vector xx stores these values for function evaluation.

The next two lines provide the y-values for the line and quadratic function. These are stored in the vectors yy1 and yy2. Notice that these functions use the values from the vector xx. Since xx is a vector, then writing a function of x evaluated at all points of the vector requires the MATLAB arithmetic operators, which operate term by term. Thus, to obtain x^2 in a function, it is written xx.^2, where the period character (.) provides the appropriate array operation. Notice that scalar multiplication does not require the period. The two lines after that are the x and y-values of the points of intersection.

The next eight lines create the graph of the functions. The plot commands start with the x-vector, xx, and the y-vector, yy1 or yy2. The third entry gives the *line color* and style. MATLAB has the predefined colors red (r), blue (b), black (k), white (w), green (g), yellow (y), magenta (m),

and cyan (c). Other colors can be defined using an RGB vector, which we will defer until later. The *line style* has several forms, including continuous (-), dotted (:), dashed (–), and dash-dot (-.). For this graph the continuous red and blue lines are used. Finally, a line width is specified, giving the point width of the line. The last plot command enters the two points of intersection, creating black circles at these points (ko). The *hold on* command has the graphs overlay each other, while the *grid* command puts grid lines on the graph. The *legend* command inserts a legend on the graph, and there are many options for locating it. The 4 places it in the lower right corner. The last two lines in this group allow the user to add text to the graph. The first two entries of the text command give the x and y component of the text, and the third entry is the text.

The next three lines are used to define the domain and range of the graph. The x-values have already been specified, so only the y-values need to be added. Following this section are five lines specifying the font type and size for the title, x and y-axes, and the size of the numbers on the axes. Finally, there are two lines for outputting the graph as EPS or JPEG files. The latter would be the most common used in a Word document or webpage. Note that the default file saving in MATLAB is a .fig file, and the user can edit this file in the MATLAB figure window. Below is the graph output from the *M-file* given above.

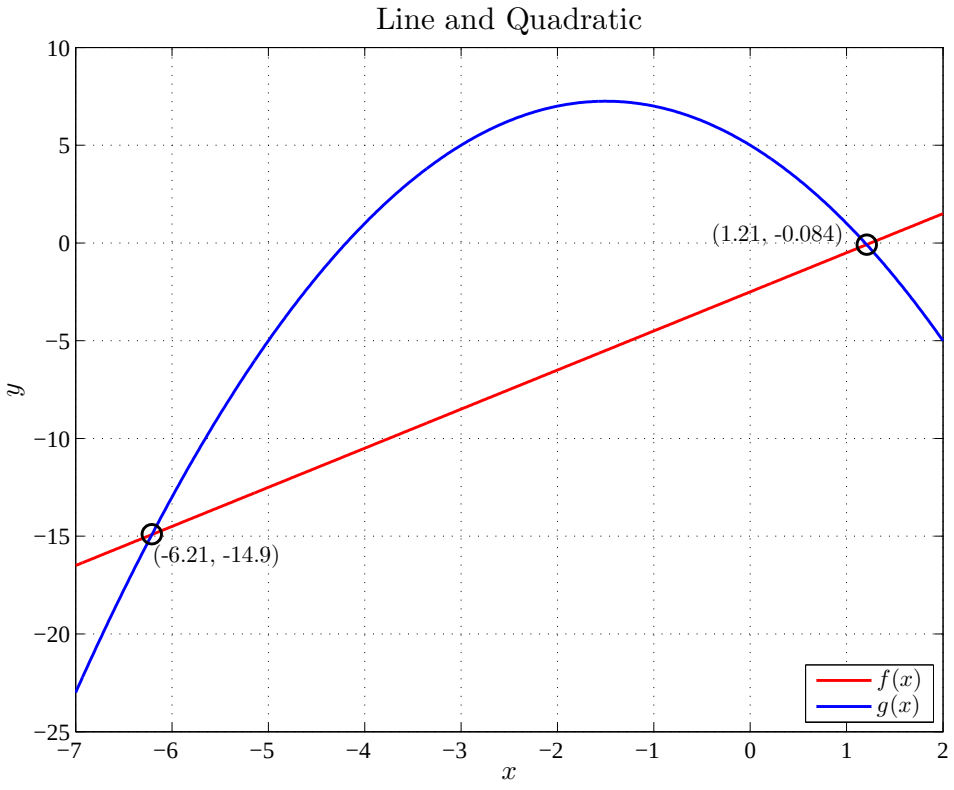

Figure 2.1: Graph of a line and quadratic from MATLAB.

MATLAB is software that produces high quality graphs in a variety of formats. The graphing information above allows one to create a basic graph in MATLAB. It should be readily apparent that the user has many options already to control the appearance of the graph. As this manual progresses, more commands are given to show the user how to add additional features to the graph

for better presentation. Also, as the user becomes more comfortable with MATLAB, personal preferences for making a graph can be adopted, including simple commands to create a quick graph in the *Command* window.

2 BASIC FUNCTIONS AND GRAPHING IN MAPLE

This section introduces a few basic commands in Maple, which allows entering functions and creating graphs. Maple is a very powerful Symbolic Algebra package with extensive abilities to handle complex mathematical problems symbolically. This lab manual is meant only as a brief introduction to Maple with the idea that users will become comfortable with basic commands and be able to solve increasingly complex problems in the future. This section shows how to enter functions and use these functions to solve equations and graph simple graphs.

Maple is an interactive line command language. Students may want to think of Maple as a very smart graphing calculator in early use. The syntax of some basic commands are shown below to make operation easy; however, note that Maple does have excellent *Help* available. The reader may want to delve deeper into what Maple can do by following the *Take a Tour of Maple* under the *Help* tab.

In this section, the problem given by Eqns. (2.1) is examined with Maple. The functions are input into Maple by typing the following commands:

```
f := x -> 2*x - 2.5
g := x -> 5 - 3*x - x^2
```

Maple will automatically modify these to appear as

$f := x \to 2 \cdot x - 2.5$

$g := x \to 5 - 3 \cdot x - x^2$

The functions $f(x)$ and $g(x)$ behave like functions, so by typing $g(2)$, Maple gives the answer -5.

The roots of $g(x)$ are found by the command:

```
solve(g(x) = 0,x)
```

which gives the answer

$$-\frac{3}{2} - \frac{1}{2}\sqrt{29}, -\frac{3}{2} + \frac{1}{2}\sqrt{29}.$$

To get a decimal answer one replaces *solve* with *fsolve*. The x-values of intersection are found and stored in a vector variable $xsoln$ by typing

```
xsoln := fsolve(f(x) = g(x),x)
```

Maple responds with the answers 1.208099244 and -6.208099244. The y-values are obtained by typing

```
f(xsoln[1]); g(xsoln[2])
```

The points of intersection found are

$$(x_1, y_1) = (-6.2081, -14.916) \qquad \text{and} \qquad (x_2, y_2) = (1.2081, -0.0838015).$$

The points above indicate a good range to graph. Maple has a very easy plot command to plot functions, and it is given by:

```
plot({f(x), g(x)}, x = -7 .. 2)
```

This command says to plot the set of functions $f(x)$ and $g(x)$ on the interval $x \in [-7, 2]$. Below is the plot that Maple produces. By right clicking on the graph, a menu appears that allows the user to edit the graph. For example, a title or axis label can be added, and the graph can be exported in a number of formats for use in a paper. Maple graphs are much harder to get the quality of output that is available with MATLAB, but the command above shows how quickly a sketch of a graph can be made in Maple. In this mode Maple acts as a very smart graphing calculator.

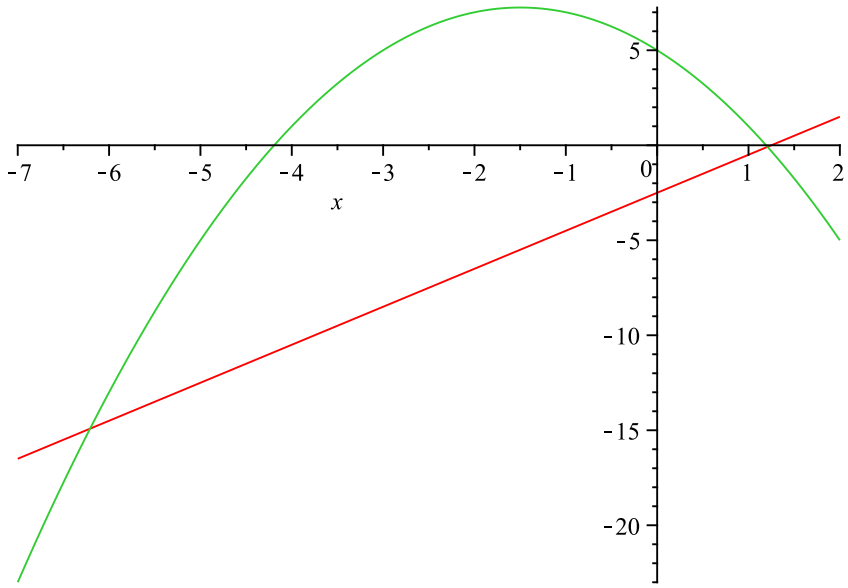

Figure 2.2: Graph of a line and quadratic from Maple.

3 PROBLEMS

Below is the first problem of the Lab Manual. The problems have been created from a collection of WeBWorK problems, and the WeBWorK file name is provided, though some modifications are necessary. This Lab Manual includes blanks corresponding to the blanks in the WeBWorK problems, and every problem has a graphing portion that is meant to be written into a Lab Report. Most problems have an interesting biological example, and a complete list of the problems is provided in an appendix at the end of the Lab Manual with a brief summary of the problem.

Additional MATLAB Commands

The problems below allow the students to either use their algebra skills or employ the symbolic tools of MATLAB or Maple. The graphing needs easy modifications of the MATLAB code provided above. To obtain a square graphing area in MATLAB (first problem), the user types:

```
axis square;
```

below the other axis command. If more digits are needed from the MATLAB output, then one can type

```
format long
```

Other MATLAB outputs can be found using *help format* in MATLAB. These problems should provide a good introduction to MATLAB and Maple.

Problem 2.1 (Two Lines)**.** *(Lab121_A1_lin_quad_graph.pg) This problem introduces graphing methods with a review of the algebra for two intersecting lines.*

a. You are given the following two lines:

$$y = 5x - 2 \qquad \text{and} \qquad y = -\frac{x}{2} + 3.$$

Find the slopes and the x and y-intercepts of these lines.

First Line:

Slope = _____ x-intercept = _____ y-intercept = _____

Second Line:

Slope = _____ x-intercept = _____ y-intercept = _____

Are these lines perpendicular? (Yes or No)

Find the point of intersection. $x =$ _____ $y =$ _____

b. In the Lab Report, create a graph with the two lines given above. Your graph needs to include a title and have the x and y axes labeled correctly. Adjust the scale of the graph so that the domain x and range y both go from -10 to 10, then expand the graph into a square. Label each line with its equation. Also, label the coordinates of the point of intersection on the graph. Write a sentence explaining how you determine if the lines are perpendicular or not. Write a short paragraph detailing how the point of intersection is found. Explain the process, not using the specific numbers for this problem.

Problem 2.2 (Line and Quadratic)**.** *(Lab121_A2_quad_line.pg) This problem studies the intersection of a line and a quadratic function.*

a. Consider the following functions, $f(x)$ and $g(x)$:

$$f(x) = 2 - 2.5x \quad \text{and} \quad g(x) = x^2 + 3x - 3.$$

Beginning with the line, find the x and y-intercepts. Also, determine the slope of the line.

x-intercept = _____ y-intercept = _____ Slope = _____

For the parabola, find the x and y-intercepts. Also, determine the coordinates of the vertex.

x-intercepts = _____ and _____ y-intercept = _____

Vertex = (_____, _____).

Find the points of intersection of the two curves.

First point of intersection = (_____, _____).

Second point of intersection = (_____, _____).

b. In the Lab Report, create a graph with the two functions given above. Be sure that the domain for the graph includes all intercepts, points of intersection, and the vertex of the parabola. Do NOT label these points, but do be sure to label the respective curves with an appropriate identifying label. Label the axes and give a title to the graph. Write a brief paragraph about how to find the points of intersection. Explain the process, not using the specific numbers for the problem.

Problem 2.3 (Line, Quadratic, and Cubic)**.** *(Lab121_C1_lin_quad_cub.pg) This problem reviews a quadratic function and studies the intersection of a line and a cubic function.*

Maple is a powerful symbolic algebra computer package and is capable of solving cubic equations, which are extremely difficult to do by hand (so are rarely taught).

a. Consider the quadratic function:

$$g(x) = x^2 + 2x - 8.$$

Find the x and y-intercepts. Also, determine the coordinates of the vertex.

x-intercepts = _____ and _____ y-intercept = _____

Vertex = (_____, _____).

b. In the Lab Report, create a graph of the parabola given above. Be sure that the domain for the graph shows all intercepts and the vertex.

c. Now consider these functions, $f(x)$ and $g(x)$.

$$f(x) = -2x - 3.1 \qquad \text{and} \qquad g(x) = 1.2 + 2.8x - 2.5x^2 - x^3.$$

Find the x and y-intercepts and the slope for the line.

x-intercept = _____ y-intercept = _____ Slope = _____

For the cubic equation, find the x and y-intercepts.

x-intercepts = _____, _____, and _____ y-intercept = _____

Find the points of intersection between $f(x)$ and $g(x)$.

Points of intersection = (_____, _____), (_____, _____), and (_____, _____).

d. In the Lab Report, graph this line and cubic function. Choose a domain such that the graph clearly shows all the points identified above.

CHAPTER 3:
FUNCTION MODELS

Keywords: Linear model, linear least squares, exponential, logarithm, power law, allometric

This section contains the laboratory problems from WeBWorK that use basic functions for modeling. The first problems use the simplest of models, the linear models of the form:

$$y = ax + b,$$

with slope, a, and y-intercept, b. The model is found using a *linear least squares* best fit to the data, which can be found readily from web sources. This manual shows how computer programs can be applied to find this best fitting model.

1 ERROR ANALYSIS

When discussing models, it is very important to know how well a particular model represents the data that is being analyzed. There are a number of ways to measure how well data fit a model. This section presents important measures of error to compare models.

Least Squares Error

The most common unbiased error used to fit a model to data is the *least squares error*. Consider a set of n data points: $(x_1, y_1), (x_2, y_2), ...(x_i, y_i), ..., (x_n, y_n)$, where (x_i, y_i) is the i^{th} data point. A linear least squares best fit to these data considers a linear model

$$y(x) = ax + b,$$

where the slope, a, and the intercept, b, are chosen to best fit the data in some sense.

The least squares best fit minimizes the square of the error in the distance between the y_i values of the data points and the $y(x_i)$ value of the line, which depends on the selection of the slope, a, and the intercept, b. Fig. 3.1 provides an illustration of this linear model fitting data. Note the the least squares best fit method can be generalized to other functions with $y(x_i) = f(x_i)$, but fitting the parameters in $f(x)$ is more difficult for nonlinear functions.

Define the *signed error* or *actual error* between each of the data points and the line as

$$e_i = y_i - y(x_i) = y_i - (ax_i + b), i = 1, ...n,$$

and the *absolute error* between each of the data points and the line as

$$|e_i| = |y_i - y(x_i)| = |y_i - (ax_i + b)|, i = 1, ...n.$$

We can see that e_i varies as a and b vary.

Fig. 3.1 displays the graph showing these error measurements.

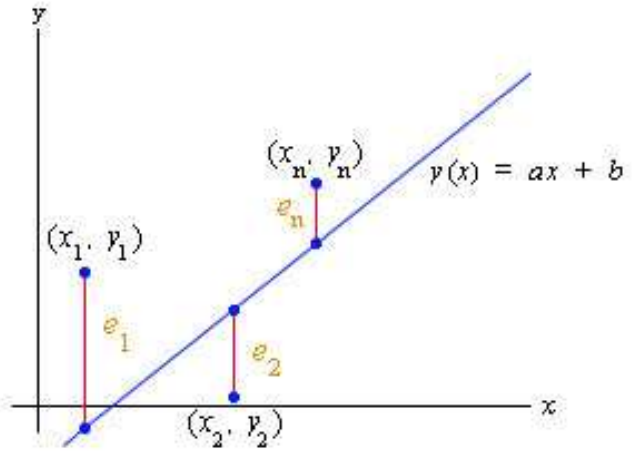

Figure 3.1: Graphic representation of the error measurements for the Least Squares Best Fit method.

The *least squares best fit* finds the minimum value of the function

$$J(a,b) = e_1^2 + e_2^2 + ... + e_n^2 = \sum_{i=1}^{n} e_i^2,$$

as the slope, a, and intercept, b, vary.

The function $J(a,b)$ sums all the squared errors between the linear model and the data points. The slope, a, and the intercept, b, of the linear model are varied until the function, $J(a,b)$, achieves its smallest value. The unique line that produces this minimum $J(a,b)$ is the least squares best fit or best linear model for the given data set. There are formulas for the best values of a and b, which are given by:

General Linear Least Squares Best Fit Method

Define the mean of the x values of the data points as

$$\bar{x} = \frac{x_1 + x_2 + ... + x_i + ... + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

The value for the slope of the line that best fits the data is given by

$$a = \frac{\sum_{i=1}^{n}(x_i - \bar{x})y_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2}.$$

With the slope computed, the intercept is found from the formula

$$b = \frac{1}{n}\sum_{i=1}^{n} y_i - a\bar{x} = \bar{y} - a\bar{x}.$$

These formulas are available in many computer programs and calculators, so the user only needs the basic program commands to obtain the best fitting slope and intercept for a linear model through a set of data.

Error Computations

There are several standard computations of error, *actual*, *absolute*, *relative*, and *percent* errors. The actual and absolute error between a data point and a model were defined above. The most common computations of error are *relative error* and *percent error*. Let X_e be an *experimental measurement* or the measurement, which is considered the *worst value*, and X_t be the *theoretical value* or the measurement, which is considered the *best value*. The problems in this lab manual will usually let the reader know which value is considered *best* or *worst*. Often the models are being tested, so the model will be considered the *worst value*.

The *relative error* is the difference between the experimental value and the theoretical value **divided** by the theoretical value, so

$$\text{Relative Error} = \frac{X_e - X_t}{X_t}.$$

The *percent error* is closely related to the *relative error*, except that the value is multiplied by 100% to change the fractional value to a percent, so

$$\text{Percent Error} = \frac{X_e - X_t}{X_t} \times 100\%.$$

Another way to think of the relative error is that it is the difference between the least accurate and most accurate values divided by the most accurate value:

$$\text{Percent Error} = \frac{Worst\ Value - Best\ Value}{Best\ Value} \times 100\%.$$

We generally divide by the most accurate value, which is usually the one provided by the data tables when comparing models.

2 LINEAR MODELS

This laboratory manual is designed to introduce students to basic modeling from a variety of examples and apply the computer programs MATLAB and Maple to graph and solve computational aspects of the problems. The simplest model is the *linear model*. The examples and problems in this section introduce linear models and show how computer programs can produce graphs for these models, showing the data and the model. The programs use the least squares best fit analysis described above. The techniques are demonstrated with an example of growing children and the Lambert-Beer law. Three problems are presented on crickets chirping, urea excretion of hummingbirds, and growth of yeast.

Example 3.1 (Juvenile Height)**.** *The rate of growth in height of children is almost linear for a range of ages.*

Age	1	3	5	7	9	11	13
Height (cm)	75	92	108	121	130	142	155

Table 3.1: The average juvenile height of American children as a function of age [6].

Table 3.1 gives data for the height of the average American child with respect to age [6]. This example shows how to use MATLAB and Maple to find the best fitting linear model and produce a graph. The model is given by

$$h(a) = ma + b,$$

where the best slope, m, and intercept, b are found by our programs. The least square error analysis showed the formulas for finding these values, which the reader may want to verify by hand.

MATLAB for Linear Least Squares Fit

The graphing template presented in the Graphing basics section is easily modified to find the best fitting linear model and produce a good graph. The key new MATLAB command is *polyfit(x,y,n)*, which finds the best fitting polynomial of order n through the data points (x, y). The user clearly changes the titles, labels, and x and y domains to match the problem of interest. The actual best linear fit comes from the following MATLAB lines:

```
ad = [1 3 5 7 9 11 13];              % Age data
hd = [75 92 108 121 130 142 155];    % Height data
p = polyfit(ad,hd,1)                 % Linear fit to data
h1 = p(1)*xx + p(2);                 % Equation of Line
```

which are inserted after the variable xx is defined. By not putting the semicolon on the line with the polyfit command, the user sees the values of the coefficients, stored in the variable p, output in the Command Window of MATLAB:

$$p = [6.4643 \quad 72.3214].$$

These values are used in the text window, where the equation is put on the graph. The plot commands are very similar to before:

```
plot(xx,h1,'r-','LineWidth',1.5);   % Red Linear Model
hold on                             % Plots Multiple graphs
plot(ad,hd,'bo','MarkerSize',10);   % Blue Data points
grid                                % Adds Gridlines
text(5,85,'$$h(a) = 6.46 a + 72.3$$','FontSize',12); % Adds Formula
```

The resulting graph is shown below on the left in Fig. 3.2.

MATLAB's ability to manage vector quantities allows one to easily construct a function to compute the sum of square errors between the model and the data. In MATLAB a function is created by selecting *New*, then *Function*. Below is a MATLAB function for computing the sum of square errors between a linear model with slope and intercept in the parameter p and the data:

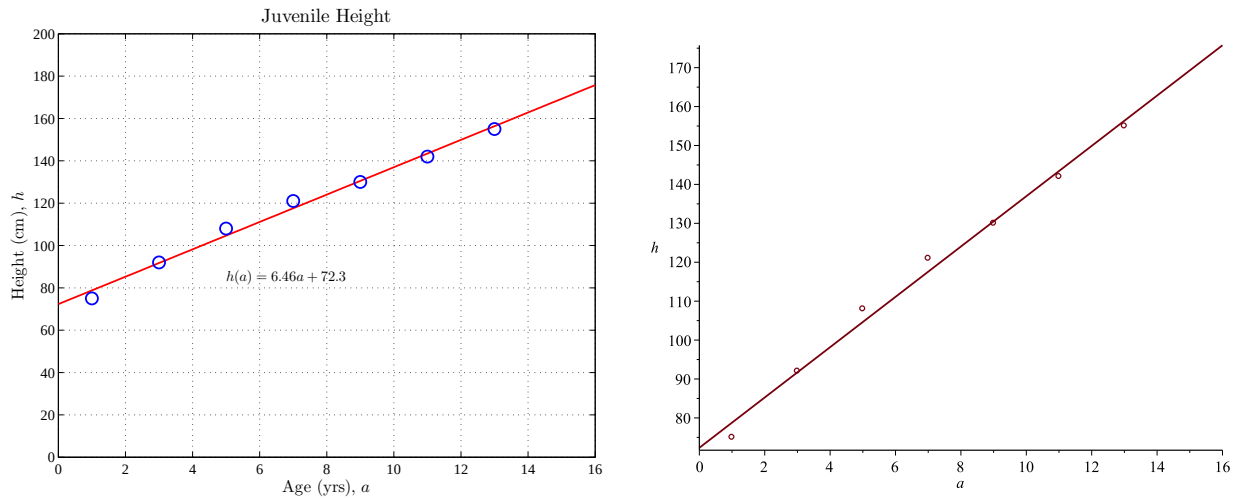

Figure 3.2: Graphs of child height data and the least squares best fit linear model from MATLAB (left) and Maple (right).

```
function J = sumsq_lin(p,xdata,ydata)
% Function to compute the least squares error for linear model
model = p(1)*xdata + p(2);  % Linear model using parameter p
error = model - ydata;      % Error between model and data
J = error*error';           % Computes sum of square error
end
```

By letting $p = [6.4643, 72.3214]$ (with the data still defined as ad and hd), then the command:

```
J = sumsq_lin(p,ad,hd)
```

yields the answer

$$J = 41.5714,$$

which is the least sum of square errors between these data and the best fitting linear model.

This example gives interesting information about child growth based on the linear model. The slope gives a good estimate of how much a child grows each year. The slope indicates that between 1 and 13 years of age, the average child grows about 6.5 cm/yr. The h-intercept is 72 cm, which gives a rough estimate of the length of a newborn child. However, the h-intercept is outside the range of the data, so it provides a poor estimate for birth length.

Maple for Linear Least Squares Fit

Maple can be used to find the linear least squares best fitting model to the juvenile height example above. Maple is organized into distinct packages to allow more efficient computing. The *polynomial fit* program is in the *Statistics* package, while techniques to graph both points and lines uses the *plots* package.

To find the best linear least squares model to the juvenile height data, the following Maple commands are issued:

```
with(Statistics):                    # Invoke Statistics package
ad := Vector([1, 3, 5, 7, 9, 11, 13]):             # Age data
hd := Vector([75, 92, 108, 121, 130, 142, 155]):   # Height data
p := PolynomialFit(1, ad, hd)      # Store Coefficients for Best Fit in p
```

The coefficients found with Maple are clearly going to have the same values as the ones found with MATLAB. The graph should have data plotted as points and the linear model as a continuous curve. To have two different types of curves, the following Maple commands are issued:

```
with(plots):                          # Invoke plots package
P1 := plot(ad, hd, style = point, symbol = circle):   # Plots the data
P2 := plot(p[2]*a+p[1], a = 0 .. 16, h):              # Plots the best model
display({P1, P2})                     # Displays both the data and model
```

Note that the hashtag, #, is used for Maple comments. For plotting graphs in Maple, one usually invokes the plot command to observe the graph, then stores the plot in a variable, which in this case is P1 and P2. Note that whenever a colon ends a line in Maple, then Maple suppresses the output. The resulting graph from the commands above is displayed on the right in Fig. 3.2.

The problems below provide applications of the *linear least squares* best fit to some biological data. These problems allow the student to practice the techniques given above in the juvenile height example. Fitting linear models is very common in scientific studies, so these techniques are important to learn.

3 DATA AND MATLAB

This lab manual includes many data sets, which will be modeled in various ways. It is important to enter the data sets accurately into MATLAB for analysis and creation of mathematical models. To accomplish this task, the reader needs to become familiar with some of MATLAB's matrix features. Below is an example of some data on crickets chirping from Bessey and Bessey [1].

Example 3.2 (Cricket Data)**.** *Experiments have shown that the rate at which a cricket chirps is approximately a linear relationship with the temperature. This example takes a set of cricket data and graphs it with the best linear model fitting the data.*

The Bessey brothers collected a very large set of data of snowy tree crickets (*Oecanthulus fultoni*) from Lincoln, NE in 1897. They carefully measured the temperature where the crickets were chirping and the rate of chirps/min. Below is a table of a small set of their measurements.

The first step in the process is entering the data into our MATLAB workspace. There are several ways to enter the data. Sometimes it is best to use Excel spreadsheets or a simple text editor, like NotePad, for easy data manipulation. However, this manual will describe how to process the data directly into MATLAB. This Lab Manual is in a PDF format, so one can simply copy the data and enter it into a variable in MATLAB.

We highlight the **numbers** in the table and copy them (**Ctrl C**). In MATLAB we create a variable *data*. We type "data = [" and paste the data (**Ctrl V**). Finally, end with a closing bracket "]" and **Enter** to complete the variable *data* as a stored *matrix of information*. In this case, the matrix has 6 rows and 6 columns. (Alternately, one can type in the data or place it in other

chirps/min	°F	chirps/min	°F	chirps/min	°F
77	54.2	123	67.5	173	76.3
85	58.5	136	69.2	168	77.2
95	59.5	146	71	162	78.2
101	61.5	144	73.5	183	79
112	64.8	158	75	208	80.9
122	66.5	172	75	195	83

Table 3.2: A collection of data from Bessey and Bessey [1] on cricket chirping rate as a function of temperature (°F).

software, like an Excel spreadsheet, to have it in the form we eventually want of two vectors of the c and T data.) (**Note**: If the data has some missing entries, then these entries must be filled with some digit to complete the matrix and later removed before the data is analyzed.)

Assume that the 6×6 matrix of data is stored in the variable *data*. Below are some matrix manipulation commands to extract the two data vectors *cd* and Td that will be needed for graphing and analyzing the data. MATLAB uses standard *matrix notation*, so $data(i, j)$ refers to the element in row i and column j. To create the *cd* vector from *data*, we note that these data appear in the 1^{st}, 3^{rd}, and 5^{th} columns. Since all of the rows are desired, they are given by a colon representing the wild card. The appropriate columns are designated by their numbers. The variable $data(:, 1)'$ takes the first column of data, then transposes it into a row vector. Note that the $'$ is MATLAB's way of transposing a matrix to get the columns into row vectors. This vector is concatenated onto the other two row vectors $data(:, 3)'$ and $data(:, 5)'$ to produce the complete vector of chirps/min, *cd*. It follows that the following MATLAB commands produce our data vectors, *cd* and Td:

```
cd = [data(:,1)',data(:,3)',data(:,5)']
Td = [data(:,2)',data(:,4)',data(:,6)']
```

The MATLAB command

```
p = polyfit(cd,Td,1)
```

produces coefficients $p = [0.2136, 40.2221]$ for the best fitting line

$$T = 0.2136\, c + 40.22.$$

With this information the graph of the data and the model is easily formed.

```
c1 = 0;          % Left x-value
c2 = 250;        % Right x-value
Npts = 50;       % Number of points plotted between x1 and x2
cc = linspace(c1,c2,Npts); % Creates vector of x-values
TT = 0.2136*cc + 40.22;     % Creates linear model
plot(cc,TT,'r-','LineWidth',1.5);  % Plots the linear model
hold on
plot(cd,Td,'bo','MarkerSize',6);   % Plots the data
grid
```

```
ax=[0 250 50 100];
```

The resulting graph is shown in Fig. 3.3.

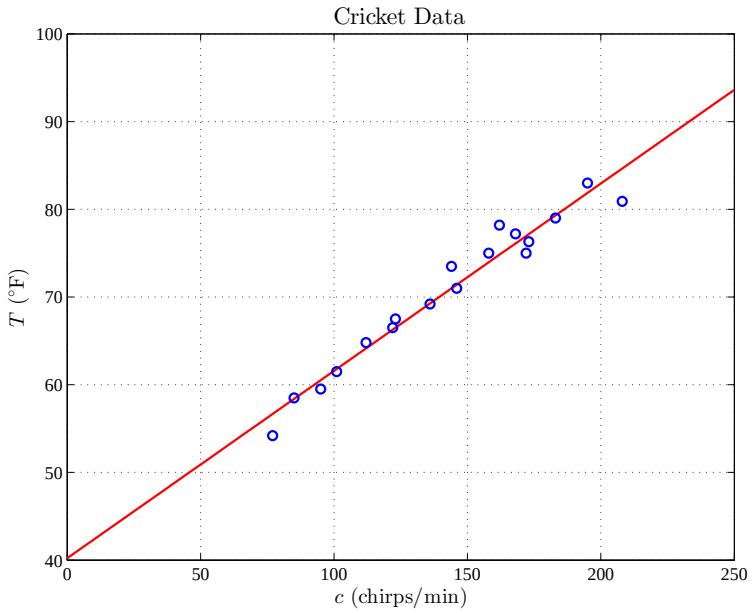

Figure 3.3: Graph showing the data and linear model for crickets chirping as temperature varies.

If we want to save all this information for later use, then we type

```
save 'cricket'
```

This creates the file `cricket.mat`, which has all the variables, like $data$, cd, Td, p, $c1$, $c2$, $Npts$, cc, and TT. If we later return to MATLAB, directing MATLAB to the directory where `cricket.mat` is stored, then typing

```
load 'cricket'
```

restores all the saved variables to the workspace. Typing `who` gives a list of the variables.

Problem 3.1 (Cricket Thermometer)**.** *(Lab121_A3_cricket.pg) Data indicate that the rate at which crickets are chirping is a linear relationship with temperature.*

For many years people have recognized a relationship between the temperature and the rate at which crickets are chirping. The folk method of determining the temperature in degrees Fahrenheit is to count the number of chirps in a minute and divide by 4, then add 40. In 1897, A. E. Dolbear [5] noted that "crickets in a field [chirp] synchronously, keeping time as if led by the wand of a conductor." In his paper, he appears to be the first person to write down a formula in a scientific publication giving a linear relationship for the temperature based on the chirp rate of crickets.

a. The classic folk "cricket thermometer," formalized by Dolbear, satisfies the linear relationship:

$$T = \frac{N}{4} + 40,$$

where T and N were the temperature and the number of chirps/minute, respectively. Determine the slope and T-intercept for this line.

Slope = _____ T-intercept = _____

The Bessey brothers [1] later made careful measurements and did a linear least squares best fit to their data and obtained the linear relationship

$$T = 0.213\,N + 40.4.$$

Determine the slope and T-intercept for this line.

Slope = _____ T-intercept = _____

b. Below are recordings of four crickets chirping at different temperatures. In this question, you time the number of chirps/minute of four crickets. By clicking on the different sound tracks for the crickets, the hyperlink goes to a recording for each of the four crickets.

Sound for Cricket A **Sound for Cricket B**

Sound for Cricket C **Sound for Cricket D**

(These sounds can be obtained through the hyperlinks:
http://www-rohan.sdsu.edu/∼jmahaffy/courses/s00a/math121/labs/laba/X
where
X = **w585_21.html** X = **w585_5.html** X = **w585_25.html** X = **w585_30.html**)

Below are a list of chirping rates for crickets, and you select the answer that is closest to the chirping rate for the sounds that you measured above.

_____ Cricket A	A. 221 chirps/min
	B. 69 chirps/min
_____ Cricket B	C. 99 chirps/min
	D. 39 chirps/min
_____ Cricket C	E. 131 chirps/min
	F. 276 chirps/min
_____ Cricket D	G. 161 chirps/min
	H. 199 chirps/min

c. In the Lab Report, create a graph of each of the models (one graph with both models). Show clearly the data points that are gathered in Part b. Write a short paragraph on how the data was collected to find the correct answers.

d. The cricket thermometer model is a linear model. Each of the variables has physical units. For example, T has units °F. Use the units of °F, chirps, and min.

What are the units for the coefficient representing the slope? _____

What are the units for the coefficient representing the T-intercept? _____

e. Errors are introduced when counting the number of chirps/min. Suppose that the error for counting the chirps/min for Cricket B is ±5 chirps/min.

Find the range of temperatures predicted by the Dolbear model

_____ $\leq T \leq$ _____

Also, find the range of temperatures predicted by the Bessey model

_____ $\leq T \leq$ _____

Suppose that the error for counting the chirps/min for Cricket D is ±10 chirps/min.

Find the range of temperatures predicted by the Dolbear model

_____ $\leq T \leq$ _____

Also, find the range of temperatures predicted by the Bessey model

_____ $\leq T \leq$ _____

f. Below is a table of a subset of the Bessey brother data [1].

chirps/min	°F	chirps/min	°F
77	54.2	145	73.5
85	58.5	150	74.8
95	59.5	158	75
105	61	174	76.3
101	61.5	168	77.2
111	65.5	189	78
123	67.5	162	78.2
128	68	183	79
132	69	191	80.9
141	69.5	198	82.5

Find the equation of the least squares best fit line through these data.

$T =$ _____

Find the percent error between slope and T-intercept of this least squares best fit line and the Bessey model given in Part a. (Assume the model in Part a is the better value.)

Percent Error in slope = _____% Percent Error in T-intercept = _____%

g. In the Lab Report, create a graph showing the data above with its least squares best fit line. On the graph include both the Dolbear and Bessey brother models from Part a. Write a brief paragraph discussing the accuracies of the models from the lab experience, what are the major sources of error (list at least two). How much agreement is there between the different models have in predicting the temperature.

Problem 3.2 (Hummingbirds and Urea)**.** *(Lab121_B2_absorb.pg) Spectrophotometry is used to quantitatively measure urea concentration, and this quantitative analysis gives information about animal excretion patterns.*

This problem examines some physiological data from the laboratory of Professor Carol Beuchat at San Diego State University [2]. Animals have evolved different mechanisms for excreting waste nitrogen. The principle means of excreting nitrogen are uric acid, urea, and ammonia. Unfortunately, the latter two are toxic so require larger volumes of water for excretion. Uric acid uses less water, but it requires more energy (ATP) to produce. Thus, animals must weigh their needs of water versus energy when selecting a means of excretion.

a. Here we only examine the amount of urea excreted. First, a standard is run to determine the absorbance at 570 nm as a function of the concentration of urea. (This is a standard technique using spectrophotometry.) The data are listed below:

Urea (mg/dl)	Absorbance	Urea (mg/dl)	Absorbance
2	0.139	70	2.17
10	0.3	100	3.117
20	0.624	130	3.826
40	1.236	150	4.545
50	1.51	200	6.149

Find the best straight line through the data, where

$$A = mu + b,$$

is the straight line describing absorbance, A, as a function of the concentration of urea, u, determining the slope, m, and intercept, b.

$m =$ _____ $b =$ _____

Find the expected absorbance for a sample containing 59 mg/dl of urea

$A =$ _____

and a sample containing 184 mg/dl of urea.

$A =$ _____

b. In the Lab Report, create a graph showing both the data and this linear model. Describe how well the linear model fits the data.

c. Use the formula for the best fitting line in Part a to find the predicted value of absorbance, A, for the concentrations of urea, u, at 50, 100, and 200.

$A(50) =$ _____ $A(100) =$ _____ $A(200) =$ _____

Determine the absolute and percent errors for the concentrations of urea, u, at 50, 100, and 200. In this case, assume that the absorbance given by the best fitting line is the best value.

When $u = 50$, the absolute error = _____ and percent error = _____%

When $u = 100$, the absolute error = _____ and percent error = _____%

When $u = 200$, the absolute error = _____ and percent error = _____%

Compute the sum of square errors, $J(m, b)$, that least squares function minimized to obtain the best fitting line to the data.

$J(m, b) =$ _____

d. In practice, one uses the spectrograph to measure the absorbance, and use the relationship between the two to calculate the urea levels. In order to do this you must now solve for u as a function of A, the inverse function. That is find a function $f(A)$ such that

$$u = f(A)$$

$u =$ _____$A+$ _____

e. In Professor Beuchat's laboratory they found that the urine from a hummingbird kept at 10°C had an absorbance of 0.138.

Concentration of urea = _____mg/dl.

When the hummingbird was kept at 20°C, the absorbance for a urine sample was 0.206.

Concentration of urea = _____mg/dl.

When the hummingbird was kept at 40°C the absorbance was 0.29.

Concentration of urea = _____mg/dl.

f. In the Lab Report, give an explanation of these results with regard to either energy or water conservation by these hummingbirds?

g. We would like to see if the analysis of urine samples tells us about other species. The table below lists different animals and the corresponding absorbances measured.

Animal	Absorbance
Chicken	3.145
Duck (Fresh Water)	0.442
Duck (Salt Water)	0.768
Frog	0.266
Turtle	1.082
Tortoise	6.908

Concentration of urea for the chicken = _____ mg/dl.

Concentration of urea for the duck (fresh water) = _____ mg/dl.

Concentration of urea for the duck (salt water) = _____ mg/dl.

Concentration of urea for the frog = _____ mg/dl.

Concentration of urea for the turtle = _____ mg/dl.

Concentration of urea for the tortoise = _____ mg/dl.

A hummingbird gets its energy from nectar, which is high in water content.
Which animal has the excretion concentration most similar to a hummingbird? _____

h. In the Lab Report, write a paragraph discussing why the one animal's excretion pattern is similar to a hummingbird's excretion. Also discuss why you would see the different concentrations of urea for the different animals in the table above. Is there a pattern between the different animals here, and can you offer some explanations?

Additional Maple Commands for Function Fitting

There are examples where a general polynomial fit is not appropriate. The most common case occurs when the model must necessarily pass through the origin. This removes the constant term from a polynomial. Other cases include situations when there is special symmetry, so the polynomial has only even or odd terms. These cases can be programmed into MATLAB, but MATLAB does not have a special function for doing this now. (There is a program, *polyfitZero*, available to download, which does this.) Maple has the function *LinearFit* in its Statistics package, which can easily find these special forms of a polynomial. The *LinearFit* function takes the form:

```
LinearFit(flst,X,Y,v)
```

where *flst* is the list of functions being used, X and Y are the x and y data entries, and v is the variable in the functions.

Example 3.3 (Lambert-Beer Law)**.** *The Lambert-Beer law for absorbance of light by a spectrophotometer is a linear relationship, which can have the form*

$$c = mA,$$

where c is the concentration of the sample, A is absorbance, and m is the slope that must be determined from experiments.

The ion dichromate forms an orange/yellow solution, which has a maximum absorbance at 350 nm and is often used in oxidation/reduction reactions. Below is a table of data from known samples.

A	0.12	0.32	0.50	0.66
c (mM)	0.05	0.14	0.21	0.30

Table 3.3: Absorbance, A, readings from samples of known concentrations, c of dichromate.

Use the data in Table 3.3 to find the best fit for parameter m in the Lambert-Beer law, then determine the concentration of dichromate in an unknown sample with $A = 0.45$.

Solution: A solution to this problem is found with the following set of Maple commands:

```
with(Statistics):
X := Vector([0.12,0.32,0.50,0.66]):
Y := Vector([0.05,0.14,0.21,0.30]):
LinearFit([A], X, Y, A)
```

These commands open the Maple Statistics package, store the data in the vectors X and Y, and apply a model with only the linear term A. The resulting best model given by Maple is

$$c = 0.440927218344965\, A.$$

It follows that an unknown sample with $A = 0.45$ gives $c = 0.1984172482$.

Since the Lambert-Beer law has a simple form for the linear function, the sum of square errors is easily written from the data.

$$J(m) = (0.05 - 0.12m)^2 + (0.14 - 0.32m)^2 + (0.21 - 0.5m)^2 + (0.30 - 0.66m)^2,$$

which can be easily input into Maple and simplified with the following commands:

```
J := m -> (0.05-0.12*m)^2+(0.14-0.32*m)^2+(0.21-0.5*m)^2+(0.30-0.66*m)^2
simplify(J(m))
```

producing

$$J(m) = 0.1562 - 0.7076\, m + 0.8024\, m^2.$$

Maple's symbolic algebra capability reduces the quadratic function to standard form for easy evaluation.

Problem 3.3 (Yeast Growth)**.** *(Lab121_C3_yeast_growth.pg) The growth rate of yeast at low densities is linear, which satisfies the conditions for Malthusian growth.*

In 1913, Carlson[3] studied the growth of a culture of yeast, *Saccharomyces cerevisiae*. Over time this culture levels off, but its initial growth is exponential or Malthusian. A Malthusian growth model is given by

$$P_{n+1} = P_n + rP_n.$$

Simply put, the population at the next time $(n+1)$ is equal to the population at the current time (P_n) plus some growth term, which is simply proportional (r) to the current population. Thus, we have a growth function

$$g(P) = rP.$$

Below is a table from Carlson's data showing the population, P, and the rate of growth, $g(P)$, at that particular population

Population	Growth/hr
18,600	10,800
29,500	18,200
48,200	24,600
71,800	48,100

a. Use the data above to find the best straight line passing through the origin. (This is a special case of the linear least squares best fit, with the line having a slope and a zero intercept.) What is the slope of the line that best fits through the data?

Slope $r =$ _____

Find the sum of square errors with this model and the data.

Sum of Square Errors = _____

b. In the Lab Report, create a graph of the data and the best fitting linear model. Briefly describe how well the line fits the data.

c. For the linear model (passing through the origin) given above, it is easy to find the sum of squares function. Consider a data point $(P_i, g(P_i))$. The absolute error between this data point and our model is given by

$$e_i = |g(P_i) - rP_i|.$$

Thus, $e_1 = |10800 - 18600r|$. Similarly, you can find e_2, e_3, and e_4. The sum of squares function is given by

$$J(r) = e_1^2 + e_2^2 + e_3^2 + e_4^2.$$

Find the expression for the quadratic function of the slope of the model, r (in simplest form).

$$J(r) =_____\ r^2 +_____\ r +_____\ .$$

Find the coordinates of the vertex.

$$(r_v, J(r_v)) = (_____,_____)$$

d. In the Lab Report, create a graph of $J(r)$ for $r \in [0.3, 0.8]$. Compare the value of the vertex of the parabola, r_v, and the slope of the best fitting line through the origin found by Maple's LinearFit. Also, compare the sum of square errors and the value of $J(r_v)$.

e. From the best model, find the growth for a population of 100,000 yeast.

Growth for 100,000 yeast $g(100,000) =$ _____

Determine the population of another culture of yeast, given that their growth rate is measured to be 75,000 yeast/hour.

For $g(P) = 75,000$, $P =$ _____

4 EXPONENTIAL AND LOGARITHMIC FUNCTIONS

The exponential and logarithmic functions are very important in a number of modeling applications. In Calculus it is natural to use the exponential function to the base e. Similarly, this is the natural logarithmic base. Most computer languages also rely on e as the natural base. For computer languages we generally write e^x in the following manner:

```
exp(x)
```

Thus, in either Maple or MATLAB, we have the following:

```
exp(1)
        2.718281828
```

For Maple, we need to type `evalf(exp(1))` or `exp(1.)` to obtain the decimal value. The Maple command `evalf(·)` is used to force Maple to produce decimal answers. Note that neither language recognizes e as a character alone.

For the natural logarithm, one generally writes $\log(x)$ in a computer language. However, many computer languages recognize the form $\ln(x)$, also. Maple recognizes both forms, while MATLAB only recognizes `log(x)`. Thus, to obtain $\ln(2)$, we type

```
log(2.)
        0.6931471806
```

Below is a problem to explore the asymptotic size of exponential and logarithmic functions compared to power functions. This problem also reviews *vertical* and *horizontal asymptotes* for e^x and $\ln(x)$. Since the domain of $\ln(x)$ is $x > 0$, there is a *vertical asymptote* at $x = 0$. From properties of the exponential function, e^x, there is a *horizontal asymptote* at $y = 0$ as $x \to +\infty$.

Problem 3.4 (Exponential, Logarithmic, and Power Functions)**.** *(Lab121_E1_exp_ln_pwr.pg) This problem explores the exponential, logarithmic, and power functions.*

This problem is intended to compare the exponential function to high order polynomials and compare the logarithmic function to fractional powers of x.

a. Consider the functions:

$$f(x) = 0.9e^{1.1x} \qquad \text{and} \qquad g(x) = 3.1x^8.$$

The domain for $f(x)$ is _____ and for $g(x)$ is _____.

(Write the answers in interval notation.)

Find all points of intersection (x and y values).

Points of intersection = (_____, _____), (_____, _____),

and (_____, _____).

Which of the following are statements are appropriate for $f(x)$?

A. This function has a Vertical Asymptote at $x = 0$.
B. This function has a Horizontal Asymptote as $x \to -\infty$.
C. This function has a Vertical Asymptote at some $x \neq 0$.
D. This function has a Horizontal Asymptote as $x \to \infty$.
E. None of the Above

If $f(x)$ has a Horizontal Asymptote or a Vertical Asymptote with $x \neq 0$, then give the x or y value of the asymptote or type DOES NOT APPLY.

Which of the following are statements are appropriate for $g(x)$?

A. This function has a Horizontal Asymptote as $x \to -\infty$.
B. This function has a Horizontal Asymptote as $x \to \infty$.
C. This function has a Vertical Asymptote at some $x \neq 0$.
D. This function has a Vertical Asymptote at $x = 0$.
E. None of the Above.

If $g(x)$ has a Horizontal Asymptote or a Vertical Asymptote with $x \neq 0$, then give the x or y value of the asymptote or type DOES NOT APPLY .

---

b. In the Lab Report, create two graphs for $f(x)$ and $g(x)$. For the first graph, choose a symmetric domain ($x \in [-A, A]$) that clearly shows the first two points of intersection, *i.e.*, take a value of A that is slightly larger than either of the x-values (in absolute value) for the first two points of intersection. For the second graph choose a larger symmetric domain that includes the third point of intersection. Note that this point of intersection has a very large y value, so the other two points of intersection are not very visible. Briefly describe the appearance of the graphs of each of these functions. Briefly discuss how any vertical or horizontal asymptotes are found. Also, describe how the points of intersection are found. Which function dominates (has larger y-values) for large values of x?

c. Now consider the functions:

$$h(x) = 1.2\ln(x) \qquad \text{and} \qquad k(x) = 1.2x^{1/7}.$$

The domain for $h(x)$ is _____ and for $k(x)$ is _____.

Find all points of intersection (x and y values).

Points of intersection = (_____, _____) and (_____, _____)

Which of the following statements are appropriate for $h(x)$?

A. This function has a Horizontal Asymptote as $x \to \infty$.
B. This function has a Vertical Asymptote at $x = 0$.
C. This function has a Horizontal Asymptote as $x \to -\infty$.
D. This function has a Vertical Asymptote at some $x \neq 0$.
E. None of the Above

If $h(x)$ has a Horizontal Asymptote or a Vertical Asymptote with $x \neq 0$, then give the x or y value of the asymptote or type DOES NOT APPLY.

Which of the following statements are appropriate for $k(x)$?

A. This function has a Vertical Asymptote at $x = 0$.
B. This function has a Vertical Asymptote at some $x \neq 0$.
C. This function has a Horizontal Asymptote as $x \to \infty$.
D. This function has a Horizontal Asymptote as $x \to -\infty$.
E. None of the Above.

If $k(x)$ has a Horizontal Asymptote or a Vertical Asymptote with $x \neq 0$, then give the x or y value of the asymptote or type DOES NOT APPLY.

---

d. In the Lab Report, create two graphs for $h(x)$ and $k(x)$. For the first graph, choose the domain $x \in [0, 10]$. For the second graph choose a much larger domain, $x \in [0, A]$, that includes the second point of intersection. Note that this point of intersection has a very large x value, which again masks the first point of intersection. Briefly describe the appearance of the graphs of each of these functions. Briefly discuss how any vertical or horizontal asymptotes are found. Also, describe how the points of intersection are found. Which function dominates (has larger y-values) for large values of x?

5 ALLOMETRIC MODELING

Many modeling problems are nonlinear. One very important class fits a model in the form of an *allometric* or *power law* model. These models have the form

$$y = kx^a. \tag{3.1}$$

By taking logarithms of this equation and using the properties of logarithms, Eqn. (3.1) can be written in the following form

$$\ln(y) = a \ln(x) + \ln(k).$$

By defining the variables $Y = \ln(y)$, $X = \ln(x)$, and $K = \ln(k)$, then this equation has the form

$$Y = aX + K,$$

which is a straight line with slope a and intercept K. The techniques from before can be used to find the best linear least squares best fit through the logarithm of the data to find the best *allometric model*. This gives an easy method for finding a and K or equivalently k. Thus, a best fitting model is found by fitting a linear model through the logarithms of the data.

The technique above gives the easiest way to fit an allometric model, since virtually all computer applications have linear least squares fitting algorithms. However, this method is finding the least squares best fit to the logarithms of the data and not the least squares best fit directly from the data. The latter is a nonlinear least squares best fit of the data. Though the nonlinear least squares best fit is technically a better unbiased fit to the data, it is much more challenging to find. In particular, it usually requires a reasonable initial guess of the parameter values. Below we show how to use a MATLAB program to find this nonlinear least squares fit to data.

Example 3.4 (Cumulative AIDS Cases)**.** *In the early years the numbers of AIDS cases grew rapidly, and the cumulative cases of AIDS could be modeled by an allometric model of the form*

$$c = kt^r,$$

where t is the number of years after 1980, c is cumulative number of AIDS cases, and k and r are parameters that fit the data.

Table 3.4 shows the data for the cumulative number of AIDS cases in the U. S. over the time period of 1981-1992. During these early years the disease was not understood, and its spread was very rapid.

Year	Cumulative AIDS Cases	Year	Cumulative AIDS Cases	Year	Cumulative AIDS Cases
1981	97	1985	15,242	1989	120,612
1982	709	1986	29,944	1990	161,711
1983	2,698	1987	52,902	1991	206,247
1984	6,928	1988	83,903	1992	257,085

Table 3.4: Data of the Cumulative AIDS (in thousands of cases) throughout the U.S.A. from 1981 until 1992.[4]

The data clearly cannot be fit well with a linear model, so one option is to use an allometric model. If we define the variable t to be the number of years after 1980 and c to be the cumulative number of AIDS cases, then an appropriate allometric model is given by:

$$c = kt^r,$$

where the parameters k and r are found from the data above. With the use of logarithms this model can be written:

$$\ln(c) = \ln(k) + r\ln(t).$$

If we let $C = \ln(c)$, $K = \ln(k)$, and $T = \ln(t)$, then the above model is written:

$$C = K + rT,$$

which is just a linear equation. Thus, the logarithms of the data can be fit to a straight line, as was done before, to give the allometric model.

MATLAB for Power Law Fit

The analysis above shows that the linear least squares fit of a line to the logarithms of the data provides the allometric model. This suggests that MATLAB's polyfit routine can be used to fit these types of models. The following MATLAB function allows the easy input of data to produce the parameters for the allometric model:

```
function [k,a] = powerfit(xdata,ydata)
% Power law fit for model y = k*x^a
% Uses linear least squares fit to logarithms of data
Y = log(ydata);      % Logarithm of y-data
X = log(xdata);      % Logarithm of x-data
p = polyfit(X,Y,1);  % Linear fit to X and Y
a = p(1);            % Value of exponent
k = exp(p(2));       % Value of leading coefficient
end
```

For the cumulative AIDS case example above, if the *powerfit* function is defined in our workspace, then the following MATLAB commands will produce the desired coefficients of the model:

```
t = 1:12;
c = [97 709 2698 6928 15242 29944 52902 83903 120612 161711 206247 257085];
[k,r] = powerfit(t,c)
```

This produces the coefficients

$$k = 82.7003 \qquad \text{and} \qquad r = 3.2741,$$

so this allometric model is given by

$$c = 82.7003\, t^{3.2741}.$$

The sum of square errors for this allometric model can be readily computed in a similar manner to before. The MATLAB function for computing this sum of square errors is given by:

```
function J = sumsq_pow(p,xdata,ydata)
% Function to compute the least squares error for allometric model
model = p(1)*xdata.^p(2);  % Power law model using parameter p
error = model - ydata;      % Error between model and data
J = error*error';           % Computes sum of square error
end
```

If we define $p = [82.7003, 3.2741]$, then the command in MATLAB of

```
J = sumsq_pow(p,t,c)
```

where t and c are the data above, gives the result $J = 9.3173 \times 10^8$.

Since the sum of square errors function is already created and the powerfit function gave a parameter fit to the model, we can take advantage of MATLAB's *fminsearch* function to find the nonlinear least squares best fit to the data. The command

```
[p1,J,flag] = fminsearch(@sumsq_pow,p,[],t,c);
```

where p, t, and c are the parameters and data from above, results in the variable $p1 = [212.8464, 2.8643]$ and $J = 2.1054 \times 10^8$. The variable $flag$ is 1, saying the routine converged to an acceptable value according to MATLAB. (If this variable is 0, then the answer is probably wrong.) This nonlinear fit to the allometric model produces a result with a sum of square errors less than a quarter of the error produced by fitting the logarithm of the data.

The techniques above produce the best fitting models to the logarithms of the data and the best nonlinear least squares fit to the data itself, giving the respective models:

$$c = 82.700\, t^{3.2741} \qquad \text{and} \qquad c = 212.85\, t^{2.8643}.$$

These models are shown in Fig. 3.4. In MATLAB, the graph can have logarithmic scales by replacing the *plot* command with the *loglog* command. For example, typing:

```
loglog(xx,yy,'b-','LineWidth',1.5);
```

where `xx` and `yy` are the points on the blue curve. One can also make this change by editing the figure and changing the x and y scales. From a modeling perspective, if data for a particular problem are plotted with a log-log scale and the plot of the data appears to follow a straight line, then an allometric model is likely a good choice.

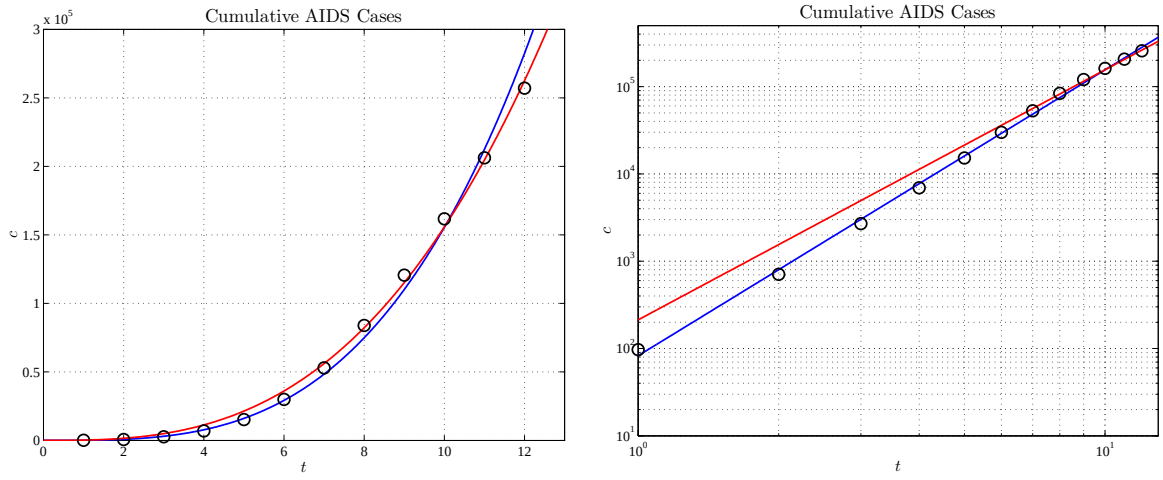

Figure 3.4: Graph of cumulative AIDS cases showing the data with the linear least squares best fit to the logarithms of the data (blue) and the nonlinear least square best fit to the data (red). The graph on the right is the same graph with logarithmic scales for the t and c axes.

In Fig. 3.4, both models fit the data quite well. However, the logarithmic fit to the data does a poorer job fitting the larger t values. Thus, the second model provides a better one for projecting

future trends for cumulative AIDS cases. It is often a subjective decision as to which model is better. The most important parameter in the allometric models is the exponent, r, and for the two fits in Fig. 3.4 these differ by about 15%, which is not too bad in many biological models.

Problem 3.5 (Dog Measurement Study)**.** *(Lab121_D3_dog_study.pg) Allometric models are used to study the relationship between length, weight, and surface area of several dogs.*

Allometric modeling provides a method of understanding physical relationships between different methods of measurement. This problem examines the relationship of the length of a dog to its weight and surface area. Physiologically, it is important to note that all dogs have a basic shape that one recognizes as a dog. This is a type of self-similarity. A collection of dogs was measured and weighed, producing the following table of data[11]. (Note that the length is from nose to anus.)

Length (cm)	Body Weight (gm)	Surface Area (cm^2)
48	3460	2245
63	5150	3250
72	5350	3865
76	9980	5000
97	17220	7900
101	26070	8995
105	33260	10535

a. The first model examines Weight, w, as a function of the Length, u. Use the allometric modeling technique of fitting the logarithms of the data to find the best model fitting the data for a model of the form:

$$w = ku^a.$$

The best fit coefficients, k and a, are given by

$k =$ _____ and $a =$ _____

Determine the sum of square errors between the weight model and the data.

Sum of Square Errors = _____

If we assume that this model gives the ideal weight for a dog, then a dog with a length of 72 cm should weigh

Weight of 72 cm dog = _____gm.

Assuming this model gives the best value, compute the percent error that the 72 cm dog in the table varies from the ideal dog in weight.

Percent Error (weight) for 72 cm dog = _____%.

Suppose a dog has a length of 101 cm, then according to the model it should weigh

Weight of 101 cm dog = _____gm.

Again, assuming this model gives the best value, compute the percent error that the 101 cm dog in the table varies from the ideal dog in weight.

Percent Error (weight) for 101 cm dog = _____%.

Suppose you have a dog that weighs 14460 gm. According to the model, what would you predict would be its length?

Length of 14460 gm dog = _____cm.

b. In the Lab Report, create a plot of this best fitting curve with the data. What are the appropriate units for the coefficient k? Write a brief paragraph to explain why the coefficient a has the value it does when fitting the weight data for the dogs.

c. Repeat the process in Part a for the surface area data, s, as a function of length, u. In this case, the model satisfies the allometric model:

$$s = ku^a,$$

where the coefficients k and a differ from Part a. The best fit coefficients, k and a, are given by

$k =$ _____ and $a =$ _____

Determine the sum of square errors between the surface area model and the data.

Sum of Square Errors = _____.

If we assume that this model gives the ideal surface area for a dog, then a dog with a length of 72 cm should have a surface area

Surface area of 72 cm dog = _____cm^2

Assuming this model gives the best value, compute the percent error that the 72 cm dog in the table varies from the ideal dog in surface area.

Percent Error (surface area) for 72 cm dog = _____%.

Suppose a dog has a length of 101 cm, then according to the model it should have a surface area

Surface area of 101 cm dog = _____cm^2.

Again, assuming this model gives the best value, compute the percent error that the 101 cm dog in the table varies from the ideal dog in surface area.

Percent Error (surface area) for 101 cm dog = _____%.

Suppose you have a dog that has a surface area 6945 cm^2. According to the model, what would you predict would be its length?

Length of 6945 cm^2 dog = _____cm.

d. In the Lab Report, create a plot of this best fitting curve with the data. What are the appropriate units for the coefficient k? Write a brief paragraph to explain why the coefficient a has the value it does when fitting the surface area data for the dogs.

e. Repeat the process in Part a for finding Weight, w, as a function of the Length, u. However, use the allometric modeling technique of nonlinear least squares to fit the data to the model:

$$w = ku^a.$$

The best fit coefficients, k and a, and sum of square errors are given by

$k =$ _____, $a =$ _____, and Sum of Square Errors = _____

Assume that this model gives the ideal weight (best value) for a dog. Find the weight of a dog with a length of 72 cm and compute the percent error that the 72 cm dog in the table varies from the ideal dog in weight.

Weight of 72 cm dog = _____gm Percent Error (weight) = _____%.

Find the weight of a dog with a length of 101 cm and compute the percent error that the 101 cm dog in the table varies from the ideal dog in weight.

Weight of 101 cm dog = _____gm Percent Error (weight) = _____%.

Suppose you have a dog that weighs 14460 gm. According to the model, what would you predict would be its length?

Length of 14460 gm dog = _____cm.

f. Repeat the process in Part c for finding the surface area, s, as a function of the Length, u. However, use the allometric modeling technique of nonlinear least squares to fit the data to the model:

$$s = ku^a,$$

where the coefficients k and a differ from Part e. The best fit coefficients, k and a, and sum of square errors are given by

$k =$ _____, $a =$ _____, and Sum of Square Errors = _____

Assume that this model gives the ideal surface area (best value) for a dog. Find the surface area of a dog with a length of 72 cm and compute the percent error that the 72 cm dog in the table varies from the ideal dog in surface area.

Surface area of 72 cm dog = _____gm Percent Error (weight) = _____%.

Find the surface area of a dog with a length of 101 cm and compute the percent error that the 101 cm dog in the table varies from the ideal dog in surface area.

Surface area of 101 cm dog = _____gm Percent Error (weight) = _____%.

Suppose you have a dog that surface area 6945 cm^2. According to the model, what would you predict would be its length?

Length of 6945 cm^2 dog = _____ cm.

g. In the Lab Report add the nonlinear least squares fitting allometric models to the graphs created in Parts b and d. Write a brief paragraph comparing the graphs from the logarithmic fit to the data and the nonlinear least squares fit to these allometric models.

Problem 3.6 (Island Biodiversity)**.** *(Lab121_E2_biodiversity.pg) An allometric model is fitted to data on herpetofauna to study biodiversity on islands of the Caribbean.*

Currently there is a debate on the importance of preserving large tracts of land to maintain biodiversity. Many of the arguments for setting aside large tracts are based on studies of biodiversity on islands. In this problem you apply the power rule to determine the number of species of herpetofauna (amphibians and reptiles) as a function of island area for the given Caribbean islands. You are given the following data[11]:

Island	Area (km^2)	Species
Redunda	2.4	3
Saba	13	7
Montserrat	86.8	10
Jamaica	11,434	40
Hispaniola	76,490	91

a. Let N be the number of species and A be the area of the island. Use the allometric modeling technique of fitting the logarithms of the data to find the best model fitting the data for a model. The power law expression relating the number of species to the area of the island is given by

$$N = kA^a.$$

$k =$ _____ and $a =$ _____

Determine the sum of square errors between the species model and the data.

Sum of Square Errors = _____.

b. In the Lab Report, plot the data and the best power law fit. Show the formula for the best fitting model on the graph. How well does the graph match the data?

c. Next, we want to fit a straight line to the logarithms of the data. From the allometric model above, we obtain the formula

$$\ln(N) = \ln(k) + a\ln(A).$$

In the table above, take the logarithm of the Number of Species ($\ln(N)$) and the logarithm of the Island Area ($\ln(A)$). Graph the logarithms of the data ($\ln(N)$ vs $\ln(A)$), then apply a linear least squares fit of the data.

What is the value of the slope of this best fitting line?

Slope = _____.

What is the value of the $\ln(N)$-intercept?

Intercept = _____.

d. In the Lab Report, plot the logarithm of the data and the best straight line fit to these data. Show the formula for the best fitting linear model on the graph. How well does the graph match the data? Write a brief discussion of how the coefficients obtained in this manner compare to the ones found in Part a.

e. The Caribbean island of Saint Croix has an area of 214 km^2. Use the allometric model found above to determine the predicted number of species of herpetofauna on this island

Number of species on Saint Croix = _____.

If it is found that Saint Croix has 13 species of herpetofauna on the island, then determine the percent error between the model and the actual number of species found. (Use the actual number of species as the best value in the percent error formula.)

Percent Error (species) for Saint Croix = _____.

Suppose that the island of Martinique has an area of 1133 km^2. Use the allometric model found above to determine the predicted number of species of herpetofauna on this island

Number of species on Martinique = _____.

If it is found that Martinique has 30 species of herpetofauna on the island, then determine the percent error between the model and the actual number of species found.

Percent Error (species) for Martinique = _____.

Now suppose that the Caribbean island of Puerto Rico is found to have 42 species of herpetofauna. Use the allometric model found above to determine the predicted an area km^2 of Puerto Rico.

Area of Puerto Rico = _____ km^2

The actual area of Puerto Rico is 8899 km^2. Use this information to determine the percent error between the model and the actual area. (Use the actual area as the best value in the percent error formula.)

Percent Error (area) for Puerto Rico = _____

Suppose that the Caribbean island of Cuba is found to have 96 species of herpetofauna. Use the allometric model found above to determine the predicted an area km^2 of Cuba.

Area of Cuba = _____ km^2

The actual area of Cuba is 120961 km^2. Use this information to determine the percent error between

the model and the actual area.

Percent Error (area) for Cuba = _____

f. This problem helps address the importance of maintaining a large tract of land for the maintenance of biodiversity. Based on the allometric model computed above, determine a factor of how much a land area must increase to double the number of species supported by the environment.

Area is multiplied by _____ to double the number of species of herpetofauna in the Caribbean region.

g. Write a brief paragraph about this mathematical model. What does this model say about the size of land tracts needed to preserve biodiversity? How can this model be extended to different regions on Earth and different types of animals? Would you expect a similar power law relationship, and if so, which of the coefficients in a new model for a different set of species in a different geographic location is most likely to be similar to the one computed for the herpetofauna data?

h. (Extended Study) In the discussion of the allometric modeling technique, there was the alternate method of fitting the data using nonlinear least squares. This problem can be approached using MATLAB's *fminsearch* to solve this highly nonlinear problem. The coefficients in Part a and least squares values can be found, then applied to the specific examples in Part e and the land area factor is Part f. The percent errors will suggest which model is fitting the data better. Finally, one can add the graph of this nonlinear least squares best fit allometric model to the graph from Part b to visually compare the differences.

Problem 3.7 (Allegheny Forest)**.** *(Lab121_E3_allegheny.pg) Data from the Allegheny National Forest is used to model the volume of trees as a function of diameter or height. Both linear and allometric models are examined.*

The data below came from the Allegheny National Forest in Pennsylvania[13]. The data were collected from Black Cherry trees with the diameter measured 4.5 ft above ground. One issue was whether either the diameter or the height of a tree accurately predicts the volume of wood in the tree. The volume is measured in cubic feet. This problem studies both linear and allometric models.

a. The simplest model is a linear model, so find the best line through the data. Begin with a linear model for the volume, V_d, in ft^3 as a function of the diameter, d, in inches

$$V_d = md + b.$$

The best linear model has a slope, m, and intercept, b with

$m =$ _____ and $b =$ _____

Determine the sum of square errors between this model and the data.

Sum of Square Errors = _____.

Diameter (in)	Height (ft)	Volume (ft^3)	Diameter (in)	Height (ft)	Volume (ft^3)	Diameter (in)	Height (ft)	Volume (ft^3)
8.3	70	10.3	11.4	76	21	14.2	80	31.7
8.6	65	10.3	11.4	76	21.4	14.5	74	36.3
8.8	63	10.2	11.7	69	21.3	16	72	38.3
10.5	72	16.4	12	75	19.1	16.3	77	42.6
10.7	81	18.8	12.9	74	22.2	17.3	81	55.4
10.8	83	19.7	12.9	85	33.8	17.5	82	55.7
11	66	15.6	13.3	86	27.4	17.9	80	58.3
11	75	18.2	13.7	71	25.7	18	80	51.5
11.1	80	22.6	13.8	64	24.9	18	80	51
11.2	75	19.9	14	78	34.5	20.6	87	77
11.3	79	24.2						

Table 3.5: Data for Black Cherry trees from the Allegheny Forest. More information on entering these data is available.54

Repeat this process for a linear model for the volume, V_h, in ft^3 as a function of the height, h, in ft

$$V_h = kh + c.$$

This best linear model has a slope, k, and intercept, c with

$k =$ _____ and $c =$ _____

Determine the sum of square errors between this model and the data.

Sum of Square Errors = _____.

Consider the tree that has a volume of 19.9 ft^3. The data give this tree a diameter of 11.2 in and a height of 75 ft. Use the models above to predict the expected volume and the percent error for each of the models (assuming the measured value to be the best value).

For $d = 11.2$, $V_d =$ _____ft^3 with percent error = _____%.

For $h = 75$, $V_h =$ _____ft^3 with percent error = _____%.

Based on the two linear models and the graphs with the data, which data set (DIAMETER or HEIGHT) better predicts the volume of a tree? _____

b. In the Lab Report, create a graph of the data and a linear model for volume as a function of diameter. Show the formula for the model on the graph. Also, create a second graph of the data and a linear model for volume as a function of height with its formula on the graph. Write a brief explanation for why you would expect one of these measurements to better predict the volume based on the biology of trees. What happens with both models as the diameter or height gets close to zero? Is this biologically feasible?

c. For this part of the problem, we examine an allometric model for the volume as a function of the diameter. Consider the model given by

$$V = Kd^A.$$

The best fitting coefficients using a linear fit to the logarithms of the data, K and A, are given by

$K =$ _____ and $A =$ _____.

Determine the sum of square errors between the weight model and the data.

Sum of Square Errors = _____.

Consider the tree that has a volume of 19.9 ft^3. The data give this tree a diameter of 11.2 in. Use this model to predict the expected volume and the percent error for the model (assuming the measured value to be the best value).

For $d = 11.2$, $V =$ _____ ft^3 with percent error = _____%.

d. In the Lab Report, plot the data and the best allometric model. Be sure to include the formula for the model on the graph. Find a reasonable explanation for the power that is obtained with the allometric model. (Round the power to the nearest integer and explain why that integer value is relevant to relating volume of a tree to its diameter to this power.) Also, have MATLAB plot a log-log plot of the data and the model that is found in Part c. Do the data roughly fall on a straight line in this log-log plot?

e. The method above of fitting an allometric or power law model to a set of data uses the best straight line through the logarithms of the data. In the case above, we have an allometric model of the form $V = Kd^A$, which can be written in the form

$$\ln(V) = \ln(K) + A\ln(d).$$

In the table above, find the logarithm of the volume ($\ln(V)$) and the logarithm of the diameter($\ln(d)$). Graph the ($\ln(V)$) vs ($\ln(d)$) data, then apply a linear least squares best fit to the logarithms of the data.

What is the value of the slope of this best fitting line? Slope = _____.

What is the value of the $\ln(V)$-intercept?

Intercept = _____.

f. In the Lab Report, plot the logarithms of the data and the best straight line fit to these logarithms. Show the formula for the best fitting linear model on the graph. How well does the graph match the data? Write a brief discussion of how the coefficients obtained in this manner compare to the ones found in Part c. Write a brief discussion comparing these different forms of

the allometric model graphs shown in Parts d and f.

h. (Extended Study) In the discussion of the allometric modeling technique, there was the alternate method of fitting the data using nonlinear least squares. This problem can be approached using MATLAB's *fminsearch* to solve this highly nonlinear problem. The coefficients in Part c and least squares values can be found with this method, then applied to the specific example in Part c. Finally, one can add the graph of this nonlinear least squares best fit allometric model to the graph from Part d to visually compare the differences.

6 OTHER NONLINEAR FUNCTIONS

The techniques above have examined models with linear and power law functions. In the process, information has been given to enter quadratic, exponential, and logarithmic functions into MATLAB and Maple. In this section, several other examples are explored that use different special functions. The first section explores weak acid chemistry and introduces the square root and $\log_{10}$ functions. Later in this section rational functions are applied to an example of enzyme kinetics.

6.1 WEAK ACID CHEMISTRY

Many of the organic acids found in biological applications are weak acids. Formic acid (HCOOH) is a relatively strong weak acid that ants use as a defense[1]. The strength of this acid makes the ants very unpalatable to predators. The chemistry of dissociation is given by the following equation:

$$\mathrm{HCOOH} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathrm{H^+ + HCOO^-}.$$

Each acid has a distinct equilibrium constant K_a that depends on the properties of the acid and the temperature of the solution. For formic acid, $K_a = 1.77 \times 10^{-4}$. Let [X] denote the concentration of a particular chemical species X, then assuming that the formic acid is in equilibrium, it satisfies the following equation:

$$K_a = \frac{[\mathrm{H^+}][\mathrm{HCOO^-}]}{[\mathrm{HCOOH}]}.$$

The *normality* of a weak acid solution is a concentration measurement for reactions. In fact, the normality x = the molarity of the weak acid (undissociated + the number of protons from dissociated weak acid). If formic acid is added to water, then dissociation causes the acid ion ($\mathrm{H^+}$) to be approximately the same as the formic ion ($\mathrm{HCOO^-}$), this means $[\mathrm{H^+}] \approx [\mathrm{HCOO^-}]$. Also, if x is the normality of the solution, then $x = [\mathrm{HCOOH}] + [\mathrm{HCOO^-}]$. (The formic anion $\mathrm{HCOO^-}$ must be either in the bound form HCOOH or ionized form $\mathrm{HCOO^-}$ with the total representing the normality of the formic acid added to solution.) It follows that $[\mathrm{HCOOH}] \approx x - [\mathrm{H^+}]$. From the approximations, the equation for K_a is written:

$$K_a = \frac{[\mathrm{H^+}][\mathrm{H^+}]}{x - [\mathrm{H^+}]},$$

[1] en.wikipedia.org/wiki/Formic_acid, last visited 10/19/13

which is easily solved using the quadratic formula to give,

$$[\mathrm{H}^+] = \frac{1}{2}\left(-K_a + \sqrt{K_a^2 + 4K_a x}\right).$$

Notice that we only take the positive solution from the quadratic equation to make physical sense.

MATLAB and Maple for Square Root and $\log_{10}$

The formula above is used to find the concentration of $[\mathrm{H}^+]$ for a solution of formic acid with a normality of 0.1 N and determine its pH. Since formic acid has $K_a = 1.77 \times 10^{-4}$ and $x = 0.1$, the formula given above yields:

$$[\mathrm{H}^+] = \frac{1}{2}\left(-0.000177 + \sqrt{(0.000177)^2 + 4(0.000177)(0.1)}\right) \approx 0.00412.$$

In either MATLAB or Maple, the square root function is written `sqrt`, so

```
0.5*(-0.000177+sqrt(0.000177^2+4*0.000177*0.1))
```

gives the answer that the concentration of acid in the 0.1 N solution is $[\mathrm{H}^+] = 4.12 \times 10^{-3}$. Since the pH is defined to be $-\log_{10}[\mathrm{H}^+]$, then using either MATLAB or Maple, where $\log_{10}$ is written `log10`, so

```
log10(0.00412)
```

gives a pH of 2.385.

Problem 3.8 (Weak Acid Study)**.** *(Lab121_C2_weak_acid.pg) A quadratic approximation of weak acid chemistry is used to find the* $[\mathrm{H}^+]$ *and the pH of a solution, as the normality of the solution varies.*

Acetic acid (CH_3COOH) is a weak acid with an equilibrium constant $K_a = 1.75 \times 10^{-5}$. Use the techniques developed above to find an expression for $[\mathrm{H}^+]$ as a function of the normality, x, of the weak acid solution.

a. In the Lab Report, write the expression for $[\mathrm{H}^+]$ as a function of x using the value of $K_a = 1.75 \times 10^{-5}$. One should write this formula using Microsoft Equation or some other technical writing tool.

b. Fill in the following table:

$x = 0.05$ N	$x = 0.45$ N	$x = 1.45$ N
$[\mathrm{H}^+] =$ _____	$[\mathrm{H}^+] =$ _____	$[\mathrm{H}^+] =$ _____

Also, when $[\mathrm{H}^+] = 0.004$, find

$x =$ _____ N.

c. In the Lab Report, create a graph of the $[H^+]$ as a function of the normality x for $x \in [0.001, 2]$. Be sure to properly label the axes.

d. The pH of a solution is given by

$$\text{pH} = -\log_{10}([H^+]).$$

Fill in the following table:

$x = 0.05$ N	$x = 0.45$ N	$x = 1.45$ N
pH = _____	pH = _____	pH = _____

Also, when the pH = 2.8, find

$x =$ _____ N.

e. In the Lab Report, create a graph of the pH as a function of the normality x for $x \in [0.001, 2]$. Be sure to properly label the axes.

6.2 RATIONAL FUNCTIONS

A rational function, $r(t)$, is defined by the quotient of two polynomials, $p(t)$ and $q(t)$, with

$$r(t) = \frac{p(t)}{q(t)},$$

and defined for all t provided $q(t) \neq 0$. This form of function is very important in enzyme kinetic models and some types of population models. Rational functions are characterized by having *vertical asymptotes*, where $q(t) = 0$ (unless $p(t) = 0$ at the same t), and *horizontal asymptotes*, if the degree of $p(t)$ is less than or equal to the degree of $q(t)$. The asymptotes can complicate the graphing of this function.

Example 3.5 (Rational Function)**.** *This example considers the intersection of a line and a rational function and shows how to create a detailed graph of the functions.*

Consider the two functions:

$$f(x) = x - 1 \qquad \text{and} \qquad g(x) = \frac{x}{x^2 - 4}.$$

The function $f(x)$ is a linear function with a slope of 1 and intercept of -1. The rational function has a domain of all $x \neq \pm 2$, which indicates there are vertical asymptotes at $x = \pm 2$. Since the degree of the numerator is less than the degree of the denominator, there is a horizontal asymptote at $y = 0$. Since numerator of $g(x)$ has only odd powers and the denominator has only even powers, $g(x)$ is an odd function and is symmetric about the origin. The only x and y-intercept is the origin.

From this information, it should be easy to sketch the graph of these functions. The graph would show these functions intersecting at three points. To find these points of intersection would require solving a cubic equation, so it is time to turn to Maple to find the points of intersection. We show the graphing commands for both Maple and MATLAB.

Maple and MATLAB Commands for Example

As seen earlier, Maple provides an easy means of graphing the functions, and this graph is useful for estimating the points of intersection. The functions are entered in the following manner:

$f := x \to x - 1; \qquad g := x \to \dfrac{x}{x^2 - 4}$

Since $g(x)$ has vertical asymptotes, the following command makes a good graph, which is shown in Fig. 3.5:

```
plot({f(x), g(x)}, x = -8 .. 8, y = -15 .. 15, discont = true)
```

Notice that we have added the additional options that limit the range of the function and a special command to tell Maple that it should allow for vertical asymptotes (*discont = true*).

Either from our intuition about the functions or the graph, we recognize there are three points of intersection, which are separated by the vertical asymptotes. Once again, Maple provides an easy tool to find these points of intersection. However, unlike its capability of finding points of intersection for polynomials, where it can find all solutions in one line, Maple generally only looks for a single solution when there is a nonpolynomial function. It does this to prevent falling into an infinite loop. For this problem, there are three points of intersection to find, so the following Maple commands obtain these points of intersection:

```
x1 := fsolve(f(x) = g(x), x = -5 .. -2); f(x1)
x2 := fsolve(f(x) = g(x), x = -2 .. 2);  f(x2)
x3 := fsolve(f(x) = g(x), x = 2 .. 5);   f(x3)
```

The second expression in the *fsolve* command shows a range where the user has Maple search to find the x value of the intersection point, and this value is stored in one of the variables, `x1`, `x2`, or `x3`. The command `f(x1)` provides the corresponding y-value for `x1`. The resulting points of intersection are:

$$(x_1, y_1) = (-2.1642, -3.1642), \quad (x_2, y_2) = (0.7729, -0.2271), \quad \text{and} \quad (x_3, y_3) = (2.3914, 1.3914).$$

The MATLAB graph shown in Fig. 3.5 is significantly more complex. Many more curves were added to this graph. The basic graphing template from before is used, but many more lines are needed to set up the domain for the graphing (3 separate domains for the rational function), the function evaluations, the plotting, and the text entries in the field. We included additional options for adjusting the color of fonts and rotating the text. This script demonstrates the amount of control and detail, which can be entered into a MATLAB graph to produce high quality graphs. Below is the script for this more complex graph.

```
mytitle = 'Rational Function and Line';  % Title
xlab = '$$x$$';  % X-label
ylab = '$$y$$';  % Y-label

x1 = -8;       % Left x-value
x2 = 8;        % Right x-value
Npts = 500;    % Number of points plotted between x1 and x2
xx = linspace(x1,x2,Npts); % Creates vector of x-values
xx1 = linspace(-8,-2.001,Npts); % Creates vector of x-values
xx2 = linspace(-1.999,1.999,Npts); % Creates vector of x-values
```

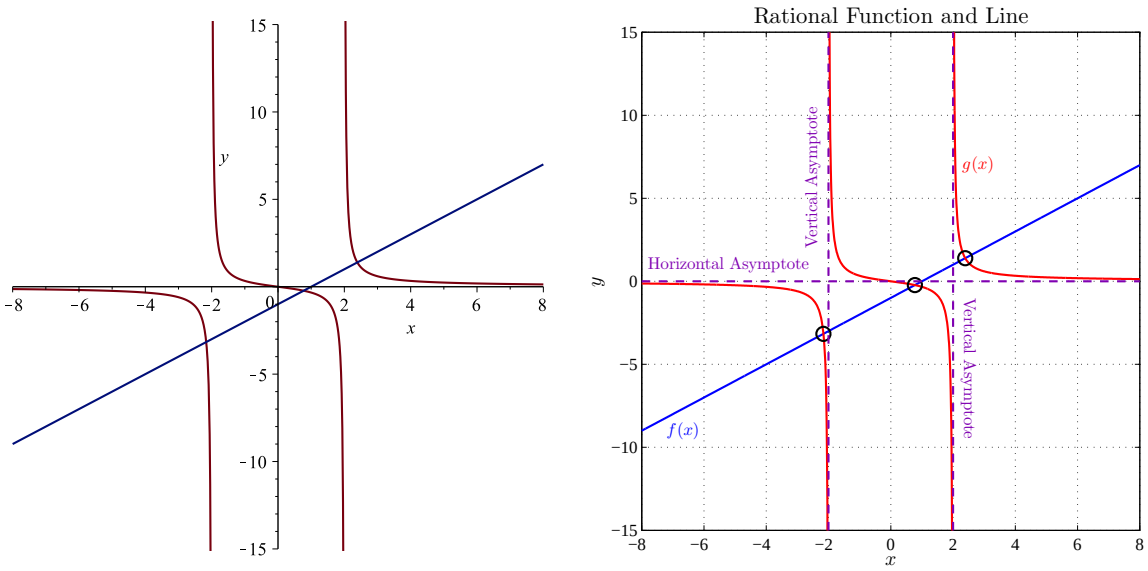

Figure 3.5: Graph of $f(x) = x - 1$ and $g(x) = \frac{x}{x^2-4}$.

```
xx3 = linspace(2.001,8,Npts);     % Creates vector of x-values

yy = xx - 1;              % Equation of Line
yy1 = xx1./(xx1.^2-4);  % Equation of g(x)
yy2 = xx2./(xx2.^2-4);  % Equation of g(x)
yy3 = xx3./(xx3.^2-4);  % Equation of g(x)
xpt = [-2.164,0.773,2.391];     % x points of intersection
ypt = [-3.164,-0.227,1.391];    % y points of intersection
xv1 = [-2,-2,-2];   % x of Vertical Asymptote
yv1 = [-15,0,15];   % y of Vertical Asymptote
xv2 = [2,2,2];      % x of Vertical Asymptote
yv2 = [-15,0,15];   % y of Vertical Asymptote
xh1 = [-8,0,8];     % x of Horizontal Asymptote
yh1 = [0,0,0];      % y of Horizontal Asymptote

plot(xx,yy,'b-','LineWidth',1.5);   % Graph of Line
hold on                             % Plots Multiple graphs
plot(xx1,yy1,'r-','LineWidth',1.5); % Graph of g(x)
plot(xx2,yy2,'r-','LineWidth',1.5); % Graph of g(x)
plot(xx3,yy3,'r-','LineWidth',1.5); % Graph of g(x)
plot(xpt,ypt,'ko','MarkerSize',10); % Intersection Points
plot(xv1,yv1,'--','LineWidth',1.5,'Color',[0.5 0 0.7]); % Vertical Asymptote
plot(xv2,yv2,'--','LineWidth',1.5,'Color',[0.5 0 0.7]); % Vertical Asymptote
plot(xh1,yh1,'--','LineWidth',1.5,'Color',[0.5 0 0.7]); % Horizontal Asymptote
grid                                % Adds Gridlines

text(-7.2,-9,'$$f(x)$$','FontSize',12,'Color','blue');
text(2.3,7,'$$g(x)$$','FontSize',12,'Color','red');
```

```
text(-7.8,1,'Horizontal Asymptote','FontSize',12,'Color',[0.5 0 0.7]);
text(-2.5,2,'Vertical Asymptote','Rotation',90,'FontSize',12,'Color',[0.5 0 0.7]);
text(2.5,-1,'Vertical Asymptote','Rotation',270,'FontSize',12,'Color',[0.5 0 0.7]);

y1 = -15;      % Lower y-value
y2 = 15;       % Upper y-value
axis([x1 x2 y1 y2]);  % Defines limits of graph
axis square;

fontlabs = 'Times New Roman';   % Font type used in labels
xlabel(xlab,'FontSize',14,'FontName',fontlabs);  % x-Label size and font
ylabel(ylab,'FontSize',14,'FontName',fontlabs);  % y-Label size and font
title(mytitle,'FontSize',16,'FontName','Times New Roman'); % Title size/font
set(gca,'FontSize',12);         % Axis tick font size

print -depsc ratl_ML.eps % Create figure as EPS file
```

Problem 3.9 (Rational and Linear Functions)**.** *(Lab121_D1_ratl.pg) This problem explores a line and rational function. The points of intersection are found, and a graph is created.*

Consider the functions

$$f(x) = 3.7 - x \qquad \text{and} \qquad g(x) = \frac{1.8x}{6.5 - x - x^2}.$$

a. Find the x and y-intercepts for both of these functions. Find the slope of the linear function, $f(x)$. List any asymptotes (vertical and horizontal) for the rational function, $g(x)$.

For the linear function, $f(x)$:

x-intercept = _____, y-intercept = _____, and Slope = _____

For the rational function, $g(x)$:

x-intercept = _____ and y-intercept = _____

Vertical Asymptotes: x = _____, _____

Horizontal Asymptote: y = _____

b. Find all points of intersection between the graphs of $f(x)$ and $g(x)$.

Points of intersection for $f(x)$ and $g(x)$,

(_____, _____), (_____, _____), and (_____, _____).

c. In the Lab Report, create a graph of these functions for $x \in [-10, 10]$ with the range restricted so that $y \in [-10, 10]$. On the graph, use points to highlight the points of intersection and clearly, show and label the vertical and horizontal asymptotes with dashed lines. Briefly discuss how the vertical and horizontal asymptotes are found. Describe how the points of intersection are found.

Problem 3.10 (Rational and Quadratic Functions)**.** *(quadrat.pg) This problem explores properties of a quadratic and rational function. The points of intersection are found, and graphs of these functions are created.*

Consider the quadratic function

$$f(x) = x^2 - 3x - 5$$

and the rational function

$$g(x) = \frac{20x}{1.4 + x}.$$

a. Find the x and y-intercepts for both of these functions. Find the vertex of the quadratic function, $f(x)$. List any asymptotes (vertical and horizontal) for the rational function, $g(x)$.

For the Quadratic Function:

x-intercepts = _____, _____ and y-intercept = _____

The vertex (x_v, y_v) = (_____, _____)

For the Rational Function:

x-intercept = _____ and y-intercept = _____

Vertical Asymptote: x = _____

Horizontal Asymptote: y = _____

b. Find all points of intersection between the graphs of $f(x)$ and $g(x)$.

Points of intersection for $f(x)$ and $g(x)$,

(_____, _____), (_____, _____), and (_____, _____).

c. In the Lab Report, create a graph of these functions for $x \in [-10, 10]$ with the range restricted to $y \in [-20, 50]$. On the graph, use circles to highlight the points of intersection, and clearly, show and label the vertical and horizontal asymptotes with dashed lines.

6.3 MICHAELIS-MENTEN ENZYME KINETICS

Proteins are considered the primary building blocks of life, and *enzymes* are an important class of proteins that catalyze many of the reactions occurring inside the cell. An enzyme facilitates a biochemical reaction with both speed and specificity to break down or build up key biochemical components. The most basic enzymatic reaction, known as the *Michaelis-Menten* mechanism[12], has an enzyme, E, combine with a substrate, S, to form a product, P:

$$\mathrm{E+S} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathrm{ES} \xrightarrow{k_2} \mathrm{E+P}.$$

The rates described in the biochemical equation above can be described in a mathematical equation known as a differential equation.

It has been observed that the enzyme-substrate forms very rapidly, while the forward reaction (also known as *turnover number*), k_2, occurs more slowly. This allows a simplifying assumption for a Michaelis-Menten (MM) enzyme to describe the rate of the forward reaction to create the product, P, as

$$V([S]) = V = \frac{V_{max}[S]}{K_m + [S]},$$

where $[S]$ is the concentration of the substrate. This rate function is a rational function with important constants, V_{max} and K_m, which characterize a MM enzyme. V_{max} is the maximum rate of the reaction and depends primarily on the enzyme concentration and temperature, while K_m is the Michaelis-Menten constant, which characterizes a particular enzyme, and depends primarily on temperature. Measuring these constants is very important for understanding the efficiency of a MM enzyme. In this form, a nonlinear least squares fit to the data is necessary for finding the constants, V_{max} and K_m.

Lineweaver-Burk plot

The rational function describing the forward reaction of a MM enzyme is a nonlinear function, which complicates finding V_{max} and K_m from experiments. However, Lineweaver and Burk[7] discovered an easy transformation to change the nonlinear relationship into a linear function. Previously, when logarithms of the data appeared linear, then an allometric model was appropriate. In this case, the transformation of the data uses inverses. We define the change of variables:

$$x = \frac{1}{[S]} \qquad \text{and} \qquad y = \frac{1}{V([S])} = \frac{1}{V},$$

so x is the inverse of the substrate concentration and y is the inverse of the forward reaction rate. The transformation gives:

$$\begin{aligned} \frac{1}{y} &= \frac{V_{max}(1/x)}{K_m + (1/x)}, \\ y &= \frac{K_m}{V_{max}}x + \frac{1}{V_{max}}, \end{aligned}$$

which is a linear relationship between x and y. It follows that if the plot of these transformed variables, the *Lineweaver-Burk plot*, is approximately linear, then the enzyme is likely a MM enzyme and a *linear least squares best fit* provides a means to obtain V_{max} and K_m. This method has long been used by biochemists to find the parameters V_{max} and K_m to characterize an enzyme.

Problem 3.11 (Michaelis-Menten Enzyme Kinetics)**.** *(mm_enzyme.pg) Many enzymes follow Michaelis-Menten kinetics, and this problem takes data from an enzymatic reaction (demethylase) to fit the classic rational reaction formula for enzymes.*

Table 3.6 provides data from Schmider *et al.*[14] for cytochrome P450 mediated demethylation of amitriptyline (AMI) to nortriptyline (N) by human liver microsomes. From this table, let $[S] =$ [AMI] and $V([S]) =$ N formation. We want to find the best values of V_{max} and K_m to characterize the enzyme cytochrome P450.

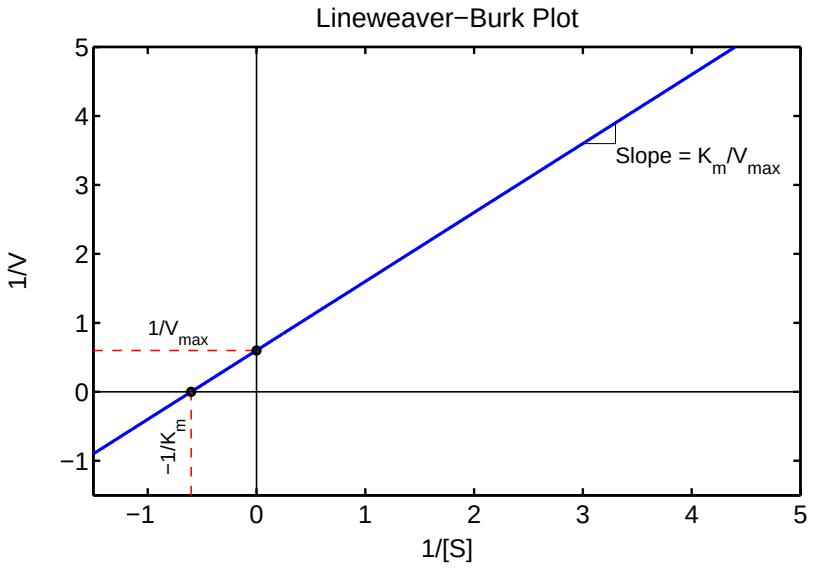

Figure 3.6: Graph for the Lineweaver-Burk function obtained from the Michaelis-Menten rate function.

[AMI] (μM)	N formation (nmol/min/mg)	[AMI] (μM)	N formation (nmol/min/mg)
5	0.20	100	2.17
10	0.45	200	2.68
15	0.60	260	2.89
25	0.80	330	3.05
50	1.35	500	3.12

Table 3.6: Data for formation of nortriptyline (N) by human liver microsomes from cytochrome P450 mediated demethylation of amitriptyline (AMI)[14].

a. The next step in this analysis is to take the data in the table above and create a new table with the values $x = 1/[S]$ and $y = 1/V([S])$. Create this new table, then plot the data y vs x. Use a linear least squares technique to find the straight line that best fits these transformed data.

x-intercept = _____ and y-intercept = _____

The slope = _____

Use this information to find V_{max} and K_m.

V_{max} = _____(nmol/min/mg) and K_m = _____(μM)

b. Use the values of V_{max} and K_m calculated above to obtain the model predications from the MM reaction formula for the production of nortriptyline (N) when [AMI] is $[S] = 15, 50, 100, 200,$ and 500. Assuming that the actual data are the best values, then calculate the percent error

between the model and the actual data. Fill in the following table:

$[S]$	$V([S])$	Percent Error
15		%
50		%
100		%
200		%
500		%

c. In the Lab Report, create a Lineweaver-Burk plot showing the transformed data and the best fitting linear model through these data. Write a brief paragraph to describe how well the linear model fits the data.

d. The transformation results in certain biases of the data. In the section for *allometric modeling*, techniques were developed to find the nonlinear least squares best directly to the model. Create a MATLAB function for the MM reaction dynamics where the parameter p in the function substitutes for the parameters V_{max} and K_m, *i.e.*,

$$V([S];p) = \frac{p(1)[S]}{p(2)+[S]} = \frac{V_{max}[S]}{K_m+[S]},$$

and this function computes the sum of square errors between the data and the MM reaction model. (Hint: Simply exchange the allometric model for the MM model in the given program.) Use the values from Part a for initial guesses for $p(1) = V_{max}$ and $p(2) = K_m$. The MATLAB routine *fminsearch* is used to find the best fitting parameters V_{max} and K_m to this nonlinear least squares routine. Record both this sum of square errors with the parameters V_{max} and K_m from both Part a and the *Least Sum of Square Errors* found by MATLAB's *fminsearch*.

$V_{max} =$ _____(nmol/min/mg) and $K_m =$ _____(μM)

Sum of Square Errors (from Part a) = _____

Least Sum of Square Errors (*fminsearch*) = _____

Find all intercepts and any asymptotes for this rational function for $[S] \geq 0$. (If no asymptote exists, then write 'None,')

$[S]$-intercept = _____and V-intercept = _____

Vertical Asymptote: $[S] =$ _____

Horizontal Asymptote: $V =$ _____

e. Use the values of V_{max} and K_m calculated in Part d to obtain the model predications from the MM reaction formula for the production of nortriptyline (N) when [AMI] is $[S] = 15, 50, 100, 200,$ and 500. Assuming that the actual data are the best values, then calculate the percent error between the model and the actual data. Fill in the following table:

$[S]$	$V([S])$	Percent Error
15		%
50		%
100		%
200		%
500		%

f. In the Lab Report, create a plot showing the original data. On this plot add the two MM models, which use the the parameters from Parts a and d. (Be sure to label which plot goes with which model.) Write a brief paragraph to describe how well each of the models fits the data and decide which model is better. Include the information from comparing the sum of square errors. Discuss some of the advantages and disadvantages of each of the models.

CHAPTER 4:
DISCRETE DYNAMICAL MODELS

Keywords: Discrete dynamical model, Malthusian growth, exponential growth, nonautonomous growth, logistic growth

This chapter introduces basic discrete dynamical models. The behavior of these models varies in complexity from simple exponential solutions to chaotic dynamics. From a computational perspective, the modeling of discrete systems with MATLAB introduces some elementary programming techniques. The models begin with simple linear models, then extend to both nonlinear discrete dynamical models and ones that include time-varying dependence. Analysis of these models can be quite complex, but finding solutions numerically is relatively easy. This chapter extends methods from before to fit model parameters to data using MATLAB.

1 DISCRETE MALTHUSIAN GROWTH

This section begins with the simplest of discrete dynamical models. Discrete dynamical models are iterations of functions at discrete time steps. *Malthusian growth*, which is often called *exponential growth* of a population, uses the concept that over a fixed period of time, each individual has a fixed probability, r, of giving birth to a new individual. That is, the new population at time $n+1$ equals the old population plus r times the population at time n. If we let the population at time n be denoted P_n, then the discrete Malthusian growth model satisfies:

$$P_{n+1} = P_n + rP_n = (1+r)P_n.$$

For this model the solution is easy to obtain. Assume that the initial population is given by P_0, then

$$\begin{aligned} P_1 &= (1+r)P_0 \\ P_2 &= (1+r)P_1 = (1+r)^2 P_0 \\ &\cdots \\ P_n &= (1+r)P_{n-1} = (1+r)^n P_0. \end{aligned}$$

It follows that given an r and P_0, then the population is known for all times n with

$$P_n = (1+r)^n P_0.$$

This expression clearly shows why Malthusian growth is also known as exponential growth. Malthusian growth occurs for most organisms when there are plentiful resources and other conditions are relatively unchanged. Note that few discrete models of the form $x_{n+1} = f(x_n)$ have a closed form solution like this.

Example 4.1 (Early U. S. Population)**.** *The population of the U. S. in its early years follows a*

pattern of Malthusian growth very well.

Let us apply the Malthusian growth model to the population in the U. S. over the first 50 years where census data are available, then see how this model works on the subsequent data. Table 4.1 of the U. S. population in millions based on the census over the last 220 years.

Year	Population	Year	Population	Year	Population
1790	3.93	1870	39.82	1950	150.70
1800	5.31	1880	50.19	1960	179.32
1810	7.24	1890	62.95	1970	203.30
1820	9.64	1900	76.21	1980	226.55
1830	12.87	1910	92.23	1990	248.71
1840	17.07	1920	106.02	2000	281.42
1850	23.19	1930	122.78	2010	308.75
1860	31.44	1940	132.16		

Table 4.1: The history of the U. S. census from 1790 to 2010. Population in millions[15].

Since the discrete Malthusian growth model has a solution, we will demonstrate **two** methods of fitting the data that are equivalent. The first method finds the nonlinear least squares best fit to the solution of this model. The second method will program the solution of the discrete Malthusian growth model, then fit it to the data. In both cases the parameters r and P_0 are found that best fit the data over 50 years.

More on Data Entry

The Table 4.1 has some blank entries, so requires some modifications to the procedure presented earlier, 18. (As noted before, it is often easier to manage the data first in spreadsheet software.) To have the population data accessible in MATLAB directly from Table 4.1, then create a variable as shown earlier, 18. **Copy** from a PDF version of the table into a variable, "data = [", and **paste** the data. Since there are two blanks, we enter "0 0];" to complete the matrix of data. The variable *data* will be an (8×6) matrix. The U. S. population data can be placed into vector, *uspop*

```
uspop = [data(:,2)',data(:,4)',data(:,6)'];
uspop = uspop(1:23);
```

Suppose we only want the population data from 1800 to 1890 and have it stored in the variable *pdata*, then the MATLAB command is

```
pdata = uspop(2:11);
```

MATLAB for Nonlinear Fit to Malthusian Growth Model

The first method follows the techniques developed in the previous chapter for finding the nonlinear least squares best fit to the data with the model given by

$$P_n = (1 + r)^n P_0,$$

where n is in decades. From the Malthusian growth model the growth in the first decade is computed by solving

$$P_1 = (1+r)P_0 \qquad \text{or} \qquad r = \frac{P_1}{P_0} - 1 = \frac{5.31}{3.93} - 1 = 0.351.$$

It follows that reasonable estimates for $P_0 = 3.93$ and $r = 0.351$. The MATLAB function for computing this sum of square errors is given by:

```
function J = sumsq_mal(p,pdata)
% Function to compute the least squares error for Malthusian growth model
N = length(pdata);          % Find dimensions of vector pdata
n = [0:N-1];                % Vector of n values
model = p(1)*(1+p(2)).^n;   % Malthusian model with p = [P_0,r]
error = model - pdata;      % Error between model and data
J = error*error';           % Computes sum of square error
end
```

Let $pdata = [3.93, 5.31, 7.24, 9.64, 12.87, 17.07]$ and define $p = [3.93, 0.351]$, then the command in MATLAB of

```
J = sumsq_mal(p,pdata)
```

gives the result $J = 0.4379$.

The proceeding MATLAB function computes the sum of square errors between the Malthusian growth model and the population data. MATLAB's *fminsearch* function finds the parameter values (P_0 and r) that give the nonlinear least squares best fit to the data. The command

```
[p1,J,flag] = fminsearch(@sumsq_mal,p,[],pdata)
```

results in the variable $p1 = [4.0180, 0.3364]$ and $J = 0.0241$, which is substantially smaller than the J-value from the initial guess. The variable $flag$ is 1, saying the routine converged to an acceptable value according to MATLAB with $P_0 = 4.0180$, $r = 0.3364$, and the sum of square errors, $J = 0.0241$.

MATLAB for Simulating Malthusian Growth Model

The discrete Malthusian growth model is an iterative model, which depends only on the previous population. It follows that a discrete population model can be written:

$$P_{n+1} = F(P_n),$$

where for the Malthusian growth model uses the linear function $F(P_n) = (1+r)P_n$. The function, $F(P_n)$, is called an *updating function*. Computer programs are particularly well-suited for this type of iterative simulation. Below we introduce the **MATLAB for-loop** to create a simple simulation of the Malthusian growth model.

```
function pop = Mal_sim(p0,r,N)
% Function simulates Malthusian growth model
%   User supplies initial population p0, growth rate r, and iterations N
```

```
pop = [p0];            % Add initial population to solution
pn = p0;               % Let pn start at p0
for i = 1:N            % Run the simulation for N iterations
    pn = (1+r)*pn;     % Update pn to new value with Malthusian growth
    pop = [pop, pn];   % Add new solution to vector solution pop
end
end
```

This function creates a vector solution, which starts with p_0 and completes N iterations of the Malthusian growth model to produce the vector solution. Notice that we could easily replace the line with the Malthusian growth model to satisfy any functional growth, including nonlinear growth, which may have no closed form solution.

In MATLAB the command

```
Pop = Mal_sim(3.93,0.351,5)
```

produces the vector solution $Pop = [3.9300, 5.3094, 7.1730, 9.6908, 13.0922, 17.6876]$, which tracks the data reasonably well, but is growing too fast. Again a least squares program using the simulation program above is needed to fit the population data with MATLAB's *fminsearch*. A minor change in the previous least squares program solves this problem.

```
function J = sumsq_mal_sim(p,pdata)
% Function to compute the least squares error for Malthusian growth model
N = length(pdata);                % Find dimensions of vector pdata
model = Mal_sim(p(1),p(2),N-1);   % Simulate Malthusian model with p = [P_0,r]
error = model - pdata;            % Error between model and data
J = error*error';                 % Computes sum of square error
end
```

The same result for the best fitting parameters is obtained using the simulation by entering the command:

```
[p1,J,flag] = fminsearch(@sumsq_mal_sim,p,[],pdata)
```

However, this result is more general because a different model that does not have a closed form solution can be used.

MATLAB for Exponential Fit

The Malthusian or exponential growth model does have the special form of an exponential function. It follows that an alternative method can be used to fit the data. Since

$$\begin{aligned} P_n &= (1+r)^n P_0, \\ \ln(P_n) &= n\ln(1+r) + \ln(P_0), \end{aligned}$$

which shows that the logarithm of the population is a linear function of n. Thus, an alternative means of finding the parameters P_0 and r is using MATLAB's linear least squares best fit to the data, where we use the logarithm of the population, $\ln(P_n)$. This technique is very similar to the method used in finding parameters for an allometric model in the previous chapter. Below we present a MATLAB program for finding an exponential model.

```
function [r,a] = exponentialfit(xdata,ydata)
% Exponential fit for model y = a*exp(r*x)
% Linear least squares fit to x data and logarithm of y data
Y = log(ydata);      % Logarithm of y-data
p = polyfit(xdata,Y,1);  % Linear fit to X and Y
r = p(1);            % Value of exponent
a = exp(p(2));       % Value of leading coefficient
end
```

When this exponential fit function is applied to the U. S. population data over the first five decades, then the best exponential fit is given by $P_n = 3.9669e^{0.2939n}$. By solving the equation $1 + r = e^{0.2939}$, we obtain the best parameters, which are $P_0 = 3.9669$ and $r = 0.3416$, which are slightly different from the nonlinear least squares best fit parameters. These parameters give a sum of square errors, $J = 0.0449$. As we found for the allometric model, the nonlinear least squares and exponential fits give slightly different parameter values because the weighting of the data are different, with the exponential fit working better for early in the simulation and the nonlinear least squares fit working better over the entire range of the data. Table 4.2 shows the best fitting Malthusian and exponential growth models until 1870. This period of history satisfies the appropriate conditions for Malthusian growth. Thus, the population is expected to grow at a fairly constant rate. The small errors between the models and the census data confirm this.

Year	Census	Model 1	% Error	Model 2	% Error
1790	3.93	4.018	2.24	3.967	0.94
1800	5.31	5.370	1.12	5.322	0.23
1810	7.24	7.176	-0.88	7.140	-1.38
1820	9.64	9.590	-0.52	9.579	-0.63
1830	12.87	12.816	-0.42	12.851	-0.15
1840	17.07	17.127	0.34	17.241	1.00
1850	23.19	22.889	-1.30	23.131	-0.26
1860	31.44	30.589	-2.71	31.032	-1.30
1870	39.82	40.879	2.66	41.633	4.55

Table 4.2: This table shows the values of the Malthusian growth model (Model 1) and the exponential fit model (Model 2) with their percent error compared to the actual census data. The models were fit through the data in 1850 and can be seen to start diverging by 1870.

A script is prepared to visualize the solution compared to the data. The script assumes that a file with the census data has been saved under **uspop.mat** with a variable *uspop* storing the data.

```
xx = 1790 + 10*[0:11];      % Creates vector of t-values
load 'uspop';               % Loads file with U. S. population data
pdata = uspop(1:12);        % Assigns pdata with actual data
pmodel = Mal_sim(4.0180,0.3364,11);   % Simulates model with best parameters
pmodela = Mal_sim(3.9669,0.3416,11);  % Simulates model with exp fit

plot(xx,pdata,'ro','MarkerSize',7);    % Plots population data
```

```
hold on
plot(xx,pmodel,'b-o','MarkerFaceColor','b','MarkerSize',5,'LineWidth',1.0);
% Plots nonlinear best fit model
plot(xx,pmodela,'m-*','MarkerFaceColor','b','MarkerSize',5,'LineWidth',1.0);
% Plots exponential model
grid
legend('Census', 'Nonlinear Best Fit', 'Exponential Fit',2);
axis([1790 1910 0 100]);  % Defines limits of graph
```

Note that this script does not include the title and axis labeling or print commands, which can be found in the **Basic Graphing** chapter. If this script has the MATLAB *plot* command replaced with *semilogy*, then a plot is produced with a logarithmic scale on the y-axis. The resulting graph shows the exponential model appears as a straight line with these axes.

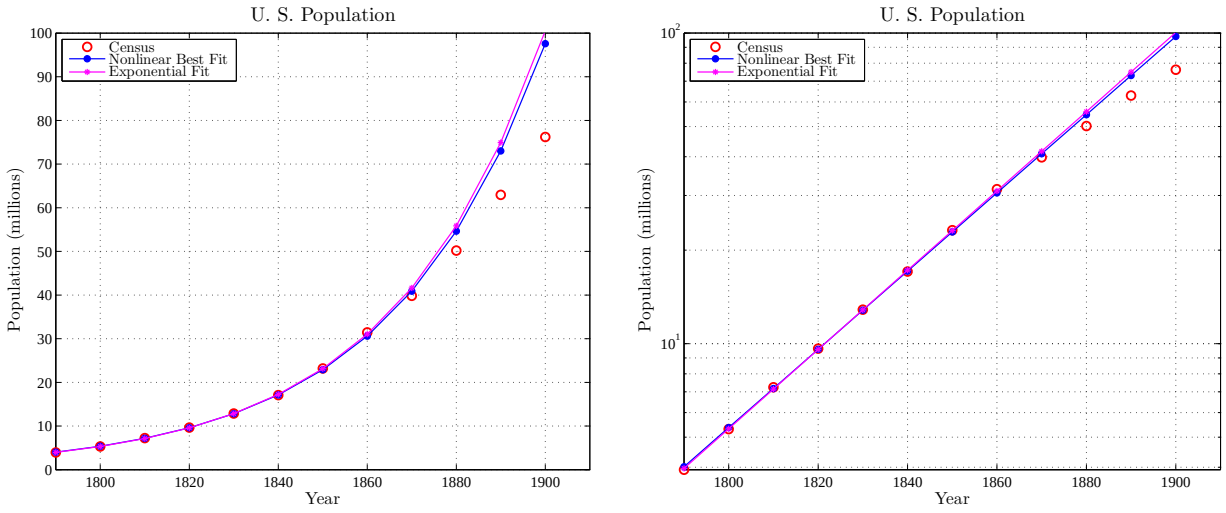

Figure 4.1: Plots of U. S. population with nonlinear least squares best fit and exponential fit. Data fit over the first 50 years, then simulated for 100 years. The graph on the right is a semilog plot.

The graph shows that the data and the model are very close from 1790 to 1850 (error $< 2.3\%$). The errors in 1860 and 1870 are only off by 2.7% (low then high). During this period of time, the society is primarily agrarian with plenty of land into which to expand (forgetting the impact on Native Americans) and stable family size, so the Malthusian growth model with its assumptions fits quite well. Growth rates of approximately 30% per decade are fairly typical of agrarian societies. Following 1870, the errors rapidly rise to over 28% by 1900. Thus, as the U. S. enters the industrial revolution, the population growth starts declining, which means that taking a constant value for r is inappropriate. We will model corrections to the growth term later, adding a time dependency.

Visualizing the Sum of Square Errors

The MATLAB programs above found the best fitting discrete Malthusian growth models using a nonlinear least sum of square errors fit and an exponential fit to the census data from 1790 to 1850. The discrete Malthusian growth model only had two parameters, initial population, P_0, and

growth rate, r, which could be varied. The MATLAB program `sumsq_mal_sim(p,pdata)` can be used to compute the sum of square errors for any set of parameters, $p = [P_0, r]$. The program below shows how to create a surface in these parameters to illustrate that the nonlinear least SSE is the best fit. This program introduces the 3D graphics capabilities of MATLAB for visualizing data.

```
% Create a least squares surface for Example 4.1
% Include points for best fit and exp fit
clear;
load 'uspop'               % Obtain census data from stored file
pdata = uspop(1:6);        % Census data from 1790 to 1850
p0=3.9+0.005*[0:40];       % Initial population from 3.9 to 4.1
r=0.32+0.0005*[0:60];      % Growth rate from 0.32 to 0.35
for i=1:41
    for j=1:61
        J(j,i)=sumsq_mal([p0(i),r(j)],pdata); % Compute SSE at each p
    end
end
surf(p0,r,J);              % Create surface over parameter range above
colormap(jet);             % Color scheme for the surf
%set(gca,'ZScale','log');  % Used for log scale of J
fontlabs = 'Times New Roman';
title('Sum of Square Errors','Fontsize',16);
xlabel('$$P_0$$','Fontsize',14); ylabel('$$r$$','Fontsize',14);
zlabel('SSE','Fontsize',14);
set(gca,'FontSize',12);
axis([3.9 4.1 0.32 0.35 0 4]); % Viewing scale in 3D
ms=6;
P1=[4.018,0.3364,0.0241];
P2=[3.9669,0.3416,0.0449];
hold on
plot3(P1(1),P1(2),P1(3),'ro','MarkerSize',ms,'MarkerFaceColor','r')
plot3(P2(1),P2(2),P2(3),'yo','MarkerSize',ms,'MarkerFaceColor','y')
hold off
view([75,35])    % normal view position
%view([150,18])  % log view
print -depsc SSE_mal_surf.eps
```

Fig. 4.2 shows the output from this program (one normal scale and another with a logarithmic scale). The surfaces illustrate how the sum of square errors varies as the parameters change. The surface with the logarithmic scale shows that the point of the nonlinear least squares fit is the lowest possible sum of square errors. With the normal scale it is easy to see that there is a long trough of values which have similar magnitude. However, going in a direction perpendicular to the trough the sum of square errors rises quite rapidly. The figure shows that the exponential fit is relatively close to the best fit, which is what we observed in Fig. 4.1.

The program above has several important new commands. The first command is the `surf(p0,r,J)`, which produces a 3D plot of the SSE, J, as the parameters P_0 and r vary. The `colormap` command

produces the color scheme for the 3D surface. Another 3D plot command is the `plot3`, which allows one to plot points in 3D. The `view` command gives the perspective for viewing the surface. The figure box allows the user to rotate the 3D figure and observe different views.

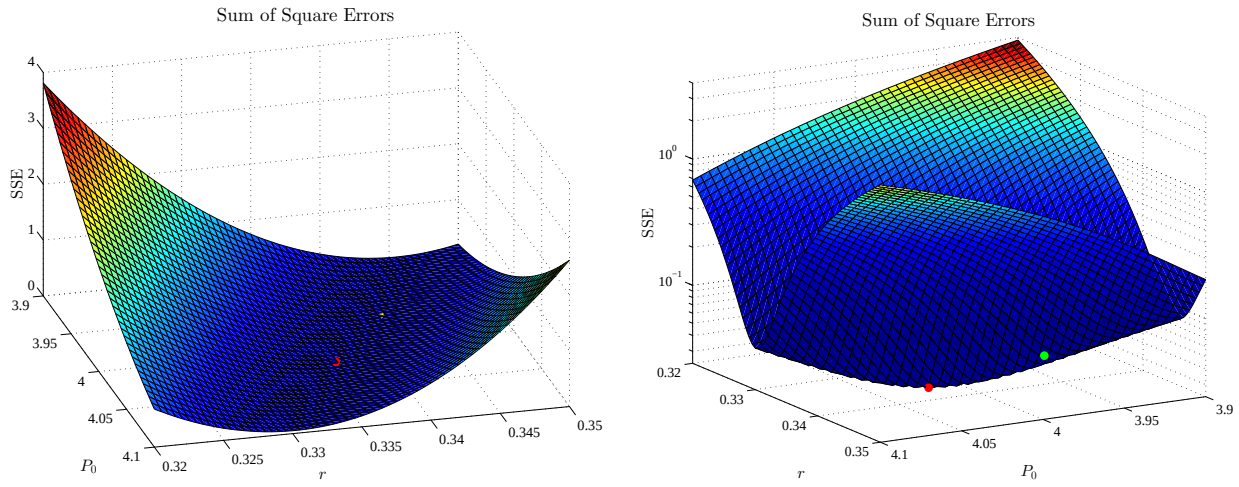

Figure 4.2: Surface plots show how the sum of square errors for the discrete Malthusian growth model as the parameters, P_0 and r, vary when compared to data from 1790 to 1850. The plot on the right has a logarithmic vertical axis to better view the least SSE. The red dot shows the SSE for the nonlinear fit to the data, while the yellow dot shows the exponential fit.

Problem 4.1 (Malthusian Growth for Two Countries)**.** *(Lab121_F2_mal_country.pg) The discrete Malthusian growth model is applied to two different periods in the history of the U. S. Parameters are fitted to the model, and graphs are made to compare the solutions to the census data.*

The population of Japan [16] was $P_0 = 116.79$ million in 1980, while in 1990 it was $P_{10} = 123.53$ million. The population in Bangladesh [16] was $Q_0 = 87.97$ million in 1980, while it was $Q_{10} = 111.45$ million in 1990.

a. Over a limited range of years, the population of most countries can be estimated using the Malthusian growth law, which is given by:

$$P_{n+1} = (1+k)P_n \qquad \text{and} \qquad Q_{n+1} = (1+r)Q_n,$$

where n is the number of years since 1980 with P_n the population of Japan and Q_n the population of Bangladesh. Use the data above to find the annual growth rate of Japan

$k =$ _____

and annual growth rate of Bangladesh

$r =$ _____.

Find the Malthusian growth model for the population of Japan,

$P_n =$ _____ million.

(The expression above uses the values of P_0 and k and depends on n in years.)

Find the Malthusian growth model for the population of Bangladesh,

$Q_n =$ _____ million.

(The expression above uses the values of Q_0 and r and depends on n in years.)

Use the Malthusian growth model to predict the population of Japan in 2000,

$P_{20} =$ _____ million.

If the actual population of Japan was 126.76 million in 2000, then the percent error between the population and the model is

Percent Error = _____ %.

Use the Malthusian growth model to predict the population of Bangladesh in 2000,

$Q_{20} =$ _____ million.

If the actual population of Bangladesh was 136.66 million in 2000, then the percent error between the population and the model is

Percent Error = _____ %.

b. Use the Malthusian growth models to determine how long it takes for each country's population to double.

Population of Japan doubles in _____ years

Population of Bangladesh doubles in _____ years

According to the Malthusian growth models, when is the population of Japan equal to the population of Bangladesh.

Populations are equal in _____ years.

c. In your Lab Report, create a graph of the Malthusian growth models for both countries from 1980 to 2030. Include the data values of the populations for 1980, 1990, and 2000 on your graph. Briefly discuss how well you believe the models track the populations of these countries over the 50 years of simulation. Include some discussion of the strengths and weaknesses of using the Malthusian growth model for this simulation.

Problem 4.2 (Malthusian Growth and U. S. Population)**.** *(Lab121_F1_US_Malthus.pg) The discrete Malthusian growth model is applied to two different periods of U. S. history. Parameters are fitted to the model, and graphs are made to compare the solutions to the census data.*

In this problem we investigate Malthusian growth models applied to the U. S. population over various time intervals. The discrete Malthusian growth model is given by

$$P_{n+1} = (1 + r)P_n, \qquad P_0 \quad \text{given},$$

where we assume that r is the **annual growth rate** and P_0 is the initial population at $n = 0$. The general solution to this model is given by

$$P_n = (1 + r)^n P_0,$$

and our goal is to find the best fitting parameters r and P_0 to this Malthusian growth model.

The models are fitted to the data in two different ways. One method of fitting uses an exponential fit to the data (linear fit to the logarithm of the population data), and the other uses a nonlinear least squares best fit to the data.

a. For our first modeling effort we consider the U. S. population in Table 4.1 from 1790 to 1870. The initial year, 1790, is associated with $n = 0$. Plot the U. S. population data from 1790 to 1870 with $n = 0$ corresponding to 1790 and n measured in years, then apply the *Exponential fit* function above to the data to create a model. This model has an exponential form, so use your knowledge of algebra to convert this exponential model into the same form as the general solution of the Malthusian growth model given above and identify the best fitting values of the initial population, P_0, and the annual growth rate, r. The best values for the initial population and best **annual growth rate** are:

$P_0 =$ _____ $r =$ _____.

Determine the sum of square errors between the Malthusian growth model and the data.

Sum of Square Errors = _____.

Use this model to find the population at 1790 and determine the percent error from the population given in Table 4.1 (assuming that the value in the Table is the most accurate):

Model Population in 1790 = _____ Percent Error in 1790 = _____%.

Also, use this model to find the population at 1840 and determine the percent error from the population given in Table 4.1 (assuming that the value in the Table is the most accurate):

Model Population in 1840 = _____ Percent Error in 1840 = _____%.

Finally, project this model to find the population at 1890 and determine the percent error from the population given in Table 4.1 (assuming that the value in the Table is the most accurate):

Model Population in 1890 = _____ Percent Error in 1890 = _____%.

b. For our second modeling effort we again consider the U. S. population in Table 4.1 from 1790 to 1870. However, this time we use MATLAB's *fminsearch* to find the nonlinear least squares best fit to the data. This directly finds the best fitting parameters P_0 and r to the general solution of the Malthusian growth model to the given data. Using this method, give the best values for the initial population and **annual growth rate**

$P_0 =$ _____ $r =$ _____.

Determine the sum of square errors between the Malthusian growth model and the data.

Sum of Square Errors = _____.

Use this model to find the population at 1790 and determine the percent error from the population given in Table 4.1 (assuming that the value in the Table is the most accurate):

Model Population in 1790 = _____ Percent Error in 1790 = _____%.

Also, use this model to find the population at 1840 and determine the percent error from the population given in Table 4.1 (assuming that the value in the Table is the most accurate):

Model Population in 1840 = _____ Percent Error in 1840 = _____%.

Finally, project this model to find the population at 1890 and determine the percent error from the population given in Table 4.1 (assuming that the value in the Table is the most accurate):

Model Population in 1890 = _____ Percent Error in 1890 = _____%.

c. In your Lab Report, plot the data for 1790 to 1890. Include the graph of both Malthusian growth models found in Parts a and b. In addition, create a graph with a logarithmic scale on the population axis, like in Fig. 4.1. Write a brief paragraph describing how well the models fit the data. Which model fits the data better? Briefly describe any discrepancies that you observe on the graph between the models and the data and put these errors in the context of what you know about U. S. history. Are there significant differences in these two models? What are the advantages and disadvantages of using one model over the other?

d. We repeat the modeling effort above but use the the U. S. population in Table 4.1 from 1840 to 1920. Now let the initial year, 1840, be associated with $n = 0$. Plot the U. S. population data from 1840 to 1920 with $n = 0$ corresponding to 1840 and n measured in years, then apply the *Exponential fit function* above to the data to create a model. This model has an exponential form, so convert this exponential model into the same form as the general solution of the Malthusian growth model given above and identify the best fitting values of the initial population, P_0, and the growth rate, r. The best values for the initial population and annual growth rate are:

P_0 = _____ r = _____.

Determine the sum of square errors between the Malthusian growth model and the data.

Sum of Square Errors = _____.

Use this model to find the population at 1840 and determine the percent error from the population given in Table 4.1 (assuming that the value in the Table is the most accurate):

Model Population in 1840 = _____ Percent Error in 1840 = _____%.

Also, use this model to find the population at 1890 and determine the percent error from the population given in Table 4.1 (assuming that the value in the Table is the most accurate):

Model Population in 1890 = _____ Percent Error in 1890 = _____%.

Finally, project this model to find the population at 1940 and determine the percent error from the population given in Table 4.1 (assuming that the value in the Table is the most accurate):

Model Population in 1940 = _____ Percent Error in 1940 = _____%.

e. Now consider the U. S. population in Table 4.1 from 1840 to 1920. This time we use MATLAB's *fminsearch* to find the nonlinear least squares best fit to the data. This directly finds the best fitting parameters P_0 and r to the general solution of the Malthusian growth model to these data. Using this method, give the best values for the initial population and **annual growth rate**

$P_0 =$ _____ $r =$ _____.

Determine the sum of square errors between the Malthusian growth model and the data.

Sum of Square Errors = _____.

Use this model to find the population at 1840 and determine the percent error from the population given in Table 4.1 (assuming that the value in the Table is the most accurate):

Model Population in 1840 = _____ Percent Error in 1840 = _____%.

Also, use this model to find the population at 1890 and determine the percent error from the population given in Table 4.1 (assuming that the value in the Table is the most accurate):

Model Population in 1890 = _____ Percent Error in 1890 = _____%.

Finally, project this model to find the population at 1940 and determine the percent error from the population given in Table 4.1 (assuming that the value in the Table is the most accurate):

Model Population in 1940 = _____ Percent Error in 1940 = _____%.

f. In your Lab Report, plot the data for 1840 to 1940. Include the graph of both Malthusian growth models found in Parts d and e. Also, create a graph with a logarithmic scale on the population axis, like in Fig. 4.1. Write a brief paragraph describing how well the models fit the data. Which model fits the data better? Briefly describe any discrepancies that you observe on the graph between the models and the data and put these errors in the context of what you know about U. S. history over this period of time. What are the differences and similarities of these modeling efforts compared to the growth models over the earlier period studied at the beginning of this Computer Lab? Which range of data provides the better information using these models for predicting next two decades and why?

2 OTHER GROWTH MODELS: NONAUTONOMOUS, LOGISTIC, AND IMMIGRATION

The growth of the population in the U. S. has not continued at a constant rate. In fact, most populations have a decline in growth rate as populations increase. For human populations the decline in growth rate often results from improved health care and education or from severe crowding and disease. Studies have shown that improved education of women results in the largest declines in growth rates because of better family planning and delayed reproduction. Technological advances, especially in medicine, cause reproduction rates to decline as families are more assured that their offspring will survive. It follows that a simple Malthusian growth rate cannot track populations for extended periods of time.

This section introduces three very important discrete dynamical population models. For human populations, the growth rate often depends on the time in history that the population is being

studied. This leads to the time dependent discrete dynamical model, which has the form:

$$P_{n+1} = F(t_n, P_n).$$

This very general form implies that the population at the next discrete time depends on both the time, t_n, and previous population, P_n. This population model is known as a *nonautonomous discrete population model*. Our study below examines the simplest nonautonomous model, where $F(t_n, P_n)$ is the product of a linear time varying growth rate times the current population or $F(t_n, P_n) = (a + bt_n)P_n$.

Animal populations often depend only on the existing population with limited time dependence. This leads to the general *autonomous discrete population model*:

$$P_{n+1} = F(P_n),$$

which states that the population at the next time only depends on the current population. One of the most important autonomous discrete population models is the *logistic growth model*. This model is very commonly used in animal studies, where the animal population is limited by resources. This model often demonstrates a saturation effect, where the population levels off at a *carrying capacity*. The logistic growth model assumes a Malthusian growth at low population density, then a decline in the growth rate for higher densities using a quadratic form. This logistic growth model has the form:

$$F(P_n) = aP_n - bP_n^2 = P_n + rP_n\left(1 - \frac{P_n}{M}\right),$$

where the linear term $a = 1 + r$ as in the Malthusian growth model and the quadratic term $b = \frac{r}{M}$. The functional form on the right above assumes a Malthusian growth rate of r with a carrying capacity of M.

An important feature of *autonomous growth models* is the existence of population *equilibria*. A population, P_e, is at *equilibrium* if successive populations remain the same, so $P_n = P_{n+1} = P_e$. The equilibria for a discrete dynamical model reduces to solving the algebraic equation

$$P_e = F(P_e).$$

In a *closed ecological model* (no immigration or emigration), there is almost always the *extinction equilibrium*, $P_e = 0$. The logistic growth model has two equilibria, the extinction equilibrium and the equilibrium at carrying capacity. The equilibria are found by solving:

$$P_e = P_e + rP_e\left(1 - \frac{P_e}{M}\right).$$

It follows that

$$0 = rP_e\left(1 - \frac{P_e}{M}\right) \qquad \text{or} \qquad P_e = 0, M,$$

assuming $r \neq 0$. Thus, the logistic growth model has a non-extinction equilibrium at the carrying capacity, $P_e = M$.

Example 4.2 (U. S. Population)**.** *This example studies the population of the U. S. from 1790 to present by fitting Malthusian, nonautonomous, and logistic growth models to the census data. The models are compared and examined to see what predictions result for the future.*

This example begins with the same programs and analysis of Example 4.1. In this case, we fit the entire U. S. census data with the discrete Malthusian growth model:

$$P_{n+1} = (1 + r)P_n, \qquad P_0 \quad \text{given},$$

where we assume that r is the **decade growth rate** and P_0 is the initial population at $n = 0$. Starting with any reasonable guesses for $p = [P_0, r]$ and letting *pdata* be the entire set of census data for the U. S., we invoke MATLAB's *fminsearch* to find the best fitting parameters. By typing,

```
[p1,J,flag] = fminsearch(@sumsq_mal,p,[],pdata)
```

results in the variable $p1 = [16.3456, 0.1460]$ and $J = 2875.2$, so $P_0 = 16.3456$ and $r = 0.1460$. The best fitting Malthusian growth model over the entire U. S. census data is

$$P_{n+1} = 16.3456 \cdot (1.1460)^n,$$

where n is in decades after 1790. This model starts with an initial population, which is not achieved in the actual census data until almost 50 years later. The growth rate is substantially below the average growth rate ($r_{ave} = 0.2233$). Thus, this model tracks the U. S. census data rather poorly, which is seen in the sum of square errors.

A nonautonomous growth model can be expressed as

$$P_{n+1} = (1 + k(t_n))P_n, \qquad P_0 \quad \text{given},$$

where $k(t_n) = a + b\,t_n$ and $t_n = 10\,n$ with n and t_n being decades and years after 1790, respectively. Note that this model has one more parameter than the simple Malthusian growth model. The function $k(t_n)$ is a linear function, so we need to find the linear least squares best fit to the growth rate of the U. S. population data. The growth rate (per decade), g, between successive census dates is computed by

$$g = \frac{P_{n+1}}{P_n} - 1.$$

We create an array of the growth rate data from the population *data* using the MATLAB function:

```
function g = growth_rate(data)
% Create an array of growth rate data
N = length(data);
g = [];            % Null array to store growth rate
for i=1:N-1
    gr = data(i+1)/data(i)-1; % Find growth rate
    g = [g, gr];   % Array of growth rate data
end
```

With the growth data, MATLAB's linear fit routine

```
polyfit(tt,gr,1)
```

gives the best fitting line through the population data, where *tt* are the years after 1790 and *gr* are the computed growth rates, associated with the time at P_n. The best fitting line is given by

$$k(t_n) = -0.0014387\,t_n + 0.37436,$$

with $t_n = 10\,n$ and n being the decades after 1790. Fig. 4.3 shows a graph of $k(t_n)$ with the growth rate data and the mean growth rate. (The mean growth rate is $r_{ave} = 0.2233$ and is easily computed in MATLAB with the command *mean(gr)*.)

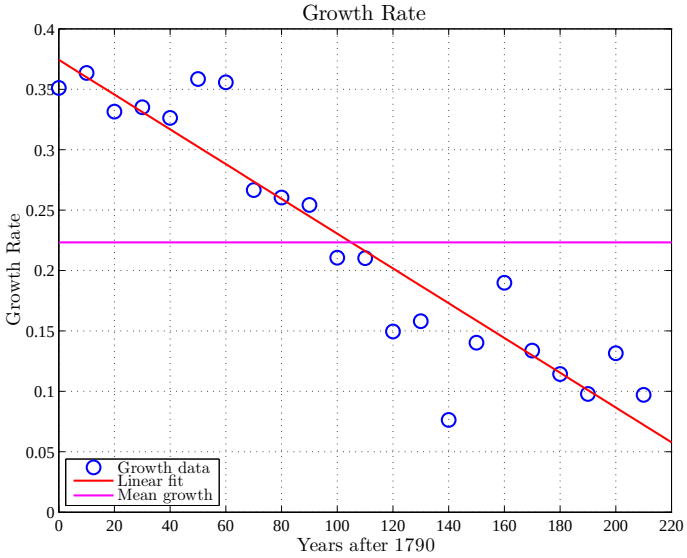

Figure 4.3: Graph showing the grow rate for the U. S. for each decade after 1790. The best linear fit is shown along with the mean growth rate.

From the information above, the best fitting nonautonomous model with the growth data satisfies the equation:

$$P_{n+1} = (1.37436 - 0.014387\,n)P_n.$$

This model is easily simulated with the MATLAB function

```
function pop = nonauto_sim(p0,gr,N)
% Simulates nonautonomous growth model
%   Insert initial population p0, growth parameters gr, and iterations N
pop = [p0];
pn = p0;
for i = 1:N-1
    pn = (1+gr(2)+gr(1)*(i-1))*pn;
    pop = [pop, pn];
end
end
```

where $gr = [0.37436, -0.014387]$.

We noted in Example 4.1 that the techniques for simulating and fitting a more general discrete dynamical model were very similar, requiring few modifications to the programs listed in the MATLAB for Simulating Malthusian Growth Model (p. 55). If we fix the growth function $k(t_n)$ and find the nonlinear least squares best fit of the nonautonomous model to the census data with *fminsearch* by varying only P_0, then the best least squares fit gives $P_0 = 3.7588$ with the sum of

square errors being 739.97, which is substantially smaller than for the Malthusian growth model. Since this model has one additional parameter, it is expected to fit better, and the sum of square errors shows a significantly better fitting of this model to the census data.

The logistic growth model is the final model fit to the U. S. census data in this example. This model has the two parameters, growth, r, and carrying capacity, M. As in the previous two models, the initial population, P_0, is also fit in the nonlinear least squares fit. The logistic growth model is easily simulated with the MATLAB function:

```
function pop = log_sim(p,N)
% Simulates logistic growth model
%   Insert p(1) initial population, p(2) growth, and carrying capacity p(3)
pop = [p(1)];
pn = p(1);
for i = 1:N-1
    pn = pn + p(2)*pn*(1 - pn/p(3));
    pop = [pop, pn];
end
end
```

As before, the sum of square errors is computed with a MATLAB function of the form:

```
function J = sumsq_log_sim(p,pdata)
% Function to compute the least squares error for logistic growth model
N = length(pdata);                % Find dimensions of vector pdata
model = log_sim(p,N);   % Simulate logistic model with p = [P_0,r,M]
error = model - pdata;            % Error between model and data
J = error*error';                 % Computes sum of square error
end
```

The least sum of square errors is computed by the MATLAB command:

```
[p1,J,flag] = fminsearch(@sumsq_log_sim,p,[],uspop)
```

Reasonable initial guesses were $p = [4, 0.25, 400]$. The best fitting parameters for the U. S. census data are found to be $p1 = [8.5751, 0.22446, 451.725]$, which implies $P_0 = 8.5751$, $r = 0.22446$, and $M = 451.725$. The sum of square errors is $J = 557.37$. This implies that the logistic growth model fits the census data best of these three models. The census data and the models are presented in Fig. 4.4.

The models in the graph are simulated for an additional 60 years to show the differences in the projected future populations for the U. S. Not surprisingly, the Malthusian growth model grows rapidly to expand beyond the range of the graph with its exponential growth. This model was not expected to predict the future population of the U. S. very well. Both of the other models are shown to fit the census data quite well over the range of the simulations, including very close fits to the actual censuses taken in 2000 and 2010. The nonautonomous model fits the early census data very well, while the logistic growth model does better in the 20^{th} century. However, the nonautonomous and logistic growth models predict very different futures over the next century. Consider the nonautonomous growth function $k(t_n)$ and solve $k(t_n) = 0$. This is the predicted year after 1790 when the growth rate goes to zero and the population starts declining. The solution is

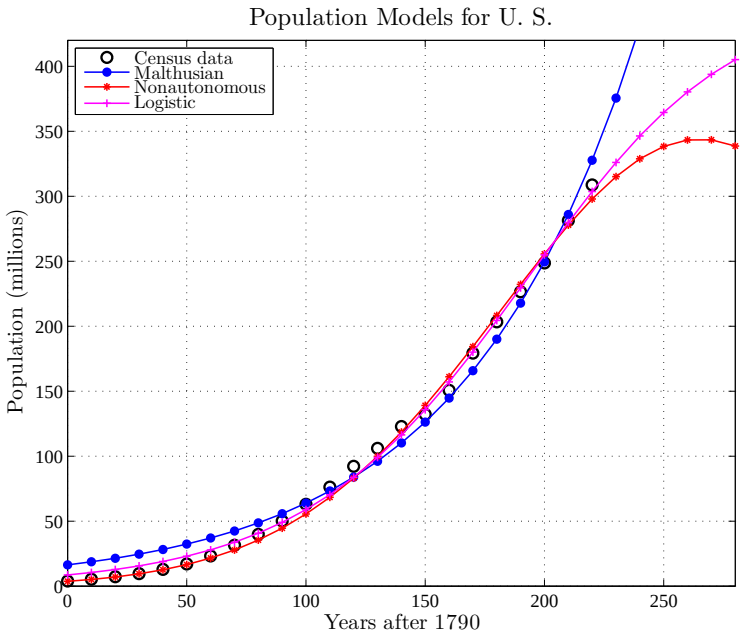

Figure 4.4: Graph showing the population for the U. S. for each decade after 1790. The best fitting Malthusian, nonautonomous, and logistic models are shown.

$t_n = 260.21$, which would be shortly after 2050. It follows that the nonautonomous model predicts that the population of the U. S. will achieve a maximum of a little over 343.4 million people between 2050 and 2060, then begin to decline. A number of developed countries around the world have seen this maximum population occur recently with Italy being the first country to achieve this milestone in 2001. Fig. 4.4 shows the nonautonomous model peaking at this predicted population 260-270 years after 1790.

The logistic growth model shows the population of the U. S. continuing to rise, but the growth rate declining to zero. The best fitting parameters give the carrying capacity, $M = 451.7$, which means that this model predicts the population of the U. S. to level off in the future around 450 million people. Fig. 4.4 shows the population at 405 million in 2070 with significantly higher predictions than the nonautonomous model for most of this century. As noted earlier, it is generally expected that a nonautonomous model for human growth is better than the logistic growth model. However, neither model includes a term for immigration, which is significant for the U. S. population. Hence, better models are needed to obtain an accurate prediction for the future, which is one of the goals of the U. S. Census Bureau. Their combination of numerous models with careful statistical analysis, including the age structure of the population, provides the best future predictions. However, the nonautonomous and logistic growth models do give some reasonable estimates of the future.

Problem 4.3 (Nonautonomous Growth and Vietnam's Population)**.** *(Lab121_F4_nonauto.pg) A discrete nonautonomous growth model is applied to census data for Vietnam from 1950 to 2000. The model assumes a declining linear growth, and it is used to predict future populations for Vietnam.*

Using data from the U. S. census bureau [16], the table below presents the population (in

millions) for Vietnam. Human populations tend to have growth rates vary with time because of advancement of living conditions from technology. Below we develop a nonautonomous model for the population growth of Vietnam based on the following data.

Year	Population	Year	Population
1950	25.34	1980	53.66
1960	31.66	1990	66.66
1970	42.59	2000	78.53

a. We begin our modeling effort by finding the growth rate of Vietnam for each decade. Let 1950 correspond to $t = 0$. Find the growth rate for each decade using the data above by dividing the population from one decade by the population of the previous decade and subtracting 1 from this ratio. Associate each growth rate with the earlier value of t from the two census dates. Find the growth rate at $t = 30$ (the decade 1980-1990).

Growth at $t = 30$ is _____.

Determine the average (mean) growth rate, r_{ave}, (per decade) from the data above.

Average growth rate $r_{ave} =$ _____.

Find the best straight line through the growth data

$$k(t) = a + bt,$$

where

$a =$ _____ and $b =$ _____.

Assuming that the data produces the best value, find the percent error between the model growth rate, $k(t)$, and the actual growth rate at $t = 30$.

Percent Error = _____%

b. In your Lab Report, plot the growth data for Vietnam as a function of t (years after 1950). Include on this graph a line representing the average growth rate, r_{ave}, over the 50 years of data and the best fitting line through the data, $k(t)$. Briefly discuss what you observe about the growth rate for Vietnam. How well does $k(t)$ fit the growth rate data?

c. The Discrete Malthusian growth model is given by

$$P_{n+1} = (1 + r_{ave})P_n, \qquad P_0 = 25.34.$$

where r_{ave} is the average growth rate (computed above) and P_0 is the population in 1950 $t = 0$. Write the general solution to this model, where n is in decades. Your answer should include the values for r_{ave} and P_0, but depend on the variable n.

$P_n =$ _____

Find the sum of square errors between this discrete Malthusian growth model and the data in the Table.

Sum of Square Errors for discrete Malthusian model = _____

Use this model to estimate the populations in 1960, 1980, and 2000. Then with the actual population being accurate, find the percent error from this Malthusian growth model

P_1 = _____ Percent error = _____%

P_3 = _____ Percent error = _____%

P_5 = _____ Percent error = _____%

Use the model to predict the population in 2020 and 2050.

P_7 = _____

P_{10} = _____

c. The nonautonomous Malthusian growth model is given by

$$P_{n+1} = (1 + k(t_n))P_n, \qquad P_0 = 25.34.$$

where $k(t_n)$ is computed above (with $t_n = 10n$) and P_0 is again the population in 1950. Simulate this nonautonomous discrete dynamical model from 1950 to 2050 ($t_n \in [0, 100]$). Find the sum of square errors between this nonautonomous Malthusian growth model and the data in the Table.

Sum of Square Errors for nonautonomous Malthusian model = _____

Use this model to estimate the populations in 1960, 1980, and 2000. Then with the actual population being accurate, find the percent error from this nonautonomous Malthusian growth model

P_1 = _____ Percent error = _____%

P_3 = _____ Percent error = _____%

P_5 = _____ Percent error = _____%

Use the model to predict the population in 2020 and 2050.

P_7 = _____

P_{10} = _____

d. The nonautonomous Malthusian growth model has a declining linear growth rate. Hence, its growth rate will eventually go to zero, then become negative. According to this model, the population of Vietnam achieves a maximum during the decade following the time when the growth rate is zero. Use the growth rate function $k(t)$ to predict when the growth rate becomes zero.

$k(t_M) = 0$ when t_M = _____

Simulate the nonautonomous Malthusian growth model until it achieves a maximum, and list the maximum value from the simulation along with the value of t_n when that maximum occurs.

Maximum population = _____ with t_n = _____

e. In your Lab Report, create a graph of the population data for Vietnam. Show the Malthusian growth and the nonautonomous Malthusian growth models on the this graph with modeling projections for the period from 1950 to 2050. Briefly discuss how well these models predict the population over this period. How believable are the modeling projections for the future? List some strengths and weaknesses of each of the models and how you might obtain a better means of predicting the future population of Vietnam.

Problem 4.4 (Discrete and Logistic Growth Models)**.** *(Lab121_G1_bac_grow) Discrete Malthusian and Logistic growth models are simulated and analyzed.*

This problem examines Discrete Malthusian and Logistic growth models, which are appropriate for studying simple organisms over limited time periods. The Malthusian growth model is given by the equation:

$$B_{n+1} = B_n + r\,B_n = (1+r)B_n,$$

where n is the time in minutes and r is the rate of growth. The Logistic growth equation is given by

$$B_{n+1} = B_n + r\,B_n\left(1 - \frac{B_n}{M}\right),$$

where M is the carrying capacity of the population.

a. Begin with a simulation of the Malthusian growth model starting with 1000 bacteria (or $B_0 = 1000$). Assume that the growth rate $r = 0.023$/min. Write an expression for the number of bacteria at each min, n.

B_n = _____

Simulate this dynamical system for $n = 350$ min, then find the number of individuals at $n = 60$, 180, and 300 min (1, 3, and 5 hr).

B_{60} = _____ B_{180} = _____ B_{300} = _____.

How long does it take for this population to double? (Use your formula to solve exactly.)

Doubling time, t_d = _____ min.

b. Next we examine a population of bacteria that satisfies the Logistic growth law. Start again with $B_0 = 1000$ bacteria, but use a growth rate of $r = 0.026$/min. Assume that $M = 1,500,000$. Simulate this model for 350 min, then find the number of individuals after $n = 60$, 180, and 300 min (1, 3, and 5 hr).

B_{60} = _____ B_{180} = _____ B_{300} = _____.

Since we cannot solve the logistic growth model exactly, we cannot determine the exact doubling time like it is possible to do for the Malthusian growth model. Use the simulation to determine the value of n when the logistic growth model first exceeds double the initial population.

Approximate doubling time, $n_d =$ _____ min.

The logistic growth model has a carrying capacity, which is the maximum population sustainable. Find how long it takes for this model to first exceed 0.5 times the maximum population of this model.

Approximate time for achieving 0.5 of carrying capacity, $n_M =$ _____ min.

c. It is clear that the growth rate of the discrete Malthusian growth model is less than the growth rate for the discrete logistic growth model, so initially the discrete logistic growth model grows more rapidly. However, the simulations should show that eventually the discrete Malthusian growth model's population exceeds that of the discrete logistic growth model. Use your data to determine the first time that the population from Malthusian growth model exceeds the one growing according to the Logistic growth model.

Approximate time for discrete Malthusian population exceeding discrete logistic growth population, $n_e =$ _____ min.

d. In your Lab Report, create a single graph with the populations of both bacterial cultures (Malthusian and Logistic) for n from 0 to 350. (Be sure to use lines to represent these simulations and not points, labeling which line represents which model.) Write a brief paragraph describing the shapes of the two population growth curves. Which model better represents a culture of bacteria on a fixed medium? Give a brief explanation for why the population from the discrete Malthusian growth model must eventually exceed the population of the discrete logistic growth model, despite the growth rate of the discrete logistic growth model being larger.

Logistic Growth Behavior

The examples above have shown the *logistic growth model* in the classic model behavior of populations growing exponentially at low density, then leveling off at a carrying capacity as the population grows. The discrete logistic growth model exhibits this behavior whenever the growth rate is not too large. Robert May[10] found that the discrete logistic growth model had some very interesting behaviors when the growth rate, r, is increased. As noted above, when this model has a low growth rate, then the population monotonically grows to the carrying capacity. As the growth rate increases, the model next exhibits a stable oscillatory behavior with the population oscillating around, but converging to the carrying capacity. Further increases in the growth rate show the population going through a series of *period doublings*. With increasing r, the population first oscillates with a stable period of two distinct populations. Next it oscillates with a stable period of four distinct populations, then eight, sixteen, etc. Eventually, at a certain value of r, the discrete logistic growth model goes into *chaos*. The behavior becomes very unpredictable. This is a very complex behavior and was quite unexpected for such a simple model. It also helps explain why one would not expect a closed form solution to the discrete logistic growth model. The problem below explores this behavior.

Problem 4.5 (Logistic Model)**.** *(Lab121_H2_log_grow.pg) Simulations are performed to observe the behavior of the logistic growth model as it goes from stable behavior to chaos.*

This problem studies the behavior of the discrete logistic growth model as the growth parameter varies. For certain parameter values, it is possible for this discrete model to exhibit chaotic behavior. The discrete logistic growth model satisfies:

$$P_{n+1} = f(P_n) = P_n + r\,P_n\left(1 - \frac{P_n}{M}\right).$$

This problem explores some of the complications that can arise as the parameter r varies.

a. Let $M = 10000$. The first step in studying this model is to find all equilibria (where the population stays the same). Determine the equilibria, P_e, by solving

$$P_e = f(P_e).$$

The equilibria are

$P_{1e} =$ _____ and $P_{2e} =$ _____.

b. Let $r = 0.89$ with $P_0 = 400$. Simulate the discrete logistic growth model for $n = 50$ generations. Give the values from your simulations at $n = 2$, 5, 10, 20, 35, and 50.

$P_2 =$ _____, $P_5 =$ _____, $P_{10} =$ _____,

$P_{20} =$ _____, $P_{35} =$ _____, $P_{50} =$ _____.

c. Now let $r = 1.69$ with $P_0 = 400$. Simulate the discrete logistic growth model for $n = 50$ generations. Give the values from your simulations at $n = 2$, 5, 10, 20, 35, and 50.

$P_2 =$ _____, $P_5 =$ _____, $P_{10} =$ _____,

$P_{20} =$ _____, $P_{35} =$ _____, $P_{50} =$ _____.

d. Next let $r = 2.15$ with $P_0 = 400$. Simulate the discrete logistic growth model for $n = 50$ generations. Give the values from your simulations at $n = 2$, 5, 10, 20, 35, and 50.

$P_2 =$ _____, $P_5 =$ _____, $P_{10} =$ _____,

$P_{20} =$ _____, $P_{35} =$ _____, $P_{50} =$ _____.

e. Next let $r = 2.51$ with $P_0 = 400$. Simulate the discrete logistic growth model for $n = 50$ generations. Give the values from your simulations at $n = 2$, 5, 10, 20, 35, and 50.

$P_2 =$ _____, $P_5 =$ _____, $P_{10} =$ _____,

$P_{20} =$ _____, $P_{35} =$ _____, $P_{50} =$ _____.

f. Next let $r = 2.57$ with $P_0 = 400$. Simulate the discrete logistic growth model for $n = 50$ generations. Give the values from your simulations at $n = 2$, 5, 10, 20, 35, and 50.

$P_2 =$ _____, $P_5 =$ _____, $P_{10} =$ _____,

$P_{20} =$ _____, $P_{35} =$ _____, $P_{50} =$ _____.

g. Place the letter of the behavior of the logistic growth model next to each r value listed below:

_____Model behavior $r = 2.15$	A. Chaos
_____Model behavior $r = 1.69$	B. Stable, Monotonic
_____Model behavior $r = 2.51$	C. Stable, Oscillatory
_____Model behavior $r = 0.89$	D. Periodic, Period 4
_____Model behavior $r = 2.57$	E. Periodic, Period 2
	F. None of the Above

h. Find a parameter value, r, that gives an oscillation with period 3. It has been shown that if the discrete logistic growth model gives a period 3 solution, then this discrete dynamical system has gone through chaos.

Period 3 oscillation when $r =$, _____.

i. In your Lab Report, create a graph showing the simulations for $r = 0.89$ and 1.69, then write a brief description of what you observe in these solutions. Create another graph showing the simulations for $r = 2.15$ and 2.51, then write another brief description of what you observe in these solutions. Finally, create a graph showing the simulation for $r = 2.57$ and your period 3 simulation, then write a brief description of what you observe in these simulations.

Population Model with Immigration

The models described above are all considered *closed* population models. A *closed population model* means that the model accounts for all of its population from within the model, and it must have a *zero* or *extinction equilibrium*. However, when considering the populations of countries, then an important source of population change comes about from immigration or emigration. The U. S. population has long history of immigration affecting its population demographics. The problem below examines the simplest population model with immigration.

Problem 4.6 (Immigration Model)**.** *(Lab121_H3_imm_log.pg) The population of the U. S. in the 20^{th} century is fit with a discrete Malthusian growth model, a Malthusian growth model with immigration, and a logistic growth model. These models are compared for accuracy and used to project future behavior of the population.*

We have examined the discrete Malthusian, a nonautonomous Malthusian, and the logistic growth models for several countries. This question examines three population models for studying the population of the U. S. during the 20^{th} century. For comparison we include the discrete

Malthusian growth model. The two other models are a Malthusian growth model with immigration (or emigration) and the logistic growth model. Use Table 4.1 for census data for the U. S. from 1910 to 2010.

a. Consider the census data for the U. S. from 1910 to 1990 in Table 4.1. (**Note** that the calculation to find the best fitting parameters does not include the data from 2000 or 2010.) The discrete Malthusian growth model is given by

$$P_{n+1} = (1 + r)P_n,$$

where r is the decade growth rate and P_0 is the initial population in 1910. Use MATLAB's *fminsearch* to find the best fitting parameters, r and P_0. For initial estimates of the parameters, consider r to be the average growth rate over the years and P_0 to be the actual population in 1910. Compute the least sum of square errors between the census data and the discrete Malthusian growth model for the years 1910 to 1990. The best values for the initial population and growth rate (per decade) are:

$P_0 =$ _____ and $r =$ _____.

Sum of Square Errors = _____.

Use this model to find the population at 1950, 2000, and 2010, then determine the percent error from the population given in Table 4.1 (assuming that the value in Table 4.1 is the most accurate):

Model Population in 1950 = _____ and Percent Error in 1950 = _____.

Model Population in 2000 = _____ and Percent Error in 2000 = _____.

Model Population in 2010 = _____ and Percent Error in 2010 = _____.

Finally, project this model to find the population at 2050.

Model Population in 2050 = _____.

b. In the U. S., immigration has played an important role. During the 20^{th} century, it has been more or less tightly regulated and maintained a relatively constant value. The discrete Malthusian growth model with immigration is given by

$$P_{n+1} = (1 + r)P_n + \mu,$$

where we assume that r is the decade growth rate, μ is the immigration rate (assumed constant), and P_0 is the initial population in 1910. Use MATLAB's *fminsearch* to find the best fitting parameters, r, μ, and P_0. For initial estimates of r, μ, and P_0, take r to be the value computed in Part a, $\mu = 4.5$, and P_0 to be the actual population in 1910. Compute the least sum of square errors between the census data and the discrete Malthusian growth model with immigration for the years 1910 to 1990. The best values for r, μ, and P_0 are:

$P_0 =$ _____ and $r =$ _____ and $\mu =$ _____.

Sum of Square Errors = _____.

Use this model to find the population at 1950, 2000, and 2010, then determine the percent error from the population given in Table 4.1 (assuming that the value in Table 4.1 is the most accurate):

Model Population in 1950 = _____ and Percent Error in 1950 = _____.

Model Population in 2000 = _____ and Percent Error in 2000 = _____.

Model Population in 2010 = _____ and Percent Error in 2010 = _____.

Finally, project this model to find the population at 2050.

Model Population in 2050 = _____.

c. The two previous models grow without bound. One question is where the U. S. population will level off, and there are many estimates on what this might be. The logistic growth model is a classic model of population growth that grows like a Malthusian growth model for low population densities. However, the growth slows with increasing density, and the logistic growth model has the property of leveling off at the carrying capacity of the population. The logistic growth model is given by the discrete dynamical model of the form:

$$P_{n+1} = P_n + rP_n\left(1 - \frac{P_n}{M}\right),$$

where we assume that r is the growth rate, M is a significant constant relating to density dependent decreases in growth, and P_0 is the initial population in 1910. Use MATLAB's *fminsearch* to find the best fitting parameters, r, M, and P_0. For initial estimates of r, M, and P_0, take r to be the value computed in Part a, $M = 700$, and P_0 to be the actual population in 1910. Compute the least sum of square errors between the census data and the logistic growth model for the years 1910 to 1990. The best values for r, M, and P_0 are:

P_0 = _____ and r = _____ and M = _____.

Sum of Square Errors = _____.

Use this model to find the population at 1950, 2000, and 2010, then determine the percent error from the population given in Table 4.1 (assuming that the value in Table 4.1 is the most accurate):

Model Population in 1950 = _____ and Percent Error in 1950 = _____.

Model Population in 2000 = _____ and Percent Error in 2000 = _____.

Model Population in 2010 = _____ and Percent Error in 2010 = _____.

Finally, project this model to find the population at 2050.

Model Population in 2050 = _____.

The carrying capacity is an equilibrium population. The equilibria of a discrete population model is found by substituting $P_e = P_n$ and $P_e = P_{n+1}$. Make this substitution into the logistic growth

model, leaving you with a quadratic equation. Solve that quadratic equation to find two equilibria.

$P_{e1} =$ _____ and $P_{e2} =$ _____.

The larger of these is the carrying capacity.

d. In your Lab Report, create a single graph with all three models and the census data on the interval 1910 to 2050. Looking at the three models above, determine which model you believe best predicts the population for the year 2050. Which model do you believe is the best and why? What are the advantages and disadvantages of using one model over the other? Describe two ways that you could improve the best model to make a better prediction for 2050. Does the immigration rate in the second model seem realistic for the U. S.? Is the carrying capacity that you computed believable for the U. S.? How might you obtain a better estimate of the immigration rate and carrying capacity?

3 BREATHING MODEL

There are numerous other applications of discrete dynamical models, other than the population models illustrated above. The physiological process of breathing is certainly a discrete process[9]. *Pulmonary ventilation* or breathing is the first step to bringing oxygen to the cells of the body and removing the metabolic waste product, carbon dioxide. Contracting the muscles of the diaphragm results in an inflow of fresh air or *inspiration*, while relaxation of these muscles or contraction of the abdominals causes *expiration* of air with the waste product CO_2. During normal respiration, the lungs exchange about 500 ml of air 12 times a minute. This is the *tidal volume* of air inspired or expired. In young adult males, there is an *inspiratory reserve volume* of about 3000 ml that can be inspired above the tidal volume, while the *expiratory reserve volume* is about 1100 ml, which can be forcefully expired. The *vital capacity* includes all of the above yielding about 4600 ml. Well-trained athletes may have values 30-40% higher, while females generally have 20-25% less for the quantities listed above. The *residual volume* represents the amount of air that cannot be expelled even by forceful expiration and averages about 1200 ml. The *functional residual capacity* is the amount of air that remains behind during normal breathing, which amounts to 2300 ml. Fig. 4.5 diagrams these physiological quantities.

There are several respiratory diseases. The respiratory muscles can be damaged by spinal paralysis or poliomyelitis, which can decrease the vital capacity to as low as 500 ml, barely enough to maintain life. Pulmonary compliance reduces vital capacity in diseases like tuberculosis, emphysema, chronic asthma, lung cancer, chronic bronchitis, cystic fibrosis, or fibrotic pleurisy. Several of the diseases above and heart disease can cause pulmonary edema, which decreases vital capacity from fluid build up in the lungs. It is important to know about the physiological state of the lungs to determine what diseases are affecting any particular patient.

One non-invasive method of determining the tidal volume and the functional residual capacity is for the subject to breathe a mixture including an inert gas, which for this experiment is Argon (Ar). The subject breathes the mixture until the lungs are essentially filled with this mixture, then the physiologist measures the amount of the inert gas in a series of breaths after the subject is removed from the gas mixture to normal air.

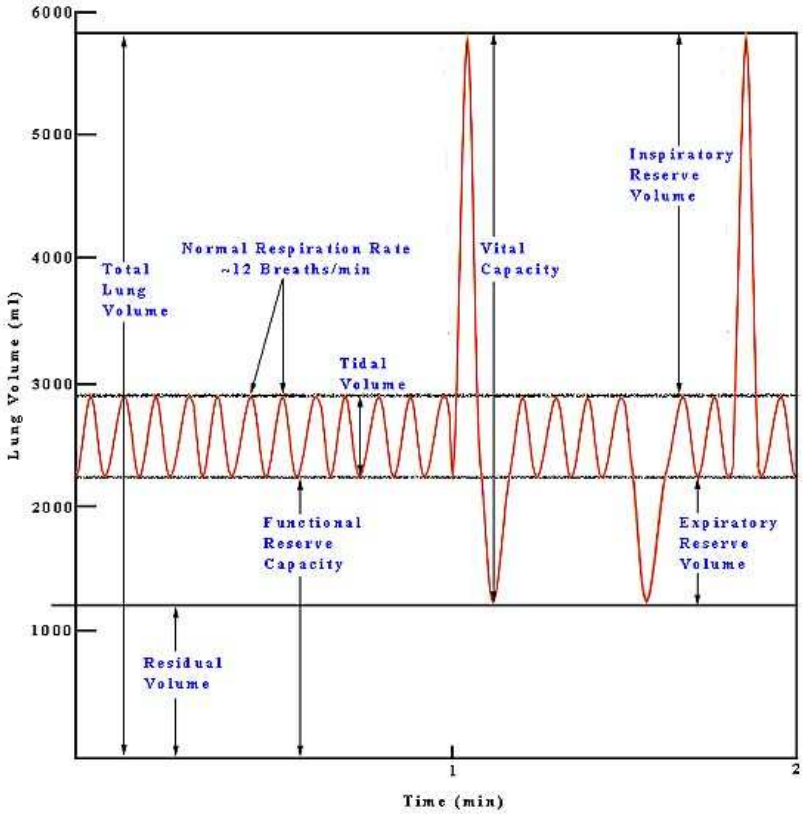

Figure 4.5: Diagram showing the graph of different lung air volumes and capacities.

Let c_n be the concentration of Ar in the n^{th} breath of the subject. The physiological parameters needed for the discrete dynamical model are V_i for the tidal volume (air normally inhaled and exhaled), V_r for the functional residual volume, and γ for the concentration of Ar in the atmosphere. Define the fraction of atmospheric air exchanged in each breath, q, and fraction of air remaining in the lungs, $1-q$, as

$$q = \frac{V_i}{V_i + V_r} \qquad \text{and} \qquad 1 - q = \frac{V_r}{V_i + V_r},$$

respectively. Upon exhaling, there remains behind the functional residual volume, which contains the amount of Ar given by $V_r c_n$. The inhaled air during this cycle contains the amount of Ar given by $V_i \gamma$. (Quantities or amounts of Ar are given by the volume times concentration, and it's the amounts that are conserved.) Thus, the amount of Ar in the next breath is given by

$$V_r c_n + V_i \gamma.$$

To find the concentration in the next breath we divide by the total volume, $V_i + V_r$, then use the definition of q. Thus, the linear discrete dynamical model for breathing an inert gas is:

$$c_{n+1} = \frac{V_r c_n}{V_i + V_r} + \frac{V_i \gamma}{V_i + V_r} = (1 - q)c_n + q\gamma. \tag{4.1}$$

Problem 4.7 (Breathing Model)**.** *(Lab121_G2_breathing.pg) A linear discrete model is examined for determining vital lung functions of normal and diseased subjects following breathing an enriched source of argon gas.*

If c_n represents the concentration of the inert gas argon (Ar) in the lungs, then a mathematical model for breathing is given by the linear discrete dynamical model, (4.1), where q is the fraction of the lung volume exchanged with each breath and $\gamma = 0.0093$ (fraction of Ar in dry air) is the concentration of Ar in the atmosphere. Normal breathing usually exchanges a volume of air, known as the *tidal volume*, V_i. The space remaining in the lung after exhaling from a normal breath is known as the *functional residual volume*, V_r. The fraction of air exchanged is $q = \frac{V_i}{V_i+V_r}$.

a. Assume that a normal subject breathes an enriched mixture of air that contains 10% Ar, so that $c_0 = 0.1$ (fraction of Ar in dry air). Suppose that the tidal volume is measured at $V_i = 505$ ml for this subject, while another measurement gives the functional residual volume, $V_r = 2350$ ml. Determine the fraction of the lung volume q exchanged for this subject.

$q =$ _____

Find the concentration of Ar in the first, third, fifth, and tenth breaths

$c_1 =$ _____ $c_3 =$ _____ $c_5 =$ _____ $c_{10} =$ _____

What is the equilibrium concentration, c_e, of Ar in this subject?

$c_e =$ _____.

Determine how many breaths are required until the concentration of Ar drops below 0.01.

Number of Breaths = _____.

b. In your Lab Report, create a graph showing the concentration of Ar in the first 10 breaths.

c. A patient with emphysema is given the same mixture of Ar (so again $c_0 = 0.1$ (fraction of Ar in dry air). The tidal volume for this patient is measured at $V_i = 205$ ml. The concentration of Ar in the first breath in found to contain 0.0899 (fraction of Ar in dry air) for this patient or $c_1 = 0.0899$. Find the fraction of the lung volume exchanged q and the functional residual volume, V_r.

$q =$ _____ $V_r =$ _____.

Find the concentration of Ar in the second, third, fifth, and tenth breaths for this subject

$c_2 =$ _____ $c_3 =$ _____ $c_5 =$ _____ $c_{10} =$ _____.

What is the equilibrium concentration, c_e, of Ar in this subject?

$c_e =$ _____.

Determine how many breaths are required until the concentration of Ar drops below 0.01.

Number of Breaths = _____.

d. In your Lab Report, create a graph showing the concentration of Ar in the first 10 breaths of this subject with emphysema. Write a brief paragraph discussing the differences between the breathing of a normal subject and a patient with emphysema based on the results above.

Appendix A:
LIST OF PROBLEMS

This appendix provides a complete listing of the problems in the Lab Manual.

Problem 2.1 Two intersecting lines (A1). 9

Problem 2.2 The intersection of a line and a quadratic function (A2). 9

Problem 2.3 A quadratic function and the intersection of a line and a cubic function (C1). 10

Problem 3.1 Cricket thermometer – linear model (A3). 20

Problem 3.2 Hummingbirds and urea – linear model (B2). 23

Problem 3.3 Yeast growth – linear growth model (C3). 26

Problem 3.4 Exponential, Logarithmic, and Power Functions (E1). 28

Problem 3.5 Dog measurement study – allometric models (D3). 33

Problem 3.6 Island Biodiversity – allometric model (E2). 37

Problem 3.7 Allegheny Forest – linear and allometric models (E3). 39

Problem 3.8 Weak Acid Study – square root function (C2). 43

Problem 3.9 Line and Rational Function (D1). 47

Problem 3.10 Quadratic and Rational Function. 47

Problem 3.11 Michaelis-Menten Enzyme Kinetics – rational function. 49

Problem 4.1 Malthusian Growth – Two countries (F2). 60

Problem 4.2 Malthusian Growth – United States (F1). 61

Problem 4.3 Nonautonomous Growth – Vietnam (F4). 69

Problem 4.4 Discrete Malthusian and Logistic Growth – Bacteria (G1). 72

Problem 4.5 Logistic Growth Model – Qualitative Behavior (G1). 73

Problem 4.6 Malthusian, Immigration, and Logistic Growth Models – U. S. Census (H3). 75

Problem 4.7 Linear Discrete Model – Breathing (G2). 80

BIBLIOGRAPHY

[1] C. A. Bessey and E. A. Bessey. Further notes on thermometer crickets. *American Naturalist*, 32:263–264, 1898.

[2] C. A. Beuchat, W. A. Calder III, and E. J. Braun. The integration of osmoregulation and energy balance in hummingbirds. *Physiological Zoology*, 63:10591081, 1990.

[3] T. Carlson. Über Geschwindigkeit und Größe der Hefevermehrung in Würze. *Biochem. Z.*, 57:313–334, 1913.

[4] AIDS cases in the U. S.: HIV/AIDS Surveillance Report. U. S. Department of Health and Human Services, Centers for Disease Control, Division of HIV/AIDS, Atlanta, GA, 1993.

[5] A. E. Dolbear. The cricket as a thermometer. *American Naturalist*, 31:970–971, 1897.

[6] David N. Holvey, editor. *The Merck Manual of Diagnosis and Therapy*. Merck Sharp and Dohme Research Laboratories, Rahway, NJ, 15th edition, 1987.

[7] H. Lineweaver and D. Burk. The determination of enzyme dissociation constants. *J. Am. Chem. Soc.*, 56 (3):658666, 1934.

[8] J. M. Mahaffy and A. Chávez-Ross. *Calculus: A Modeling Approach for the Life Sciences*, volume 2. Pearson Custom Publishing, Upper Saddle River, NJ, 2006.

[9] J. M. Mahaffy and A. Chávez-Ross. *Calculus: A Modeling Approach for the Life Sciences*, volume 1. Pearson Custom Publishing, Upper Saddle River, NJ, 2nd edition, 2009.

[10] R. M. May. Deterministic models with chaotic dynamics. *Nature*, 256:165–166, 1975.

[11] J. Mazumdar. *An Introduction to Mathematical Physiology and Biology*. Australian Mathematical Society Lecture Note Series. Cambridge University Press, Cambridge, 1989.

[12] L. Michaelis and M. L. Menten. Die Kinetik der Invertinwirkung. *Biochem Z*, 49:333369, 1913.

[13] T. A. Ryan, B. L. Joiner, and B. F. Ryan. *The Minitab Student Handbook*. Duxbury Press, 1976.

[14] J. Schmider, D. J. Greenblatt, J. S. Harmatz, and R. I. Shader. Enzyme kinetic modelling as a tool to analyse the behaviour of cytochrome P450 catalysed reactions: application to amitriptyline N-demethylation. *Brit. J. Clin. Pharmoco.*, 41:593–604, 1996.

[15] United States Census Bureau: Historical Census Data. http://www.census.gov. (Accessed 2/05/2014).

[16] United States Census Bureau: International Data Base. http://www.census.gov. (Accessed 12/19/2013).

Index

absorbance, 25
Allegheny forest, 39
allometric, 30
    AIDS, 31
    Allegheny forest, 39
    dog measurements, 33
    island biodiversity, 37
apostrophe, 5
asymptote, 29
    horizontal, 28, 44
    vertical, 28, 44

biodiversity, 37
breathing, 78
breathing model, 79

Caribbean islands, 37
carrying capacity, 65
chaos, 73
closed model, 65
cricket thermometer, 20
cubic function, 10

data
    MATLAB, 18
data entry, 18
    modify, 54
discrete dynamical models, 53
dog measurements, 33

enzyme, 48, 49
equilibrium, 65
error
    absolute, 13
    percent, 15
    relative, 15
exponential
    Maple or MATLAB, 28
exponential fit, 56
exponential growth, 53
extinction equilibrium, 65

format long, 9
formic acid, 42
fsolve, 7
function
    cubic, 10
    exponential, 28
    linear, 3
    logarithmic, 28
    MATLAB, 16
    quadratic, 3
    rational, 44
    square root, 43

graph
    Maple, 7
    MATLAB, 4
    MATLAB 3D, 59
growth
    discrete Malthusian, 53
    juvenile height, 15
    logistic, 65
    nonautonomous, 65
    yeast, 26

herpetofauna, 37
horizontal asymptotes, 28, 44
hummingbird, 23

immigration, 75

Lambert-Beer Law, 25
least squares, 13
linear function, 3
linear least squares, 13
linear model, 15
Lineweaver-Burk plot, 49

logistic growth, 65
    chaos, 73

Malthusian growth, 53
Maple
    evalf, 28
    exp, 28
    fsolve, 7, 45
    graph, 7
    LinearFit, 25
    $\log_{10}$, 42
    plot, 7
    polynomial fit, 17
    simplify, 26
    solve, 7
    square root, 42
    Statistics package, 17
MATLAB
    axis square, 8
    data, 18
    exp, 28
    exponential fit, 56
    fminsearch, 32, 51, 55
    for-loop, 55
    format, 9
    function, 16
    graph, 4, 45
    graph labels, 5
    graph legend, 5
    graph print, 5
    graph text, 5, 16
    grid, 5
    linear fit, 16
    load, 20
    $\log_{10}$, 42
    log-log graph, 33
    logistic growth, 68
    mean, 67
    nonautonomous growth, 67
    nonlinear least squares, 32
    plot, 4
    plot3, 60
    polyfit, 16
    Power Law function, 32
    save, 20
    semilog plot, 58
    square root, 42
    sum of square errors, 16, 55
    surf, 59
    Symbolic Math Toolbox, 3
    who, 20
matrix, 18
mean, 67
Michaelis-Menten, 48
Michaelis-Menten kinetics, 49
model
    breathing, 79

nonautonomous discrete model, 65

period doubling, 73
pH, 43
population
    United States, 54
    Vietnam, 69
population model
    autonomous, 65
    carrying capacity, 65
    closed, 75
    immigration, 75
    logistic growth, 65
    Malthusian growth, 53
    nonautonomous, 65
power law, 30
pulmonary ventilation, 78

quadratic function, 3

rational function, 44

solve, 7
spectrophotometry, 23, 25
Symbolic Math Toolbox
    MATLAB, 3

updating function, 55
urea, 23

vertical asymptotes, 28, 44

weak acid, 43