

# Exploring modeling with data and differential equations using R

John M. Zobitz

2021-07-28



# Contents

<b>1</b>	<b>Welcome Creating this from Bookdown to Github</b>	<b>5</b>
<b>I</b>	<b>Models with Differential Equations</b>	<b>7</b>
<b>2</b>	<b>Models of rates with data</b>	<b>9</b>
2.1	Rates of change in the world: a model is born . . . . .	9
2.2	Modeling in context: the spread of a disease . . . . .	9
2.3	Model solutions . . . . .	13
2.4	Which model is best? . . . . .	13
2.5	Start here . . . . .	14
2.6	Exercises . . . . .	15
<b>3</b>	<b>Introduction to R</b>	<b>19</b>
3.1	R and RStudio . . . . .	19
3.2	First steps: getting acquainted with R . . . . .	20
3.3	Increasing functionality with packages . . . . .	21
3.4	Working with R: variables, data frames, and datasets . . . . .	22
3.5	Visualization with R . . . . .	24
3.6	Defining functions . . . . .	27
3.7	Concluding thoughts . . . . .	29
3.8	Exercises . . . . .	30
<b>4</b>	<b>Modeling with rates of change</b>	<b>33</b>
4.1	Lynx and Hares . . . . .	33
4.2	The Law of Mass Action . . . . .	37
4.3	Establishing species . . . . .	38
4.4	Other types of functional responses . . . . .	38
4.5	Exercises . . . . .	40
<b>5</b>	<b>Euler's method</b>	<b>43</b>
5.1	Defining an Algorithm . . . . .	44
5.2	Building an iterative method . . . . .	45
5.3	Euler's method applied to systems . . . . .	48
5.4	More refined numerical solvers . . . . .	50
5.5	Exercises . . . . .	53
<b>6</b>	<b>Phase lines and equilibrium solutions</b>	<b>57</b>
6.1	Equilibrium solutions . . . . .	57
6.2	Phase lines for differentail equations . . . . .	58
6.3	A stability test for equilibrium solutions . . . . .	60
6.4	Exercises . . . . .	62
<b>7</b>	<b>Coupled systems of equations</b>	<b>65</b>
7.1	Model redux: flu with quarantine . . . . .	65
7.2	Determining stability . . . . .	69
7.3	Generating a phase plane in R . . . . .	70

7.4	Exercises . . . . .	71
<b>8</b>	<b>Exact Solutions to Differential Equations</b>	<b>77</b>
8.1	Separable Differential Equations . . . . .	77
8.2	Integrating factors . . . . .	78
8.3	Guess and Check . . . . .	79
8.4	Superposition of solutions . . . . .	80
8.5	Applying guess and check more broadly . . . . .	80
8.6	Exercises . . . . .	83

## Chapter 1

# Welcome Creating this from Bookdown to Github

I stumbled on how to do this by going to this website and built this out from here. I am glad that I will be able to work on this.

Who to thank Tidyverse Augsburg students sources waterparks family kids



## Part I

# Models with Differential Equations





## Chapter 2

# Models of rates with data

### 2.1 Rates of change in the world: a model is born

The focus of this textbook is understanding *rates of change* and how you can apply them to model real-world phenomena. In addition this textbook focuses on *using* equations with data, building both your competence and confidence to construct a mathematical model from data and a context.

Perhaps you analyzed rates of change in calculus course when answering the following types of questions:

- If  $y = xe^{-x}$ , what is the derivative function  $f'(x)$ ?
- What is the equation of the tangent line to  $y = x^3 - x$  at  $a = 1$ ?
- Where is the graph of  $\sin(x)$  increasing at an increasing rate?
- What is the largest area that can be enclosed with 100 feet of fencing to enclose, with one side being along a wall?
- If you release a ball from the top of a skyscraper 500 meters above the ground, what is its speed when it impacts the ground?

The first three questions do not appear to be connected in a real-world context - but the last two questions do. For the fencing problem, perhaps a person raises chickens and wants to care for their well-being. Perhaps a rectangular pen is more aesthetically pleasing than a circular pen. In the last example the ball falling off the skyscraper assumes that acceleration of the ball is constant.

In both of these last two cases the context may reveal underlying assumptions or physical principles, which are the starting point to build a mathematical model. For the chicken coop problem the next step is to use the assumed geometry (rectangle) with the 100 feet of fencing to develop a function for the area. For the ball problem, the velocity (or the antiderivative of acceleration) can be found, from which the position function can be calculated through antidifferentiation.

Here is another approach. Let's say we have observational data and several different (perhaps conflicting) assumptions. These assumptions describe models that involve rates of change. Which model is the best one to approximate the data? The short answer: it depends. To understand why, let's take a look at a problem in context.

### 2.2 Modeling in context: the spread of a disease

Consider the data in Figure 2.1, which come from an Ebola outbreak in Sierra Leone in 2014. Ebola is a fatal disease so we can also consider the vertical axis in Figure 2.1 to represent total *infections* due to Ebola.

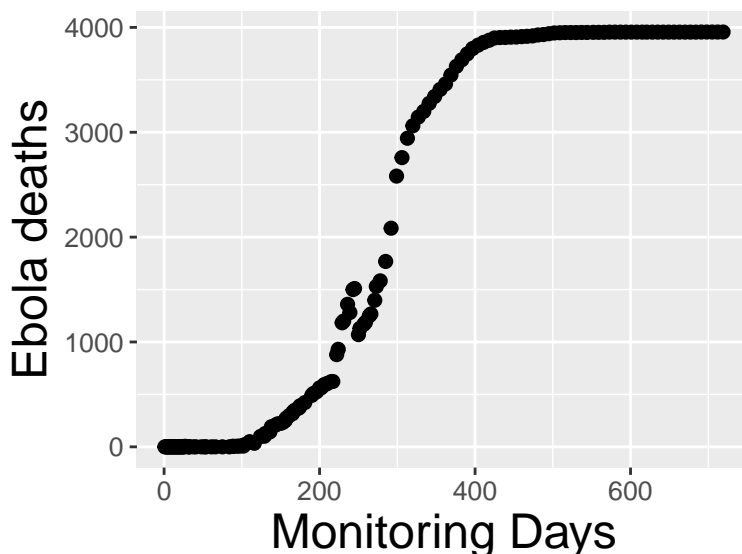


Figure 2.1: An Ebola outbreak in Sierra Leone

Constructing a model from disease dynamics is part of the field of mathematical epidemiology. How we construct a mathematical model of the spread of this outbreak largely depends on the assumptions underlying the dynamics of the disease, such as considering the rate of spread of Ebola. For our purposes here we focus on the population level (person to person) spread of Ebola. Other types of models could focus on the immune response in a single person - perhaps with a goal to design effective types of treatments to reduce the severity of infection.

Here are three initial assumptions one can make regarding the spread of Ebola:

1. The infection rate is proportional to the number of people infected.
2. The infection rate is proportional to the number of people **not** infected.
3. The infection rate is proportional to the number of infected people coming into contact with those not infected.

Let's see what each of these mathematical models would look like if we wrote down a mathematical equation. Since we are discussing *rates* of infection, this means we will need a *rate of change* or derivative. Let's use the letter  $I$  to represent the number of people that are infected.

### 2.2.1 Model 1: Infection rate proportional to number infected.

The first assumption states that the infection rate is proportional to the number of people infected. Translated into an equation this would be the following:

$$\frac{dI}{dt} = kI \quad (2.1)$$

In Equation (2.1)  $k$  can be thought of as a proportionality constant, with units of  $\text{time}^{-1}$  for consistency. Equation (2.1) is an example of a *differential equation*, which is just a mathematical equation with rates of change.

The *solution* to a differential equation is a function  $I(t)$ .<sup>1</sup> When we “solve” a differential equation we determine the family of functions consistent with our rate equation. There are a lot of techniques we can use to do that, and we will examine a few later.

Back to this proportionality constant  $k$  - another term for it is a *parameter*. We can always try to solve an equation without specifying the parameter - and then if we wanted to plot a solution the parameter would also be specified. In some situations we may not be as concerned with the particular *value* of the parameter but rather its influence on the long-term behavior of the system (this is one aspect of bifurcation theory). Otherwise we can use the collected data shown above with the given model to determine the value for  $k$ . This combination of a mathematical model with data is called *data assimilation* or *model-data fusion*.

<sup>1</sup>You may be used to working with *algebraic equations* (e.g. solve  $x^2 - 4 = 0$  for  $x$ ). In that case the solution can be points (for our example,  $x = \pm 2$ ).

Before we think about possible solutions let's try to reason out if the first model would be plausible. This model states that the rate of change (the amount of increase) gets larger the more people that are sick. While this may seem reasonable initially, but perhaps grows quickly unreasonable. In the case of Ebola or any other infectious disease, stringent public health measures would be enacted if the number of people infected become too large<sup>2</sup>. Following public health measures we would expect that the rate of infection would decrease and the number of deaths to slow. So perhaps the second model might be a little more plausible. At some point the number of people who are *not* sick will reach zero, making the rate of infection be zero (or no increase).

### 2.2.2 Model 2: Infection rate proportional to number NOT infected.

In this description notice how we are talking about people who are sick (which we have denoted as  $I$ ) and people who are *not* sick. This looks like we might need to introduce another variable for the “not sick” people, which we will call  $S$ , or susceptible. So the differential equation we would write down would be:

$$\frac{dI}{dt} = kS \quad (2.2)$$

We are still using the parameter  $k$  as with the previous model. Also note we introduced the second variable  $S$  is in Equation (2.2). Because we have introduced another variable  $S$  we should also include a differential equation for how  $S$  changes as well. One way that we can do this is by considering our entire population as consisting of two groups of people:  $S$  and  $I$ . Infection brings someone over from  $S$  to  $I$ , which we have in Figure 2.2:

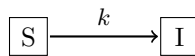


Figure 2.2: Schematic diagram for Model 1

There are three reasons why I like to use diagrams like Figure 2.2: (1) they help organize my thinking about a mathematical model (2) any assumed parameters are listed, and (3) they help me to see that rates can be conserved. If I enter into the box for  $I$ , then someone is leaving  $S$ . So then the rate of change equation for  $S$  is  $\frac{dS}{dt} = -kS$ . Together with the differential equation for  $I$  I have the following:

$$\begin{aligned} \frac{dS}{dt} &= -kS \\ \frac{dI}{dt} &= kS \end{aligned} \quad (2.3)$$

Equation (2.3) is an example of a *coupled differential equation*. In order to “solve” the system we need to determine functions for  $S$  and  $I$ . This coupled set of equations looks a little clunky, but there is something interesting going on if we add the rates  $\frac{dS}{dt}$  and  $\frac{dI}{dt}$  together. Algebraically we have:

$$\frac{dS}{dt} + \frac{dI}{dt} = \frac{d(S+I)}{dt} = 0 \quad (2.4)$$

Recall from calculus that if a rate of change equals zero then the function is constant. In this case, the variable  $S+I$  is constant, or we can also call  $S+I = N$ , the number of people in the population. This means that  $S = N - I$ , so we can re-write Equation (2.3) with a single equation:

$$\frac{dI}{dt} = k(N - I) \quad (2.5)$$

This second model does have some limiting behavior to this model as well. As the number of infected people reaches  $N$  (the total population size), the values of  $\frac{dI}{dt}$  approaches zero, meaning  $I$  doesn't change. There is one caveat to this - if there are no infected people around ( $I = 0$ ) *the disease can still be transmitted*, which might make not good biological sense.

<sup>2</sup>The COVID-19 pandemic that began in 2020 is an example of the heroic efforts of public health officials.

### 2.2.3 Model 3: Infection rate proportional to infected meeting not infected.

The third model rectifies some of the shortcomings of the second model (which rectified the shortcomings of the first model). This model states that the rate of infection is due to those who are sick, actually infecting those who are not sick. This would sort of scenario would also make some sense, as it focused on that *transmission* of the disease are between susceptibles and infected people. So if nobody is sick ( $I = 0$ ) then the disease is not spread. Likewise if there are no susceptibles ( $S = 0$ ), the disease is not spread as well.

In this case the diagram outlining this approach looks something like this:

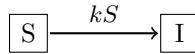


Figure 2.3: Schematic diagram for Model 3

Notice how in Figure 2.3 there is an additional  $S$  associated with  $k$  to show how the rate of infection depends on  $S$ . The differential equations that describe the scenario outlined in Figure 2.3 are the following:

$$\begin{aligned}\frac{dS}{dt} &= -kSI \\ \frac{dI}{dt} &= kSI\end{aligned}$$

Just like before for Model 2 we can combine the two equations to yield a single differential equation:

$$\frac{dI}{dt} = k \cdot I \cdot (N - I) \quad (2.6)$$

Look's pretty similar to model 2, doesn't it? In this case notice the variable  $I$  outside the expression. Notice this seems to be appropriate - if  $I = 0$ , then there is no increase in infection. If  $I = N$  (the total population size) then there is no increase in the infection.

Let's compare these two different models graphically. For both models let's plot  $\frac{dI}{dt}$  versus  $I$ , and just so we can plot let's  $k = 1$  and  $N = 10$  respectively. Plots of these functions are shown in Figure 2.4.

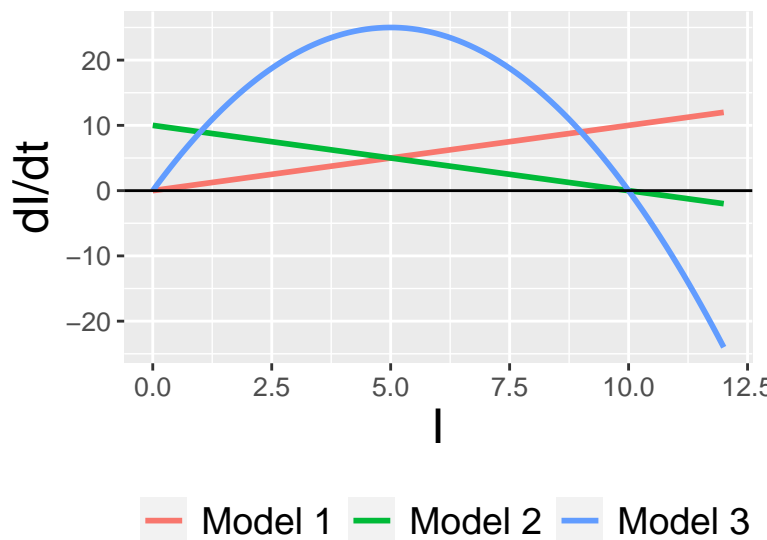


Figure 2.4: Comparing rates of change for three models

Figure 2.4 has a lot to unpack, but we can use some of our understanding of rates of change in calculus to compare the three models. Notice how the sign of  $\frac{dI}{dt}$  is always positive for Model 1, indicating that the solution ( $I$ ) is always increasing. For

Models 2 and 3,  $\frac{dI}{dt}$  equals zero when  $I = 10$ , which also is the value for  $N$ . After that case,  $\frac{dI}{dt}$  turns negative, meaning that  $I$  is decreasing.

In summary, examining the graphs of the rates can tell a lot about the *qualitative behavior* of a solution to a differential equation even without the solution.

## 2.3 Model solutions

Let's return back to possible solutions (in this case formulas for  $I(t)$ ) for our models. Usually a differential equation also has a starting or an initial value (typically at  $t = 0$ ) that actualizes the solution. When we state a differential equation with a starting value we have an **initial value problem**. We will represent that initial value as  $I(0) = I_0$ , where could be considered another parameter.

With that assumption, we can (and will solve later!) the following solutions for these models:

$$\text{Model 1 (Exponential): } I(t) = I_0 e^{kt}$$

$$\text{Model 2 (Saturating): } I(t) = N - (N - I_0)e^{-kt}$$

$$\text{Model 3 (Logistic): } I(t) = \frac{N \cdot I_0}{I_0 + (N - I_0)e^{-kt}}$$

Notice how I assigned the names to each model (Exponential, Saturating, and Logistic). That may not mean much at the moment, but Figure 2.5 plots the three functions  $I(t)$  together when  $I_0 = 5$ ,  $k = 0.03$ , and  $N = 4000$ .

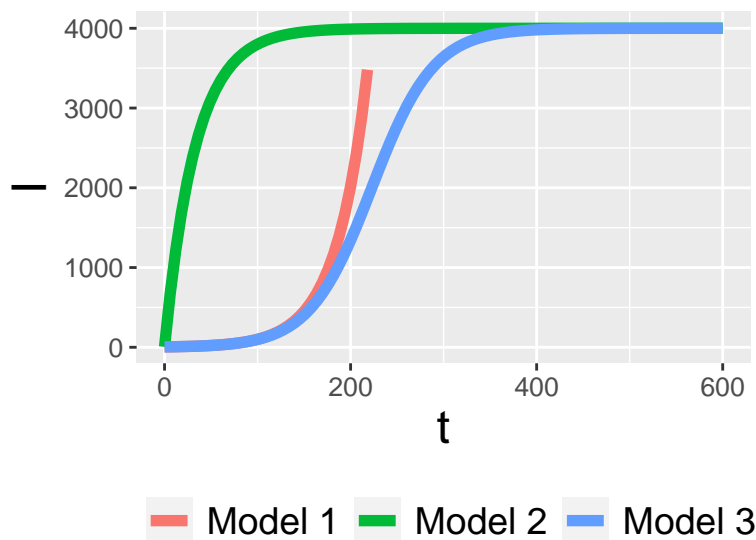


Figure 2.5: Three models compared

Notice how in Figure 2.5 Model 1 increases quickly - it actually grows without bound off the chart! Model 2 and Model 3 have saturating behavior, but it looks like Model 3 might be the one that actually captures the trend of the data.

## 2.4 Which model is best?

All three of these scenarios describe different modeling scenarios. With the saturating and logistic models (Models 2 and 3) we have some limiting behavior the possibility that the the rate of infection slows. Of the two models, which one is the *best* one? Here could be some possible criteria we could evaluate:

- Do the model outputs match the data?
- (For timeseries data) are the trends accurately represented?
- Is the model easy to use?

- How will model outputs compare with newly collected measurements? Here is a list of several things that go into a model:
- The model complexity - how many equations do we have?
- The number of parameters - a few or many?

We will also address that question later on in this textbook when we discuss *model selection* (Chapter ??). Model selection is one key part of the modeling hypothesis - where we investigate the implications of a particular model analyzed. If we don't do this, we don't have an opportunity to test out what is plausible and what is believable in our models.

## 2.5 Start here

In summary, it turns out that even with some initial assumptions we can very quickly build up a mathematical model to explain data. We have a lot more to uncover:

- How would you determine the parameters  $k$  and  $N$  with the collected data?
- Are there other more complicated models?
- What techniques are used to determine the formulas  $I(t)$ ?
- Are there other numerical techniques to approximate the solution  $I(t)$ ?
- What happens to our solutions when the parameters  $k$  and  $N$  change?
- What happens to our solutions when the number of infected people change randomly for some reason?

We will study answers to these questions and more. Let's get started!

## 2.6 Exercises

**Exercise 2.1.** Solutions to an outbreak model of the flu are the following:

$$\text{Saturating model: } I(t) = 3000 - (2990)e^{-.1t}$$

$$\text{Logistic model: } I(t) = \frac{30000}{10 + (2990)e^{-.15t}},$$

where  $t$  is in days. Use these two functions to answer the following questions:

- Plot the saturating and logistic models when  $0 \leq t \leq 100$ .
- For both models, how would you describe the growth of the outbreak as  $t$  increases? How many people will be infected overall?
- Finally, for both models evaluate  $\lim_{t \rightarrow \infty} I(t)$ . How do these results compare to values found on your graph?

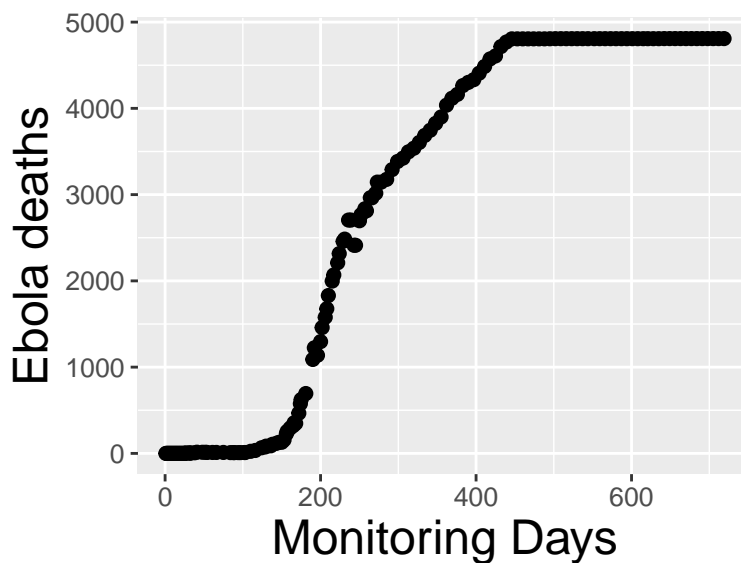


Figure 2.6: An Ebola outbreak in Liberia in 2014

**Exercise 2.2.** Figure 2.6 shows the Ebola outbreak for the country of Liberia in 2014. If we were to apply the logistic model (Model 3) based on this graphic what would be your estimate for  $N$ ?

**Exercise 2.3.** The general solution for the saturating and the logistic models are:

$$\text{Saturating model: } I(t) = N - (N - I_0)e^{-kt}$$

$$\text{Logistic model: } I(t) = \frac{N \cdot I_0}{I_0 + (N - I_0)e^{-kt}},$$

where  $I_0$  is the initial number of people infected and  $N$  is the overall population size. Using the functions from Exercise 2.1 for both models, what are  $N$  and  $I_0$ ?

**Exercise 2.4.** The general solution for the saturating and the logistic models are:

$$\text{Saturating model: } I(t) = N - (N - I_0)e^{-kt}$$

$$\text{Logistic model: } I(t) = \frac{N \cdot I_0}{I_0 + (N - I_0)e^{-kt}},$$

where  $I_0$  is the initial number of people infected and  $N$  is the overall population size. For both models carefully evaluate the limits to show  $\lim_{t \rightarrow \infty} I(t) = N$ . How do your limiting values these compare to the steady-state values you found for Models 2 and 3 in Figure 2.5, where  $N = 4000$ ?

**Exercise 2.5.** A model that describes the growth of sales of a product in response to advertising is the following:

$$\frac{dS}{dt} = .55\sqrt{1-S} - S,$$

where  $S$  is the product's share of the market (scaled between 0 and 1). Use this information to answer the following questions:

- Make a plot of the function  $f(S) = .55\sqrt{1-S} - S$  for  $0 \leq S \leq 1$ .
- Interpret your plot to predict when the market share will be increasing and decreasing. At what value is  $\frac{dS}{dt} = 0$ ? (This is called the *steady-state* value.).
- A second campaign is has the following differential equation:

$$\frac{dS}{dt} = .2\sqrt{1-S} - S$$

What is the steady-state value and how does it compare to the previous one?

**Exercise 2.6.** A more general form of the advertising model is

$$\frac{dS}{dt} = r\sqrt{1-S} - S, \quad (2.7)$$

where  $S$  is the product's share of the market (scaled between 0 and 1). The parameter  $r$  is related to the effectiveness of the advertising (between 0 and 1).

- Solve  $\frac{dS}{dt} = r\sqrt{1-S} - S$  for the steady state value (where  $\frac{dS}{dt} = 0$ ). Your final answer should be expressed as a function  $S(r)$ .
- Make a plot of the steady state value as a function of  $r$ , where  $0 \leq r \leq 1$ .
- Based on your plot, what can you conclude about the steady state value as the effectiveness of the advertising increases?

**Exercise 2.7.** A common saying is “You are what you eat.” This saying is mostly true and can be related in a mathematical model! Here's how: an equation that relates a consumer's nutrient content (denoted as  $y$ ) to the nutrient content of food (denoted as  $x$ ) is given by:

$$y = cx^{1/\theta}, \quad (2.8)$$

where  $\theta \geq 1$  and  $c$  are both constants is a constant. Units on  $x$  and  $y$  are expressed as a proportion of a given nutrient (such as nitrogen or carbon). Let's start with an example when  $c = 1$  and  $\theta = 1$ . Our function then is  $y = x$ . In this case the point  $(0.05, 0.05)$  would say that if an animal ate food that was 5% nitrogen, their body composition would be 5% as well.

- Now assume that  $c = 1$ . How does the nutrient content of the consumer compare to the food when  $\theta = 2$ ? Draw a sample curve and interpret it, contrasting it to when  $\theta = 1$ .
- Now assume that  $c = 1$ . How does the nutrient content of the consumer compare to the food when  $\theta = 5$ ? Draw a sample curve and interpret it, contrasting this curve to the previous two.
- What do you think will happen when  $\theta \rightarrow \infty$ ? Draw some sample curves to help illustrate your findings.

**Exercise 2.8.** A model for the outbreak of a cold virus assumes that the rate people get infected is proportional to infected people contacting susceptible people, as with Model 3 (the Logistic model). However people who are infected can also recover and become susceptible again with rate  $\alpha$ . Construct a diagram similar Model 3 for this scenario and also write down what you think the system of differential equations would be.



**Exercise 2.9.** A model for the outbreak of the flu assumes that the rate people get infected is proportional to infected people contacting susceptible people, as in Model 3. However people also account for recovering from the flu, denoted with the variable  $R$ . Assume that the rate of recovery is proportional to the number of infected people with parameter  $\beta$ . Construct a diagram like Model 3 for this scenario and also write down what you think the system of differential equations would be.

**Exercise 2.10.** Organisms that live in a saline environment biochemically maintain the amount of salt in their blood stream. An equation that represents the level of  $S$  in the blood is the following:

$$\frac{dS}{dt} = I + p \cdot (W - S),$$

where the parameter  $I$  represents the active uptake of salt,  $p$  is the permeability of the skin, and  $W$  is the salinity in the water. Use this information to answer the following questions:

- What is that value of  $S$  at *steady state*, or when  $\frac{dS}{dt} = 0$ ? Your final answer should be a function  $S(I, p, W)$ .
- With the steady state solution, Use your what parameters ( $I$ ,  $p$ , or  $W$ ) cause the steady state value  $S$  to increase?

**Exercise 2.11.** The immigration rate of bird species (species per time) from a mainland to an offshore island is  $I_m \cdot (1 - S/P)$ , where  $I_m$  is the maximum immigration rate,  $P$  is the size of the source pool of species on the mainland, and  $S$  is the number of species already occupying the island. Further, the extinction rate is  $E \cdot S/P$ , where  $E$  is the maximum extinction rate. The growth rate of the number of species on the island is the immigration rate minus the extinction rate.

- Make representative plots of the of the immigration and the extinction rates as a function of  $S$ . You may set  $I_m$ ,  $P$ , and  $E$  all equal to 1.
- Determine the number of species for which the net growth rate is zero, or the number of species is in equilibrium. Express your answer as  $S$  as a function of  $I_m$ ,  $P$ , and  $E$ .
- Suppose that two islands of the same size are at different distances from the mainland. Birds arrive from the source pool and they have the same extinction rate on each island. However the maximum immigration rate is larger for the island farther away. Which of the two islands will have the larger number of species at equilibrium?

**Exercise 2.12.** This problem relates to animal size and volume. Assume that an animal assimilates nutrients at a rate  $R$  proportional to its surface area. Also assume that it uses nutrients at a rate proportional to its volume. You may assume that the size of the animal is implicitly a function of the nutrient intake and usage. Determine the size of the animal if its intake and use rates were in balance (meaning  $R$  is set to zero), assuming the animal is the following shapes:

- A sphere (assume size is measured with radius  $r$ ) *Note:* first determine the geometric formulas for surface area and volume.
- A cube (assume size is measured with length  $l$ )

*Hint:* For both of these problems your goal is to determine a numeric value of  $r$  and  $l$ .



# Chapter 3

## Introduction to R

The primary tool we have to analyze models will be R and RStudio, which are commonly used for scientific and statistical computations. This is an exciting program and powerful program to learn! Admittedly learning a software requires a learning curve, however I think it is worth it. With R you will have enormous flexibility to efficiently utilize data, design effective visualizations, and process statistical models.

### 3.1 R and RStudio

We will use a lot with RStudio, which is an Integrated Development Environment for R. These are two separate downloads and files, which can be found here:

- R: (<https://cran.r-project.org/mirrors.html>)[Link to download mirror] (you need to select a location to download from; choose any one that is geographically close to you.)
- RStudio: (<https://www.rstudio.com/products/rstudio/download/>)[Choose RStudio desktop - the free version].

#### 3.1.1 Why do we have two programs?

Think of R as your basic program - this is the engine that does the computation. RStudio was developed as an Integrated Development Environment - meaning a place where you can see everything you are working on in one place. Figure 3.1 shows an example of an RStudio pane that I have:

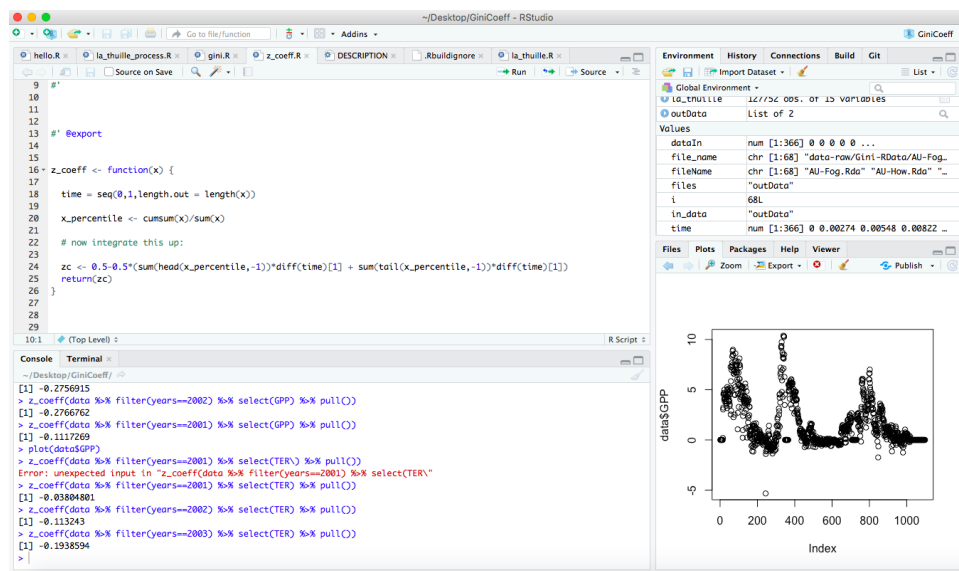


Figure 3.1: A sample RStudio pane from one of my projects.

There are 4 key panels that I work with, clockwise from the top:

- The **source** window is in the upper left - notice how those have different tabs associated with them. You can have multiple source files that you can toggle between. For the moment think of these as commands that you will want to send to R.
- The **environment** and **history** pane - these tables allow you to see what variables are stored locally in your environment, or the history of commands.
- The **files** and **plots** command (a simple plot I was working on is shown currently), but you can toggle between the tabs. The files tab shows the files in the current Rstudio project directory.
- Finally, the **console** pane is the place where R works and runs commands. You can type in there directly, otherwise we will also just “send” commands from the source down to the console.

Now we are ready to work with R and RStudio!

## 3.2 First steps: getting acquainted with R

Let’s get started! Open up RStudio. Task one will be to create a project. A project is a central place to organize your materials for this course. You may do this already, but R can be picky about its working directory - and navigating to it. I found creating a project file is an easy way to avoid some of that fussiness. Let me describe steps in how to do this.

1. In RStudio select “File” then “New Project”
2. Next select the first option “New Directory” in the window - this will create a new folder on your computer.
3. At the next window choose New Directory or Existing Directory - it depends on where you want to place this project.
4. Name the project as you like.
5. Click the “Create Project” button.

### 3.2.1 “Working” with R

Our next step: where do we get R to do something? For example if we wanted to compute of  $4+9$  (yeah, it is 13, but this is an illustrative example), we could type this command in the R console (lower left) window. Let’s try this now.

1. In the console type  $4+9$
2. Then hit enter (or return)
3. Is the result 13?

Success! Now let me show you another way that works well if you have multiple lines of code to evaluate or save. Working with a script (.R file) is better. This will utilize the upper left hand corner of your RStudio window. (You may not have anything there when you start working on a project, so let’s create one.)

1. In RStudio select “File” then “New File”
2. Next select the first option “New Script”
3. A new window called “UntitledX” should appear, where X is a number. You are set to go!

I like to use this window as a file to type stuff in and then evaluate it, which we will do next.

**Pro tip:** There are shortcuts to creating a new file: Ctrl+Shift+N (Windows and Linux) or Command+Shift+N (Mac)

### 3.2.2 Sending commands down to console.

Now we want to type and evaluate a command for R to do something. Click anywhere in the source file that you created and type the following:

```
4+9
```

```
## [1] 13
```

You have several options:

1. Copying and pasting the command to the window. Shortcuts are Ctrl+C / Command+C for copying and Ctrl+V / Command+V for Windows / Mac.
2. Run the line. This means that your cursor is at the line in your source file, then clicking the ‘Run’ button in the upper right hand side of the source window. Shortcuts are Ctrl+Enter / Command+Enter.

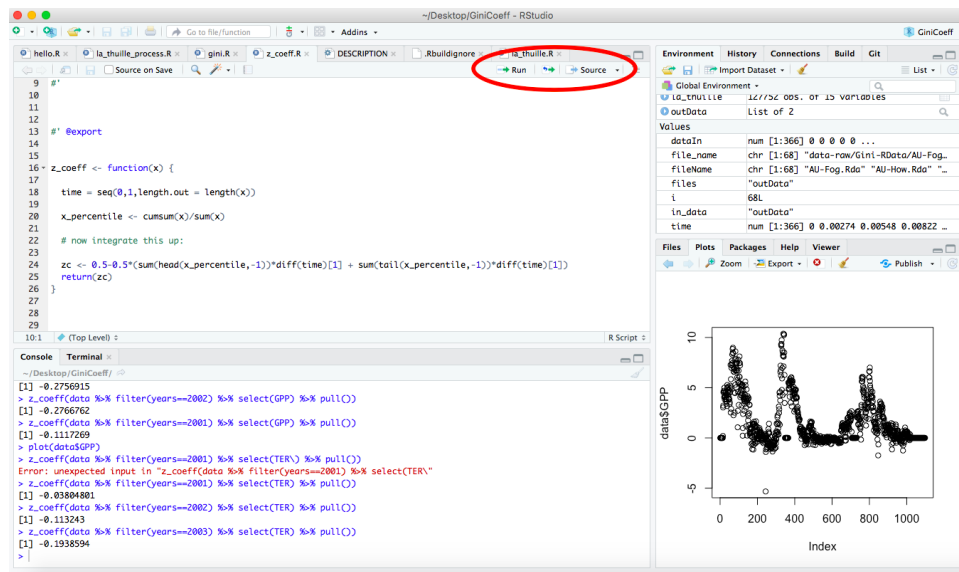


Figure 3.2: Sending a command to the console.

3. You can also source the whole file, which means all the lines from top to bottom. You do this by clicking the source button, or with shortcuts `Ctrl+Shift+Enter` / `Cmd+Shift+Enter` (Windows / Mac). You can imagine this makes things easier when you have SEVERAL lines of commands to evaluate.

Why do I like working with a script file? Well if I run some commands that have an error then it is much easier to just fix a quick mistake and then re-run the code.

It is also helpful to annotate your code with comments (`#`), which appear as green text in RStudio.

### 3.2.3 Saving your work

The neat part about a source file is that it allows you to save the file (`Ctrl+S` / `Cmd+S`). The first time you do this you may need to give this a name. The location where this file will be saved is in the same directory as your `.Rproj` project file. Now you have a file that you can come back to! In general I try to use descriptive names of files so I can refer back to them later.

## 3.3 Increasing functionality with packages

One awesome versatility with R is the ability to add packages - these packages extend the functionality of R with contributed, specialized code. These are similar to apps on your phone. You can get packages from a few different places:

- CRAN, which stands for Comprehensive R Archive Network. This is the clearing house for many contributed packages - and allows for easy cross platform functionality.

One key package is `tidyverse`, which is actually a collection of packages. If you take an introductory data science course you will most likely be learning more about this package, but to install this at the command line you type the following:

```
install.packages('tidyverse')
```

Typing this line will connect to the CRAN download mirrors and install this set of packages locally to your computer. It may take some time, but be patient.

Another package you should install is `devtools`:

```
install.packages('devtools')
```

Sometimes when you are installing packages you may be prompted to install additional packages. In this case just say yes.

- Github. This is another place where people can share code and packages. The code here has not been vetted through CRAN for compatibility, but if you trust the person sharing the code, it should work. In order to do this we will need to first load up the `devtools` library:

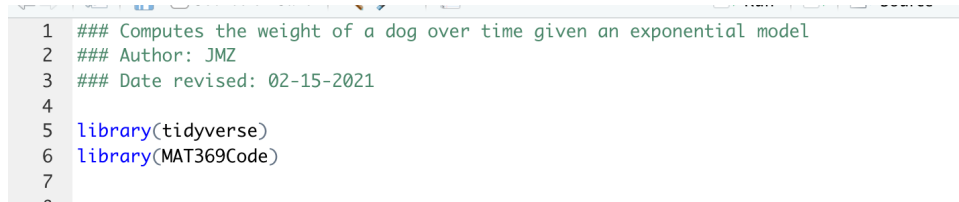
```
library(devtools)
install_github("jMZobitz/MAT369Code",build_vignettes= TRUE)
```

What this command will do is pull in the package structure from my github page and install it locally.

Here is the good news: *you only need to install a package once before using it!* To load the package up into your workspace you use the command `library`:

```
library(tidyverse)
library(MAT369Code)
```

You need to load up your these libraries *each time you restart your R session*. This is part of the benefit of a script file - at the start I always declare the libraries that I will need at the start of the script file, as shown in the following figure:



```
1  ### Computes the weight of a dog over time given an exponential model
2  ### Author: JMZ
3  ### Date revised: 02-15-2021
4
5  library(tidyverse)
6  library(MAT369Code)
7
8
```

Figure 3.3: A sample R script.

Also notice here that the first few lines of the script file I used comments `#` to denote the basic purpose of the file, who wrote it, and the date it was last revised. This type of information is good programming practice at the start.

## 3.4 Working with R: variables, data frames, and datasets

### 3.4.1 Creating variables

The next thing we will want to do is to define variables that are stored locally. This is pretty easy to do:

```
my_result <- 4+9
```

The symbol `<-` is assignment (you can use equals (`=`), but it is good coding practice to use the arrow for assignment). Notice how I named the variable called `my_result`. Generally I prefer using *descriptive* names for variables for the context at hand (In other words, `x` would be an odd choice - too ambiguous.) I also used snake case to string together multiple words. In practice you can use snake case, or alphabetic cases (`myResult`) or even `my.result` (although that may not be preferred practice in the long run). You can't use `my-result` because it looks like subtraction.

Here is the good news, then we can compute with the new variable, so for example `10*my_result` should yield 130. Cool, no? As an example, let's define a sequence, spaced from 0 to 5 with spacing of 0.05. Store this in a variable called `my_sequence`.

To do this we use the `seq` command and requires the starting value, ending value, and step size:

```
my_sequence <- seq(from=0,to=5,by=0.05)
```

The format for the function `seq` is `seq(from=start,to=end,by=step_size)`. The `seq` command is a pretty flexible - there are alternative ways you can generate a sequence by specifying the starting and the end values along with the number of points. If you want to know more about `seq` you can always use `?` followed by the command - that will bring up the help values:

```
?seq
```

Once you get more comfortable with syntax in R, you will see that `seq(0,5,0.5)` gives the same result as `seq(from=0,to=5,by=0.05)`, but it is helpful to write your code *so that you can understand what it does*.

### 3.4.2 Data frames

A key structure in R is that of a data frame, which allows different types of data to be collected together. A data frame is like a spreadsheet where each column is a value and each row a value (much like you would find in a spreadsheet):

Table 3.1: A data frame

	mpg	disp
Mazda RX4	21.0	160
Mazda RX4 Wag	21.0	160
Datsun 710	22.8	108
Hornet 4 Drive	21.4	258
Hornet Sportabout	18.7	360

Table 3.2: Model solutions

time	model_1	model_2	model_3
0.000000	5.000000	5.0000	5.000000
6.060606	5.996981	669.1571	5.995486
12.121212	7.192755	1222.9000	7.188814
18.181818	8.626962	1684.5848	8.619147
24.242424	10.347145	2069.5158	10.333332

Table 3.1 shows the miles per gallon in one column (the variable `mpg` and the engine size (the variable `disp`) for different types of cars. The row names (`Mazda RX4`) just tell you the type of the car. Sometimes row names are not shown.

Another data frame may list solutions to a differential equation, like we did with our three infection models in Chapter 2:

Data frames are an example of *tidy* data, where each row is an observation, each column a variable (which can be quantitative or categorical). There are several different ways to define a data frame in R. I am going to rely on the approach utilized by the `tidyverse`, which calls data frames `tibbles`. So for example, here is I am going to define a data frame that computes the quadratic function  $y = 3x^2 - 2x$  for  $-5 \leq x \leq 2$ .

```
x <- seq(from=-5,to=2,by=0.05)
y <- 3*x^2-2x

my_data <- tibble(x=x,
                  y=y) # Notice I sam specifically defining x and y
```

Notice that the data frame `my_data` uses the column (variable) names of `x` and `y`. You could have also used `tibble(x,y)`, but it is helpful to name the columns in the way that you would like them to be named.

### 3.4.3 Reading in datasets

R has a lot of built in datasets! In fact to see all the datasets, type `data()` at the console. This will popup a new window in RStudio with the names. Cool!

If you want to see the datasets for a specific package (such as `MAT369Code`) you type `data(package = "MAT369Code")` at the console.

Perhaps what is most important is being able to read in datasets provided to you. Data come in several different types of formats, but one of the more versatile ones are csv (comma separated values). What you need to do is the following:

- Where you have your .Rproj file located, create a folder called `data` or `datasets`
- Save the file locally on your computer. Take note where you have it saved on your computer, and drag the file to your `data` folder.
- To read in the file you will use the command `read_csv`, which has the following structure:

```
in_data <- read_csv(FILENAME)
```

The data gets assigned to the variable `in_data` (You can call this variable what you want.) For example I have the following csv file of ebola data, which I read in via the following:

```
ebola <- read_csv("data/ebola.csv")
```

Notice the quotes around the `FILENAME`. **Pro tip:** If you have the data files in the data folder, in RStudio you can type “data” and it may start to autocomplete - this is hand (you can also use tab.)

Table 3.3: Weight of a dog over time

days	mass
31	6.25
62	10.00
93	20.00
99	23.00
107	26.00
113	27.60
121	29.80
127	31.60
148	37.20
161	41.20
180	48.70
214	54.00
221	54.00
307	63.00
452	66.00
482	72.00
923	72.20
955	76.00
1308	75.00

## 3.5 Visualization with R

Now we are ready to begin visualizing data frames. Two types of plots that we will need to make will be a scatter plot and a line plot. We are going to consider both of these separately, with examples that you should be able to customize.

### 3.5.1 Making a scatterplot

One dataset we have is the mass of a dog over time, adapted from here. We have two variables here:  $D$  = the age of the dog in days and  $W$  = the weight of the dog in pounds. I have the data loaded into the **MAT369Code** package, which you can investigate by typing the following at the command line (I display it below as well):

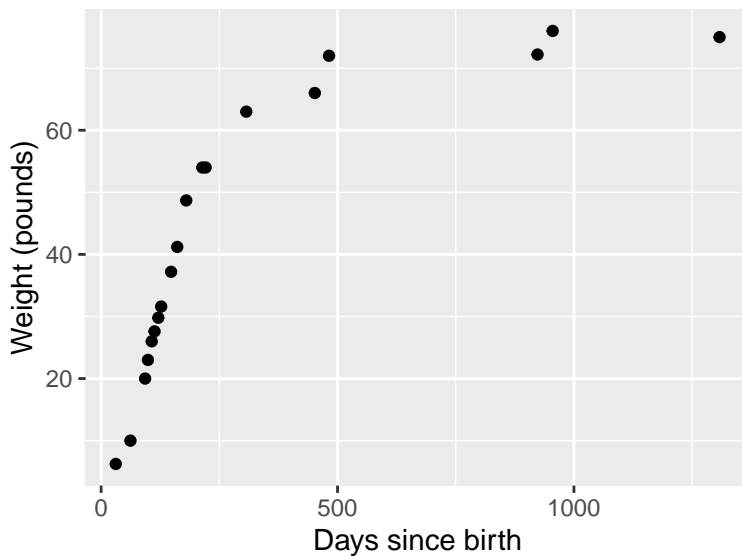
```
glimpse(wilson)
```

(Notice that I have assumed you have the **MAT369Code** library loaded.) You can also explore the documentation for this dataset by typing `?wilson` at the console.

Notice that this data frame has two variables: `days` and `mass`. To make a scatter plot of these data we are going to use the command `ggplot`:

```
ggplot(data = wilson) +
  geom_point(aes(x = days, y = mass)) +
  labs(x='Days since birth',
       y='Weight (pounds)')
```





Wow! This looks complicated. Let's break this down step by step:

- `ggplot(data = wilson)` + sets up the graphics structure and identifies the name of the data frame we are include.
- `geom_point(aes(x = days, y = mass))` defines the type of plot we are going to be making.
- `geom_point()` defines the type of plot geometry (or *geom*) we are using here - in this case, a point plot.
- `aes(x = days, y = mass)` determines the *aesthetics* of the plot. On the x axis is the days variable, on the y axis is the mass variable.
- The statement beginning with `labs(x=...)` defines the labels on the x and y axes.

I know this seems like a lot to write for a plot, but this structure is actually used for some more advanced data visualization. Trust me - learning how to make informative plots can be a useful skill!

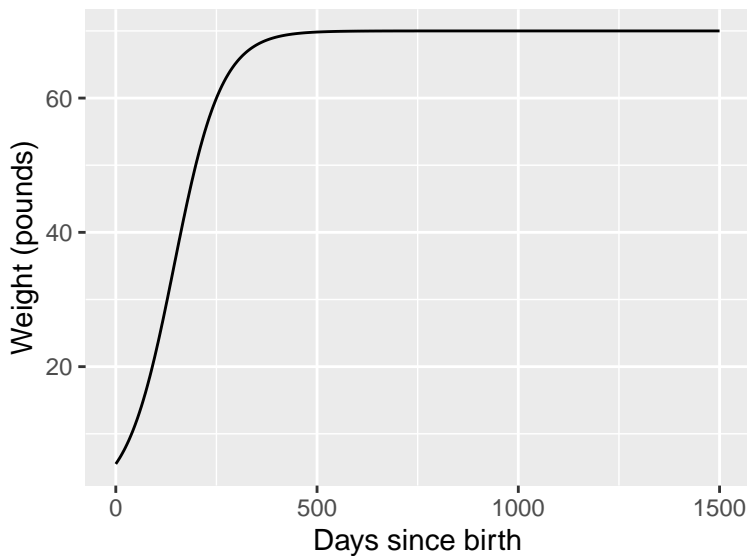
### 3.5.2 Making a line plot

Using the same `wilson` data, later on we will discover that the function  $W = f(D) = \frac{70}{1 + e^{2.46 - 0.017D}}$  represents these data. In order to make a plot of this function we can use need to first build a data frame, `plotFunction`, which has a data frame as inputs:

```
days <- seq(0,1500,by=1) # Choose spacing that is "smooth enough"
mass <- 70/(1+exp(2.46-0.017*days))

wilson_model <- tibble(days = days,
                       mass = mass)

ggplot(data = wilson_model) +
  geom_line(aes(x=days,y=mass)) +
  labs(x='Days since birth',
       y='Weight (pounds)')
```



Notice that once we have the data frame set up, the structure is very similar to the scatter plot - but this time we are calling using `geom_line()` than `geom_point`.

### 3.5.3 Changing options

Want a different color? Thicker line? That is fairly easy to do. For example if we wanted to make either our points or line a different color, we can just choose the following:

```
ggplot(data = wilson) +
  geom_point(aes(x = days, y = mass), color='red',size=2)
  labs(x='Days since birth',
        y='Weight (pounds)')
```

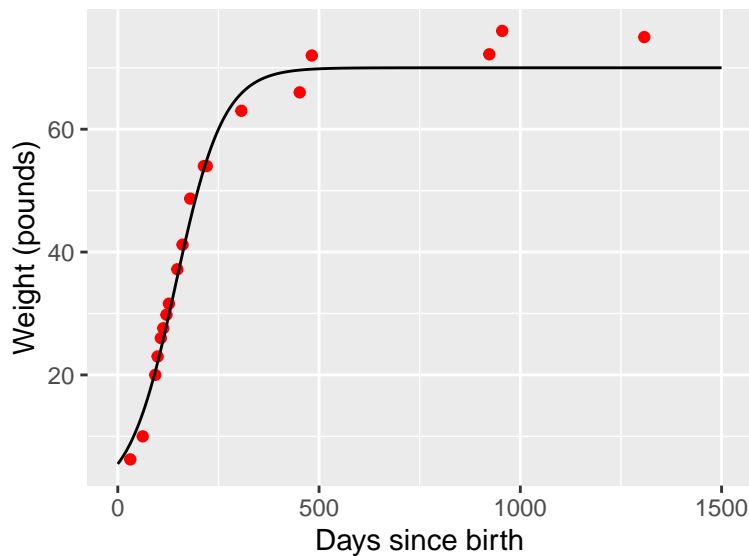
Notice how the command `color='red'` was applied *outside* of the `aes` - which means it gets mapped to each of the points in the data frame. `size=2` refers to the size (in millimeters) of the points. I've linked more options about the colors and sizes you can use here:

- **Named colors in R:** [LINK](#) Scroll down to “Picking one color in R” - you can see the list of options!
- **More colors:** [LINK](#). More information about working with colors.
- **Using hexadecimal colors:** [LINK](#) (You specify these by the code so `"#FF3300"` is a red color.)
- **Changing sizes of lines and points:** [LINK](#)

### 3.5.4 Combining scatter and line plots.

This is actually easy to do, especially since we are combining both the plot geoms together. Try running the following code (I am still using the data frame `wilson_model` as defined above:

```
ggplot(data = wilson) +
  geom_point(aes(x = days, y = mass),color='red') +
  geom_line(data = wilson_model, aes(x=days,y=mass)) +
  labs(x='Days since birth',
        y='Weight (pounds)')
```



Notice in the above code a subtle difference when I added in the dataset `wilson_model` with `geom_line`: you need to name the `data` bringing in a new data frame to a plot geom.

While it may be useful to have a legend to the plot, for this course we will make plots where this the context will be more apparent. Additional reading on legends can be found [here](#).

## 3.6 Defining functions

We will study lots of other built-in functions for this course, but you may also be wondering how you define your own function (let's say  $y = x^3$ ). We need the following construct:

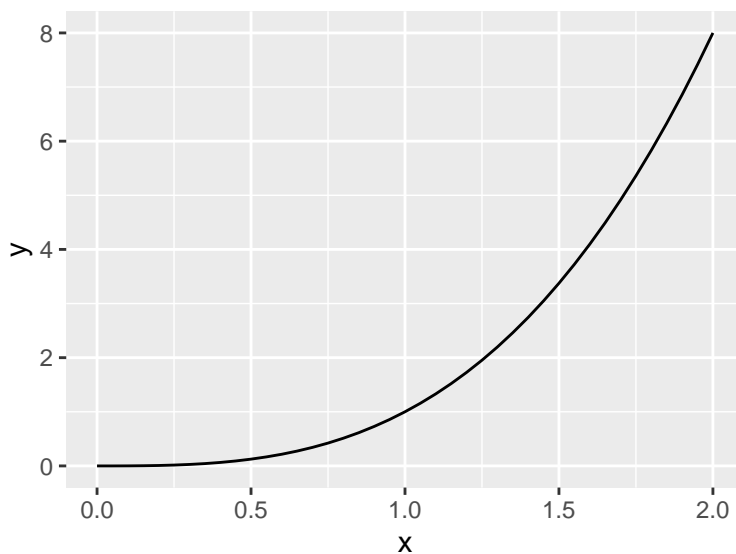
```
function_name <- function(inputs) {  
  
  # Code  
  
  return(outputs)  
}
```

Here `function_name` serves as what you call the function, `inputs` are what you need in order to run the function, and `outputs` are what gets returned. So if we are doing  $y = x^3$  then we will call that function `cubic`:

```
cubic <- function(x) {  
  y=x^3  
  return(y)  
}
```

So now if we want to evaluate  $y(2) = 2^3$  we type `cubic(2)`. Neat! Now let's make a plot of the graph  $y = x^3$  using the function defined as `cubic`. Here is the R code that will accomplish this:

```
x <- seq(from = 0, to = 2, by = 0.05)  
y <- cubic(x)  
  
my_data <- tibble(x=x,y=y)  
  
ggplot(data = my_data) +  
  geom_line(aes(x=x,y=y)) +  
  labs(x='x',  
       y='y')
```



### 3.6.1 Functions with inputs

Sometimes you may want to define a function with different input parameters, so for example the function  $y = x^3 + c$ . To define that, we can modify the function to have input variables:

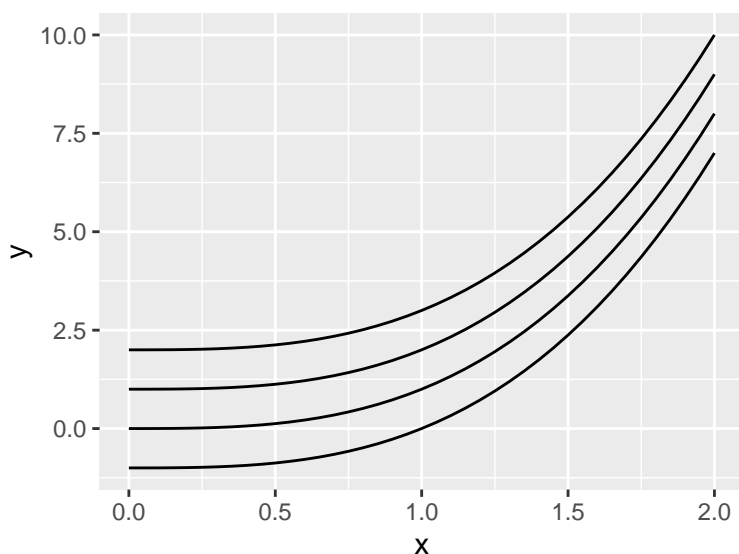
```
cubic_revised <- function(x,c) {
  y=x^3+c
  return(y)
}
```

So if we want to plot what happens for different values of  $c$  we have the following:

```
x <- seq(from = 0, to = 2, by = 0.05)

my_data_revised <- tibble(x=x,
  c_zero=cubic_revised(x,0),
  c_pos1=cubic_revised(x,1),
  c_pos2=cubic_revised(x,2),
  c_neg1=cubic_revised(x,-1))

ggplot(data = my_data_revised) +
  geom_line(aes(x=x,y=c_zero)) +
  geom_line(aes(x=x,y=c_pos1)) +
  geom_line(aes(x=x,y=c_pos2)) +
  geom_line(aes(x=x,y=c_neg1)) +
  labs(x='x',
    y='y')
```



Notice how I defined multiple columns of the data frame `my_data_revised` in the `tibble` command, and then used multiple `geom_line` commands to plot the data. Since we had combined the different values of `c` in a single data frame we didn't need to define the data with each instance of `geom_line`.

## 3.7 Concluding thoughts

This is not meant to be a self-contained chapter in R but rather one to get you - those miles have been trod by others, and here are few of my favorites that I turn to:

- **R Graphics.** This is a go to resource for making graphics. (I also use google a lot too.)
- **The Pirates Guide to R.** This book promises to build your R knowledge from the ground up.
- **R for Reproducible Scientific Analysis.** This set of guided tutorials can help you build your programming skills in R.
- **R for Data Science** this is a useful book to take your R knowledge to the next level.

The best piece of advice: DON'T PANIC! Patience and persistence are your friend. Reach out for help, and recognize that like with any new endeavor, practice makes progress. x

## 3.8 Exercises

**Exercise 3.1.** Create a folder on your computer and a project file where you will store all your R work for this textbook.

**Exercise 3.2.** Install the packages `devtools`, `tidyverse` to your R installation. Once that is done, then install the package `MAT369Code` from my github page.

**Exercise 3.3.** What are the variables listed in the dataset `phosphorous` in the `MAT369Code` library? (Hint: try the command `?phosphorous`.)

**Exercise 3.4.** Make a scatterplot (`geom_point()`) of the dataset `phosphorous` in the `MAT369Code` library. Be sure to label the axes.

**Exercise 3.5.** Change the line plot of Wilson's weight over time so the line is blue and the size is 4.

**Exercise 3.6.** Change the color of the scatterplot of Wilson's weight over time to a either a hexadecimal color or a named color of your choice.

**Exercise 3.7.** For this exercise you will do some plotting:

- Define a sequence (call this sequence  $x$ ) that ranges between -12 to 12 with spacing of .05.
- Also define the variable  $y$  such that  $y = \sin(x)$ .
- Make a scatter plot to graph  $y = \sin(x)$ . Set the points to be red.
- Make a line plot to graph  $y = \sin(x)$ . Label the x-axis with your favorite book title. Label the y-axis with your favorite food to eat.

**Exercise 3.8.** An equation that relates a consumer's nutrient content (denoted as  $y$ ) to the nutrient content of food (denoted as  $x$ ) is given by:  $y = cx^{1/\theta}$ , where  $\theta \geq 1$  and  $c$  are both constants is a constant. Let's just assume that  $c = 1$  and the  $0 \leq x \leq 1$ .

Write a function called `nutrient` that will make a sequence of `y` values for an input `x` and `theta` (*theta*). Then use that code to make a line plot (`geom_line()`) for five different values of  $\theta > 1$ , appropriately labeling all axes.

**Exercise 3.9.** Researchers measured the phosphorous content of *Daphnia* and its primary food source algae. This is the dataset `phosphorous` in the `MAT369Code` library.

Researchers believe that *Daphnia* has strict homeostatic regulation of the phosphorous in algae, and as such want to determine the value of  $\theta$  in the equation  $y = y = cx^{1/\theta}$ . They have already determined that the value of  $c = 1.737$ .

- If you haven't already, make a scatterplot (`geom_point()`) of the dataset `phosphorous` in the package library. Be sure to label the axes correctly.
- Use your function `nutrient` from the previous exercise to make an initial guess for `theta` ( $\theta$ ) that would be consistent with the data. You can evaluate your guess by plotting (with `geom_line()`) against the data.
- Use guess and check to refine the value of  $\theta$  that seems to work best.
- When you have the correct value, enter in the value of  $\theta$  by clicking on [here](https://tinyurl.com/day2theta).

**Exercise 3.10.** For this exercise you will investigate some built-in functions. Remember you can learn more about a function by typing `?FUNCTION`, where `FUNCTION` is the name.

- Explain (using your own words) what the function `runif(1,100,1000)` does.
- Explain (using your own words) what the function `ceiling()` does, showing an example of its use.

**Exercise 3.11.** For this exercise you write a sample function file.

- Create a new source file and save it as `myFunction.R`.
- Type this code the file you created: `myInteger <- ceiling(runif(1, 100, 1000))` (This will declare a variable `myInteger` that you will work with in the following steps.)
- Determine a function in R that will **compute a cumulative sum** from 1 to the value of `myInteger`. Modify your file so that it also computes the cumulative sum and then source your file.
- Copy and paste your function into your homework document for evaluation.

**Exercise 3.12.** The Ebola outbreak in Africa in 2014 severely affected the country of Sierra Leone. A model for the number of deaths  $D$  due to ebola is given by the following equation:

$$D(t) = \frac{K \cdot N_0}{N_0 + (K - N_0) \exp(-rt)},$$

where  $K = 3980$ ,  $N_0 = 5$  and  $r = 0.0234$ . The variable  $t$  is in days. Use `geom_line()` to visualize this curve from  $0 \leq t \leq 700$ .

**Exercise 3.13.** Consider the following piecewise function:

$$y = \begin{cases} x^2 & \text{for } 0 \leq x < 1, \\ 2 - x & \text{for } 1 \leq x \leq 2 \end{cases} \quad (3.1)$$

- Define a function in R that computes  $y$  for  $0 \leq x \leq 2$ .
- Use `geom_line()` to generate a graph of  $y(x)$  over the interval  $0 \leq x \leq 2$ .

**Exercise 3.14.** An insect's development rate  $r$  depends on temperature  $T$  (degrees Celsius) according to the following equation:

$$r = \begin{cases} 0.1 & \text{for } 17 \leq T < 27, \\ 0 & \text{otherwise.} \end{cases} \quad (3.2)$$

- Define a function in R that computes  $r$  for  $0 \leq T \leq 30$ .
- Use `geom_line()` to generate a graph of  $r(T)$  over the interval  $0 \leq T \leq 30$ .





## Chapter 4

# Modeling with rates of change

So far we have looked at some examples for how we can apply rates of change to develop a mathematical model, and also learned a little bit about the ways we can apply computational software such as R.

In this section we are going to look some additional examples of how we can translate equations with rates of change to understand phenomena. The focus here will be on writing differential equations from a contextual description.

Oftentimes when we construct differential equations from a contextual description we bring our own understanding and knowledge to this situation. How *you* may write down the differential equation may be different from someone else - *do not worry!* This is the fun part of modeling: models can be considered testable hypotheses that can be refined when confronted with data.

In this section I work through a few well-known examples from mathematical biology and you will apply that knowledge to develop models from context.

### 4.1 Lynx and Hares

Our first example is a *system of differential equations*. The context is between the snowshoe hare and the Canadian lynx. Figure 4.1 shows a picture of them below from [link](#)



Figure 4.1: The lynx and hare - aren't they beautiful?

Figure 4.2 timeseries of their population is shown with this figure from ([stenseth\\_population\\_1997?](#)).

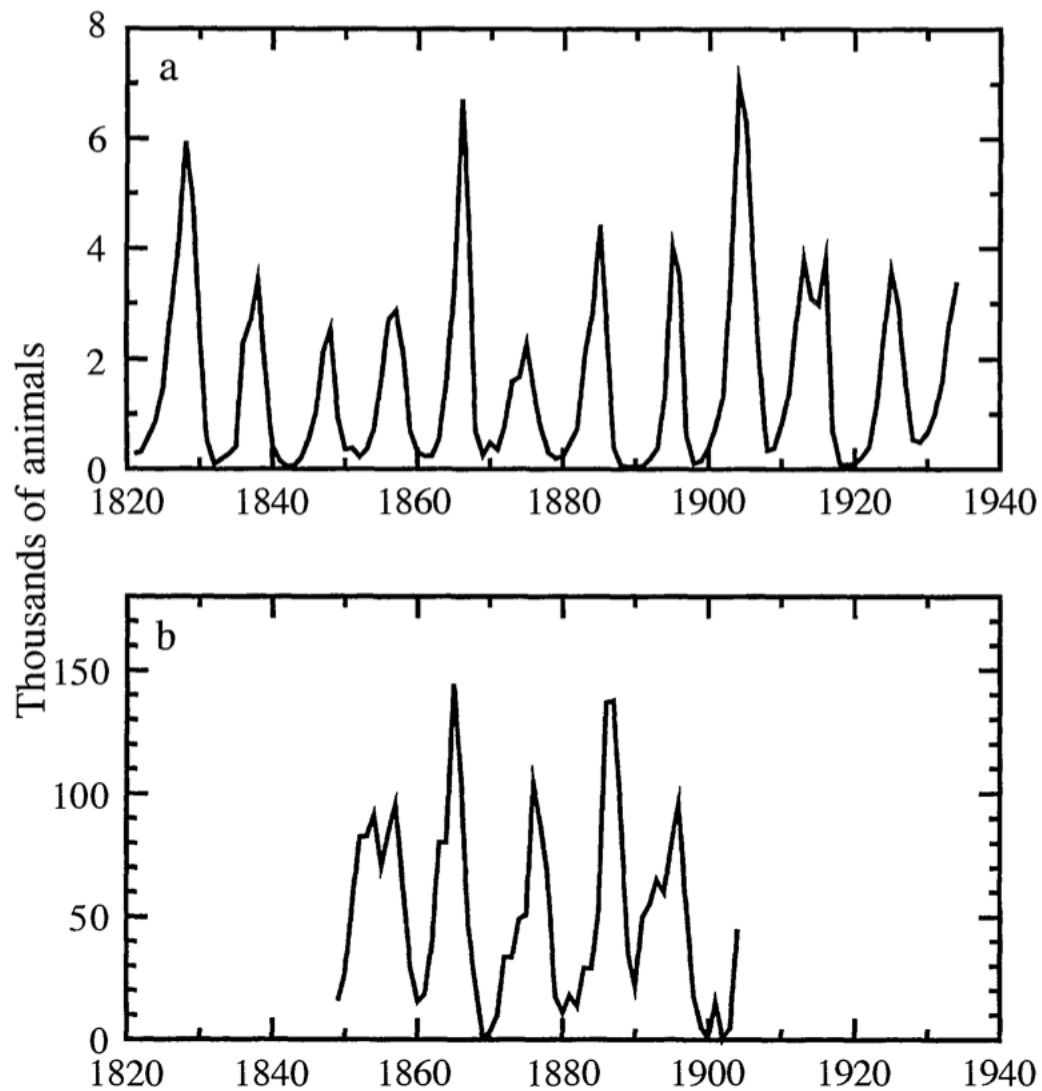


FIG. 1. Cycles in abundance of Canadian lynx and snowshoe hare as revealed in the fur harvest records of the Hudson's Bay Company. (a) Lynx fur harvest from the MacKenzie River District (after Elton and Nicholson 1942*b*). (b) Snowshoe hare harvest from the regions near Hudson's Bay (after MacLulich 1957).

Figure 4.2: A timeseries of the combined lynx and hare system. Notice how the populations are coupled with each other.

Notice how in Figure 4.2 both populations seem to fluctuate periodically. One plausible reason is that the lynx prey on the snowshoe hares, which causes the population to initially decline. Once the snowshoe hare population declines, then there is less food for the lynx to survive, so their population declines. The decline in the lynx population causes the hare population to increase, and so on it goes ...

In summary it is safe to say that the two populations are *coupled* to one another. But in order to understand how they are coupled together, first let's consider the two populations *separately*.

The hares grow much more quickly than the lynx - in fact some hares have been known to reproduce several times a year. A reasonable assumption for large hare populations is that rate of change of the hares is proportional to the hare population.

Based on this assumption Equation (4.1) describes the rate of change of the hare population, with  $H$  is the population of the hares:

$$\frac{dH}{dt} = rH \quad (4.1)$$

In this case we know that the growth rate  $r$  is positive, so then the rate of change ( $H'$ ) will be positive as well, and  $H$  will be increasing. Typical values given for  $r$  in (**stenseth population 1997?**) are between 1.8 - 2.0 year<sup>-1</sup>. You may be thinking that the units on  $r$  seem odd - (year<sup>-1</sup>). Another way to think about  $r$  is to take its inverse:  $r^{-1} \approx 0.5 - 0.55$  years. Then  $r^{-1}$  represents the amount of time that passes before the hare population increases (pretty short!)

Let's consider the lynx now. A approach is to assume their population declines exponentially, or changes at the rate proportional to the current population. Let's consider  $L$  to be the lynx population, with the following differential equation (Equation (4.2)):

$$\frac{dL}{dt} = -dL \quad (4.2)$$

We assume the death rate  $d$  in Equation (4.2) is positive, leading to a negative rate of change for the Lynx population (and a decreasing value for  $L$ ). Typical values of  $d$  are 0.9 - 2.4 year<sup>-1</sup>. Similar to  $r$ , another way  $d$  is - like  $b$  - to take its inverse (about 0.4 - 1.1 years), which represents the amount of time that passes before the lynx population decreases by one.

The next part to consider is how they interact. Since the hares are prey for the lynx, when the lynx hunt, the hare population. We can represent this with the following adjustment to our hare equation:

$$\frac{dH}{dt} = rH - bHL \quad (4.3)$$

So the parameter  $b$  represents the hunting rate. Notice how we have the term  $HL$  for this interaction. This term injects a sense of realism: if the lynx are not present ( $L = 0$ ), then the hare population can't decrease due to hunting. We say that the *interaction* between the hares and the lynx with multiplication. Typical values for  $b$  are 480 - 870 hares · lynx<sup>-1</sup> year<sup>-1</sup>. It is okay if that unit seems a little odd to you - it should be! Here is one way to think about it. The quantity  $\frac{dH}{dt}$  represents the *rate of change* of the hares, so it should have units of hares per year. Since the term  $bHL$  has both lynx and hare, the units for  $b$  need to account for this.

How does hunting affect the lynx population? One possibility is that it increases the lynx population:

$$\frac{dL}{dt} = bHL - dL \quad (4.4)$$

Notice the symmetry between the rate of change for the hares and the lynx equations. In many cases this makes sense - if you subtract a rate from one population, then that rate should be added to the receiving population. You could also argue that there is some efficiency loss in converting the hares to lynx - not all of the hare is converted into the lynx biomass. In this situation we sometimes like to adjust the lynx equation with another parameter  $e$ , representing the efficiency that hares are converted into lynx:

$$\frac{dL}{dt} = e \cdot bHL - dL \quad (4.5)$$

(sometimes people just make a new parameter  $c = e \cdot b$ , but for now we will just leave it as is). Equation (4.6) shows the coupled system of differential equations:

$$\begin{aligned} \frac{dH}{dt} &= rH - bHL \\ \frac{dL}{dt} &= ebHL - dL \end{aligned} \quad (4.6)$$

The schematic diagram representing these interactions is the following is shown in Figure 4.3:

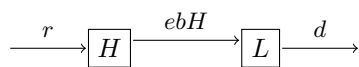


Figure 4.3: Schematic diagram Lynx-Hare system.

Equation (4.6) is a classical model in mathematical biology and differential equations - it is called the *predator prey* model, also known as the Lotka-Volterra equations. There is a lot of interesting mathematics from this system of equations that we will study later in this textbook. In later sections we will graphically and numerically analyze these equations and their solutions.

## 4.2 The Law of Mass Action

Notice in the previous section that the interaction between the lynx and the hare was of the form  $bHL$  - meaning you needed both positive values of  $H$  and  $L$  for the interaction to continue. This law states that the rate of a change is directly proportional to the *product* of the populations.

This assumption of the law of mass action is also commonly used in chemical reactions - especially in modeling enzyme dynamics. For example let's say you have a substrate  $A$  that reacts with enzyme  $B$  to form a product  $S$ . Perhaps you might have seen this represented as a reaction equation:



How we would write the product of formation, or  $\frac{dS}{dt}$  is the following:

$$\frac{dS}{dt} = kAB, \quad (4.8)$$

where  $k$  is the proportionality constant or the rate constant associated with the reaction. If we wanted to represent this as a schematic we would have the following diagram (Figure 4.4):

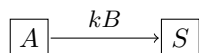


Figure 4.4: Schematic diagram of the law of mass action.

We could also consider if there was a constant decay of the substrate, which we might revise Figure 4.4 to Figure 4.5:

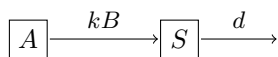


Figure 4.5: Revised schematic diagram of the law of mass action with decay.

For this case, the rate of change of  $S$  would then be:

$$\frac{dS}{dt} = kAB - dS, \quad (4.9)$$

You may be wondering about rates for  $A$  and  $B$ . When  $S$  is formed  $A$  and  $B$  are catalyzed, so the rate of formation for  $S$  (the positive term in Equation (4.9)) will be a loss for  $A$  and  $B$  (Equation (4.10)):

$$\begin{aligned} \frac{dA}{dt} &= -kAB \\ \frac{dB}{dt} &= -kAB \end{aligned} \quad (4.10)$$

Both equations are similar in this case, not pending any additional inputs or outputs.

### 4.3 Establishing species

Let's look at another example where from we will determine a differential equation model from a context:

An newly introduced plant species is introduced to a region. It competes with another established species for nutrients (and is a better competitor). However, the growth rate of the new species is proportional to the difference between the current number of established species and the number of new species. You may assume that the number of established species is a constant  $E$ .

For this problem we will start by naming our variables. Let  $N$  represent number of new species and  $E$  the number of established species. We will break this down accordingly:

- “the growth rate of the new species” means  $\frac{dN}{dt}$ .
- “is proportional to the difference between the current number of established species and the number of new species” means  $k \cdot (E - N)$ , where  $k$  is the proportionality constant. Including this parameter helps to avoid assuming we have a 1:1 correspondence between the growth rate of the new species and the population difference.
- “and is a better competitor” helps to explain why the term is  $k \cdot (E - N)$  instead of  $k \cdot (N - E)$ . We know that the newly established species will start out in much smaller numbers than  $N$ . But since it is a better competitor, we would expect its rate to increase initially. So  $\frac{dN}{dt}$  should be *positive* rather than negative. Assuming  $N < E$ , then  $E - N > 0$ , which guarantees that the new species will grow.

So this description could be modeled with Equation in summary we have the following model:

$$\frac{dN}{dt} = k(E - N) \quad (4.11)$$

Does this equation seem familiar to you? It is similar to the second model in Chapter 1 for the spread of Ebola! While this may seem surprising, it is often the case that similar equations appear in different contexts. It is far more advantageous to learn how to analyze models qualitatively rather than memorize several different types of models and not see the connections between them.

An interesting solution to a differential equation is the *steady state* or *equilibrium* solution. We find this where the rate equals zero. Let's take a look how to do that for our establishing plant model

**Example 4.1.** What is the steady state value for the differential equation  $\frac{dE}{dt} = k(E - N)$ ? (That is solve for  $E$  when  $\frac{dE}{dt} = 0$ .)

*Solution.* Let's solve  $\frac{dE}{dt} = k(E - N) = 0$ . For this equation we want to express the right hand side in terms of  $E$ . The parameter  $k$  is a constant  $k > 0$ , so really the steady state occurs when  $E - N = 0$ , or when  $N = E$ .

What this model tells us that eventually the new species will overtake the established species  $E$ .

### 4.4 Other types of functional responses

In several examples we have seen a rate of change proportional to the current population, as in the rate of growth of the hare population is  $rH$ . This is one example of what we would call a *functional response*. Another type of functional response assumes that the rate reaches a limiting value proportional to the population size, so  $\frac{dH}{dt} = \frac{rH}{1 + arH}$ . This is an example of a **type II functional response**. Finally the type II response has also been generalized (a **type III functional response**)  $\frac{dH}{dt} = \frac{rH^2}{1 + arH^2}$ . Figure 4.6 shows all three functional responses together:

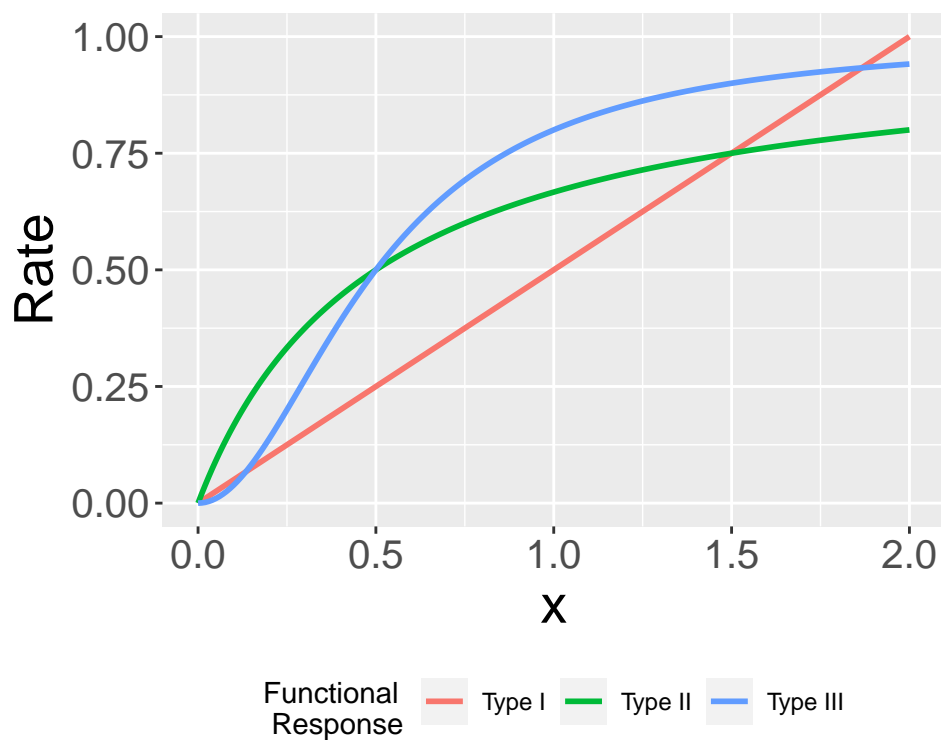


Figure 4.6: Comparison between Type I - Type III functional responses.

Notice the limiting behavior in the Type II and Type III functional responses. These responses are commonly used in ecology and predator-prey dynamics and in problems of how animals search for food.

## 4.5 Exercises

**Exercise 4.1.** Consider the following type of functional responses:

$$\text{Type I: } \frac{dP}{dt} = 0.1P \quad (4.12)$$

$$\text{Type II: } \frac{dP}{dt} = \frac{0.1P}{1 + .03P} \quad (4.13)$$

$$\text{Type III: } \frac{dP}{dt} = \frac{0.1P^2}{1 + .05P^2} \quad (4.14)$$

For each of the functional responses evaluate  $\lim_{P \rightarrow \infty} \frac{dP}{dt}$ . Since these functional responses represent a rate of change of a population, what are some examples (hypothetical or actual) would each of these responses be appropriate?

**Exercise 4.2.** A population grows according to the equation  $\frac{dP}{dt} = \frac{0.1P}{1 + .05P} - P$ .

- On the same axis, plot the equations  $f(P) = \frac{0.1P}{1 + .05P}$  and  $g(P) = P$ . What are the two positive values of  $P$  where  $f(P)$  and  $g(P)$  intersect?
- Next algebraically determine the two steady state values of  $P$ , that is solve  $\frac{dP}{dt} = 0$  for  $P$ . (*Hint:* factor a  $P$  out of the expression  $\frac{0.1P}{1 + .05P} - P$ .)
- Does your algebraic solution match your graphical solutions?

**Exercise 4.3.** A population grows according to the equation  $\frac{dP}{dt} = 2P - \frac{4P^2}{1 + P^2}$ .

- On the same axis, plot the equations  $f(P) = 2P$  and  $g(P) = \frac{4P^2}{1 + P^2}$ . What are the two positive values of  $P$  where  $f(P)$  and  $g(P)$  intersect?
- Next algebraically determine the two steady state values of  $P$ , that is solve  $\frac{dP}{dt} = 0$  for  $P$ . (*Hint:* factor a  $P$  out of the expression  $2P - \frac{4P^2}{1 + P^2}$ .)
- Does your algebraic solution match your graphical solutions?

**Exercise 4.4.** A population grows according to the equation  $\frac{dP}{dt} = \frac{aP}{1 + abP} - dP$ , where  $a$ ,  $b$  and  $d$  are parameters. Determine the two steady state values of  $P$ , that is solve  $\frac{dP}{dt} = 0$  for  $P$ .

**Exercise 4.5.** For a chemical reaction takes two chemicals  $X$  and  $Y$  to form a substrate  $Z$  through the law of mass action. However the substrate can also disassociate. The reaction schematic is the following:



where the proportionality constant  $k_+$  is associated with the formation of the substrate  $Z$  and  $k_-$  the disassociation ( $Z$  decays back to  $X$  and  $Y$ ).

Write down a differential equation that represents the rate of reaction  $\frac{dZ}{dt}$ .



**Exercise 4.6.** For each of the following exercises consider the following contextual situations modeling rates of change. Name variables for each situation and write down a differential equation describing the context. Be sure to identify and briefly describe any parameters you need for your model. For each problem you will need to:

- Name and describe all variables.
  - Write down a differential equation.
  - Identify and describe any parameters needed.
  - Write a brief one-two sentence explanation of why your differential equation models the situation at hand.
  - Hand sketch a rough graph of what you think the solution as a function of time - *note*: your solution needs to be consistent with your explanation and vice versa.
- a. The rate of change of an animal's body temperature is proportional to the difference in temperature between the environment.
  - b. A plant grows propotional to its current length  $L$ . Assume this proportionality constant is  $\mu$ , whose rate also decreases proportional to its current value. You will need to write down a system of two equation with variables  $L$  and  $\mu$ .
  - c. A patient undergoing chemotherapy receives an injection at rate  $I$ . This injection decreases the rate that a tumor accumulates mass. Independent of the injection, the tumor accumulates mass at a rate proportional to the mass of the tumor.
  - d. A cell with radius  $r$  assimilates nutrients at a rate proportional to its surface area, but uses nutrients proportional to its volume. Determine an equation that represents the rate of change of the radius.
  - e. A patient undergoing chemotherapy receives an injection at rate  $I$ . This injection decreases the rate that a tumor accumulates mass. Independent of the injection, the tumor accumulates mass at a rate proportional to the mass of the tumor.
  - f. The rate that a cancer cell divides (increases in amount) is proportional to the amount of healthy cells in its surrounding environment. You may assume that a healthy cell has a mortality  $\delta_H$  and a cancer cell has mortality  $\delta_C$ . Be sure to write down a system of differential equations for the population of cancer cells  $C$  and healthy cells  $H$ . The rate that a virus is spread to the population is proportional to the probability that a person is sick (out of  $N$  total sick and healthy individuals).
  - g.

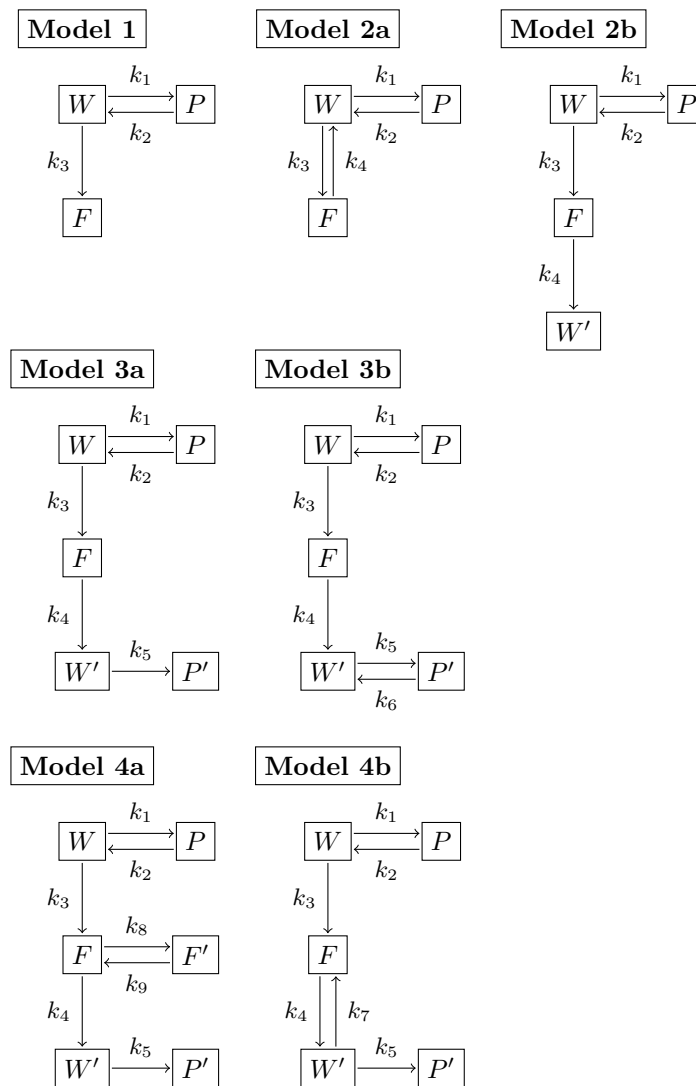


Figure 4.7: Reaction schemes.

**Exercise 4.7.** You are tasked with the job of investigating the effect of a pesticide on water quality, in terms of its effects on the health of the plants and fish in the ecosystem. Different models can be created that investigate the effect of the pesticide. Different types of reaction schemes for this system are shown in Figure 4.7, where  $F$  represents the amount of pesticide in the fish,  $W$  the amount of pesticide in the water, and  $S$  the amount of pesticide in the soil. The prime (e.g.  $F'$ ,  $W'$ , and  $S'$ ) represent other bound forms of the respective state). In all seven different models can be derived. For each of the model schematics, apply the Law of Mass Action to write down a system of differential equations.

## Chapter 5

# Euler's method

The focus of this section is on *approximation* of solutions to a differential equation via a numerical method. Typically a first numerical methods one might learn to tackle this problem is *Euler's method*, which is so fundamental it was popularized in the movie Hidden Figures.

The way we are going to do this is through expansion of the idea of a *locally linear approximation* to the tangent line. Let's start with an example.

**Example 5.1** (The flu bug). The rate of change of the flu through a population is given by the number of people infected  $t$  days after infection is,

$$\frac{dI}{dt} = 3e^{-.025t}.$$

Assuming that  $I(0) = 10$ , what is a locally linear approximation to this infection? Second, using your linear approximation, what would you predict is the value after one day ( $I(1)$ )?

*Remark.* In order to solve this problem, we know that the locally linear approximation is to  $I(t)$  at  $t = 0$  is  $L(t) = I(0) + I'(0) \cdot (t - 0)$ . Here,  $I(0) = 10$  and  $I'(0) = 3$ , so

$$L(t) = 10 + 3t$$

. Using  $L(t) \approx I(t)$ , we have  $L(1) = 10 + 3 = 13$ . So our model predicts there will be 13 people sick.

Notice in Example 5.1 we used two pieces of information: the (given) value of the function at  $t = 0$  and the estimate of the derivative from the rate of change.

It may be helpful to compare our prediction from  $L(1)$  to the actual value. The solution to the differential equation in Example 5.1 is  $I(t) = 130 - 120e^{-.025t}$  (you should verify this is the case by differentiation). Let's compare in the following table:

$t$	Linear approximation	Actual Solution
0	10	10
1	13	12.96

Not too bad, huh? Our approximation at  $L(1)$  is an *overestimate*, mainly because the actual solution is concave down, but it isn't that far off.

Let's build this solution out a little more by computing the rate of change at  $t = 1$ , assuming that thirteen people is a pretty close estimate of the  $I$  at  $t = 1$ . What we could do is to build *another* linear approximation using the differential equation. So the locally linear approximation is to  $I(t)$  at  $t = 1$  is

$$L(t) = I(1) + I'(1) \cdot (t - 1)$$

.

Here,  $I(1) = 13$  and  $I'(1) = 2.92$ , and  $L(t) = 13 + 2.92(t - 1)$ .

Assuming that  $L(t) \approx I(t)$ , we can evaluate  $L(t)$  at  $t = 2$  as an approximation for  $I(2)$ : have  $L(2) = 13 + 2.92 = 15.92$ . Comparing this to the actual solution at  $t = 2$ , we have  $I(2) = 15.85$ . Again, not too bad of a solution.

We can continue to build out the solution from there. Figure @ref(fig:eulers\_ver1) shows what we would have for a solution if we continued to build out this approach:

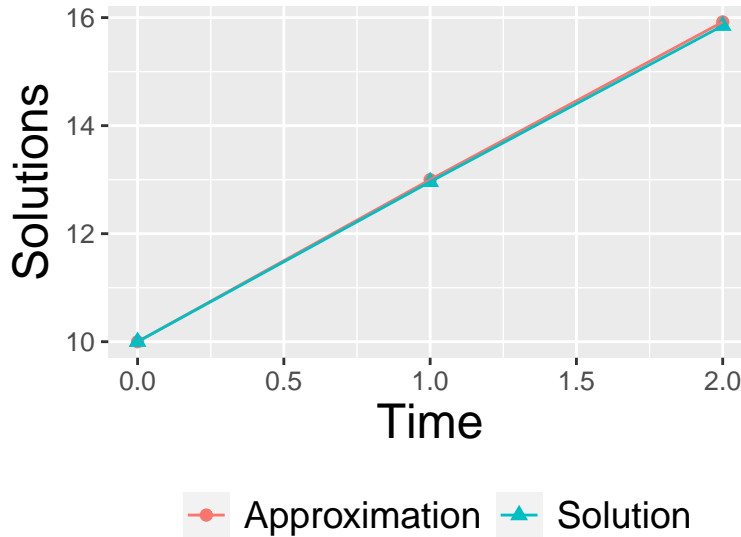


Figure 5.1: (#fig:eulers\_ver1) Approximation of a solution using local linearity

When you plot them they do look indistinguishable from each other by eye. It looks like we are onto something here!

## 5.1 Defining an Algorithm

Here would be an algorithm that would describe our process to determine a solution to a differential equation:

- Determine the locally linear approximation at a given point.
- Forecast out to another time value.
- Repeat the locally linear approximation.

If we continue on in this way, let's take a look at how our approximation would do after several days:

$t$	Approximate Solution	Actual Solution
90	118.4	117
95	119.9	118.6

Now it seems that our approximation isn't so accurate as time goes on. What if we updated the infection rate every half day? I know this means that we would be doing additional work (more iterations), but taking smaller timesteps goes hand in hand with more accurate solutions. Let's start out smaller with the first few timesteps:

$t$	$I$	$\frac{dI}{dt}$	$\frac{dI}{dt} \cdot \Delta t$
0	10	3	1.5
0.5	$= 10 + 1.5 = 11.5$	2.96	1.48
1	$= 11.5 + 1.48 = 12.98$	2.92	1.46
1.5	$= 12.92 + 1.46 = 14.38$	2.88	1.44
2	$= 14.38 + 1.44 = 15.82$		

Notice how we have started to build up a way to organize how to compute the solution. Each column is a "step" of the

method, computing the solution at a new timestep based on our step size  $\Delta t$ . The third column just computes the value of the derivative for a particular time and  $I$ , and then the fourth column is the *increment* size, or the amount we are forecasting the solution will grow by to the next timestep. (There are other ways to think about this, but if you have a *rate of change* multiplied by a time increment this will give you an approximation to the net change in a function.)

This idea of *approximate, forecast, repeat* is the heart of many numerical methods that approximate solutions to differential equations. The particular method that we have developed here is called *Euler's Method*. We display the results from additional steps in Figure 5.2.

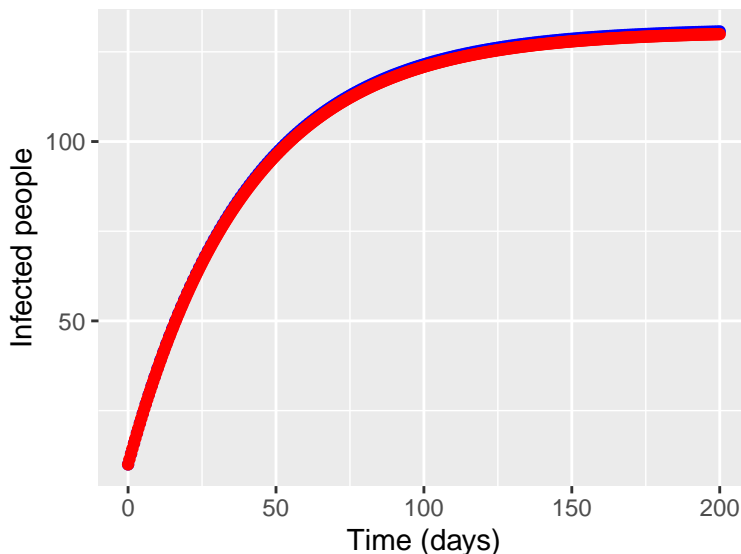


Figure 5.2: Approximation of a solution using local linearity

You may notice that the approximation in Figure @ref(fig:eulers\_ver2) compares very favorably to the actual solution function. At the end, we have the following comparisons:

$t$	Euler's Method ( $\Delta t = 1$ )	Euler's Method ( $\Delta t = 0.5$ )	Actual Solution
190	130.5	129.7	129
195	130.6	129.8	129.1
200	130.7	129.9	129.2

There is a tradeoff here - the smaller stepsizes you have the more work it will take to compute you solution. You may have seen a similar tradeoff in Calculus when you explored numerical integration and Riemann sums.

## 5.2 Building an iterative method

Now that we have worked on an example, let's carefully formulate Euler's method with another example. Consider the following differential equation:

Consider the following equation that describes the rate of change of the spread of a disease (such as Ebola, as we covered in the first section):

$$\frac{dI}{dt} = 0.003I \cdot (4000 - I)$$

Let's call the function  $f(I) = 0.003I \cdot (4000 - I)$ . In order to numerically approximate the solution, we will need to recall some concepts from calculus. One way that we can approximate the derivative is through a difference function:

$$\frac{dI}{dt} = \lim_{\Delta t \rightarrow 0} \frac{I(t + \Delta t) - I(t)}{\Delta t}$$

As long as we consider the quantity  $\Delta t$  to be small (say for this problem 0.1 days if you would like to have units attached to this), we can approximate the derivative with difference function on the right hand side. With this information, we have a reasonable way to organize the problem:

$$\begin{aligned}\frac{I(t + \Delta t) - I(t)}{\Delta t} &= 0.003I \cdot (4000 - I) \\ I(t + \Delta t) - I(t) &= 0.003I \cdot (4000 - I) \cdot \Delta t \\ I(t + \Delta t) &= I(t) + 0.003I \cdot (4000 - I) \cdot \Delta t\end{aligned}$$

The last equation  $I(t + \Delta t) = I(t) + 0.003I \cdot (4000 - I) \cdot \Delta t = f(I) \cdot \Delta t$  is a reasonable way to define an iterative system, especially if we have a spreadsheet program. Here is some code in R that can define a for loop to do this in an iterative way and then plot the solution with `plotData`:

```
# Define your timestep and time vector
deltaT <- 0.1
t <- seq(0,2,by=deltaT)

# Define the number of steps we take. This is equal to 10 / dt (why?)
N <- length(t)

# Define the initial condition
i_approx <- 10

# Define a vector for your solution: the derivative equation
for(i in 2:N) { # We start this at 2 because the first value is 10
  didt <- .003 * i_approx[i-1] * (4000-i_approx[i-1])
  i_approx[i] <- i_approx[i-1] + didt*deltaT
}

# Define your data for the solution into a tibble:
solution_data <- tibble(time = t,
                        infected = i_approx)

# Now plot your solution:
ggplot(data = solution_data) +
  geom_line(aes(x=time,y=infected))
```

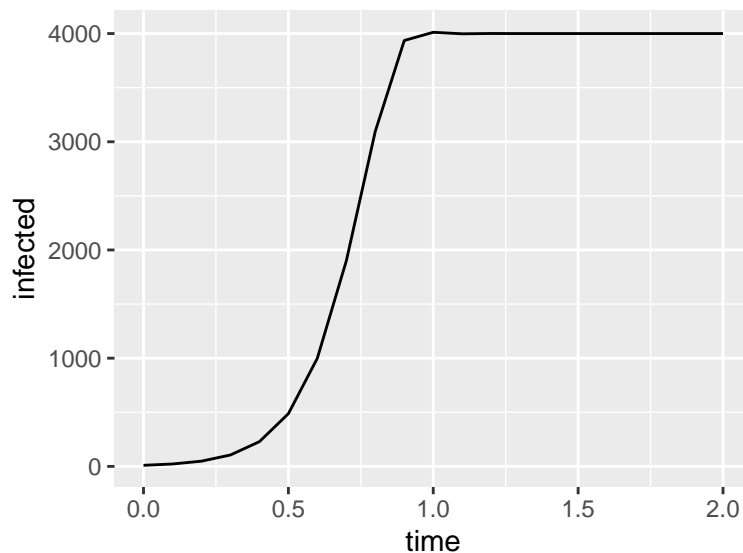


Figure 5.3: An iterative method

```
labs(x='Time',
     y='Infected')

## $x
## [1] "Time"
##
## $y
## [1] "Infected"
##
## attr("class")
## [1] "labels"
```

Ok, let's discuss the different aspects of this code:

- `deltaT <- 0.1` and `t <- seq(0,2,by=deltaT)` define the timesteps ( $\Delta t$ ) and the output time vector `t`. We also define `N <- length(t)` so we know how many steps we take.
- `i_approx <- 10` defines the initial condition to the system, in other words  $I(0) = 10$ .
- The for loop goes through this system - first computing the value of  $\frac{dI}{dt}$  and then forecasting out the next timestep  $I(t + \Delta t) = f(I) \cdot \Delta t$
- The remaining code plots the dataframe, like we learned in Section 3.

Let's recap what we've learned to summarize Euler's method. The most general form of a differential equation is:

$$\frac{d\vec{y}}{dt} = f(\vec{y}, \vec{\alpha}, t),$$

where  $\vec{y}$  is the vector of state variables you want to solve for, and  $\vec{\alpha}$  is your vector of parameters.

At a given initial condition, Euler's method applies locally linear approximations to forecast the solution forward  $\Delta t$  time units:

$$\vec{y}_{n+1} = y_n + f(\vec{y}_n, \vec{\alpha}, t_n) \cdot \Delta t$$

To generate Figure 5.3 we created the solution directly in R - but you don't want to copy and paste the code. I've created a function called `euler` that does the same process to generate the output solution:

```
# Define the rate equation:
system_eq <- c(didt ~ .003 * i * (4000-i))

# Define the initial condition (as a named vector)
init_cond <- c(i=10)

# Define deltaT and the time steps:
deltaT = 0.2
n_steps <- 10

# Compute the solution via Euler's method:
out_solution <- euler(system_eq, initial_condition=init_cond, deltaT=deltaT, n_steps = n_steps)

# Now plot your solution:
ggplot(data = out_solution) +
  geom_line(aes(x=t, y=i))
```

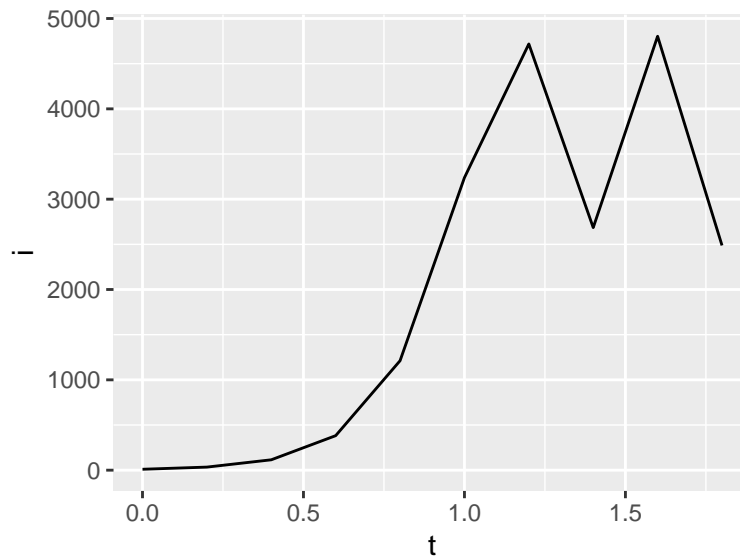


Figure 5.4: Euler's method solution

```
labs(x='Time',
     y='Infected')
```

```
## $x
## [1] "Time"
##
## $y
## [1] "Infected"
##
## attr(,"class")
## [1] "labels"
```

Let's talk through the steps of this code as well:

- The line `system_eq <- c(didt ~ .003 * i * (4000-i))` represents the differential equation, written in formula notation. So  $\frac{dI}{dt} \rightarrow \text{didt}$  and  $f(I) \rightarrow .003 * i * (4000-i)$ , with the variable `i`.
- The initial condition  $I(0) = 10$  is written as a **named vector**: `init_cond <- c(i=10)`. Make sure the name of the variable is consistent with your differential equation.
- As before we need to identify  $\Delta t$  and the number of steps  $N$ .

The command `euler` then computes the solution applying Euler's method, returning a data frame so we can plot the results. Note the columns of the data frame are the variables  $t$  and  $i$  that have been named in our equations.

### 5.3 Euler's method applied to systems

Here is a sample code that shows the dynamics for the Lotka-Volterra equations, as studied in Section 4:

$$\begin{aligned}\frac{dH}{dt} &= rH - bHL \\ \frac{dL}{dt} &= ebHL - dL\end{aligned}\tag{5.1}$$

We are going to use Euler's method to solve this differential equation. Similar to the previous example we will need to determine the  $f$

```
# Define the rate equation:
system_eq <- c(dHdt ~ r*H-b*H*L,
              dLdt ~ e*b*H*L-d*L)
```



```

# Define the parameters (as a named vector)
lynx_hare_params <- c(r = 2, b = 0.5, e = 0.1, d = 1) # parameters: a named vector

# Define the initial condition (as a named vector)
init_cond <- c(H=1, L=3)

# Define deltaT and the time steps:
deltaT <- 0.05 # timestep length
timeSteps <- 200 # must be a number greater than 1

# Compute the solution via Euler's method:
out_solution <- euler(system_eq,
                      parameters = lynx_hare_params,
                      initial_condition=init_cond,
                      deltaT=deltaT,
                      n_steps = n_steps)

# Make a plot of the solution, using different colors for lynx or hares.
ggplot(data = out_solution) +
  geom_line(aes(x=t,y=H),color='red') +
  geom_line(aes(x=t,y=L),color='blue') +
  labs(x='Time',
       y='Lynx (red) or Hares (blue)')

```

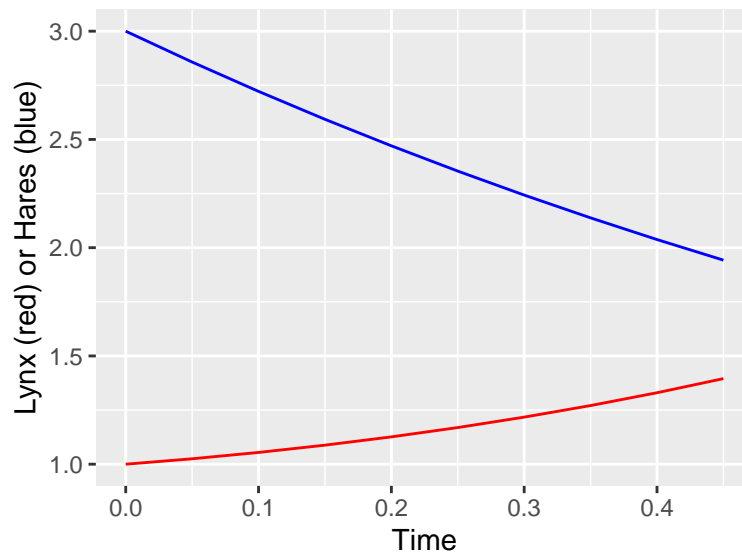


Figure 5.5: Euler's method solution for Lynx-Hare system

This example is structured similarly as a single variable differential equation, with some key changes:

- The variable `system_eq` is now a vector, with each entry one of the rate equations.
- We need to identify both variables in their initial condition.
- Most importantly, Equation (5.1) has parameters, which we define as a named vector `lynx_hare_params <- c(r = 2, b = 0.5, e = 0.1, d = 1)` that we pass through to the command `euler` with the option `parameters`. If your equation does not have any parameters you do not need to worry about it.
- We plot both solutions together at the end, or you can make two separate plots. Remember that you can choose the color in your plot.

Thankfully the code is pretty easy to adapt for systems of equations!

## 5.4 More refined numerical solvers

Perhaps in the course of working with Euler's method you encounter a differential equation that produces some nonsensical results. Take for example the following which is the implementation of our quarantine model:

```
system_eq <- c(dSdt ~ -k*S*I,
              dIdt ~ k*S*I-beta*I)

deltaT <- .1    # timestep length
timeSteps <- 15 # must be a number greater than 1

quarantine_parameters <- c(k=.05, beta=.2) # parameters: a named vector

init_cond <- c(S=300, I=1) # Be sure you have enough conditions as you do variables.

# Compute the solution via Euler's method:
out_solution <- euler(system_eq,
                      parameters = quarantine_parameters,
                      initial_condition=init_cond,
                      deltaT=deltaT,
                      n_steps = n_steps)

ggplot(data = out_solution) +
  geom_line(aes(x=t,y=S),color='red') +
  geom_line(aes(x=t,y=I),color='blue') +
  labs(x='Time',
       y='Susceptible (red) or Infected (blue)')
```

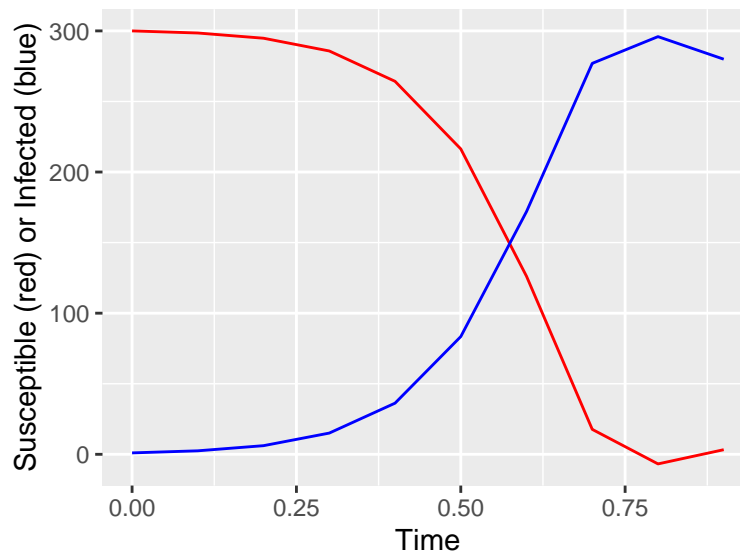


Figure 5.6: Surprising results with Euler's method.

You may notice the solution for  $S$  wiggles around  $t = 0.75$  and is negative. This is concerning because we know there can't be negative people! This requires a little more investigation.

If we take a look at  $t = 0.75$  the value for  $S \approx 1$  and the value for  $I \approx 280$ . If we let  $k = 0.05$  and  $\beta = 0.2$ , this means that  $\frac{dS}{dt} = -14$  and  $\frac{dI}{dt} = -42$ . The values of  $S$  and  $I$  are both decreasing! We know that our Euler's method update is one where the new value is the old value plus any change. So the new value for  $S = 1 - 14 \cdot 0.1 = -0.4$ . Mathematically Euler's method is doing nothing incorrect, but we know realistically this cannot happen.

It turns out that this can easily be overcome. While Euler's method is useful, it does quite poorly in cases where the solution

is changing rapidly - or we might need to make some smaller step sizes. How we remedy this is to use a *higher order solver*, and one such method is called the Runge-Kutta Method. If you take a course in numerical analysis you might study these, but for the moment you see the difference between the twoa Runge-Kutta solver implemented in the MAT369Code package, which by all intents and purposes is replaces the command `euler` with `rk4`:

```
system_eq <- c(dSdt ~ -k*S*I,
              dIdt ~ k*S*I-beta*I)

deltaT <- .1    # timestep length
timeSteps <- 15 # must be a number greater than 1

quarantine_parameters <- c(k=.05, beta=.2) # parameters: a named vector

init_cond <- c(S=300, I=1) # Be sure you have enough conditions as you do variables.

# Compute the solution via a Runge-Kutta method:
out_solution <- rk4(system_eq,
                    parameters = quarantine_parameters,
                    initial_condition=init_cond,
                    deltaT=deltaT,
                    n_steps = n_steps)

ggplot(data = out_solution) +
  geom_line(aes(x=t,y=S),color='red') +
  geom_line(aes(x=t,y=I),color='blue') +
  labs(x='Time',
       y='Susceptible (red) or Infected (blue)')
```

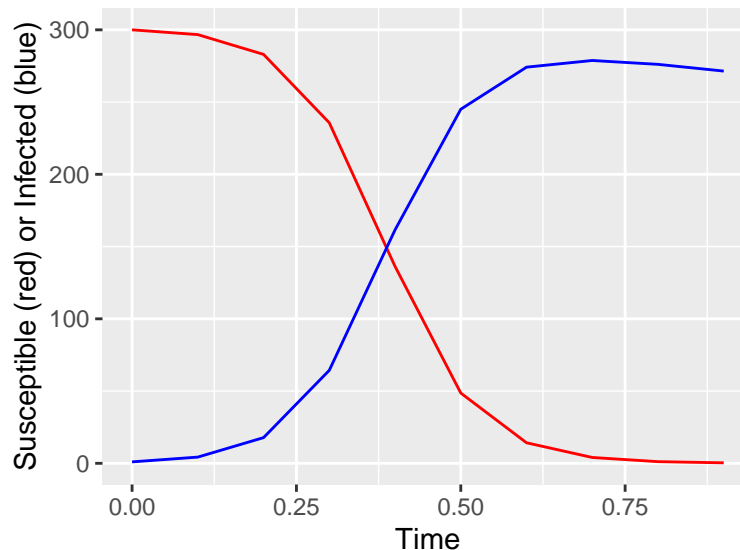


Figure 5.7: Better results with the Runge-Kutta method.

So what is going on here? Briefly, the differences between the two methods have to do with the error in the numerical method. The error is quantified as the difference between the actual solution and the numerical solution. Euler's method has an error on the order of the stepsize  $\Delta t$ , whereas the Runge-Kutta method has an error of  $(\Delta t)^4$ . For this example  $\Delta t = .1$ , so  $(\Delta t)^4 = .0001$  - that is a noticeable difference! We can improve Euler's method by taking a smaller timestep - BUT that means we need a larger number of steps  $N$  - which may take more computational time.

Does this discussion sound familiar? Perhaps you examined similar when you took calculus and studied Riemann sums to approximate the area underneath a curve (left sum, right sum, trapezoid, midpoint). It turns out that these two problems are closely related. Numerical analysis is a great field of study to examine these topics and others!

Moving ahead, I might switch between Euler's method or the Runge-Kutta method when solving a differential equation. Thankfully the bulk of the work is “under the hood” - the setup will be the same for both!

## 5.5 Exercises

**Exercise 5.1.** Verify that  $I(t) = 130 - 120e^{-0.25t}$  is a solution to the differential equation

$$\frac{dI}{dt} = 130 - 0.025I$$

with  $I(0) = 10$ .

**Exercise 5.2.** Apply the `rk4` solver with  $\Delta t = 0.1$  with  $N = 10$  to the initial value problem  $\frac{dI}{dt} = 0.003I \cdot (4000 - I)$   $I(0) = 10$ . Compare your graph to Figure 5.3. What differences do you observe? Which solution method (`euler` or `rk4`) is better (and why)?

**Exercise 5.3.** The following exercise will help you explore the relationships between stepsize, ending points, and number of steps needed. You may assume that we will start at  $t = 0$  in all parts.

- If we wish to do an Euler's method solution with step size 1 second and ending at  $t = 5$  seconds, how many steps will we take?
- If we wish to do an Euler's method solution with step size 0.5 seconds and ending at  $t = 5$  seconds, how many steps will we take?
- If we wish to do an Euler's method solution with step size 0.1 seconds and ending at  $t = 5$  seconds, how many steps will we take?
- If we wish to do an Euler's method solution with step size  $\Delta t$  and go to ending value of  $T$ , what is an expression that relates the number steps  $N$  as a function of  $\Delta t$  and  $T$ ?

**Exercise 5.4.** To get a rough approximation between error and step size, let's say for a particular differential equation that we are starting at  $t = 0$  and going to  $t = 2$ , with  $\Delta t = 0.2$ . We know that the Runge-Kutta error will be on the order of  $(\Delta t)^4 = 0.0016$ . If we want to use Euler's method with the same order of error, we could say  $\Delta t = .0001$ . For that case, how many steps will we need to take?

**Exercise 5.5.** For each of the following differential equations, apply Euler's method to generate a numerical solution to the differential equation and plot your solution. The stepsize ( $\Delta t$ ) and number of iterations ( $N$ ) are listed.

- Differential equation:  $\frac{dS}{dt} = 3 - S$ . Set  $\Delta t = 0.1$ ,  $N = 50$ . Initial conditions:  $S(0) = 0.5$ ,  $S(0) = 5$ .
- Differential Equation:  $\frac{dS}{dt} = \frac{1}{1 - S}$ . Set  $\Delta t = 0.01$ ,  $N = 30$ . Initial conditions:  $S(0) = 0.5$ ,  $S(0) = 2$ .
- Differential equation:  $\frac{dS}{dt} = 0.8 \cdot S \cdot (10 - S)$ . Set  $\Delta t = 0.1$ ,  $N = 50$ . Initial conditions:  $S(0) = 3$ ,  $S(0) = 10$ .

**Exercise 5.6.** For each of the following differential equations, apply the Runge-Kutta method method to generate a numerical solution to the differential equation and plot your solution. The stepsize ( $\Delta t$ ) and number of iterations ( $N$ ) are listed. Contrast your answers with Exercise 5.5.

- Differential equation:  $\frac{dS}{dt} = 3 - S$ . Set  $\Delta t = 0.1$ ,  $N = 50$ . Initial conditions:  $S(0) = 0.5$ ,  $S(0) = 5$ .
- Differential Equation:  $\frac{dS}{dt} = \frac{1}{1 - S}$ . Set  $\Delta t = 0.01$ ,  $N = 30$ . Initial conditions:  $S(0) = 0.5$ ,  $S(0) = 2$ .
- Differential equation:  $\frac{dS}{dt} = 0.8 \cdot S \cdot (10 - S)$ . Set  $\Delta t = 0.1$ ,  $N = 50$ . Initial conditions:  $S(0) = 3$ ,  $S(0) = 10$ .

**Exercise 5.7.** Let's do some more work with Euler's method for  $\frac{dS}{dt} = 0.8 \cdot S \cdot (10 - S)$ . This time set  $S(0) = 15$ ,  $\Delta t = 0.1$ ,  $N = 10$ . When you examine your solution, what is incorrect about the Euler's method solution based on your qualitative knowledge of the underlying dynamics? Now calculate Euler's method for the same differential equation for the following conditions:  $S(0) = 15$ ,  $\Delta t = 0.01$ ,  $N = 100$ . What has changed in your solution?

**Exercise 5.8.** Let's do some more work with Euler's method for  $\frac{dS}{dt} = \frac{1}{1-S}$ . This time set  $S(0) = 1.5$ ,  $\Delta t = 0.1$ ,  $N = 10$  and also  $S(0) = 1.5$ ,  $\Delta t = 0.01$ ,  $N = 100$ . Between these two solutions, what has changed? Do you think it is numerically possible to calculate a reasonable solution for Euler's method near  $S = 1$ ? (*note: this differential equation is an example of finite time blow up*)

**Exercise 5.9.** Similar to Exercise 5.8, let's apply the `rk4` method for  $\frac{dS}{dt} = \frac{1}{1-S}$ . This time set  $S(0) = 1.5$ ,  $\Delta t = 0.1$ ,  $N = 10$  and also  $S(0) = 1.5$ ,  $\Delta t = 0.01$ ,  $N = 100$ . Between these two solutions, what has changed? Does this numerical solver do a better job in computing solutions compared to the Euler method? (*note: this differential equation is an example of finite time blow up*)

**Exercise 5.10.** One way to model the growth rate hares is with  $f(H) = \frac{rH}{1+kH}$ , where  $r$  and  $k$  are parameters. This is in contrast to exponential growth, which assumes  $f(H) = rH$ .

- First evaluate  $\lim_{H \rightarrow \infty} rH$ .
- Then  $\lim_{H \rightarrow \infty} \frac{rH}{1+kH}$ .
- Compare your two answers. Discuss how the growth rate  $f(H) = \frac{rH}{1+kH}$  seems to be a more realistic model.

**Exercise 5.11.** In the lynx hare example we can also consider an alternative system where the growth of the hare is not exponential:

$$\begin{aligned}\frac{dH}{dt} &= \frac{2H}{1+kH} - 0.5HL \\ \frac{dL}{dt} &= 0.05HL - dL\end{aligned}\tag{5.2}$$

Set the number of timesteps to be 2000. Apply Euler's method to numerically solve this system of equations when  $k = 0.1$  and  $k = 1$ . Plot your simulation results.

For the following differential equations use the code `euler` to generate a numerical solution to the differential equation and plot your solution. The stepsize ( $\Delta t$ ) and number of iterations ( $N$ ) are listed. After your Euler's method solution is determined, explain why you would expect the behavior in the solution that you see.

**Exercise 5.12.** Consider the following:

- Differential equation:  $\frac{dS}{dt} = r \cdot S \cdot (K - S)$ .
  - Set  $r = 1.2$  and  $K = 3$ .
  - Set  $\Delta t = 0.1$ ,  $N = 50$ .
- Use the code `'euler'` to generate a numerical solution to the differential equation for the following initial conditions:  $S(0) = 1$ ,  $S(0) = 3$ ,  $S(0) = 5$ .
  - Plot your Euler's method solutions with the three initial conditions on the same plot. What do you notice when you do plot them together?

- c. Make a hypothesis regarding the long term behavior of this system. Then plot a few more solution curves to verify your guess.

**Exercise 5.13.** Consider the following:

- Differential equation:  $\frac{dS}{dt} = K - S$ .
  - Set  $K = 2$ .
  - Set  $\Delta t = 0.1$ ,  $N = 50$ .
- a. Use the code ‘euler’ to generate a numerical solution to the differential equation for the following initial conditions:  $S(0) = 0$ ,  $S(0) = 2$ ,  $S(0) = 5$ .
- b. Plot your Euler’s method solutions with the three initial conditions on the same plot. What do you notice when you do plot them together?
- c. Make a hypothesis regarding the long term behavior of this system. Then plot a few more solution curves to verify your guess.

**Exercise 5.14.** Consider the differential equation  $\frac{dS}{dt} = \frac{1}{1-S}$ . Notice that at  $S = 1$  the rate  $\frac{dS}{dt}$  is not defined.

- a. If you applied Euler’s method solution with initial condition  $S(0) = 0.9$ , what do you think your solution would approach as the number of timesteps increased?
- b. If you applied Euler’s method solution with initial condition  $S(0) = 1.1$ , what do you think your solution would approach as the number of timesteps increased?
- c. Explain how you could come to the same conclusion as the previous two problems if you graphed  $f(S) = \frac{1}{1-S}$ .





## Chapter 6

# Phase lines and equilibrium solutions

In modeling with differential equations, we want to understand how a system develops both qualitatively and quantitatively. Euler's method (and other associated numerical methods for solving differential equations) illustrate solution behavior numerically. One key thing about the qualitative analysis is we are interested in the *motion* or the “flow” of the solution at a given point. Is the solution increasing, decreasing, or staying the same?

For this section we will discuss qualitative aspects of a differential equation. We are going to focus on differential equations in one variable. Section 7 will build your understanding of the same idea with coupled systems of equations.

### 6.1 Equilibrium solutions

A great place to start is where the rate of change for a differential equation is zero, or in other words there is *no* flow. Borrowing ideas from calculus, this occurs when the rate of change is zero. We solve this by setting the left hand side of  $\frac{dy}{dt} = f(y)$  equal to zero and solving for  $y$  (or whatever dependent variable describes the problem).

**Example 6.1.** What are the equilibrium solutions to  $\frac{dy}{dt} = -y$ ?

*Solution.* For this example we know that when the rate of change is zero, this means that  $\frac{dy}{dt} = 0$ , or when  $0 = -y$ . So  $y = 0$  is the equilibrium solution.

The general solution to the differential equation  $\frac{dy}{dt} = -y$  is when  $y(t) = Ce^{-t}$ , where  $C$  is an arbitrary constant. Figure 6.1 plots different solution curves, with the equilibrium solution shown as a horizontal line:

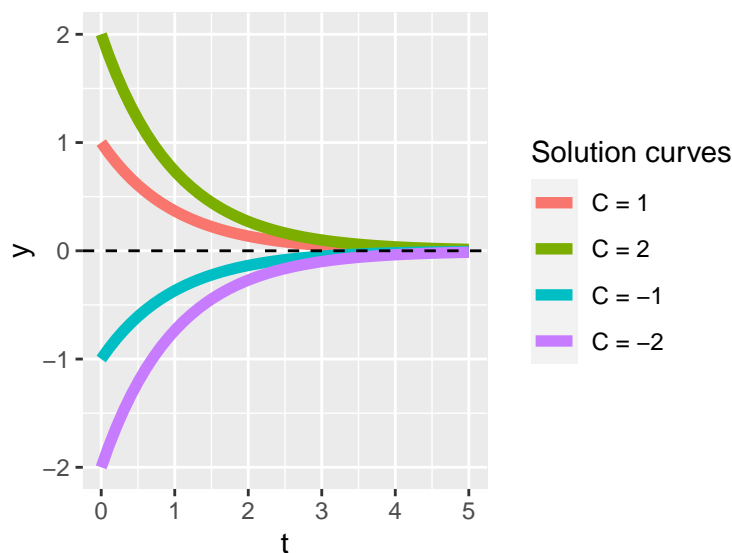


Figure 6.1: Exponential solution curves

Notice that as  $t$  increases, all solutions approach the equilibrium solution  $y = 0$  as  $t$  no matter if the initial condition is positive or negative. You can also evaluate the following limit for the solution:  $\lim_{t \rightarrow \infty} Ce^{-t} = 0$  to verify this is the case too.

**Example 6.2.** What are the equilibrium solutions to  $\frac{dN}{dt} = N \cdot (1 - N)$ ?

*Solution.* In this case the equilibrium solutions occur when  $N \cdot (1 - N) = 0$ , or when  $N = 0$  or  $N = 1$ .

The generic solution to this differential equation is

$$N(t) = \frac{N_0}{N_0 + (1 - N_0)e^{-t}}.$$

Figure 6.2 displays several different solution curves.

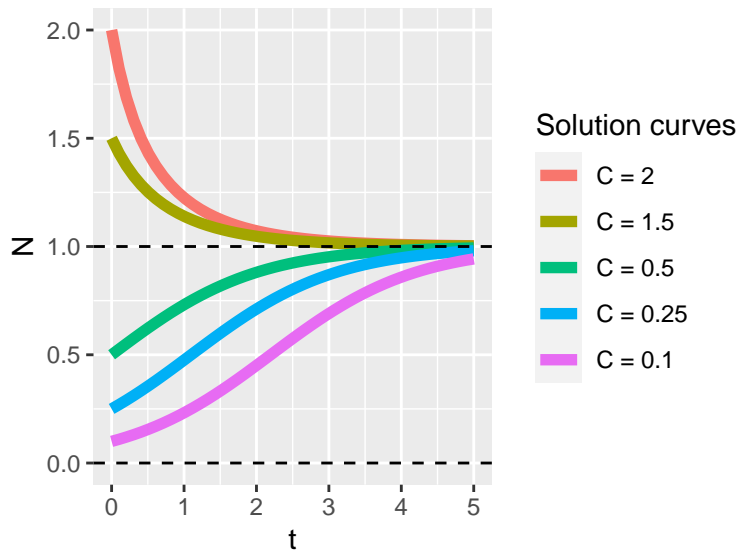


Figure 6.2: Solution curves for  $N' = N(1 - N)$

In Figure 6.2 notice how all the solutions tend towards  $N = 1$ , but even solutions that start close to  $N = 0$  seem to move away from this equilibrium solution. This example brings us to understanding classifying the *stability* of the equilibrium solutions.

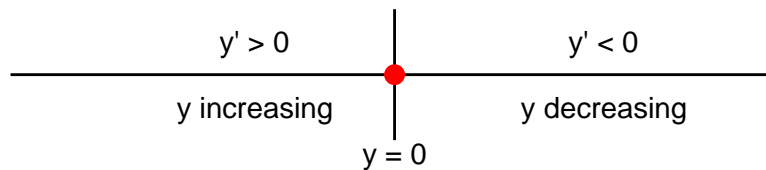
## 6.2 Phase lines for differential equations

While it is one thing to determine where the equilibrium solutions are, we are also interested in classifying the **stability** of the equilibrium solutions. To do this investigate the behavior of the differential around the equilibrium solutions, using facts from calculus:

- If  $\frac{dy}{dt} < 0$ , the function is decreasing.
- If  $\frac{dy}{dt} > 0$ , the function is increasing.

We say that the solution  $y = 0$  is a *stable* equilibrium solution in this case.

Let's apply this logic to our differential equation  $\frac{dy}{dt} = -y$ . We know that if  $y = 3$ ,  $\frac{dy}{dt} = -3 < 0$ , so we say the function is *decreasing* to  $y = 0$ . If  $y = -2$ ,  $\frac{dy}{dt} = -(-2) = 2 > 0$ , so we say the function is *increasing* to  $y = 0$ . This can be represented neatly in the phase line diagram for Figure 6.3:

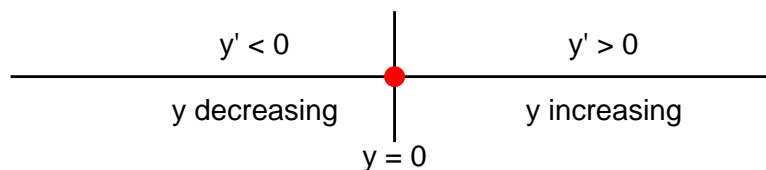
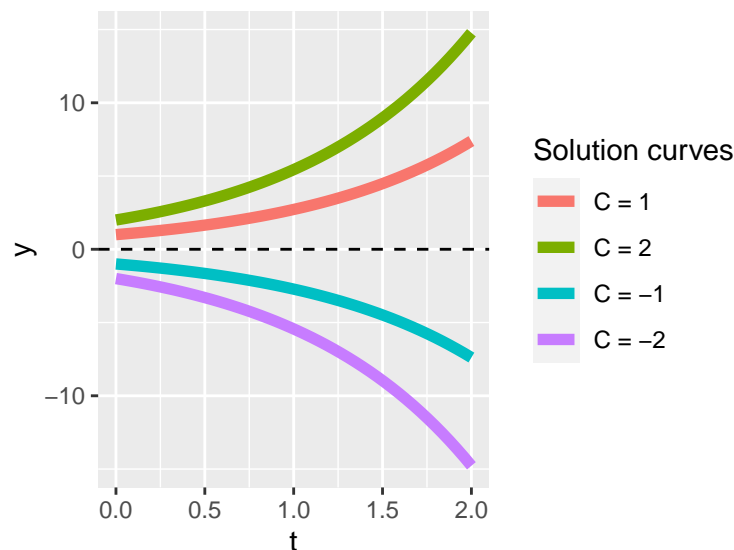
Figure 6.3: Phase line to  $y' = -y$ .

Because the solution is *increasing* to  $y = 0$  when  $y < 0$ , and *decreasing* to  $y = 0$  when  $y > 0$ , we say that the equilibrium solution is **stable**, which is also confirmed by the solutions we plotted above.

**Example 6.3.** Classify the stability of the equilibrium solutions to  $\frac{dy}{dt} = k \cdot y$ , where  $k$  is a parameter.

*Solution.* In this case the equilibrium solution is still  $y = 0$ . We will need to consider two different cases for the stability depending on the value of  $k$  ( $k > 0$ ,  $k < 0$ , and  $k = 0$ ):

- When  $k < 0$ , the phase line will be similar Figure 6.3.
- When  $k > 0$  the phase line will be is shown in the Figure 6.4. We say in this case that the equilibrium solution is *unstable*, as all solutions flow away from the equilibrium. Several different solutions are shown in Figure 6.5
- When  $k = 0$  we have the differential equation  $\frac{dy}{dt} = 0$ , which has  $y = C$  as a general solution. For this special case the equilibrium solution is neither stable or unstable<sup>1</sup>.

Figure 6.4: Phaseline for  $y' = ky$ , with  $k > 0$ .Figure 6.5: Solution curves for  $y' = ky$ , with  $k > 0$ .

**Example 6.4.** Let's investigate the phase line for the differential equation  $\frac{dN}{dt} = N \cdot (1 - N)$  and classify stability of the equilibrium solutions.

<sup>1</sup>By all intents and purposes this is a different differential equation than  $\frac{dy}{dt} = k \cdot y$ ; something peculiar is going on here - which we come back to when discuss bifurcations in Section ??.

*Solution.* This differential equation has equilibrium solutions when  $N(1 - N) = 0$ , or  $N = 0$  or  $N = 1$ . We evaluate the stability of the solutions in the following table:

Test point	Sign of $N'$	Tendency of solution
$N = -1$	Negative	Decreasing
$N = 0$	Zero	Equilibrium solution
$N = 0.5$	Positive	Increasing
$N = 1$	Zero	Equilibrium solution
$N = 2$	Negative	Decreasing

Notice how the selected test points in the first column are either the the *left* or the *right* of the equilibrium solution. We can also represent the information in the table using a phase line diagram (Figure 6.6), but in this case we need to include *two* equilibrium solutions.

The table and Figure 6.6 confirms that  $N$  is moving *away* from  $N = 0$  (either decreasing when  $N$  is less than 0 and increasing when  $N$  is greater than 0) and moving *towards*  $N = 1$  (either increasing when  $N$  is between 0 and 1 and decreasing when  $N$  is greater than one).

These results suggest that equilibrium solution at  $N = 0$  to be *unstable* and at  $N = 1$  to be *stable*.

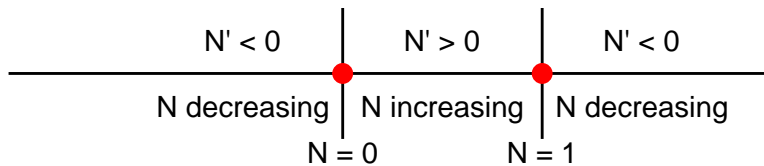


Figure 6.6: Phase line diagram for the differential equation  $N' = N(1 - N)$ .

Other than writing the words in the phase line diagram, we also use arrows to signify increasing or decreasing in the solutions.

### 6.3 A stability test for equilibrium solutions

Notice how when constructing the phase line diagram we relied on the behavior of solutions *around* the equilibrium solution to classify the stability. As an alternative we can also use the point at the equilibrium solution itself.

To do this we are going to consider the general differential equation  $\frac{dy}{dt} = f(y)$ . We are going to assume that we have an equilibrium solution at  $y = y_*$ .

We use local linearization to construct a locally linear approximation to  $L(y)$  to  $f(y)$  at  $y = y_*$ :

$$L(y) = f(y_*) + f'(y_*) \cdot (y - y_*)$$

We will use  $L(y)$  as an approximation to  $f(y)$ . There are two key things here. First, because we have an equilibrium solution,  $f(y_*) = 0$ . The other key thing is that if we define the variable  $P = y - y_*$ , then the differential equation translates to

$$\frac{dP}{dt} = f'(y_*) \cdot P \quad (6.1)$$

Does Equation (6.1) look familiar? I - it should! This is similar to the example where we classified the stability of  $\frac{dy}{dt} = k \cdot y$  - cool! So let's use what we learned in Example 6.3 above to classify the stability:

**Local linearization stability test for equilibrium solutions:** For a differential equation  $\frac{dy}{dt} = f(y)$  with equilibrium solution  $y_*$ , we can classify the stability of the equilibrium solution through the following:

- If  $f'(y_*) > 0$  at an equilibrium solution, the equilibrium solution  $y = y_*$  will be *unstable*.
- If  $f'(y_*) < 0$  at an equilibrium solution, the equilibrium solution  $y = y_*$  will be *stable*.
- If  $f'(y_*) = 0$ , we cannot conclude anything about the stability of  $y = y_*$ .

**Example 6.5.** Apply local linearization to classify the stability of the equilibrium solutions of  $\frac{dN}{dt} = N \cdot (1 - N)$

*Solution.* The locally linear approximation is  $L(N) = 1 - 2N$ . We have  $L(0) = 1 > 0$ , so  $N = 0$  is unstable. Similarly  $L(1) = -1$ , so  $N = 1$  is stable.

## 6.4 Exercises

**Exercise 6.1.** What are the equilibrium solutions to the following differential equations?

a.  $\frac{dS}{dt} = 0.3 \cdot (10 - S)$

b.  $\frac{dP}{dt} = P \cdot (P - 1)(P - 2)$

Then classify the stability of the equilibrium solutions using the local linearization stability test.

**Exercise 6.2.** Using your results from Exercise 6.1, construct a phase line for each of the differential equations and classify the stability of the equilibrium solutions.

**Exercise 6.3.** A population grows according to the equation  $\frac{dP}{dt} = \frac{P}{1 + 2P} - 0.2P$ .

- Determine the equilibrium solutions for this differential equation.
- Classify the stability of the equilibrium solutions using the local linearization stability test.

**Exercise 6.4.** A cell with radius  $r$  assimilates nutrients at a rate proportional to its surface area, but uses nutrients proportional to its volume, according to the following differential equation:

$$\frac{dr}{dt} = 4\pi r^2 - \frac{4}{3}\pi r^3.$$

- Determine the equilibrium solutions for this differential equation.
- Construct a phase line for this differential equation to classify the stability of the equilibrium solutions.

**Exercise 6.5.** The Chanter equation of growth is the following, where  $W$  is the weight of an object:

$$\frac{dW}{dt} = W(3 - W)e^{-Dt}, \quad (6.2)$$

Use this differential equation to answer the following questions.

- What happens to the rate of growth ( $\frac{dW}{dt}$ ) as  $t$  grows large?
- What are the equilibrium solutions to this model? Are they stable or unstable?
- Notice how the equilibrium solutions are the same as those for the logistic model. Based on your previous work, do you think this model would be a more realistic model of growth than the logistic model  $\frac{dW}{dt} = W(3 - W)$ ?

**Exercise 6.6.** Red blood cells are formed from stem cells in the bone marrow. The red blood cell density  $r$  satisfies an equation of the form

$$\frac{dr}{dt} = \frac{br}{1 + r^n} - cr, \quad (6.3)$$

where  $n > 1$  and  $b > 1$  and  $c > 0$ . Find all the equilibrium solutions  $r_*$  to this differential equation. *Hint:* can you factor an  $r$  from your equation first?

**Exercise 6.7.** Organisms that live in a saline environment biochemically maintain the amount of salt in their blood stream. An equation that represents the level of  $S$  in the blood is the following:

$$\frac{dS}{dt} = I + p \cdot (W - S)$$

Where the parameter  $I$  represents the active uptake of salt,  $p$  is the permeability of the skin, and  $W$  is the salinity in the water.

- First set  $I = 0$ . Determine the equilibrium solutions for this differential equation. Express your answer  $S_*$  in terms of the parameters  $p$ , and  $W$ .
- Next consider  $I > 0$ . Determine the equilibrium solutions for this differential equation. Express your answer  $S_*$  in terms of the parameters  $p$ ,  $W$ , and  $I$ . Why should your new equilibrium solution be greater than the equilibrium solution from the previous problem?
- Classify the stability of both equilibrium solutions using the local linearization stability test.

**Exercise 6.8.** The immigration rate of bird species (species per time) from a mainland to an offshore island is  $I_m \cdot (1 - S/P)$ , where  $I_m$  is the maximum immigration rate,  $P$  is the size of the source pool of species on the mainland, and  $S$  is the number of species already occupying the island. Further, the extinction rate is  $E \cdot S/P$ , where  $E$  is the maximum extinction rate. The growth rate of the number of species on the island is the immigration rate minus the extinction rate, given by the following differential equation:

$$\frac{dS}{dt} = I_m \left(1 - \frac{S}{P}\right) - \frac{ES}{P} \quad (6.4)$$

- Determine the equilibrium solutions  $S_*$  for this differential equation. Expression your answer in terms of  $I_m$ ,  $P$ , and  $E$ .
- Classify the stability of the equilibrium solutions using the local linearization stability test.

**Exercise 6.9.** A colony of bacteria growing in a nutrient-rich medium deplete the nutrient as they grow. As a result, the nutrient concentration  $x(t)$  is steadily decreasing. The equation describing this decrease is the following:

$$\frac{dx}{dt} = -\mu \frac{x \cdot (\xi - x)}{\kappa + x},$$

where  $\mu$ ,  $\kappa$ , and  $\xi$  are all parameters greater than zero.

- Determine the equilibrium solutions  $x_*$  for this differential equation.
- Construct a phase line for this differential equation and classify the stability of the equilibrium solutions.

**Exercise 6.10.** Can a solution curve cross an equilibrium solution of a differential equation?





## Chapter 7

# Coupled systems of equations

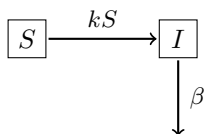
In this section we will learn how to qualitatively understand systems of differential equations. When analyzing a single differential equation we used the idea of a phase line to understand if a solution was stable or unstable. Here we extend that to equations of more than one variable and investigate what we will call the *phase plane*.

### 7.1 Model redux: flu with quarantine

In Section 2 we studied the following model for the flu as a coupled system of equations:

$$\begin{aligned}\frac{dS}{dt} &= -kSI \\ \frac{dI}{dt} &= kSI\end{aligned}$$

In this scenario we are also going to consider that those who are infected are quarantined, proportional to the number infected, according to the following schematic:



Which gives us the following system of equations:

$$\begin{aligned}\frac{dS}{dt} &= -kSI \\ \frac{dI}{dt} &= kSI - \beta I\end{aligned}\tag{7.1}$$

To find the equilibrium solutions we want to find values of  $S$  and  $I$  where the rates  $\frac{dS}{dt}$  and  $\frac{dI}{dt}$  are *both* zero. This can be done by algebraically solving the system of equations:

$$\begin{aligned}0 &= -kSI \\ 0 &= kSI - \beta I\end{aligned}$$

Let's examine the first equation ( $0 = -kSI$ ), which we can see is consistent when either  $S = 0$  and  $I = 0$ . These give us two options, which we then use in the second equation ( $0 = kSI - \beta I$ ). When  $S = 0$ , then  $0 = k \cdot 0 \cdot I - \beta I \rightarrow 0 = -\beta I$ , which is consistent when  $I = 0$ . So  $(S_*, I_*) = (0, 0)$  is one equilibrium solution. (In fact, if  $I = 0$ , then any value of  $S$  would be an equilibrium solution. Can you explain why?)

This is an interesting example. We call the equations  $S = 0$  and  $I = 0$  when  $\frac{dS}{dt} = 0$  as *nullclines* for  $S$ . In a similar manner, the equations in  $S$  and  $I$  when  $\frac{dI}{dt} = 0$  are called *nullclines* for  $I$ . Let's try to determine formulas for these equations:

$$\begin{aligned} 0 &= kSI - \beta I \\ 0 &= I \cdot (kS - \beta) \end{aligned}$$

Because the last equation is factored as a product, nullclines for  $I$  are either  $I = 0$  or  $S = \frac{\beta}{k}$ .

Nullclines are not equilibrium solutions by themselves - it is the *intersection* of two different nullclines that determine equilibrium solutions. Figure 7.1 shows the nullclines in the  $S-I$  plane (since we have two equations), with  $S$  on the horizontal axis and  $I$  on the vertical axis.

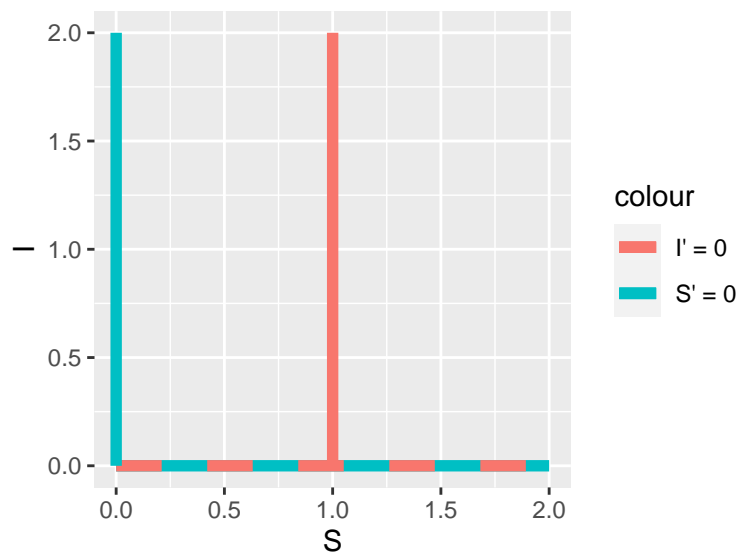


Figure 7.1: Nullclines for Equation (7.1). To generate the plot we assumed  $\beta = 1$  and  $k = 1$

```
## <ggproto object: Class ScaleDiscrete, Scale, gg>
##   aesthetics: colour
##   axis_order: function
##   break_info: function
##   break_positions: function
##   breaks: S' = 0 I' = 0
##   call: call
##   clone: function
##   dimension: function
##   drop: TRUE
##   expand: waiver
##   get_breaks: function
##   get_breaks_minor: function
##   get_labels: function
##   get_limits: function
##   guide: legend
##   is_discrete: function
##   is_empty: function
##   labels: waiver
##   limits: NULL
##   make_sec_title: function
##   make_title: function
##   map: function
```

Table 7.1: Values of  $\frac{dS}{dt}$  and  $\frac{dI}{dt}$  for the flu with quarantine model.

S	I	dS_dt	dI_dt
0.0000000	0.0000000	0.0000000	0.0000000
0.4444444	0.4444444	-0.1975309	-0.2469136
0.8888889	0.8888889	-0.7901235	-0.0987654
1.3333333	1.3333333	-1.7777778	0.4444444
1.7777778	1.7777778	-3.1604938	1.3827160
2.2222222	2.2222222	-4.9382716	2.7160494
2.6666667	2.6666667	-7.1111111	4.4444444
3.1111111	3.1111111	-9.6790123	6.5679012
3.5555556	3.5555556	-12.6419753	9.0864198
4.0000000	4.0000000	-16.0000000	12.0000000

```
## map_df: function
## n.breaks.cache: NULL
## na.translate: TRUE
## na.value: grey50
## name: Nullclines
## palette: function
## palette.cache: NULL
## position: left
## range: <ggproto object: Class RangeDiscrete, Range, gg>
##   range: NULL
##   reset: function
##   train: function
##   super: <ggproto object: Class RangeDiscrete, Range, gg>
## rescale: function
## reset: function
## scale_name: hue
## train: function
## train_df: function
## transform: function
## transform_df: function
## super: <ggproto object: Class ScaleDiscrete, Scale, gg>
```

A key thing to note is that where two different nullclines cross is an *equilibrium solution* to the system of equations. This means that **both**  $\frac{dS}{dt}$  and  $\frac{dI}{dt}$  are zero at this point. Examining Figure 7.1, there are three possibilities:

1. At  $S = 0$  and  $I = 0$  (otherwise known as the origin). This equilibrium solution makes biological sense: if there is nobody susceptible or infected (everyone is perfectly healthy - yay!) there are no flu cases.
2. Building on the last equilibrium solution, the entire horizontal axis is an equilibrium solution because the nullclines for  $S$  and  $I$  are the ( $I = 0$ ). There is a practical interpretation of this nullcline - whenever  $I = 0$ , meaning there are no infected people around, the solution is at an equilibrium.
3. There is also a third possibility where the vertical line at  $S = 1$  crosses the horizontal axis ( $S = 1, I = 0$ ), but that also falls under the second equilibrium solution.

Now that we have identified our nullclines and equilibrium solutions, we will add additional context with the *flow* of the solution.

### 7.1.1 Adding context to our phase plane: slope fields

Let's go back to the idea of a phase plane, but this time we are going to add more context to our nullcline graph by evaluating different values of  $S$  and  $I$  into our system of equations and plot the *slope field*.

First let's evaluate the derivatives  $\frac{dS}{dt}$  and  $\frac{dI}{dt}$  in Equation (7.1) for different values of  $S$  and  $I$ :

Notice how the different values of  $\frac{dS}{dt}$  and  $\frac{dI}{dt}$  at each of the  $S$  and  $I$  values. We can plot each of the coordinate pairs of  $\left(\frac{dS}{dt}, \frac{dI}{dt}\right)$  with a vector in the  $(S, I)$  plane. We associate  $\frac{dS}{dt}$  with left-right motion, so positive  $\frac{dS}{dt}$  means pointing to the right. Likewise, we associate  $\frac{dI}{dt}$  with up-down motion, so positive  $\frac{dI}{dt}$  means the vector points up. At the point  $(S, I) = (1, 1)$ , we have an arrow that points directly to the west because  $\frac{dI}{dt} < 0$  and  $\frac{dS}{dt} = 0$ . If we sequentially sample points in the  $(S, I)$  plane we get a vector field plot, superimposed with the nullclines:

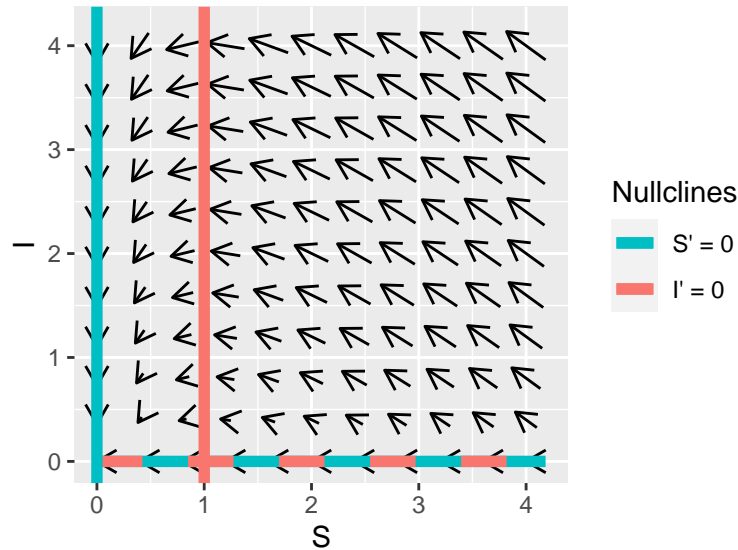


Figure 7.2: Phaseplane for Equation (7.1). To generate the plot we assumed  $\beta = 1$  and  $k = 1$

### 7.1.2 Motion around the nullclines

We can also extend the motion around the nullclines to investigate the stability. With a one dimensional differential equation we used a number line to quantify values where the solution is increasing / decreasing. The problem with several differential equations is that the notion of “increasing” or “decreasing” becomes difficult to understand - as there is an additional degree of freedom! Simply put, in a plane you can move left/right *or* up/down. The benefit for having nullclines is that they **isolate** the motion in one direction.

In general for a two dimensional system: - When a horizontal axis variable has a nullcline, the only allowed motion is up/down.  
- When a vertical axis variable has a nullcline, the only motion is left/right.

Applying this knowledge to Equation (7.1), if we choose points where  $I' = 0$  then we know that the only motion is to the left and the right because  $S$  can still change along that curve. If we choose points where  $S' = 0$  then we know that the only motion is to the up/down because  $I$  can still change along that curve (Figure 7.3).

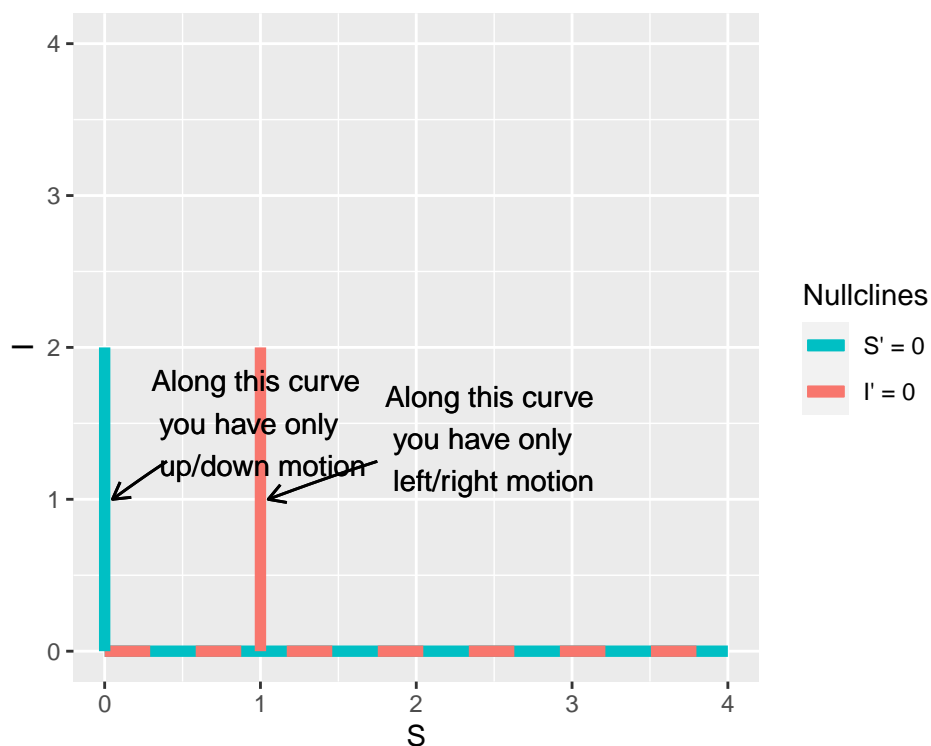
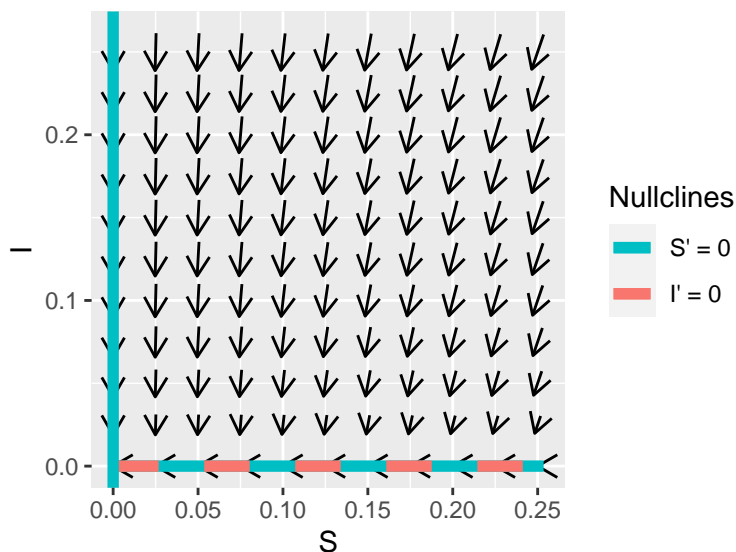


Figure 7.3: Nullclines for Equation (7.1) with context on the direction of the motion.

## 7.2 Determining stability

The picture of the phase plane with the nullcline qualitatively tells us about the stability of an equilibrium point. One of the equilibrium solutions is at the origin  $(S, I) = (0, 0)$ . As before we want to investigate if the equilibrium solution is stable or unstable.

Figure ?? zooms in the phaseplane at the equilibrium solution at  $S = 0, I = 0$ :



As you can see the arrows appear to be pointing into and towards the equilibrium solution. So we would classify this equilibrium solution as *stable*.

## 7.3 Generating a phase plane in R

Let's take what we learned from the case study of the flu model with quarantine to qualitatively analyze a system of differential equations:

- We determine nullclines by setting the derivatives equal to zero.
- Equilibrium solutions occur where nullclines for the two different equations intersect.
- The arrows in the phase plane help us characterize the stability of the equilibrium solution.

To determine the phaseplane diagram **MAT369Code** package has some basic functionality to generate a phase plane. Consider the following system of differential equations (Equation (7.2)):

$$\begin{aligned}\frac{dx}{dt} &= x - y \\ \frac{dy}{dt} &= -x + y\end{aligned}\tag{7.2}$$

In order to generate a phaseplane diagram for Equation (7.2) we need to define functions for  $x'$  and  $y'$ , which I will annotate as  $dx$  and  $dy$  respectively. We are going to collect these equations in one vector called **system\_eq**, using the tilde ( $\sim$ ) as a replacement for the equals sign:

```
system_eq <- c(dx ~ x-y,
               dy ~ x+y)
```

Then what we do is apply the command **phaseplane**, which will generate a vector field over a domain:

```
phaseplane(system_eq, 'x', 'y') # The values in quotes are the labels for the axes and to identify the variable
```

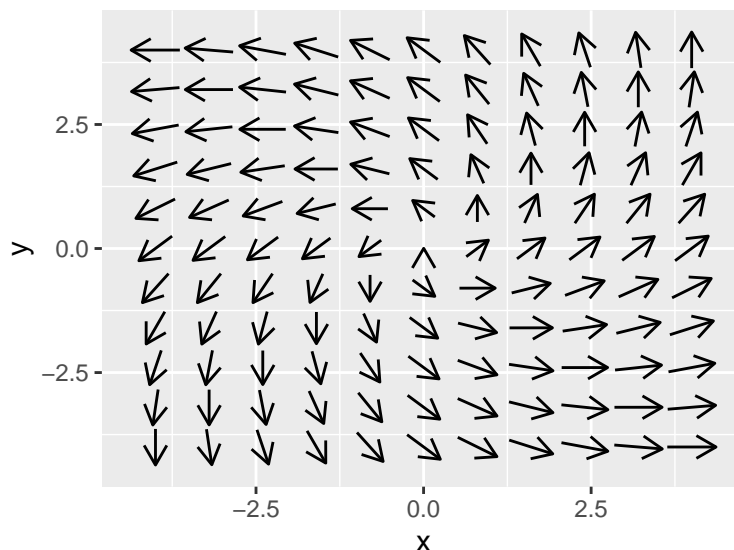


Figure 7.4: Phaseplane diagram for Equation (7.2)

The command **phaseplane** has an option called **eq\_soln** that will if there are any equilibrium solutions to be found and report them to the console. For example try running **phaseplane(system\_eq, 'x', 'y', eq\_soln=TRUE)** and see what gets output to console. While this option lists equilibrium solutions, you should confirm them with the differential equation through direct solving.

### 7.3.1 Generating a phase line in R:

From Section 6 we discussed how to construct phaselines by hand. It turns out that the command **phaseplane** can also plot phase lines. Let's take a look at an example first and then discuss how that it works.

**Example 7.1.** A colony of bacteria growing in a nutrient-rich medium deplete the nutrient as they grow. As a result, the nutrient concentration  $x(t)$  is steadily decreasing. Determine the phaseline for the following differential equation:

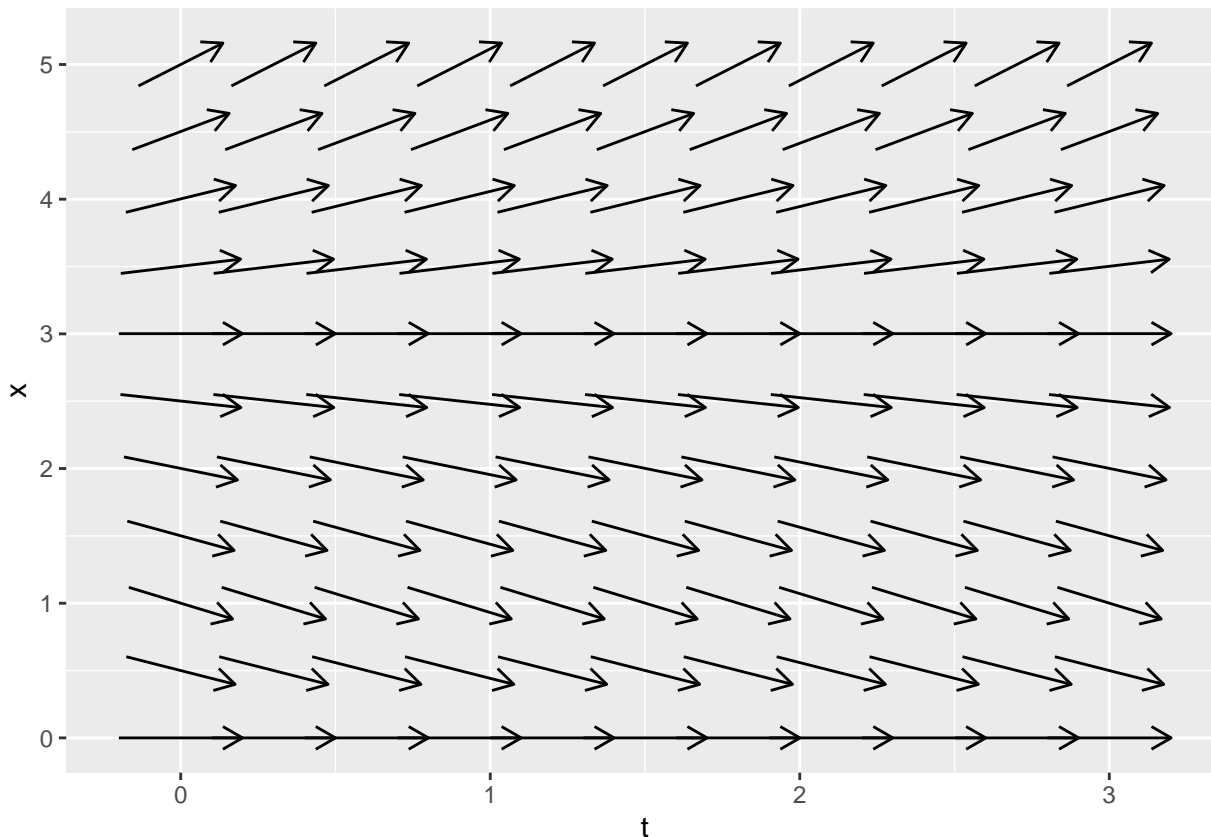
$$\frac{dx}{dt} = -0.7 \cdot \frac{x \cdot (3 - x)}{1 + x} \quad (7.3)$$

The R code to generate this phaseplane is the following:

```
# Define the windows where we make the plots
t_window <- c(0,3)
x_window <- c(0,5)

# Define the differential equation
system_eq <- c(dt ~ 1,
               dx ~ -0.7 * x*(3-x)/(1+x))

phaseplane(system_eq,"t","x",t_window,x_window)
```



Notice how we have the equation  $dt = 1$ . What we are doing is re-writing the differential equation with a new variable  $s$  (Equation (7.4)):

$$\begin{aligned} \frac{dt}{ds} &= 1 \\ \frac{dx}{ds} &= -0.7 \cdot \frac{x \cdot (3 - x)}{1 + x} \end{aligned} \quad (7.4)$$

The differential equation  $\frac{dt}{ds} = 1$  has solution  $s = t$ , so in essence is the same as Equation (7.3) (perhaps a little more complicated). However re-writing this system was a quick and handy workaround to re-use code.

## 7.4 Exercises

**Exercise 7.1.** Determine equilibrium solutions for Equation (7.2).

**Exercise 7.2.** This problem considers the following system of differential equations:

$$\begin{aligned}\frac{dx}{dt} &= y \\ \frac{dy}{dt} &= -x\end{aligned}$$

- Determine the equations of the nullclines and equilibrium point of this system of differential equations.
- Modify the function `phaseplane` to generate a phaseplane of this system.
- For each point along a nullcline, determine the resulting motion (up-down or left-right).
- Based on the work you generated, determine if the equilibrium solution is *stable* or *unstable*.
- Verify that the functions  $x(t) = \sin(t)$  and  $y = \cos(t)$  is one solution to this system of differential equations.

**Exercise 7.3.** Considers the following system of differential equations:

$$\begin{aligned}\frac{dx}{dt} &= y \\ \frac{dy}{dt} &= 3x^2 - 1\end{aligned}\tag{7.5}$$

- Determine the equations of the nullclines and equilibrium solutions for this system of differential equations.
- For each point along a nullcline, determine the resulting motion (up-down or left-right).
- Modify the function `phaseplane` to generate a phaseplane of this system.
- Make a hypothesis to classify if the equilibrium point is *stable* or *unstable*.

**Exercise 7.4.** A plant grows proportional to its current length  $L$ . Assume this proportionality constant is  $\mu$ , whose rate also decreases proportional to its current value. The system of equations that models this plant growth is the following:

$$\begin{aligned}\frac{dL}{dt} &= \mu L \\ \frac{d\mu}{dt} &= -0.1\mu\end{aligned}\tag{7.6}$$

- Explain why  $L = 0$  and  $\mu = 0$  is the only equilibrium solution to this differential equation.
- Modify the function `phaseplane` to generate a phaseplane of this system.
- Is the origin a stable equilibrium solution?

**Exercise 7.5.** Red blood cells are formed from stem cells in the bone marrow. The red blood cell density  $r$  satisfies an equation of the form

$$\frac{dr}{dt} = \frac{0.2r}{1+r^2} - 0.1r,\tag{7.7}$$

- What are the equilibrium solutions for this differential equation?
- Modify the function `phaseplane` to generate a phaseline for this differential equation for  $0 \leq t \leq 5$  and  $0 \leq r \leq 5$ .
- Based on the phaseline, are the equilibrium solutions stable or unstable?



**Exercise 7.6.** Organisms that live in a saline environment biochemically maintain the amount of salt in their blood stream. An equation that represents the level of  $S$  in the blood is the following:

$$\frac{dS}{dt} = 1 + 0.3 \cdot (3 - S)$$

- What are the equilibrium solutions for this differential equation?
- Modify the function `phaseplane` to generate a phaseline for this differential equation for  $0 \leq t \leq 10$  and  $0 \leq S \leq 10$ .
- Based on the phaseline, are the equilibrium solutions stable or unstable?

**Exercise 7.7.** The core body temperature ( $T$ ) of a mammal is coupled to the heat production (scaled by heat capacity  $Q$ ) with the following system of differential equations:

$$\begin{aligned} \frac{dT}{dt} &= Q + 0.5 \cdot (20 - T) \\ \frac{dQ}{dt} &= 0.1 \cdot (38 - T), \end{aligned} \tag{7.8}$$

- Determine the equations of the nullclines and equilibrium point of this system of differential equations.
- For each point along a nullcline, determine the resulting motion (up-down or left-right).
- Make a hypothesis to classify if the equilibrium point is *stable* or *unstable*.

**Exercise 7.8.** Consider the following system of differential equations for the lynx-hare model:

$$\frac{dH}{dt} = rH - bHL \tag{7.9}$$

$$\frac{dL}{dt} = ebHL - dL \tag{7.10}$$

- Determine the steady states of this system of differential equations.
- Determine equations for the nullclines, expressed as  $L$  as a function of  $H$ . There should be two nullclines for each rate.

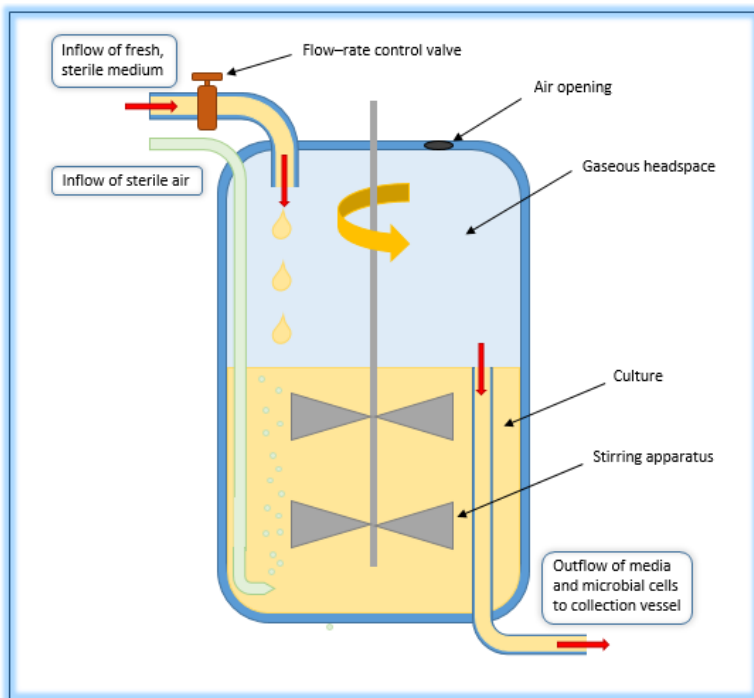


Figure 7.5: An example of a chemostat.

**Exercise 7.9.** A chemostat is a tank used to study microbes and ecology, where microbes grow under controlled conditions. Think of this like a large tank with nutrient-rich water being continuously cycled through, as shown in the following Figure 7.5 (Source: Wikipedia). Equations that describe the microbial biomass  $W$  and the nutrient concentration  $C$  (in the culture) are the following:

$$\frac{dW}{dt} = \mu W - F \frac{W}{V} \quad (7.11)$$

$$\frac{dC}{dt} = D \cdot (C_R - C) - S \mu \frac{W}{V}, \quad (7.12)$$

where we have the following parameters:  $\mu$  is the per capita reproduction rate,  $F$  is the flow rate,  $V$  is the volume of the culture solution,  $D$  is the dilution rate,  $C_R$  is the concentration of nutrients entering the culture, and  $S$  is a stoichiometric conversion of nutrients to biomass.

- Write the equations of the nullclines for this differential equation.
- Determine the equilibrium solutions for this system of differential equations.
- Generate a phaseplane for this differential equation with the values  $\mu = 1$ ,  $D = 1$ ,  $C_R = 2$  and  $S = 1$  and  $V = 1$ .
- Classify the stability of the equilibrium solutions.

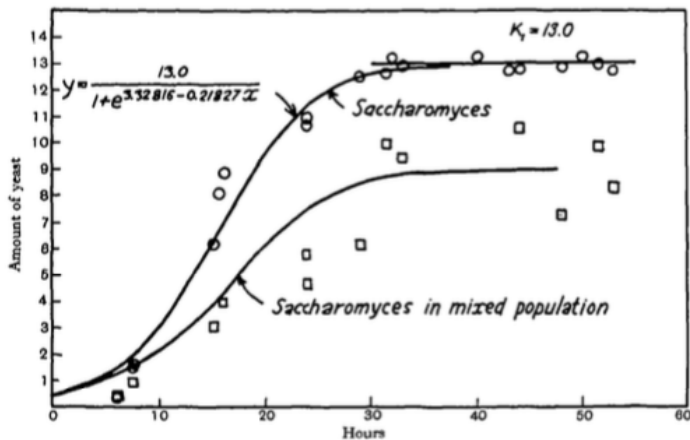


Fig. 2. The growth of the volume of *Saccharomyces cerevisiae* cultivated separately and in the mixed population according to the first and second series of experiments.

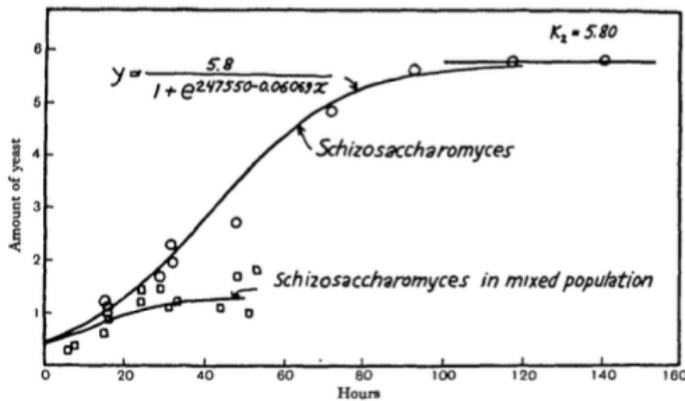


Figure 7.6: Population results from two yeast species growing in competition. From Gause (1932)

**Exercise 7.10.** A classical paper *Experimental Studies on the Struggle for Existence: I. Mixed Population of Two Species of Yeast* by Gause (1932) examined two different species of yeast growing in competition with each other. The differential equations given for two species in competition are:

$$\begin{aligned}\frac{dy_1}{dt} &= -b_1 y_1 \frac{(K_1 - (y_1 + \alpha y_2))}{K_1} \\ \frac{dy_2}{dt} &= -b_2 y_2 \frac{(K_2 - (y_2 + \beta y_1))}{K_2},\end{aligned}$$

where  $y_1$  and  $y_2$  are the two species of yeast with the parameters  $b_1$ ,  $b_2$ ,  $K_1$ ,  $K_2$ ,  $\alpha$ ,  $\beta$  describing the characteristics of the yeast species.

- Determine the equilibrium solutions for this differential equation. Express your answer in terms of the parameters  $b_1$ ,  $b_2$ ,  $K_1$ ,  $K_2$ ,  $\alpha$ ,  $\beta$ .
- Gause computed the following values of the parameters:  $b_1 = 0.21827$ ,  $b_2 = 0.06069$ ,  $K_1 = 13.0$ ,  $K_2 = 5.8$ ,  $\alpha = 3.15$ ,  $\beta = 0.439$ . Using these values, what would be the predicted values for the equilibrium solutions?
- Use the function `systems` to solve this system of differential equations numerically.
- Figure 7.6 is from Gause (1932) and shows the experimental population values “in mixed population”. Use the graph to estimate the equilibrium solutions for both species. How close (or how far from) the equilibrium solutions are Gause’s results to your computed equilibrium solutions?



## Chapter 8

# Exact Solutions to Differential Equations

We have already discussed some tools that can analyze differential equations numerically (Section 5) and qualitatively (Sections 6 and 7). The phase plane allowed us to evaluate equilibrium solutions and their stability. Beyond this graphical approach, it is also helpful to know the exact solution to an equation. In this section we will study a few techniques to find exact solutions to differential equation. We will apply some of the tools you may have learned from calculus.

### 8.1 Separable Differential Equations

One technique to solve differential equations is the method of *separation of variables*. Let's look at an example:

What is the general solution to  $\frac{dy}{dx} = yx^2$ ? To solve this expression we collect the variables involving  $x$  and one side of the equation, and the variables involving  $y$  on the other:

$$\frac{1}{y} dy = x^2 dx.$$

Now the next step is to determine the antiderivative of both sides of expression:

$$\begin{aligned}\int \frac{1}{y} dy &= \ln(y) + C. \\ \int x^2 dx &= \frac{1}{3}x^3 + C.\end{aligned}$$

Finally since both sides are equal we can solve for the dependent variable  $y$ . One thing to note: usually for antiderivatives we always include a  $+C$ . For solving differential equations it is okay just to keep only one  $+C$ , which usually is best on the side of the independent variable:

$$\ln(y) = \frac{1}{3}x^3 + C \rightarrow e^{\ln(y)} = e^{\frac{1}{3}x^3 + C} \rightarrow y = Ce^{\frac{1}{3}x^3}.$$

We are in business! So here is a general technique approach to solving a differential equation via separation of variables:

1. **Separate** the variables on one side of the equation.
2. **Integrate** both sides individually.
3. **Solve** for the dependent variable.

If we solve this equation using separation of variables we have Using your work above as a guide, solve this differential equation to determine a solution  $y(x)$ .

## 8.2 Integrating factors

One model that we have looked at is the the  $SI$  model where the spread of the disease is proportional to the number infected:

$$\frac{dI}{dt} = .03(1000 - I) = 30 - .03I$$

While this differential equation can be solved via separation of variables, let's try a different approach as an illustration of another useful technique. First let's write the terms involving  $I$  on one side of the equation:

$$\frac{dI}{dt} + .03I = 30.$$

What we are going to do is multiply both sides of this equation by  $e^{.03t}$  (I'll explain more about that later):

$$\frac{dI}{dt} \cdot e^{.03t} + .03I \cdot e^{.03t} = 30 \cdot e^{.03t}$$

Hmmm - this seems like we are making our equation harder to solve, doesn't it? However the left hand side is actually the derivative of the expression  $I \cdot e^{kt}$ ! Let's take a look:

$$\frac{d}{dt} (I \cdot e^{.03t}) = \frac{dI}{dt} \cdot e^{.03t} + I \cdot .03e^{.03t}$$

This derivative is courtesy of the product rule from calculus. Ok, so what does this do to the differential equation? Well, by re-writing the differential equation as a derivative and integrating:

$$\begin{aligned} \frac{d}{dt} (I \cdot e^{.03t}) &= 30 \cdot e^{.03t} \rightarrow \\ \int \frac{d}{dt} (I \cdot e^{.03t}) dt &= \int 30 \cdot e^{.03t} dt \rightarrow \\ I \cdot e^{.03t} &= 30 \cdot e^{.03t} + C \end{aligned} \tag{8.1}$$

Notice how by writing the left hand side in terms of the product rule and integrating we could find the solution. We added the  $+C$  to the right hand side. All that is left to do is to solve in terms of  $I(t)$  by dividing by  $e^{kt}$ . We will label this solution  $I_1(t)$ :

$$I_1(t) = 1000 + Ce^{-.03t} \tag{8.2}$$

Cool! The function  $e^{.03t}$  is called an *integrating factor*. To see what is meant by that, let's try one more example.

In this case, we are going to assume that the rate of infection is time dependent, or  $k(t) = .03t$ . How this would work in practice is that initially (at  $t = 0$ ) there is no infection, but the infection rate increases as time goes on.

Our differential equation in this case is:

$$\frac{dI}{dt} + .03t \cdot I = 30t.$$

So if we want to write the left hand side as a product, what we will do is multiply the *entire* differential equation by  $e^{\int .03t dt} = e^{0.015t^2}$ . This term is called the *integrating factor*:

$$\frac{dI}{dt} \cdot e^{0.015t^2} + .03t \cdot I \cdot e^{0.015t^2} = 30t \cdot e^{0.015t^2}$$

First we rewrite the left hand side using the product rule:

$$\frac{dI}{dt} \cdot e^{0.015t^2} + .03t \cdot I \cdot e^{0.015t^2} = \frac{d}{dt} (I \cdot e^{0.015t^2}).$$

Now we can integrate this equation by:

$$\begin{aligned}\frac{d}{dt} \left( I \cdot e^{0.015t^2} \right) &= 30t \cdot e^{0.015t^2} \rightarrow \\ \int \frac{d}{dt} \left( I \cdot e^{0.015t^2} \right) dt &= \int 30t \cdot e^{0.015t^2} dt \rightarrow \\ I \cdot e^{0.015t^2} &= N \cdot e^{0.015t^2} + C\end{aligned}\tag{8.3}$$

All right! So the last step is to write the equation in terms of  $I(t)$ , which we will label  $I_2(t)$ :

$$I_2(t) = 1000 + Ce^{-0.015t^2}\tag{8.4}$$

Figure 8.1 compares solutions when  $I(0) = 10$ .

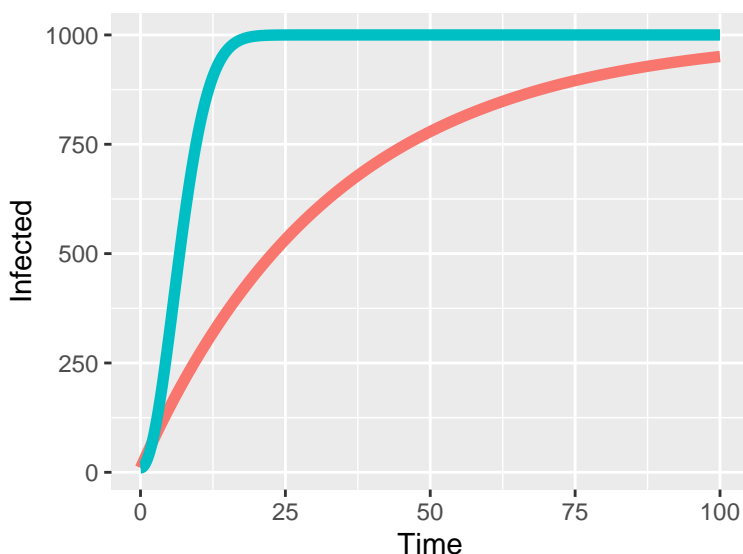


Figure 8.1: Comparison of two integrating factor solutions, Equation (8.2) in red and Equation (8.4) in blue.

Ok, let's summarize this integrating factor approach for differential equations that can be written in the form

$$\frac{dy}{dt} + f(t) \cdot y = g(t)$$

1. Calculate the *integrating factor*  $e^{\int f(t) dt}$ . Hopefully the integral  $\int f(t) dt$  is easy to compute!
2. Next multiply the integrating factor across your equation to rewrite the differential equation as  $\frac{d}{dt} \left( y \cdot e^{\int f(t) dt} \right) = g(t) \cdot e^{\int f(t) dt}$ .
3. Then compute the integral  $H(t) = \int g(t) \cdot e^{\int f(t) dt} dt$ . This looks intimidating - but hopefully is manageable to compute! Don't forget the  $+C$ !
4. Then solve for  $y(t)$ :  $y(t) = H(t) \cdot e^{-\int f(t) dt} + Ce^{-\int f(t) dt}$ .

This technique is a handy way to work with equations that aren't easily separable.

## 8.3 Guess and Check

A final approach is called the guess and check method. Say for example we have the following equation that describes the rate of change above: The first approach if a function is a solution to a differential equation is the guess and check method, or by direct substitution.

$$\frac{dS}{dt} = 0.7S$$

We know that we can apply separation of variables, but instead let's try to see if the function  $\tilde{S}(t) = 5e^{0.7t}$  is a solution in order to do that, we need to differentiate  $\tilde{S}(t)$ , which using our knowledge of calculus is  $0.7 \cdot 5e^{0.7t}$ . If we note that  $\frac{d\tilde{S}}{dt} = 0.7\tilde{S} = 0.7e^{0.7t}$  then the function  $\tilde{S}$  does solve the differential equation.

Super cool! Let's try an example:

**Example 8.1.** Verify the following functions are solutions to the differential equation  $\frac{dS}{dt} = 0.7S$ :

- $\tilde{R}(t) = 10e^{0.7t}$
- $\tilde{P}(t) = e^{0.7t}$
- $\tilde{Q}(t) = 5e^{0.7t}$
- $\tilde{F}(t) = 3$
- $\tilde{G}(t) = 0$

*Remark.* Let's apply direct differentiation to each of these functions:

- $\tilde{R}(t) = 10e^{0.7t} \rightarrow \tilde{R}'(t) = 7e^{0.7t}$
- $\tilde{P}(t) = e^{0.7t} \rightarrow \tilde{P}'(t) = e^{0.7t}$
- $\tilde{Q}(t) = 5e^{0.7t} \rightarrow \tilde{Q}'(t) = 3.5e^{0.7t}$
- $\tilde{F}(t) = 3 \rightarrow \tilde{F}'(t) = 0$
- $\tilde{G}(t) = 0 \rightarrow \tilde{G}'(t) = 0$

Now we will compare each of these solutions to the right hand side:

- $0.7\tilde{R}(t) = 0.7 \cdot 10e^{0.7t} \rightarrow 7e^{0.7t}$
- $0.7\tilde{P}(t) = 0.7e^{0.7t}$
- $0.7\tilde{Q}(t) = 0.7 \cdot 5e^{0.7t} \rightarrow 3.5e^{0.7t}$
- $0.7\tilde{F}(t) = 0.7 \cdot 3 \rightarrow 2.1$
- $0.7\tilde{G}(t) = 0.7 \cdot 0 \rightarrow 0$

Notice how the right hand sides of each equation equals the left hand sides. When that is the case, our candidate functions are indeed solutions to the differential equation!

## 8.4 Superposition of solutions

Related to the Guess and Check method is this concept called superposition of solutions. Here how this works: if you have two known solutions to a differential equation, then the sum (or difference) is a solution as well. Let's look at an example:

**Example 8.2.** Show that  $\tilde{S}(t) = 5e^{0.7t} + e^{0.7t}$  is a solution to the differential equation  $\frac{dS}{dt} = 0.7S$

*Remark.* By direct differentiation,  $\tilde{S}'(t) = 3.5e^{0.7t} + 0.7e^{0.7t}$ . Also we have that  $0.7 \cdot \tilde{S}(t) = 0.7 \cdot (5e^{0.7t} + e^{0.7t}) = 3.5e^{0.7t} + 0.7e^{0.7t}$ , which equals  $\tilde{S}'$ .

What this example illustrates is the principle that if you have two solutions to a differential equation, they can be added together and produce a new solution. This is an example of a *linear combinations* of solutions, and we can state this more formally:

If  $x(t)$  and  $y(t)$  are solutions to the differential equation  $z' = f(t, z)$ , then  $c(t) = a \cdot x(t) + b \cdot y(t)$  are also solutions, where  $a$  and  $b$  are constants.

Hopefully you were able to verify that  $\tilde{R}$  and  $\tilde{Q}$  and  $\tilde{G}$  all were solutions to the differential equation, and that  $\tilde{R} + \tilde{Q}$  was a solution as well. The most general solution to this differential equation is  $S(t) = Ce^{0.7t}$ , where the initial condition would determine the value of  $C$ .

## 8.5 Applying guess and check more broadly

As noted earlier, the guess and check method may seem to be trivial - if you have a differential equation, and solution, why verify it? Well, this method helps to introduce a useful solution technique to a differential equation, and one that we can



build up through direct verification.

We are going to revisit the lynx hare model, but simplified a little bit. Here we are going to assume that lynx and hares both decline at a rate proportional to the population size, but the lynx population increases according to the rate of hare decline:

$$\frac{dH}{dt} = -bH \quad (8.5)$$

$$\frac{dL}{dt} = bH - dL \quad (8.6)$$

Based on these simplified assumptions a good approach is to assume a solution that is exponential for both  $H$  and  $L$ :

$$\tilde{H}(t) = C_1 e^{\lambda t} \quad (8.7)$$

$$\tilde{L}(t) = C_2 e^{\lambda t} \quad (8.8)$$

The form of this solution has three unknowns:  $C_1$ ,  $C_2$ , and  $\lambda$ . If you have had Linear Algebra, you may recognize that we are assuming the solution is a vector of the form  $\vec{v} = \vec{C}e^{\lambda t}$ . Let's apply Guess and Check to solve these equations. By differentiation, we have the following:

$$\frac{d\tilde{H}}{dt} = \lambda C_1 e^{\lambda t} \quad (8.9)$$

$$\frac{d\tilde{L}}{dt} = \lambda C_2 e^{\lambda t}. \quad (8.10)$$

Comparing to our differential equation we can show that

$$\lambda C_1 e^{\lambda t} = -bC_1 e^{\lambda t} \rightarrow (\lambda + b)C_1 e^{\lambda t} = 0 \quad (8.11)$$

$$\lambda C_2 e^{\lambda t} = bC_1 e^{\lambda t} - dC_2 e^{\lambda t} \quad (8.12)$$

Let's rearrange this expression a little bit:

$$(\lambda + b)C_1 e^{\lambda t} = 0 \quad (8.13)$$

$$(\lambda + d)C_2 e^{\lambda t} = bC_1 e^{\lambda t} \quad (8.14)$$

Notice that for the second equation we can solve for  $C_1 e^{\lambda t}$ , or  $C_1 e^{\lambda t} = \frac{(\lambda + d)}{b} C_2 e^{\lambda t}$ .

This allows for something neat to happen. We can substitute this expression for  $C_1 e^{\lambda t}$  into the first equation:

$$(\lambda + b) \frac{(\lambda + d)}{b} C_2 e^{\lambda t} = 0 \quad (8.15)$$

If we assume that  $b \neq 0$ , then we have the following simplified expression:

$$(\lambda + b)(\lambda + d)C_2 e^{\lambda t} = 0 \quad (8.16)$$

Because the exponential function never equals zero, with this new equation, the only possibility is that  $(\lambda + b)(\lambda + d) = 0$ , or that  $\lambda = -b$  or  $\lambda = -d$ . Remember: if expressions multiply to zero, then the only possibility is that at least one of them is zero. This process finds the *eigenvalues* and *eigenvectors* of a system of equations. We will study this later in the course.

We now need to determine values of  $C_1$  and  $C_2$ . We can do this by going back to the equation  $(\lambda + d)C_2 e^{\lambda t} = bC_1 e^{\lambda t}$ , or  $(\lambda + d)C_2 e^{\lambda t} - bC_1 e^{\lambda t} = 0$  rearranged.

Let's analyze this equation for each of the values of  $\lambda$ :

### 8.5.1 Case $\lambda = -d$

For this situation, we have

$$(-d + d)C_2e^{-dt} - bC_1e^{-dt} = 0 \rightarrow -bC_1e^{-dt} = 0.$$

The only way for this equation to be consistent and remain zero is if  $C_1 = 0$ . We don't have any restrictions on  $C_2$ , so the general solution will be

$$\tilde{H}(t) = 0 \tag{8.17}$$

$$\tilde{L}(t) = C_2e^{-dt}. \tag{8.18}$$

### 8.5.2 Case $\lambda = -d$

For this situation, we have  $(-d + b)C_2e^{-dt} - dC_1e^{-dt} = 0$  which leads to the following equation:

$$((-d + b)C_2 - dC_1)e^{-dt} = 0 \tag{8.19}$$

The only way for this equation to be consistent and remain zero is if  $((-d + b)C_2 - dC_1) = 0$ , or if  $C_2 = \left(\frac{d}{-d + b}\right)C_1$ . In this case, the general solution will be

$$\tilde{H}(t) = C_1e^{-dt} \tag{8.20}$$

$$\tilde{L}(t) = \left(\frac{d}{-d + b}\right)C_1e^{-dt}, \tag{8.21}$$

The parameter  $C_2$  can be determined by the initial condition. Notice that we need to have  $d \neq b$  or our solution will be undefined.

Now we can write down a general solution to the system by combining our two solutions together. Here we can use the fact that two solutions can be added together (superposition) to generate a solution.

$$H(t) = C_1e^{-dt} \tag{8.22}$$

$$L(t) = \left(\frac{d}{-d + b}\right)C_1e^{-dt} + C_2e^{-bt} \tag{8.23}$$

This method only works on *linear* differential equations (i.e. it wouldn't work if there was a term such as  $kHL$  in our dynamics. Later on in the course we will look at a more systematic method (i.e. eigenvalues) to determine solutions to linear systems of equations.

## 8.6 Exercises

**Exercise 8.1.** Determine the value of  $C$  when  $I(0) = 10$  for the two equations:

$$\begin{aligned} I_1(t) &= 1000 + Ce^{-.03t} \\ I_2(t) &= 1000 + Ce^{-0.015t^2} \end{aligned} \tag{8.24}$$

**Exercise 8.2.** Verify that  $I_2(t) = N + Ce^{-0.5kt^2}$  is the solution to the differential equation  $\frac{dI}{dt} = kt(N - I)$ . Plot your solution for various values of  $k$  ranging from .001 to .1. What effect does  $k$  have on the solution?

**Exercise 8.3.** A chemical reaction  $2A \rightarrow C + D$  can be modeled with the following differential equation (scholz\_first-order\_2014?):

$$\frac{dA}{dt} = -2kA^2 \tag{8.25}$$

Apply the method of separation of variables to determine a general solution for this differential equation.

**Exercise 8.4.** Which of the following differential equations be solved via separation of variables?

1.  $\frac{dy}{dx} = x \cdot (y^2 + 2)$
2.  $\frac{dy}{dx} = x^2 + xy$
3.  $\frac{dy}{dx} = e^{x+y}$
4.  $\frac{dy}{dx} = y \cdot \cos(2 + x)$
5.  $\frac{dy}{dx} = \ln x + \ln y$

Once you have identified which ones can be solved via separation of variables, apply that technique to solve each differential equation.

**Exercise 8.5.** Solve the following differential equations by separation of variables:

1.  $\frac{dy}{dx} = \frac{y^3}{x}$
2.  $\frac{dy}{dx} = 1 + y^2$
3.  $\frac{dy}{dx} = 8 - y$

**Exercise 8.6.** Consider the following differential equation  $\frac{dP}{dt} = -\delta P$ ,  $P(0) = P_0$ , where  $\delta$  is a constant parameter.

- a. Solve this equation using the method of separation of variables.
- b. Solve this equation using an integrating factor.
- c. Your two solutions from the two methods should be the same - are they?

**Exercise 8.7.** Here we return to the problem of how animals consume food. A differential equation that relates a consumer's nutrient content (denoted as  $y$ ) to the nutrient content of food (denoted as  $x$ ) is given by:

$$\frac{dy}{dx} = \frac{1}{\theta} \frac{y}{x}, \quad (8.26)$$

where  $\theta \geq 1$  is a constant. Apply separation of variables to determine the general solution to this differential equation.

**Exercise 8.8.** Apply separation of variables to determine general solutions to the following systems of differential equations:

$$\begin{aligned} \frac{dx}{dt} &= x \\ \frac{dy}{dt} &= y \end{aligned} \quad (8.27)$$

(This system is an example of an *uncoupled* system of equations.)

**Exercise 8.9.** A plant grows proportional to its current length  $L$ . Assume this proportionality constant is  $\mu$ , whose rate also decreases proportional to its current value. The system of equations that models this plant growth is the following:

$$\frac{dL}{dt} = \mu L \quad (8.28)$$

$$\frac{d\mu}{dt} = -k\mu \quad (8.29)$$

$$(k \text{ is a constant parameter}) \quad (8.30)$$

Apply separation of variables to determine the general solutions to this system of equations.

**Exercise 8.10.** Use the method developed in this section determine the general solution to the following system of differential equations:

$$\frac{dx}{dt} = x - y \quad (8.31)$$

$$\frac{dy}{dt} = 2y \quad (8.32)$$

**Exercise 8.11.** Apply the method of integrating factors to determine the solution to the differential equation  $\frac{dI}{dt} = (N - I) = kN - kI$ , where  $k$  and  $N$  are parameters.

**Exercise 8.12.** For each of the following differential equations:

- Determine equilibrium solutions for the differential equation.
- Apply separation of variables to determine general solutions to the following differential equations:
- Choose reasonable values of any parameters and use the function `plotFunction` to generate a solution curve for an initial condition that you select.

1.  $\frac{dy}{dx} = -\frac{x}{y}$

2.  $\frac{dy}{dx} = 8 - y$

3.  $\frac{dW}{dt} = k(N - W)$  ( $k$  and  $N$  are constant parameters)

4.  $\frac{dR}{dt} = -aR \ln \frac{R}{K}$  ( $a$  and  $K$  are constant parameters)

**Exercise 8.13.** Consider the following differential equation, where  $M$  represents a population of mayflies and  $t$  is time (given in months), and  $\delta$  is a mortality rate (units % mayflies / month):

$$\frac{dM}{dt} = -\delta M \quad (8.33)$$

Determine the general solution to this differential equation and plot a few different solution curves with different values of  $\delta$ . Assume that  $M(0) = 10,000$ . Also identify the equilibrium solution to the differential equation and classify the stability of the equilibrium solution based on your solution curves.

**Exercise 8.14.** An alternative model of mayfly mortality is the following:

$$\frac{dM}{dt} = -\delta(t)M, \quad (8.34)$$

where  $\delta(t)$  is a time dependent mortality function. Determine a solution and plot a solution curve (assuming  $M(0) = 10,000$  and over the interval from  $0 \leq t \leq 1$ ) for this differential equation when  $\delta(t)$  has the following forms:

1.  $\delta(t) = t^2$
2.  $\delta(t) = 1 - t^2$

Provide a reasonable biological explanation justifying the use of this alternative mayfly model.