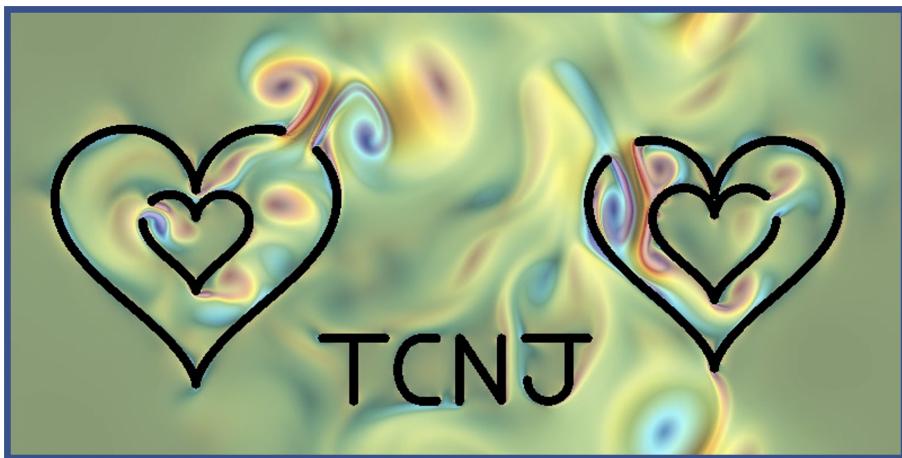


MAT/BIO 330:
Mathematical Biology¹
*(a mathematical supplement)*²



v1.8

NICHOLAS A. BATTISTA³

May 10, 2021

¹Woo course numbers and official name of the course!

²So we can all be on the same page regarding the mathematics throughout this course.

³Where I am: Dept. of Math and Stats, TCNJ, battistn@tcnj.edu, battistn.pages.tcnj.edu ... actually most likely my kitchen.

Contents

1 Discrete Dynamical Systems	3
1.1 Introduction to Discrete Dynamical Systems	3
1.2 Linear Discrete Dynamical Equations	10
1.2.1 Linear, autonomous, homogeneous equations	11
1.2.2 Linear, non-homogeneous equations	18
1.3 First-order Non-linear Discrete Dynamical Systems	23
1.3.1 The Discrete Logistic Equation	24
1.3.2 Equilibria and Stability in First-order Discrete Equations	28
1.3.3 Existence and Stability of Periodic Solutions to 1D Discrete Equations	41
1.4 Systems of Discrete Equations	51
1.4.1 Equilibria and stability in discrete systems	54
1.4.2 Further classification of equilibrium points	69
1.4.3 Linear systems of coupled discrete equations	74
1.4.4 Mathematics for stability analysis of coupled discrete equations	90
2 Continuous Dynamical Systems	107
2.1 Introduction and transition to Continuous Dynamical Systems	107
2.2 Mathematically massaging discrete models into continuous models	110
2.3 Refresher - what do derivatives tell us?	113
2.4 Equilibria & Stability in Systems of Differential Equation . .	114
2.5 Nullclines and Phase Planes	134
2.6 Holling-Type Interactions Between 2 Species	145
3 Application Hodge-Podge!	157
3.1 SIR Models and Disease Transmission Models	157
3.1.1 Base SIR Model: no deaths	161
3.1.2 Base SIR Model: including deaths	167
3.1.3 The Basic Reproduction Number, Vaccines, and Herd Immunity	172

3.2	Random Walks, Brownian Motion, and Diffusive Processes . . .	178
3.2.1	Random Walks...you're tellin' me we can just flip a coin?	179
3.2.2	Building towards Diffusion from Random Walks!	188
3.2.3	Comparing Random Walkers to Diffusion	196
3.3	Dynamics of a Neuron	198
3.3.1	Analyzing the FitzHugh-Nagumo Eqns (no applied current, nor diffusion)	207
3.3.2	Analyzing the FitzHugh-Nagumo Eqns (w/ applied current, no diffusion)	212

Preface

This book is intended to be a supplemental resource for the mathematics we will be going through in course. As this course is on an incredibly highly interdisciplinary subject, *mathematical biology*, and taken by students with highly diverse backgrounds, in terms of majors, years, and experience, I hope this helps serve as a succinct resource to provide context regarding where some of the mathematical ideas that we will use are coming from.

At this point in development, I am not creating this book to actually dive into many actual biological applications, but rather just to provide a brief overview and context for the mathematics we will be mastering together. In that vein, this is not written as a standard mathematics textbook - with the usual whimsical definition, lemma, theorem, and proof fairy tales. There might be some of that, but the focus is more on using mathematical concepts to solve time-dependent problems, whether they come in the flavors of discrete equations, differential equations, or partial differential equations. I will only provide theorems and proofs, where I believe is appropriate to backup some key ideas and concepts to which our analysis will heavily depend, e.g., linear stability theorems for equilibrium... ugh jargon. Don't worry, I'm also assuming you don't know any of that yet... the course hasn't started!

Disclaimer: I am 99.5165% sure that there are typos in this book. If you come across something funky (or that just seems off), please let me know. I am putting all of this together for you... if it doesn't help you, I'd like to know how I might be able to improve it. Thanks!

Structure of book

Each section will be broken into the main mathematical tools we will use as we go forward in the course. In particular, I am hoping that by the middle of the semester that this book will contain mathematical supplements for the entire course.

As you will see shortly, we will start the semester discussing what are

called Discrete Dynamical Systems. We will explore these as a tool for studying the interaction between different biological phenomena, rather its species interactions in ecology or cellular processes. Hence right now (and since I started putting this together around August 15, 2020) all that exists is a unit on Discrete Systems.

Thanks for bearing with me.

Kind regards,

Nicholas A. Battista
 Assistant Professor of Mathematics and Statistics
 Dept. of Mathematics and Statistics
 The College of New Jersey
battistn@tcnj.edu
<http://battistn.pages.tcnj.edu/>
<http://github.com/nickabattista/>

Acknowledgements

- I'm deeply indebted my my mother, sister, and grandparents for their support and encouragement as well as the too-numerous-to-count mathematicians and scientists that have helped me along the way. I wish to thank my Ph.D. advisor Dr. Laura Miller for her infinite patience, kindness, and help during my graduate studies and into my career at TCNJ.
- I am also thankful for the help of the following mathematicians and biologists who kindly let me know that something seemed fishy in the text (typos):
 1. Gabrielle Glavin
 2. Abhi Bhattaru
 3. Alexa Morreale
 4. Lillian Hoffman

Nicholas A. Battista
 Assistant Professor of Mathematics and Statistics
 Dept. of Mathematics and Statistics
 The College of New Jersey
battistn@tcnj.edu
<http://battistn.pages.tcnj.edu/>
<http://github.com/nickabattista/>

1

Discrete Dynamical Systems

“See, here I’m now by myself, uh, er, talking to myself. That’s, that’s chaos theory.”

– Dr. Ian Malcolm, *Jurassic Park*

1.1 Introduction to Discrete Dynamical Systems

Our first adventure takes us into the world of *iterative* maps. These will feel very similar to *sequences* from Calculus B. That is, we define iterative maps in terms of an indexing variable, say n . For example, consider the following iterative map:

$$X_{n+1} = 2X_n.$$

If we are given an *initial value* for the iterations to stem from, we can compute the other values in the iterative set. For the above example, given $X_0 = 13$, we can find the values for $X_1, X_2, \dots, \text{etc..}$

$$\begin{aligned} X_0 &= 13 \text{ (given in problem)} \\ X_1 &= 2X_0 = 2(13) = 26 \\ X_2 &= 2X_1 = 2(26) = 52 \\ X_3 &= 2X_2 = 2(52) = 104 \\ &\vdots \end{aligned}$$

So given an iterative map such as $X_{n+1} = 2X_n$ **AND** an initial value X_0 , we can find subsequent values produced by the iterative map. However, what would be very dandy is if rather than produce a sequence of successive values like $X_1, X_2, X_3, \dots, X_{117}, X_{118}, \dots$, if we could find a solution given a particular n value of interest. That is, what if we could find the solution $X(n) = X_n$ for any n without having to generate a sequence up to that point? I mean... I wouldn’t want to find X_{5254} by computing all the values

up to that point. We will dive into finding precise solutions later. First, let's kick things up a notch.

Consider another example of an iterative map:

$$Y_{n+1} = 2Y_n - 3Y_{n-1}.$$

This iterative map is slightly different than the first one we thought about - there is another term in the map! The way in which we could interpret this one is that the next value in the sequence (Y_{n+1}) depends on the previous value (Y_n) *and* the preceding value that to that (Y_{n-1}). Therefore, in order to start iterating away, we would need two pieces of initial data! As an example, let's choose $Y_0 = -1$ and $Y_1 = 3$, and find the first few iterations:

$$\begin{aligned} Y_0 &= \textcolor{blue}{-1} \quad (\textit{given in problem}) \\ Y_1 &= \textcolor{green}{3} \quad (\textit{given in problem}) \\ Y_2 &= 2Y_1 - 3Y_0 = 2(\textcolor{green}{3}) - 3(\textcolor{blue}{-1}) = \textcolor{pink}{9} \\ Y_3 &= 2Y_2 - 3Y_1 = 2(\textcolor{pink}{9}) - 3(\textcolor{green}{3}) = \textcolor{red}{9} \\ Y_4 &= 2Y_3 - 3Y_2 = 2(\textcolor{red}{9}) - 3(\textcolor{pink}{9}) = -9 \end{aligned}$$

The number of initial conditions that we would require to start iterating is how we define the order of iterative map, i.e.,

$$\begin{aligned} X_{n+1} = 2X_n &\Rightarrow \text{First-order equation} && \Leftrightarrow \text{one initial condition} \\ Y_{n+1} = 2Y_n - 3Y_{n-1} &\Rightarrow \text{Second-order equation} && \Leftrightarrow \text{two initial conditions} \end{aligned}$$

Let's now officially define a k^{th} -order iterative map:

Definition 1.1

A general k^{th} -order iterative map has the following general form:

$$x_{n+1} = f(x_n, x_{n-1}, x_{n-2}, \dots, x_{n-(k-1)}),$$

where f is a function that takes k inputs. In order to successfully iterate this map, one also needs k initial conditions, i.e., $\{x_0, x_1, x_2, \dots, x_{n-(k-1)}\}$.

To get used to the notation, let's quickly check out the following example. What I hope you take away from this example is becoming comfortable with the notation, in particular, the subscript notation.

Example 1.1.1.

- If we have a 4th-order iterative map, then we know the next value in our iteration, say x_{n+1} depends on the previous 4 values, $x_n, x_{n-1}, x_{n-2}, x_{n-3}$.
- Hence we could generally write this iterative map as

$$x_{n+1} = g(x_n, x_{n-1}, x_{n-2}, x_{n-3}),$$

where g is a known function that takes 4 input values.

- Since in order to compute x_{n+1} requires the 4 previous iteration values, to start iterating we would need 4 initial conditions, x_0, x_1, x_2 , and x_3 so that we could then find x_4 and beyond.
 - To be clear in this example, we did not explicitly define the function g , but merely wanted to point out the notation.
-

At this junction, if we had an iterative map such as

$$z_{n+1} = \frac{1 + 5 \sin(z_n) - e^{z_{n-1}}}{5 - z_{n-1} + 2z_n},$$

and two initial conditions $z_0 = 0.5$ and $z_1 = \pi$, we could (very painstakingly) find more iterations and more iterations of the map. That is, we could fine z_2 using z_1 and z_0 , and then z_3 using z_2 and z_1 . However, in a little bit we going to turn our attention to finding either *exact* solutions to such equations or develop tools to help us analyze equations further so we can understand properties about possible solutions. Plus, computers can help us find any number of iterations we desire to equations that like above. ☺

In our immediate future, the next step we will take along our classification journey will the one that guides us towards whether we will try to find exact solutions to problems or merely analyze properties of their solutions. This will take us a few sections to unpack. To point out where we will be going, we will use 2nd-order iterative maps to us point out the major differences.

Example 1.1.2. Consider the following two iterative maps:

$$X_{n+1} = \frac{2}{3}X_n + 7X_{n-1} \quad \text{and} \quad Y_{n+1} = 5Y_{n-1} \sin(Y_n) + Y_n^2.$$

- ***Similarities:***

- They are both 2nd-order equations because the next iterate depends on the two previous iteration values.
- Hence they both also require two initial values to start the iteration process.

- ***Differences:***

- The equation on the left is called a **linear** equation because X_{n+1} depends linearly on X_n and X_{n-1} . It is called linear because the right hand side of the equation only involves a linear combination of X_n and X_{n-1} . That is, we do not have any X_n^2 terms, products of $X_n X_{n-1}$, etc.. The right hand side only has a term that involves (“some coefficient”) · X_n + (“some coefficient”) · X_{n-1} .
 - The equation on the right is called **non-linear** because the right hand side depends non-linearly on either one (or both) of Y_n or Y_{n-1} . For example, Y_n^2 is nonlinear because it does not fit the mold of a coefficient Y_n and $Y_{n-1} \sin(Y_n)$ is nonlinear because it involves the product of Y_n and Y_{n-1} stuff.
-

This whole idea of linear vs. non-linear equations always confused me at first, so let's do a few more examples of just classifying equations as linear or non-linear to make sure we are all on the same page. Classification gets tricky and as you will see in the example we can also classify equations further beyond simply their order and whether they are linear or non-linear.

Example 1.1.3. *Classifying iterative equations as linear or non-linear*

1. $P_{n+1} = g(P_n, P_{n-1}) = 7P_n - 2P_{n-1}$

- Basic classification: linear, 2nd-order
- Why: P_{n+1} depends on P_n and P_{n-1} so it's 2nd-order. It is linear because $g(P_n, P_{n-1})$ only involves a linear combination of P_n and P_{n-1} .

2. $w_{n+1} = f(w_{n-2}, w_{n-5}) = \frac{3}{7}w_{n-2} + \frac{1}{2}w_{n-5}$

- *Basic classification: linear, 6th-order*
- *Why: While w_{n+1} depends on only w_{n-2} and w_{n-5} , you need to go back 6 iterative values to get w_{n-5} , so it is 6th-order. Note that you would still need 6 initial values to begin iterating this equation. However, it is also linear because $f(w_{n-2}, w_{n-5})$ only involves a linear combination of w_{n-2} and w_{n-5} .*

$$3. z_{n+1} = H(z_n, z_{n-1}) = \sin(n)z_n + 6n^2z_{n-1}$$

- *Basic classification: linear, 2nd-order*
- *Why: z_{n+1} depends on z_n and z_{n-1} so it's 2nd-order and still would only require two initial values for one to begin iterating. It is linear because $H(z_n, z_{n-1})$ still only involves a linear combination of dependent values z_n and z_{n-1} ; in this case they just have coefficients that are a function of the independent variable, n . Whenever coefficient are functions of the independent variable, n , we call them non-autonomous equations. When the coefficients are simply static numbers, we call them autonomous.*

$$4. R_{n+1} = d(R_n, R_{n-1}) = 7e^nR_n - 2nR_{n-1} + 7\sin(n)$$

- *Basic classification: linear, 2nd-order*
- *Why: R_{n+1} depends on R_n and R_{n-1} so it's 2nd-order. It is linear because the dependence of R_n and R_{n-1} in $d(R_n, R_{n-1})$ still only involves a linear combination those two dependent variables. This equation just contains an added factor of $7\sin(n)$ being added every iteration. Even though this factor is not constant (it depends on the iteration index, n), it doesn't involve the dependent variables R_n or R_{n-1} . Therefore it is still linear. However, since there is now a term in the equation that doesn't have a R_{n+1} , R_n , or R_{n-1} attached to it, we get to call it non-homogeneous. It is also non-autonomous because the coefficients are functions of the independent indexing variable, n .*

$$5. Q_{n+1} = c(Q_n, Q_{n-1}) = 7Q_n(1 - 2Q_{n-1})$$

- *Basic classification: non-linear, 2nd-order*
- *Why: Q_{n+1} depends on Q_n and Q_{n-1} so it's 2nd-order. It is non-linear because $c(Q_n, Q_{n-1})$ contains a non-linear term of $-2Q_nQ_{n-1}$.*

$$6. b_{n+1} = A(b_n, b_{n-1}) = \frac{1}{b_n + b_{n-1}}$$

- *Basic classification: non-linear, 2nd-order*
- *Why: b_{n+1} depends on b_n and b_{n-1} so it's 2nd-order. It is non-linear because the fractional term makes it have non-linear dependence on b_n and b_{n-1} . That is, the right hand side is NOT just a linearly combination of the two dependent variables b_n and b_{n-1} . Another way to look at this is to cross multiply and see that you get the following*

$$b_{n+1}(b_n + b_{n-1}) = 1,$$

which will involve products of b_{n+1} and both b_n and b_{n-1} .

In a nutshell, linear equations lend themselves to being solved by paper and pen methods MUCH easier than non-linear equations. In fact, many (most?) non-linear equations have no precise strategy for solving them exactly. Sometimes you might get lucky and find/develop a clever method (or series of tricks, I mean, substitutions and other mathematical massaging techniques) to figure out how to solve a non-linear equation. However, we can rely on other analytical techniques to help us understand the properties of solutions of non-linear equations.

Before we move on, let's quickly define homogeneous vs. non-homogeneous equations, as well as autonomous vs. non-autonomous equations. Then we'll finally start solving some *linear* equations exactly.

Definition 1.2

- For any equation of the form $X_{n+1} = F(X_n, X_{n-1}, \dots, X_{n-k})$, if one moves all of the terms involving $X_{n+1}, X_n, X_{n-1}, \dots, X_{n-k}$ to the same side of the equation and the opposite side is simply 0, that equation is called **homogeneous**.

For example: $X_{n+1} = 2X_n^2 - 7X_n \Rightarrow X_{n+1} - 2X_n^2 + 7X_n = 0$
(homogeneous)

- If after one moves all the terms to same side and the opposite side is **not** identically 0, we call that **non-homogeneous**.

For example: $X_{n+1} = 2X_n^2 - 7X_n - 22 \Rightarrow X_{n+1} - 2X_n^2 + 7X_n = -22$
(non-homogeneous)

Note that if an equation is *homogeneous*, e.g.,

$$y_{n+1} = 2y_n + n^2 y_{n-1} \Rightarrow y_{n+1} - 2y_n - n^2 y_{n-1} = 0,$$

one could also view the problem more abstractly as the following:

$$F(y_{n+1}, y_n, y_{n-1}) = y_{n+1} - 2y_n - n^2 y_{n-1} = 0$$

and hence what you are really doing is trying to find y_{n+1} given y_n and y_{n-1} such that you find a root of that function, that is, a zero of the function. We won't use this notion much in this course, but these ideas/mathematics are very close to many *root-finding* algorithms in numerical analysis (#MAT331).

Definition 1.3

- Any equation of the form $X_{n+1} = F(X_n, X_{n-1}, \dots, X_{n-k})$ is called ***autonomous*** if there is no explicit dependence on the independent variable, n (other than in the subscripts for $X_{n+1}, X_n, X_{n-1}, \dots, X_{n-k}$).
- If the equation has inherent dependence on the independent variable n , it is called ***non-autonomous***.

For example:

1. $x_{n+1} = (n + 2)X_n + 7$ (non-autonomous)
2. $x_{n+1} = x_{n-2} - 3x_{n-1} + 17$ (autonomous)
3. $x_{n+1} = x_n + 7n^3$ (non-autonomous)
4. $x_{n+1} = \frac{x_n}{1+x_{n-1}+x_{n-2}}$ (autonomous)

Lastly, we finally mention that all of these iterative maps we have been discussing over the past few pages also go by many other names. To name a few, they also go by *recursive equations*, *difference equations*, or finally what we will call them for this course, ***discrete dynamical systems***.

The main takeaways from this section:

- Iterating a discrete dynamical system is rather easy, given a map and sufficient initial values to get the process started - computers make this process even easier!
- The *order* of an iterative map is equal to the number of initial values necessary to start the iteration process.
- *Linear* vs. *Non-linear* discrete dynamical systems.
- *Homogeneous* vs. *Non-homogeneous* iterative maps.
- *Autonomous* vs. *Non-autonomous* iterative maps.

- We care about classifying equations because their classification will usually guide us into how we analyze the equation - whether we try to find exact solutions or only hope to analyze properties of solutions.

1.2 Linear Discrete Dynamical Equations

This section will focus on finding *exact* solutions to individual *linear*, autonomous discrete dynamical systems. What we mean by individual is that there is only one equation present, e.g.,

$$X_{n+1} = c_0 X_n + c_1 X_{n-1} + \dots + c_{n-k} X_{n-k},$$

and thus we only care about the solution set $\{X_j\}_{j=0}^{\infty}$. Now even though we will be able to find exact solutions in this section, there are a lot of cases for us to consider. Some possible avenues we could explore are:

- How does finding solutions for 1st, 2nd, ..., or k^{th} -order linear, autonomous equations differ from one another?
- What happens if we have a homogeneous vs. non-homogeneous equation?
- What about autonomous vs. non-autonomous equations?

Of course, there are also the combinations of all the above possible ventures as well. However, rather than jump right into the deep end head first, let's start off with basic first-order linear, autonomous, homogeneous equations.

1.2.1 Linear, autonomous, homogeneous equations

A first-order, linear, autonomous, homogeneous equation is of the general form:

$$X_{n+1} = kX_n \quad (1.1)$$

It is properly accompanied by one initial value, say $X_0 = \sigma$, where σ is some finite number. In fact, let's just make it more concrete, so let $X_0 = \sigma = 5$ and $k = -2$. Keep in mind that neither the initial value or coefficient value chosen here won't change the solution technique we'll develop, but that we could have done this process completely in general (#flexingMathMuscles).

As we saw in Section 1.1, given a first-order equation and one initial value, we could iterate away. Although, here we wish to solve the equation *exactly*. This means our goal is to find a function such that when we evaluate the function at any value of the independent variable (n), it will give us the exact value that would have been found through the iteration process. However, there is a tremendously powerful insight to be gained by observing the iteration process:

$$\begin{aligned} X_0 &= 5 \quad (\text{given}) \\ X_1 &= -2X_0 = -2(5) = 5(-2) \\ X_2 &= -2X_1 = -2(5(-2)) = 5(-2)^2 \\ X_3 &= -2X_2 = -2(5(-2)^2) = 5(-2)^3 \\ X_4 &= -2X_3 = -2(5(-2)^3) = 5(-2)^4 \\ &\vdots \end{aligned}$$

Maybe you are starting to see a pattern? It looks like for any $n > 0$, our solution to Eq. 1.1 could be written as

$$X_n = 5(-2)^n.$$

Well in this first-order case, a pattern could be found just by iterating; however, we have also stumbled upon a larger idea for linear, autonomous equations: *It appears that the n^{th} iteration value is always equal to some constant times a value raised to the n^{th} power.* What this means in practice is that we may have gotten around iterating a bunch of times if we assumed a solution to Eq. 1.1 that had the following form:

$$X_n = \lambda^n,$$

where λ is an unknown, finite number that we can solve for. Now with this solution *ansatz* (meaning an ‘educated’ guess or hypothesis to what the solution form may take), we can test to see if it works or not by plugging it into Eq. 1.1. If we can find a value for λ that works, then we have found a solution!

Since we're assuming (and hoping) that $X_n = \lambda^n$, upon substituting it into Eq. 1.1, we find that

$$X_{n+1} = -2X_n \Rightarrow \lambda^{n+1} = -2\lambda^n.$$

Once we have completed our substitution, our goal is to find the λ value that makes this equation work out!

Let's solve $\lambda^{n+1} = -2\lambda^n$:

$$\begin{aligned} \lambda^{n+1} &= -2\lambda^n \\ \frac{\lambda^n \lambda}{\lambda^n} &= -2 \frac{\lambda^n}{\lambda^n} \quad (\text{divide both sides by } \lambda^n) \\ \lambda &= -2 \end{aligned}$$

Hence we have found that $\lambda = -2$ and therefore our solution looks like $X_n = \lambda^n = (-2)^n$. Note that we could divide both sides by λ since we could safely assume that $\lambda \neq 0$. Otherwise our solution would simply be $X_n = 0^n = 0$. So we are looking for more "interesting" solutions, meaning *non-trivial* solutions to our homogeneous equation.

Now that we did that work for our finding the solution, let's test it! Using $X_n = (-2)^n$,

$$\begin{aligned} X_0 &= 5 \\ X_1 &= (-2)^1 = -2 \\ X_2 &= (-2)^2 = 4 \\ X_3 &= (-2)^3 = -8 \\ X_4 &= (-2)^4 = 16 \\ &\vdots \end{aligned}$$

Um... did you notice anything fishy when you compare these iterations to the first ones we did? They are not the same! In fact, the ones we just did didn't even use the initial condition $X_0 = 5$ at all! What happened?

Well, it turns out that we didn't use one of the main properties of this type of equation - it's *linearity*. When we assumed our solution *ansatz* we only assumed $X_n = \lambda^n$, when in reality we need a constant coefficient attached to it as well $X_n = c_0 \lambda^n$, where now both c_0 and λ are unknown values. However, due to linearity properties of the equation, that is, because it is linear, when we perform the substitutions, the c_0 constants will cancel out everywhere! Check it out, using $X_n = c_0 \lambda^n$ and assuming $c_0 \neq 0$ (otherwise we get the trivial solution, 0), we find

$$X_{n+1} = -2X_n \Rightarrow c_0 \lambda^{n+1} = -2c_0 \lambda^n \Rightarrow \lambda^{n+1} = -2\lambda^n.$$

Then we would perform the same arithmetic and algebra as we did before and find that $\lambda = -2$. This is a powerful aspect of linear equations. These

sorts of unknown constant coefficients that we need to solve for will always cancel out upon substituting. This will become more clear when we do a 2^{nd} -order, linear, autonomous equation next. But first, let's find out who c_0 actually is in our newly renovated solution guess. To do this, we get to finally use the initial value! Since we have that

$$X_n = c_0(-2)^n$$

we can use the initial condition

$$X_0 = 5$$

to solve for c_0 now. Plugging in $n = 0$ and the initial condition into the above solution we find that:

$$X_0 = 5 = c_0(-2)^0 \Rightarrow 5 = c_0(1),$$

and therefore we have that $c_0 = 5$. So our full *exact* solution is $X_n = 5(-2)^n$.

Of course this solution matches the solution we found by picking up on a pattern through all of the iterations, but it lead us to a general solution strategy for solving linear, autonomous, homogeneous discrete dynamical system equations. Let's outline it below:

Strategy 2.1

- For a general first-order, autonomous, homogeneous discrete dynamical system such as $X_{n+1} = A_0 X_n$ with initial value $X_0 = \sigma$ we can assume a solution *ansatz* of

$$X_n = c_0 \lambda^n$$

and then via substitution, algebra, and arithmetic find what λ is, followed by using the initial value to find c_0 .

- For a general linear, autonomous, homogeneous discrete dynamical system, such as $X_{n+1} = A_0 X_n + A_1 X_{n-1} + A_2 X_{n-2} + \dots + A_k X_{n-k}$, with $k+1$ initial values $\{X_j\}_{j=0}^k$ the process can be extended analogously. We can still assume solution *ansatz* of

$$X_n = \lambda^n$$

and then via substitution, algebra, and arithmetic find ALL the λ values that satisfy the equation. For a $(k+1)^{st}$ -order equation, you will find $\{\lambda_j\}_{j=0}^k$ values. Assuming these values are all unique, the solution can then be written as a linear combination involved

ALL λ values, i.e.,

$$X_n = c_0\lambda_0^n + c_1\lambda_1^n + c_2\lambda_2^n + \dots + c_k\lambda_k^n.$$

Now you must find all of the $\{c_j\}_{j=0}^k$ coefficients that help the solution satisfy the initial values. You then use all $k+1$ initial values to find the coefficient values $\{c_j\}_{j=0}^k$. It is the same process, but just a bit heavier on the subsequent algebraic and arithmetic steps.

- Note:

1. We will do an example of the latter next.
2. Upon substituting and performing algebra to find $\{\lambda_j\}_{j=0}^k$, you will inevitable need to find the roots of a polynomial equation. This is called the *characteristic equation*.
3. The $\{\lambda_j\}_{j=0}^k$ values you find, i.e., the roots of the *characteristic equation*, are also called *eigenvalues*.
4. The above approach also assumes that the algebraic multiplicity = the geometric multiplicity, i.e., that all $\{\lambda_j\}_{j=0}^k$ are unique and there are no repeated factors. For the time being, repeated factors are above our pay grade in this course.

□

We will now dive into an example of a 2^{nd} order equation to illustrate the above Strategy's latter description.

Example 1.2.1. *The Fibonacci sequence is a famous example of a 2^{nd} -order, linear, autonomous, homogeneous discrete dynamical system. It is the sequence defined by*

$$1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, 233, 377, 610, 987, 1597, \dots$$

Basically, the next number in the sequence is the sum of the previous two iterates; hence we can represent it as

$$F_{n+1} = F_n + F_{n-1}, \tag{1.2}$$

with initial values $F_0 = 1$ and $F_1 = 1$. Note that Eq. 1.2 is the discrete dynamical system governing the Fibonacci sequence but without the specific initial values of $F_0 = F_1 = 1$, Eq. 1.2 could lead to another sequence of values entirely.

Since the Fibonacci sequence is an example of a 2nd-order, linear, autonomous, homogeneous equation, we can apply the solution strategy outlined in Strategy 2.1. Therefore we will assume our solution ansatz of

$$F_n = c_0 \lambda_0^n + c_1 \lambda_1^n, \quad (1.3)$$

where λ_0 and λ_1 can be found through substitution and c_0 and c_1 are unknown coefficients that will be found by enforcing the initial values at the end of the day. Let's try it out!

Substituting (1.3) into (1.2), we get the following:

$$\begin{aligned} [F_{n+1}] &= [F_n] + [F_{n-1}] \\ &\quad \text{substitute ansatz solution} \\ [\lambda^{n+1}] &= [\lambda^n] + [\lambda^{n-1}] \\ &\quad \text{divide by } \lambda^{n-1} \\ \lambda^2 &= \lambda + 1 \\ &\quad \text{move over to one side} \\ \lambda^2 - \lambda - 1 &= 0 \\ &\quad \text{use quadratic formula to find roots of characteristic poly.} \\ \lambda &= \frac{1 \pm \sqrt{5}}{2}. \end{aligned}$$

Therefore we have two roots (eigenvalues) of the characteristic polynomial and they are both unique. We can now write our general solution to (1.2) as the following

$$F_n = c_0 \left(\frac{1 + \sqrt{5}}{2} \right)^n + c_1 \left(\frac{1 - \sqrt{5}}{2} \right)^n.$$

To find c_0 and c_1 we can use the initial values $F_0 = 1$ and $F_1 = 1$. As a spoiler, enforcing these initial values will lead us to a linear system of equations to solve for c_0 and c_1 . Check it out:

$$\begin{aligned} F_0 = 1 &\Rightarrow F_0 = 1 = c_0 \left(\frac{1 + \sqrt{5}}{2} \right)^0 + c_1 \left(\frac{1 - \sqrt{5}}{2} \right)^0 = c_0 + c_1 \\ F_1 = 1 &\Rightarrow F_1 = 1 = c_0 \left(\frac{1 + \sqrt{5}}{2} \right)^1 + c_1 \left(\frac{1 - \sqrt{5}}{2} \right)^1 = c_0 \left(\frac{1 + \sqrt{5}}{2} \right) + c_1 \left(\frac{1 - \sqrt{5}}{2} \right) \end{aligned}$$

Hence we have the following linear system of equations to solve:

$$c_0 + c_1 = 1$$

$$c_0 \left(\frac{1 + \sqrt{5}}{2} \right) + c_1 \left(\frac{1 - \sqrt{5}}{2} \right) = 1$$

There are two equations and two unknowns. At this junction you could either implement some linear algebra matrix techniques or solve by substitution alone. Let's go with the substitution approach:

From $c_0 + c_1 = 1$, we rearrange to find that $c_1 = 1 - c_0$. Now we can substitute that definition of c_1 into the other equation to get:

$$c_0 \left(\frac{1 + \sqrt{5}}{2} \right) + (1 - c_0) \left(\frac{1 - \sqrt{5}}{2} \right) = 1.$$

Now we have one equation that is fully in terms of a single unknown (c_0) so let's solve for it!

Distributing and collecting terms we find that

$$c_0 \left[\frac{1}{2} + \frac{\sqrt{5}}{2} - \frac{1}{2} + \frac{\sqrt{5}}{2} \right] + \left(\left(\frac{1 - \sqrt{5}}{2} \right) = 1 \right)$$

and then

$$\sqrt{5}c_0 = \frac{1}{2} + \frac{\sqrt{5}}{2}$$

so doing one final division by $\sqrt{5}$ gives

$$c_0 = \frac{1 + \sqrt{5}}{2\sqrt{5}}.$$

Therefore since $c_1 = 1 - c_0$, we find that

$$c_1 = \frac{\sqrt{5} - 1}{2\sqrt{5}}.$$

The exact solution to the Fibonacci sequence is given by

$$F_n = \frac{1 + \sqrt{5}}{2\sqrt{5}} \left(\frac{1 + \sqrt{5}}{2} \right)^n + \frac{\sqrt{5} - 1}{2\sqrt{5}} \left(\frac{1 - \sqrt{5}}{2} \right)^n.$$

Now you can find the value of any term in the Fibonacci sequence without having to compute all the iterates up to that value.

As long as the the discrete dynamical system is linear, autonomous, homogeneous, and that all the roots of the characteristic polynomial are unique, i.e., the eigenvalues are all unique with none repeated, the above strategy will always be successful for finding the solution. Some of the difficulties that could arise are from actually having to find the roots of a polynomial (especially if you are without computational help) or solving a big system of linear equations that could get rather tedious. Both of these can be may be avoided if we embrace computational techniques ☺.

Next along our journey for solving linear, autonomous discrete dynamical systems is exploring what happens if the equation is non-homogeneous. That is, what happens if there is a non-zero right hand side once all iterates are moved onto the left same side of the equation.

1.2.2 Linear, non-homogeneous equations

We will begin by considering a first-order, linear, but *non-homogeneous* with the following form:

$$X_{n+1} = kX_n + r \quad (1.4)$$

where k is a constant and r could be a function of the independent variable, n . That is, $k \in \mathbb{R}$ and $r = r(n)$. This equation is *non-homogeneous* because if we move all the iterative variables to the left side we get

$$X_{n+1} - kX_n = r \neq 0.$$

The key idea in solving these equations is that we will need to solve for TWO different solutions: a *general* solution and a *particular* solution.

Definition 2.1

- The **complementary solution**, denotes X_n^c solves the homogeneous equation. For the example equation (1.4) this would mean X_n^c solves

$$X_{n+1}^c - kX_n^c = 0.$$

- The **particular solution**, denoted X_n^p solves the non-homogeneous equation. For the example equation (1.4) this would mean X_n^p solves

$$X_{n+1}^p - kX_n^p = r.$$

- The full solution would then entail a sum of the complementary solution and particular solution, i.e.,

$$X_n = X_n^c + X_n^p.$$

- Note: that finding the complementary solution will assume Strategy 2.1. However, one would wait to solve the unknown coefficients until after you have combined the complementary solution and particular solution. Only once you've added them together, i.e., $X_n = X_n^c + X_n^p$, could you then invoke the initial values and then find the unknown coefficients in the complementary solution.

Since we've already gone through the rigmarole of finding the complementary solution in Section 1.2.1, let's dive into the strategy we will use to find the *particular solution*. The method we will use is called the **Method of Undetermined Coefficients**.

In a nutshell, the Method of Undetermined Coefficients can help us find only the particular solution of a non-homogeneous problem. As most things

in solving discrete dynamical systems (or differential equations), the approach is highly dependent on the form of the non-homogeneous forcing term. What this means is that, for example in the equation listed above,

$$X_{n+1} = kX_n + r(n)$$

depending on what the form of $r(n)$ takes, governs how we go about performing the Method of Undetermined Coefficients. Maybe that seems a bit cryptic at this point.

Perhaps what is better to start with is that the Method of Undetermined Coefficients has the flavor of a glorified *guess and check* method. That is, we guess a specific form of the particular solution depending on what type of functional form $r(n)$ has. For example, if $r(n) = -3n^2 + 5n + 7$, we might guess that our particular solution takes the form of $X_n^p = An^2 + Bn + C$, where it is our goal to find A, B and C that make it work. How about now I just give you some examples of how we would define X_n^p based on possible non-homogeneous forcing functions?

$r(n)$	X_n^p
5	$X_n^p = A$
$5n - 3$	$X_n^p = An + B$
$15n^3 - 3n + 1$	$X_n^p = An^3 + Bn^2 + Cn + D$
e^{7n}	$X_n^p = Ae^{7n}$
e^{-3n}	$X_n^p = Ae^{-3n}$
$\sin(3n)$	$X_n^p = A \sin(3n) + B \cos(3n)$
$\cos(8n)$	$X_n^p = A \sin(8n) + B \cos(8n)$
$e^{7n} \sin(2n)$	$X_n^p = e^{7n} (A \sin(2n) + B \cos(2n))$

Table 1.1: Possible forms of forcing functions and their associated particular form solutions when using the Methods of Undetermined Coefficients

Okay, it's probably time for a few examples. Remember - we have to find both a *particular* solution that solves the non-homogeneous equation AND a *complementary* solution that solves the homogeneous equation.

Example 1.2.2. Let's solve the following first-order, linear, autonomous, non-homogeneous discrete dynamical system:

$$X_{n+1} = -2X_n + 7n - 3. \quad (1.5)$$

with initial value $X_0 = 11$

1. First, we will find the **particular solution**, X_n^p . Notice that the forcing function $r(n) = 7n - 3$ takes the form of a line, so we can try to guess the particular solution takes the same form (see Table 1.1), i.e.,

$$X_n^p = An + B,$$

where we get to solve for unknowns A and B to make it work. Substituting $X_n^p = An + B$ into (1.5), we get that

$$X_{n+1}^p = -2X_n^p + 7n - 3 \quad \Rightarrow \quad A(n+1) + B = -2[An + B] + 7n - 3.$$

Now distributing we get

$$An + A + B = -2An - 2B + 7n - 3.$$

We can now move all the terms with A or B coefficients onto the left hand side. This gives us

$$3An + A + 3B = 7n - 3.$$

Similar to the method of Partial Fractions in Calculus B, we can collect and write equations based on different powers of n to give us a system of linear equations:

$$\begin{aligned} n^0 : &= A + 3B = -3 \\ n^1 : &= 3A = 7. \end{aligned}$$

Hence we have two equations for our two unknowns A and B . The second equation gives us that $A = 7/3$. From the first equation, when solving for B , we find that $B = -\frac{3+A}{3}$ and hence substituting $A = 7/3$ into it gives us that $B = -16/9$. Therefore the particular solution is

$$X_n^p = \frac{7}{3}n - \frac{16}{9}.$$

2. Second, we get to find the complementary solution, X_n^c . That is, the solution that solves the homogeneous equation:

$$X_{n+1}^c = -2X_n^c.$$

Recall that for homogeneous equations like this we guess a solution of the form $X_n^c = \lambda^n$. Substituting this ansatz into our homogeneous equation gives us

$$X_{n+1}^c = -2X_n^c \Rightarrow \lambda^{n+1} = -2\lambda^n.$$

Therefore dividing by λ^n gives us that $\lambda = -2$. So the complementary solution takes the form of $X_n^c = (-2)^n$. Due to the linearity of the problem, we also get to attach an unknown coefficient c_0 onto it. Hence our complementary solution is:

$$X_n^c = c_0(-2)^n.$$

However, before finding the identify of c_0 to satisfy the given initial value, we need to combine both the particular and complementary solutions.

3. Third, once we have both the particular and complementary solutions, we can combine them to get the full solution:

$$X_n = X_n^c + X_n^p \Rightarrow X_n = c_0(-2)^n + \frac{7}{3}n - \frac{16}{9}.$$

We still need to enforce the initial value given at the beginning of the problem, $X_0 = 11$. Notice that we still have one unknown coefficient, c_0 , so we have enough degrees of freedom to enforce it! That is, we have one initial value and one unknown coefficient.

Now that we have combined the particular and complementary solutions, we can substitute the initial value (when $n = 0$) and find out the identify of c_0 :

$$X_0 = 11 = c_0(-2)^0 + \frac{7}{3}(0) - \frac{16}{9}.$$

Thereby, we get

$$11 = c_0 - \frac{16}{9} \Rightarrow c_0 = 11 + 16/3 = 115/9.$$

Our final solution is then

$$X_n = \frac{115}{9}(-2)^n + \frac{7}{3}n - \frac{16}{9}.$$

At this junction we have uncovered how to solve individual linear discrete dynamical systems if they are autonomous and either homogeneous or non-homogeneous. Next we will turn our efforts into studying *nonlinear* discrete dynamical systems. As a heads up, there is where the math gets spicy - but don't be alarmed. While we may not be able to find exact solutions, like we did in Sections 1.2.1 and 1.2.2, we will develop tools to help us analyze and explore the behavior of their solutions.

1.3 First-order Non-linear Discrete Dynamical Systems

Perhaps the best place to start would be adding a non-linear term that might feel familiar to many people as their first foray into non-linear functions in high school - a quadratic term. Let's first look at the following first-order non-linear, autonomous, homogeneous, iterative map:

$$X_{n+1} = AX_n + BX_n^2 \quad (1.6)$$

with $X_0 = \sigma$ where A and B are both constants, i.e., $A, B \in \mathbb{R}$. The reason we start here with this quadratic non-linearity is because with proper choices of A and B , we could redefine (1.6) into what is known as the *Discrete Logistic Equation*.

The Discrete Logistic Equation is usually written in the following form:

$$X_{n+1} = X_n + kX_n \left(1 - \frac{X_n}{C}\right), \quad (1.7)$$

where k and C are constants ($k, C \in \mathbb{R}^+$). Notice how in both (1.6) and (1.7) there are two constants in each, either k and C or A and B . You can transform (1.6) into (1.7) with the following mathematical slight of hand, aka substitution:

Let

$$A = 1 + k \quad \text{and} \quad B = -\frac{k}{C}.$$

Substituting those values of A and B into (1.6) will allow you to transform it into the familiar form of the Discrete Logistic Equation. Note that in the Discrete Logistic Equation, the constants (called *parameters*) are assumed to be strictly positive, i.e., $k, C > 0$. The reason for this assumption is because of its application to population growth. We will focus on the Discrete Logistic Equation (or equations that resemble it) for much of the analysis in this section, as it displays a plethora of beautiful, albeit complicated behavior!

1.3.1 The Discrete Logistic Equation

Let's start off by giving our new friend, the Discrete Logistic Equation, a proper mathematical introduction (via explicitly defining it):

Definition 3.1

- The standard **Discrete Logistic Equation**, is a first-order non-linear, autonomous, homogeneous discrete dynamical system. It displays a quadratic non-linearity. It is traditionally given as the following:

$$X_{n+1} = X_n + kX_n \left(1 - \frac{X_n}{C}\right) \quad (1.8)$$

where k, C are **parameters**. That is, they are constants ($k, C \in \mathbb{R}^+$) that we get to choose at the beginning of problem to try to match our equation (model) to data, if possible. Varying parameters may result in subtle or significant changes in solution behavior, as Figure 1.1 illustrates.

- The parameter C is called the *carrying capacity*. It is thought of as the largest number that the environment can sustain due to resources. The parameter k is a *growth rate* parameter.
- Note that many times in literature this equation may be written differently with proper mathematical slights of hand (i.e., redefining parameters). We saw previously that if we define $A = 1 + k$ and $B = -\frac{k}{C}$, we could have equivalently written this equation as

$$X_{n+1} = AX_n + BX_n^2.$$

Another way that this equation is commonly studied in the literature is if we let $\tilde{k} = 1 + k$ and $\tilde{C} = \frac{c\tilde{k}}{k}$, we could also equivalently write it as:

$$X_{n+1} = \tilde{k}X_n \left(1 - \frac{X_n}{\tilde{C}}\right).$$

This last version of the Discrete Logistic Equation looks very similar to (1.8) but without the one " $+X_n$ " term.

- One reason people rewrite equations in this manner is to redefine parameters so that any analysis they do on the equation later might be more streamlined (i.e., *easier*). However, when you

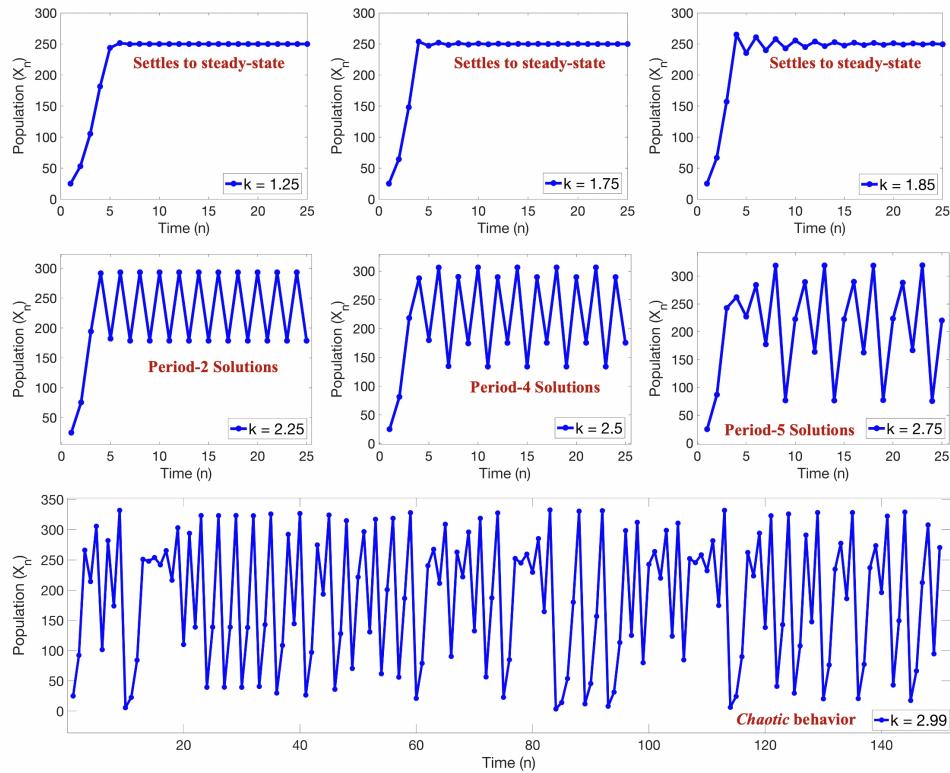


Figure 1.1: A variety of solutions to the Discrete Logistic Equation with $X_0 = 25$ and $C = 250$ for a variety of k values. Varying parameters can lead to substantial variations in the resulting dynamics.

start redefining parameters like this, the units of the parameters change. So, if you were trying to match (*estimate*) parameter values to experimental data, you have to be very careful to match units appropriately.

The first notable thing I wish to say about the Discrete Logistic Equation is that almost any type of gnarly behavior you could ever hope to dream a first-order discrete equation has, this equation will display. Some examples of possible solution behavior is seen in Figure 1.1. Some solutions tend to eventually plateau near the value of $C = 250$ ($k = 1.25, 1.75, 1.85$), while others exhibit periodic solutions ($k = 2.25, 2.5, 2.75$), and another displays what appears to be complicated, non-periodic behavior known as *chaos*. Recall that this equation arose from simply a natural extension of where we wanted to start exploring non-linear equations - a quadratic term!

Hopefully Figure 1.1 convinces you that a lot of complicated behavior

may be exhibited in solutions to non-linear equations when one (or maybe more) parameters are varied. Figure 1.1 varied the parameter, k , and kept both C constant and the initial value, X_0 , constant. Unfortunately, as you might expect, finding *exact* solutions to non-linear equations might be extremely difficult, if not impossible. The methods we used in Sections 1.2.1 and 1.2.2 for linear problems will be ineffective for solving non-linear equations. While there are some *exact* solutions to very specific (highly studied) non-linear equations, we will not focus our efforts there. Instead, we will focus on how to analyze properties of possible solution behaviors of non-linear equations. That is, we won't necessarily solve non-linear equations exactly, but instead develop a set of tools to peek into what we'd expect their solutions to do.

Before we do that, I want to briefly give you an idea of how difficult finding solutions to non-linear problems could be. Like mentioned previous, the Discrete Logistic Equation has been well-studied by many people and some have found a strategy for finding an exact solution to it. I'll briefly outline it below:

Example 1.3.1. *Finding an exact solution to the Discrete Logistic Equation:*

1. *First, we won't even consider the Discrete Logistic Equation, we look at its continuous analog*

$$\frac{dx}{dt} = kx \left(1 - \frac{x}{C}\right).$$

2. *We then use the method of separable equations from differential equations and integrate from X_n to X_{n+1} and n to $n + 1$ appropriately, i.e.,*

$$\int_{X_n}^{X_{n+1}} \frac{C}{x(C-x)} dx = \int_n^{n+1} k dt$$

3. *Using Partial Fractions you can find that*

$$\ln\left(\frac{C-X_{n+1}}{CX_{n+1}} \div \frac{C-X_n}{CX_n}\right) = -k$$

4. *You can then solve this logarithmic equation for X_{n+1} to find another non-linear discrete dynamical system*

$$X_{n+1} = \frac{Ce^k X_n}{C + (e^k - 1)X_n}.$$

Note that since the independent variable here is n , we can define a bulk parameter $\xi = e^k$.

5. Fortunately, you can then use the transformation

$$u_n = \frac{1}{X_n}$$

to transform this non-linear equation into a linear equation!

6. After using that transformation (with a bit of algebra) you can arrive at

$$u_{n+1} = \frac{1}{\xi} u_n + \frac{\xi - 1}{\xi C}.$$

This is a linear, autonomous, non-homogeneous iterative equation. We can actually solve this using the tools we developed in Sections 1.2.1 and 1.2.2!

7. Solving the above and then transforming back from $u_n \rightarrow X_n$, one can arrive at

$$X_n = \frac{\sigma C \xi^n}{C + \sigma(\xi^n - 1)},$$

where $X_0 = \sigma$

8. Phew. Glad we won't be needing to think that much outside the box regularly just to get the solution to one equation... if we need to find what a solution looks like we can just use... computational methods ☺!
-

The time is now in this course where we are going to make a transition away from calculating *exact* solutions to equations by hand and towards methods that help us *analyze what we'd expect of the behavior of solutions* to non-linear equations. That is, we're not going to solve for solutions explicitly but try to develop a set of tools so that we can predict behavior or properties that they might exhibit. Don't worry these methods will still be rigorous, and quite frankly, both allow us to use some pretty groovy math and apply it in ways that have direct biological significance.

1.3.2 Equilibria and Stability in First-order Discrete Equations

In this section we will begin our foray into developing tools for helping us understand the behavior of solutions to discrete dynamical equations, without explicitly solving for their solution. In order to do this, we will consider the following discrete equation as our test model:

$$X_{n+1} = X_n + kX_n \left(1 - \frac{X_n}{C}\right) \left(\frac{X_n}{r} - 1\right), \quad (1.9)$$

where $C, k, r \in \mathbb{R}^+$ and with $X_0 = \sigma$. Note that this looks very much of the form of the Discrete Logistic Equation (see Definition 3.1) but with an added term at the end. This term at the end adds what is called *depensation* (to fishery scientists) or the *Allee effect* (to ecologists). At this junction we won't concern ourselves with those definitions or what models with this added term hope to do. For our purposes here, we only wish to analyze the given equation.

First, to motivate our ideas in this Section I will show a few solutions for specific values of C, k, r , and σ . Here we will vary σ , the initial value, and observe what happens to the solutions. **What I want you to get from these solutions is that:**

1. Varying the initial condition can affect the resulting dynamics
2. There appears to be a few different values for which the solutions appear they could plateau, like how we saw some solutions plateau and steady-off in the Discrete Logistic Equation in Figure 1.1.

Figure 1.2 shows three different solution behaviors for different initial conditions that are all pretty close to each other $X_0 = \sigma = 99, 100$, and 101 , even though the parameters are at fixed values of $k = 0.25, C = 250$, and $r = 100$. Notice that each of these solutions tends to plateau (or steady-off) to some particular value, whether it is $0, r$, or C . Once they hit those specific values they seem stuck, or *fixed*, there. This is no coincidence - when solutions reach such states, where they appear stuck, or there are no more fluctuations or changes, they are called *equilibrium values* or *equilibrium points*. The solutions cease changing and hence the next solution value will always be equal to the prior, with the next after that the same, with the next after that the same, ..., forever.

The key idea that the solutions cease changing effectively means that once n gets large enough (say $n \geq N$ where N is *sufficiently* large), all of a

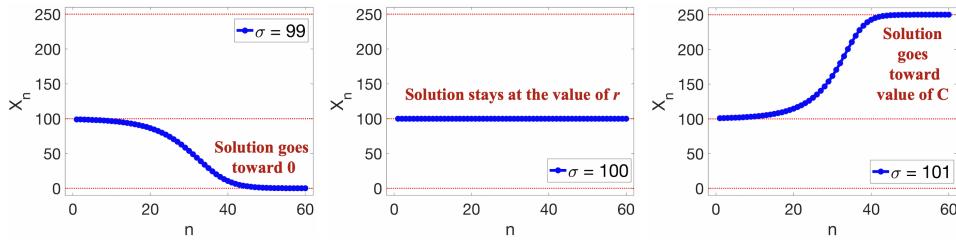


Figure 1.2: Three different solutions to the Discrete depensation/Allee effect model for $X_0 = \sigma = 99, 100, 101$ with $C = 250, k = 0.25, r = 100$. Notice that these solutions each seem to plateau or steady-off to three different values as n continues to increase, even though they initial all start off near each other.

sudden

$$X_N = X_{N+1} = X_{N+2} = X_{N+3} = \dots$$

All of a sudden from this observation that the solutions stop changing, we have a mathematical strategy for finding such *equilibrium values*:

Strategy 3.1

Finding equilibrium values to discrete dynamical systems

1. If $X_N = X_{N+1} = \dots$, define \bar{X} to be that value, i.e., $\bar{X} = X_N = X_{N+1} \dots$
2. Then in the discrete equation wherever there is an X_{n+1} , X_n , etc. we could replace them with \bar{X} .
3. Once the discrete equation is now fully in terms of \bar{X} , we can solve for \bar{X} , thus finding what equilibrium values are possible!

□

Let's try this strategy out on (1.9) above. Letting $\bar{X} = X_{n+1} = X_n$ and substituting into (1.9), we get:

$$\bar{X} = \bar{X} - k\bar{X} \left(1 - \frac{\bar{X}}{C}\right) \left(\frac{\bar{X}}{r} - 1\right).$$

Notice that the above is one equation and only has one unknown (\bar{X}), since the three parameters k, C, r are assumed to be known constants (or constants that could be determined from data presumably). Therefore, let's go ahead and try to solve for \bar{X} :

$$\begin{aligned}\bar{X} &= \bar{X} - k\bar{X} \left(1 - \frac{\bar{X}}{C}\right) \left(\frac{\bar{X}}{r} - 1\right) \\ 0 &= -k\bar{X} \left(1 - \frac{\bar{X}}{C}\right) \left(\frac{\bar{X}}{r} - 1\right) \\ 0 &= \bar{X} \left(1 - \frac{\bar{X}}{C}\right) \left(\frac{\bar{X}}{r} - 1\right)\end{aligned}$$

So, we have the product of three things equally zero, therefore one or more of those must be zero. Setting equal of them equal to 0 gives:

$$\begin{aligned}\bar{X} &= 0 \quad \Rightarrow \quad \bar{X} = 0. \\ 1 - \frac{\bar{X}}{C} &= 0 \quad \Rightarrow \quad \bar{X} = C \frac{\bar{X}}{r} - 1 = 0 \quad \Rightarrow \quad \bar{X} = r\end{aligned}$$

At this junction we have found three equilibrium values to (1.9): $\bar{X} = \{0, r, C\}$. You will find that if you compare these values we found mathematically to those predicted in Figure 1.2, that they fully agree. Moreover, this mathematical strategy also lets us determine that *there were only 3* possible equilibrium values to (1.2), which blindly solving (1.2) for A LOT of different parameter value (k, C, r) and initial value (σ) combinations may not have easily showed us.

What I want you to take away here is:

1. We can find equilibrium values using this mathematical strategy outlined above
2. However, just because we have found what equilibrium values are possible for the Discrete equation to tend towards, we still don't know what parameter values or initial values would lead us to these equilibrium values!
3. To find the equilibrium values, we had to do some sort of root-finding technique. On equations such as (1.9) this process was straight-forward and manageable; however, this process could become very difficult very fast.

Lastly, I want us to return to Figure 1.2. Notice that although we kept all the parameters constant and only varied the initial value $X_0\sigma$ that we tended towards different equilibrium values. Moreover, notice how the initial values we started with were all relatively close! You might even posit a hypothesis that even if you started really close but below the value of $\bar{X} = r$ you may have tended away from the equilibrium at $\bar{X} = r$ and towards the

equilibrium at $\bar{X} = 0$ (left-most figure), or similarly that even if you started really close but above the value $\bar{X} = r$ you may have tended away from the equilibrium at $\bar{X} = r$ and towards the equilibrium at $\bar{X} = C$ (right-most figure). Remember that there was an equilibrium value at $\bar{X} = r$, but solutions still moved away from it, even though they started near it! This idea of solutions gravitating towards or away from an equilibrium brings us to our next venture - discussing *stability* of equilibrium points.

Before we mosey onto discussing stability of equilibrium points, let's state a formal definition of an equilibrium point.

Definition 3.2

An *equilibrium point* of a k^{th} -order discrete dynamical equation

$$X_{n+1} = f(X_n, X_{n-1}, X_{n-2}, \dots, X_{n-(k-1)})$$

is a constant solution, denoted \bar{X} , such that

$$\bar{X} = f(\bar{X}, \bar{X}, \bar{X}, \dots, \bar{X}).$$

- For a general first-order equation such as $X_{n+1} = f(X_n)$ this would mean $\bar{X} = f(\bar{X})$.
- Equilibrium points are often referred to as *fixed points*, *equilibrium values*, or *steady-state solutions* interchangeably.

The idea of *stability* of an equilibrium point really gets at asking the following question: *if a solution ever goes near an equilibrium value, will it continue to gravitate towards it, or will it repel and move away?* As we saw from Figure 1.2 solutions tended to move away from the equilibrium value at $\bar{X} = r$. In fact, the only time the solution stayed at $\bar{X} = r$ was when the initial value was identically equal to that equilibrium value, i.e., when $\sigma = 100$. For this reason, we would call the equilibrium at $\bar{X} = r$ *unstable* because all solutions moved away from it, even if were near it, i.e., the cases when $\sigma = 99$ or 101 .

The concept of stability actually gets into some pretty deep mathematical analysis territory. Don't worry I'll list them at the end of this section. Fortunately for us, there is a beautiful theorem that allows us to discuss the stability of an equilibrium point by embracing our Calculus A knowledge! Spoiler: we only need to know how to compute derivatives, although it may appear more complicated than that at first glance. Without further ado, here is that theorem:

Theorem 3.1

Consider a discrete dynamical equation such as $X_{n+1} = f(X_n)$ with initial value $X_0 = \sigma$ and equilibrium point \bar{X} , i.e., $\bar{X} = f(\bar{X})$. Assume $f'(x)$ is continuous on an open interval I containing the equilibrium point \bar{X} . Then \bar{X} is said to be a *locally asymptotically stable equilibrium* of $X_{n+1} = f(X_n)$ if

$$|f'(\bar{X})| < 1$$

and *unstable* if

$$|f'(\bar{X})| > 1.$$

- People sometimes simply refer to a locally asymptotically stable equilibrium as a *stable equilibrium*. However, there are some subtle issues that can arise from being this cavalier, see the definition of *locally stable* in Definition 3.3.
- Note that in cases of $|f'(\bar{X})| = 1$, the *Schwarzian derivative* might help determine stability. We will not focus on that case in this course.

A proof of this theorem is given at the close of this section (see Proof 1.3.1). All of the models we will explore in this class will satisfy the condition that $f(x)$ is differentiable and that $f'(x)$ is continuous on an open interval. Therefore really what we need to do practically in determining stability of an equilibrium point is:

1. Take a derivative of the iterative map function, $f(x)$, to get $f'(x)$.
2. Evaluate that derivative at the equilibrium value, i.e., $f'(\bar{x})$
3. Throw absolute values around the derivative, i.e., $|f'(\bar{x})|$ and see whether it is less than or greater than 1.

Let's try out using Theorem 3.1 and see what it tells us about the equilibrium points we found for (1.9). Recall that for the equation

$$X_{n+1} = X_n + kX_n \left(1 - \frac{X_n}{C}\right) \left(\frac{X_n}{r} - 1\right)$$

we found equilibrium points at

$$\bar{X} = \{0, r, C\}.$$

Example 1.3.2. Determine the stability of the equilibrium points for (1.9).

- First, note that iterative map function $f(x)$ of (1.9) is:

$$f(x) = x + kx \left(1 - \frac{x}{C}\right) \left(\frac{x}{r} - 1\right).$$

Next we will distribute so that taking a derivative is a bit more straightforward. Expanding the above gives

$$f(x) = (1 - k)x + kx^2 \left(\frac{1}{r} + \frac{1}{C}\right) - \frac{kx^3}{rC}.$$

Taking a derivative of it, we get

$$f'(x) = (1 - k) + 2kx \left(\frac{1}{r} + \frac{1}{C}\right) - \frac{3kx^2}{rC}.$$

- Now let's evaluate at the three different equilibrium points to see what we get:

1. $\bar{X} = 0$:

$$|f'(0)| = |1 - k|$$

Hence if $|1 - k| < 1$, the equilibrium point at $\bar{X} = 0$ is locally asymptotically stable.

2. $\bar{X} = C$:

$$|f'(C)| = \left|(1 - k) + 2kC \left(\frac{1}{r} + \frac{1}{C}\right) - \frac{kC^2}{rC}\right| = 1 + k \left(1 - \frac{C}{r}\right).$$

Hence if

$$\left|1 + k \left(1 - \frac{C}{r}\right)\right| < 1,$$

the equilibrium point at $\bar{X} = C$ is locally asymptotically stable.

3. $\bar{X} = r$:

$$|f'(r)| = \left|(1 - k) + 2kr \left(\frac{1}{r} + \frac{1}{C}\right) - \frac{kr^2}{rC}\right| = 1 + k \left(1 - \frac{r}{C}\right).$$

Hence if

$$\left|1 + k \left(1 - \frac{r}{C}\right)\right| < 1,$$

the equilibrium point at $\bar{X} = r$ is locally asymptotically stable.

- At this junction, I hope that you observed that the stability properties of the equilibrium points depend on inequalities that involve the equation's parameters! Hence, changing parameter values, e.g., $k = 0.25$ to $k = 1.75$, could quite possibly substantially change the resulting stability properties of the equilibria.
- Recall that for Figure 1.2, we used the following parameter values when solving (1.9): $k = 0.25$, $r = 100$, and $C = 250$. Inserting those values into the above stability inequalities we get:
 1. $\bar{X} = 0$: $|1 - k| = |1 - 0.25| = 0.75 < 1$ therefore *locally asymptotically stable*.
 2. $\bar{X} = C$: $|1 + k\left(1 - \frac{C}{r}\right)| = |1 + 0.25\left(1 - \frac{250}{100}\right)| = 0.625 < 1$ therefore *is locally asymptotically stable*
 3. $\bar{X} = r$: $|1 + k\left(1 - \frac{r}{C}\right)| = |1 + 0.25\left(1 - \frac{100}{250}\right)| = 1.15 > 1$ therefore *is unstable*
- Note that these confirmations of which equilibria were stable or unstable matched where solutions tended to settle down in Figure 1.2. However, at no point did we actually invoke the initial value σ in any of this stability analysis, even though different values of σ are what led to solutions behaving differently. **Remember, the idea of stability of an equilibrium point only pertains to a region close enough to that equilibrium point.** It turned out that in the above example, solutions didn't need to be incredibly close to either $\bar{X} = 0$ or $\bar{X} = C$ for solutions to gravitate closer to those equilibria.
- Lastly, for this example stemming from applications in fishery science and population dynamics, I want to quickly note that all the parameters, k, r, C are assumed to be positive and also that $r < C$. With these parameter assumptions in mind you can convince yourself that the equilibrium at $\bar{X} = r$ must always be unstable. However, depending on values of k, r , and C the other two equilibrium could either be stable or unstable as well. Thus, the stability of equilibrium points is highly dependent on parameter values.

Stability analysis requires that we know what the equilibrium points are. If we didn't have those, we simply could not perform stability analysis - we need to evaluate derivatives at the equilibrium points after-all! As briefly mentioned in the preceding text, finding the equilibrium points could prove very difficult, as you likely would need to perform a root-finding scheme on a non-linear equation. In this example using Eq (1.9), we were able to find \bar{X} in a straight-forward manner; however, this likely is not the case for

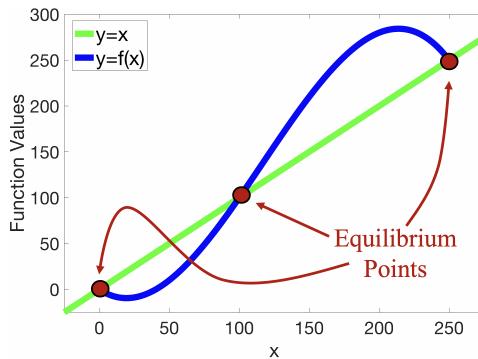


Figure 1.3: A cobweb diagram showing the equilibrium solutions for the Model given by (1.9).

many equations. The mathematics could get very tedious or involve (literally) impossible algebra (#5thOrderDegreePolynomialsOrAboveAnyone?). While turning to computers and using symbolic computation, such as those offered in Mathematica, MATLAB, or even Python can certainly alleviate a lot of this difficulty, mathematical biologists (or dynamacists) did not have these luxuries until the past 15 years or so. One method they used to both find equilibrium points and determine their stability concurrently was a visual method called *cob-webbing*.

The idea for *cob-webbing* is rather elegant. Simply plot your discrete equation's iterative function $f(x)$ against the line $y = x$. The points of intersection of the line $y = x$ and the function $y = f(x)$ are all the equilibrium values! See Figure 1.3 for an illustration of this for the example using Eq. (1.9) for the case of $k = 2$, $r = 100$, and $C = 250$.

Next start with a particular initial value *close* to one of those points of intersection and begin the cobweb iteration process. However, in the cobweb diagram what you can do move vertically from the initial condition (X_0) to where you touch the function $X_1 = f(X_0)$, see Figure 1.4a. Next move horizontally until you touch the line $y = x$. Then use that resulting x value (X_1) as your next value - that is, find the value of $X_2 = f(X_1)$ for that x value, see Figure 1.4b. Move horizontally from that value of $f(X_1)$ to where it hits $y = x$. Keep going (see Figure 1.4c). Eventually you may spiral into one point, suggested that point is both an equilibrium value and *locally asymptotically stable*. If you spiral away from a point, then it could be *unstable*. However, another interesting thing may occur where your solution doesn't always tend towards a particular equilibrium point (or away for that matter) and you can start to see the presence of periodic solutions!

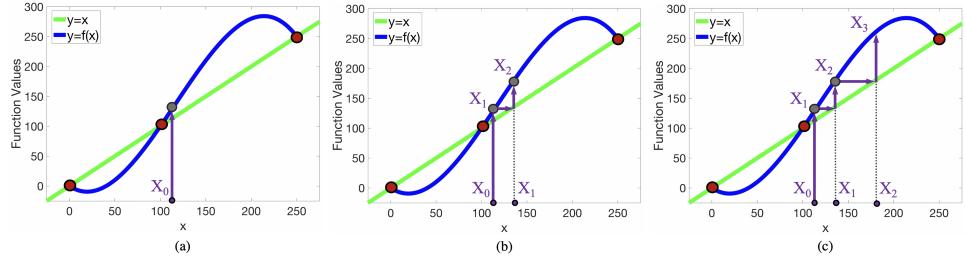


Figure 1.4: A cobweb diagram illustrating how the process evolves.

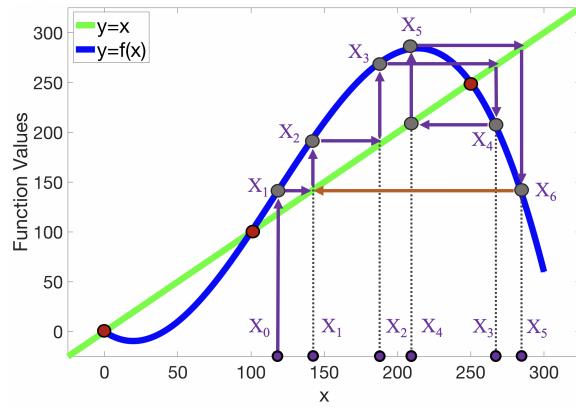


Figure 1.5: A cobweb diagram that suggests there might be a possible periodic solution with period 5!

For example, if we continued the cobwebbing process for the example outlined above in Figures 1.3 and 1.4 for (1.9) with parameters $k = 2, r = 100$, and $C = 250$, it appears that eventually we might *appear* that we loop back to an earlier value! Figure 1.5 shows this idea. In Figure 1.5, the sketch *suggests* that X_7 would be identical to X_2 , thus $X_7 = X_2$. However, if that is the case, we would just keep looping around the previous iterations along the cobweb diagram forever and never reach an equilibrium. Thus, our cobweb diagram suggests to us that these parameter combinations and initial value might give rise to periodic solutions. In this case, our sketch suggests that we would have periodic solutions with period 5. However, *we did not prove this, but rather the cobweb diagram helped us believe that periodic solutions may exist for this discrete equation.*

You know what, let's briefly explore the existence and stability of periodic solutions to discrete equations (see Section 1.3.3). But first, as promised, I will divulge some more analysis jargon heavy definitions and a theorem related to the stability of equilibrium points: what it means for an equilibrium

point to be *locally stable* as well as *locally attracting*. The mathematical hammer we get from these two definitions comes in the form of a theorem that says an equilibrium point being both locally stable and locally attractive guarantees that it is locally asymptotically stable as well, that is, stable.

Definition 3.3

An equilibrium point \bar{X} is called ***locally stable*** if for every $\epsilon > 0$ there exists a $\delta > 0$ such that if $|X_0 - \bar{X}| < \delta$ then

$$|X_n - \bar{X}| = |f(f(\cdots f(X_0))) - \bar{X}| < \epsilon$$

for every $n \geq 0$.

- The notation $f(f(\cdots f(X_0)))$ refers to applying the iterative map n times, i.e., iterating n times to find X_n .
- This definition basically says that solutions *near* an equilibrium point will always stay *near* that equilibrium point.
- If \bar{X} is *not locally stable*, then it is called ***unstable***.

Definition 3.4

An equilibrium point \bar{X} is ***locally attracting*** if there exists a $\gamma > 0$ such that for all $|X_0 - \bar{X}| < \gamma$ that

$$\lim_{n \rightarrow \infty} X_n = \lim_{n \rightarrow \infty} f(f(\cdots f(X_0))) = \bar{X}.$$

- This definition basically says that if a point is *sufficiently close enough* to an equilibrium point that the iterations will go towards that equilibrium point eventually.

If we can show that an equilibrium point is both *locally stable* and *locally attracting*, then we get the wield the following mathematical hammer (i.e., theorem):

Theorem 3.2

If an equilibrium is both *locally stable* and *locally attracting*, then it is *locally asymptotically stable*. That is, it is *stable*. In other words, the iterative map will eventually *converge* to that equilibrium point.

Note that although the above theorem gives us an alternative route for proving an equilibrium point is *stable*, we will not go about proving stability in this manner. In practical terms, we will only use Theorem 3.1 for that. The theorem above grants us more tools in our rigorous mathematical analysis toolbox, which are useful for formally developing more useful theorems, such as Theorem 3.1. In the grand scheme of mathematics, while deep mathematical analysis can surely be difficult, complex, and frightening at first, it ultimately helps provide us with practical tools in which we know when they work, why they work, and their limitations.

In my personal career, much of experience as an applied mathematician has been trying to wear two hats. The first hat is what I think of as a ‘formal mathematical analysis hat’. I wear this hat when developing strategic approaches, both analytical and numerical, to solving a real-world problem. This results in developing a new (dare I say novel?) method for solving an equation(s). I try to figure out why it may or may not work, what the strategy is limited by, and most importantly, when it would be appropriate to use. Rigorous mathematical analysis holds the key to answering these questions. On the other hand, the other hat I get to wear feels much more like ‘practitioner (or user) of analysis theorems hat’, in which I actually get to implement the strategy (new method) to solve a problem stemming from a real-world applications. The former provides us the formal strategy and the latter is when we actually get to wield it.

Finally, we end this section with a Proof of Theorem 3.1. This proof will involve ideas you’ve seen previously in Calculus and this book - namely Taylor Series, solving an individual single first-order, linear, autonomous, homogeneous discrete dynamical system, and limits.

Proof 1.3.1. (*An outline of a proof for Theorem 3.1*)

Without loss of generality, consider

$$X_{n+1} = f(X_n).$$

We need to show why an equilibrium point \bar{X} is called locally asymptotically stable if $|f'(\bar{X})| < 1$. Let’s start!

1. *First we will define a quantity e_n that gives us the distance between our n^{th} iteration, X_n , and the equilibrium point, \bar{X} :*

$$e_n = X_n - \bar{X},$$

and hence

$$e_{n+1} = X_{n+1} - \bar{X},$$

or

$$X_n = e_n + \bar{X} \quad \text{and} \quad X_{n+1} = e_{n+1} + \bar{X}.$$

2. Next we will take $X_{n+1} = f(X_n)$ and subtract \bar{X} from both sides and then substitute e_{n+1} and e_n :

$$\begin{aligned} X_{n+1} &= f(X_n) \\ X_{n+1} - \bar{X} &= f(X_n) - \bar{X} \\ e_{n+1} &= f(\bar{X} + e_n) - \bar{X}. \end{aligned}$$

3. Since we're looking for equilibrium solutions, we're assuming that e_n would be a small quantity, since we'd imagine we are in a near neighborhood of the equilibrium point. This leads to focusing on the $f(\bar{X} + e_n)$ term... we can Taylor expand it about the equilibrium point \bar{X} ! #GoTaylorSeries!

$$e_{n+1} = \left[f(\bar{X}) + f'(\bar{X})e_n + \frac{1}{2!}f''(\bar{X})e_n^2 + \dots \right] - \bar{X}.$$

4. Notice that since \bar{X} is an equilibrium point that $\bar{X} = f(\bar{X})$ and we can substitute that into the above Taylor expansion:

$$e_{n+1} = \left[\bar{X} + f'(\bar{X})e_n + \frac{1}{2!}f''(\bar{X})e_n^2 + \frac{1}{3!}f'''(\bar{X})e_n^3 + \dots \right] - \bar{X},$$

and hence we can cancel out the single \bar{X} terms, which leads us to

$$e_{n+1} = f'(\bar{X})e_n + \frac{1}{2!}f''(\bar{X})e_n^2 + \dots$$

5. If X_n is sufficiently close to the equilibrium, we can assume that e_n is small and hence assume that $e_n^2 \ll e_n$, thus making all e_n^2 , e_n^3 , e_n^4 , and higher-order terms are all negligible. This leaves us to investigate the behavior of the following:

$$e_{n+1} \approx f'(\bar{X})e_n.$$

This is a first-order, linear, autonomous, homogeneous discrete equation! Note that since \bar{X} is a constant, that $f'(\bar{X})$ is some constant value.

6. We can employ all of our linear solution techniques from Section 1.2.1 and see that

$$e_n = C [f'(\bar{X})]^n,$$

where C is a coefficient that would otherwise be determined from an initial value. Here we aren't concerned with the identify of C , but instead, what we care about is what happens to e_n as $n \rightarrow \infty$.

7. As $n \rightarrow \infty$, we fall into one of three cases:

- (a) If $|f'(\bar{X})| < 1 \Rightarrow e_n \rightarrow 0$ as $n \rightarrow \infty$. Since e_n decreases as n increases, the equilibrium point is stable, i.e., as $n \rightarrow \infty$, X_n gets closer and closer to \bar{X} .
 - (b) If $|f'(\bar{X})| > 1 \Rightarrow e_n \rightarrow \infty$ as $n \rightarrow \infty$. Since e_n increases as n increases, the equilibrium point is unstable, i.e., as $n \rightarrow \infty$, X_n gets farther and farther away from \bar{X} .
 - (c) If $|f'(\bar{X})| = 1$, this analysis doesn't help us determine whether the equilibrium point is stable or unstable. We would have to look into another technique - Schwarzian derivatives.
-

1.3.3 Existence and Stability of Periodic Solutions to 1D Discrete Equations

Thus far in our discussion of non-linear equations, we have formally discussed equilibrium points and their stability properties. However, if you remember back to when we introduced the Discrete Logistic Equation, we showed numerous possible solutions behaviors that it could produce, based on different parameter values. These were seen in Figure 1.1. It's okay, please flip back a few pages to remind yourself. I'll wait.

A few of those solutions demonstrated *periodic* behavior. We have not defined what it means to be periodic, but hopefully those plots give an idea - in a nutshell, the solution continually repeats the same values (in the same order) over and over, for forever. For example, a solution with period-2 would imply the solution bounces back and forth between two values forever. A period-4 solution would imply 4 values and in a particular order. Hopefully this gives you an intuitive idea of what it periodic solutions means.

What I will discuss in this section is the *existence* of periodic solutions, or more specifically, a recipe for how we can *try to show the existence* of solutions with any periodicity with want. We actually have all the mathematical machinery we need at this point - all we need to do is one tiny leap (iteration? Ok, bad pun) forward!

To do this, let's consider the following Discrete Logistic Equation

$$X_{n+1} = \tilde{k}X_n \left(1 - \frac{X_n}{\tilde{C}}\right),$$

with $\tilde{C} = 1$. Again, we are getting so much mileage out of the Discrete Logistic Equation because it demonstrates so much different kind of dynamical behavior. The approach we use can generalize for any individual single non-linear discrete equation. We simply are calling on our friend, the Discrete Logistic Equation, to help us out (once again). Before we dive into searching for period-2 solutions, let's quickly find the equilibrium points for this equation.

To find the equilibrium points we assume that $\bar{X} = X_{n+1} = X_n$ and upon substituting get to solve:

$$\bar{X} = \tilde{k}\bar{X} \left(1 - \bar{X}\right),$$

which after subtracting \bar{X} from both sides is equivalent to the following:

$$0 = \tilde{k}\bar{X} \left(1 - \bar{X}\right) - \bar{X}$$

and then

$$0 = \bar{X} \left[\tilde{k}(1 - \bar{X}) - 1 \right].$$

Since we have the product of two things equaling 0, one or more of them must be zero and therefore we get two equilibrium solutions:

$$\bar{X} = 0 \quad \text{AND} \quad \tilde{k}(1 - \bar{X}) - 1 = 0 \quad \Rightarrow \quad \bar{X} = \frac{\tilde{k} - 1}{\tilde{k}}.$$

The reason we computed these equilibrium points was because we'll see them pop up as potential period-2 solutions... we'll get into that later, but for right now make note of those equilibrium points.

Okay, let's finally begin searching for solutions with period-2.

Example 1.3.3. *Finding Solutions with Period-2:*

Using

$$X_{n+1} = \tilde{k}X_n(1 - X_n), \quad (1.10)$$

we will show the existence of solutions with period-2. The idea hopefully we feel very natural. Recall that Figure 1.1 showed that period-2 solutions seem to exist for the Discrete Logistic Equation. However, we note that it appeared that it took a few iterations before the periodic behavior started. Therefore, it seems that there must exist some iteration number, say N , such that if $n \geq N$ that the periodic behavior starts. Therefore, let's assume without loss of generality that n is sufficiently large enough to grant us this.

Since we are trying to show the existence of solutions with period-2, we will begin by starting with two iterations (assuming $n \geq N$ is large enough),

$$X_n : \text{ for some } n \geq N$$

$$X_{n+1} = f(X_n) = \tilde{k}X_n(1 - X_n)$$

$$X_{n+2} = f(X_{n+1}) = \tilde{k}X_{n+1}(1 - X_{n+1}).$$

Now since this is an iterative map, note that we could fully rewrite X_{n+2} fully in terms of X_n , using composition of functions i.e., using $X_{n+1} = \tilde{k}X_n(1 - X_n)$, we can get

$$X_{n+2} = f(X_{n+1}) = f(f(X_n)) = f(\tilde{k}X_n(1 - X_n)),$$

and then via composition, we get that

$$X_{n+2} = \tilde{k}(\tilde{k}X_n(1 - X_n)) \left[1 - \tilde{k}X_n(1 - X_n) \right]. \quad (1.11)$$

Fortunately, the only intuitive leap we make next is that since we trying to find solutions with period-2, what that would mean is that

$$X_{n+2} = X_n \quad \forall n \geq N.$$

Hence we can take (1.11) and equivalently set it equal to $\textcolor{red}{X}_n$!

$$X_{n+2} = \textcolor{red}{X}_n \quad \Rightarrow \quad \textcolor{red}{X}_n = \tilde{k}(\tilde{k}X_n(1 - X_n)) \left[1 - \tilde{k}X_n(1 - X_n) \right]. \quad (1.12)$$

At this junction, (1.12) gives us an equation fully in terms of X_n . Hence we have one equation and one unknown (X_n). If we can find X_n that satisfies (1.12), we have found our period-2 solution(s)! Note that we will see shortly that the equilibrium points will also creep into the solutions as well. The only thing left standing in our way for those period-2 solutions is algebra and, yet another, root-finding problem. Let's get into it:

- First, note that we can pull out one X_n and group the remaining terms to re-write (1.12) equivalently as

$$X_n = X_n \left[r \left[\tilde{k}(1 - X_n) \right] \left[1 - \tilde{k}X_n(1 - X_n) \right] \right].$$

- Next we can perform one clever (it took me a while to figure this out) algebraic manipulation (that will thankfully make our lives much easier). Let's forcibly subtract $\textcolor{magenta}{X}_n$ from both sides:

$$X_n - \textcolor{magenta}{X}_n = X_n \left[r \left[\tilde{k}(1 - X_n) \right] \left[1 - \tilde{k}X_n(1 - X_n) \right] \right] - \textcolor{magenta}{X}_n$$

which is equivalent to

$$0 = X_n \left[r \left[\tilde{k}(1 - X_n) \right] \left[1 - \tilde{k}X_n(1 - X_n) \right] - 1 \right]. \quad (1.13)$$

- Now solving the above equation looks like one big ole mess (remember how I mentioned root-finding problems are hard?). Fortunately, we can use previous information we know to make this situation easier - **remember those equilibrium points we found?**

Recall that equilibrium solutions are defined as

$$\bar{X} = X_n = X_{n+1} = X_{n+2} = X_{n+3} = \dots$$

Now if you notice careful from the above definition of an equilibrium point you will see that $\bar{X} = X_{n+2} = X_n$. So it appears that equilibrium points satisfy our requirement for period-2 solutions. We could keep extending this idea - **equilibrium solutions artificially satisfy any arbitrary periodic behavior**, but they are not unique periodic solutions since the solution never changes.

From this logic, we can deduce that those equilibrium solutions must also satisfy (1.13) above! That is, those equilibrium solutions we found previously, i.e.,

$$\bar{X} = 0 \quad \text{and} \quad \bar{X} = \frac{\tilde{k}-1}{\tilde{k}},$$

are solutions to (1.13). Welp, since they are solutions to that root-finding problem, they are roots to that equation. Therefore, we can pull them out of the equation!

Pulling the those equilibrium solutions out of (1.13) gives us the following equation to solve:

$$0 = X_n \left(\tilde{k}X_n - (\tilde{k}-1) \right) \left[\tilde{k}^2 X_n^2 - \tilde{k}(\tilde{k}+1)X_n + (\tilde{k}+1) \right]. \quad (1.14)$$

Note that I have spared you from a bit of the algebraic manipulation it took to pull out those two factors above; however, if you wish you can show by distribution that (1.14) is equivalent to (1.13).

- Since we know that within (1.14) that $X_n = 0$ and $(\tilde{k}X_n - (\tilde{k}-1))$ give us the equilibrium solutions, that means that our period-2 solutions are given by solutions to

$$\tilde{k}^2 X_n^2 - \tilde{k}(\tilde{k}+1)X_n + (\tilde{k}+1) = 0.$$

Notice that the above equation is actual a quadratic in X_n - we can solve those! #QuadraticFormulaForTheWin

Using the quadratic formula, we arrive at the following:

$$X_n = \frac{\tilde{k}(\tilde{k}+1) \pm \sqrt{\tilde{k}^2(\tilde{k}+1)^2 - 4\tilde{k}^2(\tilde{k}+1)}}{2\tilde{k}^2}.$$

- Simplifying the above quadratic roots, we fine the period-2 solutions can be written as

$$\begin{aligned} X_n &= \frac{\tilde{k}(\tilde{k}+1) \pm \sqrt{\tilde{k}^2(\tilde{k}+1)(\tilde{k}+1-4)}}{2\tilde{k}^2} \\ &= \frac{\tilde{k}(\tilde{k}+1) \pm \tilde{k}\sqrt{\tilde{k}+1}\sqrt{\tilde{k}-3}}{2\tilde{k}^2} \\ &\rightarrow X_n = \frac{(\tilde{k}+1) \pm \sqrt{\tilde{k}+1}\sqrt{\tilde{k}-3}}{2\tilde{k}}. \end{aligned}$$

Therefore if we call our two solutions with period-2 X_{P1} and X_{P2} , we found that they can be explicitly written as:

$$\begin{aligned} X_{P1} &= \frac{(\tilde{k}+1)+\sqrt{\tilde{k}+1}\sqrt{\tilde{k}-3}}{2\tilde{k}} \\ X_{P2} &= \frac{(\tilde{k}+1)-\sqrt{\tilde{k}+1}\sqrt{\tilde{k}-3}}{2\tilde{k}}. \end{aligned} \tag{1.15}$$

To illustrate the existence of these solutions with period-2, check out Figure 1.6. This figure illustrates the existence of these period-2 solutions for 3 different values of \tilde{k} : $\tilde{k} = \{5, 3.5, 3.01\}$. Our formula for the period-2 solutions is able to perfectly predict the periodic solutions - hooray math!

On an interesting side note, notice that all choices of \tilde{k} were greater than 3 and that as $\tilde{k} \rightarrow 3$ the difference between X_{P1} and X_{P2} got smaller and smaller, i.e., it appears that as $\tilde{k} \rightarrow 3^+$ that $X_{P1} - X_{P2} \rightarrow 0$. This can be seen directly from our period-2 solutions in (1.15).

Notice that in both X_{P1} and X_{P2} there is a term involving $\sqrt{\tilde{k}-3}$. Well if $\tilde{k} = 3$, this term vanishes and hence both

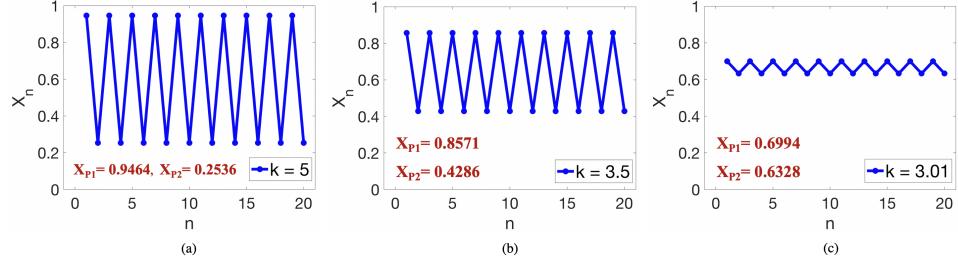


Figure 1.6: A variety of period-2 solutions to the Discrete Logistic Equation for (a) $\tilde{k} = 5$, (b) $\tilde{k} = 3.5$, and (c) $\tilde{k} = 3.01$. Note that the initial X_0 was chosen to be equal to X_{P1} in each case to start the period-2 cycle from the git-go.

$X_{P1} = X_{P2}$. However, this is no longer a period-2 solution... since this implies that $X_{P1} = X_{P2}$ and hence $X_n = X_{n+1} = X_{n+2} = \dots$ forever, this must in fact be an equilibrium solution! Well, what is the value of this equilibrium value when $\tilde{k} = 3$?

$$\tilde{k} = 3 \quad \Rightarrow \quad X_{P1} = X_{P2} = \frac{2}{3}.$$

How does this compare to the equilibrium solution we found for this equation previously? Recall $\bar{X} = \{0, \frac{\tilde{k}-1}{\tilde{k}}\}$. Well if $\tilde{k} = 3$, we actually find that $\bar{X} = \frac{3-1}{3} = \frac{2}{3}$. Therefore, we found that in the limit as $\tilde{k} \rightarrow 3^+$, the period-2 solutions to (1.10) actually collapse and converge to the equilibrium solution of $\bar{X} = \frac{\tilde{k}-1}{\tilde{k}}$. Although that this is only true as \tilde{k} approaches 3 from above, i.e., $\tilde{k} \rightarrow 3^+$. When $\tilde{k} < 3$ our periodic solutions from (1.15) appear to dive into imaginary number land, that is, the complex plane.

Although we've only gone through the exact steps on how to find the explicit period-2 solutions to the Discrete Logistic Equation in its form in (1.10), this methodology could be extended to searching for period-2 solutions for any given single discrete equation, such as $X_{n+1} = f(X_n)$. To find period-2 solutions we will always try to set

$$X_{n+2} = f(X_{n+1}) = f(f(X_n)) = X_n$$

and see if there exist any X_{P1} and X_{P2} that satisfy the equation. The only thing that stands in our way is really difficult (or impossibly hard) algebra

⊕. I also want to clearly state that some discrete equations may not have period-2 solutions, either, so you might run into some mathematical fallacies - this is the equation's way of letting us know that there aren't any.

However, this approach is not limited to finding period-2 solutions - we can generalize this to searching procedure for periodic solutions with any period we are interested in finding. For example, if we were interested in finding solutions with period **3**, we would approach it in the following manner using the idea of composition with our iterative map:

$$X_{n+3} = f(X_{n+2}) = f(f(X_{n+1})) = f(f(f(X_n))) = X_n.$$

One reason we are interested in periodic solutions, might be the following theorem by Li and Yorke. Simply put, if we find the existence of period-3 solutions, our discrete equation is also capable of *chaotic behavior*, that is, *chaos*.

Theorem 3.3

Suppose we are considering the following first-order discrete dynamical equation

$$X_{n+1} = f(X_n).$$

If $f(x)$ is a *continuous* function on a domain $[a, b]$ who images falls within the range $[a, b]$, then if $f(x)$ has a solution with period-3, it $f(x)$ is a chaotic map.

There are a few different ways we could formally define the concept of *chaos*; however, the following definition is appropriate for this section:

Definition 3.5

Suppose we have a first-order discrete dynamical system such as

$$X_{n+1} = f(X_n).$$

A function $f(x)$ is called *chaotic* if for any n , it has a periodic solution with period- n .

We won't discuss the implications of *chaos* or *chaos theory* much in this course. We'll leave that up to our Seminar in Dynamical Systems class #MAT454.

The last thing we will do in this section is (as promised) discuss the *stability* of periodic solutions to first-order discrete equations. Again, we will fall back onto our friend, the Discrete Logistic Equation (Eq. 1.10), and the period-2 solutions we found (Eq. 1.15). Recall that generally the concept of stability lets us know whether solution behavior will remain consistent, e.g., as was the case with equilibrium solutions. Here we think of it analogously - it *perturbed* will period-2 solutions remain period-2 solutions? Meaning, if we vary parameters (or initial conditions) how will that affect our periodic solutions?

In that manner, we will very much follow the steps for determining stability using Theorem 3.1:

1. Find the periodic solutions
2. Take a derivative of the iterative map
3. Evaluate that derivative at the periodic solutions and take its absolute value
4. If that is less than one, it is stable. If it is greater than one, it is unstable.

The only kink in the steps above is in Step 2 - we do not have an explicit iterative map for periodic solutions. For example, for the Discrete Logistic Equation in our example, we only have the original map $X_{n+1} = f(X_n)$ and thus $f(x) = \tilde{k}x(1-x)$. Using $f'(x)$ in the above steps for period-2 solutions wouldn't make sense because the periodic solutions we would evaluate it at X_{P1} and X_{P2} are not equilibrium points. They are periodic solutions with period-2.

Not to fear, we can figure out the proper iterative map for the steps above by using composition of functions once again - hey, we've already done this ☺. Let's explicitly do this for our period-2 solutions that we found before for Eq. 1.10.

Example 1.3.4. Stability of Period-2 Solutions for (1.10).

Recall that the equation we are analyzing was

$$X_{n+1} = f(X_n) = \tilde{k}X_n(1 - X_n)$$

, where the iterative map is seen to be $f(x) = \tilde{k}x(1-x)$, and to whom we found the following period-2 solutions for:

$$X_{P1} = \frac{(\tilde{k} + 1) + \sqrt{\tilde{k} + 1}\sqrt{\tilde{k} - 3}}{2\tilde{k}}$$

$$X_{P2} = \frac{(\tilde{k} + 1) - \sqrt{\tilde{k} + 1}\sqrt{\tilde{k} - 3}}{2\tilde{k}}.$$

Furthermore, recall that to get these solutions we necessitated that $X_{n+2} = X_n$. This lead us to using composition of functions to write the following relation:

$$X_{n+2} = f(X_{n+1}) = f(f(X_n)) = X_n.$$

We can use the above composition of functions to find the periodic iterative function we need to use for stability analysis! That is, we can use the following:

$$X_{n+2} = f(f(X_n))$$

and hence consider the following function

$$F(x) = f(f(x)) = f(\tilde{k}x(1-x)),$$

which finally gives us after one more evaluation

$$F(x) = \tilde{k} \left[\tilde{k}x(1-x) \right] \left(1 - \tilde{k}x(1-x) \right).$$

Now we are ready to determine the Stability of the Period-2 solutions we found! Let's follow the steps outlined above.

1. Find period-2 solutions ✓
2. Take a derivative of $F(x)$: Let's have a stroll down our Calculus memory lane. We had been using the idea of composition a lot... now we want to differentiate... chain rule! Notice our $F(x)$ function was defined using the following composition:

$$F(x) = f(f(x))$$

which we could invoke more general math language and let $u = f(x)$ to see that

$$F(x) = f(u),$$

so F now takes the following form:

$$F = \tilde{k}u(1-u) \quad \text{where } u \text{ is defined as } u = f(x) = \tilde{k}x(1-x).$$

We now get to use the general **chain rule** formula,

$$\begin{aligned} F'(x) &= \frac{dF}{dx} = \frac{df}{du} \cdot \frac{du}{dx} \\ &= \left[\tilde{k}(1-u) - \tilde{k}u \right] \left[\tilde{k}(1-x) - \tilde{x} \right] \\ &= \left[\tilde{k} - 2\tilde{k}u \right] \left[\tilde{k} - 2\tilde{k}x \right] \\ F'(x) &= \tilde{k} \left[1 - 2\tilde{k}x(1-x) \right] \left[1 - 2x \right]. \end{aligned}$$

Finally you can expand out the above $F'(x)$ to get

$$F'(x) = \tilde{k} \left[1 - 2x(1+\tilde{k}) + 6\tilde{k}x^2 - 2\tilde{k}x^3 \right].$$

3. Next we wish to evaluate $F'(x)$ at both period-2 solutions, i.e., $F'(X_{P1})$ and $F'(X_{P2})$.
4. Finally, if either $|F'(X_{P1})| > 1$ or $|F'(X_{P2})| > 1$, it is unstable.

In the case of $\tilde{k} = 3.25$, we find that $|F'(X_{P1})| = |F'(X_{P2})| \approx 0.3571 < 1$ and therefore the solutions with period-2 are stable.

In the case of $\tilde{k} = 5$, we find that $|F'(X_{P1})| = |F'(X_{P2})| \approx 2.2 > 1$ and therefore the solutions with period-2 are unstable.

Again, we see that the **stability properties really depend on the parameter values**.

At this junction, there is **so much** more we could do in the business of analyzing only first-order discrete equations; however, we will turn our sights onto a new endeavor - studying coupled systems of 2 discrete equations... or higher. This will bring us into the world of coupling equations, i.e., exploring how different variables interact with or affect the other(s). You can imagine scenarios in which animals interact with each other - say a predator (fox) and a prey (rabbit). These interactions result in one population affecting another. We're at the point that we can start modeling these dynamical behaviors!

1.4 Systems of Discrete Equations

Up to this point we have only analyzed individual discrete equations - that is we've only considered one iterative map, e.g.,

$$X_{n+1} = f(X_n),$$

where $f(X_n)$ is a function that describes how to iterate to the next value in the sequence, X_{n+1} . If we let X_n be the population of, say, squirrels, and tried to model how the population of squirrels changes over time at TCNJ, at some point we'd have to take into account things such as food availability, environmental changes, and its *predators*, like raptors or hawks. Fortunately, there is no reason why we wouldn't introduce another dependent variable, like Y_n , and let it describe the hawk population or squirrel. What I am getting at here is that we can start coupling discrete dynamical equations together. You'll see that at the end of this section that there is no mathematical limit on how many equations we can couple together; however, there may be an unfortunate limit to how much ecological, experimental, or survey data is available to find/fit parameter values to make the model system realistic.

For now, we will consider systems of 2 coupled discrete equations. All of the math we will develop and discuss in this section can generalize for any number of coupled equations. The only difference is that the nitty-gritty mathematics necessary just gets a bit (*ok a lot*) more tedious, but it will involve the same strategies. To give you the general idea, you can think of a general coupled system of 2 discrete equations as the following:

$$\begin{aligned} X_{n+1} &= f(X_n, Y_n) \quad \text{with initial value, } X_0 \\ Y_{n+1} &= g(X_n, Y_n) \quad \text{with initial value, } Y_0. \end{aligned} \tag{1.16}$$

We will assume that both $f(x, y)$ and $g(x, y)$ have all the nice properties a function can have, like continuity and differentiability. Don't worry, we'll see why in a little bit \odot . The way we can interpret (1.16) is that the next iterations, X_{n+1} and Y_{n+1} , both can depend on each other's previous values, X_n and Y_n , depending on what the form of the functions $f(x, y)$ and $g(x, y)$ take. Now $f(x, y)$ and $g(x, y)$ could be either *linear* or *non-linear* functions of x and y . In fact, just for clarity let's do a brief example just to illustrate those differences.

Example 1.4.1. Are these coupled equations linear or non-linear?

Let's consider a few different examples of coupled systems and determine whether they are linear or non-linear systems. Here we go!

- **Problem 1:** Let

$$\begin{aligned} X_{n+1} &= f(X_n, Y_n) = 2X_n - 7Y_n + 17 \\ Y_{n+1} &= g(X_n, Y_n) = -2Y_n + 13X_n \end{aligned}$$

The above equations are linear because f and g only depend linearly on X_n and Y_n . That is, f and g only have a linear combination of the variables, i.e., each variable only has a coefficient attached to it and they are added together.

- **Problem 2:** Let

$$\begin{aligned} X_{n+1} &= f(X_n, Y_n) = 13X_n \\ Y_{n+1} &= g(X_n, Y_n) = 5X_n + 7Y_n + \frac{1}{Y_n} \end{aligned}$$

This system is non-linear because of the $1/Y_n$ term in the g function. Note that the function f is a linear function, but because one function is non-linear that is all it takes for the system to be deemed non-linear.

- **Problem 3:** Let

$$\begin{aligned} X_{n+1} &= f(X_n, Y_n) = 13X_n + 3\sin(X_n) \\ Y_{n+1} &= g(X_n, Y_n) = -3Y_n + 4\cos(Y_n) \end{aligned}$$

Well, in this case both f and g are non-linear because of the sine and cosine functions, respectively. However, this isn't actually a system of 2 discrete equations... each equation only depends on one variable.

- **Problem 3:** Let

$$\begin{aligned} X_{n+1} &= f(X_n, Y_n) = 2e^{-Y_n} \\ Y_{n+1} &= g(X_n, Y_n) = 13 \ln(X_n) \end{aligned}$$

This is a non-linear system because both the exponential function and natural logarithm that are found in $f(x, y)$ and $g(x, y)$ are non-linear functions; they are composed of solely linear combinations, like $ax + by$ type terms. For this system, it turns out the next iteration X_{n+1} only depends on the previous Y_n value and vice versa. That's fine.

Non-linear systems of discrete equations are incredibly difficult to find exact solutions to by purely analytical techniques. In fact as I write this, I cannot think of one example where a non-linear discrete system has known solutions. I could be wrong, but hopefully that statement alludes to the fact

that when most folks study these sorts of equations they term to strategies to study the behavior of solutions rather than finding the explicit exact solutions themselves. In the previous section for individual first-order discrete dynamical systems, we developed tools for studying properties of solutions to a discrete equation without ever explicitly solving for it - *equilibrium points* and the concept of *stability*. We will use those notions here as well. **Equilibria and stability are the bread and butter of much of the analysis of systems of discrete equations.**

A nice thing about systems of discrete equations is that if we have a computer accessible, we can use it to numerically solve for its solutions in a very straight-forward manner. We could do this in Google Sheets, Microsoft Excel, MATLAB, Mathematica, . . . , the list goes on. However, to compute a solution in this manner, we would have to give explicit values to each parameter in the system and choose particular initial values. Trying to find all the solutions that are possible in this manner could be very tricky - it would involve constantly changing parameters in a systematic fashion (e.g., one at a time) and observing how solutions change. For example, the Discrete Logistic Equation alone demonstrates multiple equilibrium solutions, periodic solutions, and chaos! *Equilibria and stability* offers us a strategy to explore solution behavior across a wide range of parameter values, without explicitly ever defining them - in fact, we can even understand parameter ranges that lead to certain solution behavior.

Lucky for us, we have done most of the work already building up the ideas of equilibrium points and stability, and those ideas extend analogously for systems of discrete equations! Let's dive into that now.

1.4.1 Equilibria and stability in discrete systems

The first I'll do is list the the strategies for finding equilibrium points and their stability for individual (*un-coupled*) discrete equations and compare it to that of discrete coupled systems. As a word of warning there may be some mathematical concepts or ideas presented that you may not recognize - please don't fret, we will go through all of these concepts together.

The main mathematical leap between the single discrete equation and coupled discrete equations scenario is that we are stepping from Calculus A to Calculus C differentiation ideas, but all the ideas and greater strategis are analogous. For example, to determine stability in a discrete system, we will again get to take some kind of derivative, evaluate it at the equilibria, and check a specific inequality, which ultimately tells us whether that equilibria is stable or unstable. Okay, let's get into it.

	Single Discrete Equation	Coupled Discrete Equations
	$X_{n+1} = f(X_n, Y_n)$ $X_0 = \sigma$	$X_{n+1} = f(X_n, Y_n)$ $Y_{n+1} = g(X_n, Y_n)$ $X_0 = \sigma_X, Y_0 = \sigma_Y$
EQUILIBRIA	$\bar{x} = f(\bar{x})$	$\bar{x} = f(\bar{x}, \bar{y})$ $\bar{y} = g(\bar{x}, \bar{y})$
STABILITY	(1) Take derivative, $f'(x)$ (2) Evaluate $f'(x)$ at equilibria, $f'(\bar{x})$ (3) $ f'(\bar{x}) < 1 \rightarrow \text{stable}$	(1) Compute Jacobian Matrix, $J(x, y)$ (matrix of all partial derivatives) (2) Evaluate $J(x, y)$ at equilibria, $J(\bar{x}, \bar{y})$ (3.1) Find eigenvalues of $J(\bar{x}, \bar{y})$, $\{\lambda_j\}$ (3.2a) If all λ_j are real, $\max_j \lambda_j < 1 \rightarrow \text{stable}$ (3.2b) If λ_j are complex, $\max_j \operatorname{Re}(\lambda_j) < 1 \rightarrow \text{stable}$

Table 1.2: Comparing the strategy for equilibria and stability for single discrete equations and systems of discrete equations.

Table 1.2 provides a comparison of finding equilibria the stability thereof

for single discrete equations and systems of discrete equations. Please, if you haven't already, take 30s and read through Table 1.2. Notice how the ideas from the single equation case generalize analogously to the coupled discrete equations (systems) case. In essence, this section starts off by presenting us the only recipe we need for untangling these concepts. The only thing left to go through the recipe and provide context and all necessary background on certain mathematical elements that may not be so familiar to you. We'll go through each portion of the table piece-by-piece.

To do this, let's go through an example together rather than work in generality. We will consider the following 2 coupled discrete equations

$$\begin{aligned} X_{n+1} &= f(X_n, Y_n) = X_n Y_n - 2X_n \\ Y_{n+1} &= g(X_n, Y_n) = Y_n(X_n - 1), \end{aligned} \quad (1.17)$$

with $X_0 = 5$ and $Y_0 = 2$. Before we begin going through the steps of this problem, we will offer some spoilers of what is to come. *At this point there is no reason any of the information I am about to list should be obvious without doing the analysis we're about to do.* I just want to provide a sense of where this problem is going to take us:

- We will see that the initial values do not matter for this equilibria and stability analysis
- We will uncover that there are two equilibria for this system
- We will find that one is stable while the other is unstable
- The system of equations studied in this example (1.17) do not have any inherent parameters. If we were to introduce parameters, depending on parameter values themselves the equilibrium points and their stability properties could change.

Enough background mumbo-jumbo, let's get to work.

Example 1.4.2. Finding equilibria and their stability properties for the discrete system given in (1.17)

1. **Finding equilibria:**

Recall that an equilibrium point means that the solutions cease changing. For two equations that would mean that $X_n = X_{n+1} = X_{n+2} = \dots$ and $Y_n = Y_{n+1} = Y_{n+2} = \dots$ once n is large enough that both quantities have reached their steady-state solution (equilibrium value). (Of

course mathematically what the latter means mathematically is that if that “there exists some N such that once $n \geq N$ then $X_{n+1} = X_n$ and $Y_{n+1} = Y_n$, both equations have reached their equilibria value.”)

Assuming that both solutions have reached their equilibrium, we can define (as we had done previously for single equations) those solutions to be \bar{x} and \bar{y} , i.e.,

$$\bar{x} = X_n = X_{n+1} = \dots \quad \text{and} \quad \bar{y} = Y_n = Y_{n+1} = \dots$$

When we substitute this equilibria assumption into our governing equations, we get that this must mean

$$\bar{x} = f(\bar{x}, \bar{y}) \quad \text{and} \quad \bar{y} = g(\bar{x}, \bar{y}),$$

and specifically for our governing equations given in (1.17), we get the following

$$\begin{aligned} \bar{x} &= \bar{x}\bar{y} - 2\bar{x} \\ \bar{y} &= \bar{y}(\bar{x} - 1). \end{aligned}$$

Glance at those equations for a minute. Realize that there is no longer any n dependence. What we have is actually now an algebraic problem to solve - we have two equations for two unknowns, \bar{x} and \bar{y} . Moreover, notice that this **isn't** an algebraic linear system of equations; it is non-linear. Therefore, our best bet for solving this is solving for one variable in terms of the other and substitute.

Let's first take a look at $\bar{y} = \bar{y}(\bar{x} - 1)$. We need to try to find \bar{x} and/or \bar{y} values that make the right hand side equal the left hand side. In essence we are looking for as many ways that we can make this happen. Let's dive in.

- If we let $\bar{y} = 0$, then we get $0 = 0$. Yay! But wait... what does that mean \bar{x} is? Spoiler: we'll get to use the other equation and the fact that $\bar{y} = 0$ and see what that means \bar{x} must be.
- If we assume $\bar{y} \neq 0$, then we get

$$\begin{aligned} \bar{y} &= \bar{y}(\bar{x} - 1) \\ 1 &= (\bar{x} - 1) \\ \Rightarrow \bar{x} &= 2. \end{aligned}$$

Therefore we know that another equilibrium seems to have $\bar{x} = 2$ but we still don't know the identify of \bar{y} for this track. Again, we'll get to use the other equation, substitute $\bar{x} = 2$ into it, and see what that means \bar{y} must be.

At this junction, we have found that one equilibria seems to have $\bar{y} = 0$ (still unclear what \bar{x} is) while the other has $\bar{x} = 2$ (still unclear what \bar{y} is). Let's use that information and substitute it into the other equation:

- For $\bar{y} = 0$:

Substituting $\bar{y} = 0$ we get

$$\begin{aligned}\bar{x} &= \bar{x}\bar{y} - 2\bar{x} \\ \bar{x} &= \bar{x}(0) - 2\bar{x} \\ \bar{x} &= -2\bar{x} \\ \Rightarrow \bar{x} &= 0.\end{aligned}$$

So, we found that if $\bar{y} = 0$, then \bar{x} must also be equal to 0. Thus, one equilibria point is $(\bar{x}, \bar{y}) = (0, 0)$.

- For $\bar{x} = 2$:

Substituting $\bar{x} = 2$ we get

$$\begin{aligned}\bar{x} &= \bar{x}\bar{y} - 2\bar{x} \\ 2 &= 2\bar{y} - 2(2) \\ 6 &= 2\bar{y} \\ \Rightarrow \bar{y} &= 3.\end{aligned}$$

So, we found that if $\bar{x} = 2$, then $\bar{y} = 3$. Therefore the identify of the other equilibrium point is $(\bar{x}, \bar{y}) = (2, 3)$.

So we have found that for this system of 2 discrete equations given in (1.17), there are two equilibria:

$$(\bar{x}, \bar{y}) \in \{(0, 0), (2, 3)\}.$$

Before proceeding with stability analysis, I want to quickly comment that in this particular example for a system of 2 coupled discrete equations, it was a merely a coincidence that we ended up with 2 equilibria; sometimes there might only be one, sometimes there may be many more, even an infinite amount. The number of equilibria really depend on the relationships between the variables (i.e., the non-linearities) that are present within the system.

Alright, without further ado, let's go explore the stability of these equilibria.

2. Stability properties:

As of now, we have our equilibria. If you take a look at Table 1.2, we still get to do a few more things before declaring stable or unstable. There may also be some unfamiliar terms for you in the procedure that is outlined. As promised, we'll be going through all of these procedures together. There is also a supplemental section dedicated to these mathematical ideas given in Section 1.4.4 if you'd like any further explanation. Here, we will implore a more cavalier just in time approach.

Before diving into the explicit mathematical procedure, remember that for stability of a single equation (like $X_{n+1} = f(X_n)$) that we considered the iterative as a continuous, differentiable function, $f(x)$. We will do the same thing here. The big difference is that for a coupled system that our continuous, differentiable functions are functions of multiple variables. For this example, that would give us the following two functions

$$f(x, y) = xy - 2x \quad \text{and} \quad g(x, y) = y(x - 1).$$

Remember, I mentioned a leap from Calculus A to Calculus C? Fortunately, the only mathematical machinery we will utilize from Calculus C here in the notion of a partial derivative. If you are unfamiliar with this term, I suggest reading Section 1.4.4. However, for now you can think of a partial derivative as obeying the same differentiation rules but that we get to choose which independent variable to differentiate with respect to (here either x or y) and then assume the other is like a constant in the equation. For example, if you take a partial derivative of $f(x, y)$ with respect to y , you only differentiate the function with respect to y and treat all x related stuff as though they are constants, like a parameter, or even a number, say 5.

We get to denote the partial derivative of $f(x, y)$ with respect to y as either f_y , $f_y(x, y)$, or $\frac{\partial f}{\partial y}$, where the latter resembles the same kind of Leibniz differentiation notation in 1D but where the d 's have a little more swagger. Since we are using a lot of subscripts already throughout this course, I'm going to use the latter notation for partial derivative throughout as not to confuse subscripts for partial derivatives. Good notation is the backbone of effective mathematical communication ☺.

Step 1 : Compute the Jacobian Matrix, $J(x, y)$.

The Jacobian matrix is the matrix of all first-order partial derivatives of our functions. In our problem at hand we have two functions: $f(x, y)$ and $g(x, y)$, each depending on two independent

variables x and y . We define the Jacobian matrix, $J(x, y)$ as the following:

$$J(x, y) = \begin{bmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \\ \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} \end{bmatrix}. \quad (1.18)$$

Notice how the Jacobian matrix contains all possibilities of each function and a partial derivative with respect to each variable. Each row of $J(x, y)$ looks at a different function and each column corresponds to a partial derivative with respect to a different variable.

Therefore, to compute $J(x, y)$, we first compute the following partial derivatives:

$$\begin{aligned} \frac{\partial f}{\partial x} &= y - 2 \\ \frac{\partial f}{\partial y} &= x \\ \frac{\partial g}{\partial x} &= y \\ \frac{\partial g}{\partial y} &= x - 1. \end{aligned}$$

and then can substitute them into $J(x, y)$ to get

$$J(x, y) = \begin{bmatrix} y - 2 & x \\ y & x - 1. \end{bmatrix}$$

Hopefully that process wasn't too scary!

Step 2 : Evaluate the Jacobian matrix at the equilibria:

Here we will evaluate the $J(x, y)$ at each equilibrium point. Recall we had two equilibrium points: $(\bar{x}, \bar{y}) \in \{(0, 0), (2, 3)\}$. Hence we will end up with two different matrices that we get to analyze. Substituting those equilibrium points in give us:

For $(0, 0)$:

$$J(0, 0) = \begin{bmatrix} 0 - 2 & 0 \\ 0 & 0 - 1 \end{bmatrix} = \begin{bmatrix} -2 & 0 \\ 0 & -1 \end{bmatrix}$$

For $(2, 3)$:

$$J(2, 3) = \begin{bmatrix} 3 - 2 & 2 \\ 3 & 2 - 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 3 & 1 \end{bmatrix}.$$

Step 3.1 : Find the **eigenvalues** of each Jacobian matrix evaluated that the equilibrium points.

If you are unfamiliar with **eigenvalues**, I highly encourage you to check out Section 1.4.4. For each evaluated Jacobian above, we will solve the following algebraic problem:

$$\det(\lambda I - J(\bar{x}, \bar{y})) = 0.$$

That is, we will subtract each Jacobian from λ times an identity matrix, then take its determinant, and set it equal to zero. Taking the determinant will give us a **characteristic polynomial** to whom once we set it equal to zero, its roots are the **eigenvalues** that we are after.

Let's rock 'n roll.

$$\begin{aligned} - \text{ For } (0, 0): \text{ recall that } J(0, 0) &= \begin{bmatrix} -2 & 0 \\ 0 & -1 \end{bmatrix}. \\ \lambda I - J(0, 0) &= \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} - \begin{bmatrix} -2 & 0 \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} \lambda + 2 & 0 \\ 0 & \lambda + 1 \end{bmatrix}. \end{aligned}$$

Now taking the determinant we get the following:

$$\det(\lambda I - J(0, 0)) = \begin{vmatrix} \lambda + 2 & 0 \\ 0 & \lambda + 1 \end{vmatrix} = (\lambda + 2)(\lambda + 1) - (0)(0).$$

Finally setting that equal to zero, gives us the **characteristic polynomial** for $J(0, 0)$:

$$\det(\lambda I - J(0, 0)) = (\lambda + 2)(\lambda + 1) = 0,$$

where we see that the roots are $\lambda = \{-1, -2\}$. These roots are the **eigenvalues** of $J(0, 0)$.

Let's do the same procedure for the other equilibrium point.

- For $(2, 3)$: recall that $J(2, 3) = \begin{bmatrix} 1 & 2 \\ 3 & 1 \end{bmatrix}$.
- $$\lambda I - J(2, 3) = \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} - \begin{bmatrix} 1 & 2 \\ 3 & 1 \end{bmatrix} = \begin{bmatrix} \lambda - 1 & -2 \\ -3 & \lambda - 1 \end{bmatrix}.$$

Now taking the determinant we get the following:

$$\det(\lambda I - J(2, 3)) = \begin{vmatrix} \lambda - 1 & -2 \\ -3 & \lambda - 1 \end{vmatrix} = (\lambda - 1)(\lambda - 1) - (-2)(-3).$$

Finally setting that equal to zero, gives us the **characteristic polynomial** for $J(2, 3)$:

$$\det(\lambda I - J(2, 3)) = (\lambda - 1)^2 - 6 = \lambda^2 - 2\lambda - 5 = 0.$$

Using our pal, the quadratic formula, we find the roots (**eigenvalues**) in this case are:

$$\lambda = \frac{2 \pm \sqrt{24}}{2} = 1 \pm \sqrt{6}.$$

where we see that the **eigenvalues** of $J(2, 3)$ are $\lambda = \{1 - \sqrt{6}, 1 + \sqrt{6}\}$.

Step 3.2 : Determine stability properties

All eigenvalues found in this example happened to be fully real - none had imaginary parts. Thus we can determine the stability of the equilibria by taking the max eigenvalue in each case in absolute values and comparing it to 1.

- For $(0, 0)$: recall that the eigenvalues of $J(0, 0)$ were $\lambda = \{-1, -2\}$.

So, $\max_j |\lambda_j| = 2 > 1$ so the equilibrium point at $(0, 0)$ is **unstable**.

- For $(2, 3)$: recall that the eigenvalues of $J(2, 3)$ were $\lambda = \{1 - \sqrt{6}, 1 + \sqrt{6}\}$.

In this case, $\max_j |\lambda_j| = 1 + \sqrt{6} > 1$ so the equilibrium point at $(2, 3)$ is also **unstable**.

- Hence both equilibria in this problem turn out to be unstable. Again, this is just a coincidence that this happened.

A few takeaways from Example 1.4.2 are the following:

1. A non-linear system of algebraic equations probably needs to be solved when determining all the equilibrium points, (\bar{x}, \bar{y}) .
2. The number of equilibrium points does not have to equal the number of equations in the discrete system. It was a coincidence in this problem that there were 2 discrete coupled equations and 2 equilibria were found. Any number of equilibria could exist even in a single discrete equation! It really depends on the non-linearities present in the governing equations.
3. Consistency is necessary for filling in the Jacobian matrix, J . Each column must only correspond to partial derivatives with respect to a specific variable, e.g., column 1 is x , column 2 is y , etc., and that each row corresponds to a specific function.
4. Finding the eigenvalues involves a few matrix processes (matrix addition/subtraction and taking determinants). Ultimately, when taking the determinant you will arrive at a **characteristic polynomial** whose roots are the **eigenvalues**.
5. Thus, to get the eigenvalues you need to perform a root-finding problem on a polynomial.

Similar to the single discrete equation case, we have a theorem that supports the procedure we just did for finding stability properties of an equilibrium point for a system of coupled discrete equations. In fact, we can generalize this procedure from just two coupled equations to any number we wish, say M .

Theorem 4.1

Consider a coupled set of M discrete dynamical equations such as

$$\begin{aligned} X_{n+1}^1 &= f_1(X_n^1, X_n^2, X_n^3, \dots, X_n^M) \\ X_{n+1}^2 &= f_2(X_n^1, X_n^2, X_n^3, \dots, X_n^M) \\ X_{n+1}^3 &= f_3(X_n^1, X_n^2, X_n^3, \dots, X_n^M) \\ &\vdots \\ X_{n+1}^M &= f_M(X_n^1, X_n^2, X_n^3, \dots, X_n^M) \end{aligned} \tag{1.19}$$

with appropriate initial values and equilibrium point $(\bar{x}^1, \bar{x}^2, \dots, \bar{x}^M)$. Assume that f_1, f_2, \dots, f_M are all continuous on an open interval I containing the equilibrium point. Then let $\{\lambda_j\}$ be the eigenvalues of the Jacobian matrix associated with (1.17).

If $\max_j |\operatorname{Re}(\lambda_j)| < 1$, then the equilibrium point is said to be a **stable**.
If $\max_j |\lambda_j| > 1$ **unstable**.

Note the following:

- If $\lambda_j \in \mathbb{R}$ for all j , then $\operatorname{Re}(\lambda_j) = \lambda_j$ anyways.
- This theorem just guarantees an equilibrium point as **stable** or **unstable**; however, there are different flavors of **stable/unstable** equilibria in the world of couple discrete equations. We will dive in that shortly!

Lastly, let's make a mathematical connection with the process for finding equilibrium points and determining their stability of a single discrete equation (say $X_{n+1} = f(X_n)$), as in Section 1.3.2. By taking the derivative of $f(x)$ and evaluating it at the equilibrium point \bar{x} , it actually gives us the eigenvalue for the problem, it is $f'(\bar{x})$. How is that the same? Well, the connection lies in the extension of the proof for Theorem 3.1 to that of Theorem 4.1 above. Let's go through that proof now.

But first, let's *spoil it*. Recall that the proof of Theorem 3.1 used the following mathematical tools to help us out: Taylor Series and computing exact solutions to single discrete equations. Guess what we're going to do? Welp, Taylor Series... but in higher dimensions and an iterative process... except this time those iterations will involve a matrix, and more specifically, repeatedly performing matrix-vector multiplying by continuing multiply that matrix over and over. Note that this proof is a bit heavy on some concepts from Calculus C and Linear Algebra (#MAT205). **What I hope you get out of this proof is that the proof of Theorem 3.1 for single discrete equations generalizes analogously to systems of coupled discrete equations.** The difference is that once we extend into higher dimensions, we now get to talk about *partial derivatives* instead of just derivatives, Jacobian matrices instead of a first derivative of a single variable function, and eigenvalues of said Jacobian matrices instead of simply evaluating single variable function derivatives.

In fact, it might help to quickly revisit Proof 1.3.1 before jumping in here. Seriously, take a minute or two to remind yourself what we did together. I'll wait as long as you'd like.

Proof 1.4.1. (An outline of a proof for Theorem 4.1)

Consider the system of M -coupled discrete equations from Theorem 4.1:

$$\begin{aligned} X_{n+1}^1 &= f_1(X_n^1, X_n^2, X_n^3, \dots, X_n^M) \\ X_{n+1}^2 &= f_2(X_n^1, X_n^2, X_n^3, \dots, X_n^M) \\ X_{n+1}^3 &= f_3(X_n^1, X_n^2, X_n^3, \dots, X_n^M) \\ &\vdots \\ X_{n+1}^M &= f_M(X_n^1, X_n^2, X_n^3, \dots, X_n^M) \end{aligned} \quad (1.20)$$

We will now introduce vector notation for this system. Let

$$\vec{\mathcal{X}}_{n+1} = \begin{pmatrix} X_{n+1}^1 \\ X_{n+1}^2 \\ \vdots \\ X_{n+1}^M \end{pmatrix} \quad \text{and} \quad \vec{G}(\vec{\mathcal{X}}_n) = \begin{pmatrix} f_1(X_n^1, X_n^2, X_n^3, \dots, X_n^M) \\ f_2(X_n^1, X_n^2, X_n^3, \dots, X_n^M) \\ \vdots \\ f_M(X_n^1, X_n^2, X_n^3, \dots, X_n^M) \end{pmatrix},$$

and hence we can write (1.20) as the following compact notation:

$$\vec{\mathcal{X}}_{n+1} = \vec{G}(\vec{\mathcal{X}}_n).$$

We can then note that we can represent the Jacobian matrix, \mathbf{J} as the gradient of the vector function $\vec{G}(\vec{\mathcal{X}})$, that is,

$$\mathbf{J} = \nabla \vec{G}(\vec{\mathcal{X}}),$$

where ∇ is the gradient operator, defined as the vector of all partial derivative operators, $\nabla = \left(\frac{\partial}{\partial X_{n+1}^1}, \frac{\partial}{\partial X_{n+1}^2}, \dots, \frac{\partial}{\partial X_{n+1}^M} \right)$. As an example, for the system of 2 coupled equations described in Example 1.4.2, the gradient would be $\nabla = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y} \right)$.

Furthermore, let's introduce \vec{z} represent the vector describing an equilibrium point of (1.20):

$$\vec{z} = \begin{pmatrix} \bar{x}^1 \\ \bar{x}^2 \\ \vdots \\ \bar{x}^M \end{pmatrix}.$$

Now we have all the notation we need describing the items we have at our disposal already to begin the proof (i.e., the assumptions made for theorem - the “if” part). Phew.

In parallel with the proof involving stability of an equilibrium for a single discrete equation, to start off the proof here, we will define a vector quantity \vec{e}_n ,

$$\vec{e}_n = \vec{\mathcal{X}}_n - \vec{\bar{z}} = \begin{pmatrix} X_{n+1}^1 - \bar{x}^1 \\ X_{n+1}^2 - \bar{x}^2 \\ \vdots \\ X_{n+1}^M - \bar{x}^M \end{pmatrix}.$$

Therefore, this also gives us that

$$\vec{\mathcal{X}}_n = \vec{\bar{z}} + \vec{e}_n.$$

Now let’s begin the substitution and Taylor Series game, starting with \vec{e}_{n+1} and recalling that $\vec{\mathcal{X}}_{n+1} = \vec{G}(\vec{\mathcal{X}}_n)$:

$$\begin{aligned} \vec{e}_{n+1} &= \vec{\mathcal{X}}_{n+1} - \vec{\bar{z}} \\ &= \vec{G}(\vec{\mathcal{X}}_n) - \vec{\bar{z}} \\ &= \vec{G}(\vec{\bar{z}} + \vec{e}_n) - \vec{\bar{z}}. \end{aligned}$$

Welp, we’ve arrived at my favorite point in this proof where we get to do a multi-variable Taylor expansion of $\vec{G}(\vec{\bar{z}} + \vec{e}_n)$. We assume that each element of \vec{e}_n is sufficiently small after enough iterations, i.e., n is large enough, so that we’re in a neighboring of the our equilibrium $\vec{\bar{z}}$. The logic here is analogous to the single equation case. Actually performing the Taylor expansion is where things get a bit spicy. I’ll do my best to describe the journey.

$$\begin{aligned} \vec{e}_{n+1} &= \vec{G}(\vec{\bar{z}} + \vec{e}_n) - \vec{\bar{z}} \\ &= \left[\vec{G}(\vec{\bar{z}}) + \nabla \vec{G}(\vec{\bar{z}}) \cdot \vec{e}_n + \frac{1}{2} \vec{e}_n^T \nabla (\nabla \vec{G}(\vec{\bar{z}})) \cdot \vec{e}_n + \dots \right] - \vec{\bar{z}}. \end{aligned}$$

Now since $\vec{\bar{z}}$ is defined to be an equilibrium point of (1.20), we know that $\vec{\bar{z}} = \vec{G}(\vec{\bar{z}})$ and from how the Jacobian can be written in gradient form we have that $\mathbf{J} = \nabla \vec{G}(\vec{\bar{z}})$. Therefore we arrive at the following

$$\begin{aligned}
e_{n+1} &= \left[\vec{G}(\vec{\bar{z}}) + \nabla \vec{G}(\vec{\bar{z}}) \cdot \vec{e}_n + \frac{1}{2} \vec{e}_n^T \nabla (\nabla \vec{G}(\vec{\bar{z}})) \cdot \vec{e}_n + \dots \right] - \vec{\bar{z}} \\
&= \left[\vec{\bar{z}} + \mathbf{J} \cdot \vec{e}_n + \frac{1}{2} \vec{e}_n^T \nabla (\nabla \vec{G}(\vec{\bar{z}})) + \dots \right] - \vec{\bar{z}} \\
&= \mathbf{J} \cdot \vec{e}_n + \frac{1}{2} \vec{e}_n^T \nabla (\nabla \vec{G}(\vec{\bar{z}})) \cdot \vec{e}_n + \dots
\end{aligned}$$

As previously mentioned, (and as in our argument for the single discrete equation case), we assume that all entries in \vec{e}_n are sufficiently small enough. Thus, we are able to get rid of higher-order terms, i.e., all terms involving products of entries of \vec{e}_n . In this case, this means throwing away the second term and higher listed above, i.e., $\frac{1}{2} \vec{e}_n^T \nabla (\nabla \vec{G}(\vec{\bar{z}})) \cdot \vec{e}_n$ and any terms beyond.

Also for fun, note that the notation $\nabla (\nabla \vec{G}(\vec{\bar{z}}))$ denotes computing what is known as a Hessian, \mathbf{H} . Generally, mathematicians don't enjoy computing such things, even numerically with a computer as the process isn't efficient and is a bit costly to do mid-algorithm, but on the flip side they can actually provide quite a bit of insight into some problems. Anyways, throwing away that Hessian term and all others beyond it, we now have

$$\vec{e}_{n+1} \approx \mathbf{J} \cdot \vec{e}_n.$$

At this point we are virtually finished with the proof, modulo actually iterating the above discrete equation. Notice that as we would iteration the above relation, one would get

$$\vec{e}_{n+5} \approx \mathbf{J}(\mathbf{J}(\mathbf{J}(\mathbf{J}(\mathbf{J}(\mathbf{J} \cdot \vec{e}_n)))),$$

or equivalently

$$\vec{e}_{n+5} \approx \mathbf{J}^5 \cdot \vec{e}_n.$$

More generally,

$$\vec{e}_{n+c} \approx \mathbf{J}^c \cdot \vec{e}_n.$$

Wait, we haven't made a connection to the eigenvalues of \mathbf{J} yet... welp, here we go! If we were to take $c \rightarrow \infty$ an interesting result would arise, which you may have seen in #MAT205, or if not there, then #MAT331: Numerical Analysis or #MAT370: Linear Algebra II. If you haven't, now is the day you get to learn about it - **matrix decompositions**.

The main idea of a **matrix decomposition** is that we can rewrite our matrix at hand, i.e., \mathbf{J} , in terms of other matrices. For our purposes here, we will assume that our matrix \mathbf{J} has distinct eigenvalues, where each has

its own linearly independent eigenvectors. Recall that the mathematical definition of an eigenvalue (λ) and eigenvector (\vec{v}) of a matrix J satisfy:

$$J\lambda = J\vec{v}.$$

In the case of a full set of linearly independent eigenvectors, we can perform what is known as **matrix diagonalization**, where we define the following matrices:

$$\Lambda = \begin{bmatrix} \lambda_1 & & & \\ & \ddots & & \\ & & \lambda_M & \end{bmatrix} \quad \text{and} \quad V = \begin{bmatrix} | & | & | & | \\ \vec{v}_1 & \vec{v}_2 & \dots & \vec{v}_M \\ | & | & | & | \end{bmatrix},$$

where Λ is a diagonal matrix with the eigenvalues along the main diagonal and V is a matrix composed of the eigenvectors (in column form). We can then rewrite J ("decomposed J ") in the following manner:

$$J = V\Lambda V^{-1}.$$

Why is that useful? Let's revisit that iterative process from our proof:

$$\vec{e}_{n+1} \approx J \cdot \vec{e}_n.$$

Substituting our matrix diagonalization for J into that iterative equation, we get

$$\vec{e}_{n+1} \approx V\Lambda V^{-1} \cdot \vec{e}_n.$$

Now if we continue this iterative process for, say, 5 iterations, as we did before, we arrive at the following

$$\vec{e}_{n+5} \approx (V\Lambda V^{-1})^5 \cdot \vec{e}_n.$$

If we take a quick look at that scary looking term $(V\Lambda V^{-1})^5$, we can actually convince ourselves that it reduces rather nicely...

$$\begin{aligned} (V\Lambda V^{-1})^5 &= (V\Lambda V^{-1})(V\Lambda V^{-1})(V\Lambda V^{-1})(V\Lambda V^{-1})(V\Lambda V^{-1}), \\ &= V\Lambda V^{-1}V\Lambda V^{-1}V\Lambda V^{-1}V\Lambda V^{-1}V\Lambda V^{-1}, \end{aligned}$$

which when using what it means to be an inverse of a matrix, i.e., $V^{-1}V = I$, we arrive at

$$(V\Lambda V^{-1})^5 = V\Lambda^5 V^{-1}.$$

Hence if we iterated, say, c times, we would still get

$$(V\Lambda V^{-1})^c = V\Lambda^c V^{-1}.$$

Therefore, in the limit as $c \rightarrow \infty$, we see the following happens:

$$\lim_{c \rightarrow \infty} e_{n+c} = V \Lambda^c V^{-1} \vec{e}_n = V \begin{bmatrix} \lambda_1^c & & \\ & \ddots & \\ & & \lambda_M^c \end{bmatrix} V^{-1} \vec{e}_n.$$

If the **dominant eigenvalue**, i.e., the largest eigenvalue in magnitude, is less than one, i.e., $|\operatorname{Re}(\lambda_j)| < 1$ for all j , then this limit would equal zero, i.e., $\lim_{c \rightarrow \infty} e_{n+c} = 0$. This implies **stability of the equilibrium**.

However, if the **dominant eigenvalue** was greater than one in magnitude, i.e., there exists a λ_j such that $|\operatorname{Re}(\lambda_j)| > 1$ for at least one j , then $\lim_{c \rightarrow \infty} e_{n+c} \neq 0$ and can get arbitrarily large. This implies that the equilibrium is **unstable**.

Oof. What a proof!

So far we have found equilibrium points for systems of coupled discrete equations and have been able to deem them either *stable* or *unstable* based on the Jacobian of the system's eigenvalues. Upon that analysis, we've seen that since we have multiple eigenvalues, some might be real, some could be complex, or some might even be fully imaginary. However, our definition and strategy for stability gave us only two possible avenues for classification - stable or unstable. As you might imagine, depending on whether eigenvalues are fully real, complex, or fully imaginary helps us further classify a equilibrium point, and beyond that, the dynamics we might expect in a neighborhood about said point. We're going to explore this idea in the following section - how to further classify equilibria.

1.4.2 Further classification of equilibrium points

We are going to start this section by showing two different solutions to the same system of coupled discrete equations. The solutions differ due differences in the parameters for each case, in particular the parameter d . The discrete system we will use is the following:

$$\begin{aligned} X_{n+1} &= X_n(k - qX_n - drY_n) \\ Y_{n+1} &= dY_n(1 + rX_n), \end{aligned} \tag{1.21}$$

with $X_0 = 50$ and $Y_0 = 5$. Note that this discrete system is an example of a *predator-prey* system. See if you can put it back into it's more familiar form. The left side of Figure 1.7 illustrates different solution behavior for different values of d . Check it out - take a minute. What I want you to recognize is that we see again changing, even one, parameter can significantly change a system's resulting dynamics **AND** that we can see these differences through the eigenvalues computed in stability analysis.

The right side of Figure 1.7 are called *phase portraits* of the solution trajectories. In phase portraits, the solution values X_n and Y_n are plotted against each other. Using some lingo from Calculus B, we are essentially thinking of the quantities X_n and Y_n are a parameterized set of values, where the parameter that changes is n . Therefore using those ideas, changing n moves us around the parametric curve described by (X_n, Y_n) . You can see this by the arrows indicating the direction along the parametric curve, err phase portrait. Phase portraits are an invaluable tool for studying dynamical systems.

Hey, where is the connection with those eigenvalues of the Jacobian of our system that we worked so hard to find?! Glad you asked. It turns out that if we knew the eigenvalues, we would be able to predict the shape (and direction) of the trajectories in the phase portraits. Similarly, knowing what the phase portraits look like, we have a pretty good idea what the eigenvalues are like for those particular solutions.

Okay, but if we can get information about the eigenvalues from the phase portraits, why-oh-why, did we spend time computing eigenvalues!? Also, a fantastic question. Remember to get solutions such as those in Figure 1.7a, c, and e, we had to declare specific values for the parameters of our system. Thus, we only know what the solution does for that very particular set of parameters!

By not giving parameters specific values and computing the eigenvalues

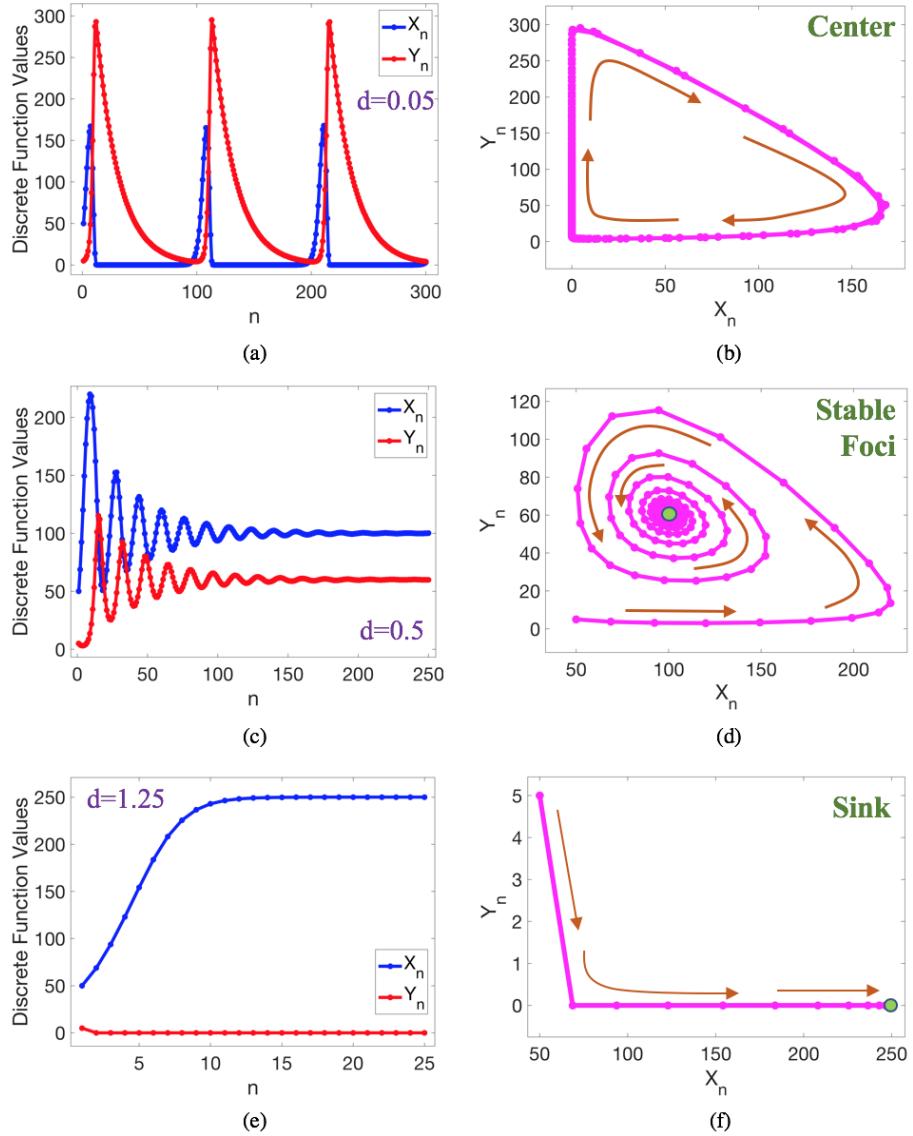


Figure 1.7: Numerical solutions and their associated phase portraits to (1.21) for (a,b) $d = 0.05$, (c,d) $d = 0.5$, and (e,f) $d = 1.25$ with $k = 1.5$, $q = 0.002$, and $r = 0.01$. Note that (a),(c), and (e) provide numerical solutions on the left side, while (b),(d), and (f) on the right side are their associated phase portraits.

of the system in terms of those “undeclared” parameters, we can observe where solutions significantly change behavior as parameter values change. For example, we see that by changing d for the system given in (1.21) that the solutions can substantially change. By finding our eigenvalues in terms of the parameters, we may have been able to predict what ranges of each parameter and what combinations thereof would lead to particular solution behavior \odot .

I’m getting a little ahead of myself. Sorry about that; we are just on the brink of such gnarly mathematics. For now, I hope you get the picture that our solution behavior about equilibria is all about the eigenvalues of the Jacobian. Changing parameters can change that behavior. What a world. Knowing the eigenvalues we can further classify the equilibrium, beyond solely either *stable* or *unstable*. Figure 1.8 lists these further classification beyond stable or unstable.

Some general ideas to notice for coupled systems of two discrete equations:

- Stable equilibria are always associated with eigenvalues in which all have real part is **less than one** in magnitude.
- Whenever a real part of an eigenvalue is **greater than one** in absolute value for at least one eigenvalue, the equilibrium will be unstable.
- Whenever eigenvalues are complex with non-zero real and imaginary parts, there will be ‘spiraling’ behavior in the phase plane.

To close this brief foray into further classification of equilibrium points, I want to remind us all that all of these ideas about stability and how they help us classify only work in a *neighborhood* about the equilibrium. In practical terms this means that if an equilibria is classified as a **sink** for particular parameter values, the system’s solutions do not necessarily have to approach that equilibrium. At first this might seem strange or incorrect because, hey, we showed that it should be a sink! However, the subtle issue at hand is that the system’s solutions *need to be close enough* to that equilibrium point for them to actually approach it.

For example, if the initial values drove the system far away from that equilibrium point, there is not reason the solutions should circle back or approach it. In all of our stability analysis we have only proven properties of *local stability* and have not made any claims regarding the *global* properties of solutions. If you remember, we used our pal from Calculus B, the Taylor Series, to help us prove *local stability* properties. We had to include phrases

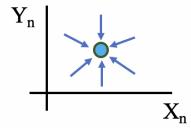
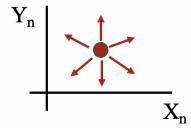
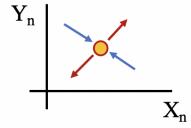
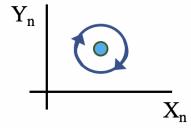
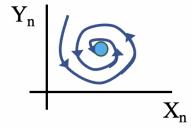
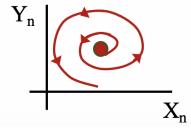
Case	Classification	Name	Phase Plane Depiction
$ \lambda_1 < 1, \lambda_2 < 1$	Asymptotically Stable	Sink	
$ \lambda_1 > 1, \lambda_2 > 1$	Unstable	Source	
$ \lambda_1 < 1, \lambda_2 > 1$	Unstable	Saddle-point	
$\lambda_1 = a + bi, \lambda_2 = a - bi$ ($Re(\lambda_1) = Re(\lambda_2) = a = 0$)	Stable (neutrally stable) (not asymptotically stable)	Center	
$\lambda_1 = a + bi, \lambda_2 = a - bi$ ($0 < Re(\lambda_1) = Re(\lambda_2) = a < 1$)	Stable	Stable foci	
$\lambda_1 = a + bi, \lambda_2 = a - bi$ ($ Re(\lambda_1) = Re(\lambda_2) = a > 1$)	Unstable	Unstable foci	

Figure 1.8: Classifying equilibria further than stable or unstable depending on the eigenvalues.

in our proofs and definitions like *close enough* or *in a neighborhood of* to be clear (and honest) that we were only thinking of small local regions surrounding equilibrium points. Taylor Series are wonderful approximations if we're *close enough* to something... unfortunately they aren't sufficient to help us prove *global* properties.

1.4.3 Linear systems of coupled discrete equations

It might seem a bit backward to be talking about **linear** systems of coupled discrete equations after we've spent time discussing **non-linear** systems, but I have a few reasons for doing so - we're going to need some information about *eigenvalues* to help us analyze these things. If we consider a general *linear* autonomous homogeneous system of 3 coupled discrete equations, it would take the following form:

$$\begin{aligned} X_{n+1} &= a_1 X_n + a_2 Y_n + a_3 Z_n \\ Y_{n+1} &= b_1 X_n + b_2 Y_n + b_3 Z_n \\ Z_{n+1} &= c_1 X_n + c_2 Y_n + c_3 Z_n \end{aligned}$$

where $\{a_j\}_{j=1}^3$, $\{b_j\}_{j=1}^3$, and $\{c_j\}_{j=1}^3$ are all constants. We also assume initial values X_0, Y_0, Z_0 . To refresh our memory the above set of equations is linear because there are no X_n^2 or $X_n Y_n$ type terms; the system is comprised entirely by linear combinations of the variables.

Due to this form of linear combinations (maybe a lightbulb went off in your head?), we can actually rewrite this system using matrix notation. We can define what we call a **state vector** and system matrix as the following:

$$\vec{x}_n = \begin{pmatrix} X_n \\ Y_n \\ Z_n \end{pmatrix} \quad \text{and} \quad A = \begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix}.$$

Having done so, we can equivalently write our system as an iterative matrix equation:

$$\vec{x}_{n+1} = A \vec{x}_n. \quad (1.22)$$

Maybe this helps give you a better intuition on what this system actually does each iteration → we keep applying the same matrix (linear operator) over and over! This might feel very similar to the concept of a *Markov chain*, if you're familiar with that term (don't worry if you're not). They are similar in that the process of continually multiplying a matrix over and over is the same; however, a Markov chain's matrix has a very specific property that ours here does not generally have - the elements of each row *must* add to 1 in a Markov chain process.

Beyond writing this as a matrix equation, we can actually take this another step forward. The matrix equation (1.22) bears an uncanny resemblance to a first-order linear, homogeneous, autonomous discrete equation from Section 1.2.1. Similar to those first-order linear equations we've already seen, we can actually solve this with a completely analogous approach.

Writing our initial values in vector notation, i.e.,

$$\vec{x}_0 = \begin{pmatrix} X_0 \\ Y_0 \\ Z_0 \end{pmatrix},$$

We see we can write a solution to the linear system defined in (1.22) as

$$\vec{x}_n = A^n \vec{x}_0,$$

The notation A^n means multiplying A onto itself n times, e.g., $A^3 = AAA$. Also, if you don't see where the solution came from here, I jacked the same ideas we used back in Section 1.2.1 for a single linear autonomous homogeneous discrete equation ☺.

If you remember for an equation of the form

$$X_{n+1} = \lambda X_n,$$

where $\lambda \in \mathbb{R}$ with initial value X_0 . There were only a few possibilities:

1. If $|\lambda| > 1$, then $|X_n| \rightarrow \infty$ as $n \rightarrow \infty$,
2. If $|\lambda| < 1$, then $|X_n| \rightarrow 0$ as $n \rightarrow \infty$, or
3. If $|\lambda| = 1$, then $|X_n|$ is constant as $n \rightarrow \infty$ (note that this case is *boring*)...the equation would simply be $X_{n+1} = \pm X_n$.

For a system, such as (1.22), these possibilities actually directly extend. However, the distinction in each case will be determined by what the **dominant eigenvalue** of the matrix is. Now, we have already seen eigenvalues pop up in our discussion of stability of equilibrium where we were concerned with the eigenvalues of the Jacobian matrix, J . However, here, we don't need the eigenvalues of a Jacobian. We aren't discussing equilibria nor their associated stability. *We are looking for solutions to (1.22)*.

In fact, if you look at (1.22), there is no indication what space each vector or matrix belongs to (unless you read back a tad further than that equation). We can extend this linear system idea from 3 to N dimensions without breaking a sweat here. Let's do to this both flex our math muscles as well as see that there isn't anything particularly special about a 3-dimensional vs. N -dimension linear system. We we first introduce the following notation:

X_n^j : the population of Stage j at iteration n

Caution: the superscripts in this section do not mean exponents. They are to differentiate different stages (age classifications) within a population.

Now we will consider an N dimension system of linear, homogeneous, autonomous coupled discrete equations:

$$\begin{aligned} X_{n+1}^1 &= a_{11}X_n^1 + a_{12}X_n^2 + a_{13}X_n^3 + \dots + a_{1N}X_n^N \\ X_{n+1}^2 &= a_{21}X_n^1 + a_{22}X_n^2 + a_{23}X_n^3 + \dots + a_{2N}X_n^N \\ &\vdots \\ X_{n+1}^N &= a_{N1}X_n^1 + a_{N2}X_n^2 + a_{N3}X_n^3 + \dots + a_{NN}X_n^N \end{aligned}$$

where $a_{ij} \in \mathbb{R}$ for $i, j = 1, 2, \dots, N$ with initial values $\{X_0^j\}_{j=1}^N$. We can rewrite this general N -dimensional system into a compact matrix equation using a similar idea as before. Let

$$\vec{x}_n = \begin{pmatrix} X_n^1 \\ X_n^2 \\ \vdots \\ X_n^N \end{pmatrix} \quad \text{and} \quad A \in \mathbb{R}^{N \times N}.$$

Then we can write this matrix system equivalent as

$$\vec{x}_{n+1} = A\vec{x}_n, \tag{1.23}$$

with initial values $\vec{x}_0 = (X_0^1, X_0^2, \dots, X_0^N)$.

Here's where a little knowledge of Linear Algebra helps us out. If $A \in \mathbb{R}^{N \times N}$ has N -distinct eigenvalues (or as long as the dimension of A 's *eigenspace* is N), we can write any vector \mathbb{R}^N as a linear combination of A 's eigenvectors. Rad!

Recall that if λ is an eigenvalue of A and \vec{v} is its associated eigenvector, that λ and \vec{v} satisfy

$$A\vec{v} = \lambda\vec{v}.$$

Therefore, by our assumptions above, if we have N -eigenvalues, $\{\lambda_j\}_{j=1}^N$, each λ_j must be associated with a particular eigenvector, \vec{v}_j , and thus we must have N -eigenvectors, $\{v_j\}_{j=1}^N$, too.

If we can write any vector in \mathbb{R}^N as a linear combination of A 's eigenvectors, let's choose to write out \vec{x}_0 , the vector of all initial values for our system. So, if we can do this, there must exist N -coefficients $\{c_j\}_{j=1}^N$, with $c_j \in \mathbb{R}$ for $j = 1, 2, \dots, N$, such that

$$\vec{x}_0 = c_1\vec{v}_1 + c_2\vec{v}_2 + c_3\vec{v}_3 + \dots + c_N\vec{v}_N. \tag{1.24}$$

Why on Earth would we want to do that?! Thank you for asking. We can apply our iteration from (1.23) directly to this! Check it out:

$$\begin{aligned}\vec{x}_1 &= A\vec{x}_0 \\ &= A \left[c_1 \vec{v}_1 + c_2 \vec{v}_2 + c_3 \vec{v}_3 + \dots + c_N \vec{v}_N \right] \\ &= c_1 A \vec{v}_1 + c_2 A \vec{v}_2 + c_3 A \vec{v}_3 + \dots + c_N A \vec{v}_N.\end{aligned}$$

Okay...so? Well, we can use what it means to be an eigenvector now! That is, we can use the fact that for each $A\vec{v}_j$ product, we know that it is equal to $A\vec{v}_j = \lambda_j \vec{v}_j$. Substituting that into the above gives us

$$\begin{aligned}\vec{x}_1 &= c_1 A \vec{v}_1 + c_2 A \vec{v}_2 + c_3 A \vec{v}_3 + \dots + c_N A \vec{v}_N \\ \vec{x}_1 &= c_1 \lambda_1 \vec{v}_1 + c_2 \lambda_2 \vec{v}_2 + c_3 \lambda_3 \vec{v}_3 + \dots + c_N \lambda_N \vec{v}_N.\end{aligned}$$

I'm still not sure where you're going with this.... Hey, that's okay. Let's try iterating one more time and notice is you pick up on any patterns that arise.

$$\begin{aligned}\vec{x}_2 &= A\vec{x}_1 \\ &= A(A\vec{x}_0) \\ &= A \left[c_1 \lambda_1 \vec{v}_1 + c_2 \lambda_2 \vec{v}_2 + c_3 \lambda_3 \vec{v}_3 + \dots + c_N \lambda_N \vec{v}_N \right] \\ &= c_1 \lambda_1 A \vec{v}_1 + c_2 \lambda_2 A \vec{v}_2 + c_3 \lambda_3 A \vec{v}_3 + \dots + c_N \lambda_N A \vec{v}_N.\end{aligned}$$

We were able to move A right by both c_j and λ_j in each term in the above because they are only *scalars* (constants, not vectors nor matrices). We're at the point where we can use the same eigenvalue-eigenvector statement as before ($A\vec{v}_j = \lambda_j \vec{v}_j$) and trade all of the A 's for eigenvalues! Doing this we get

$$\begin{aligned}\vec{x}_2 &= c_1 \lambda_1 A \vec{v}_1 + c_2 \lambda_2 A \vec{v}_2 + c_3 \lambda_3 A \vec{v}_3 + \dots + c_N \lambda_N A \vec{v}_N \\ &= c_1 \lambda_1 \lambda_1 \vec{v}_1 + c_2 \lambda_2 \lambda_2 \vec{v}_2 + c_3 \lambda_3 \lambda_3 \vec{v}_3 + \dots + c_N \lambda_N \lambda_N \vec{v}_N \\ \vec{x}_2 &= c_1 \lambda_1^2 \vec{v}_1 + c_2 \lambda_2^2 \vec{v}_2 + c_3 \lambda_3^2 \vec{v}_3 + \dots + c_N \lambda_N^2 \vec{v}_N\end{aligned}$$

We could continue iterating in this same way going forward; however, a pattern will emerge:

$$\vec{x}_{\textcolor{red}{n}} = c_1 \lambda_1^{\textcolor{red}{n}} \vec{v}_1 + c_2 \lambda_2^{\textcolor{red}{n}} \vec{v}_2 + c_3 \lambda_3^{\textcolor{red}{n}} \vec{v}_3 + \dots + c_N \lambda_N^{\textcolor{red}{n}} \vec{v}_N.$$

If you don't believe me, please try iterating once or twice more ☺.

Whoa, whoa, whoa... weren't we trying to get a better idea about solutions!? It seems like we just did some fancy Linear Algebra-y eigenvalue-y slight of hand. Well, you're not wrong. Except, remember when we talked about the system's behavior being governed by a **dominant eigenvalue**? We're going to use that little nugget of information now! Let's quickly define what it means to be a **dominant eigenvalue**.

Definition 4.1

The **dominant eigenvalue** of a matrix A is the largest eigenvalue in magnitude.

For example if $A \in \mathbb{R}^{3 \times 3}$ and $\lambda_1 = -3$, $\lambda_2 = 0.5$, and $\lambda_3 = 2.5$, the dominant eigenvalue would be $\lambda_1 = -3$ since $|\lambda_1| > |\lambda_2|$ and $|\lambda_1| > |\lambda_3|$.

Without loss of generality, let's saftely assume that λ_1 is the **dominant eigenvalue** of A from our matrix system (1.23). Next from our n^{th} iterated solution, we are going to pull out a λ_1^n from each term:

$$\vec{x}_n = \lambda_1^n \left[c_1 \vec{v}_1 + c_2 \frac{\lambda_2^n}{\lambda_1^n} \vec{v}_2 + c_3 \frac{\lambda_3^n}{\lambda_1^n} \vec{v}_3 + \dots + c_N \frac{\lambda_N^n}{\lambda_1^n} \vec{v}_N \right].$$

Using a smidgen of algebra, we can raise those fractions to the same powers of n :

$$\vec{x}_n = \lambda_1^n \left[c_1 \vec{v}_1 + c_2 \left(\frac{\lambda_2}{\lambda_1} \right)^n \vec{v}_2 + c_3 \left(\frac{\lambda_3}{\lambda_1} \right)^n \vec{v}_3 + \dots + c_N \left(\frac{\lambda_N}{\lambda_1} \right)^n \vec{v}_N \right].$$

Now as $n \rightarrow \infty$ each of those fractions are going to continually get closer and closer to 0, since $|\lambda_k| < |\lambda_1|$ for $k \geq 2$. Therefore, the main contribution to our solution must come from that first term, i.e.,

$$\vec{x}_n \sim c_1 \lambda_1^n \vec{v}_1.$$

We call this an *asymptotic approximation*, since it becomes more and more accurate as n gets larger. For us, though, we can more focus on what happens in the limit as n goes to ∞ of \vec{x}_n . We will fall into one of the following cases:

1. If $|\lambda_1| > 1$:

$$\lim_{n \rightarrow \infty} \vec{x}_n \rightarrow \pm\infty \quad (\text{solutions blow up}).$$

2. If $|\lambda_1| < 1$:

$$\lim_{n \rightarrow \infty} \vec{x}_n = \vec{0} \text{ (solutions all go to 0).}$$

3. If $|\lambda_1| = 1$:

$$\lim_{n \rightarrow \infty} \vec{x}_n = \pm c_1 \vec{v}_1.$$

Thus the behavior is analogous to what happens in the single discrete equation case that we briefly mentioned earlier in this section. The **dominant eigenvalue** drives the solution behavior overall.

This derivation we did in expanding a vector in terms of linear combination of eigenvectors, continually multiplying the system by the matrix A, and upon doing so, doing a lot of swapping A's for λ 's, actually has a name. It is called the **Power Iteration**. The Power Iteration is a numerical algorithm for determining the dominant eigenvalue of a matrix, A by successively multiplying the matrix A to itself. Note you must begin the process by selecting an initial vector to be applying A onto. Although, this algorithm only helps you find the dominant eigenvalue, there is another closely-related algorithm called the **Inverse Power Iteration**, which helps you find others. However, the Inverse Power Iteration uses the same principles as the Power Iteration, but with one slight twist... for more information, consider taking #MAT331: Numerical Analysis.

To close this section, we will briefly explore two different models, where each lends itself to a particular matrix form, i.e., the models are different based on A having a specific form. This models are useful when modeling different classes within a population, such as different ages or developmental stages. The first example we will focus the **Leslie age-structured model**, associated with the **Leslie matrix**, and the second will be a slight variant called the **Usher model**.

1. **Leslie's age-structured models:**

Consider humans for a second. Depending on a human's age, there are different reproduction rates or likelihoods for reproduction. For example, when humans are prepubescent (or preadolescence), they are not capable of reproduction. Similarly, when humans are greater than 60 years old, the likelihood of reproducing is very small. One might consider breaking the population up in the following manner

(while defining their associated age-population variable):

$$\begin{aligned} X_n^1 &: \text{ages 0 to 14} \\ X_n^2 &: \text{ages 15 to 29} \\ X_n^3 &: \text{ages 30 to 44} \\ X_n^4 &: \text{ages 45 to 59} \\ X_n^5 &: \text{ages 60 to 75} \end{aligned}$$

Of course you might think breaking the population up into groups spanning 10, 5, 2, or 1 years would be more accurate, rather than jumps on 15 years. Hey, you're not wrong. By the way, there also isn't a real reason I'm stopping at 75 age for this model... we could add more classes. If you're wondering demographers that consider age-structured population classes like this generally use groups that span 5 years. Since our model assumes groupings of 15 years, this assumes that $n \rightarrow n + 1$ really means time has progresses 15 years in that single iteration.

Anyways, the way in which these models go would be that anyone born into the human population would immediately be put into the X_n^1 group. So if b_i is a birth rate (over a 15 year period) for the parents in age group X_n^i , we could write a discrete equation for X_n^1 in the following manner:

$$X_{n+1}^1 = b_1 X_n^1 + b_2 X_n^2 + b_3 X_n^3 + b_4 X_n^4 + b_5 X_n^5.$$

Next the way in which the other age-classes can change is by moving to the next age group over one time-step, n . Remember moving $n \rightarrow n + 1$ is a leap forward of 15 years in time since we broke the groups into age spans of 15 years. We define parameters τ_{ij} to be the survival rate for those in population group X^i moving to X^j for $i, j = 1, 2, 3, 4, 5$. Note the following:

- Values of $\tau_{i,i+1}$ are not non-zero for $i = 1, 2, 3, 4$ - you can only move to the next age class in the model.
- All other values, i.e.,

$$\begin{aligned} &\tau_{11}, \tau_{13}, \tau_{14}, \tau_{15}, \\ &\tau_{21}, \tau_{22}, \tau_{24}, \tau_{25}, \\ &\tau_{31}, \tau_{32}, \tau_{33}, \tau_{35}, \\ &\tau_{41}, \tau_{42}, \tau_{43}, \tau_{44}, \\ &\tau_{51}, \tau_{52}, \tau_{53}, \tau_{54}, \tau_{55} \end{aligned}$$

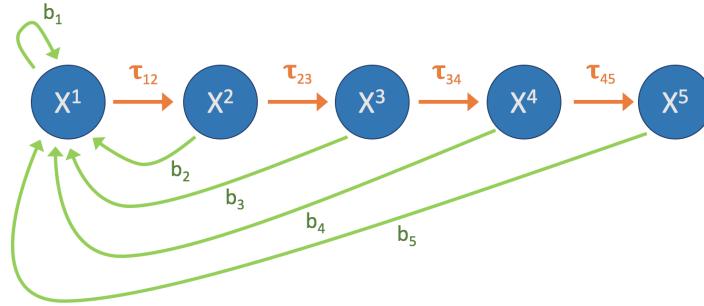


Figure 1.9: A graphical representation of a 5 category Leslie age-structured model. Notice that no matter what class (node) you start in, you can trace a trajectory (following the arrows) to any other class (node) in the graph (assuming all $\tau_{i,i+1} \neq 0$ and $b_i \neq 0$). This graphical property will actually grant us powerful hammer (a theorem!) to analyze the system.

must be zero. This is because you can't go from say adolescence straight into retirement aged adult. Or similarly, you can Benjamin Button it, and go from retirement aged adult back into adolescence.

- Death is taken into account in the model by these survival rates moving from one age group to the next.

Using the definition of $\tau_{i,i+1}$ described above, we can write the entire system of coupled linear, autonomous, homogeneous discrete equations as the following:

$$\begin{aligned} X_{n+1}^1 &= b_1 X_n^1 + b_2 X_n^2 + b_3 X_n^3 + b_4 X_n^4 + b_5 X_n^5 \\ X_{n+1}^2 &= \tau_{12} X_n^1 \\ X_{n+1}^3 &= \tau_{23} X_n^2 \\ X_{n+1}^4 &= \tau_{34} X_n^3 \\ X_{n+1}^5 &= \tau_{45} X_n^4 \end{aligned}$$

We can visualize this system in a graph, see Figure 1.9. This type of graph is called a *directed graph*, since every node has a directed edge relating it to the other in the graph. Figure 1.9 illustrates a strikingly powerful property of Leslie age-structured systems (assuming all $\tau_{i,i+1} \neq 0$ and $b_i \neq 0$). It is **strongly connected**, since you can start at any node and reach any other node.

We will use this property once we write the above system in its equivalent matrix-form:

$$\vec{x}_{n+1} = \begin{pmatrix} X_{n+1}^1 \\ X_{n+1}^2 \\ X_{n+1}^3 \\ X_{n+1}^4 \\ X_{n+1}^5 \end{pmatrix} = \begin{bmatrix} b_1 & b_2 & b_3 & b_4 & b_5 \\ \tau_{12} & 0 & 0 & 0 & 0 \\ 0 & \tau_{23} & 0 & 0 & 0 \\ 0 & 0 & \tau_{34} & 0 & 0 \\ 0 & 0 & 0 & \tau_{45} & 0 \end{bmatrix} \begin{pmatrix} X_n^1 \\ X_n^2 \\ X_n^3 \\ X_n^4 \\ X_n^5 \end{pmatrix} = L\vec{x}_n. \quad (1.25)$$

The matrix form of the system (1.25) shows that the linear operator of the system, i.e., matrix, L has a very particular form. The only non-zero entries in L are its first row and the first sub-diagonal. To circle us back (*pun*), the directed graph illustrating L gives us a nice visualization of how to interpret this matrix.

Where we are hoping to go next is to say something about L 's eigenvalues, since we previously saw that for systems of linear autonomous, homogeneous coupled discrete equations, that their solutions are driven by the eigenvalues of the matrix (flip back a few pages and check out the analysis directly preceding this discussion of Leslie age-structured models). Fortunately, the directed graph representation of L being **strongly connected** and the fact that all parameters must be positive, i.e., $\tau_{i,i+1} \geq 0$ and $b_i > 0$ for all i , gives us a powerful mathematical hammer - Theorem 4.2.

Theorem 4.2

Assume $\tau_{i,i+1} \geq 0$ and $b_i > 0$ for all i . Therefore L is non-negative matrix A whose graphical representation shows that it is **strongly connected**.

If a non-negative matrix A and whose directed graph is **strongly connected**, then A always has a real, positive eigenvalue, λ , that is a simple root (multiplicity one) of its characteristic equation. Moreover, this real, positive eigenvalue λ is special because it is greater than or equal to the magnitude of all of the other eigenvalues.

*Note: This is called the **Frobenius Theorem**.*

Therefore, we are only a few short steps away from showing that the matrix L has a **dominant eigenvalue**. Fortunately, if we assume

that $\tau_{i,i+1} \geq 0$ and $b_i > 0$ for all i , we can show that the matrix L is what we in the business call **primitive**. That is, L will only have one dominant eigenvalue, and no others have equal maximum magnitude. In general, there is a powerful theorem regarding this aspect:

Theorem 4.3

A non-negative matrix A is **primitive** if and only if some power of A is positive.

Note:

- A **non-negative matrix** is a matrix whose entries are all greater than or equal to zero
- A **positive matrix** is a matrix whose entries are all strictly greater than zero.
- Not all Leslie matrices are primitive if you relax the assumptions and do not require that $\tau_{i,i+1} \geq 0$ and $b_i > 0$ for all i .

In a moment, these two qualities - a directed graph representation that is **strongly connected** and being **primitive**, will help us out even more! ...you might see a trend: *knowing more linear algebra is always advantageous*.

If we search for the eigenvalues of L , i.e., subtract L from λI_5 , from that difference's determinant, and all of that jazz we can actually write down the characteristic polynomial for our general Leslie matrix as the following:

$$p_L(\lambda) = \lambda^5 - b_1\lambda^4 - b_2\tau_{12}\lambda^3 - b_3\tau_{12}\tau_{23}\lambda^2 - b_4\tau_{12}\tau_{23}\tau_{34}\lambda - b_5\tau_{12}\tau_{23}\tau_{34}\tau_{45}. \quad (1.26)$$

Recall that all $b_i > 0$ and $\tau_{i,i+1} > 0$ for all appropriate i . Therefore this characteristic polynomial shows that there is only one sign change - the first term λ^5 has the only positive coefficient and then those that follow are negative. We can use **Descartes's Rule of Signs** to help us guarantee that there is only one positive, real root. Thus, making that single positive, real root also the dominant root, and thereby the dominant eigenvalue.

Theorem 4.4

If the nonzero terms of a single-variable polynomial with real coefficients are ordered by descending powers of the variable, then the number of positive roots of the polynomial is either equal to the number of sign changes between consecutive (non-zero) coefficients, or is less than it by an even number.

At this junction, we know that our system (with $b_i > 0$ and $\tau_{i,i+1} > 0$) has one dominant eigenvalue who is both real and positive. Let's call this dominant eigenvalue λ_D . In fact, we can use the form of the characteristic polynomial to help us even further analyze what this eigenvalue could be. At this point we can see from the characteristic polynomial in (1.26) that the following are true:

- $\lim_{\lambda \rightarrow \infty} p_L(\lambda) = \infty$
- $p_L(\lambda_D) = 0$ since λ_D is the dominant eigenvalue and thereby a root of the characteristic polynomial.
- $p_L(\lambda) > 0$ for $\lambda > \lambda_D$
- $p_L(\lambda) < 0$ for $\lambda < \lambda_D$

For our exploration of the solutions what we need to know is whether this dominant eigenvalue, λ_D is greater than 1 or less than 1. Recall that for **any** system of linear autonomous, homogeneous coupled discrete equations that the solution behavior will fall into one of 3 camps:

- (a) If $|\lambda_D| > 1$:

$$\lim_{n \rightarrow \infty} \vec{x}_n \rightarrow \pm\infty \quad (\text{solutions blow up}).$$

- (b) If $|\lambda_D| < 1$:

$$\lim_{n \rightarrow \infty} \vec{x}_n = \vec{0} \quad (\text{solutions all go to 0}).$$

- (c) If $|\lambda_D| = 1$:

$$\lim_{n \rightarrow \infty} \vec{x}_n = \pm c_1 \vec{v}_1.$$

Therefore, it is important for us to figure out what λ_D could be! What would be even gnarlier is if we could figure out what λ_D would be based off of our model's input parameters $\{b_i\}_{i=1}^5$ and $\{\tau_{i,i+1}\}_{i=1}^4$. Lucky for us because of the specific form that the Leslie matrix, L, takes, we can actually do this!

Since we care whether $\lambda_D > 1$ or $\lambda_D < 1$, lets evaluate our characteristic polynomial at 1,

$$p_L(1) = 1 - b_1 - b_2\tau_{12} - b_3\tau_{12}\tau_{23} - b_4\tau_{12}\tau_{23}\tau_{34} - b_5\tau_{12}\tau_{23}\tau_{34}\tau_{45}.$$

Actually, this big ole sum on the right hand side, we give a name - we define it as the **inherent net reproductive number** for our system, and denote it by R_0 . More explicitly, it is written as:

$$R_0 = b_1 + b_2\tau_{12} + b_3\tau_{12}\tau_{23} + b_4\tau_{12}\tau_{23}\tau_{34} + b_5\tau_{12}\tau_{23}\tau_{34}\tau_{45}. \quad (1.27)$$

This R_0 quantity has a very nice intuitive way to think about it. R_0 represents the expected number of offspring per individual per lifetime. Thus is $R_0 > 1$, is assumes that on average a person will have more than one child in their lifetime. When we hop into the analysis of epidemics, in particular disease transmission models, we will have a similar reproduction number to study, although mathematically defined differently. Hence using our mathematical definition of R_0 , we can write $p_L(1)$ as the following:

$$p_L(1) = 1 - R_0.$$

We also get a beautiful (helpful?) theorem relating R_0 , our strictly dominant eigenvalue λ_D , and our solutions \vec{x}_n to (1.25).

Theorem 4.5

Assume the Leslie matrix, L, of the age-structured Leslie model

$$\vec{x}_{n+1} = L\vec{x}_n,$$

has a directed graph representation that is strongly connected and that L is primitive. Then L has a strictly dominant eigenvalue that is both real and positive that satisfies one of the following relationships:

- (a) $\lambda_D = 1$ if and only if $R_0 = 1$.

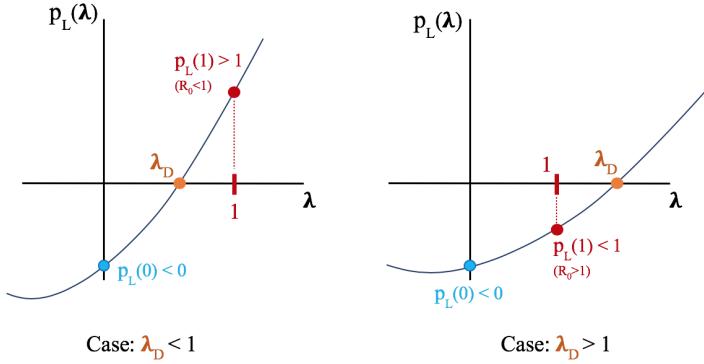


Figure 1.10: Plots of $p_L(\lambda)$ for the two different cases: $\lambda_D < 1$ and $\lambda_D > 1$. These plots illustrate how the two cases fall out due to the constraints placed on the characteristic polynomial, $p_L(\lambda)$, via what we've uncovered in Theorems 4.2, 4.3, and 4.5.

- (b) $\lambda_D > 1$ if and only if $R_0 > 1$.
In this case $\lim_{n \rightarrow \infty} \bar{x}_n \rightarrow \infty$ (solutions eventually blow up).
- (c) $\lambda_D < 1$ if and only if $R_0 < 1$.
In this case $\lim_{n \rightarrow \infty} \bar{x}_n \rightarrow \vec{0}$ (solutions go toward extinction).

Plots of $p_L(\lambda)$ to depict the case when $\lambda_D < 1$ and $\lambda_D > 1$ are provided in Figure 1.10. These plots depict the constraints placed on the characteristic polynomial stemming from our Leslie matrix, L.

2. Usher age-structured models:

The other age-structure model I'll share is called an **Usher** model. Usher models feel very similar to Leslie age-structure models, but with one major difference - a member of a class could stay in the same class at the next time-step.

For example, flip back to our human example for Leslie age-structure models. Remember how we broke the human categories up by 15 year chunks? Then when we iterated our model, we assumed that going from $n \rightarrow n + 1$ meant a 15 year jump in time? In Leslie models, you tend to assume that the time jumps $n \rightarrow n+1$ are the same as the age range in each category (15 year human chunks, and a 15 year time-step); however, Usher models relax this assumption. For an Usher model version of that human problem, we could use the same 15 year age span for each human grouping, but use a time-step of say 5

years. Thus if someone is say in category X_n^2 (age 15-29), they would stay in that group for 3 successive iterations, since each iteration is 5 years. Therefore, people could stay in the same category for numerous discrete jumps from in time, i.e., $n \rightarrow n+1$, $n+1 \rightarrow n+2$, etc..

What this would mean is that we would get the same model as that in (1.25), except with additional terms that allow one to stay in the same class. Since τ_{ij} was defined to be the survival rate for those in population group X^i moving to X^j for $i, j = 1, 2, 3, 4, 5$, these terms would all be of the flavor $\tau_{i,i}X_n^i$. Therefore (1.25) would change to:

$$\vec{x}_{n+1} = \begin{pmatrix} X_{n+1}^1 \\ X_{n+1}^2 \\ X_{n+1}^3 \\ X_{n+1}^4 \\ X_{n+1}^5 \end{pmatrix} = \begin{bmatrix} b_1 + \tau_{11} & b_2 & b_3 & b_4 & b_5 \\ \tau_{12} & \tau_{22} & 0 & 0 & 0 \\ 0 & \tau_{23} & \tau_{33} & 0 & 0 \\ 0 & 0 & \tau_{34} & \tau_{44} & 0 \\ 0 & 0 & 0 & \tau_{45} & \tau_{55} \end{bmatrix} \begin{pmatrix} X_n^1 \\ X_n^2 \\ X_n^3 \\ X_n^4 \\ X_n^5 \end{pmatrix} = U\vec{x}_n. \quad (1.28)$$

Note that the general form of the Leslie matrix, L, is inherently different from that of the Usher matrix, U. Therefore, much of the analysis that guaranteed us certain properties of the Leslie age-structured models may not be applicable to Usher models. However, again, since these Usher models are still coupled linear, autonomous, homogeneous discrete equations, what drives their solutions is still the eigenvalues of the matrix describing the change from one iteration to the next. Therefore, the story still depends on the eigenvalues of U. If you can prove that U has a dominant eigenvalue, that dominant eigenvalue will tell you what to expect of solutions as $n \rightarrow \infty$.

It is important to note that if you move from a Leslie model to an Usher model that although the τ_{ij} and b_i have consistent definitions among the 2 models, the explicit parameter values might be different due to the differences in what each time-step means for each model.

Let's quickly do an example to show the utility of Usher models. This example will involve the life stages of a whale.

Example 1.4.3. Usher models for life stages of a female killer whale.

A female killer whale has 3 very different life stages - the immature stage, its breeding stage, and its post-breeding stage. We can assign each of those stages a discrete variable, such as:

$$\begin{aligned} \text{Stage 1: } & X_n^1 \text{ (immature)} \\ \text{Stage 2: } & X_n^2 \text{ (breeding)} \\ \text{Stage 3: } & X_n^3 \text{ (post-breeding).} \end{aligned}$$

In our Usher model, a whale might stay in one of these stages for more than one time-step at a time. For example, killer whales reach sexual maturity in the time they are about 10-13 years old and will continue to reproduce until they are roughly 41 years old. A female killer whale will live to about 50 years old on average and have about 5 progeny. You could use these estimates to estimate the parameters τ_{ij} and b_i for this model. Just from the above ideas we can assume that

- $b_1 = b_3 = 0$ since whales are only breeding capable when they are in Stage 2, X_n^2 . Therefore we also know that $b_2 > 0$.
- $\tau_{12} > 0$ and $\tau_{23} > 0$ since there is a non-zero likelihood that some whales may progress to the next stage of life
- $\tau_{11} > 0$, $\tau_{22} > 0$, and $\tau_{33} > 0$ since whales might stay in each stage of life for more than one time-step forward, i.e., more than one iteration, $n \rightarrow n + 1$.
- In order to give the parameters specific values, we would need to make sure we take into account what the time-step $n \rightarrow n + 1$ actually is, in order to get estimate their values.
- More data about killer whales and their life stages could only help us estimate these parameters more accurately.

Applying the idea of (1.28) to this model for killer whales would give the following linear system of equations:

$$\vec{x}_{n+1} = \begin{pmatrix} X_{n+1}^1 \\ X_{n+1}^2 \\ X_{n+1}^3 \end{pmatrix} = \begin{bmatrix} \tau_{11} & b_2 & 0 \\ \tau_{12} & \tau_{22} & 0 \\ 0 & \tau_{23} & \tau_{33} \end{bmatrix} \begin{pmatrix} X_n^1 \\ X_n^2 \\ X_n^3 \end{pmatrix} = U \vec{x}_n. \quad (1.29)$$

A graphical representation of the matrix U is given in Figure 1.11. A few more interesting facts (for no other reason than they're interesting): Female killer whales typically live longer than males, and have seen to be in the upwards of 90 years old in some cases, while males typically only live about 30 years on average (but have been

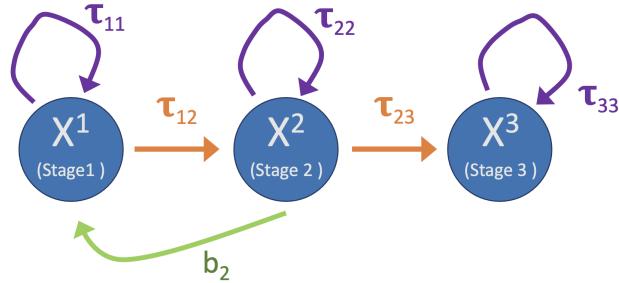


Figure 1.11: A graphical representation of the 3 category Usher age-structured model for killer whales. Notice that only the whales in Stage 2 of their life cycle can reproduce, and that each time-step they can either stay in the same life cycle category or move to the next.

seen as old as 50-60).

1.4.4 Mathematics for stability analysis of coupled discrete equations

Here we will give an brief overview of the mathematics necessary to determine stability properties of equilibria from systems of coupled discrete equations. In particular, I will discuss the following topics here:

- Partial derivatives of a multi-variable function
- Finding the Jacobian matrix, J , for a system
- Determining the **eigenvalues** of the Jacobian matrix, J
 - Matrix addition/subtraction
 - Taking the determinant of a matrix

Disclaimer: this section is going to provide *just enough* background to help us get started to use this tools; however, this is *PLENTY* more than could be studied on this topics. This quick synopsis will not provide these mathematical concepts the justice they deserve. For a deeper understanding of these concepts, please consider taking #MAT229: Multivariable Calculus and/or #MAT205: Linear Algebra.

1. Partial Derivatives:

To begin discussing *partial derivatives*, we must first define a multi-variable *scalar* function. A multi-variable *scalar* function is a function to which takes in more than one input and maps it to one output. That is, a multi-variable *scalar* function of say N *real* variables (independent variables), is a function f , such that $f : \mathbb{R}^N \rightarrow \mathbb{R}$.

For example consider the following multi-variable *scalar* function is a function of 3 input variables, x , y , and z :

$$f(x, y, z) = x^2 y^3 + z^4 \sin(x^5) - \cos(yz),$$

and hence given a $(x, y, z) \in \mathbb{R}^3$, we see that $f : \mathbb{R}^3 \rightarrow \mathbb{R}$.

To introduce the notion of a *partial derivative* of a multi-variable *scalar* function means that we are trying to understand how the function changes with respect to variations of *one* specific input variable.

If you call that for a single valued function, like $g(x) = x^3 + \sin(x)$, that the rigorous definition of a derivative presented in Calculus A is:

$$\frac{dg}{dx} = \lim_{\Delta x \rightarrow 0} \frac{g(x + \Delta x) - g(x)}{\Delta x}.$$

Fortunately, like many ideas in mathematics (but not all), this definition of derivative extends naturally into the higher dimension settings (i.e., multi-variable setting). In a nutshell, *when taking a partial derivative with respect to a particular independent variable, we only consider that specific independent variable to be the only one that is changing and that all others are acting as though they are constant quantities.*

Definition 4.2

The **partial derivative** of a multi-variable *scalar* function $f(x_1, x_2, x_3, \dots, x_N)$ with respect to x_j where $j \in \{1, 2, \dots, N\}$, is defined to be:

$$\frac{\partial f}{\partial x_j} = \lim_{\Delta x_j \rightarrow 0} \frac{f(x_1, x_2, \dots, x_j + \Delta x_j, \dots, x_N) - f(x_1, x_2, \dots, x_N)}{\Delta x_j}.$$

That is, only the variable x_j is considered to be changing, by perturbing it some distance infinitesimal distance Δx_j .

The way we can interpret Definition 4.2 is that it is intuitively the same thing as the single variable function derivative case - there are just more input variables present; however, we treat them as constants. Thus, if we are taking a derivative with respect to, say x_3 , all other x_j (with $j \neq 3$) we would just treat as though they are simply a number, e.g., we could think of them just as having constant values like $x_1 = 2$, $x_2 = 2$, $x_4 = 4$, etc.. Let's consider the multi-variable *scalar* function from before and find the partial derivatives of it with respect to each of its input variables to make sure we are all of the same page.

Example 1.4.4. Partial derivatives of

$$f(x, y, z) = x^2 y^3 + z^4 \sin(x^5) - \cos(yz)$$

- *Finding $\frac{\partial f}{\partial x}$: hold all variables constant, except x . Take derivatives with respect to x only:*

$$\frac{\partial f}{\partial x} = 2x \cdot y^3 + 5x^4 \cdot z^4 \cos(x^5).$$

To examine this a bit further, let's look at each individual addend in $f(x, y, z)$ and consider their individual partial derivatives with respect to x :

(a) $\frac{\partial}{\partial x} (\textcolor{blue}{x^2} y^3) \rightarrow$ we only need to differentiate the $\textcolor{blue}{x^2}$ factor since y^3 does not depend on x . Thus we get

$$\frac{\partial}{\partial x} (\textcolor{blue}{x^2} y^3) = \textcolor{violet}{2x} \cdot y^3$$

(b) $\frac{\partial}{\partial x} (z^4 \sin(\textcolor{blue}{x^5})) \rightarrow$ we only need to differentiate the $\sin(\textcolor{blue}{x^5})$ factor since z^4 does not depend on x . Thus we get

$$\frac{\partial}{\partial x} (z^4 \sin(\textcolor{blue}{x^5})) = z^4 \cdot \textcolor{violet}{5x^4} \cos(\textcolor{violet}{x^5}).$$

(c) $\frac{\partial}{\partial x} (-\cos(yz)) \rightarrow$ there are no terms containing x . Therefore the partial derivative with respect to x would be zero, since we are treating y and z like constants, i.e.,

$$\frac{\partial}{\partial x} (-\cos(yz)) = 0$$

- Finding $\frac{\partial f}{\partial y}$: hold all variables constant, except y . Take derivatives with respect to y only:

$$\frac{\partial f}{\partial y} = x^2 \cdot 3y^2 + z \sin(yz).$$

- Finding $\frac{\partial f}{\partial z}$: hold all variables constant, except z . Take derivatives with respect to z only:

$$\frac{\partial f}{\partial z} = 4z^3 \cdot \sin(\textcolor{blue}{x^5}) + y \sin(yz).$$

Finally, I will briefly mention the gradient operator. The gradient operator is an operator that contains the partial derivative operator for each independent variable. It is denoted by the mathematical symbol ∇ . The main idea that will be helpful for us is that for a multi-variable *scalar* function, such as $f = f(x, y, z)$, the gradient of f produces a 3 component vector that is comprised of all the partial derivatives of f , that is, in this case involving 3 independent variables, $\nabla f : \mathbb{R}^3 \rightarrow \mathbb{R}^3$.

Definition 4.3

The **gradient** of a multi-variable scalar function that takes N input variables, say x_1, x_2, \dots, x_N , is defined to be

$$\nabla f(x_1, x_2, \dots, x_N) = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_N} \end{pmatrix},$$

where the gradient operator is defined as

$$\nabla = \left(\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \frac{\partial}{\partial x_3}, \dots, \frac{\partial}{\partial x_N} \right).$$

Let's compute the gradient of the multi-variable *scalar* function from Ex. 1.4.4.

Example 1.4.5. The gradient of

$$f(x, y, z) = x^2 y^3 + z^4 \sin(x^5) - \cos(yz)$$

Since we have computed all the partial derivatives of this function in Example 1.4.4, we can simply substitute them appropriately here into the gradient of $f(x, y, z)$, i.e., ∇f :

$$\nabla f(x, y, z) = \begin{pmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \\ \frac{\partial f}{\partial z} \end{pmatrix} = \begin{pmatrix} 2x \cdot y^3 + 5x^4 \cdot z^4 \cos(x^5) \\ x^2 \cdot 3y^2 + z \sin(yz) \\ 4z^3 \cdot \sin(x^5) + y \sin(yz) \end{pmatrix}.$$

The gradient of a function is useful to compute because it points in the direction of greatest increase along the surface of the multi-variable function. For example, this is useful in the field of *optimization*, where the gradient can help iterative algorithms move towards

minima (or maxima). Such algorithms are called **gradient descent** algorithms. In fact, neural networks would not work without the assistance of optimization algorithms such as these based on gradients and/or calculus.

Lastly, before we move onto some topics from Linear Algebra, I want to briefly discuss why I have been careful to mention that the multi-variable functions we discussed here were multi-variable **scalar** functions. All of the functions discussed above had the form of N input variables (N independent variables) and produced one output, thus $f : \mathbb{R}^N \rightarrow \mathbb{R}$. The gradient of that multi-variable **scalar** function f had a different kind of output, where it mapped the input to the a space of equal dimension, i.e., $\nabla f : \mathbb{R}^N \rightarrow \mathbb{R}^N$.

The reason I hope to bring your attention to this is because there are also multi-variable **vector** functions, say \vec{f} , where each component of \vec{f} is a multi-variable **scalar** function. For example,

$$\vec{f}(\vec{x}) = \vec{f}(x_1, x_2, \dots, x_N) = \begin{pmatrix} f_1(\vec{x}) \\ f_2(\vec{x}) \\ \vdots \\ f_M(\vec{x}) \end{pmatrix},$$

where f_1, f_2, \dots, f_M are all multi-variable **scalar** functions. Also note that the dimension of the multi-variable **vector** function, \vec{f} , i.e., $\vec{f} \in \mathbb{R}^M$ does not have to be equal to the dimension on the input space, $\vec{x} \in \mathbb{R}^N$.

Also this may seem like a small subtle issue, where we are going next in this section - computing the Jacobian matrix, is actually a consequence of these ideas. It turns out that we can write our system of coupled discrete equations as an iterative **vector** process. If we do that, the Jacobian matrix, J , is actually the gradient of the iterative **vector** function, e.g.,

$$J = \nabla \vec{f}(\vec{x}).$$

For our considerations, to compute J we only really need to know how to take partial derivatives of multi-variable **scalar** functions and where to place them in forming the matrix, J . Let's get into it.

2. Finding the Jacobian matrix:

Consider a **vector** function \vec{f} whose input is a vector from \mathbb{R}^3 , i.e., $(x, y, z) \in \mathbb{R}^3$, and is comprised of 3 multi-variable **scalar** functions, i.e.,

$$\vec{f}(\vec{x}) = \vec{f}(x, y, z) = \begin{pmatrix} f_1(x, y, z) \\ f_2(x, y, z) \\ f_3(x, y, z) \end{pmatrix}.$$

The Jacobian matrix is then defined to be the matrix comprised of all first-order partial derivatives of each component function with respect to each independent variable. In the case listed above, it is defined as:

$$J(x, y, z) = \begin{bmatrix} \frac{\partial f_1}{\partial x} & \frac{\partial f_1}{\partial y} & \frac{\partial f_1}{\partial z} \\ \frac{\partial f_2}{\partial x} & \frac{\partial f_2}{\partial y} & \frac{\partial f_2}{\partial z} \\ \frac{\partial f_3}{\partial x} & \frac{\partial f_3}{\partial y} & \frac{\partial f_3}{\partial z} \end{bmatrix}.$$

Notice that each column corresponds to partial derivatives with respect to a specific independent variable and that each row corresponds to a particular component function, f_j , of \vec{f} .

This notion of a Jacobian matrix generalizes to any dimensions we'd like to explore.

Definition 4.4

Suppose $\vec{f} : \mathbb{R}^N \rightarrow \mathbb{R}^M$. That is, \vec{f} requires N input variables and produces an output vector that exists in \mathbb{R}^M . Assume that all first order partial derivatives exist for each component function of \vec{f} on \mathbb{R}^N . Thus we have,

$$\vec{f}(\vec{x}) = \begin{pmatrix} f_1(x_1, x_2, \dots, x_N) \\ f_2(x_1, x_2, \dots, x_N) \\ \vdots \\ f_M(x_1, x_2, \dots, x_N) \end{pmatrix}.$$

Then the **Jacobian matrix** of \vec{f} is defined to be an $M \times N$ matrix, denoted by J , as the following:

$$J = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_N} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_M}{\partial x_1} & \cdots & \frac{\partial f_M}{\partial x_N} \end{bmatrix},$$

whose (i, j) entry is $J_{ij} = \frac{\partial f_i}{\partial x_j}$.

Let's do an example of this to make sure we're all on the same page. Looking ahead and thinking about what an example will entail - hopefully you're thinking at this point that we'll need some vector function and that we'll be taking A LOT of partial derivatives. That's basically it in a nutshell!

Example 1.4.6. *Find the Jacobian matrix for the following multi-variable vector function,*

$$\vec{f}(\vec{x}) = \vec{f}(x, y, z) = \begin{pmatrix} f_1(x, y, z) \\ f_2(x, y, z) \\ f_3(x, y, z) \end{pmatrix} = \begin{pmatrix} x + x^2y^3 - 3z \\ 5y - x^3z^2 + x \sin(y) \\ 2x + 3y + 4z \end{pmatrix}$$

Taking all the necessary partial derivatives, we get

$$\begin{aligned} \frac{\partial f_1}{\partial x} &= 1 + 2xy^3 \\ \frac{\partial f_1}{\partial y} &= 3x^2y^2 \\ \frac{\partial f_1}{\partial z} &= -3 \\ \frac{\partial f_2}{\partial x} &= -3x^2z^2 + \sin(y) \\ \frac{\partial f_2}{\partial y} &= 5 + x \cos(y) \\ \frac{\partial f_2}{\partial z} &= -2x^3z \\ \frac{\partial f_3}{\partial x} &= 2 \\ \frac{\partial f_3}{\partial y} &= 3 \\ \frac{\partial f_3}{\partial z} &= 4. \end{aligned}$$

Substituting these into the Jacobian matrix form, we get

$$J(x, y, z) = \begin{bmatrix} \frac{\partial f_1}{\partial x} & \frac{\partial f_1}{\partial y} & \frac{\partial f_1}{\partial z} \\ \frac{\partial f_2}{\partial x} & \frac{\partial f_2}{\partial y} & \frac{\partial f_2}{\partial z} \\ \frac{\partial f_3}{\partial x} & \frac{\partial f_3}{\partial y} & \frac{\partial f_3}{\partial z} \end{bmatrix} = \begin{bmatrix} 1 + 2xy^3 & 3x^2y^2 & -3 \\ -3x^2z^2 + \sin(y) & 5 + x \cos(y) & -2x^3z \\ 2 & 3 & 4 \end{bmatrix}.$$

3. Adding/subtracting vectors and matrices:

The first noteworthy thing about adding/subtracting vectors and matrices is that *you can only do this if both vectors or both matrices are the same size!* The way in which you add/subtract vectors or matrices is component wise. For example, consider two vectors

$$\vec{v}_1 = \begin{pmatrix} a_1 \\ b_1 \\ c_1 \end{pmatrix} \quad \text{and} \quad \vec{v}_2 = \begin{pmatrix} a_2 \\ b_2 \\ c_2 \end{pmatrix}.$$

If we were to add them together like $\vec{v}_1 + \vec{v}_2$ we would do it in the following manner:

$$\vec{v}_1 + \vec{v}_2 = \begin{pmatrix} a_1 \\ b_1 \\ c_1 \end{pmatrix} + \begin{pmatrix} a_2 \\ b_2 \\ c_2 \end{pmatrix} = \begin{pmatrix} a_1 + a_2 \\ b_1 + b_2 \\ c_1 + c_2 \end{pmatrix}.$$

Thus, we only can add the same components together. Even if we wanted to subtract, say $\vec{v}_1 - 5\vec{v}_2$, we would do it the same way, except first we would *scale* ("multiple") \vec{v}_2 by 5, i.e., multiply every component in \vec{v}_2 by 5:

$$\vec{v}_1 - 5\vec{v}_2 = \begin{pmatrix} a_1 \\ b_1 \\ c_1 \end{pmatrix} - 5 \begin{pmatrix} a_2 \\ b_2 \\ c_2 \end{pmatrix} = \begin{pmatrix} a_1 - 5a_2 \\ b_1 - 5b_2 \\ c_1 - 5c_2 \end{pmatrix}.$$

The same notions apply to matrices. For example, if we consider

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix},$$

and wish to compute $A - 5B$, we would do it analogously:

$$A - 5B = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} - 5 \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix} = \begin{bmatrix} a_{11} - 5b_{11} & a_{12} - 5b_{12} & a_{13} - 5b_{13} \\ a_{21} - 5b_{21} & a_{22} - 5b_{22} & a_{23} - 5b_{23} \\ a_{31} - 5b_{31} & a_{32} - 5b_{32} & a_{33} - 5b_{33} \end{bmatrix}.$$

The last thing we will define here is the N -dimensional identity matrix, I_N . This is a matrix with zeros everywhere, except 1's on the main diagonal, i.e.,

$$I_N = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{bmatrix}.$$

4. Determinants of a matrix:

The determinant of a **square** matrix A , denoted by either $\det(A)$ or $|A|$, is a scalar value that encodes properties related to the linear transformation that A represents. Okay, a lot to unpack there. Note the following:

- Determinants can only be computed for square matrices, that $A \in \mathbb{R}^{N \times N}$.
- We will see [spoiler] that the Determinant is equal to the product of all the matrix A 's eigenvalues.
- The last bullet is related to the *non-singular* properties of the matrix. That is, if the matrix is invertible, then the determinant does *not* equal zero, and thus no eigenvalue is identically zero.
- Therefore, if $\det(A) = 0$, then the matrix is *singular*, meaning it has no inverse. On the other hand if $\det(A) \neq 0$, then matrix has an inverse, denoted A^{-1} .

Computing a determinant is a nested process. For example if you were to compute a determinant for a 3×3 matrix, during the process you would need to compute determinant for 2×2 sub-matrices within the greater matrix. If you started with a 4×4 , you would compute 3×3 determinants, and thus in the process get to compute 2×2 matrix determinants.

Rather than do this in generality, let's start off with the general method for the determinant of a 2×2 matrix.

Definition 4.5

Consider a general 2×2 real-valued matrix, $A \in \mathbb{R}^{2 \times 2}$, such as

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

Then the **determinant** of A is:

$$\det(A) = ad - bc.$$

Note that the determinant of a matrix will always be a *scalar* value, meaning a number.

In order to compute the determinant of a 3×3 , we will need this definition of the determinant of a 2×2 matrix. Moreover, we have a few different choices in how we could approach taking the determinant of a 3×3 matrix. Let's consider the following general 3×3 matrix:

$$A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix},$$

where all quantities in A are real numbers (constants). To take a determinant of $\det(A)$, we get to choose which row (or column) we would like to expand over. For example, we will choose (for no reason other to illustrate what this means) to expand along the first column of A, i.e.,

$$\begin{bmatrix} \textcolor{blue}{a} & b & c \\ \textcolor{blue}{d} & e & f \\ \textcolor{blue}{g} & h & i \end{bmatrix}.$$

Expanding along this column means that we will choose each component of the column and “block off” all other entries in that row and column and then consider the submatrix leftover. For example, if we first choose to expand along a , we would consider the submatrix:

$$\text{Expand } a : \begin{bmatrix} \textcolor{red}{a} & - & - \\ | & e & f \\ | & h & i \end{bmatrix} \Rightarrow \text{get submatrix } \begin{bmatrix} e & f \\ h & i \end{bmatrix}.$$

If we expanded along d , we would get

$$\text{Expand } d : \begin{bmatrix} - & b & c \\ \textcolor{red}{d} & - & - \\ | & h & i \end{bmatrix} \Rightarrow \text{get submatrix } \begin{bmatrix} b & c \\ h & i \end{bmatrix},$$

and similarly if we expanded along g , we would get

$$\text{Expand } g : \begin{bmatrix} | & a & b \\ | & e & f \\ \textcolor{red}{g} & - & - \end{bmatrix} \Rightarrow \text{get submatrix } \begin{bmatrix} a & b \\ e & f \end{bmatrix}.$$

You might be wondering why we would do that...I don't blame you. We find those submatrices because this expansion is apart of the algorithm for finding determinants. It turns out that to find the determinant of that general 3×3 matrix listed above, we would take the value we've expanded upon and multiply it onto the determinant of that nested submatrix and add all of these together in a specific way. Check it out.

Definition 4.6

Consider a general 3×3 real-valued matrix, $A \in \mathbb{R}^{3 \times 3}$. We will compute its determinant by expanding along the first column:

$$\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}.$$

Expanding in this manner gives that the *determinant* of this 3×3 matrix A is:

$$\det(A) = \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = a \begin{vmatrix} e & f \\ h & i \end{vmatrix} + (-1)d \begin{vmatrix} b & c \\ h & i \end{vmatrix} + g \begin{vmatrix} b & c \\ e & f \end{vmatrix}.$$

Note the following:

- (a) When doing this in practice, we would try to expand along a row or column that typically has the most 0's → that way it would minimize the amount of 2×2 determinants we would need to do.
- (b) The value of the determinant will not change depending on which row or column we compute it along (as long as our arithmetic is correct).
- (c) The (-1) coefficient would go onto any coefficient of an even row (or column) term that we are expanding along.
- (d) If we were expanding along an even row or column entirely, we would multiply our formula above by -1, e.g.,

$$\det(A) = \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = (-1)b \begin{vmatrix} d & f \\ g & i \end{vmatrix} + e \begin{vmatrix} a & c \\ g & i \end{vmatrix} + (-1)h \begin{vmatrix} a & c \\ d & f \end{vmatrix}.$$

Let's do an example of a 3×3 matrix determinant. In this example, notice how we will expand along a row or column that has the most 0's in it to save us some work.

Example 1.4.7. Compute the determinant for the following 3×3 matrix,

$$A = \begin{bmatrix} 1 & 5 & 3 \\ 2 & -7 & 8 \\ 0 & 4 & -2 \end{bmatrix}$$

For this matrix we will expand along the bottom row of the matrix, that is,

$$\begin{bmatrix} 1 & 5 & 3 \\ 2 & -7 & 8 \\ 0 & 4 & -2 \end{bmatrix},$$

with the reason being that we see a 0 there and the second reason being that we have not done an example expanding along a row. It will be the same process as described in Definition 4.6. Because of that 0 in the first entry of that row, we won't have to compute the associated 2×2 determinant.

Using our formula we get that

$$\begin{aligned} \det(A) &= \begin{vmatrix} 1 & 5 & 3 \\ 2 & -7 & 8 \\ 0 & 4 & -2 \end{vmatrix} \\ &= (-1)^4 \begin{vmatrix} 1 & 3 \\ 2 & 8 \end{vmatrix} + (-2) \begin{vmatrix} 1 & 5 \\ 2 & 7 \end{vmatrix} \\ &= -4((1)(8) - (3)(2)) - 2((1)(7) - (5)(2)) \\ &= -4(2) - 2(-3) \\ &= -2. \end{aligned}$$

The process for computing the determinant of a 4×4 matrix (and higher) would follow analogously...as you can tell, there would be a lot of arithmetic operations needed.

At this stage, now that we have a little more Linear Algebra machinery at our disposal, let's finally use it to find the eigenvalues of a matrix!

5. Determining eigenvalues of a matrix:

Finally, we are ready to find those glorious eigenvalues of matrices. We have all the math necessary at this junction. The only thing left to do is combine it all. Recall that an eigenvalue (λ) of a matrix $A \in \mathbb{R}^{N \times N}$ satisfies the following equation:

$$\det(\lambda I_N - A) = 0.$$

Hence, our strategy for finding eigenvalues is the following:

Strategy 4.1

- (a) Let λ be an unknown eigenvalue, i.e., a variable, to whom we are trying to find.
- (b) Subtract the matrix A from the appropriate scaled identity matrix λI . Recall we still don't know what λ is but we've now inserted a variable into the mix.
- (c) Compute the determinant of $\lambda I - A$, that is, compute

$$\det(\lambda I - A) = |\lambda I - A|,$$

- (d) That resulting determinant will be in terms of the unknown variable λ ; it will give us some polynomial function where λ is the variable being raised to different powers. This polynomial is called the ***characteristic polynomial*** of A .
- (e) Setting the characteristic polynomial equal to 0 and finding its roots gives the values of λ . That is, the roots of the characteristic polynomial are the matrix A 's eigenvalues. Note that the number of roots should equal the degree of the polynomial, which is always equal to the dimension of the square matrix A .

□

Let's do a few examples of this. We will do one involving a 2×2 matrix as well as one involving a 3×3 matrix.

Example 1.4.8. *Find the eigenvalues of the following 2×2 matrix,*

$$\begin{bmatrix} 0 & 1 \\ 2 & 3 \end{bmatrix}$$

First, we subtract A from λI_2 ,

$$\lambda I_2 - A = \begin{bmatrix} \lambda & -1 \\ -2 & \lambda - 3 \end{bmatrix}.$$

Next, we compute its determinant,

$$\det(\lambda I_2 - A) = \begin{vmatrix} \lambda & -1 \\ -2 & \lambda - 3 \end{vmatrix} = \lambda(\lambda - 3) - (-1)(-2). = \lambda^2 - 3\lambda - 2$$

Hence $p_A(\lambda) = \lambda^2 - 3\lambda - 2$ is the characteristic polynomial of the matrix A. To find its roots (and hence eigenvalues) we will set it equal to zero:

$$p_A(\lambda) = \lambda^2 - 3\lambda - 2 = 0.$$

We can now call on our friend, the quadratic formula, to help us out once again:

$$\lambda = \frac{-(-3) \pm \sqrt{(-3)^2 - 4(1)(-2)}}{2} = \frac{3 \pm \sqrt{17}}{2}.$$

Hence the eigenvalues of A are

$$\lambda = \left\{ \frac{3 - \sqrt{17}}{2}, \frac{3 + \sqrt{17}}{2} \right\}.$$

Next we will turn our sights to computing the eigenvalues of a 3×3 matrix. We note that it will be the same exact process, but calculating the determinant takes a tad more effort and at the end of the day our characteristic polynomial should be a cubic (a 3^{rd} degree polynomial) and hence we'll have to find the roots of a cubic.

Example 1.4.9. Find the eigenvalues of the following 3×3 matrix,

$$\begin{bmatrix} 2 & -17 & -3 \\ 0 & 4 & 5 \\ 0 & 4 & 3 \end{bmatrix}$$

Using Strategy 4.1 once again, we first subtract A from λI_3 ,

$$\lambda I_3 - A = \begin{bmatrix} \lambda - 2 & 17 & 3 \\ 0 & \lambda - 4 & -5 \\ 0 & -4 & \lambda - 3 \end{bmatrix}.$$

Now finding the determinant by expanding along the first column of that resulting matrix we get

$$\begin{aligned} \det(\lambda I_3 - A) &= \begin{vmatrix} \cancel{\lambda - 2} & 17 & 3 \\ 0 & \lambda - 4 & -5 \\ 0 & -4 & \lambda - 3 \end{vmatrix} \\ &= (\cancel{\lambda - 2}) \begin{vmatrix} \lambda - 4 & -5 \\ -4 & \lambda - 3 \end{vmatrix} + (-1)(0) \begin{vmatrix} 17 & 3 \\ -4 & \lambda - 3 \end{vmatrix} + (0) \begin{vmatrix} 17 & 3 \\ \lambda - 4 & -5 \end{vmatrix} \\ &= (\cancel{\lambda - 2})((\lambda - 4)(\lambda - 3) - (-5)(-4)) \\ &= (\lambda - 2)((\lambda - 3)(\lambda - 4) - 20). \end{aligned}$$

Hence the characteristic polynomial of A is $p_A(\lambda) = (\lambda - 2)((\lambda - 3)(\lambda - 4) - 20)$. Next we wish to find its roots, and thereby find the eigenvalues.

Note we can do this cleverly since looking at $p_A(\lambda)$ we can tell that one of its roots must be $\lambda = 2$, since we have the product of two things equaling zero $\rightarrow (\lambda - 2)$ and $((\lambda - 3)(\lambda - 4) - 20)$. Therefore we only have to focus on setting the other part equal to zero, i.e.,

$$\begin{aligned} (\lambda - 3)(\lambda - 4) - 20 &= 0 \\ \lambda^2 - 7\lambda + 12 - 20 &= 0 \\ \lambda^2 - 7\lambda + -8 &= 0 \\ (\lambda - 8)(\lambda + 1) &= 0. \end{aligned}$$

Thus, we see the other two eigenvalues of A are $-1, 8$. Dang, we didn't get to call on our pal, the quadratic formula, here \odot . The three eigenvalues of A are:

$$\lambda = \{-1, 2, 8\}.$$

Now the story about eigenvalues does not end here by any means! There is *SO* much more we could discuss regarding them, like for example, what happens if we have repeated eigenvalues? Or, what about the eigenvectors that go along with eigenvalues? I'm choosing to end this section with a small (*and by no means exhaustive in any way...literally just a few on my mind right now*) list of where eigenvalues pop up:

- (a) Image compression \rightarrow singular value decomposition
- (b) Special relativity \rightarrow the speed of light is an eigenvalue of the Lorentz transformation
- (c) Clustering algorithms \rightarrow understanding connections between fields (items) in business, marketing, biology, etc.
- (d) Principle Component Analysis (PCA) and dimensionality reduction \rightarrow very related to eigenvalues and eigenvectors
- (e) Prediction algorithms \rightarrow streaming services to predict your rating and thus offer recommendations (these algorithms only care about the largest eigenvalues)
- (f) Google Page Rank algorithms and search engine optimization
- (g) Lastly, for us, linear stability analysis of dynamical systems, both discrete and continuous.

2

Continuous Dynamical Systems

"If you assume continuity, you can open the well-stocked mathematical toolkit of continuous functions and differential equations, the saws and hammers of engineering and physics for the past two centuries (and the foreseeable future)."

– Benoit Mandelbrot

2.1 Introduction and transition to Continuous Dynamical Systems

In this chapter, we will discuss *continuous* dynamical systems. No longer will we be iterating in discrete jumps, but assume we have continuous functions and all of the mathematical machinery that can go along with them - namely differentiation and integration. We will transition away from discrete equations like $X_{n+1} = f(X_n)$ and move towards equations that involve relationships between derivatives and independent and dependent variables, like $\frac{dx}{dt} = f(x)$.

If you haven't had any differential equations since Calculus B, that is a-okay! Like discrete dynamical systems, we will build up all of the relevant mathematical tools that we use in this course. No prior knowledge is necessary, other than those from Calculus A and B ☺.

The first thing I will do here it try to motivate why one may want to step away from discrete dynamical systems, with a good ole pros/cons list. Keep in mind that this list is only to help us intuitively differentiate (pun intended?) some of the benefits or drawbacks of each modeling approach.

- **Discrete Models:**

Positives	Negatives
<ul style="list-style-type: none"> · Can iterate by hand (but ugh.) · Easy to solve with computer · Data comes in discrete form 	<ul style="list-style-type: none"> · Hard to find exact solutions analytically · May portray unrealistic behavior (eg., negative populations, chaos) · "finicky" (see below)

- **Continuous Models:**

Positives	Negatives
<ul style="list-style-type: none"> · We have derivatives! · Solutions are continuous · Preserve realistic behavior (no negative populations) · Lots of methods exist for solving on a computer · More theory exists! - Dr. M. Mizuhara 	<ul style="list-style-type: none"> · Hard to find exact solutions analytically · Slightly higher learning curve to solve on a computer

For example, let's compare a few solutions of the Discrete Equation vs. the Continuous Equation for a few cases. Try to see what I meant by "finicky" - the continuous model better preserves realistic behavior, while the discrete equation' solution behavior more significantly varies (*bifurcate*)! We'll begin by showing how the equations look in both modeling frameworks.

- **Logistic Population Model**

$$X_{n+1} = X_n + kX_n \left(1 - \frac{X_n}{C}\right) \quad \frac{dX}{dt} = kX \left(1 - \frac{X}{C}\right)$$

(discrete)

(continuous)

Figure 2.1 provides solutions of the two above Logistic models for a variety of k values. Notice that for some ranges of k the solutions look almost identical; however, the behavior substantially differs depending on what value of k is chosen.

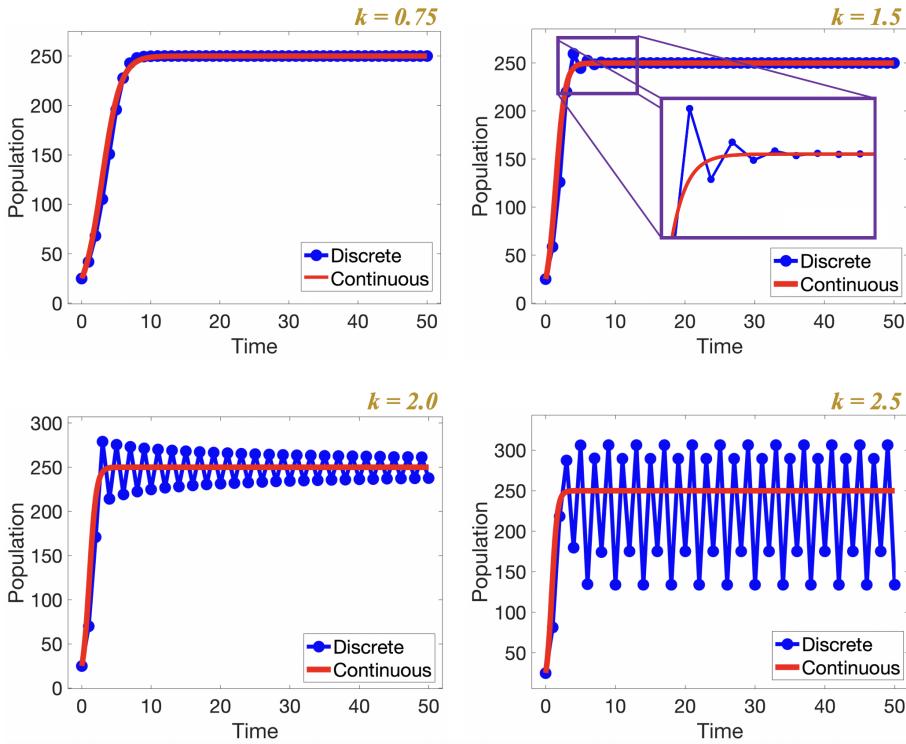


Figure 2.1: Comparing solutions of the discrete and continuous Logistic equation for a variety of k , $C = 250$, and $X_0 = X(0) = 25$.

- **Predator Prey**

$$X_{n+1} = X_n + kX_n \left(1 - \frac{X_n}{C}\right) - q_1 X_n Y_n$$

$$Y_{n+1} = Y_n - dY_n + q_2 X_n Y_n$$

$$\frac{dX}{dt} = kX \left(1 - \frac{X}{C}\right) - q_1 XY$$

$$\frac{dY}{dt} = -dY + q_2 XY$$

(discrete)

(continuous)

Figure 2.2 provides solutions of the two above Predator-Prey models for a variety of q_2 values. Notice that for some ranges of q_2 the solution behavior is remarkably similar; however, the behaviors can substantially differ depending on what value of q_2 was chosen.

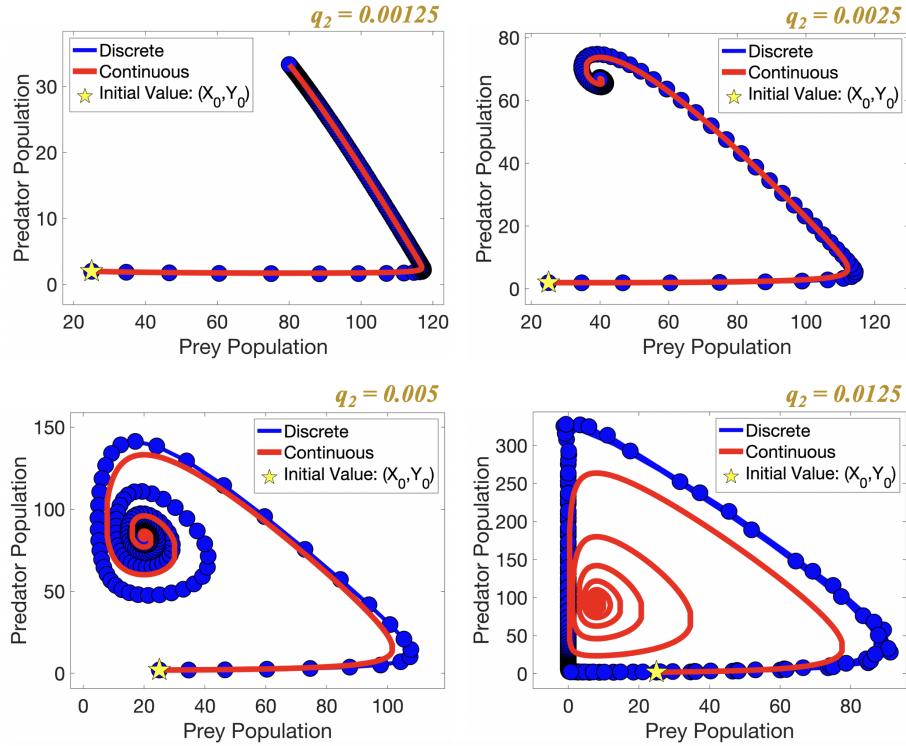


Figure 2.2: Comparing solutions of the discrete and continuous Logistic equation for a variety of q_2 , with $X_0 = X(0) = 25$, $Y_0 = Y(0) = 2$, $k = 0.5$, $C = 120$, $d = 0.1$, and $q_1 = 0.005$.

- **NOTE:** we used identical parameter values to simulate both the continuous and discrete versions of the two models above. What are the parameters units in the above equations? Even if the parameters had the same values would they represent the same thing?

2.2 Mathematically massaging discrete models into continuous models

In this section we will briefly discuss how to take a discrete model and transform it into a continuous model. We will see some familiar friends help us out: linear approximation (or truncated Taylor series), limit definition of the derivative, and dimensional analysis (aka what are the units?!). We will use the basic human population growth model to illustrate these ideas,

$$P_{n+1} = P_n + bP_n - dP_n.$$

Now recall that n is our indexing variable and can only take integer values ($n = 0, 1, 2, \dots$). We will make a few mathematical slight of hands.

First, let's substitute t for n :

$$P_{t+1} = P_t + \textcolor{blue}{b}P_t - \textcolor{red}{d}P_t.$$

We really didn't make any significant change. For all we know that t variable can only take integer values. Next, since P_t really represents what the population value is at the index value t , we could really re-write it as $P(t) = P_t$. This of course assuming that t only takes integer values. Our model equation then becomes

$$P(t+1) = P(t) + \textcolor{blue}{b}P(t) - \textcolor{red}{d}P(t). \quad (2.1)$$

From this equation (2.1), we will derive how to get a continuous model analog to the discrete equation. Before we dive into that, we will briefly explore the units for $\textcolor{blue}{b}$ and $\textcolor{red}{d}$, the birth and death rates, respectively. The square bracket notation $[\cdot]$ is a short hand for discussing the units that a quantity has. For example, below are the units for birth and death rates:

$$[\textcolor{blue}{b}] = \frac{\# \text{ of total births in population}}{(\text{unit of time}) \cdot (\text{total people})}$$

$$[\textcolor{red}{d}] = \frac{\# \text{ of total deaths in population}}{(\text{unit of time}) \cdot (\text{total people})}.$$

Note that this uses the language that ($\#$ of total deaths/deaths in population) and (total people) have the same units - those units being $\#$ of people (*number of people*). That is,

$$[P(t)] = [P(t+1)] = \# \text{ of people}.$$

But, wait! Where did **ALL** of those units come from for $\textcolor{blue}{b}$ and $\textcolor{red}{d}$? Let's take a look back at (2.1) and investigate what the units are on both sides of the equation:

$$\begin{aligned} [P(t+1)] &= [P(t) + \textcolor{blue}{b}P(t) - \textcolor{red}{d}P(t)] \\ [P(t+1)] &= [P(t)] + [\textcolor{blue}{b}P(t)] - [\textcolor{red}{d}P(t)] \\ [P(t+1)] &= [P(t)] + [\textcolor{blue}{b}][P(t)] - [\textcolor{red}{d}][P(t)] \\ (\# \text{ of people}) &= (\# \text{ of people}) + \left(\frac{\# \text{ of total births in population}}{(\text{unit of time}) \cdot (\text{total people})} \right) (\# \text{ of people}) - \\ &\quad \left(\frac{\# \text{ of total deaths in population}}{(\text{unit of time}) \cdot (\text{total people})} \right) (\# \text{ of people}), \end{aligned}$$

and then when you cancel out units (using algebra), you arrive at the following

$$\underbrace{P(t+1)}_{\substack{\text{units of} \\ \# \text{ people}}} = \underbrace{P(t)}_{\substack{\text{units of} \\ \# \text{ people}}} + \underbrace{bP(t)}_{\substack{\text{units of} \\ \# \text{people/time}}} - \underbrace{dP(t)}_{\substack{\text{units of} \\ \# \text{people/time}}} .$$

Although, there is something else that looks fishy - check out those two terms of the far right! The units do not appear consistent across the entire equation - we *need* the units to be # of people, but those two last terms appear to have units of (# of people)/time. That doesn't work - we need consistent units throughout! *Every term that we add or subtract in an equation needs to have the same units!*

Simply put, looking at (2.1) with a quick glance there is no indication of whether t is a continuous or discrete variable, without more information. With this information about units, i.e., *dimensional analysis* conducted, let's take a closer look into each of those possibilities...

1. t is a discrete variable:

If t is a discrete variable, that means t can only take on integer values, such as $t = 0, 1, 2, \dots$. Therefore we get to define what it means for t to go from $t \rightarrow t + 1$. Since we're talking about (2.1) in reference to human populations, let's consider that $t \rightarrow t + 1$ means 1 year has passed.

We can use this additional information to help us fix the issue we our dimension-alization above. We are actually just doing a basic multiply by 1 operation. That is, all we need to do is need to multiply the problematic unit terms (the ones with seemingly units of people/time) by 1 year, i.e.,

$$P(t+1) = P(t) + bP(t)(1 \text{ year}) - dP(t)(1 \text{ year}). \quad (2.2)$$

Did we actually change the equation? It really doesn't look like it, but it actually helps us land the correct units!

2. t is a continuous variable:

We will do the same thing here, except rather than use a time unit of 1 year, we're going to think much smaller... days... minutes... seconds... nanoseconds... hey, you know what, let's just let that small unit of time be Δt . That way, we can explicitly declare what it is later, but either way, it has units of time!

Also, we no longer will consider a jump forward in time from $t \rightarrow t + 1$, but rather $t \rightarrow t + \Delta t$. Using this information, we see that (2.1) would become

$$P(t + \Delta t) = P(t) + bP(t)\Delta t - dP(t)\Delta t. \quad (2.3)$$

Compare (2.2) to (2.3). Convince yourself of the following:

- (a) If $\Delta t = 1$, then the equations are the exact same and it appears that (2.3) would be a discrete equation... in this event t would certainly not be a continuous variable.
- (b) If $\Delta t \neq 1$ then the equations are different!
- (c) The smaller Δt is, the "more continuous" the function $P(t)$ looks. Smaller and smaller Δt seems to make it appear more and more continuous. I guess that makes sense, since the distance between solution points are getting closer and closer together as Δt gets smaller. For example, consider what would

happen if we updated our population everyday versus every minute versus every second versus every nanosecond *is it just me, or does it feel like we might be heading down a path do use something from Calculus A - limits, anyone?*

For right now, let's just roll with the idea that Δt is *something small but not identically zero*. Let's rearrange (2.3) to get the following:

$$P(t + \Delta t) - P(t) = \Delta t(bP(t) - dP(t)),$$

and if we divide both sides by Δt , we get

$$\frac{P(t + \Delta t) - P(t)}{\Delta t} = (b - d)P(t).$$

Take a look at the above equation? What does the left hand side look like? What if Δt is small, *wait*, smaller than that, *actually*, smaller than that, too, . . . , actually if we just let $\Delta t \rightarrow 0$ then we have a derivative of the function $P(t)$. Let's do that by taking the limit of both sides as $\Delta t \rightarrow 0$:

$$\lim_{\Delta t \rightarrow 0} \frac{P(t + \Delta t) - P(t)}{\Delta t} = \lim_{\Delta t \rightarrow 0} \underbrace{(b - d)P(t)}_{\text{does not depend on } \Delta t}$$

$$\frac{dP}{dt} = (b - d)P.$$

That last equation is our continuous population growth model. Good job - we were able to derive a continuous model for population growth from discrete modeling ideas! Fortunately, for us, we won't be doing this transformation process routinely, but it is only proper mathematical form as mathematical biologists in training to at least see these ideas once, or twice.

Next let's remind ourselves what derivatives mean in both a practical and intuitive sense.

2.3 Refresher - what do derivatives tell us?

Let's start by considering a general continuous, differentiable function over the interval $t \geq 0$, $P(t)$. Say this function $P(t)$ gives the number of squirrels at TCNJ at time t (where t is measured in years since 1987); thus $[P(t)] = \#$ of squirrels and $[t] = \text{years}$. Taking a derivative of $P(t)$ with respect to t , we get the rate of change of $P(t)$ with respect to t . That is, $\frac{dP}{dt}$. Note that the derivative has the following units:

$$\left[\frac{dP}{dt} \right] = \frac{\#\text{ of squirrels}}{\text{year}}.$$

Depending on the sign of the derivative, one of the following situations could arise:

1. $\frac{dP}{dt} > 0$: the function $P(t)$ is increasing as t increases. So, the population of squirrels appears to be increasing - it's going up ☺!
2. $\frac{dP}{dt} < 0$: the function $P(t)$ is decreasing as t increases. So, the population of squirrels appears to be decreasing - it's going down ☹.
3. $\frac{dP}{dt} = 0$: the population of squirrels is not changing. *What does case this remind you of?*

We can extend these ideas to *differential equations*. If we have a differential equation, such as

$$\frac{dP}{dt} = f(t, P),$$

we can interpret $f(t, P)$ as some function that depends on the independent variable, t , as well as the dependent variable P , to which we are trying to solve for! You could also read this as the rate that the function P is changing with respect to t , depends on the function P itself.

This might seem like a strange idea, but we can think of it in terms of a practical example with squirrel populations. If there are only 10 squirrels ($P = 10$), the following year there might be slightly more squirrels from those who reproduce (and successfully survive). However, the population from year to year would increase much more rapidly if there were 1000 squirrels ($P = 1000$) to begin with. So the rate in which a population changes should depend on what the current population level is! It would be a little strange if every year the population just grew by 5 squirrels, whether there were 10 squirrels the year before or 1000.

Next we will be diving head first into a similar kind of analysis that we conducted for Discrete Dynamical Systems - searching for equilibrium points and their stability properties. We already have the first bit of the puzzle in place based on what we know about derivatives. We've already (subtly) pointed out how to find equilibrium points - they occur when the right hand side of our (first-order) differential equation is zero!

2.4 Equilibria & Stability in Systems of Differential Equation

We will dive right into equilibria and their associated stability properties for systems of differential equations. From our Calculus A toolbox we already know how to find equilibrium points for a first-order differential equation - it's when the derivative equals zero! This idea even extends to systems of first-order differential equations.

Definition 4.1

An *equilibrium point* of a first-order autonomous differential equation

$$\frac{dx}{dt} = f(x)$$

is a constant solution, denoted \bar{x} , such that

$$\frac{dx}{dt} = 0 \Rightarrow f(\bar{x}) = 0.$$

An *equilibrium point* of a system of first-order autonomous differential equations

$$\begin{aligned}\frac{dx}{dt} &= f(x, y, z) \\ \frac{dy}{dt} &= g(x, y, z) \\ \frac{dz}{dt} &= h(x, y, z)\end{aligned}$$

is a point $(\bar{x}, \bar{y}, \bar{z})$ such that all of their right-hand sides are zero, i.e.,

$$\begin{aligned} f(\bar{x}, \bar{y}, \bar{z}) &= 0 \\ g(\bar{x}, \bar{y}, \bar{z}) &= 0 \\ h(\bar{x}, \bar{y}, \bar{z}) &= 0. \end{aligned}$$

- Note that this idea extends into any amount of coupled first-order autonomous differential equations
- The more coupled equations there are, the more difficult it might be to find an equilibrium point → you're most likely solving a non-linear root finding problem.

Let's do an example just to make sure we're on the same page.

Example 2.4.1. Equilibrium Points for the Depensation model (Allee effect model).

Find the equilibrium points for the following equation governing depensation (fishery scientists) or the Allee effect (ecologists):

$$\frac{dP}{dt} = f(P) = kP \left(1 - \frac{P}{C}\right) \left(\frac{P}{r} - 1\right),$$

where $k, P, r \in \mathbb{R}^+$, $r < C$, and initial value $P(0) = P_0 > 0$.

We are looking for a point \bar{P} such that $f(\bar{P}) = 0$. Hence, we get to solve the following (beautifully factored) cubic polynomial problem:

$$k\bar{P} \left(1 - \frac{\bar{P}}{C}\right) \left(\frac{\bar{P}}{r} - 1\right) = 0.$$

Therefore we find that the equilibrium points are $\bar{P} = \{0, C, r\}$.

You might imagine we are now going to jump into the stability of these equilibria - correct! Before formally giving you the mathematical hammer (theorem) for stability of equilibria in first-order differential equation systems, let's take a moment to reflect on some ideas from Calculus A that we already quickly reviewed in Section 2.3. When the derivative is positive, the function is increasing. When it's negative, the function is decreasing, and when the derivative is zero, the function is constant. *Okay, so what's your point?*

Let's quickly look at the Depensation (Allee effect) model from Example 2.4.1 above. If we were to plot Equilibrium values vs. time, we get the following plot, shown in Figure 2.3. This might not look like much, really just a few horizontal lines illustrating the equilibrium solutions; however, the power of calculus and differential equations provides much more insight. Check it out:

- We know that at those three equilibrium solutions that $\frac{dP}{dt} = 0$.
- Due to the implied continuity (and differentiability... we're working with a derivative after-all) we know that if a solution starts, say, somewhere between r and C , i.e., $r < P_0 < C$, we know that the solution $P(t)$ will be stuck between r and C for all time t .

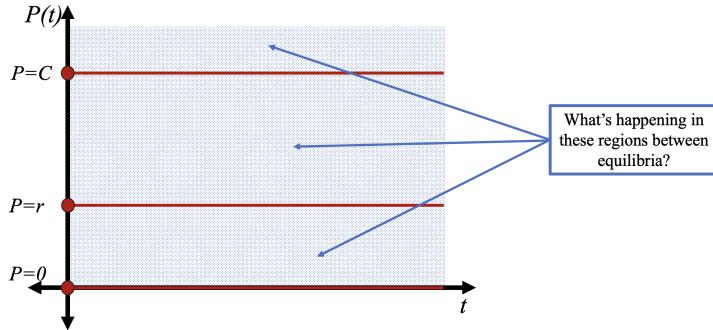


Figure 2.3: Plot of the equilibrium solutions for the Depensation (Allee effect) Model from Example 2.4.1.

- These solutions are “stuck” in this range because if they ever get near one of those points r or C , the derivative is nearly zero. Therefore the solution must stop changing since both r and C happen to be equilibrium points in this case.
- This is true of all equilibrium points - a solution can never get onto the other side of an equilibrium solution - it can’t hop across. Those derivatives near the equilibrium solutions are nearly zero.
- And, and, and since we have its derivative, we can tell what the solution does between all of the equilibrium solutions! This is an amazing feature.
- We know that since the solution to the differential equation must be continuous (and differentiable) that in each of these regions between equilibria that the derivative can never be zero. Therefore the derivative can only have one sign in each of these regions! It can only ever be increasing or decreasing between equilibrium solutions. Rad!

For the Depensation (Allee effect) model above, since we assume that $P(t)$ can only be greater than or equal to zero, this gives us three regions to investigate: $0 < P_0 < r$, $r < P_0 < C$, and $P_0 > C$. We effectively choose which region our solution would be in based off the initial value $P(0) = P_0$. Whichever region our solution starts in, it must stay! Plus, it can only ever be doing the same thing in those regions - increasing or decreasing. Let’s walk through those regions now:

1. $0 < P_0 < r$:

If we assume an initial value is in $(0, r)$, all $P(t)$ will forever be somewhere in this region. Let’s check out the sign of each factor in the derivative in its governing equation!

$$\frac{dP}{dt} = \underbrace{kP}_{+} \left(1 - \underbrace{\frac{P}{C}}_{+}\right) \left(\underbrace{\frac{P}{r} - 1}_{-}\right).$$

$kP > 0$ and $\left(1 - \frac{P}{C}\right) > 0$ since $0 < P < r$. However, $\left(\frac{P}{r} - 1\right) < 0$ because $\frac{P}{r} < 1$ for $0 < P < r$.

Therefore in this region we see that

$$\frac{dP}{dt} < 0 \quad \text{when } 0 < P < r \quad \text{and solutions must decrease.}$$

2. $r < P_0 < C$:

If we assume an initial value is in (r, C) , all $P(t)$ will forever be somewhere in this region. Checking the sign of each of the factors in the derivatives gives us

$$\frac{dP}{dt} = \underbrace{kP}_{+} \underbrace{\left(1 - \frac{P}{C}\right)}_{+} \underbrace{\left(\frac{P}{r} - 1\right)}_{+}.$$

$kP > 0$, $\left(1 - \frac{P}{C}\right) > 0$, and $\left(\frac{P}{r} - 1\right) > 0$ since $r < P < C$.

Therefore in this region we see that

$$\frac{dP}{dt} > 0 \quad \text{when } r < P < C \quad \text{and solutions must increase.}$$

3. $P_0 > C$:

If we assume an initial value is $P_0 > C$, all $P(t)$ will forever be somewhere in this region. Again, checking the sign of each of the factors in the derivatives gives us

$$\frac{dP}{dt} = \underbrace{kP}_{+} \underbrace{\left(1 - \frac{P}{C}\right)}_{-} \underbrace{\left(\frac{P}{r} - 1\right)}_{+}.$$

$kP > 0$ and $\left(\frac{P}{r} - 1\right) > 0$ since $P > C$. However, $\left(1 - \frac{P}{C}\right) < 0$, since $P > C$ now which implies that $1 < \frac{P}{C}$.

Therefore in this region we see that

$$\frac{dP}{dt} < 0 \quad \text{when } P > C \quad \text{and solutions must decrease.}$$

Making a visual representation of the above information we can amend Figure 2.3 to what is shown in Figure 2.4. In any of these regions, once we put in an initial value P_0 somewhere into it, we know the general behavior that the solutions will exhibit - either increasing or decreasing. For example, if we consider an initial value P_0 such that $r < P_0 < C$, we know that the solution will increase, but can only increase as far as the next equilibrium solution above it (at $P = C$), as Figure 2.4 illustrates. We effectively know what solutions will do in this equation, regardless of where we start (as long as $P_0 > 0$). All solutions look like they will approach one of the three possible equilibrium solutions. However, while we know all solutions will tend to an equilibrium based on our Calculus reasoning above, we do not know exactly *how* they tend towards it. What I mean is that we generally have the idea of where the solutions will go, but don't know the exact functional form of the solutions - only that they are either increasing or decreasing.

Okay, cool. I get how Calculus A works, but what does this have to do with stability? Remember back to when we thought about stability in discrete equations. Our intuitive idea of a *stable equilibrium* was that if the solution ever got close enough to that equilibrium point, it would remain in a small neighborhood of it, while simultaneously getting

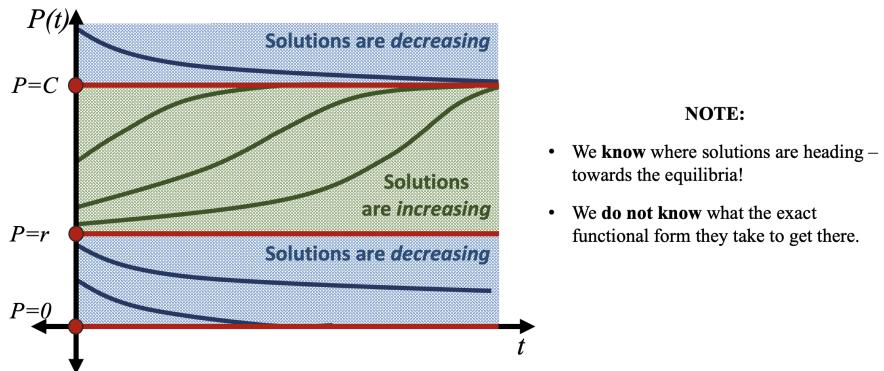


Figure 2.4: An illustration of what solutions (between two equilibria) must do in the Depensation (Allee effect) model.

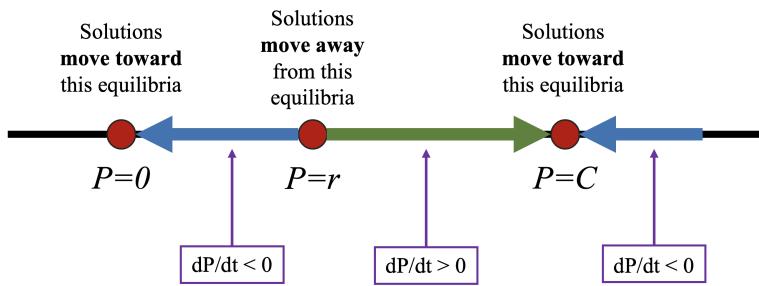


Figure 2.5: The 1D Phase Portrait for the Depensation (Allee effect) Model. Compare this to Figure 2.4.

closer and closer to it as the the number of iterations increased. In the differential equation example above, we can see from Figure 2.4 where solutions will tend over time. Moreover, if we are very close to one of those equilibrium solutions ever, we can tell whether we will get closer to it (*stable*) or move far away from it (*unstable*). All of this is tied up in what the derivative is - is it positive or negative!?

Another way to see this is from what is called a **1D Phase Portrait**. Since we are only dealing with one differential equation in this case, we can plot the equilibrium points along a horizontal line and then use arrows to illustrate whether you could move towards or away from each equilibrium. Check this out in Figure 2.5. From observing this phase portrait you can see that if solutions are moving away from an equilibria, it must be unstable. On the other hand, if all solutions are moving towards an equilibrium, it must be stable!

Furthermore, we can think quite geometrically about the stability of equilibrium points for differential equation models, based on the ideas shown in Figure 2.5. From Calculus A and our discussion of equilibrium points above, we know that when the derivative is equal to zero, our function (or solution) will not change. If we look at the slopes surrounding that *extremum* in Calculus A (or equilibrium here), we are able to determine if it is a maximum or a minimum. Recall that a maximum and minimum occur when:

- Maximum: a point \bar{x} is a maximum if the slope before the point is positive and

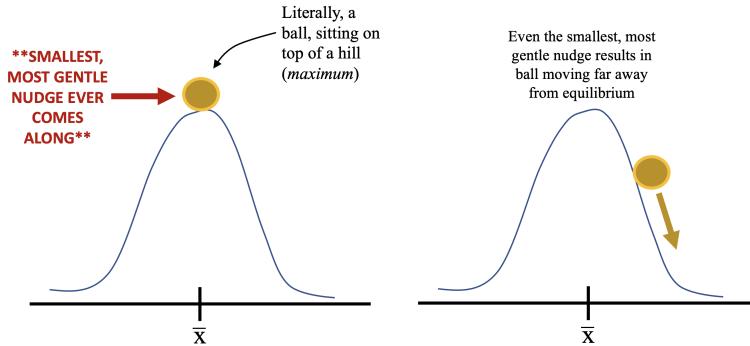


Figure 2.6: The geometric interpretation behind an unstable equilibrium using Calculus A ideas.

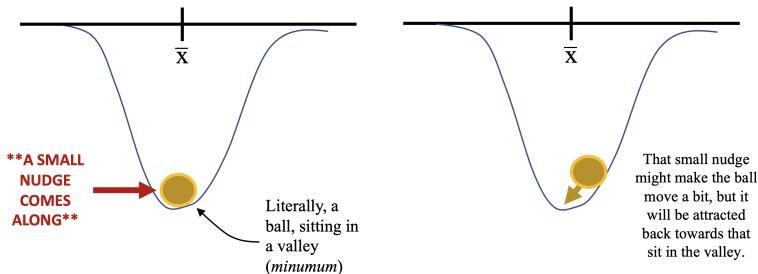


Figure 2.7: The geometric interpretation behind a stable equilibrium using Calculus A ideas.

the slope after it is negative. Also, the second derivative evaluated at \bar{x} is positive.

- Minimum: a point \bar{x} is a minimum if the slope before the point is negative and the slope after it is positive. Also, the second derivative evaluated at \bar{x} is negative.

What does this have to do with stability and a geometric interpretation? Welp, if our equilibrium point is a “maximum”, like the ideas above - solutions increase before the equilibrium point and then decrease thereafter, we can imagine a ball sitting at the top of a hill (see Figure 2.6). If the smallest, most gentle nudge came along and tapped on the ball, it would roll down the hill and far away from that equilibrium (maximum) value. This is a geometric interpretation of being an unstable equilibrium. If you are the smallest bit away from that equilibrium, you are going to move far away from it!

On the other hand, if our equilibrium point is a “minimum”, where solutions decrease coming up to it and then increase thereafter, we can envision the scenario in Figure 2.7. A small nudge may push the ball away from the bottom of the valley, but it would simply roll back towards its original equilibrium position at the bottom of the valley.

Not only are we able to use our Calculus A knowledge to gain this geometric interpretation of the stability of equilibrium points for equilibrium solutions to differential equations, but we can go one step further. Calculus A also gives us the recipe for analytically deter-

mining the stability properties! Recall above, where we reminded ourselves about what it meant to be a maximum or minimum. We had that additional useful method through the second derivative - we can use this to determine the stability of our equilibrium points!

If we're given a differential equation in the following form

$$\frac{dx}{dt} = f(x),$$

note that we already have the derivative of the function $x(t)$ there - it is literally the right hand side of the differential equation. Thus, we already have a first derivative of sorts. It might not be a standard explicit derivative, since the right hand side depends on x and not t , but it's a derivative nonetheless. Moreover, when $f(x) = 0$, we call those solutions equilibrium solutions, \bar{x} as above. Thus if we take another derivative (akin to a second derivative of sorts), we can determine whether that equilibrium is a minimum or maximum, or more importantly for us, a stable equilibrium or unstable equilibrium, respectively. We get the following Stability theorem:

Theorem 4.1

Consider an autonomous differential equation of the form

$$\frac{dx}{dt} = f(x)$$

with equilibrium point \bar{x} , i.e., $f(\bar{x}) = 0$. If

1. $f'(\bar{x}) < 0$: the equilibrium point is **locally asymptotically stable**.
2. $f'(\bar{x}) > 0$: the equilibrium point is **unstable**.
3. $f'(\bar{x}) = 0$: the stability is indeterminate. The differential equation is called *non-hyperbolic*.

Note: the cases in which $f'(\bar{x}) \neq 0$, the differential equation is called *hyperbolic*.

It's probably not too surprising that the stability of an equilibrium point of a differential equation model is so highly dependent on a derivative. Continuity and differentiability of a function are extremely powerful tools to have at your disposal. Here they helped us guarantee that solutions would be stuck in a particular region between two equilibrium solutions and that those solutions could only ever be increasing or decreasing in those regions. As you might expect, once we start getting into coupled systems of differential equations, we might start seeing other stability behavior as in the coupled discrete dynamical systems case.

Before we do that, I offer you yet another proof using Taylor Series. This time for the stability of an equilibrium point from a differential equation. Enjoy!

Proof 2.4.1. (An outline of a proof for Theorem 4.1)

Consider the following linear autonomous differential equation

$$\frac{dx}{dt} = f(x),$$

with initial value $x(0) = x_0$ and equilibrium solution \bar{x} , i.e., $f(\bar{x}) = 0$, where $f(x)$ is continuous over an open interval containing \bar{x} . We will show that if the solution $x(t)$ is ever close enough to \bar{x} and that $f'(\bar{x}) < 0$, it guarantees \bar{x} is a **stable equilibrium**.

First, we must define the distance between our solution $x(t)$ and the equilibrium itself,

$$e(t) = x(t) - \bar{x}.$$

If our solution $x(t)$ is close to the equilibrium, then $e(t)$ is small. Spoiler - this quantity $e(t)$ is how we can rigorously make sure we are sufficiently close to the equilibrium. Next we are going to differentiate both sides of the above definition of $e(t)$ with respect to t :

$$\frac{d}{dt} e(t) = \frac{d}{dt} [x(t) - \bar{x}]$$

$$\frac{de}{dt} = \frac{dx}{dt} - \frac{d\bar{x}}{dt}$$

$$\frac{de}{dt} = \underbrace{\frac{dx}{dt}}_{f(x)}$$

$$\frac{de}{dt} = f(x(t)).$$

Welp, we're at a crossroads now. The final differential equation looks like we just swapped out the left hand side from the original. However, recall that we have a relationship between $e(t)$ and $x(t)$ - we can rewrite $x(t)$ in terms of $e(t)$ and \bar{x} , i.e., $x(t) = \bar{x} + e(t)$. Let's substitute that into the equation above

$$\frac{de}{dt} = f(\bar{x} + e(t)).$$

This is where we can now invoke the "close enough" clause. We are assuming that our solution $x(t)$ is close enough to the equilibrium \bar{x} for some t . This is also where we could make this proof more (properly) rigorous using the language of Real Analysis and define an appropriate ϵ -neighborhood about the equilibrium. However, we won't be that formal here.

Looking at the right hand side of the above equation $f(\bar{x} + e(t))$, if $e(t)$ is a small quantity, we can Taylor expand about \bar{x} . Upon Taylor expanding about \bar{x} we get

$$\frac{de}{dt} = \underbrace{f(\bar{x})}_0 + f'(\bar{x})e(t) + \frac{1}{2!}f''(\bar{x})[e(t)]^2 + \dots$$

From the above expression, we know one of those terms is zero $\rightarrow f(\bar{x}) = 0$. This was the definition of an equilibrium point \bar{x} for our differential equation! So our equation reduces to

$$\frac{de}{dt} = f'(\bar{x})e(t) + \frac{1}{2!}f''(\bar{x})[e(t)]^2 + \dots$$

Now since we are assuming that we are close enough to the equilibrium, we must have that

$$e(t) \gg [e(t)]^2 \gg [e(t)]^3 \gg [e(t)]^4 \gg \dots$$

Therefore, to first-order our equation takes the following form

$$\frac{de}{dt} \approx \underbrace{f'(\bar{x})}_{\text{constant}} e(t).$$

Since the first derivative evaluated at the equilibrium is a constant, let's define it to be k to make this look a bit easier to understand, i.e., let $k = f'(\bar{x})$. Hence we need to explore the following differential equation

$$\frac{de}{dt} \approx ke.$$

Recall that $e(t)$ is a function that gives the distance between our solution $x(t)$ to the governing differential equation and the equilibrium. As t increases, ideally $e(t) \rightarrow 0$. More formally, what this means is that in order to be a stable (or unstable) equilibrium we need the following to happen:

$$\begin{aligned} e(t) &\rightarrow 0 \text{ as } t \rightarrow \infty \quad \text{stable} \\ e(t) &\rightarrow \infty \text{ as } t \rightarrow \infty \quad \text{unstable}. \end{aligned}$$

In order to figure out what happens, let's solve that differential equation above for $e(t)$. We can do this using tools from Calculus B - solving differential equations via the separable equations technique!

$$\frac{de}{dt} = ke$$

$$\frac{de}{e} = kdt$$

$$\int \frac{de}{e} = \int kdt$$

$$\ln|e(t)| = kt + C \quad \text{where } C \text{ is the integration constant}$$

$$e(t) = e^{kt} e^C$$

$$e(t) = Ae^{kt}.$$

Using our knowledge of exponential functions, we know that the only way $e^{kt} \rightarrow 0$ as $t \rightarrow \infty$ is if $k < 0$. Therefore we have

$$\lim_{t \rightarrow \infty} e(t) = \lim_{t \rightarrow \infty} e^{kt} = \lim_{t \rightarrow \infty} e^{f'(\bar{x})t} = 0$$

if $k = f'(\bar{x}) < 0$. Thus for local asymptotic stability we have the condition that $f'(\bar{x}) < 0$.

Note:

- If $k = f'(\bar{x}) > 0$, this implies that $e(t) \rightarrow \infty$ as $t \rightarrow \infty$. Therefore the equilibrium would be unstable.
- If $k = f'(\bar{x}) = 0$, the stability properties of that equilibrium are indeterminate. This is what is called a non-hyperbolic equation.
- On that note, when $f'(\bar{x}) \neq 0$, the equation is called hyperbolic.

Before we move onto coupled systems of differential equations, I want to bring your attention to difference between the stability conditions for discrete equations and differential (continuous) equations. Both depended on the first derivative of the governing function ("right hand side") evaluated at the equilibrium point. However, they are subtly different! Check it out:

	Discrete Equation	Differential Equation
Equation	$x_{n+1} = f(x_n)$	$\frac{dx}{dt} = f(x)$
Stability Condition	$ f'(\bar{x}) < 1$	$f'(\bar{x}) < 0$

Now, the above Table 2.4 illustrates that the conditions for stability for first-order discrete and first-order differential equations have a similar feel, but with ultimately a different condition. The same will be true for coupled systems of differential equations vs. coupled systems of discrete equations. Actually, I think we're ready for this now - we've done most of the work already when exploring discrete systems. We're going to see the following pop-up:

1. Solving a system of algebraic equations to find the equilibrium point(s), as in Definition 4.1 above.
2. Computing a Jacobian matrix for the system of coupled equations.
3. Evaluating that Jacobian matrix at the equilibrium point.
4. Finding the eigenvalues of the Jacobian matrix (for a specific equilibrium point)
5. Based on the values of those eigenvalues, determining stability.

Let's dive right now and try to recreate this process for a general coupled system of 2 differential equations.

Example 2.4.2. Generic coupled system of 2 differential equations, equilibria, and stability.

Consider the following dynamical system

$$\frac{dx}{dt} = f(x, y) \quad (2.4)$$

$$\frac{dy}{dt} = g(x, y) \quad (2.5)$$

where $f(x, y)$ and $g(x, y)$ are continuous and differentiable functions of x and y on an open set that will contain an equilibrium point for the system, (\bar{x}, \bar{y}) . Let's now go through our recipe for finding equilibria and their stability properties.

1. Finding Equilibria:

We know that equilibria for differential equations occur when the derivatives are all zero. Here that would mean that we need

$$\frac{dx}{dt} = f(x, y) = 0$$

$$\frac{dy}{dt} = g(x, y) = 0.$$

Therefore, we must find values \bar{x} and \bar{y} such that $f(\bar{x}, \bar{y}) = 0$ and $g(\bar{x}, \bar{y}) = 0$. This is an algebraic problem, but nonetheless, this may not be an easy task.

2. Compute the Jacobian Matrix:

To find the Jacobian matrix, we need to take all first-order partial derivatives of our functions $f(x, y)$ and $g(x, y)$ and arrange them in a matrix appropriately, as follows

$$J = \begin{bmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \\ \frac{\partial g}{\partial x} & \frac{\partial f_g}{\partial y} \end{bmatrix}.$$

Each row of J specifies a particular function - either f or g . Each column of J specifies a partial derivative with respect to a specific variable - either x or y .

3. Evaluate the Jacobian Matrix at the Equilibrium Point:

Next we simply must remember to evaluate the Jacobian matrix, J , at the equilibrium point (\bar{x}, \bar{y}) .

$$J \Big|_{(\bar{x}, \bar{y})} = \begin{bmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \\ \frac{\partial g}{\partial x} & \frac{\partial f_g}{\partial y} \end{bmatrix} \Big|_{(\bar{x}, \bar{y})}.$$

4. Find eigenvalues of the Jacobian Matrix:

Finally, we can find the eigenvalues of the Jacobian matrix. Recall that this involves taking the determinant of the matrix difference $\lambda I - J$ and setting it equal to zero, i.e.,

$$|\lambda I - J| = 0.$$

This determinant will result in a polynomial in λ (the characteristic polynomial) to which its roots are the eigenvalues. At the end of the day, finding the eigenvalues is a root-finding problem.

If we generalize the elements of $J(\bar{x}, \bar{y})$ in the following way (mostly to hide the partial derivatives evaluated at the equilibrium):

$$J \Big|_{(\bar{x}, \bar{y})} = \begin{bmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{bmatrix},$$

we can walk through the steps for this generic 2x2 system. Let's begin...

$$\begin{aligned} \lambda I - J(\bar{x}, \bar{y}) &= \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} - \begin{bmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{bmatrix} \\ &= \begin{bmatrix} \lambda - J_{11} & -J_{12} \\ -J_{21} & \lambda - J_{22} \end{bmatrix}. \end{aligned}$$

Now computing the determinant gives us

$$|\lambda I - J(\bar{x}, \bar{y})| = \begin{vmatrix} \lambda - J_{11} & -J_{12} \\ -J_{21} & \lambda - J_{22} \end{vmatrix} = (\lambda - J_{11})(\lambda - J_{22}) - J_{12}J_{21}$$

This is the characteristic polynomial in λ . In order to find its roots (the eigenvalues), we set it equal to zero

$$(\lambda - J_{11})(\lambda - J_{22}) - J_{12}J_{21} = 0.$$

Distributing the above terms in the polynomial above we get the following equivalent root-finding problem

$$\lambda^2 - \lambda(J_{11} + J_{22}) + (J_{11}J_{22} - J_{12}J_{21}) = 0.$$

Interesting you might recognize some of the characters in the above equation if you've been doing a lot of linear algebra lately...

$$\lambda^2 - \lambda \underbrace{(J_{11} + J_{22})}_{\text{Trace of } J} + \underbrace{(J_{11}J_{22} - J_{12}J_{21})}_{\text{Determinant of } J} = 0.$$

I'm pointing this out because we might be able to use those later (spoiler: we will use that information in a Theorem in a few moments). For now, we can simply find the roots of the characteristic polynomial using quadratic formula!

Using quadratic formula we find the eigenvalues (roots) are:

$$\lambda = \frac{(J_{11} + J_{22}) \pm \sqrt{(J_{11} + J_{22})^2 - 4(J_{11}J_{22} - J_{12}J_{21})}}{2}.$$

Now depending on the values of the eigenvalues λ determine the stability properties for the specific equilibrium (\bar{x}, \bar{y}) we are investigating, the one we evaluated the Jacobian matrix at above.

Now that we've gone through a familiar recipe for finding the stability properties of equilibrium points in continuous systems (at least coupled systems of 2 equations), let's quickly discuss what the eigenvalues must be for stability.

Some general ideas to notice for two coupled differential equations:

- **Stable** equilibria are always associated with eigenvalues in which all have real part is **less than zero**.
- Whenever a real part of an eigenvalue is **greater than zero** for at least one eigenvalue, the equilibrium will be **unstable**.
- Whenever eigenvalues are complex with non-zero real and imaginary parts, there will be 'spiraling' behavior in the phase plane near the equilibrium.

These ideas are summarized in Figure 2.8. Compare Figure 2.8 (systems of differential equations) to Figure 1.8 (systems of discrete equations). Stability properties of equilibria always depend on the eigenvalues of the Jacobian of the system, evaluated at that specific equilibrium point. However, the conditions on the eigenvalues that determine stability are different depending on whether it is a discrete or continuous (differential equations) system.

Okay, two things - (1) Why did you point out the determinant and trace of J before in Example 2.4.2? And (2), what if we have a system of 3 or more differential equations?

Oh boy, great questions. Let's explore both of those questions. First, it turns out that for a coupled system of 2 differential equations, we actually have a lovely theorem that can guarantee when an equilibrium is locally asymptotically stable that doesn't require us explicitly finding the eigenvalues of the Jacobian matrix, J . Check it out:

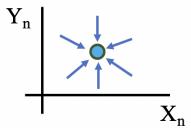
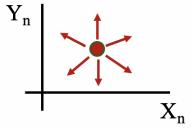
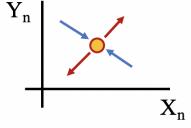
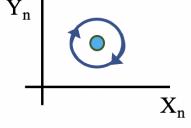
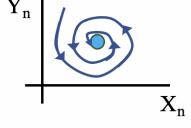
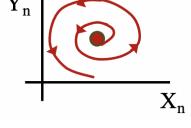
Case	Classification	Name	Phase Plane Depiction
$\lambda_1 < 0, \lambda_2 < 0$	Asymptotically Stable	Sink	
$\lambda_1 > 0, \lambda_2 > 0$	Unstable	Source	
$\lambda_1 < 0, \lambda_2 > 0$	Unstable	Saddle-point	
$\lambda_1 = a + bi, \lambda_2 = a - bi$ ($\text{Re}(\lambda_1) = \text{Re}(\lambda_2) = a = 0$)	Stable (neutrally stable) (not asymptotically stable)	Center	
$\lambda_1 = a + bi, \lambda_2 = a - bi$ ($\text{Re}(\lambda_1) = \text{Re}(\lambda_2) = a < 0$)	Stable	Stable foci	
$\lambda_1 = a + bi, \lambda_2 = a - bi$ ($\text{Re}(\lambda_1) = \text{Re}(\lambda_2) = a > 0$)	Unstable	Unstable foci	

Figure 2.8: Classifying equilibria further than stable or unstable depending on the eigenvalues.

Theorem 4.2

Consider a system of 2 coupled differential equations, such as

$$\begin{aligned}\frac{dx}{dt} &= f(x, y) \\ \frac{dy}{dt} &= g(x, y).\end{aligned}$$

Assume that the partial derivative of $f(x, y)$ and $g(x, y)$ are continuous in an open set containing an equilibrium point (\bar{x}, \bar{y}) . Then the equilibrium point is **locally asymptotically stable** if the trace and determinant of the Jacobian evaluated at that equilibrium point are less than zero and greater than zero, respectively, i.e.,

$$\text{Tr}(J) < 0 \quad \text{and} \quad \det(J) > 0.$$

Note:

- The proof almost writes itself, as we did most of the work above. Recall for the generic coupled system of two differential equations, we got to the following characteristic polynomial:

$$p_\lambda = \lambda^2 - \lambda \left(\underbrace{J_{11} + J_{22}}_{\text{Trace of } J} \right) + \left(\underbrace{J_{11}J_{22} - J_{12}J_{21}}_{\text{Determinant of } J} \right) = \lambda^2 - \text{Tr}(J)\lambda + \det(J),$$

whose roots would then be

$$\lambda = \frac{\text{Tr}(J) \pm \sqrt{[\text{Tr}(J)]^2 - 4 \det(J)}}{2}.$$

- If $\det(J) < 0$, then the eigenvalues are real but of opposite sign. Therefore, it is a saddle point, which is always unstable.
- $0 < \det(J) < \text{Tr}(J)^2/4$, the eigenvalues are real and distinct. The eigenvalues would either be both positive or both negative. Hence if $\text{Tr}(J) < 0$, it would be a stable equilibrium, otherwise if $\text{Tr}(J) > 0$, it would be unstable.
- If $0 < \text{Tr}(J)^2/4 < \det(J)$, the eigenvalues are not real but also not purely imaginary. Hence the phase portrait would show a spiral. It would be stable if $\text{Tr}(J) < 0$ and unstable if $\text{Tr}(J) > 0$.
- These ideas can compactly be summarized in the as the geometric interpretation that Figure 2.9 represents.

Unfortunately, the above theorem only applies for differential equations systems of two coupled equations. Now let's seek out to answer the white elephant in the room question: *what if we have a system of 3 or more differential equations?* Simply put, in order for an equilibrium to be stable **all** eigenvalues of system's Jacobian matrix must either be less than zero themselves or have real parts less than zero. If one eigenvalue of the Jacobian is itself greater than 0 or has a real part greater than 0, it is unstable. These ideas extend naturally from the ideas explored above. However, the phase portraits in more than 2 dimensions may a bit trickier to visualize or interpret.

We can cast these stability ideas/conditions as another mathematical hammer (theorem) for us to use:

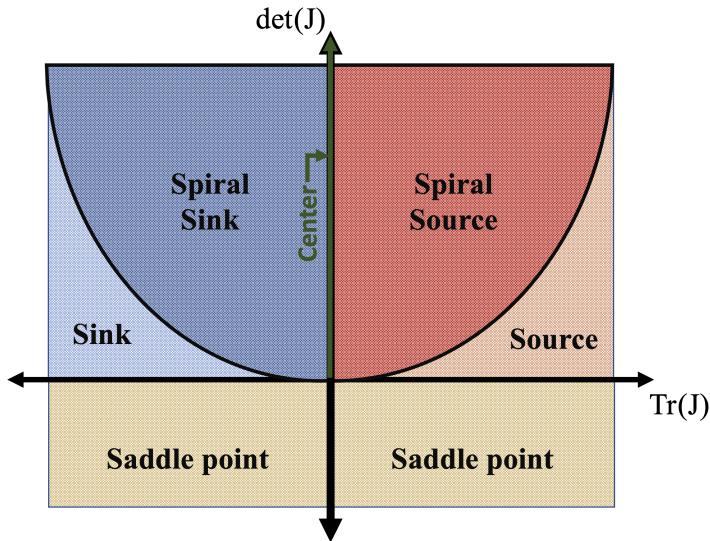


Figure 2.9: Geometric interpretation of the $\text{Tr}(J)$ and $\det(J)$ theorem for stability of equilibria in systems of 2 coupled differential equations.

Theorem 4.3

Consider a system of N -coupled differential equations, such as:

$$\begin{aligned} \frac{dx_1}{dt} &= f_1(x_1, x_2, \dots, x_N) \\ \frac{dx_2}{dt} &= f_2(x_1, x_2, \dots, x_N) \\ &\vdots \\ \frac{dx_N}{dt} &= f_N(x_1, x_2, \dots, x_N). \end{aligned}$$

where $\{f_1, f_2, \dots, f_N\}$ are all differentiable on an open set containing an equilibrium point $(\bar{x}_1, \bar{x}_2, \dots, \bar{x}_N)$. Then if all the eigenvalues of the Jacobian matrix for the system, when it is evaluated at such equilibrium point, are either less than 0 or have real part less than zero, the equilibrium point is **stable**. If one eigenvalue is greater than zero or has real part greater than zero, the equilibrium point is **unstable**.

The proof of Theorem 4.3 we've actually *almost* seen before numerous times in this book. Spoiler: it'll rely on Taylor Series once again and follow a very similar approach to Proof 2.4.1 for stability of equilibrium from individual differential equations, but with the multi-variable flair previously seen in Proof 1.4.1 for stability of equilibrium points in coupled discrete dynamical systems. Alright. Let's power through it!

Proof 2.4.2. (*An outline of a proof for Theorem 4.3*)

Consider a system of N -coupled differential equations, such as:

$$\begin{aligned} \frac{dx_1}{dt} &= f_1(x_1, x_2, \dots, x_N) \\ \frac{dx_2}{dt} &= f_2(x_1, x_2, \dots, x_N) \\ &\vdots \\ \frac{dx_N}{dt} &= f_N(x_1, x_2, \dots, x_N). \end{aligned} \quad \left\{ \Rightarrow \underbrace{\frac{d}{dt} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{pmatrix}}_{\vec{x}} = \underbrace{\begin{pmatrix} f_1(x_1, x_2, \dots, x_N) \\ f_2(x_1, x_2, \dots, x_N) \\ \vdots \\ f_N(x_1, x_2, \dots, x_N) \end{pmatrix}}_{\vec{F}(x_1, x_2, \dots, x_N)} \right.$$

where $\{f_1, f_2, \dots, f_N\}$ are all differentiable on an open set containing an equilibrium point $\bar{\mathcal{Z}} = (\bar{x}_1, \bar{x}_2, \dots, \bar{x}_N)$. Note that the choice to call this point $\bar{\mathcal{Z}}$ was just to avoid any mistakes with too many \mathcal{X} 's floating around. We will now follow similar steps as in Proof 2.4.1 but with a multi-variable twist!

First, we define a "distance" function to give us how far each quantity is away from the equilibrium point,

$$\vec{e} = \vec{x} - \bar{\mathcal{Z}}.$$

We assume that we are in a neighborhood of the equilibrium so $\|\vec{e}\|$ is small. Next, we differentiate both sides with respect to t , and get

$$\frac{d\vec{e}}{dt} = \underbrace{\frac{d\vec{x}}{dt}}_{\vec{F}} - \underbrace{\frac{d\bar{\mathcal{Z}}}{dt}}_0,$$

since $\bar{\mathcal{Z}}$ is a constant vector. That is, since it's an equilibrium point, it's not a function of any of the variables. Hence we are left with

$$\frac{d\vec{e}}{dt} = \frac{d\vec{x}}{dt} = \vec{F}(x_1, x_2, \dots, x_N).$$

Following the same steps as Proof 2.4.1, our next move is to Taylor expand $\vec{F}(x_1, x_2, \dots, x_N)$ about the equilibrium point $\bar{\mathcal{Z}} = (\bar{x}_1, \bar{x}_2, \dots, \bar{x}_N)$. Yay - we get to use a multi-variable Taylor expansion once again!

$$\frac{d\vec{e}}{dt} = \vec{F}(\bar{\mathcal{Z}}) + \underbrace{\nabla \vec{F}(\bar{\mathcal{Z}})}_{\text{Jacobian, } J} \cdot \vec{e} + \frac{1}{2} \vec{e}^T \underbrace{\nabla(\nabla \vec{F}(\bar{\mathcal{Z}}))}_{\text{Hessian, } H} \cdot \vec{e} + \dots$$

Recall that the definition of an equilibrium point for our system implies that

$$\vec{F}(\bar{\mathcal{Z}}) = \vec{F}(\bar{x}_1, \bar{x}_2, \dots, \bar{x}_N) = 0.$$

Therefore the above equation becomes

$$\frac{d\vec{e}}{dt} = J\vec{e} + \frac{1}{2} \vec{e}^T H \vec{e} + \dots$$

Again, invoking that our solution is close enough to the equilibrium point, we can get rid of higher-order terms, meaning the quadratic and higher terms. Thus, we are left with

$$\frac{d\vec{e}}{dt} = J\vec{e}.$$

Hence we are left with a first-order **linear** system of coupled differential equations. The long and the short of it is that all of the solutions to this system will involve "exponentials raised to eigenvalues times t ", i.e., a vector involving combinations of $e^{\lambda_1 t}, e^{\lambda_2 t}, \dots, e^{\lambda_N t}$. Following Proof 2.4.1, if we want our solution to approach the equilibrium as $t \rightarrow \infty$, this implies that we need

$$e^{\lambda_1 t}, e^{\lambda_2 t}, \dots, e^{\lambda_N t} \rightarrow 0 \text{ as } t \rightarrow \infty,$$

in order for $\vec{e}(t) \rightarrow \vec{0}$ as $t \rightarrow \infty$. This can only happen if either $\lambda_j < 0$ or $\operatorname{Re}(\lambda_j) < 0$ for all $j = 1, 2, \dots, N$. The equilibrium would then be **stable**.

On the other hand, if there exist at least one λ_k such that $\lambda_k > 0$ or $\operatorname{Re}(\lambda_k) > 0$, then $\vec{e}(t) \not\rightarrow \vec{0}$ as $t \rightarrow \infty$. Therefore the equilibrium would be **unstable**.

Without getting into all the nuts of bolts of the solution to a system of first-order linear differential equations, the solution appears analogous to that of a single first-order linear differential equation:

	Single	System
Governing Equation	$\frac{dx}{dt} = ax$	$\frac{d\vec{x}}{dt} = J\vec{x}$
Constants	$a \in \mathbb{R}$	$J \in \mathbb{R}^{N \times N}$
Initial Value	$x(0) = x_0$	$\vec{x}(0) = \vec{x}_0$
Solution	$x(t) = x_0 e^{at}$	$\vec{x}(t) = e^{Jt} \cdot \vec{x}_0$

So, where do the eigenvalues come into play if the solution to that system of linear first-order differential equations is: $\vec{x}(t) = e^{At} \cdot \vec{x}_0$? Well, it has to do with matrix decomposition using its eigen-basis. If J is diagonalizable, we could express J in terms of linear combinations of its eigenvectors and eigenvalues. This idea was introduced in Proof 1.4.1 for the stability of equilibrium in systems of discrete equations. We can “decompose” J in terms of its eigen-basis in the following way:

$$J = V\Lambda V^{-1}$$

where we have defined the following matrices:

$$\Lambda = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_M \end{bmatrix} \quad \text{and} \quad V = \begin{bmatrix} | & | & | & | \\ \vec{v}_1 & \vec{v}_2 & \dots & \vec{v}_M \\ | & | & | & | \end{bmatrix},$$

such that Λ is a diagonal matrix with the eigenvalues along the main diagonal and V is a matrix composed of the eigenvectors (in column form).

We can then write our solution to the system of linear first-order differential equations as

$$\vec{x}(t) = e^{Jt} \cdot \vec{x}_0 = e^{V\Lambda V^{-1}t} \cdot \vec{x}_0.$$

How does this help? Here is the really cool part - more Taylor Series! We’re going to expand that exponential term in a Taylor Series... even though its argument involves a product of matrices - just so cool!

$$\begin{aligned} e^{V\Lambda V^{-1}t} &= \underbrace{I}_{\substack{\text{Identity} \\ \text{Matrix}}} + V\Lambda V^{-1}t + \frac{1}{2!}(V\Lambda V^{-1})^2 t^2 + \frac{1}{3!}(V\Lambda V^{-1})^3 t^3 + \dots \\ &= I + V\Lambda V^{-1}t + \frac{1}{2!}(V\Lambda V^{-1})(V\Lambda V^{-1})t^2 + \frac{1}{3!}(V\Lambda V^{-1})(V\Lambda V^{-1})(V\Lambda V^{-1})t^3 + \dots \end{aligned}$$

Now since $I = V^{-1}V = VV^{-1}$, this reduces to

$$\begin{aligned}
e^{V\Lambda V^{-1}t} &= I + V\Lambda V^{-1}t + \frac{1}{2!}(V\Lambda V^{-1})(V\Lambda V^{-1})t^2 + \frac{1}{3!}(V\Lambda V^{-1})(V\Lambda V^{-1})(V\Lambda V^{-1})t^3 + \dots \\
&= I + V\Lambda V^{-1}t + \frac{1}{2!}V\Lambda^2 V^{-1}t^2 + \frac{1}{3!}V\Lambda^3 V^{-1}t^3 + \dots \\
&= VV^{-1} + V\Lambda V^{-1}t + \frac{1}{2!}V\Lambda^2 V^{-1}t^2 + \frac{1}{3!}V\Lambda^3 V^{-1}t^3 + \dots
\end{aligned}$$

If we pull out a V from the left side and a V^{-1} from the right this reduces further to

$$\begin{aligned}
e^{V\Lambda V^{-1}t} &= VV^{-1} + V\Lambda V^{-1}t + \frac{1}{2!}V\Lambda^2 V^{-1}t^2 + \frac{1}{3!}V\Lambda^3 V^{-1}t^3 + \dots \\
&= V \underbrace{\left[I + \Lambda t + \frac{1}{2!}\Lambda^2 t^2 + \frac{1}{3!}\Lambda^3 t^3 + \dots \right]}_{e^{\Lambda t}} V^{-1} \\
&= Ve^{\Lambda t}V^{-1}.
\end{aligned}$$

Therefore, we can now see that in order for $\vec{e}(t) \rightarrow \vec{0}$ as $t \rightarrow \infty$, we need that $e^{Jt} = Ve^{\Lambda t}V^{-1} \rightarrow \vec{0}$ as $t \rightarrow \infty$. In order for that to happen, all λ_{jj} along the main diagonal of Λ must either be less than zero or have real part less than zero.

To make sure we're all on the same page, let's go through an explicit example together! While we are going through this example, keep in mind the similarities, but also the differences between how we approach equilibria and stability for coupled differential equations versus coupled discrete equations.

Example 2.4.3. Equilibria and Stability in a System of Differential Eqns.

Consider the following coupled system of differential equations

$$\begin{aligned}
\frac{dx}{dt} &= (x - a)(b - y) \\
\frac{dy}{dt} &= cxy
\end{aligned}$$

with $a, b, c \in \mathbb{R}^+$ and initial values $x(0) = x_0$ and $y(0) = y_0$.

First, to find the equilibria point (\bar{x}, \bar{y}) , we substitute \bar{x} and \bar{y} into the right hand side of the differential equations and set them equal to zero:

$$\begin{aligned}
(\bar{x} - a)(b - \bar{y}) &= 0 \\
c\bar{x}\bar{y} &= 0
\end{aligned}$$

From the second equation we see that either $\bar{x} = 0$ OR $\bar{y} = 0$. That is, either \bar{x} or \bar{y} MUST be zero. Therefore, we will consider each case separately to find what the other value must be to satisfy the top equation, either \bar{y} or \bar{x} , respectively.

1. Assuming $\bar{x} = 0$:

If we assume that $\bar{x} = 0$, we find that the other equation then becomes

$$(\bar{x} - a)(b - \bar{y}) = (0 - a)(b - \bar{y}) = -a(b - \bar{y}) = 0.$$

Therefore, we see that \bar{y} must be b for this equation to be satisfied, i.e.,

$$\bar{y} = b,$$

and hence one equilibrium point is:

$$(\bar{x}, \bar{y}) = (0, b).$$

2. Assuming $\bar{y} = 0$:

If we assume that $\bar{y} = 0$, we find that the other equation then becomes

$$(\bar{x} - a)(b - \bar{y}) = (\bar{x} - a)(b - 0) = b(\bar{x} - a) = 0.$$

Therefore, we see that \bar{x} must be a for this equation to be satisfied, i.e.,

$$\bar{x} = a,$$

and hence the other equilibrium point is:

$$(\bar{x}, \bar{y}) = (a, 0).$$

Now we have our two equilibrium points for dynamical system:

$$(\bar{x}, \bar{y}) \in \{(0, b), (a, 0)\}.$$

In order to begin the process for determining the stability of these points, we recall a familiar recipe:

1. Compute the Jacobian matrix, $J(x, y)$

2. Evaluate the Jacobian matrix at the Equilibrium point, $J(\bar{x}, \bar{y})$.

3. Find the eigenvalues of the Jacobian matrix. Depending on whether the absolute value (if real) or absolute value of the real part (if complex) are less than or greater than 1 will determine stability (see Figure 2.8).

Let's roll through this recipe! First, we'll compute the Jacobian matrix for our system. If it helps, we can rewrite our dynamical system to give function names to each right hand side,

$$\begin{aligned}\frac{dx}{dt} &= f(x, y) = (x - a)(b - y) \\ \frac{dy}{dt} &= g(x, y) = cxy.\end{aligned}$$

Now to compute the Jacobian matrix, we find all the first-order partial derivatives of $f(x, y)$ and $g(x, y)$ with respect to x and y and order them in a matrix in the following manner:

$$J = \begin{bmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \\ \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} \end{bmatrix}.$$

Performing those partial derivatives and substituting them into that matrix, we will obtain

$$\mathbf{J} = \begin{bmatrix} b - y & -(x - a) \\ cy & cx \end{bmatrix}.$$

At this point, we get to determine the stability of each equilibrium individually. Let's jump into both equilibrium cases.

Equilibria: $(0, b)$

Evaluating the Jacobian at the equilibrium $(0, b)$ gives:

$$\mathbf{J}(0, b) = \begin{bmatrix} 0 & a \\ bc & 0 \end{bmatrix}.$$

Next we will begin the process of finding the eigenvalues of $\mathbf{J}(0, b)$. To do this, we take the determinant of $\lambda I - \mathbf{J}(0, b)$ to get the characteristic polynomial of the system. The roots of this polynomial are our eigenvalues!

$$\det(\lambda I - \mathbf{J}(0, b)) = \begin{vmatrix} \lambda & -a \\ -bc & \lambda \end{vmatrix} = \lambda^2 - abc.$$

Setting the characteristic polynomial equal to 0 gives:

$$\lambda^2 - abc = 0 \Rightarrow \lambda = \pm\sqrt{abc}.$$

Hence the eigenvalues are $\lambda = \{-\sqrt{abc}, \sqrt{abc}\}$. Therefore, this equilibrium is a **saddle-point** and hence **unstable** because one eigenvalue is positive and the other is negative (see Figure 2.8).

Equilibria: $(a, 0)$

Evaluating the Jacobian at the equilibrium $(a, 0)$ gives:

$$\mathbf{J}(a, 0) = \begin{bmatrix} b & 0 \\ 0 & ac \end{bmatrix}.$$

Next we will begin the process of finding the eigenvalues of $\mathbf{J}(a, 0)$. Again, to do this, we take the determinant of $\lambda I - \mathbf{J}(a, 0)$ to get the characteristic polynomial for the system. The roots of the characteristic polynomial are our eigenvalues!

$$\det(\lambda I - \mathbf{J}(a, 0)) = \begin{vmatrix} \lambda - b & 0 \\ 0 & \lambda - ac \end{vmatrix} = (\lambda - b)(\lambda - ac).$$

Setting the characteristic polynomial equal to 0 gives:

$$(\lambda - b)(\lambda - ac) = 0 \Rightarrow \lambda = b \text{ and } \lambda = ac.$$

There since both eigenvalues are positive (remember our assumption that $a, b, c > 0$ at the beginning of the problem?), this equilibrium is **unstable** for any combination of parameters (again, assuming they are positive)!

2.5 Nullclines and Phase Planes

Let's take a quick journey into exploring phase planes in a bit more depth. To paint the picture a bit more, consider two coupled differential equations, such as

$$\begin{cases} \frac{dx}{dt} = f(x, y) \\ \frac{dy}{dt} = g(x, y) \end{cases} \quad (2.6)$$

Now if we were to solve such a system numerically and plot a phase plane, one such possible phase plane could look the curve given in Figure 2.10. Note that on the phase plane we are showing x on the horizontal axis and y on the vertical axis. That is, the curve given on Figure 2.10 is actually a *parametric curve*, like those studied in Calculus B. Therefore, as the independent variable t changes, we move to different x and y positions along the curve. On that note the curve has a particular direction along it. Now from our knowledge of Equilibria and Stability it would appear that this phase-plane is showing a stable spiral (*stable foci*) based on our equilibrium classification from Figure 2.8.

The reason we are reminding ourselves of this is to pull in some of the ideas that we developed in Calculus B regarding parametric curves - namely the slope along one. And, furthermore we can view our solutions to a coupled differential equation such as those from (2.6) as a parametric curve representation on our phase plane. From (2.6), we would get two solutions $\rightarrow x(t)$ and $y(t)$ that satisfy those differential equations for some given initial values $x(0) = x_0$ and $y(0) = y_0$. Therefore, we can plot those solutions as a parametric curve $(x(t), y(t))$. Again, as mentioned previously above, we can view the independent variable t as the parameter that moves us to different positions along this curve. Now that we have a parametric curve, we can compute the slope of the curve at different specified points, quantified by different t values.

From Calculus B, for a parametric curve $(x(t), y(t))$, we saw that we could compute the slope at any point along the curve by computing the following:

$$\frac{dy}{dx} = \frac{\frac{dy}{dt}}{\frac{dx}{dt}} = \text{slope ('change in } y \text{ over change in } x\text')} \quad (2.7)$$

Note that we are still computing a slope via the old adage \rightarrow *the slope is given by the change in y over change in x* . Now, dissecting this a bit more, what we are really saying in the *change in the vertical direction over change in the horizontal direction* so we are not restricted to always x and y variables. Anywho, looking back at (2.7), we know what to substitute into those time derivatives \rightarrow use can simply substitute the right hand sides of our differential equations from (2.6) above! Thus, whenever computing the slope of the our solution curve from a phase plane, we don't actually need to compute any derivatives... they are immediately given at the start of the problem! Yay!

So for us here, our slope can be compute via

$$\begin{cases} \frac{dx}{dt} = f(x, y) \\ \frac{dy}{dt} = g(x, y) \end{cases} \Rightarrow \frac{dy}{dx} = \frac{\frac{dy}{dt}}{\frac{dx}{dt}} = \frac{g(x, y)}{f(x, y)}.$$

Now we can infer a lot of information about the phase plane solution trajectory $(x(t), y(t))$ of (2.6) by checking out the slope using this parametric calculus machinery. For example, on our phase plane, we know that the solution curve $(x(t), y(t))$ will have a slope of 0 ($dy/dx = 0$), and hence have a horizontal tangent line, if

$$g(x, y) = 0 \quad \text{and} \quad f(x, y) \neq 0,$$

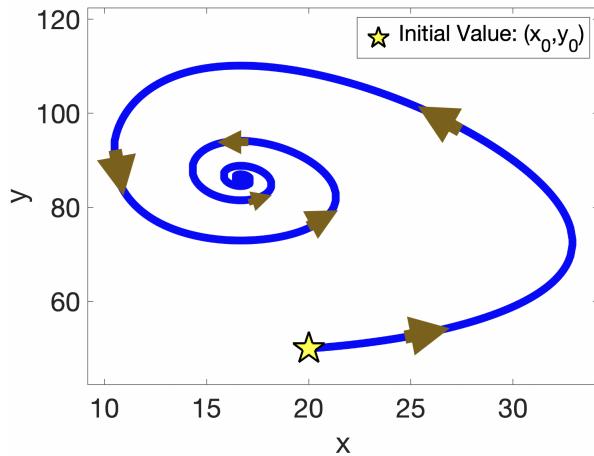


Figure 2.10: An illustrative example of what one plausible phase plane trajectory could look like, given by the solution curves $(x(t), y(t))$.

that is the numerator of (2.7) is 0 but the denominator is not zero. On the other hand if the denominator is zero and the numerator is non-zero then the solution curve $(x(t), y(t))$ has infinite slope ($dy/dx \rightarrow \infty$), i.e.,

$$f(x, y) = 0 \quad \text{and} \quad g(x, y) \neq 0.$$

Let's abstract some of the ideas a bit more that were just touched upon. If we tie back some of these ideas to differential equations, we can note the following:

1. $f(x, y)$ determines the motion in the x direction at a location (x, y) in the phase plane. That is, it governs how the **horizontal** solution's trajectory $x(t)$ changes.
2. $g(x, y)$ determines the motion in the y direction at a location (x, y) in the phase plane. That is, it governs how the **vertical** solution's trajectory $y(t)$ changes.

We are at a point in which we can introduce a new definition to help us better understand these ideas - **nullclines**.

Definition 5.1

Consider a differential equation system such as

$$\begin{cases} \frac{dx}{dt} = f(x, y) \\ \frac{dy}{dt} = g(x, y) \end{cases}$$

We can define **nullclines** in both the horizontal and vertical directions to be a collection of points in space in which we know the solution must be moving strictly only vertical or horizontal, respectively. More specifically, consider them separately:

1. The **horizontal-nullcline** is a set of points in the phase plane such that $\frac{dx}{dt} = 0$. Geometrically, these are the points where the solution is strictly moving either straight up or straight down. Algebraically, we find these nullclines by solving

$$f(x, y) = 0.$$

2. The **vertical-nullcline** is a set of points in the phase plane such that $\frac{dy}{dt} = 0$.

Geometrically, these are the points where the solution is strictly moving either left or right. Algebraically, we find these nullclines by solving

$$g(x, y) = 0.$$

Note:

- A nullcline exists whenever one of the differential equations is equal to zero
- Thus, when setting the differential equations equal to zero, we get independent relationships between the dependent variables for each nullcline. These give rise to the *nullcline curves* for either horizontal or vertical nullclines.
- The direction along the nullcline curves are given by the sign of the other function being evaluated along the curve! For example, along the horizontal nullcline in which $f(x, y) = 0$, we know the solution is moving either straight up or straight down by evaluating $g(x, y)$ along the points where $f(x, y) = 0$.
- The point in which vertical and horizontal nullclines intersect are **equilibrium points!**
- The nullclines subdivide the phase plane into regions in which different dynamics arise. Since nullclines are defined by derivatives being equal to zero, they can create barriers in which a solution trajectory cannot cross, depending on the flow along the nullcline!

The idea of nullclines might seem a bit abstract, but they can help us divide the phase plane into regions in which we expect to see different solution behavior. There might be a few moving parts (pun intended) but they all involve a mixture of tools from Calculus A and B, just used in a slightly different way than what we may be accustomed to. To solidify these ideas, let's all hop into an example together and go through each aspect piece by piece ☺.

Example 2.5.1. Nullclines Example!

This example will illustrate the following:

1. How to compute nullcline curves and how to plot them
2. The intersection of horizontal and vertical nullclines is where equilibrium occur
3. How to determine the flow direction along a nullcline
4. Putting it all together!

Consider the following coupled differential equations system:

$$\begin{cases} \frac{dx}{dt} = y(x - 1) \\ \frac{dy}{dt} = x(2 - y) \end{cases}, \quad (2.8)$$

so based on our previous notation $f(x, y) = y(x - 1)$ and $g(x, y) = x(2 - y)$. Let's go through this analysis step-by-step.

1. **Computing the nullclines:**

- **Horizontal Nullcline:**

To compute the horizontal nullcline, we need set $f(x, y) = y(x - 1) = 0$. Therefore we find that there are two curves that give rise to this:

$$y = 0 \quad \text{and} \quad x = 1.$$

Let's plot these horizontal nullclines in Figure 2.11a. Recall that these horizontal nullclines imply that the solution does not change horizontally along these nullcline curves. Therefore, the solution may only change vertically, i.e., either straight up or straight down.

- **Vertical Nullcline:**

To compute the vertical nullcline, we need set $g(x, y) = x(2 - y) = 0$. Therefore we find that there are two curves that give rise to this:

$$x = 0 \quad \text{and} \quad y = 2.$$

Let's plot these vertical nullclines in Figure 2.11b. These vertical nullclines imply that the solution does not change vertically along these nullcline curves. Along the vertical nullclines, the solution may only change horizontally, i.e., either left or right.

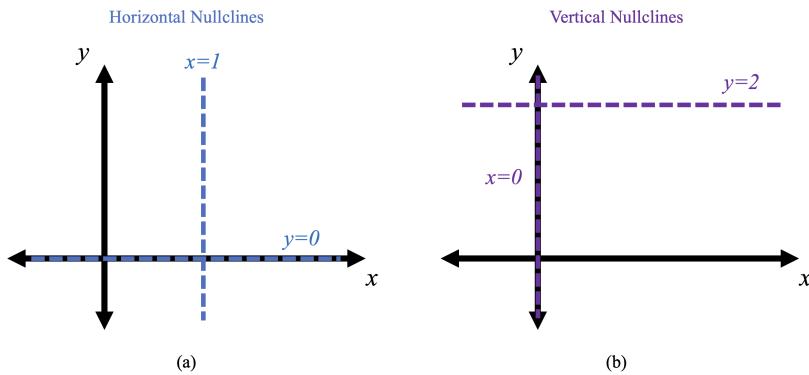


Figure 2.11: The (a) horizontal and (b) vertical nullclines for the example given in (2.8).

2. **Determine the equilibrium points:**

The equilibrium points are where the horizontal and vertical nullclines intersect each other. Figure 2.12 illustrates the two equilibrium points that arise for the system given in (2.8).

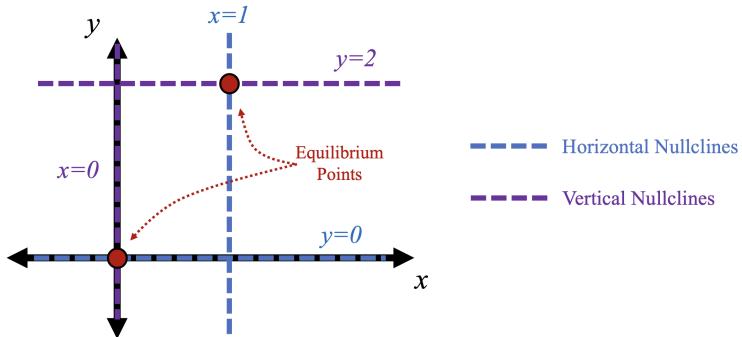


Figure 2.12: The equilibrium points for (2.8) arise where the horizontal and vertical nullclines intersect.

3. Determining the flow direction along nullclines:

The main idea here is that we are wondering what happens on the nullclines themselves. For example, if we are considering the **horizontal nullcline** (when $\frac{dx}{dt} = f(x, y) = 0$), we are wondering what the flow direction will be on the nullcline itself. If we are considering the horizontal nullcline, the only way the flow direction could point is either straight up or straight down along the horizontal nullcline. The flow direction is then determined by looking at the sign of $\frac{dy}{dt} = g(x, y)$. That is when $f(x, y) = 0$, if $g(x, y) > 0$ the flow goes up, while if $g(x, y) < 0$ the flow goes downward.

Let's go ahead and consider each case separately - what are the flow directions along each horizontal nullcline and what happens along each vertical nullcline.

(a) Flow direction along Horizontal Nullclines:

Again, these horizontal nullclines are found by setting $\frac{dx}{dt} = f(x, y) = 0$, which gives rise to horizontal nullclines when either $y = 0$ or $x = 1$.

Therefore, to determine the flow direction along these nullclines, we must evaluate $\frac{dy}{dt} = g(x, y)$ along these horizontal nullcline curves. Recall the flow direction must be in the vertical direction along these nullclines, either straight up or straight down, since we are considering the case in which $\frac{dx}{dt} = f(x, y) = 0$. If think back in terms of the overall slope from (2.7), this then implies that $\frac{dy}{dx} \rightarrow \pm\infty$ since $\frac{dx}{dt} = 0$.

Let's consider each of these two nullclines separately:

$$x = 1:$$

$$g(x, y) = x(2 - y)$$

$$= (1)(2 - y)$$

$$= 2 - y$$

Therefore, next we want to explicit look at the different sign possibilities of the above function $g(1, y) = 2 - y$. Hence we see that:

$$\text{if } y < 2 \Rightarrow g(1, y) > 0$$

$$\text{if } y > 2 \Rightarrow g(1, y) < 0$$

$$\text{if } y = 2 \Rightarrow g(1, y) = 0.$$

Therefore along the horizontal nullcline associated with $x = 1$, we can draw flow direction arrows given by the above inequalities. These are shown on Figure 2.13

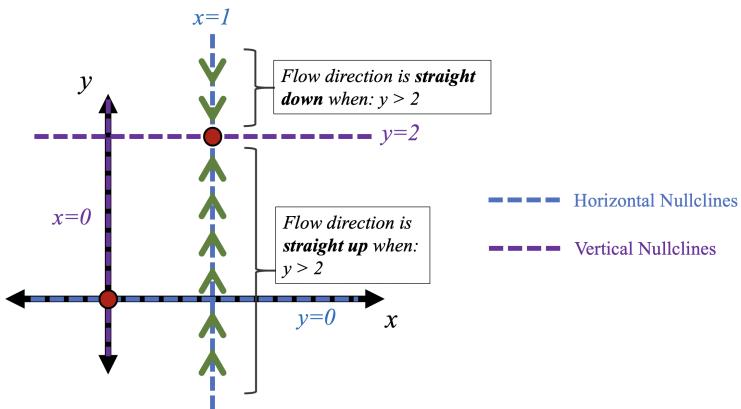


Figure 2.13: The flow direction along the horizontal nullcline associated with $x = 1$.

$$y = 0:$$

$$g(x, y) = x(2 - y)$$

$$= x(2 - 0)$$

$$= 2x$$

Therefore, next we want to explicit look at the different sign possibilities of the above function $g(x, 0) = 2x$. Hence we see that:

$$\text{if } x < 0 \Rightarrow g(x, 0) > 0$$

$$\text{if } x > 0 \Rightarrow g(x, 0) > 0$$

$$\text{if } x = 0 \Rightarrow g(x, 0) > 0.$$

Therefore along the horizontal nullcline associated with $y = 0$, we can draw flow direction arrows given by the above inequalities. These are shown on Figure 2.14

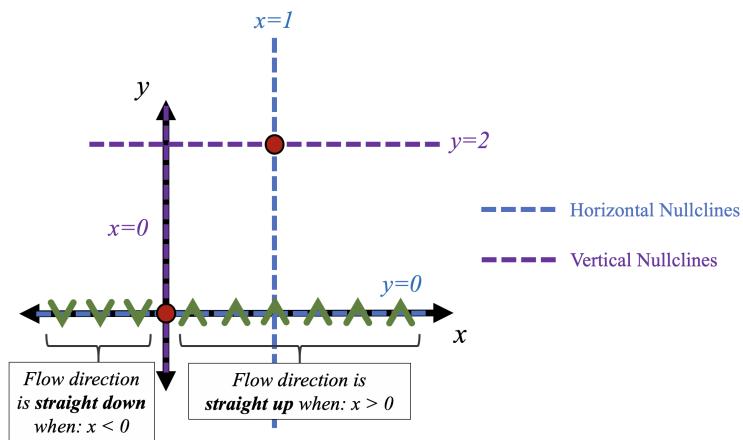


Figure 2.14: The flow direction along the horizontal nullcline associated with $y = 0$.

Now we can put **both** flow directions together for each horizontal nullcline and produce a plot like Figure 2.15.

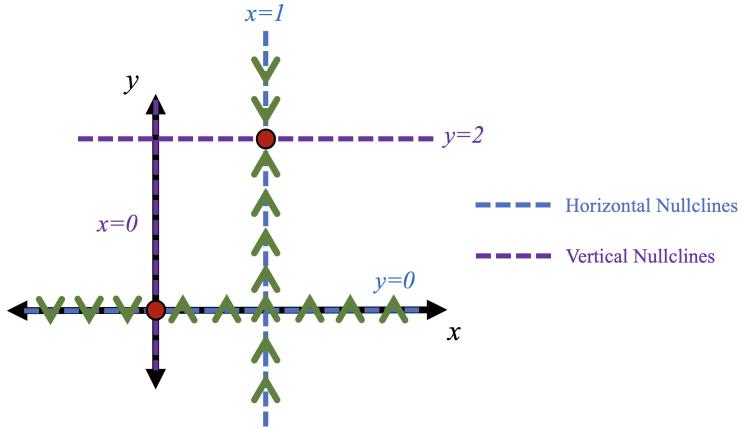


Figure 2.15: The flow direction along each horizontal nullcline associated with $x = 1$ and $y = 0$.

(b) *Flow direction along Vertical Nullclines:*

To quickly recap, these vertical nullclines are found by setting $\frac{dy}{dt} = g(x, y) = 0$, which gives rise to vertical nullclines in this example when either $x = 0$ or $y = 2$.

Therefore, to determine the flow direction along these nullclines, we must evaluate $\frac{dx}{dt} = f(x, y)$ along these vertical nullcline curves. Recall the flow direction must be in the horizontal direction along these nullclines, either left or right only, since we are considering the case in which $\frac{dy}{dt} = g(x, y) = 0$. If think back in terms of the overall slope from (2.7), this then implies that $\frac{dy}{dx} = 0$ since $\frac{dy}{dt} = 0$.

Let's consider each of these two nullcline curves separately:

$$x = 0:$$

$$f(\textcolor{violet}{x}, y) = y(\textcolor{violet}{x} - 1)$$

$$= y(\textcolor{violet}{0} - 1)$$

$$= -y$$

Therefore, next we want to explicit look at the different sign possibilities of the above function $f(\textcolor{violet}{0}, y) = -y$. Hence we see that:

$$\text{if } y < 0 \Rightarrow f(\textcolor{violet}{x}, y) > 0$$

$$\text{if } y > 0 \Rightarrow f(\textcolor{violet}{x}, y) < 0$$

$$\text{if } y = 0 \Rightarrow f(\textcolor{violet}{x}, y) = 0.$$

Therefore along the vertical nullcline associated with $\textcolor{violet}{x} = 0$, we can draw flow direction arrows given by the above inequalities. These are shown on Figure 2.16a.

$y = 2$:

$$f(x, \textcolor{violet}{y}) = \textcolor{violet}{y}(x - 1) = 2(x - 1)$$

Therefore, next we want to explicit look at the different sign possibilities of the above function $f(x, \textcolor{violet}{y}) = 2(x - 1)$. Hence we see that:

$$\text{if } x < 1 \Rightarrow f(x, \textcolor{violet}{y}) < 0$$

$$\text{if } x > 1 \Rightarrow f(x, \textcolor{violet}{y}) > 0$$

$$\text{if } x = 0 \Rightarrow f(x, \textcolor{violet}{y}) > 0.$$

Therefore along the vertical nullcline associated with $\textcolor{violet}{y} = 2$, we can draw flow direction arrows given by the above inequalities. These are shown on Figure 2.16b.

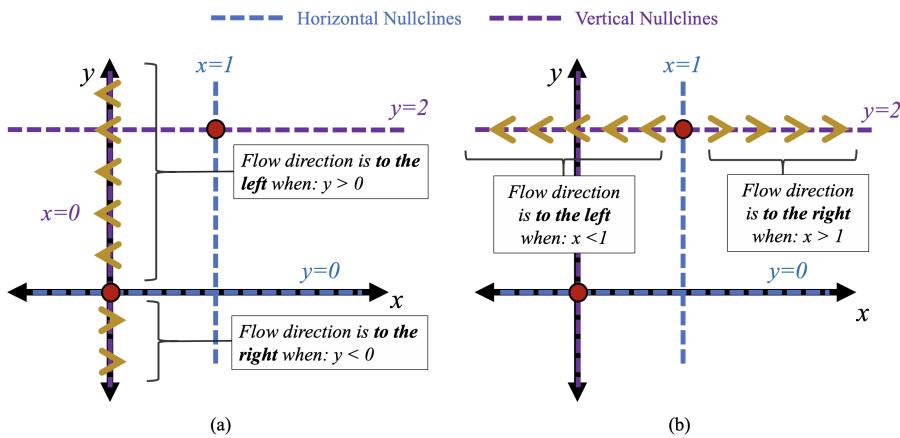


Figure 2.16: The flow direction along the vertical nullcline associated with (a) $\textcolor{violet}{x} = 0$ and (b) $\textcolor{violet}{y} = 2$.

(c) **Putting it all together:**

We can now deduce the flow direction along each nullcline, whether a vertical or horizontal nullcline. Putting together all of the information from Figures 2.15 and 2.16, we can depict the flow directions in the following manner:

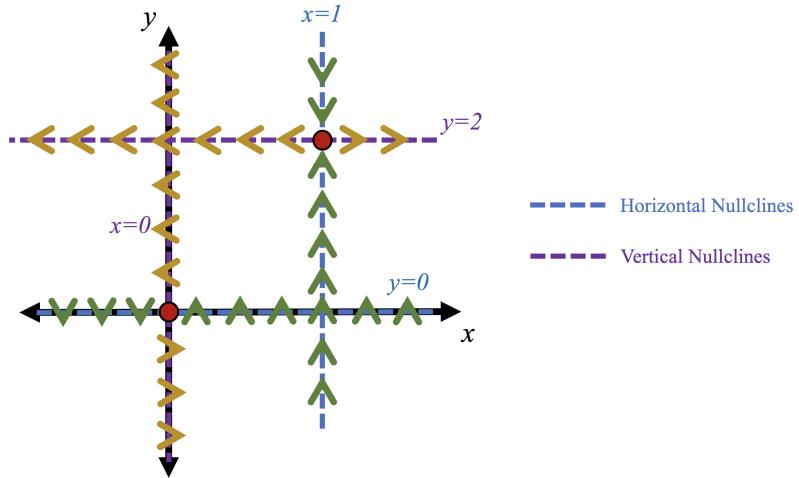


Figure 2.17: The flow direction along each horizontal nullcline and vertical nullcline.

Now that we have the flow direction along each nullcline, we can see where a point would move within the phase plane (xy -plane), as dictated by the flow directions. Figure 2.18 gives such possible solution trajectories from a particular initial value in the xy -plane. The solution trajectories follow the nullcline flow directions.

Based on Figure 2.18, we can determine that the equilibrium point at $(x, y) = (1, 2)$ must be unstable since there are both flow directions moving towards and away from the equilibrium point. Therefore the equilibrium at $(x, y) = (1, 2)$ appears to be saddle point.

Similarly, we can try to classify the other equilibrium based on the flow trajectories. In this case, the $(x, y) = (0, 0)$ equilibrium appears to be either some sort of spiral sink or source, or even possibly a center based on the flow trajectories surrounding it. Performing the standard analytical stability analysis (the Jacobian matrix and eigenvalue story) will show that it is actually, in fact, a center.

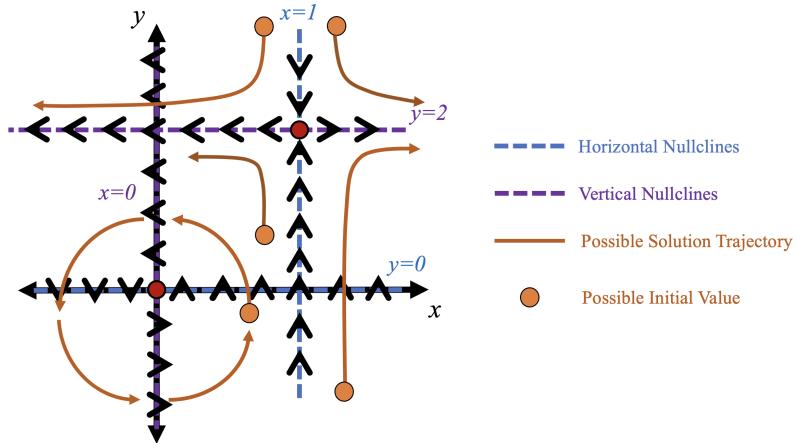


Figure 2.18: Plausible solution trajectories from initial values placed around the xy -plane, as determined by the flow directions along the nullclines.

2.6 Holling-Type Interactions Between 2 Species

Species-species interactions play a vital role in every ecosystem. Perhaps the first way we think about interacting species is through predator-prey type relationships, or varying **trophic levels**. A trophic level is the position in which an organism sits in the food web. The varying trophic levels go from the producers (green plants), followed by the second-level organisms that eat the plants (herbivores), then the third-level of primary carnivores that eat the herbivores, and lastly the fourth-level of secondary carnivores who eat the primary carnivores. These levels are not strictly defined; some organisms may depend on several trophic levels for sustenance. Moreover, another trophic level contains organisms like bacteria or fungi in which break down wastes and dead organisms into nutrients that are useful by the bottom - the producers. Biodiversity plays a vital role in the health of an ecosystem at every trophic level [1]

The Lotka-Volterra equations were first discussed in the early 20th century to describe predator-prey dynamics. This story begins in 1910 when Alfred Lotka wrote down equations for describing a chemical reaction [2], although his system was mostly the Logistic equation finding a new application. Later in 1920, Lotka extended his equations with the help of Andrey Kolmogorov to describe what they referred to as an, "organic system". For them this was an example involving a plant and herbivore [3].

Independently around the same time, Vito Volterra published the same set of equations in the context of mathematical biology [4]. While Lotka was inspired from understanding chemical processes, Volterra became interested in species interactions by his soon-to-be son-in-law, Umberto D'Ancona, who was a marine biologist. D'Ancona was interested in fishery science and was perplexed as to why an unexpectedly high percentage of all fish caught were predatory fish during the years of World War I, even though fishing effort was heavily reduced during the war years. Together, Volterra and D'Ancona developed the same system of equations as Lotka and Kolmogorov to help them understand such dynamics. On a side note, Volterra resigned from his university position and forfeited his membership in a number of scientific academies in Italy in the 1940s when he refused

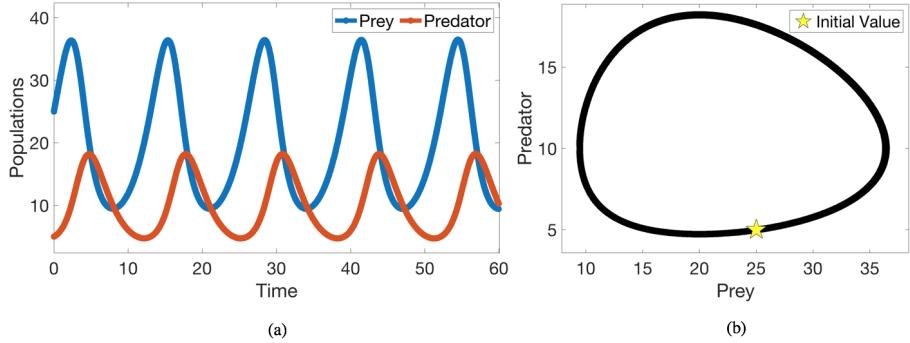


Figure 2.19: A possible periodic solution to the Lotka-Volterra equations when $\alpha = 0.5$, $\beta = 0.05$, $\delta = 0.025$, and $\gamma = 0.5$ with initial values $x(0) = 25$ and $y(0) = 5$.

to sign the oath of allegiance to his at-the-time fascist controlled government regime [5].

What became known as the Predator-Prey equations are given as the following system of differential equations

$$\begin{cases} \frac{dx}{dt} = \alpha x - \beta xy \\ \frac{dy}{dt} = \delta xy - \gamma y, \end{cases} \quad (2.9)$$

where x and y are the prey and predator populations, respectively, and all the parameters $\alpha, \beta, \delta, \gamma > 0$. Each of the terms of this model can be described in the following way:

- αx : rate in which the prey population grows from births
- βxy : reduction rate in the prey population due to being consumed
- δxy : rate of increase in the predator population due to consuming prey
- γy : reduction rate in the predator population due to natural deaths.

As an example, Figure 2.19 shows periodic solutions to the system above when $\alpha = 0.5$, $\beta = 0.05$, $\delta = 0.025$, and $\gamma = 0.5$ with initial values $x(0) = 25$ and $y(0) = 5$. You can notice that the predator population lags behind the prey population when it is growing. However, once the predator population grows too large, the prey population starts collapsing, which then in turn causes the predator population to decrease, i.e., less food to sustain it.

However, since the 1920s many ecologists have embraced the initial ideas of Lotka, Kolmogorov, Volterra, and D'Ancona and extended their models for other types of species interactions, such as competition, parasitism, mutualism, and commensalism. Moreover, in the late 1950s, only two years after earning his Doctorate in Ecology, Crawford Stanley (Buzz) Holling mathematically formalized different functional responses for predator-prey interactions [6], [7], [8]. These functional responses describe how predation may differ based upon the available prey in a system. His logic was that if there are a large number or small number of prey in the system, that would effect predation dynamics.

Let's first begin with a skeleton version of predator-prey reactions based upon Lotka-Volterra, where we leave the interaction term as general as possible:

$$\begin{cases} \frac{dx}{dt} = \alpha x - f(x, y) \\ \frac{dy}{dt} = g(x, y) - \gamma y, \end{cases} \quad (2.10)$$

Holling's idea was to investigate possible mathematical ways to write $f(x, y)$ (or $g(x, y)$) based upon how many prey there are. He developed three different possible functional responses. Let's take a look at each of those now.

Type I: The Holling-Type I response is actually we have been using for interactions up to this point, whether for discrete or continuous equations, i.e.,

$$f(x, y) = axy = (\textcolor{red}{ax})y,$$

where the term $(\textcolor{red}{ax})$ is what we call the functional response. It describes the prey capture rate by the predator and is inherently a function of how many prey there are. We can note the following dynamics:

1. When there are minimal prey, the capture rate is near zero.
2. When the number of preys grow from zero, the capture rate linearly increases.
3. As the number of prey goes to infinity, the capture rate continues increasing towards infinity as well.

However, the way in which Holling envisioned this term was that the capture rate would only grow linearly until a particular threshold value for prey, and then it would stay at one saturation level, see Figure 2.20.

Type II: The Holling-Type II response introduces a new dynamic in which the prey capture rate will saturate when the number of prey continues to increase. It is represented by the following term:

$$f(x, y) = \frac{axy}{1 + bx^2} = \left(\frac{\textcolor{red}{ax^2}}{1 + \textcolor{red}{bx}} \right) y, \quad (2.11)$$

where the term $\left(\frac{\textcolor{red}{ax}}{1 + \textcolor{red}{bx}} \right)$ is the Type II functional response. The terms a and b can be thought of as:

- a : the attack time, searching efficiency, or capture efficiency
- b : the “handling time”, which could be the time it takes to attack, consume, and digest the prey, or a learning rate associated with predation.

It gives rise to the following dynamics:

1. When there are minimal prey, the capture rate is near zero.
2. When the number of prey grows from zero, the capture rate linearly increases.
3. However, in contrast to the Type I response, as the number of prey goes to infinity, the capture rate approaches $\frac{a}{b}$.

This Type II response is illustrated in Figure 2.20.

Type III: The Holling-Type III response is similar to the Type II response in that for a large number of prey, the capture rate saturates; however, it displays different dynamics for lower numbers of prey. It is represented by the following term:

$$f(x, y) = \frac{axy^2}{1 + bx^2} = \left(\frac{\textcolor{red}{ax^2}y}{1 + \textcolor{red}{bx^2}} \right) y, \quad (2.12)$$

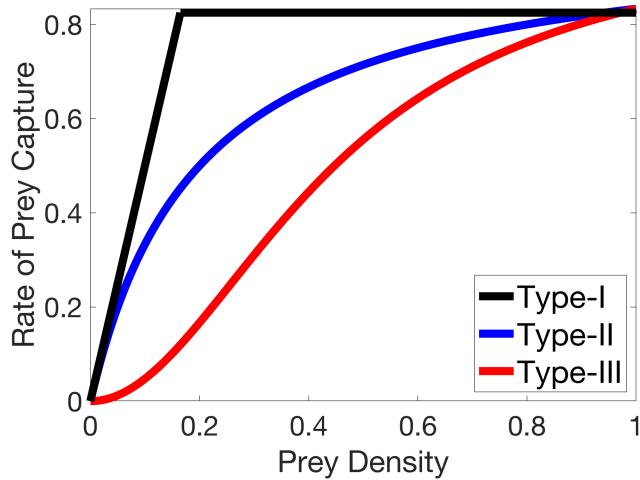


Figure 2.20: A comparison between the three Holling-Type Functional Response terms. These terms illustrate how the prey capture rate varies for differing levels of prey in the system.

where the term $\left(\frac{ax^2}{1+bx^2}\right)$ is the Type III functional response. It gives rise to the following dynamics:

1. When there are minimal prey, the capture rate is near zero.
2. When the number of prey grows from zero, the capture rate slowly increases from zero, in contrast to either the Type I or II responses.
3. Similar to the Type II response, as the number of prey goes to infinity, the capture rate approaches $\frac{a}{b}$.

This Type III response is illustrated in Figure 2.20.

Note: In chemical reaction contexts, particularly in biochemistry or pharmacology, the above Holling-type interactions go by a different name - [Hill functions](#) [9], [10]. The Hill function still represents a response curve for a particular interaction event. However, in the context of biochemistry and pharmacology, it tends to refer to event involving receptor proteins being bound by the ligand. The conceptual idea remains the same in which saturation occurs when there is too high of a ligand binding concentration.

However, unlike commonly used Holling-types described above, the Hill function tends to generalize the nonlinear power in the response function, i.e., rather than only $f(x,y) = \frac{ax^1y}{1+bx^1}$ or $f(x,y) = \frac{ax^2y}{1+bx^2}$, we consider a power of α :

$$f(x,y) = \frac{ax^\alpha y}{1 + bx^\alpha}, \quad (2.13)$$

where α is known as the [Hill coefficient](#). Moreover, in this context of biochemistry, $f(x,y)$ represents the fraction of receptor proteins being bound by ligand molecules for a particular concentration of unbound ligands. Varying α can give rise to different *binding curves*, like the one found in Figure

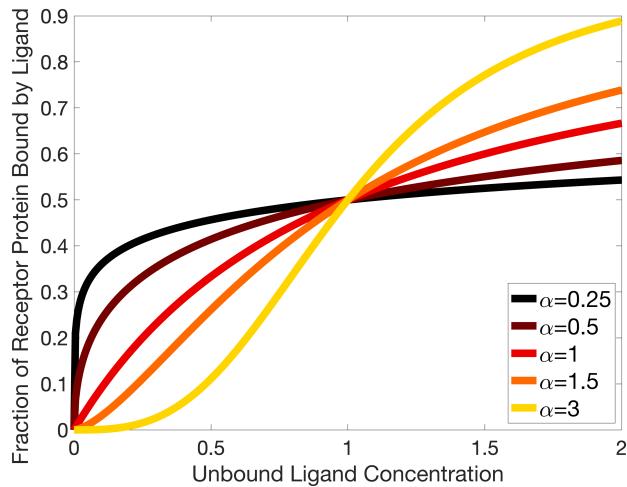


Figure 2.21: An illustration of the effect of varying α on the *binding curves* arising from (2.13). Smaller α result in faster saturation.

2.21. The Hill coefficient is a tuning parameter that helps estimate what concentration of ligand molecules are necessary to bind the receptor bind for a particular saturation level 11. Furthermore, the Hill equation, like that in (2.13) is useful for constructing dose-response curves in pharmacology.

We note that we can easily use this idea of a Hill function to model ecological interactions in the same way that Holling-Type functions were used. The important thing to recognize is that whether you use Holling-Type II or III functional responses or Hill functions, you are incorporating aspects of saturation effects into your modeling framework. Moreover, sometimes people refer to Holling-Type III function responses as Hill functions interchangeably, where they include a power of α rather than a specific power of $\alpha = 2$, like we had before in (2.12) above. Either way, the functions in which incorporate such saturation effects are commonly referred to as **sigmoidal curves**.

Sigmoidal curves have the common feature in that for large enough values of the independent variable, the curve flattens out. In fact, sigmoidal curves are solutions to the logistic differential equation [12]. Furthermore, these types of Hill functions and sigmoidal functions are the backbone of **Michaelis-Menton kinetics** [12], [13]. Michaelis-Menton kinetics is a famous way to model enzymatic reactions, i.e., chemical reactions involving enzymes attaching to a substrate that ignites a chemical reaction to take place (*catalysis*) that produces a new chemical product.

Let's briefly look at solutions to a predator-prey system with the following form for different Holling-Type interaction functions, for the same interaction term in both the predator and prey equations:

$$\begin{cases} \frac{dx}{dt} = \alpha x \left(1 - \frac{x}{C}\right) - f(x, y) \\ \frac{dy}{dt} = f(x, y) - \gamma y, \end{cases} \quad (2.14)$$

where $f(x, y) \in \left\{ qxy, \frac{qxy}{1+bx}, \frac{qx^2y}{1+bx^2} \right\}$. For our comparisons, we will fix the following parameter values:

$$\alpha = 0.5, C = 120, d = 0.1, q = 0.005,$$

and vary the handling time, b , i.e., we will compare solutions for

$$b \in \{0.0015, 0.00375, 0.0125, 0.025, 0.0375, 0.05\}.$$

Figure 2.22 shows the differences in solutions for those various b values. When b is small enough, the dynamics for the Holling-Type I and Holling Type-II solutions are very similar. As b increases, the solutions display different behavior. However, the Holling-Type III dynamics are substantially different for small b . This is due to the fact that as in the limit as $b \rightarrow 0$, we see the Holling Type-I and Holling-Type II terms are equivalent, while the limiting case for the Holling-Type III is different. Let's quickly go through that calculation:

$$\lim_{b \rightarrow 0} (\text{Holling-Type II term}) = \lim_{b \rightarrow 0} \frac{qxy}{1+bx} = qxy = \text{Holling-Type I term},$$

and

$$\lim_{b \rightarrow 0} (\text{Holling-Type III term}) = \lim_{b \rightarrow 0} \frac{qx^2y}{1+bx^2} = qx^2y \neq \text{Holling-Type I term}.$$

However, as b increases significantly, the Holling-Type II and Holling-Type III solutions begin to look similar. This is because the asymptotic behavior begins to look like:

$$\frac{qxy}{1+bx} \sim \frac{qxy}{bx} = \frac{qy}{b} \quad \text{for large } b, \text{ Holling-Type II}$$

$$\frac{qx^2y}{1+bx^2} \sim \frac{qx^2y}{bx^2} = \frac{qy}{b} \quad \text{for large } b, \text{ Holling-Type III}$$

Thus, Holling-Type II and III have the same asymptotic behavior for large b .

While it typically goes that Type I and II responses are more indicative for invertebrates, e.g., Type I has been applied to many filter feeders and Type II to insects or parasitoids, and Type III for vertebrates, this is not set in stone. Many ecologists will use whichever Holling-Type seems appropriate for their specific situation, i.e., which model will best describe the population data they have. For example, sometimes a Holling-Type III response may be used for phytoplankton and zooplankton interactions, even though phytoplankton (P) are producers in aquatic systems and small zooplankton (Z) are invertebrates. An example of such a system is given below:

$$\begin{cases} \frac{dP}{dt} = \beta P(1 - P) - \left(\frac{P^2}{\alpha^2 + P^2}\right) Z \\ \frac{dZ}{dt} = \gamma Z \left[\left(\frac{P^2}{\alpha^2 + P^2}\right) Z - C \right], \end{cases}$$

where P and Z are the density of phytoplankton and zooplankton, respectively, and parameters $\alpha, \beta, \gamma, C > 0$.



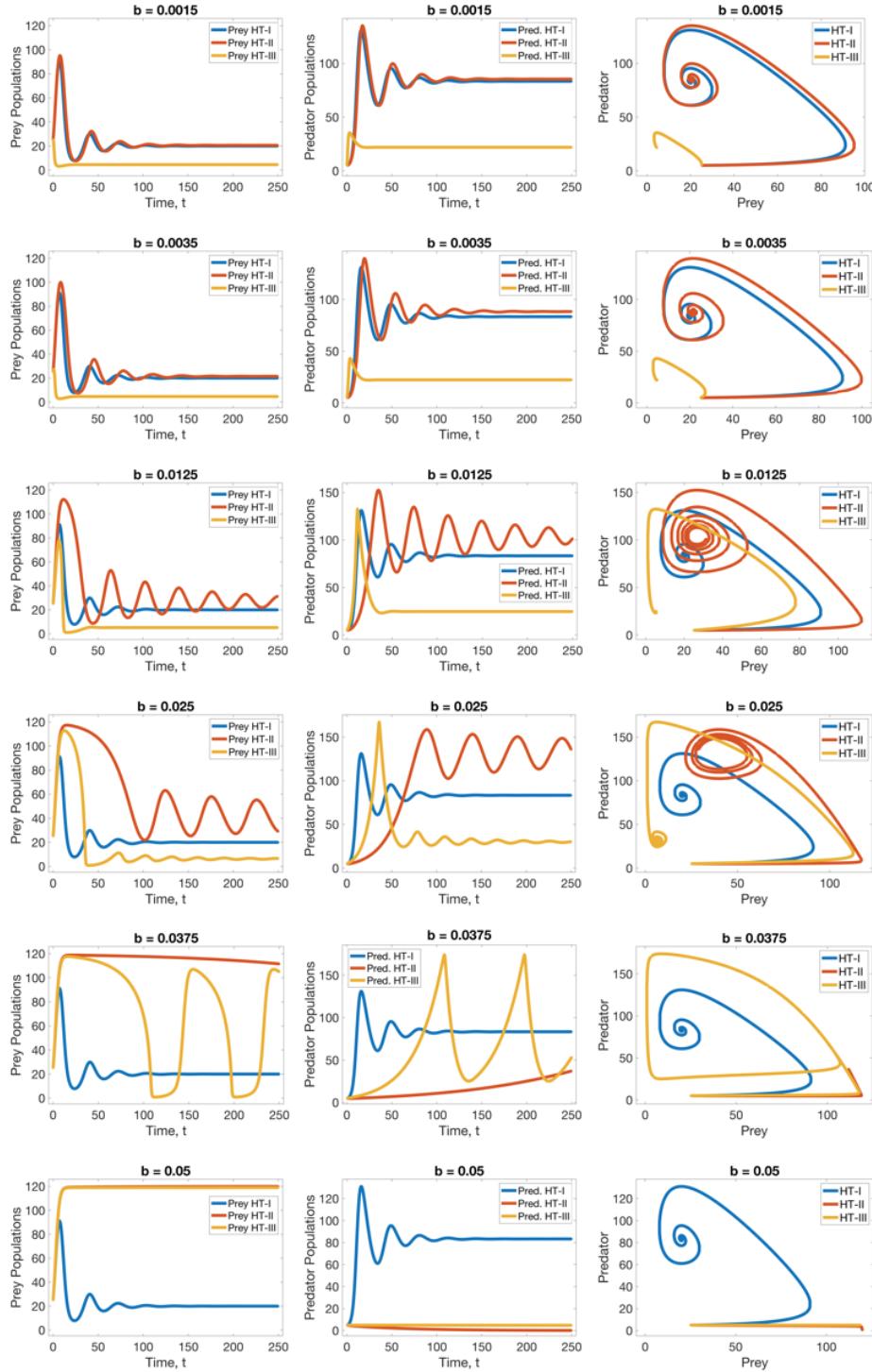


Figure 2.22: A comparison between solutions for the predator-prey system given in (2.14) for different levels of the parameter, b , i.e., the parameter that represents the handling time or learning time by the predator.

Example 2.6.1. Algae blooms!

There are many important models of trophic level dynamics in aquatic food webs. Here we will consider one such model describing phytoplankton and zooplankton interactions:

$$\begin{cases} \frac{dP}{dt} = \beta P(1 - P) - \left(\frac{P^2}{\alpha^2 + P^2}\right) Z \\ \frac{dZ}{dt} = \gamma Z \left[\left(\frac{P^2}{\alpha^2 + P^2}\right) Z - C \right], \end{cases} \quad (2.15)$$

where P and Z are the density of phytoplankton and zooplankton, respectively, and parameters $\alpha, \beta, \gamma, c > 0$. The zooplankton prey upon the phytoplankton for sustenance. We are ignoring the energy transfer dynamic involving the phytoplankton depending on water nutrients (and/or photosynthesis processes). Thus, β encapsulates the necessary conditions for phytoplankton density growth.

First, we will search for possible equilibria of the model. Starting with the second equation for $\frac{dZ}{dt}$ and setting it equal to 0, we can see that

$$\gamma \bar{Z} = 0 \Rightarrow \bar{Z} = 0.$$

and that

$$\left(\frac{P^2}{\alpha^2 + P^2}\right) Z - C = 0,$$

which then gives

$$\frac{P^2}{\alpha^2 + P^2} = C$$

$$P^2 = C(\alpha^2 + P^2)$$

$$P^2 - CP^2 = C\alpha^2$$

$$P^2(1 - C) = C\alpha^2$$

$$P^2 = \frac{C}{1 - C} \alpha^2$$

$$\bar{P} = \pm \sqrt{\frac{C}{1 - C}} \alpha. \quad (\text{we only need the '+' root})$$

Hence we have possible values of equilibria involving $\bar{Z} = 0$ and $\bar{P} = \frac{C}{1 - C} \alpha$. We will then plug them into the other differential equation for $\frac{dP}{dt}$ and see what value of the other variable gives a zero right hand side.

- $\bar{Z} = 0$:

$$\frac{dP}{dt} = \beta P(1 - P) - (\textcolor{blue}{0}) \frac{P^2}{\alpha^2 + P^2} = 0$$

$$\beta P(1 - P) = 0$$

$$\Rightarrow \bar{P} \in \{0, 1\}.$$

So we get the following equilibrium points:

$$(\bar{P}, \bar{Z}) \in \left\{ \underbrace{(0, 0)}_{\substack{\text{total} \\ \text{extinction}}}, \underbrace{(1, 0)}_{\substack{\text{algae} \\ \text{bloom}}} \right\}.$$

- $\bar{P} = \frac{C}{1-C} \alpha$:

Now, we know what \bar{P} is equal to. Therefore we can solve for \bar{Z} in terms of \bar{P} and then substitute what we found for \bar{P} into that.

$$\frac{dP}{dt} = \beta \bar{P}(1 - \bar{P}) - (\textcolor{pink}{Z}) \frac{\bar{P}^2}{\alpha^2 + \bar{P}^2} = 0$$

$$\beta \bar{P}(1 - \bar{P}) = (\textcolor{pink}{Z}) \frac{\bar{P}^2}{\alpha^2 + \bar{P}^2}$$

$$(\textcolor{pink}{Z}) \frac{\bar{P}^2}{\alpha^2 + \bar{P}^2} \cdot \frac{\alpha^2 + \bar{P}^2}{\bar{P}^2} = \beta \bar{P}(1 - \bar{P}) \cdot \frac{\alpha^2 + \bar{P}^2}{\bar{P}^2}$$

$$\bar{Z} = \beta(1 - \bar{P}) \frac{\alpha^2 + \bar{P}^2}{\bar{P}}.$$

$$\Rightarrow \bar{Z} = \beta \left[1 - \frac{\frac{C}{1-C} \alpha}{1 - \frac{C}{1-C} \alpha} \right] \frac{\left(\frac{C}{1-C} \alpha \right)^2 + \alpha^2}{\frac{C}{1-C} \alpha}.$$

Hence the non-zero equilibrium we get is:

$$(\bar{P}, \bar{Z}) = \left(\frac{C}{1-C} \alpha, \beta \left[1 - \frac{C}{1-C} \alpha \right] \frac{\left(\frac{C}{1-C} \alpha \right)^2 + \alpha^2}{\frac{C}{1-C} \alpha} \right).$$

Note that if $C \leq 0$, or $C \geq 1$ this equilibrium cannot exist, i.e., we will get either imaginary values or a divide by zero. However, if $C = 0$, we see that $\bar{P} = 0$, which would pop us over to the total extinction equilibrium. Therefore, it seems that there is a clear parameter restriction on C in order for this equilibrium to exist: $0 < C < 1$.

Therefore, this phytoplankton and zooplankton model gives rise to three possible equilibria:

1. $(\bar{P}, \bar{Z}) = (0, 0)$: total extinction. Both population densities go towards zero.
2. $(\bar{P}, \bar{Z}) = (1, 0)$: an algae bloom occurs! The zooplankton population goes to zero and there is nothing to suppress the booming increase in phytoplankton.
3. $(\bar{P}, \bar{Z}) = \left(\frac{C}{1-C} \alpha, \beta \left[1 - \frac{C}{1-C} \alpha \right] \frac{\left(\frac{C}{1-C} \alpha \right)^2 + \alpha^2}{\frac{C}{1-C} \alpha} \right)$: both populations are healthy and thriving.

For our considerations here, we are interested in investigating the stability properties for the algae bloom equilibrium. If the zooplankton population goes to 0 while the phytoplankton population is thriving, it implies that there must severe water toxicity issues arising. Thus, healthy zooplankton populations are an indicator of healthy water quality. Let's see what parameter relationships in this model can give rise to this unfortunate situation.

First, we get to compute the Jacobian matrix. Let's walk through this together. Spoiler: we will get to perform many product rules! Before diving into any partial derivatives, let's quickly define the right hand sides of our differential equations from (2.15):

$$\begin{cases} \frac{dP}{dt} = \beta P(1 - P) - \left(\frac{P^2}{\alpha^2 + P^2} \right) Z = \mathbf{f}(P, Z) \\ \frac{dZ}{dt} = \gamma Z \left[\left(\frac{P^2}{\alpha^2 + P^2} \right) Z - C \right] = \mathbf{g}(P, Z), \end{cases}$$

Hence, we will fill the Jacobian matrix using the following

$$\mathbf{J}(P, Z) = \begin{bmatrix} \frac{\partial f}{\partial P} & \frac{\partial f}{\partial Z} \\ \frac{\partial g}{\partial P} & \frac{\partial g}{\partial Z} \end{bmatrix}.$$

To avoid quotient rule differentiations, we can rewrite $\mathbf{f}(P, Z)$ in the following manner using algebraic rules involving powers

$$\mathbf{f}(P, Z) = \beta P(1 - P) - Z P^2 (\alpha^2 + P^2)^{-1}.$$

Now let's compute those partial derivatives

- $\frac{\partial f}{\partial P} = \beta - 2\beta P - \frac{2ZP}{\alpha^2 + P^2} + \frac{2ZP^3}{(\alpha^2 + P^2)^2}$
- $\frac{\partial f}{\partial Z} = \frac{-P^2}{\alpha^2 + P^2}$
- $\frac{\partial g}{\partial P} = \gamma Z \left[\frac{2P}{\alpha^2 + P^2} - \frac{2P^3}{(\alpha^2 + P^2)^2} \right]$
- $\frac{\partial g}{\partial Z} = \gamma \left[\frac{P^2}{\alpha^2 + P^2} - C \right]$.

Substituting those partial derivatives into the Jacobian matrix appropriately, we get the following Jacobian in all of its glory:

$$\mathbf{J}(P, Z) = \begin{bmatrix} \beta - 2\beta P - \frac{2ZP}{\alpha^2 + P^2} + \frac{2ZP^3}{(\alpha^2 + P^2)^2} & \frac{-P^2}{\alpha^2 + P^2} \\ \gamma Z \left[\frac{2P}{\alpha^2 + P^2} - \frac{2P^3}{(\alpha^2 + P^2)^2} \right] & \gamma \left[\frac{P^2}{\alpha^2 + P^2} - C \right] \end{bmatrix}.$$

Recall that we're interested in the algae bloom cases when $(\bar{P}, \bar{Z}) = (1, 0)$. Let's substitute those values into the Jacobian. Doing so gives us

$$J(1,0) = \begin{bmatrix} \beta - 2\beta(\textcolor{violet}{1}) & \frac{-(\textcolor{violet}{1})^2}{\alpha^2 + (\textcolor{violet}{1})^2} \\ 0 & -\gamma \left(\frac{(\textcolor{violet}{1})^2}{\alpha^2 + (\textcolor{violet}{1})^2} - C \right) \end{bmatrix} = \begin{bmatrix} -\beta & \frac{-1}{\alpha^2 + 1} \\ 0 & \gamma \left(\frac{1}{\alpha^2 + 1} - C \right) \end{bmatrix}.$$

Welp, that reduced quite a bit I'd say! Let's compute the eigenvalues of the above matrix to determine the stability properties of this equilibrium. Notice that this matrix is an upper triangular matrix \rightarrow therefore we know its eigenvalues are the values along the main diagonal! (Also, for the record, the same is true for either diagonal matrices or lower-triangular matrices.)

Hence, our eigenvalues are:

$$\lambda = -\beta \quad \text{and} \quad \lambda = \gamma \left(\frac{1}{\alpha^2 + 1} - C \right).$$

Since all the parameters are assumed to be positive at the start of the problem, one eigenvalue is always negative ($\lambda = -\beta$) and thereby satisfies the stability condition for differential equations. However, from the other eigenvalue we can find conditions that will determine when this equilibrium is stable. Let's take a look at that. For stability we need $\lambda < 0$, so

$$\lambda = \gamma \left(\frac{1}{\alpha^2 + 1} - C \right) \Rightarrow \frac{1}{\alpha^2 + 1} < C.$$

Therefore, if $\frac{1}{\alpha^2 + 1} < C$, this algae bloom equilibrium in which the zooplankton population goes to zero but the phytoplankton thrives is stable. If we glance at the differential equation model once again for the zooplankton, we can actually see that this condition alone implies the zooplankton population must decrease, even if there was a healthy zooplankton density population to begin with! Recall

$$\frac{dZ}{dt} = \gamma Z \left[\left(\frac{P^2}{\alpha^2 + P^2} \right) Z - C \right].$$

Now if $P \approx 1$ and even if $Z \approx 1$, we see that this equation looks like

$$\frac{dZ}{dt} \approx \gamma \left(\frac{(\textcolor{violet}{1})^2}{\alpha^2 + (\textcolor{violet}{1})^2} \right) (\textcolor{blue}{1}) - C$$

and then

$$\frac{dZ}{dt} \approx \gamma \left[\frac{1}{\alpha^2 + 1} - C \right].$$

Therefore, if $\frac{1}{\alpha^2 + 1} < C$, we see that this implies that $\frac{dZ}{dt} < 0$, even if there were healthy zooplankton populations to begin with! However, we did not prove with this quick side calculation whether the zooplankton population would continue decreasing towards 0 or if it would decrease to some other non-zero equilibrium value.

3

Application Hodge-Podge!

“Don’t let anyone rob you of your imagination, your creativity, or your curiosity. It’s your place in the world; it’s your life. Go on and do all you can with it, and make it the life you want to live.”

– Dr. Mae Jemison ([bio](#))

3.1 SIR Models and Disease Transmission Models

Epidemiology is the study of the incidence, distribution, patterns, and possible control of health and disease within well-defined populations. It shapes both health and public policy based on data-driven efforts and evidence based scientific practices. Epidemiologists may focus either on preventive aspects of disease, by identifying risk factors, or retrospective aspects, to which they determine future control efforts based on previous policy action (or inaction). Epidemiology is the cornerstone of public health.

Mathematical modeling formally entered the public health space in 1927 by W.O. Kermack and A.G. McKendrick in 1927 ([link](#)). They were not mathematicians; Kermack was a Military Physician and Epidemiologist and McKendrick was a biochemist. They were able to mathematically develop a differential equations approach for understanding how a disease spreads through a population, based off the collaborative work of medical doctor Sir Ronald Ross ([bio](#)) and algebraic geometer Dr. Hilda Phoebe Hudson ([bio](#)). Ross and Hudson's work focused on preventing the spread of malaria a decade prior. Sir Ronald Ross was the first to discover that the malaria parasite is transmitted by mosquitoes.

The groundbreaking work of Kermack-McKendrick from 1927, 1932, and 1933 is still widely used today as the foundation of mathematical modeling for disease transmission. In a nutshell, they break the population into distinct categories (*compartments*), such as a susceptible (S: *healthy persons*), infected (I: *persons who have the disease*), and removed (R, R: *persons who have recovered from the disease and are immune*) category. The idea is that if there is at least one individual in the I category, they could infect a healthy person in the S class to catch the disease. From there the individual could recover and join the R category, or in the case of a lethal disease, they may pass away from having the disease. These ideas are illustrated schematically in Figure 3.1.

Figure 3.1 shows 3 compartments - S, I, and R, each corresponding to a different population category. The direction of the arrows indicate whether a population is increasing or decreasing, i.e., arrows pointed toward a compartment implies that population is

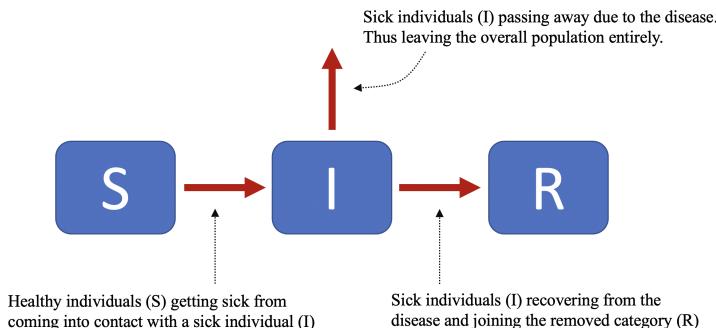


Figure 3.1: Original SIR Model of Kermack-McKendrick.

increasing from members of the compartment to which the arrow is stemming from. Similarly, arrows pointed away imply that the population is decreasing and the arrow shows where members of that population subcategory are heading. Thus the flow through the model is $S \rightarrow I$ and then from I , the members of the I category either (i) recover from the disease and join the R class, $I \rightarrow R$, or (ii) they pass away from the disease and leave the overall population within the model itself.

There are many variations of an SIR-type model, as the one shown in Figure 3.1. Some models may assume that new members can be added into the S class via natural birth, or that natural death could happen to members of any compartment, or they could add other dynamics entirely, such as different compartments depending on the incidence and severity of contracting a disease or vaccinated classes. In a grander view, each SIR model that gets constructed is specific for a particular disease. That is, a model that accurately describes the transmission of the common cold (acute viral nasopharyngitis) may not be appropriate for the measles (rubeola), a much more deadly, virulent, and infectious disease.

One common goal among all disease transmission mathematical models is to quantify/compute that model's **basic reproduction number**.

Definition 1.1

The **basic reproduction number**, denoted R_0 , gives the expected number of new infected cases directly generated by one individual in the infected category (I).

NOTE:

- This assumes a population where all individuals are susceptible to infection
- It is pronounced *R-naught*
- $R_0 > 1$: bad (epidemic occurs)
- $R_0 < 1$: desired (disease will die away)
- An **epidemic** occurs when a disease perpetuates itself throughout a particular community infecting many people very rapidly.
- A **pandemic** has a similar definition, except it happens at a much larger scale - either across many countries, continent, or worldwide.

Thus, R_0 gives us a mechanism to understand whether we would expect an epidemic to occur or whether a disease will go away over time. For example, a $R_0 > 1$, means

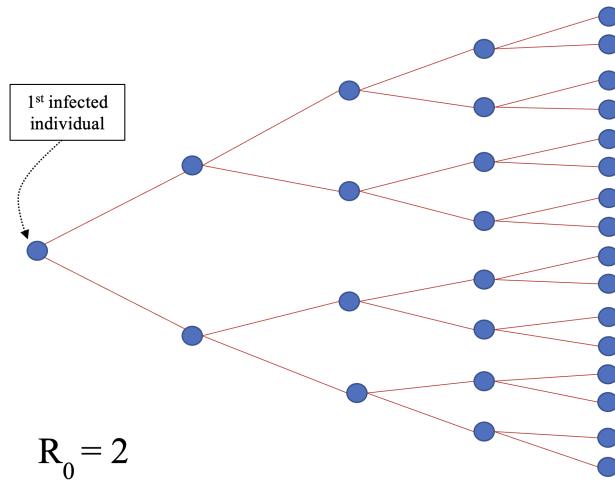


Figure 3.2: An illustration of a situation involving a disease with known basic reproduction number of 2 ($R_0 = 2$).

that one infected individual is likely to infect more than one other person to contract the disease. Those new individuals who contract the disease will then themselves go and infect more than person on average to contract the disease. Therefore, an epidemic is likely where many people quickly become infected! For example, Figure 3.2 illustrates a situation in which there is a single infected person for a disease in which is known to have a basic reproduction number of 2 ($R_0 = 2$). The individuals who contract the disease then go and (unfortunately) transmit the disease to 2 other healthy individuals on average. Shortly, the number of infected individuals *geometrically* increases.

On the other hand, if a disease has a $R_0 < 1$, an infected person (on average) transmits the disease to less than one other person. That is, not every sick person will make another person sick! Eventually, this would lead to the disease going away. It would do so *geometrically*.

You may be wondering why I used the word *geometrically* twice in the above two previous paragraphs. That is because this idea is strongly tied to the notion of a **geometric series** from Calculus B - where the multiplicative factor is the basic reproduction number (R_0), see Definition 1.2. If $R_0 > 1$ we see a *geometric increase* in the number of infected cases, which we can compute by partial sums of a geometric series. In the same vein, if $R_0 < 1$, we could see a *geometric decrease* in the number of infected cases.

Definition 1.2

A **geometric series** has the form of

$$\sum_{n=0}^{\infty} ar^n = a + ar + ar^2 + ar^3 + \dots$$

If

- $|r| < 1$: the geometric series *converges*. Furthermore, we even know what is

converges to! It converges to

$$\sum_{n=0}^{\infty} ar^n = \frac{a}{1-r}.$$

- $|r| \geq 1$: the geometric series *diverges*.

NOTE:

- In a context of disease epidemics we assume that the multiplicative factor r in a geometric series is actually R_0 .
- In the same context, we would assume that a is equal to the original number of infected individuals.
- Thus, if $r = R_0 < 1$, using geometric series we can approximate how many individuals we would expect to get sick.

At this point you might be wondering what some known approximate basic reproduction numbers are for a few diseases. Table 3.1 lists some of them. For many there are wide ranges in what the basic reproduction number is. Think for a moment about how an epidemiologist, mathematician, or scientist might compute R_0 using the data available. There are *many* factors that could affect the R_0 value - certainly the demographics of the population considered (age), the geographic location (city, county, state, country), access to healthcare options, and public health policy! Therefore, R_0 could be substantially different depending on those factors - the who and the where. Moreover, Table 3.1 gives the R_0 values estimated from available data. Unfortunately, collecting data can be difficult, especially when you're trying to collect data on number of overall infected cases, rather than overall fatalities. When we compute the basic reproduction number from our SIR-type mathematical models, we will be able to see how different dynamics could affect R_0 , e.g., we can see how do different intervention or control strategies affect it, such as vaccines, social distancing, or masks.

Moreover, this type of SIR-modeling (or *compartment modeling* in general) gets applied to many problems in mathematical biology. For example, outside of the population-based disease transmission models for diseases such as those listed in Table 3.1, mathematicians and scientists have used them to model other systems, such as:

- the opioid crisis and other drug epidemics
- rumor (and misinformation) propagation
- disease models via cellular processes in medicine (like HIV/AIDS, cancer)
- designing new drugs, drug therapies, or pharmaceuticals
- math anxiety across generations of students and teachers
- the evolution of languages and linguistics
- ...these were only a few off the top of my head and by no means an exhaustive list!

Disease	R_0	Transmission Mode(s)
Measles	12-18	Aerosol
Chickenpox (varicella)	10-12	Aerosol
Mumps	10-12	Respiratory Droplets
Polio	5-7	Fecal-oral
Rubella	5-7	Respiratory Droplets
Whooping cough(pertussis)	5.5	Respiratory Droplets
Smallpox	3.5-6	Respiratory Droplets
SARS	0.19-1.08	Respiratory Droplets
COVID-19	2-6	Respiratory Droplets / possible Aerosol
Common Cold	2-3	Respiratory Droplets
Influenza (1918)	1.4-2.8	Respiratory Droplets
Ebola (2014)	1.5-1.9	Bodily fluids
Influenza (2009)	1.4-1.6	Respiratory Droplets
Influenza (seasonal)	0.9-2.1	Respiratory Droplets
MERS	0.3-0.8	Respiratory Droplets

Table 3.1: Values of the basic reproduction number, R_0 , for a variety of diseases based off of data. Note the distinction between respiratory droplets and aerosols is that aerosols are assumed to remain suspended in air for an extended period of time (Sources: [1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13])

Okay, that is enough background information on disease transmission models for now. We've introduced the idea of a **compartment model** and the **basic reproduction number**, R_0 . Now seems like the perfect time to introduce some differential equation based SIR-models, compute their R_0 , and relay all of this information in the context of what we've been doing for dynamical systems - *equilibria* and *stability analysis*.

3.1.1 Base SIR Model: no deaths

For the first SIR-type model that we will consider, we will again assume that there are three categories (compartments): susceptible (S), infected (I), and recovered (R). Moreover, we will assume that no deaths occur due to the disease. For that matter, it inherently assumes that anyone who contracts the disease, will recover. Actually, let's explicitly list out our assumptions:

1. 3 different sub-populations: S, I, and R
2. People get sick, but do not die from the disease \rightarrow everyone who contracts the disease will recover
3. Once someone recovers from the disease, they cannot get it again.
4. This disease takes place on a timescale in which we can ignore natural births and deaths within the overall population
5. No vaccines or preventative measures are established within the population

Our model will include two terms. One term involves the transmission of the disease from infected individuals (I) to healthy, susceptible individuals (S). Thus, we need a term

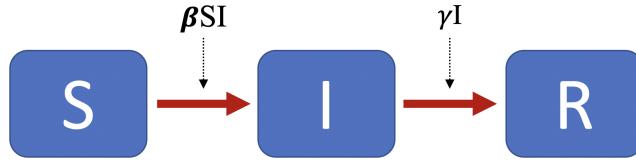


Figure 3.3: Compartment model diagram (box model) for the base SIR model described in Section 3.1.1 with no deaths due to the disease, nor natural births or deaths in the overall population.

that gives the per capita (per person) rate that a healthy person gets sick when interacting with a sick person. The other term our model will include involves the rate in which sick individuals will recover. We can mathematically describe these two terms in the following manner:

- βSI : per capita rate that a healthy person gets sick when interacting with a sick person
- γI : rate at which sick individuals recover from the disease

The *compartment model diagram* (box model) that describes this scenario is shown in Figure 3.3. This figure shows that healthy individuals move over to the infected class once they become infected, at a rate of βSI . Next, the infected individuals move over into the recovered class, once they recover at some rate γI . Hence, what leaves one compartment, enters another! Therefore, we can write the dynamical system describing this scenario in the following way:

$$\left. \begin{array}{rcl} \frac{dS}{dt} & = & -\beta SI \\ \frac{dI}{dt} & = & \beta SI - \gamma I \\ \frac{dR}{dt} & = & \gamma I \end{array} \right\} \quad (3.1)$$

Notice that if let $N = S + I + R$, that is, we let N equal the total number of people in the population, we can show that the overall population N must remain constant. How can we show that? Welp, take a derivative of N with respect to time, t , and see what happens! Check it out:

$$N = S + I + R$$

$$\begin{aligned}\frac{d}{dt}N &= \frac{d}{dt}(S + I + R) \\ \frac{dN}{dt} &= \frac{d}{dt}S + \frac{d}{dt}I + \frac{d}{dt}R \\ &= \frac{dS}{dt} + \frac{dI}{dt} + \frac{dR}{dt} \\ &= -\beta SI + \beta SI - \gamma I + \gamma I \\ &= 0.\end{aligned}$$

Hence we have showed that the overall population N is constant in this model for all time, since $\frac{dN}{dt} = 0$. This also let us do a logic test on our model...remember those assumptions that there are no natural births or deaths nor deaths due to the disease? You may be wondering, *how does this help us?* Glad you asked! One tool that is useful for many math models is making an assumption that the overall population doesn't change. Note that this only requires that $\frac{dN}{dt} = 0$, but it doesn't necessarily make us make the same assumptions above again (see later SIR-model sections \circledcirc). What this constant population assumption buys us is the ability to work in terms of population fractions rather than bulk population numbers. That is, we will model how the fractions of the overall population who are susceptible, infected, or recovered changes over time, rather than an explicit number of individuals. In a nutshell, it takes out the overall population size from consideration. Enough rambling on, this is what I mean:

Since

$$S + I + R = N,$$

we can divide both sides by N to get

$$\frac{S}{N} + \frac{I}{N} + \frac{R}{N} = 1.$$

If we define $s = \frac{S}{N}$, $i = \frac{I}{N}$, and $r = \frac{R}{N}$, we now have that

$$s + i + r = 1.$$

Next we can rewrite our original dynamical in terms of $S, I, R \rightarrow s, i, r$:

$$\left. \begin{array}{l} \frac{ds}{dt} = -\beta si \\ \frac{di}{dt} = \beta si - \gamma i \\ \frac{dr}{dt} = \gamma i \end{array} \right\}. \quad (3.2)$$

This may not seem like that big of deal, but it will help us out in our future analyses. For example, we will be able to reduce our system of 3 coupled differential equations in terms of 2! Check it out. Revisiting $s + i + r = 1$, we can solve for r in terms of s and i , i.e.,

$$r = 1 - s - i. \quad (3.3)$$

Stare at that for a moment. It might not look like much but just some algebraic relation; however (!), what it says is that if we known the values of s and i , we can just use (3.3) to get the associated value of r ! So. We. Never. Have. To. Actually. Solve. The. Differential. Equation. For. $\frac{dr}{dt}$. *Gnarly!*

Our system of 3 coupled differential equations, actually becomes a system of 2 equations now:

$$\left. \begin{array}{l} \frac{ds}{dt} = -\beta si \\ \frac{di}{dt} = \beta si - \gamma i \end{array} \right\}. \quad (3.4)$$

When presented with dynamical systems previously, what did we do? We searched for their equilibria and their stability properties. Let's do that!

Example 3.1.1. Equilibria for the Base SIR-model given in (3.4).

To find the equilibria for this system, we need to find the values of \bar{s} and \bar{i} that make the differential equations have right-hand-side of zero, so that $\frac{ds}{dt} = 0$ and $\frac{di}{dt} = 0$. Let's start with the $\frac{di}{dt}$ equation:

$$\begin{aligned} \frac{di}{dt} &= \beta \bar{s} \bar{i} - \gamma \bar{i} = 0 \\ \bar{i}(\beta \bar{s} - \gamma) &= 0. \end{aligned}$$

Since we have the product of two things equaling zero, one or more of those factors must be zero. Setting them both equal to zero, we can find that this happens when either $\bar{i} = 0$ or $\bar{s} = \frac{\gamma}{\beta}$. Let's put these two values into our other equation for $\frac{ds}{dt}$ and see what happens!

1. $\bar{i} = 0$:

$$\frac{ds}{dt} = -\beta \bar{s} \bar{i} = -\beta \bar{s}(0) = 0.$$

Therefore it doesn't actually matter what \bar{s} is...? Welp, we could determine whether this equilibrium has any chance at being stable. Before we do that, let's check out the other case for $\bar{s} = \frac{\gamma}{\beta}$.

2. $\bar{s} = \frac{\gamma}{\beta}$:

$$\frac{ds}{dt} = -\beta \bar{s} \bar{i} = -\beta \left(\frac{\gamma}{\beta}\right) \bar{i} = -\gamma \bar{i}.$$

Therefore if we want to enforce $\frac{ds}{dt} = 0$, too, we need that either $\gamma = 0$ (which can't happen based off our modeling assumptions that $\gamma > 0$) or that $\bar{i} = 0$, too.

Hence this equilibrium is: $(\bar{s}, \bar{i}) = (0, 0)$. Note that we find that \bar{r} is equal to:

$$\bar{r} = 1 - \bar{s} - \bar{i} = 1 - 0 - 0 = 1.$$

Therefore, in this equilibrium, everyone is in the recovered class. This would imply that everyone in the healthy population got sick and recovered.

Hence the dynamical system for our Base SIR-model in (3.4) gave rise to two potential equilibrium values: $(\bar{s}, \bar{i}) = (0, 0)$, which lead to everyone getting sick, recovering, and then being put into the recovered class, and the other which appears more interesting. The other equilibrium only said that $\bar{i} = 0$ - that is, we didn't find a value for \bar{s} . Let's check out what this other equilibrium would imply.

Let's think for a moment. Does $\bar{i} = 0$ with no other conditions on \bar{s} or \bar{r} seem like a good thing? If $\bar{i} = 0$, this implies that there are no sick people in the infected class - this seems great! For many epidemiologists and mathematical biologists, this is the equilibrium we're after. If $i = 0$, we give this a special name... the **disease free equilibrium (DFE)**. This equilibrium occurs when there are no more infected (sick) individuals in the population, thus everyone is either still healthy or has recovered! Hopefully it goes without saying that we will want this equilibrium to be stable, so that the population stays disease free. On that note, let's check out its stability properties.

Example 3.1.2. *Stability properties of the DFE for the Base SIR-model given in (3.4).*

To check our the stability properties of this disease free equilibrium, we get to (1) compute a Jacobian matrix, J , for the system, (2) evaluate J at the equilibrium value $(\bar{s}, \bar{i}) = (\bar{s}, 0)$, and (3) find the eigenvalues of J .

Recall our system from (3.4) was:

$$\begin{aligned} \frac{ds}{dt} &= -\beta si \\ \frac{di}{dt} &= \beta si - \gamma i \end{aligned} \quad \left. \right\}.$$

The resulting Jacobian matrix for this system is:

$$J(\bar{s}, \bar{i}) = \begin{bmatrix} -\beta i & -\beta s \\ \beta i & \beta s - \gamma \end{bmatrix}.$$

If we evaluate this Jacobian at the DFE, we get:

$$J(\bar{s}, 0) = \begin{bmatrix} 0 & -\beta \bar{s} \\ 0 & \beta \bar{s} - \gamma \end{bmatrix}.$$

Next, we will compute the eigenvalues of this matrix.

$$\begin{aligned} |\lambda I - J(\bar{s}, 0)| &= \begin{vmatrix} \lambda & -\beta s \\ 0 & \lambda - (\beta \bar{s} - \gamma) \end{vmatrix} \\ &= \lambda(\lambda - (\beta \bar{s} - \gamma)). \end{aligned}$$

Thus the eigenvalues are:

$$\lambda \in \{0, \beta \bar{s} - \gamma\}.$$

Since $\lambda = 0$ is an eigenvalue, in order for the DFE to have a chance at being stable, we need that:

$$\lambda = \beta \bar{s} - \gamma < 0.$$

This would then require that we require that

$$\beta \bar{s} < \gamma.$$

If we divide both sides of the above inequality by γ , we obtain the following condition to make the DFE stable:

$$\frac{\beta}{\gamma} \bar{s} < 1. \quad (3.5)$$

Perhaps, you might be wondering why we did that extra algebra after finding the eigenvalues to give us (3.5) in the above example. I don't blame you. Welp, the reason is particularly subtle, and it goes back to our desire to compute a basic reproduction number, R_0 . Notice that in (3.5), we have a quantity that is *less than one* to which when satisfied states that the disease free equilibrium will be stable. This term of the left hand side of (3.5) is actually our basic reproduction number, R_0 , for this system!!

Thus, for our base SIR-model without natural births and deaths nor deaths due to the infectious disease, we found the basic reproduction number to be

$$R_0 = \frac{\beta}{\gamma} \bar{s}.$$

Furthermore, remember that since \bar{s} describes a fraction of the population that is susceptible to catching the disease (the healthy population), \bar{s} can only take on values between 0 and 1, i.e., $0 \leq \bar{s} \leq 1$. Larger values of \bar{s} actually make R_0 larger in this case (see the dependence of \bar{s} in R_0)! Let's think about that for a minute. Does that make sense? If R_0 describes how many secondary infections will occur due to one individual being infected, it makes sense that with a larger susceptible population, the chance of infecting one of those healthy individuals will be greater. *The more people that could be infected, the greater the chance someone gets infected when exposed.*

Let's examine what happens when we make \bar{s} as large as it can get in this case ($\bar{s} = 1$). This will result in maximizing R_0 for different values of \bar{s} . Moreover, it would mean that the entire population could remain healthy and that no one gets sick. Assuming $\bar{s} = 1$, the largest R_0 can get in this case is:

$$R_0 = \frac{\beta}{\gamma}.$$

Observing the form of R_0 above, we see that the numerator has a term that describes the likelihood of infection when a sick individual comes into contact with a healthy individual, while the denominator contains the recovery rate in which sick people recover from the disease. From an epidemiological perspective, this says that in order for $R_0 < 1$ to prevent an epidemic, we need that the recovery rate must be greater than the infection rate \odot ! On the other hand, if the infection rate is higher than the recovery rate, more and more individuals will get sick, resulting in an epidemic \odot .

Interestingly, for this specific SIR-type model with no natural births or deaths nor deaths due to the disease, mandating that $\beta < \gamma$ will also cause the the population of i to always decrease! We can see that through the differential equation describing the fraction of the population that are infected, $\frac{di}{dt}$, i.e.,

$$\frac{di}{dt} = \beta s i - \gamma i = i(\beta s - \gamma) < 0 \quad \text{if } \beta < \gamma \quad \text{and since } 0 \leq s \leq 1 \quad \text{and } 0 \leq i \leq 1.$$

For the next SIR-type model, let's add dynamics to govern natural births and deaths as well as deaths due to the disease into the model.

3.1.2 Base SIR Model: including deaths

For the next SIR-type model that we will again assume that there are three categories (compartments): susceptible (S), infected (I), and recovered (R). However, unlike the model presented in Section 3.1.1, we will permit natural births and deaths as well as deaths due to the disease. That is, we will modify the existing model in (3.2) to include terms governing such dynamics. Let's begin by explicitly listing our assumptions:

1. 3 different sub-populations: S, I, and R
2. When people get sick, they might either recover or pass away due to the disease at an *enhanced death rate*. This enhanced death rate includes the natural death rate and death rate due to the disease.
3. Natural births and deaths are permissible dynamics within the population.
4. Once someone recovers from the disease, they cannot get the disease again.
5. We will assume the overall population remains constant over-time, i.e.,

$$N = S + I + R \quad \text{where} \quad \frac{dN}{dt} = 0.$$

Hence we will once again work in terms of population fractions,

$$s = \frac{S}{N}, \quad i = \frac{I}{N}, \quad \text{and} \quad r = \frac{R}{N}.$$

6. No vaccines or preventative measures are established within the population

Our model will include more terms than those of (3.2). Before we list out the terms, let's visualize the dynamics through a compartment diagram (box model), see Figure 3.4. This box model has the following new dynamics: (1) arrows pointed away from the S, I, and R classes signifying death and (2) an arrow pointed towards the S class illustrating an increase in the S class due to new births. Thus our new model includes the following terms:

- βSI : per capita rate that a healthy person gets sick when interacting with a sick person
- γI : rate at which sick individuals recover from the disease
- μS and μR : terms describing folks leaving the S or R classes due to natural death
- $\mu^* I$: term describing folks leave the I class due to either natural death or death due to the disease. Therefore we assume that $\mu^* > \mu$.
- Λ : term describing new births into the susceptible population. We will determine what the value of Λ is below.

Thus, we can write our model in the following way:

$$\begin{cases} \frac{dS}{dt} = \Lambda - \beta SI - \mu S \\ \frac{dI}{dt} = \beta SI - \gamma I - \mu^* I \\ \frac{dR}{dt} = \gamma I - \mu R \end{cases} \quad (3.6)$$

Compare those equations to what we saw in our box model for this situation. Observe once again that the direction of the arrows determines the sign of a term in the differential

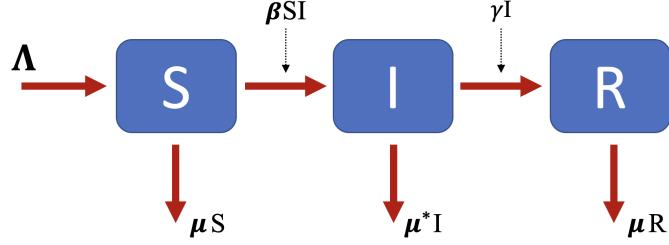


Figure 3.4: Compartment model diagram (box model) for a SIR-type model described in Section 3.1.2, which includes natural birth and death terms as well as death due to the infectious disease.

equation. Next we will again let $N = S + I + R$, that is, we let N equal the total number of people in the overall population. Here we will again desire that the overall population remains constant over time. However, in this case we will see shortly that upon doing so, it will give us conditions on our parameters → in particular it will help us define what Λ must be! Welp, let's begin again by taking a derivative of N with respect to time, t , and see what happens! Check it out:

$$N = S + I + R \quad \text{let's differentiate both sides w/ respect to } t$$

$$\begin{aligned} \frac{d}{dt}N &= \frac{d}{dt}(S + I + R) \\ \frac{dN}{dt} &= \frac{d}{dt}S + \frac{d}{dt}I + \frac{d}{dt}R \\ &= \frac{dS}{dt} + \frac{dI}{dt} + \frac{dR}{dt} \\ &= \Lambda - \beta SI - \mu S + \beta SI - \gamma I - \mu^* I + \gamma I - \mu R \\ &= \Lambda - \mu S - \mu^* I - \mu R. \end{aligned}$$

Since we wish to mandate that the overall population remains constant over time, we need to then require that

$$\frac{dN}{dt} = \Lambda - \mu S - \mu^* I - \mu R = 0.$$

Therefore, in order to enforce this condition we can find a birth rate that will satisfy this requirement, i.e.,

$$\Lambda = \mu S + \mu^* I + \mu R.$$

Note that this would mean that Λ varies over time; however, it is done so to ensure a constant overall population over time. Using this idea, we find that our system of

equations becomes:

$$\begin{cases} \frac{dS}{dt} = \Lambda - \beta SI - \mu S \\ \frac{dI}{dt} = \beta SI - \gamma I - \mu^* I \\ \frac{dR}{dt} = \gamma I - \mu R \end{cases} \Rightarrow \begin{cases} \frac{dS}{dt} = (\mu S + \mu^* I + \mu R) - \beta SI - \mu S \\ \frac{dI}{dt} = \beta SI - \gamma I - \mu^* I \\ \frac{dR}{dt} = \gamma I - \mu R \end{cases} \quad (3.7)$$

Note that in the equation for $\frac{dS}{dt}$ we will be able to cancel the μS terms. Since we wish to work in terms of population fractions once again, i.e.,

$$s = \frac{S}{N}, \quad i = \frac{I}{N}, \quad \text{and } r = \frac{R}{N},$$

we can cast the system of equations in (3.7) into the following:

$$\begin{cases} \frac{ds}{dt} = \mu^* i + \mu r - \beta s i \\ \frac{di}{dt} = \beta s i - \gamma i - \mu^* i \\ \frac{dr}{dt} = \gamma i - \mu r \end{cases} \quad (3.8)$$

When we actually do this change of variables, we divide both sides by N first and then see the new system fall out. For example, let's specifically look at the case for $\frac{ds}{dt}$. We'll start out by dividing both sides by $\textcolor{red}{N}$:

$$\frac{1}{\textcolor{red}{N}} \left[\frac{dS}{dt} \right] = \frac{1}{\textcolor{red}{N}} \left[\mu^* I + \mu R - \beta S I \right] = \mu^* \frac{I}{\textcolor{red}{N}} + \mu \frac{R}{\textcolor{red}{N}} - \beta \frac{S}{\textcolor{red}{N}} I.$$

A bunch of those terms we see just become our population fraction terms now:

$$\frac{1}{\textcolor{red}{N}} \left[\frac{dS}{dt} \right] = \mu^* i + \mu r - \beta s I;$$

however, we still have the I quantity on the far right and that derivative term to take care of. We can perform successfully finish this change of variables by noticing the following:

1. We can bring constants inside of differentiation operations without worrying due to the linearity property of differentiation:

$$\frac{1}{\textcolor{red}{N}} \frac{dS}{dt} = \frac{d}{dt} \left(\frac{S}{N} \right) = \frac{ds}{dt}$$

2. We can use the old multiple-by-one card up our sleeve:

$$\beta s I = \beta s \frac{N}{\textcolor{red}{N}} I = \beta s \textcolor{red}{N} \frac{I}{N} = \beta \textcolor{red}{N} s \textcolor{red}{N} i.$$

Wait, what do we do about that $\textcolor{red}{N}$ in that term now?!

We can simply refine our parameter β to encapsulate it. Notationally speaking the proper way to do this is by defining a **new** parameter, say $\tilde{\beta}$, as the following $\tilde{\beta} = \beta N$. Thus that nonlinear term will become:

$$\tilde{\beta} s i.$$

However, what most math biologists do at this point is revert back to their old notation (just recycle it) with the understanding that our newly revised β parameter in equation (3.8) just has different units than the original β value in (3.7).

So ultimately after that change of variables $S \rightarrow s$, $I \rightarrow i$, and $R \rightarrow r$, our new system of equations (3.4) looks virtually identically to the original system (3.7), but with s, i , and r and some parameters that have different units than the original system.

At this point you might be wondering why we had to do that change of variables in the first place - remember we did it to capitalize on the constant population assumption. Since we have enforced that the system will maintain a constant population overtime, we can actually reduce it from three equations to two equations. So, rather than have equations for s , i , and r , we only need to consider two equations, say one for s and i since we can always find the value of r from s and i , i.e.,

$$r = 1 - s - i.$$

Therefore, we can rewrite our system of equations as the following:

$$\begin{cases} \frac{ds}{dt} = \mu^* i + \mu r - \beta s i \\ \frac{di}{dt} = \beta s i - \gamma i - \mu^* i \\ \frac{dr}{dt} = \gamma i - \mu r \end{cases} \Rightarrow \begin{cases} \frac{ds}{dt} = \mu^* i + \mu(1 - s - i) - \beta s i \\ \frac{di}{dt} = \beta s i - \gamma i - \mu^* i \end{cases} \quad (3.9)$$

For our disease modeling considerations, we really want to explore the equilibrium and stability properties for the **disease free equilibrium (DFE)**. Let's go ahead and do this subsequent analysis!

Example 3.1.3. The DFE and its Stability Properties for the SIR-type system with deaths given in (3.9).

In order to find this disease free equilibrium, we need to mandate that $\bar{i} = 0$. This enforces that the number of infected individuals is zero so that the disease has effectively died out. To do this, we will get to find values of \bar{s} that could give rise to this DFE situation. Let's first start by setting (assuming) $\bar{i} = 0$. In this case the $\frac{di}{dt}$ equation looks like the following:

$$\frac{di}{dt} = \beta s \bar{i} - \gamma \bar{i} - \mu^* \bar{i} = \bar{i}(\beta s - \gamma - \mu^*) = 0,$$

since $\bar{i} = 0$. So by setting $\bar{i} = 0$, we have effectively made $\frac{di}{dt} = 0$ as well. Let's see what that means for the $\frac{ds}{dt}$ equation:

$$\frac{ds}{dt} = \mu^* \bar{i} + \mu(1 - s - \bar{i}) = \mu(1 - s).$$

Therefore to necessitate an equilibrium we need $\frac{ds}{dt} = 0$ which implies that the only \bar{s} values that satisfies that is $\bar{s} = 1$. So our **DFE** in this case is:

$$(\bar{s}, \bar{i}) = (1, 0).$$

Now we will investigate the stability properties of this DFE. Recall that we wish for this equilibrium to be stable so that if the number of infected individuals is ever very low, i.e., where i is close to zero, the disease itself will simply die out and go away. This does not mean that the individuals with the disease pass away, but simply that people can no longer become infected!

As usual, for stability analysis, we begin by computing the Jacobian of the system given in (3.9):

$$J(s, i) = \begin{bmatrix} -\mu - \beta i & \mu^* - \mu - \beta s \\ \beta i & \beta s - \gamma - \mu^* \end{bmatrix}.$$

Substituting our DFE into the above Jacobian gives:

$$J(1, 0) = \begin{bmatrix} -\mu - \beta(0) & \mu^* - \mu - \beta(1) \\ \beta(0) & \beta(1) - \gamma - \mu^* \end{bmatrix} = \begin{bmatrix} -\mu & \mu^* - \mu - \beta \\ 0 & \beta - \gamma - \mu^* \end{bmatrix}.$$

Once having evaluated the Jacobian at the DFE, we can compute its eigenvalues. We can go through the explicit steps of computing eigenvalues, which culminates in

$$\det(\lambda I - J(1, 0)) = 0,$$

or we can notice that the matrix J to which we want to find eigenvalues is upper triangle. That is, it only has non-zero entries along (or above) the main diagonal. For these types of matrices, our eigenvalues lie along the main diagonal! Therefore the eigenvalues of $J(1, 0)$ are:

$$\lambda = -\mu \quad \text{and} \quad \lambda = \beta - \gamma - \mu^*.$$

Next we will check the stability conditions for differential equations; we need these eigenvalues to be less than zero (or have real part less than zero). Upon doing so, we will be able to define the basic reproduction number, R_0 , for this SIR-type model with deaths. Checking those stability conditions gives us:

$$1. \quad -\mu < 0 \quad (\text{since } \mu > 0 \text{ by assumption})$$

$$2. \quad \text{If } \beta - \gamma - \mu^* < 0 \text{ implies that}$$

$$\beta < \gamma + \mu^*.$$

We can divide both sides by $\gamma + \mu^*$ to get:

$$\frac{\beta}{\gamma + \mu^*} < 1.$$

This fraction on the left hand side of the above inequality is actually our basic reproduction number, R_0 , for this SIR-type model with deaths! That is, it is the term that arises from the stability condition of the DFE. Hence

$$R_0 = \frac{\beta}{\gamma + \mu^*}.$$

From our analysis of the DFE for this previous SIR-type model with deaths, we found its basic reproduction number

$$R_0 = \frac{\beta}{\gamma + \mu^*}. \tag{3.10}$$

Let's glance at this R_0 quantity for a moment. How can we interpret this result? Recall that we need $R_0 < 1$ in order for the disease to go away over time. That would mean that in this case involving natural births and deaths and deaths due to the disease, we need that $\beta < \gamma + \mu^*$. That is, we need the rate in which people recover from the disease (γ) when added to the enhanced death rate due to the disease (μ^*) to be greater than the rate in which people become infected (β).

So, if people are recovering very quickly from the disease or dying from the disease for that matter, those two mechanisms could cause a decrease in R_0 that might be able to

push R_0 to fall below 1... which would then lead to the disease going away. In a positive light this says that if people are recovering quickly, it would imply access to the appropriate, effective healthcare is paramount would have suppress the disease from spreading. Increasing access and the effectiveness of medical treatment would drive γ upwards (and thereby decrease R_0). On the hand, a much darker scenario presents itself - a very large value of μ^* . A large value of μ^* implies a scenario in which the disease is so incredibly dangerous and deadly, that those who get it, pass away extremely quickly. Why would this suppress spread of the disease? Well, this dark scenario implies that since people are passing away from the disease so quickly, they don't have the time to infect anyone else.

Fortunately, for this model to have a $R_0 < 1$, we need the sum of both γ and μ^* to be greater than β . Therefore, *we would hope* that γ is as high as possible and only a minor contribution from μ^* helps make

$$\beta < \gamma + \mu^* \quad \Rightarrow \quad R_0 = \frac{\beta}{\gamma + \mu^*} < 1.$$

3.1.3 The Basic Reproduction Number, Vaccines, and Herd Immunity

At this point we have both seen basic reproduction numbers for a variety of infectious diseases (Table 3.1) and computed R_0 for a few SIR-type models. We saw R_0 incredibly high values for the measles (12-18), chicken pox (10-12), mumps (10-12), and polio (5-7). It may not come as a surprise that due to this level of infectiousness, scientists invested a lot of time and effort into creating vaccines to effectively eradicate them. Many vaccines exist today; however, scientists are not done designing vaccines. Many are continually designing and developing new vaccines that are more safe, more effective, and for diseases to which no known vaccines exist. We define the term **vaccine efficacy** to be the percentage reduction of disease in a vaccinated group of people when compared to an unvaccinated group. In a nutshell, it can be interpreted as the percentage of vaccinated individuals to whom gain immunity from having taken the vaccine.

In a perfect world where vaccine efficacy is 100% (or as high as humanly possible), what epidemiologists and scientists hope for is that a large number of population becomes vaccinated. Through these efforts, epidemiologists believe **herd immunity** can be achieved. Herd immunity is *community protection*. That is, it only occurs when a large portion of the community ("the herd") becomes immune to the disease, thereby substantially lowering the probability (to almost zero) of spreading the disease from person to person. This can be accomplished through the use of safe, effective vaccines. However, consider the following cases illustrated in Table 3.2 for vaccine efficacy and vaccination rates and their possible combined effect on communities.

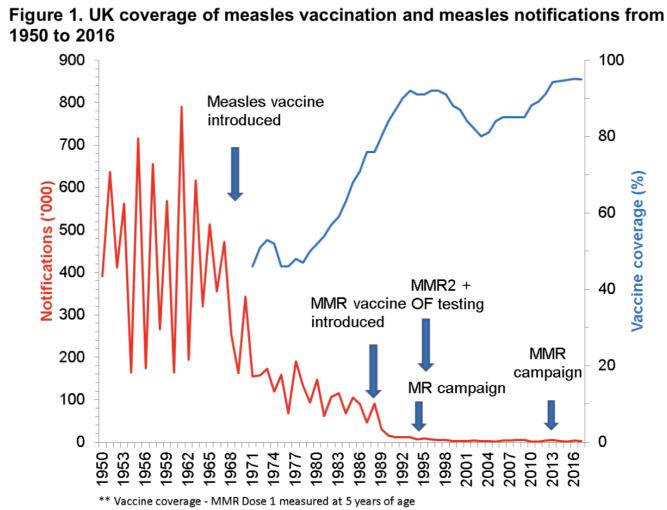


Figure 3.5: Number of Measles cases (notifications) in the UK over time as the vaccine was introduced (as measured in the 100,000's of people) and the percentage of people in the population that were vaccinated [15].

Vaccine Efficacy	Vaccination Rate			Effect on Community
low	+	low	=	No herd immunity
low	+	high	=	Little chance of herd immunity
high	+	low	=	No herd immunity
high	+	high	=	Herd immunity

Table 3.2: Table illustrating how both vaccine efficacy and vaccination rate need to happen in order for herd immunity to be achieved [14].

For example, Figure 3.5 shows how the number of cases of the measles went from ~ 500,000 to nearly 0 from the 1950s to the 2010s due to the introduction of the measles vaccine. Also, it illustrates how significantly the number of cases dropped off once certain percentages of the population became vaccinated over time. Moreover, it shows the effort of the public health community to campaign for the MR and MMR vaccines to ensure that the population remained vaccinated, i.e., children continued to be vaccinated so that there would not be a resurgence of the measles. Once the number of cases is nearly 0, public health officials worry that the general population will forget about the severity and danger of such a disease and stop getting vaccinated. This false sense of security could lead to unfortunate resurgences of an infectious disease. Unfortunately, it seems to have happened in the United States as the number of measles cases in 2019 were the highest there have been since measles was virtually eliminated by 2000 [16].

There are a few ways in which epidemiologists and mathematicians have incorporated the effects of vaccines into the basic reproduction number, R_0 , as well as interpreted

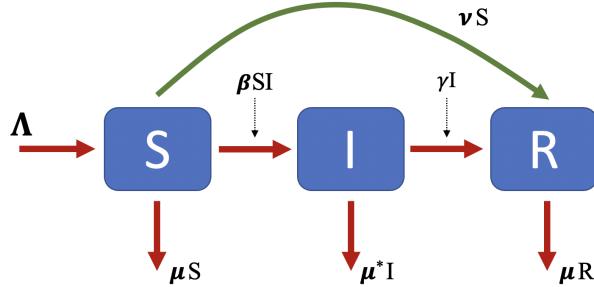


Figure 3.6: Compartment model diagram (box model) for a SIR-type model, which includes natural birth and death terms, an enhanced death dynamic due to the infectious disease, as well as the effect of vaccinations.

how these two ideas may mend together. Namely, there is a very SIR-based mathematical modeling approach to see the influence of vaccination rate on a theoretically derived basic reproduction number. However, there is always an approach in which uses the abstract idea of what a basic reproduction number represents and combines it with vaccine efficacy to seek out herd immunity thresholds. Let's take a look at both!

1. The Mathematical Modeling Approach:

One way is to simply modify our SIR-based model to include a dynamic for vaccines, i.e., modify equations, such as (3.6), to include a term describing the susceptible population getting vaccinated. Let's define ν to be the rate in which the susceptible population gets vaccinated. Furthermore, we can even encapsulate within ν the efficacy of a vaccine as well. With these new vaccination rate, we could modify our general SIR-type model with deaths to the following to include the effect of vaccinations:

$$\begin{cases} \frac{dS}{dt} = \Lambda - \beta SI - \mu S \\ \frac{dI}{dt} = \beta SI - \gamma I - \mu^* I \\ \frac{dR}{dt} = \gamma I - \mu R \end{cases} \Rightarrow \begin{cases} \frac{dS}{dt} = \Lambda - \beta SI - \mu S - \nu S \\ \frac{dI}{dt} = \beta SI - \gamma I - \mu^* I \\ \frac{dR}{dt} = \gamma I - \mu R + \nu S \end{cases} \quad (3.11)$$

This new SIR dynamical system incorporating the effects of vaccines is illustrated through a box diagram in Figure 3.6. The inherent assumptions of the model are as follows:

- (a) There are 3 different sub-populations: S, I, and R
- (b) When people get sick, they might either recover or pass away due to the disease at an *enhanced death rate*. This enhanced death rate includes the natural death rate and death rate due to the disease.
- (c) Natural births and deaths are permissible dynamics within the population.
- (d) Once someone recovers from the disease, they cannot get the disease again.
- (e) We will assume the overall population remains constant over-time, i.e.,

$$N = S + I + R \quad \text{where} \quad \frac{dN}{dt} = 0.$$

Hence we will once again work in terms of population fractions,

$$s = \frac{S}{N}, \quad i = \frac{I}{N}, \quad \text{and} \quad r = \frac{R}{N}.$$

- (f) When a person becomes vaccinated, they are protected from getting the disease. This puts them into the "recovered" class, since here we assume people who have recovered from the disease cannot get it again.

Enforcing a constant population assumption and going through the same rigmarole of finding the disease free equilibrium (working in population fractions, reducing from 3 equations to 2 equations, etc.), gives the following equilibrium values:

$$\bar{s} = \frac{\mu}{\nu + \mu}, \quad \bar{i} = 0, \quad \text{and} \quad \bar{r} = 1 - \bar{s} - \bar{i} = \frac{\nu}{\nu + \mu}.$$

The Jacobian matrix, J , for the system is:

$$J = \begin{bmatrix} -\beta i - \mu - \nu & \mu^* - \mu - \beta s \\ \beta i & \beta s - \gamma - \mu^* \end{bmatrix}.$$

Evaluating J at the equilibrium values, $\bar{s} = \frac{\mu}{\nu + \mu}$ and $\bar{i} = 0$ gives:

$$J = \begin{bmatrix} -(\mu + \nu) & \mu^* - \mu - \beta \left(\frac{\mu}{\nu + \mu} \right) \\ 0 & \beta \left(\frac{\mu}{\nu + \mu} \right) - \gamma - \mu^* \end{bmatrix}.$$

We find that the eigenvalues of the above Jacobian evaluated at the DFE, are the following:

$$\lambda \in \left\{ -(\nu + \mu), \beta \left(\frac{\mu}{\nu + \mu} \right) - \gamma - \mu^* \right\}.$$

Hence we see that the eigenvalue $\lambda = -(\nu + \mu) < 0$ since $\nu > 0$ and $\mu > 0$, so we need to see whether the other eigenvalue is less than zero for stability of the disease free equilibrium. Requiring this, we obtain the following inequality:

$$\beta \left(\frac{\mu}{\nu + \mu} \right) - \gamma - \mu^* < 0$$

$$\frac{\beta \mu}{\nu + \mu} < \gamma + \mu^*$$

$$\frac{\beta \mu}{(\nu + \mu)(\gamma + \mu^*)} < 1.$$

The left hand side of the last inequality above, we can define to be the basic reproduction number of this system:

$$R_0 = \frac{\beta \mu}{(\nu + \mu)(\gamma + \mu^*)}. \quad (3.12)$$

What do we notice from the mathematical form of the above basic reproduction number, R_0 , from (3.12)? Well, let's take a look at the parameter relationships.

- First, we see that the numerator gets larger if β increases. This makes sense - the higher the rate of infection, the higher we'd expect R_0 to be.
- Next, we see that in the denominator that if ν , or γ is very large, that will cause R_0 to decrease. This makes sense since that would insist that many people are becoming vaccinated and that people quickly recover from the disease, leaving fewer people to infect others.
- If μ^* is very large, this will also cause R_0 to decrease. Again, this would be associated with the unfortunate event where an infectious disease is so deadly that people who contract it die at such a high rate that there isn't as much opportunity to spread the disease to others.
- If we know the true values of μ , β , and γ , we can calculate the vaccination rate ν that is required in order for R_0 to be less than one. Thus, finding the critical vaccination rate threshold. However, be careful regarding what the units of ν are when interpreting your findings!

2. Using the concept of R_0 :

If we have an estimated value of R_0 based on medical and survey data, we can try to estimate what the herd immunity thresholds are with a particular vaccine. We will see that that we can take into account a vaccine's efficacy, so this analysis will not assume that vaccines are 100% effective for everyone. Let's begin by defining a few key parameters for us:

- (a) Let p by the percentage of fully immune people needed to eradicate the disease. That is, p , represents the number of fully immune individuals to suppress an epidemic from happening. Therefore, based on this definition, the total number of people left in the susceptible class are $(1 - p)S$. Moreover, if we assume that everyone is susceptible at the start of an epidemic, this just gives us $1 - p$, since $s \approx 1$.
- (b) We can define what is called the **effective reproduction number** (R_E), as the product of the basic reproductive number, R_0 , and the fraction of the population that is susceptible, i.e.,

$$R_E = R_0 \cdot s = R_0(1 - p).$$

Similarly, to eradicate a disease, R_E must be less than one.

Therefore, we can incorporate our definition of p and the effective reproduction number, R_E , in the following manner:

$$R_E = R_0(1 - p) < 1.$$

We can go ahead and solve this inequality for p to give us what percentage of people need to be fully immune in order to eradicate the disease:

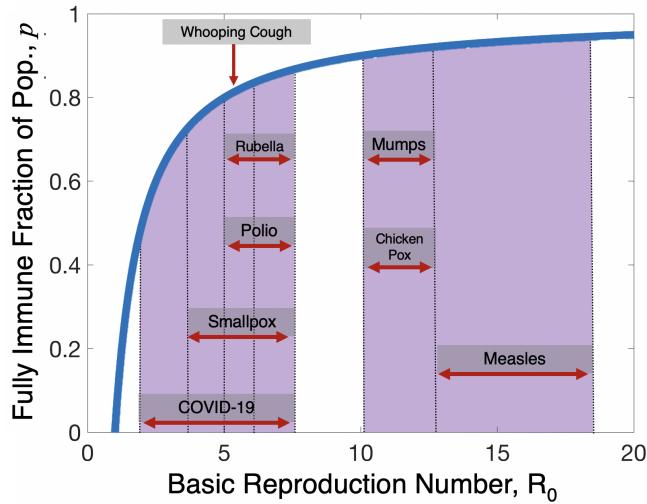


Figure 3.7: Showing what fraction of the population must have full immunity in order to achieve herd immunity as a function of the basic reproduction number.

$$R_0(1 - p) < 1$$

$$1 - p < \frac{1}{R_0}$$

$$-p < \frac{1}{R_0} - 1$$

$$p > 1 - \frac{1}{R_0}.$$

Thus $p > 1 - \frac{1}{R_0}$ is the immunization threshold to eliminate a disease from a population, i.e., the **herd immunity threshold**. However, we have not yet incorporated the effect of non-100% vaccine efficacy. Figure 3.7 shows what fraction of the population must have full immunity in order to achieve herd immunity for various basic reproduction numbers. Therefore Figure 3.7 could be interpreted as showing what fraction of the population must be vaccinated in order to achieve herd immunity as a function of the basic reproduction number, if assuming 100% vaccine effectiveness. However, since vaccines unfortunately do not have 100% efficacy, the real life threshold values must be larger than what Figure 3.7 illustrates!

Let's now incorporate vaccine efficacy into these calculations.

- (c) Let's define E to be the vaccine efficacy, i.e., the fraction of people who get the vaccine and develop immunity. If $E = 1$, then 100% of

individuals who take the vaccine develop immunity, while if $E = 0.25$, then only 25% of the individuals who take the vaccine develop immunity. Thus, if a vaccine is very effective, E would be very close to one.

- (d) We can now define V_c to be the vaccine coverage that is required for herd immunity. We can now mathematically define V_c in terms of p and E :

$$V_c = \frac{p}{E} \quad \rightarrow \quad p = V_c \cdot E.$$

Substituting this new mathematical definition of p in terms of V_c and E into our herd immunity threshold from before we get that:

$$p = V_c \cdot E > 1 - \frac{1}{R_0}.$$

Therefore, we find that for a vaccine with an efficacy of E , that to achieve herd immunity in a community that the vaccination coverage must satisfy:

$$V_c > \frac{1}{E} \left(1 - \frac{1}{R_0} \right). \quad (3.13)$$

For example, in terms of measles, let's assume a R_0 on the lower end of its estimated range: $R_0 = 12$. The measles vaccine is known to be roughly 97.5% effective, i.e., $E = 0.975$. Therefore to achieve herd immunity, from (3.13) we see that

$$V_c > \frac{1}{0.975} \left(1 - \frac{1}{12} \right) \approx 0.94.$$

Therefore, over 94% of the population needs to be vaccinated for herd immunity to take place.

3.2 Random Walks, Brownian Motion, and Diffusive Processes

In every section of this book until now, we've modeled biological systems using either discrete or differential equation model systems. In fact, all of the models we've analyzed so far are what we call **deterministic** models. Simply put, **deterministic** equations are equations with no inherent randomness or *stochastic* elements involved. In practical terms this means that if we solve these equations today, we will find the same exact solutions as we will tomorrow or if we solved them yesterday, last week, or 17 days from now (assuming the same numerical methods and parameter values, of course). However, the world as we know it is inherently random.

For example, in the nanoworld of cells, things are inherently random. Protein–protein interactions (PPIs) drive many molecular and regulatory processes in your body. These interactions result from biochemical reactions that occur due to electrostatic forces or hydrogen bonding and/or hydrophobic effects. Such PPIs give rise to the notion of a *molecular machine*, e.g., a thing that performs the specific tasks that are essential for life, like DNA replication and ATP synthesis. However, such protein–protein interactions seem to occur stochastically (randomly). In fact, many quantitative research efforts have tried to describe these random events [1], [2] [3] [4]. Therefore, even the way our body works

3.2. RANDOM WALKS, BROWNIAN MOTION, AND DIFFUSIVE PROCESSES 179

at the smallest scales appears to hinge on random processes.

There are a few interesting things to think about regarding these ideas of randomness:

- **Biological question:** If everything is so random in the nanoworld of cells, how can we say anything predictive about what's going on there?
- **Physical Idea:** Could the *collective* activity of *many* randomly moving things be effectively predictable, even if the individual motions are not...?

Here we will start dabbling in the world of modeling random behaviors. We will develop intuition and a theory for what are called **random walks**. That is, what if every choice or decision in life was governed by the flip of a coin? Eventually this will lead us to modeling these such random processes using one collective notion, called **diffusion**. In a biological context, diffusion is the process that governs the bulk movement of particles, molecules, or even animals from an area of high concentration to an area of low concentration. You can think of diffusion as a *spreading out* of some quantity. How will we go from coin flips to diffusion, welp - read on! We're going to get to use ideas from probability theory and calculus ☺, but first I'll leave us with a few examples of diffusive processes in biology:

- Bacteria moving within the gut.
 - Plant dispersal (dandelions, trees).
 - Insects moving through a field.
 - Brine shrimp movement and many other zooplankton movements.
 - Some animals' foraging behavior and/or method for finding their way home
 - Chemical signals given off by an animal looking for a mate
 - Forest fires spreading
 - Heat dissipation
 - Air pollution mixing with the gases in our atmosphere
 - Proteins moving into membranes and organelles.
 - Net movement of glucose down the concentration gradient
 - Oxygen and carbon dioxide across the alveolar-capillary membrane of mammalian lungs
 - Thyroid gland secreting a chemical causing your bones to release calcium, which then diffuses into your blood when there isn't enough calcium in your blood
 - Nephrons in your kidney separating blood from waste chemicals and toxins, which then reabsorb the water and nutrients in the blood through diffusion.
 - Your body is in a constant state of diffusing water from one place to another
- ... *diffusive processes are all around us, all the time.*

3.2.1 Random Walks... you're tellin' me we can just flip a coin?

Let's begin by playing a game. Flip a **fair, unbiased** coin. By fair and unbiased, I mean a coin in which the likelihood of a heads or tails is exactly equal at 50%. Let's lay down a few rules. If you get heads, you step one unit to the right, tails, one unit to the left. Say we flip the count 7 times and get in succession the following:

1. heads

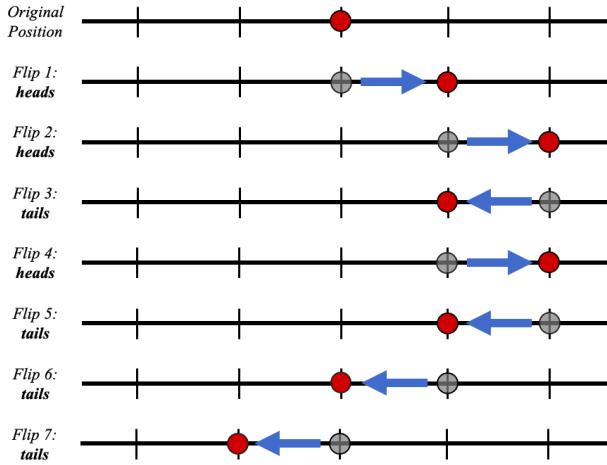


Figure 3.8: An illustration of the random walk sequence based on flipping a coin 7 times. Notice that after 7 flips, we eventually end up the left the one where we started - one position over.

2. heads
3. tails
4. heads
5. tails
6. tails
7. tails

If we try to place that on a grid, starting at a central point, we could illustrate our movement as the following sequence of events as in Figure 3.8. The process of our left or right movements being dictated by the flip of a coin is precisely how a *random walker* would move in one dimension (along one line left or right). Let's define these conditions a bit more stringently.

Definition 2.1

A **random walk** is a path that consists of a succession of random steps. These steps must obey the following conditions:

1. The location of the walker can only jump to neighboring sites of the lattice - that is, they cannot skip steps in-between locations (in this example - it can only move left or right).
2. The probability of the location of the next step is based only on the current time. There is no memory. The previous steps do not influence the next step. It is a *Markov process*.

After 7 coin flips we eventually end up one position over to the left from where we started. You might imagine that if we did this 7 coin flip experiment over and over, say 1000 times, that on average we might start ending up directly back where we started more often than any other ending position after 7 flips. Actually, mathematically speaking -

that is exactly what we'd expect. Let's investigate this idea in a very mathematical fashion. To start, let's define some quantities to help get us started:

- Let's define the unit in which we move to the right or the left to be a unit L for every coin flip.
- Let's define l_j to be how far we move in the j^{th} coin flip. The only possible values of l_j are either a $+L$ or $-L$, i.e.,

$$l_j = \begin{cases} +L & (\text{heads}) \\ -L & (\text{tails}) \end{cases},$$

since we can only move left or right by a unit of L .

- Let x be the distance traveled by a random walker after N flips of a coin. Hence the total distance x could be represented as the following sum:

$$x = l_1 + l_2 + l_3 + \dots + l_N.$$

Note that in the summation above, the right hand side is a big ole sum of simply items that are either $+L$ or $-L$.

- For example, for our 7 coin flip example above, we would find

$$\begin{aligned} x &= \sum_{j=1}^{N=7} l_j \\ &= l_1 + l_2 + l_3 + l_4 + l_5 + l_6 + l_7 \\ &= (+L) + (+L) + (-L) + (+L) + (-L) + (-L) + (-L) \\ &= L + L - L + L - L - L - L \\ &= -L, \end{aligned}$$

since $l_1 = l_2 = l_4 = +L$ (heads) and $l_3 = l_5 = l_6 = l_7 = -L$ (tails).

Now to compute where we'd *expect* to end up after N coin flips, we get to introduce an idea from probability theory - the *expected value*.

Definition 2.2

The *expected value* (or *expectation*) refers to the value of random variable one would "expect" to find if one could repeat the random variable process an infinite number of times and take the average of the values obtained.

Note:

- The expected value is a weighted average of all possible values.
- The notation for expected values are angle bracket, e.g., the expected value of an event q is denoted: $\langle q \rangle$.
- The expected value operator is a **linear operator**, i.e.,

$$\langle q_1 + q_2 + q_3 + q_4 \rangle = \langle q_1 \rangle + \langle q_2 \rangle + \langle q_3 \rangle + \langle q_4 \rangle.$$

- As an example, consider $\langle l_1 \rangle$: *the average position after the first coin toss*. Another way to view this is the average position after one coin toss by summing the probability of each outcome with the distance traveled, i.e.,

$$\langle l_1 \rangle = \left(\frac{1}{2}\right)(+L) + \left(\frac{1}{2}\right)(-L) = \frac{L}{2} - \frac{L}{2} = 0,$$

where the $(1/2)$'s are because the probability of going to the left or the right is 50% or $(1/2)$ for a **fair, unbiased coin**.

- For example, consider $\langle l_1 \rangle$, but this time the coin is **biased** and lands on heads 70% of the time (and tails only 30% of the time). This time when we compute $\langle l_1 \rangle$, we find:

$$\langle l_1 \rangle = \left(\frac{7}{10} \right) (+L) + \left(\frac{3}{10} \right) (-L) = \frac{7L}{10} - \frac{3L}{10} = \frac{4L}{10}.$$

Thus, for this **biased** coin we would expect that after one single coin flip we might have moved a distance of $\frac{4L}{10}$ to the right.

Okay, let's find out on average where we expect to end up after N coin tosses for a **fair, unbiased coin**. Therefore, let's get to computing $\langle x \rangle$:

$$\begin{aligned} \langle x \rangle &= \langle l_1 + l_2 + l_3 + \dots + l_N \rangle \\ &= \langle l_1 \rangle + \langle l_2 \rangle + \langle l_3 \rangle + \dots + \langle l_N \rangle \quad (\text{via linearity}) \\ &= \left[\left(\frac{1}{2} \right) (+L) + \left(\frac{1}{2} \right) (-L) \right] + \left[\left(\frac{1}{2} \right) (+L) + \left(\frac{1}{2} \right) (-L) \right] + \left[\left(\frac{1}{2} \right) (+L) + \left(\frac{1}{2} \right) (-L) \right] + \dots \\ &\quad \dots + \left[\left(\frac{1}{2} \right) (+L) + \left(\frac{1}{2} \right) (-L) \right] \\ &= (0) + (0) + (0) + \dots + (0) \\ &= 0. \end{aligned}$$

Therefore, on average we would expect to go nowhere if our movements were dictated by a **fair, unbiased coin**. That is, our average displacement would be 0. If we applied this formalism to our random walker, it would *on average* be right back where it started. *Okay, so what?* Welp, at least our intuitive idea was confirmed from earlier. But wait! An interesting thing happens when we stop only considering the average position (displacement) and instead consider the average **square** of the displacement, i.e., $\langle x^2 \rangle$. Let's go ahead and compute that quantity...

$$\begin{aligned}
\langle x^2 \rangle &= \langle (l_1 + l_2 + l_3 + \dots + l_N)^2 \rangle \\
&= \langle (l_1 + l_2 + l_3 + \dots + l_N)(l_1 + l_2 + l_3 + \dots + l_N) \rangle \quad (\text{distributing} \dots) \\
&= \langle (l_1^2 + l_2^2 + l_3^2 + l_N^2) + (l_1 l_2 + l_1 l_3 + \dots + l_1 l_N + l_2 l_1 + l_2 l_3 + l_2 l_4 + \dots) \rangle \\
&= \left\langle \sum_{j=1}^N l_j^2 + \sum_{j \neq k} l_j l_k \right\rangle \quad (\text{using series notation} \dots) \\
&= \sum_{j=1}^N \langle l_j^2 \rangle + \sum_{j \neq k} \langle l_j l_k \rangle \quad (\text{via linearity})
\end{aligned}$$

Alright, it looks like we have a little work ahead of us. So far we have found that

$$\langle x^2 \rangle = \sum_{j=1}^N \langle l_j^2 \rangle + \sum_{j \neq k} \langle l_j l_k \rangle.$$

We will not set our sights on figuring out what those two summations are equal to on the right hand side of the above sum.

$$1. \sum_{j=1}^N \langle l_j^2 \rangle:$$

To do this, it all hinges on us figuring out what $\langle l_j^2 \rangle$ is. Now, while we may read this at first as the expectation value of two of the same flip occurring, we *must* realize that we are only considering cases in which the same flip occurs.

That is to say, we must assume that we either have a heads-heads or a tails-tails situation occur. Those are the only two possibilities here. Therefore if there are only those two possibilities available, each must have an associated probability of 1/2. Since we are using a fair, unbiased coin and that these are the only two possibilities that could occur, each must have an occurrence probability of 50% associated with them. To emphasize, we are not asking what the probability is of getting two heads or two tails in a row, but rather that we know we must have gotten either two heads or two tails in a row.

Hence for a **fair, unbiased** coin, it would be:

$$\langle l_j^2 \rangle = \left(\frac{1}{2} \right) (+L)(+L) + \left(\frac{1}{2} \right) (-L)(-L) = \frac{L^2}{2} + \frac{L^2}{2} = L^2.$$

Therefore, the entire summation reduces to:

$$\sum_{j=1}^N \langle l_j^2 \rangle = \sum_{j=1}^N L^2 = N L^2.$$

$$2. \sum_{j \neq k} \langle l_j l_k \rangle:$$

Similarly, to understand what this summation is equal to, let's begin by first considering one specific case, say $\langle l_j l_k \rangle$. Note that while we assume that $j \neq k$ in the summation, for the notation that $\langle l_j l_k \rangle$ we must still consider the possibilities of landing two heads or tails in a row for a coin toss, not only a head then a tails or vice versa. By the $j \neq k$ what we are saying is that we are considering the case separately where we *only* consider the case of two heads or two tails.

Therefore, if we consider all possible combinations, for a **fair, unbiased** coin this leads us to:

$$\begin{aligned} \langle l_j l_k \rangle &= \left(\frac{1}{4}\right)(+L)(+L) + \left(\frac{1}{4}\right)(+L)(-L) + \left(\frac{1}{4}\right)(-L)(-L) + \left(\frac{1}{4}\right)(-L)(+L) \\ &= \frac{1}{4} [L^2 - L^2 + L^2 - L^2] \\ &= 0. \end{aligned}$$

Therefore, all together we find that this implies that

$$\sum_{j \neq k} \langle l_j l_k \rangle = 0.$$

If we put both of these computations together we find that expected square of the displacement (or **mean squared displacement**, MSE) $\langle x^2 \rangle$ is:

$$MSE = \langle x^2 \rangle = \sum_{j=1}^N \langle l_j^2 \rangle + \sum_{j \neq k} \langle l_j l_k \rangle = NL^2 + 0 = NL^2,$$

and thereby we obtain

$$MSE = \langle x^2 \rangle = NL^2.$$

Finally, we can compute what is called the **root-mean squared displacement**, the **RMS**. It is given by \sqrt{MSE} or equivalently, $\sqrt{\langle x^2 \rangle}$:

$$RMS = \sqrt{MSE} = \sqrt{\langle x^2 \rangle} = \sqrt{NL}. \quad (3.14)$$

In a practical sense what this result means is that a typical random walker will have moved a distance of \sqrt{NL} from its starting point after N steps. It effectively measures how far it is at the end of the N steps from its initial starting position. Note that distance cannot be a negative value, i.e., we basically computed it via the *square root of the mean squared distance*. The factor of \sqrt{N} in the RMS is the **fundamental feature of a random walker**.

For example, say we know that for a particular tiny random walker that it takes 1 minute to travel 1 meter. We can then use (3.14) to determine that it will take 100 minutes to travel 10 meters from its starting point. To do this, we would consider N to be the

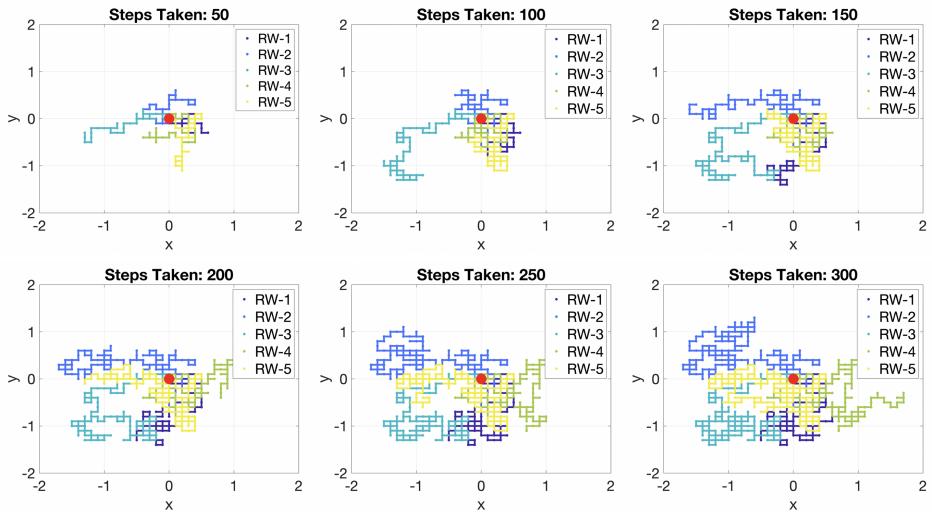


Figure 3.9: Snapshots from a two-dimensional random walk process for 5 different random walkers!

"time" give by the number of minutes (steps) and L is the step-size, here 1 meter. Thus we can set up the problem in the following manner, assuming we are looking for an $RMS = 10$:

$$RMS = 10 \text{ meters} = \sqrt{N \text{ minutes}} \cdot (1 \text{ meters}) \Rightarrow N = (10)^2 = 100.$$

Interestingly enough, for a 2D random walk these results also apply. A two-dimensional random walk on a lattice is when a random walker gets the option to move right, left, up, or down, thus moving on a 2D grid. The steps for a two-dimensional random walk involve:

Step 1: Start at an initial position, usually the origin in an xy -plane.

Step 2: Randomly choose 1 or 4 nearest neighboring grid points to move to.
This could involve rolling a $d4$ - a tetrahedron shaped die.

Step 3: Make the jump to the randomly selected nearest neighbor site - that is, move up, down, right or left accordingly to Step 2.

Step 4: Go back to Step 2, and repeat for the desired amount of steps.

An illustration of 2D random walks is presented in Figure 3.9. There are 5 different random walkers and the successive images illustrates how the random walkers move as they take more and more steps. Notice how in this random walk formulation, each random walker only walks along certain (x,y) grid points. That is, they *cannot* cut across a grid cell diagonally.

You are welcome to go through the same type of analysis above for the two-dimensional random walker case or run numerical experiments by simulating large amounts of random walkers, to illustrate the **fundamental feature of a random walker** still holds true for two-dimensional random walks. It will even be true in any number of dimensions we want to randomly walk, whether it be 3, 4, 5, ..., 716, etc., but what does that even physically mean? As long as the way in which we determine which direction we step is **unbiased**

this result scales to any number of dimensions we wish. In fact, we can write it as the following mathematical hammer (theorem):

Theorem 2.1

For a random walk process in 1D, 2D, or any dimensions in which we perform N steps, each of a step-size of L , the root mean squared displacement, RMS , is equal to:

$$RMS = \sqrt{\langle x^2 \rangle} = \sqrt{NL}.$$

While we are on the topics of random walks in different amount of dimensions. The Hungarian mathematician George Pólya proved an interesting theorem regarding random walks and whether or not they are ever expected to cross the starting point again. To summarize his findings, he proved that the probability of ever returning back to the starting place for a n -dimensional random walk would be far less than 100% when $n \geq 3$. Later, other mathematicians computed approximate probability values for numerous dimensions, as shown in Table 3.3. Note that these results do not just apply to re-encountering the initial starting place, but any point on these n -dimensional lattices.

Therefore based on these findings, as a human if we were to find ourselves in a new city, we could guarantee ourselves that we could randomly walk around and eventually we would reach the place we wanted to go. However, if a fish or bird tries the same, there is only a $\sim 34\%$ guarantee they could find the place they were looking for (of course this assumes that the skies or oceans were infinite in size and that they were *randomly* moving around without any preferential direction). The Japanese-American mathematician, Shizuo Kakutani put it much more eloquently, “A drunk man will find his way home, but a drunk bird may get lost forever.”

# of Dimensions	Probability of returning to the starting point
1	100%
2	100%
3	$\approx 34.05\%$
4	$\approx 19.32\%$
5	$\approx 13.52\%$
6	$\approx 10.47\%$
7	$\approx 8.58\%$
8	$\approx 7.29\%$

Table 3.3: Table of probabilities associated with ever crossing the starting place for a random walk in an increasing number of dimensions [5].

Before leaving George Pólya and random walk theory, I want to mention some excerpts from his famous book How To Solve it [6]. In this book he offers a list of four principle steps for trying to solve any problem. While his book is meant to be applied for solving mathematical problems, the abstract ideas can be broadly applied. In short, they are the following:

Strategy 2.1

Pólya's Problem Solving Principles

1. Understand the problem

This seems so obvious that it is often not even mentioned, yet many efforts are stymied simply because people don't fully understand what a problem is asking. Before attempting a solution, Pólya suggests reinterpreting what a problem is, by asking yourself the following questions:

- Are you able to restate the problem in your own words?
- Do you understand all the words used in stating the problem?
- What are you asked to find or show?
- Can you offer a picture or diagram that might help someone understand the problem?
- Is there actually enough information to find a solution?

2. Devise a plan

There are many reasonable ways to solve problems; however the skill of choosing an appropriate strategy is only learned by solving many problems. This principle asks you to be patient and think whether you have ever seen or know of a related problem? If so, could you apply that strategy or parts thereof to your current problem?

In general, here is where we try to find the connection between what we known and what we may not know. It is most certainly tricky.

Over time, you will develop more and more strategies for solving problems. Some possible ideas include:

- Guess and check - why not try something out and see what happens?
- Make an orderly list of ideas
- Eliminate possibilities
- Consider special cases
- Look for a pattern
- Draw a picture(s)
- Solve a similar, but simpler problem
- Work backward
- Use direct reasoning
- Use a formula and/or solve an equation

3. Carry out the plan

This step is usually easier than devising the plan. In general, all you need is care and patience to carry out your idea. Persist with the plan that you have chosen. Develop grit - if this plan didn't lead anywhere or solve your problem, step back and choose another. Don't be fooled, this is how the world's

most famous mathematicians and scientists work - they are actually experts at figuring out what went wrong... and then trying to figure out why.

4. *Reflect on what you did; look back*

This principle often gets overlooked. It's not surprising why - it requires further time investment beyond having actually solved the problem; however, here is where much growth can occur!

Much can be gained by taking the time to reflect and look back at what you have done, what worked, and what didn't. This can enlighten you to seeing the applicability of this procedure or strategy to other problems. Investing the time to do this will help enable you to predict what strategy to use to when being presenting with future problems.

□

3.2.2 Building towards Diffusion from Random Walks!

Consider M trees surrounding a circular lake, as shown in Figure 3.10. We will assume that in each tree there are insects to whom are only allowed to jump tree-to-tree around the lake. Much in the same vein as our random walk process above, each step forward in time, we will assume that the insects have the option to either:

1. Move to the tree to their right
2. Move to the tree to their left
3. Stay in the same tree

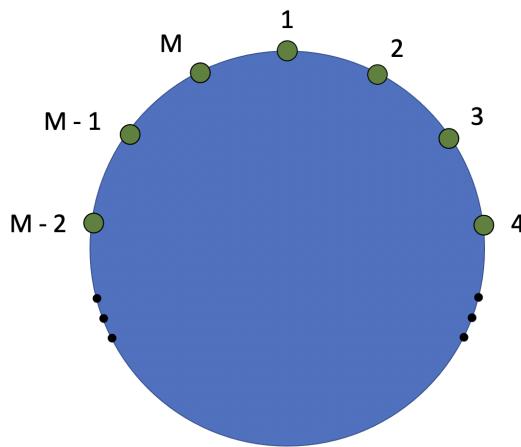


Figure 3.10: M trees all sitting around a circular lake. In each tree some fraction of the overall insect population is sitting at a particular moment in time.

Let's start by setting some ground rules (pun intended):

Rules and Definitions:

1. Each forward step in time, the insect can either move left, move right, or stay put in the same tree. For our considerations here, we will assume that all the insects must be in one of the trees. Further, we will not assume any births or deaths take place.
2. We will assume that if the insects choose to jump to another tree that they choose either the left or the right in an unbiased fashion. In other words, there is no difference between trees in terms of food sources or predators.
3. Since it is an unbiased decision, some proportion q will elect to jump to the tree to their left
4. Similarly, some proportion q will choose to jump to the tree to their left
5. Therefore the proportion of insects that decided to stay in the same tree is $1 - 2q$.
6. We will let $U_{m,n}$ denote the proportion of total insects in *tree m* at *time n*.
7. At any given time n , ALL the insects must be in one tree or another around the lake. Therefore, if we add together ALL of the proportions $U_{m,n}$ for each every tree, it must sum to 1. That is, adding together the insect population fractions in every tree around the lake must add to 1.

$$\sum_{j=1}^M U_{j,n} = U_{1,n} + U_{2,n} + U_{3,n} + \dots + U_{M,n} = 1 \quad \text{for any time } n.$$

Next we wish to see what happens when we update the insect proportions in each tree other time. That is, we are going to analyze $U_{m,n+1}$ for each tree m , as we march forward in time from time $n \rightarrow n + 1$. We are going to do this in two parts. At the end of the second part we will see an equation for diffusion pop out!

Part 1: Analyze $U_{m,n+1}$

To get $U_{m,n+1}$, we must add together all three possibilities of where an insect may move from to get into tree m at time $n + 1$. Thus, there are 3 cases to consider of where an insect may come from, i.e.,

1. They were in tree m and stay in tree m .

We can describe this dynamic of the insects staying in the same tree by the following proportion:

$(1-2q)U_{m,n}$: Proportion of insects in tree m that stay in tree m .

2. They were in tree $m + 1$ and move left to go into tree m .

We can describe this dynamic of the insects moving from tree $m+1$ to m as:

$qU_{m+1,n}$: Proportion of insects in tree $m+1$ moving to tree m .

3. They were in tree $m - 1$ and move right to go into tree m .

We can describe this dynamic of the insects moving from tree $m-1$ to m as:

$qU_{m-1,n}$: Proportion of insects in tree $m-1$ moving to tree m .

Putting these three ideas together we get the total proportion of insects in tree m at time $n + 1$ as:

$$U_{m,n+1} = (1 - 2q)U_{m,n} + qU_{m+1,n} + qU_{m-1,n}. \quad (3.15)$$

We can use (3.15) to get the proportion of insects at any tree at any time in the future based off the previous insect locations. Note that if we had an initial population fraction for each of the m trees and a value of q , we could update (3.15) just like any discrete equation we did previously. That is, (3.15) is a discrete equation!

Similar to the idea we first encountered in Section 2.2, where we transitioned from discrete to continuous equations, we can apply the same principles to (3.15). This will allow us to see its continuous (differential) equation analog.

...spoiler - this is how the equation governing diffusion will show up!

Part 2: Transition (3.15) to its continuous counterpart formulation

First we will try to rewrite $U_{m,n+1}$ in terms of continuous variables x and t . How will we do that? Let's start by defining some quantities to help the transition go smoother. We will let:

- Δx = distance between adjacent trees, and
- Δt = the duration of each step forward in time.

Therefore, after time n has passed, we can define the actual time to be

$$t = n\Delta t,$$

that is, the total amount of time that has passed after n steps. Similarly, we see that we can define a distance x as

$$x = m\Delta x.$$

This distance x gives the distance to the m^{th} tree around the lake. Now we can define a multivariable function u in terms of x and t . . . err, we can see u defined in terms of x and t by way of variables $m, \Delta x, n$, and Δt . Check it out:

$$u(x, t) = u(m\Delta x, n\Delta t) = U_{m,n}.$$

Let's revisit our discrete pal back in (3.15) and try to rewrite each of its terms in terms of our newly defined variables $m, \Delta x, n$, and Δt and thereby casting it towards x and t .

Recall that (3.15) says:

$$\underbrace{U_{m,n+1}}_{\substack{\text{fraction in} \\ \text{tree } m \\ \text{time } n+1}} = \underbrace{(1-2q)U_{m,n}}_{\substack{\text{fraction staying} \\ \text{in tree } m \\ \text{from time } n \rightarrow n+1}} + \underbrace{qU_{m+1,n}}_{\substack{\text{fraction moving} \\ \text{from tree } m+1 \\ \text{to tree } m \\ \text{from time } n \rightarrow n+1}} + \underbrace{qU_{m-1,n}}_{\substack{\text{fraction moving} \\ \text{from tree } m-1 \\ \text{to tree } m \\ \text{from time } n \rightarrow n+1}} \quad (3.16)$$

We will now cast each of those 4 terms in the above equation into the form using the newly defined variables. Doing so gives the following relationships:

$$\begin{aligned} U_{m,n} &= u(m\Delta x, n\Delta t) = u(x, t) \\ U_{m,n+1} &= u(m\Delta x, (n+1)\Delta t) = u(x, t + \Delta t) \\ U_{m+1,n} &= u((m+1)\Delta x, n\Delta t) = u(x + \Delta x, t) \\ U_{m-1,n} &= u((m-1)\Delta x, n\Delta t) = u(x - \Delta x, t). \end{aligned}$$

Therefore we can rewrite (3.16) as the following by substituting in the above relations

$$u(x, t + \Delta t) = (1-2q)u(x, t) + qu(x + \Delta x, t) + qu(x - \Delta x, t),$$

which we can then write as what's below after distributing the $(1-2q)$ term

$$u(x, t + \Delta t) = u(x, t) + qu(x + \Delta x, t) - 2qu(x, t) + qu(x - \Delta x, t).$$

We will next subtract $u(x, t)$ from both sides to get

$$u(x, t + \Delta t) - u(x, t) = qu(x + \Delta x, t) - 2qu(x, t) + qu(x - \Delta x, t),$$

and finally after pulling a factor of q out of all three terms on the right hand side of the above equation, we arrive at

$$u(x, t + \Delta t) - u(x, t) = q \left[u(x + \Delta x, t) - 2u(x, t) + u(x - \Delta x, t) \right]. \quad (3.17)$$

Believe or not, we are well on our way! The remaining steps we have are to mathematically massage both sides of this equation into a type of difference quotient that will (*spoiler*) give us a first derivative or second derivative related quantity! Actually, let's just go ahead and ruin the surprise. Remember our definition of a partial derivative? Yup - we're going to use that right now (see Definition 2.3)!

Definition 2.3

The **first-order partial derivative** of a function $f(x, t)$ with respect to t , denoted as either $\frac{\partial f}{\partial t}$ or $f_t(x, t)$, is defined as

$$f_t(x, t) = \lim_{\Delta t \rightarrow 0} \frac{f(x, t + \Delta t) - f(x, t)}{\Delta t}.$$

In fact, we can piggyback off of Definition 2.3 to help us define what we mean by the second-order partial derivative of a function with respect to an independent variable, see Definition 2.4. In essence, we will apply the same difference quotient type logic, but to the first-order partial derivative of a function! It is the *derivative of a derivative* after all.

Definition 2.4

The **second-order partial derivative** of a function $f(x, t)$ with respect to x , denoted as either $\frac{\partial^2 f}{\partial x^2}$ or $f_{xx}(x, t)$, is defined as

$$f_{xx}(x, t) = \lim_{\Delta x \rightarrow 0} \frac{f_x(x, t) - f_x(x - \Delta x, t)}{\Delta x}.$$

How in the world does this help us see (3.17) in a new way? We'll get to that very shortly! First, let's play with the definition of the second-order partial derivative, as given in Definition 2.4. What I would like us to do together is mathematically massage the difference quotient given on the right hand side of the term in Definition 2.4 into differences involving only the original function f and not any partial derivatives (so no f_x terms). This will involve substituting what we know from Definition 2.3 regarding a first-order partial derivative into Definition 2.4, which involves second-order partials.

From Definition 2.4, we have the following

$$f_{xx}(x, t) = \lim_{\Delta x \rightarrow 0} \frac{f_x(x, t) - f_x(x - \Delta x, t)}{\Delta x}$$

Let's use the definition of a first-order partial derivative to rewrite the terms on the right hand side of the above equation, i.e.,

$$f_x(x, t) = \lim_{\Delta x \rightarrow 0} \frac{f(x + \Delta x, t) - f(x, t)}{\Delta x}$$

and

$$f_x(x - \Delta x, t) = \lim_{\Delta x \rightarrow 0} \frac{f(x, t) - f(x - \Delta x, t)}{\Delta x}.$$

Substituting these quantities into our definition of a second-order partial derivative gives us:

$$f_{xx}(x, t) = \lim_{\Delta x \rightarrow 0} \frac{\left[\frac{f(x + \Delta x, t) - f(x, t)}{\Delta x} \right] - \left[\frac{f(x, t) - f(x - \Delta x, t)}{\Delta x} \right]}{\Delta x}.$$

We can further reduce the right hand side of the above equation by taking out a factor of Δx in each denominator of the fractions in the numerator to combine with the one already in the overall denominator to arrive at the following

$$f_{xx}(x, t) = \lim_{\Delta x \rightarrow 0} \frac{\left[f(x + \Delta x, t) - f(x, t) \right] - \left[f(x, t) - f(x - \Delta x, t) \right]}{\Delta x^2}.$$

From here we can simply the term inside the large square brackets to get

$$f_{xx}(x, t) = \lim_{\Delta x \rightarrow 0} \frac{1}{\Delta x^2} \left[f(x + \Delta x, t) - 2f(x, t) + f(x - \Delta x, t) \right],$$

and the final form we will use that is:

$$f_{xx}(x, t) = \lim_{\Delta x \rightarrow 0} \frac{f(x + \Delta x, t) - 2f(x, t) + f(x - \Delta x, t)}{\Delta x^2}. \quad (3.18)$$

Oof. Welp, we're finally ready to revisit that insect equation (3.17). We'll provide the equation again below. Note that the name of the game now is pushing to substitute those definitions of first-order and second-order partial derivatives into (3.17). But wait, where are the limits? Where are those small quantities Δt or Δx^2 ? We're going to use rules of Calculus to help us bring them in! Here is the equation we're working with from earlier (3.17):

$$u(x, t + \Delta t) - u(x, t) = q \left[u(x + \Delta x, t) - 2u(x, t) + u(x - \Delta x, t) \right].$$

First, let's divide both sides by Δt . Recall that Δt is a non-zero quantity that eventually we will drive towards zero in a limit. However, for now we will simply divide both sides by Δt :

$$\frac{u(x, t + \Delta t) - u(x, t)}{\Delta t} = \frac{q}{\Delta t} \left[u(x + \Delta x, t) - 2u(x, t) + u(x - \Delta x, t) \right].$$

Hopefully the left-hand side of the above equations looks somewhat familiar → first-order partial derivative anyone? However, there isn't a limit... yet. Next let's take the limit of both sides of the equation as $\Delta t \rightarrow 0$:

$$\underbrace{\lim_{\Delta t \rightarrow 0} \frac{u(x, t + \Delta t) - u(x, t)}{\Delta t}}_{\text{equal to}} = \lim_{\Delta t \rightarrow 0} \frac{q}{\Delta t} \left[u(x + \Delta x, t) - 2u(x, t) + u(x - \Delta x, t) \right].$$

From Definition 2.3 we see that the left hand side is actually $\frac{\partial u}{\partial t}$, i.e.,

$$\frac{\partial u}{\partial t} = \lim_{\Delta t \rightarrow 0} \frac{q}{\Delta t} \left[u(x + \Delta x, t) - 2u(x, t) + u(x - \Delta x, t) \right].$$

Looking at the right hand side of the above equation, we might notice that the term in square brackets looks awfully similar to the numerator in a second-order partial derivative, i.e., compare to (3.18). However, we are missing a division by Δx^2 and a limit as Δx goes to 0. Let's use our math powers, to make those things happen!

To get a division by Δ^2 on the right hand side, let's first multiply by $\frac{\Delta x^2}{\Delta x^2}$, as in

$$\begin{aligned} \frac{\partial u}{\partial t} &= \frac{\Delta x^2}{\Delta x^2} \lim_{\Delta t \rightarrow 0} \frac{q}{\Delta t} \left[u(x + \Delta x, t) - 2u(x, t) + u(x - \Delta x, t) \right] \\ &= \lim_{\Delta t \rightarrow 0} \frac{q \Delta x^2}{\Delta t} \left[\frac{u(x + \Delta x, t) - 2u(x, t) + u(x - \Delta x, t)}{\Delta x^2} \right]. \end{aligned}$$

From here we will take a limit as Δx goes to zero on both sides, i.e.,

$$\lim_{\Delta x \rightarrow 0} \frac{\partial u}{\partial t} = \lim_{\Delta x \rightarrow 0} \lim_{\Delta t \rightarrow 0} \frac{q \Delta x^2}{\Delta t} \left[\frac{u(x + \Delta x, t) - 2u(x, t) + u(x - \Delta x, t)}{\Delta x^2} \right].$$

Here is where a little analysis (or faith) is required. The left hand side of the above equation involves no Δx quantities. Therefore, that limiting operation is negligible and does not affect anything. However, we can notice a couple things on the right hand side of the above equation. The first thing you might notice from staring at it is that there are two limits happening in succession. One limit involves Δx and the other Δt . Second, we might notice that we could view that right hand side as the product of two quantities. Let's check that idea out below:

$$\frac{\partial u}{\partial t} = \lim_{\Delta x \rightarrow 0} \lim_{\Delta t \rightarrow 0} \left[\frac{q \Delta x^2}{\Delta t} \right] \left[\frac{u(x + \Delta x, t) - 2u(x, t) + u(x - \Delta x, t)}{\Delta x^2} \right].$$

Now we might notice that only one of those two factors on the right hand side involves any Δt 's. We can now implore a theorem from Calculus, involving taking the limit of the product of a few factors. Let's take a walk down memory road and recall that theorem together.

Theorem 2.2

The Product Law for Limits

If we know that $\lim_{x \rightarrow c} f(x) = A$ and $\lim_{x \rightarrow c} g(x) = B$, then it is guaranteed that

$$\lim_{x \rightarrow c} [f(x)g(x)] = AB.$$

Note:

- We can only apply this theorem if both limits exist!
- In practical terms, if both individual limits are known to exist, what this means is that

$$\lim_{x \rightarrow c} [f(x)g(x)] = \left[\lim_{x \rightarrow c} f(x) \right] \left[\lim_{x \rightarrow c} g(x) \right].$$

Now **ASSUMING** that the limits of both of these quantities exist, we can rewrite our in the following manner

$$\frac{\partial u}{\partial t} = \left[\lim_{\Delta x \rightarrow 0} \lim_{\Delta t \rightarrow 0} \frac{q\Delta x^2}{\Delta t} \right] \underbrace{\lim_{\Delta x \rightarrow 0} \frac{u(x + \Delta x, t) - 2u(x, t) + u(x - \Delta x, t)}{\Delta x^2}}_{\text{equal to}} \cdot \frac{\partial^2 u}{\partial x^2}$$

Noticing that second-order partial derivative sitting there on the right hand side, our equation then becomes

$$\frac{\partial u}{\partial t} = \left[\lim_{\Delta x \rightarrow 0} \lim_{\Delta t \rightarrow 0} \frac{q\Delta x^2}{\Delta t} \right] \left[\frac{\partial^2 u}{\partial x^2} \right].$$

Lastly, in this mathematical modeling business that remaining term in square brackets, which involves the two back-to-back limits, is given a special name. It is what is known as the **diffusivity** and is denoted by D , i.e.,

$$D = \lim_{\Delta x \rightarrow 0} \lim_{\Delta t \rightarrow 0} \frac{q\Delta x^2}{\Delta t}. \quad (3.19)$$

This formulation assumes that D is a constant value. That is, it is a number, not a function of space or time. If it were, we would have to modify our entire approach above, starting at the very beginning of how insects move from tree to tree. The situation involving a non-constant diffusivity would imply there was bias in how the insects moved from tree to tree, perhaps based on scent, whether indicative of a food source or predator.

So at long last, we finally arrive at what we set out to do → obtain a diffusion equation using our knowledge of a random walk! Our hard work in all of its glory is given below

$$\frac{\partial u}{\partial t} = D \frac{\partial^2 u}{\partial x^2}. \quad (3.20)$$

Eq. (3.20) is formally known as the **diffusion equation**, aka the *heat equation*. It is a second-order, linear **partial differential equation** (PDE). That is, it describes the *diffusion* or movement of a quantity u in both time AND space due to random fluctuations. There is quite a lot to unpack here!

For starters, up to this point in the book, we have only considered what are called **ordinary differential equations**. That is, an equation relating derivatives of a function with respect to only one independent variable, usually time, t . Eq. (3.20) involves both a derivative with respect to time, t , and space, x . Hence it cannot be an ordinary differential equation.

However, to bump this equation up to two spatial dimensions, say x and y , the process is straight-forward. We could repeat the same procedure we just did here and find that we would arrive at an analogous looking diffusion equation,

$$\frac{\partial u}{\partial t} = D \left[\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \right]. \quad (3.21)$$

Solutions of these diffusion equations are a bit more tricky to compute, whether in *1D* or *2D*, and whether we approach them analytically or numerically. To give you an idea of how the solutions behave, consider the snapshots of the solution over time as shown in Figure 3.11. Initially there is a source term only at the origin, e.g., at $t = 0$,

$$u(x, y, t = 0) = \begin{cases} 1 & \text{if } x = y = 0 \\ 0 & \text{elsewhere} \end{cases}.$$

You can notice how the source (heat, insects, oxygen concentration, etc.) spreads out from the origin evenly in all directions. This is indicative of having a constant diffusivity, D . That is, D is not a function of space or time. Therefore, we expect the spreading out of the concentration u to be unbiased, that is, evenly in all directions with no preferential direction.

3.2.3 Comparing Random Walkers to Diffusion

At this point you may be wondering just how a random walker model compares against a continuous diffusion model. On one hand, the random walker model has inherent stochasticity - if we have 1000 random walkers in 2D, there is virtually a 0% chance that after 100 steps that any of the random walkers would have traced out the same exact trajectory. Let alone if we were to repeat this experiment there is no chance that they would all trace out the same trajectory! If we're removing the inherent randomness and using a *deterministic* partial differential equation model like those we derived in Section 3.2.2, such as (3.21), how could they properly encapsulate such randomness?

Really, this boils down to asking how can we compare the dynamics between Figures 3.9 and 3.11. Well, let's take a step back (pun intended) and only consider random walks in 1D and the appropriate 1D diffusion equation, previously derived in (3.20). Take a moment and flip back to glance at that equation and those figures.

In a nutshell, these two modeling approaches are consistent with one another. Check out Figure 3.12! This figure provides a histogram of the random walker positions, given as

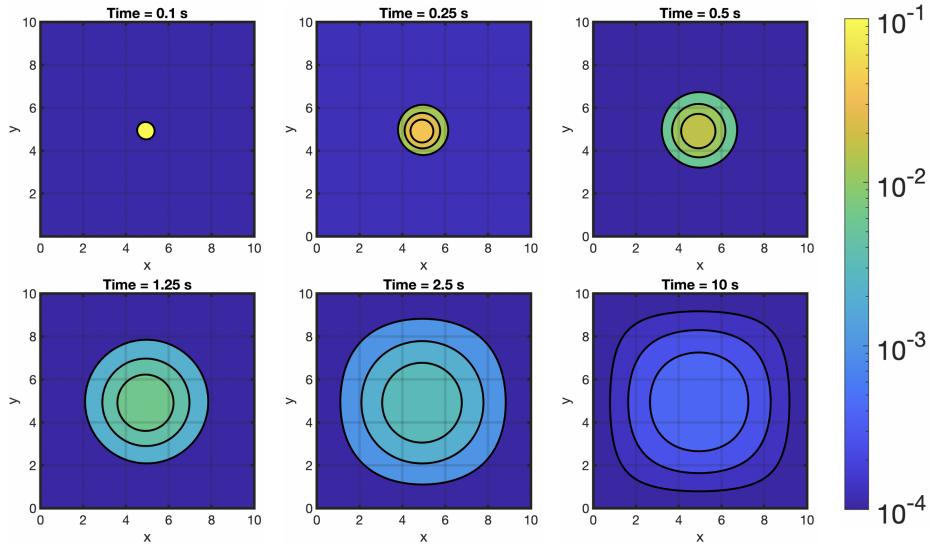


Figure 3.11: Numerically computed solutions to the diffusion equation, when initially there is only a source at $(x, y) = (0, 0)$ given by $u(x = 0, y = 0, t = 0) = 1$ and that everywhere else $u = 0$ at $t = 0$. The solution was computed using a simple finite difference scheme with a grid resolution of $dx = dy = 0.025$ and time-step $dt = 10^{-3}$. The diffusive coefficient used was $D = 0.25$.

the percentage of the overall number of random walkers, against the numerical solutions of the 1D diffusion PDE. To perform this comparison, the random walkers all start at position $x = 500$. There are $M = 2000$ random walkers. As they continue taking step after step, they begin slowly spreading out, err moving away from their initial position. We then directly compare those spreading dynamics via a 1D diffusion equation as given previous in (3.20), or repeated below here,

$$\frac{\partial u}{\partial t} = D \frac{\partial^2 u}{\partial x^2}.$$

Similarly, the initial value of for the PDE model is one such that at $t = 0$ all of the non-zero information is solely at $x = 500$, i.e.,

$$u(x, t = 0) = \begin{cases} M & \text{if } x = 500 \\ 0 & \text{elsewhere} \end{cases}.$$

While there is inherent randomness in the random walkers, you can see that the bulk general spreading dynamics of the random walkers are captured pretty well by our PDE model. To compare these two different modeling approaches, the parameters were chosen in a consistent fashion, e.g., a random walker takes a step every 0.01 s to match the time-step of the finite difference PDE solver, and the diffusive coefficient D was chosen appropriately to be $D = \frac{\Delta x^2}{\Delta t} p$, where $p = 1/3$, which is the probability that the walker moves left, right, or stays in the same place.

While we are able to numerically simulate a random walk process in 1D, 2D, or 3D, the process of solving the diffusion PDE such as (3.20) or (3.21) is above the scope of this course. Some buzzwords if you're interested in investigating this further are *finite difference* methods for solving *linear* PDEs. Moreover, solving these sort of equations

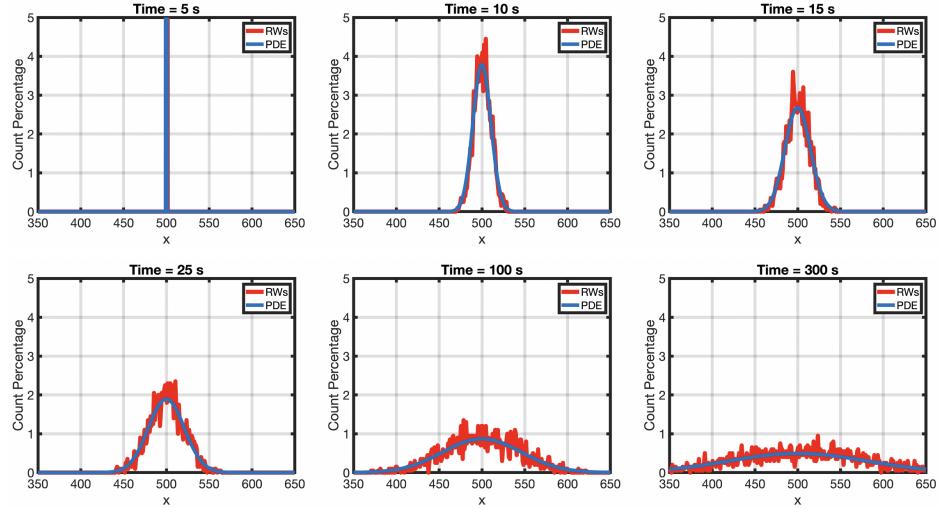


Figure 3.12: A comparison of $M = 2000$ random walkers starting at position $x = 500$ and slowly spreading outwards and the numerical solutions of a 1D diffusion equation with a similar initial condition, i.e., only a source at $x = 500$ given by $u(x = 0, t = 0) = M$ and everywhere else $u = 0$ at $t = 0$.

analytically are also above the scope of our course here. To learn how to solve PDEs such as the diffusion equation introduced here, please consider taking #MAT426: Partial Differential Equations.

In summary, although there is inherent randomness in a random walk, the average bulk behavior of the spreading out process by *many* random walkers continually taking steps can be captured through a deterministic partial differential equation. Or, on the other hand, you equivalently view this in that vein that you could use random walk processes to approximate solutions to certain partial differential equations.

3.3 Dynamics of a Neuron

Did you see that? Um... you read it. As you read that incredibly brilliant (obnoxious) pun, you may have cringed. Well, you have your *neurons* (the cells in your nervous system) to thank for that response! Neurons are the basic working units in your brain. They are specifically designed cells that are responsible for transmitting information along to other brain cells, muscles, or glands.

You might imagine that the mathematics necessary to model such an entity would be incredibly complex. Well, yes and no. Like any biological phenomenon, there is a large spectrum of models in this area. From rather simplistic models that try to capture the general bulk system dynamics to extremely detailed, intricate models that try to model the very nitty-gritty details of a system, we can ask ourselves one question, *what is our goal? what are we trying to do with this model?* This question will drive us to develop and use one sort of model over the other. For our considerations here, we will start off with a rather simplistic, humble modeling goal: *let's try to model the action potential behavior of a neuron... err, let's try to model how a neuron's voltage changes over time.*

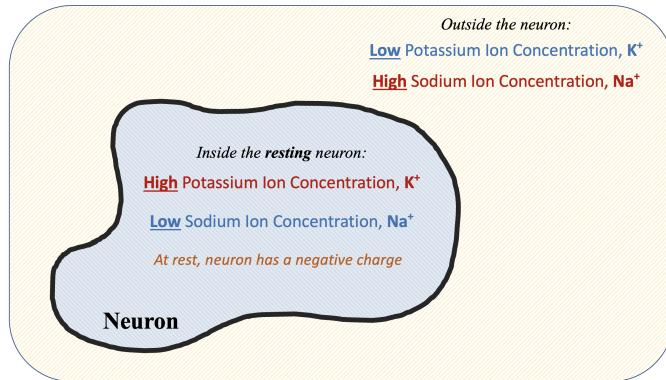


Figure 3.13: A quick overview of a neuron at rest. The overall charge (and potential) inside of a resting neuron is negative. There is a higher concentration of potassium ions, K^+ , than sodium ion concentrations, Na^+ inside the resting neuron. The opposite is true of the environment outside the resting neuron.

To begin this story of neurons, action potentials, and voltage, we will get to introduce some key biological cellular ideas (characters). Perhaps the best place to start is with a much simplified image of what a neuron might look like for us mathematical biologists at this moment in our journey. Figure 3.13 depicts a neuron *at rest* (not firing) in which inside the neuron there is an excess of potassium ions, K^+ , than sodium ions, N^+ . Moreover, the opposite case is true outside of the neuron - there is an abundance of Na^+ rather than K^+ . Although both sodium and potassium ions are positively charged, the a resting neuron maintains a negative resting **potential** (voltage). The potential is measured across the membrane of the neuron, thus sometimes it is referred to as the membrane potential. Thus, even though Na^+ and K^+ are positively charged, there is a far greater amount of negative charges inside to which the sodium and potassium ions cannot outweigh. Ions that are negatively charged are called *anions* (and ions to which are positively charged are called *cations*). Therefore, a resting neuron is negatively charged due to all of the other anions inside of it. In fact, neurons (and cells in general) invest a significant amount of energy to maintain this charge distribution!

For our modeling considerations, what we can recognize from this information so far is that *when a neuron is at rest, it maintains a negative voltage*. However, if this was the whole story, life would be boring - literally, our cells would remain static and nothing would ever change. Therefore, something must happen to these ions. In fact, cellular membranes are full of little paths, called *ion channels*, which provide a road for ions to dynamically move in (or out) of cells! Neurons are no different. Just like there are many different kinds of road and structures for us to use to get place to place (streets, highways/expressways, interstates, etc.), there are many different kinds of ion channels. The ion channels we're going to consider here for our neuron are similar to that of a pay bridge, in that we are only allowed to cross certain bridges if we pay them upfront to use them. It turns out that some ion channels only allow ions to pass when a voltage signal is high enough! So, like us trying to cross a pay bridge with money we're scrambling to find in a car, ions will not be allowed to flow unless voltage is above a certain threshold.

Ion channels in which only allow the flow of ions once voltage levels pass a certain

threshold value are referred to as *voltage-gated ion channels*. Also, there are certain voltage-gated ion channels for specific ions, too. For example, potassium ions trying to flow out of the neuron could only pass through voltage-gated potassium channels, and not those specific for sodium ions. Now, remember from our discussion of Random Walks and Diffusion in Section 3.2, things (including ions) will flow (diffuse) from higher concentrations to lower concentrations. From Figure 3.13, we can gather that this means that potassium ions will want to move from inside the neuron to outside, and in contrast, sodium ions would want to move from outside the neuron to inside the neuron.

Let's quickly paint of the situation. If our neuron is at rest, it has a negative potential (it is negatively charged). If sodium channels opened, more Na^+ ions would flow into the neuron. Since the sodium ions are positively charged, the voltage would increase inside the neuron, as the Na^+ ions are causing the neuron to become increasingly more positively charged. However, this cannot happen indefinitely. Remember, the cell wants to maintain its negative resting charge distribution. In response to this increasing voltage, the neuron does two things. First, it acts to block voltage-gated *sodium* channels, thus ceasing the flow of Na^+ into the neuron. Second, at this point the voltage is high enough to open the voltage-gated *potassium* channel. Remember that higher concentration of K^+ ions inside the cell? Well, opening this voltage-gated *potassium* channel, causes those positively charged K^+ ions to flow out of neuron, thereby decreasing the positive charge inside the neuron, which corresponds to a drop in the voltage potential.

Both of these neuron mechanisms to decrease the voltage potential are considered slow mechanisms. Imagine these *blocking* mechanisms as the ground when you roll a ball along it. The ground eventually slows down the ball but doesn't do so immediately. Similarly, both of these blocking mechanisms move to slow down and stop the increasing positive charge of the neuron, but it cannot do it instantaneously. Once the voltage falls below a certain threshold, these mechanisms will also turn off gradually. These events will bring the neuron back to its original, preferred resting state. A slightly amended Figure 3.13 shows these ideas in Figure 3.14.

So, what? Well these dynamics governing how the voltage starts off low (*neuron at rest*), spikes and increases (Na^+ flow into cell) and decreases (K^+ flow out of cell, Na^+ channel gets blocked), result in electrical excitation for the cell. The peak in these activity is what is referred to as an *action potential*. These excitation signals (electrical impulses) can then be passed to other neurons and transmit all of the starting information our bodies need to perform any tasks! Action potentials are a big part of the language of our brains and nervous system.

The lowdown of action potential dynamics can be summarized in the following steps. These dynamics are show in Figure 3.15, which gives an idea of what the voltage potential is during an entire cycle.

1. The neuron starts off in a resting state (negative potential, more K^+ ions inside than outside, fewer Na^+ ions outside than inside)
2. Cell membranes have lipids (fats) that block ions from entering and leaving on their own
3. Fortunately, cell membranes have protein, some of which are *ion channels* that mitigate the flow of ions from inside to outside the cell and vice versa.
4. When an ion "attaches" itself to an ion channel, the ion channels morphs in shape and carries it to the other side of the membrane.
5. Ions want to flow from regions of higher concentration to lower concentration.

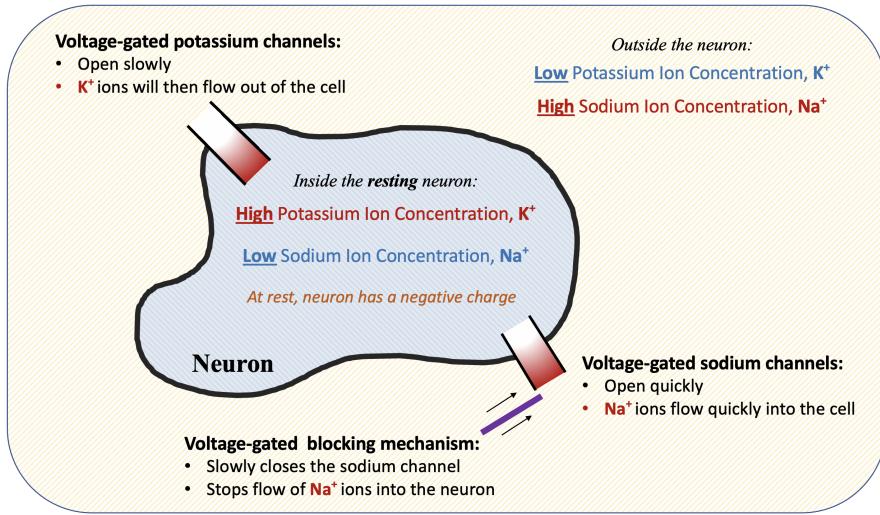


Figure 3.14: An amended version of Figure 3.13 to include the voltage-gated ion channels as well as blocking mechanism to illustrate the voltage dynamics that can occur.

6. Sodium ion channels allow Na^+ ions to flow into the neuron. The voltage increases, becoming more positive, and this results in more voltage-gated Na^+ channels to open. Thus, more and more positively charged sodium ions continue flowing into the neuron (**positive feedback**)! This happens *quickly*!
7. The neuron *slowly* combats this increase in its voltage (potential) via a **negative feedback** process. This happens by two different mechanisms:
 - Initiating a blocking mechanism that slowly halts the entry of Na^+ ions from outside to inside the cell.
 - Voltage-gated potassium channels open allowing K^+ ions to flow from inside the neuron to outside, resulting in a drop in the voltage potential. This slowly causes a change in the charge of the neuron, from more positively to more negatively charged.
8. Once the voltage V is less than or equal to the neuron's preferred resting potential V_{rest} , these two mechanisms turn off and the cycle gets performed again.

Okay, so this process involves many things - a variety of ions, different ion channels, a voltage potential, and multiple neuron blocking mechanisms. Where to begin a mathematical model of these action potential dynamics? Welp, let's start with a more simplified model. If we step away and abstract the general concepts of the above, there appear to be two major components:

- A quick process involving the flow of positive ions into the neuron in order to increase the voltage potential and
- Slower mechanisms that try to suppress that flow of positive ions to drop the voltage potential.

Hence there appears to be a battle between these two mechanisms that try to increase and decrease the *voltage potential* (internal charge of the neuron), respectively. Let's try to model these two mechanisms to begin! In fact, we will try to model the overall neuron

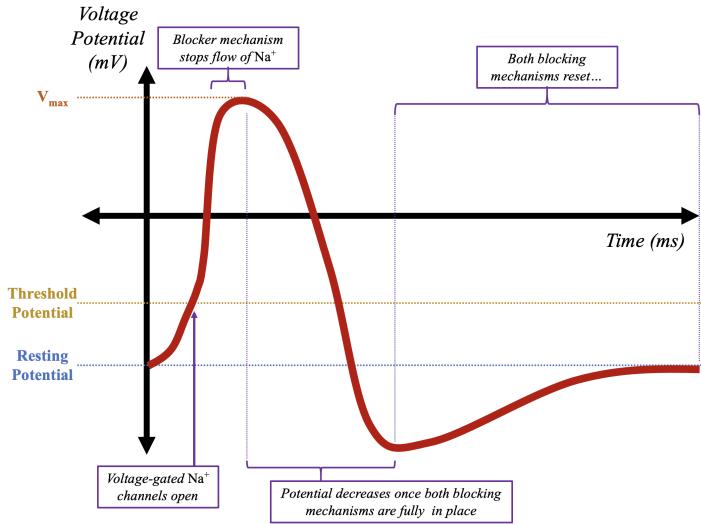


Figure 3.15: A full action potential cycle. For reference, these cellular voltage potentials are on the order of milli-volts (mV) and a full takes place in around ~ 3 milliseconds (ms).

voltage potential and a blocking mechanism that tries to decrease it.

To simplify things more, we will scale our voltage potential from milli-volt units in $V(t)$ to dimensionless units in terms of \tilde{v} . The purpose of this move will be so that:

- $\tilde{v} = 1$ corresponds to when $V = V_{max}$, and
- $\tilde{v} = 0$ corresponds to when $V = V_{rest}$.

To do this, we can find a linear relationship between \tilde{v} and V , as we have two bits of information we want to enforce (those two values of \tilde{v} listed above corresponding to V_{max} and V_{rest}). Hence we have the following:

$$\tilde{v}(V) = mV + b,$$

where m and b are our unknown scaling variables. Hence from those two specific values of \tilde{v} above, we get the following two equations:

$$\tilde{v} = 1 : \Rightarrow \tilde{v}(V_{max}) = 1 = mV_{max} + b$$

$$\tilde{v} = 0 : \Rightarrow \tilde{v}(V_{rest}) = 0 = mV_{rest} + b.$$

Solving this system of equations leads us to:

$$m = \frac{-1}{V_{max} + V_{rest}},$$

$$b = \frac{V_{rest}}{V_{rest} + V_{rest}}.$$

Hence we know exactly how to scale our original voltage values V , which are measured in milli-volts, to get our dimensionless voltage potential, \tilde{v} . Further, we can easily define $\tilde{v} = a$ to be the *threshold potential*, which says when $\tilde{v}(t) > a$, the neuron fires and more Na^+ ions flow into the neuron causing the voltage to further increase. Hence we have the following three important values of \tilde{v} :

$$\begin{cases} \tilde{v} = 0 & \text{resting potential} \\ \tilde{v} = a & \text{threshold potential} \\ \tilde{v} = 1 & \text{max potential} \end{cases}. \quad (3.22)$$

From these ideas, let's start building our mathematical model for a voltage potential and a blocking mechanism. We will build each equation separately.

Voltage Equation :

To begin, let's consider our information at hand. We expect that the cell wants to be at a resting state involving $\tilde{v} = 0$. Therefore, we need to ensure that $\tilde{v} = 0$ is an equilibrium point, so that the system pushes itself toward such a point. Hence we will build our model to include such an equilibrium point value, therefore our voltage model so far looks like:

$$\frac{d\tilde{v}}{dt} = \tilde{v}.$$

Next, we want to include an idea such that if $\tilde{v} < a$, that the neuron does not fire, i.e., the voltage-gated sodium channels do not open up! Therefore, the voltage should decrease back towards $\tilde{v} = 0$. To do this, we could include an equilibrium point at $\tilde{v} = a$; however, we wish to make this equilibrium unstable so that if the voltage is below a it will move away from a and towards 0. Therefore, including this idea in our model gives us:

$$\frac{d\tilde{v}}{dt} = \tilde{v}(\tilde{v} - a).$$

Finally, we want to ensure that if $\tilde{v} > a$ that the voltage would continue increasing until $\tilde{v} = 1$. Note that we always know that $a < 1$. Similarly, if $\tilde{v} > 1$, we want the voltage to decrease back towards $\tilde{v} = 1$. Hence we want the voltage potential to be attracted to a value of $\tilde{v} = 1$ if either $\tilde{v} > a$ or $\tilde{v} > 1$. Thus to make $\tilde{v} = 1$ an equilibrium point of our model, we can attach a term such as

$$\frac{d\tilde{v}}{dt} = \tilde{v}(\tilde{v} - a)(1 - \tilde{v}). \quad (3.23)$$

Note that the $(1 - \tilde{v})$ will give the appropriate stability properties as described above! We can visualize these stability properties as in the 1D phase portrait given in Figure 3.16.

We can plot these ideas by graphing the right hand side of our modeling equation for the voltage potential vs. \tilde{v} , (3.23), i.e.,

$$\frac{d\tilde{v}}{dt} = \tilde{v}(\tilde{v} - a)(1 - \tilde{v}) = f(\tilde{v}).$$

Such a plot is shown in Figure 3.17. Note that the right hand side is a cubic polynomial. Hence, we expect there to be similar behavior



Figure 3.16: 1D phase portrait corresponding to our modeling assumptions for what happens when the dimensionless voltage, \tilde{v} , falls within specific regions of the overall line.

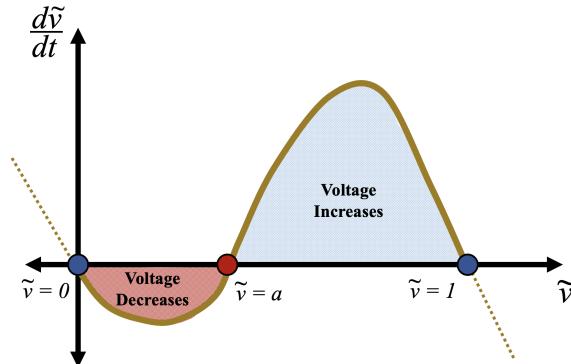


Figure 3.17: A plot of the right hand side of the differential equation governing the voltage dynamics from (3.23).

to a model, such as the *Allee effect* (or Depensation). These ideas were introduced in Example 2.4.1. This model is actually a natural extension of the Logistic model, which in this case would have only granted us an equilibrium at $\tilde{v} = 0$ and $\tilde{v} = 1$. To add in that unstable equilibrium at $\tilde{v} = a$, we bumped up the order of the polynomial from a quadratic to cubic to give our desired dynamics.

At this junction our model of the voltage potential does not include any influence from the blocking mechanism, nor exactly how an initial input signal might cause the voltage potential to cross the threshold barrier of $\tilde{v} = a$.

Blocking Equation :

Let's let w represent the strength magnitude in all of the blocking mechanisms. There is a lot swept under the rug here; however, we will see that this variable w will be sufficient for us to start. To begin, we recognize that we want this blocking mechanism to turn on as the neuron tries to fire. Therefore, we could try to resemble such dynamic in the following way:

$$\frac{dw}{dt} = \epsilon \tilde{v}.$$

Hence as long as $\epsilon > 0$, the blocking strength increases when $\tilde{v} > 0$. However, this will not produce the behavior we want where the blocking mechanism turns off once the voltage potential decreases far enough. We can do this by creating an equilibrium that occurs once the voltage potential is equal to a scaled value of the blocking

mechanism, i.e.,

$$\frac{dw}{dt} = \epsilon(\tilde{v} - \gamma w). \quad (3.24)$$

The above equation (3.24) basically reflects a battle between the potential voltage increasing the value of the blocking mechanism, while a large value of the strength of the blocking mechanism will actually cause its strength to decrease. We believe this to be the case because once the blocking mechanism has grown strong enough, it should correspond to a small voltage potential. Thus, the blocking mechanism will have completed its task! Furthermore, note that when $\frac{dw}{dt} = 0$, implies that $\bar{w} = \frac{\tilde{v}}{\gamma}$. Therefore, a larger value of γ creates a smaller equilibrium value of \bar{w} . Moreover, the parameter ϵ will not affect any equilibrium values.

Put it all together :

Lastly, to incorporate the blocking mechanism into (3.23), we can simply include a term such as $-w$ to exhibit the exponential decay effects of the blocking mechanisms, i.e.,

$$\frac{d\tilde{v}}{dt} = \tilde{v}(\tilde{v} - a)(1 - \tilde{v}) - w. \quad (3.25)$$

We might also wish to include a term describing an applied current (electrical impulse) to our neuron. We could mathematically call this applied current, I_{app} . This could be a thought of as a flux of ions at could be modeling whether as a time dependent ($I_{app} = I_{app}(t)$) or time independent term ($I_{app} \in \mathbb{R}^+$). Since this would be a rate in which positively ions enter the cells, it would be proportional to the derivative of the potential itself. Hence we could include this applied current term in the following manner:

$$\frac{d\tilde{v}}{dt} = \tilde{v}(\tilde{v} - a)(1 - \tilde{v}) - w + I_{app}. \quad (3.26)$$

On more thing... we mentioned that action potentials carry information via the nervous system throughout our entire body. So far our models are including any spatial dependence! What did we do so far with spatial dependence? Oh, that's right! Our discussion of the Diffusion equation from Sections 3.2.2 and 3.2.3. Recall the Diffusion equation that describes the spread of some quantity U in one spatial dimension is given by the following partial differential equation:

$$\underbrace{\frac{\partial U}{\partial t}}_{\substack{\text{change in } U \\ \text{with respect} \\ \text{to time, } t}} = \underbrace{D \frac{\partial^2 U}{\partial x^2}}_{\substack{\text{spatial dependence} \\ \text{about how a quantity} \\ \text{is spreading in } x}}.$$

The simplest thing we could do to include spatial dependence is attach a term like the right hand side of the above equation, i.e., $D \frac{\partial^2 U}{\partial x^2}$, onto our voltage potential equation. This will turn our voltage potential equation into a *reaction-diffusion* equation. Therefore, an equation giving the voltage potential in our neuron that includes

the blocking mechanism, applied current, and spatial dependence could be the following:

$$\frac{\partial \tilde{v}}{\partial t} = \tilde{v}(\tilde{v} - a)(1 - \tilde{v}) - w + I_{app} + \frac{\partial^2 \tilde{v}}{\partial x^2}. \quad (3.27)$$

All of these possible voltage potential equations (3.25, (3.26), and (3.27) along with the blocking mechanism equation (3.24) form variations on what are called the *FitzHugh-Nagumo Equations*.

The FitzHugh-Nagumo equations come in many forms. To quickly summarize, we've derived the following variations shown in Definition (3.1).

Definition 3.1

Below are three different variations on what are called the *FitzHugh-Nagumo Equations*.

Here \tilde{v} and w are a neuron's voltage potential, as measured across its cellular membrane, and its blocking mechanisms, respectively. The blocking mechanism try to drive the voltage potential towards its preferred resting voltage potential, i.e., $\tilde{v} = 0$ in dimensionless units here or $V = V_{rest}$ in the original milli-volt units.

1. **Original:** the original equations do not include an applied current nor spatial dependence:

$$\frac{d\tilde{v}}{dt} = \tilde{v}(\tilde{v} - a)(1 - \tilde{v}) - w \quad (3.28)$$

$$\frac{dw}{dt} = \epsilon(\tilde{v} - \gamma w). \quad (3.29)$$

2. **With Applied Current:** the equations in which there is an applied current, I_{app} , that increases the voltage potential, are given as

$$\frac{d\tilde{v}}{dt} = \tilde{v}(\tilde{v} - a)(1 - \tilde{v}) - w + I_{app} \quad (3.30)$$

$$\frac{dw}{dt} = \epsilon(\tilde{v} - \gamma w). \quad (3.31)$$

3. **With Applied Current:** the equations in which there is an applied current, I_{app} , that increases the voltage potential, are given as

$$\frac{\partial \tilde{v}}{\partial t} = \tilde{v}(\tilde{v} - a)(1 - \tilde{v}) - w + I_{app} + D \frac{\partial^2 \tilde{v}}{\partial t^2} \quad (3.32)$$

$$\frac{dw}{dt} = \epsilon(\tilde{v} - \gamma w). \quad (3.33)$$

Let's explore each of these different models to see what type of solutions they offer!

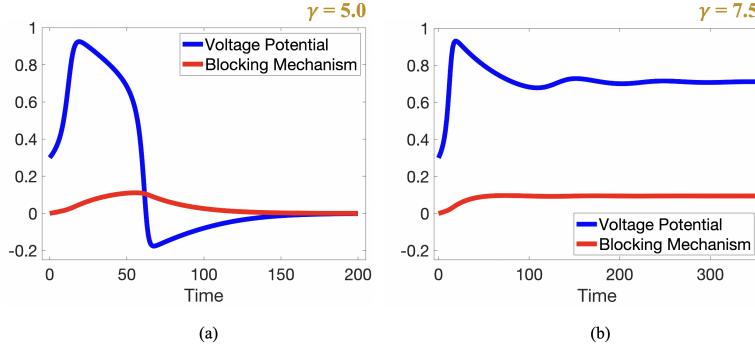


Figure 3.18: Some solutions to the original FitzHugh-Nagumo equations given in (3.34). These solutions were created with $a = 0.25$, $\epsilon = 0.005$, $\tilde{v}(0) = 0.3$, and $w(0) = 0$, for (a) $\gamma = 5.0$ and (b) $\gamma = 7.5$.

3.3.1 Analyzing the FitzHugh-Nagumo Eqns (no applied current, nor diffusion)

Let's start off by analyzing the original FitzHugh-Nagumo equations, which do not include an applied current nor spatial dependence (diffusion).

$$\begin{cases} \frac{d\tilde{v}}{dt} = \tilde{v}(\tilde{v} - a)(1 - \tilde{v}) - w \\ \frac{dw}{dt} = \epsilon(\tilde{v} - \gamma w) \end{cases} \quad (3.34)$$

Some solutions of these equations are shown in Figure 3.18. Notice that these solutions were found for the specific parameters values $a = 0.25$, $\epsilon = 0.005$, $\tilde{v}(0) = 0.3$, and $w(0) = 0$, and two different γ values - 5.0 and 7.5 in Figures 3.18a and b, respectively. In Figure 3.18a, there appears to be an emerging equilibrium for $\tilde{v}, \bar{w} = (0, 0)$, while Figure 3.18b shows an emergent non-zero \tilde{v} -related equilibrium. Although, you might surmise that these equilibria appear to both be stable, we should perform our favorite - stability analysis, to confirm. Moreover, we will be able to see when non-zero equilibrium even exist! Hint: we will see that a relationship between parameters a and γ will govern whether a non-zero equilibrium is even a possibility!

Looking for equilibrium values to (3.34), we will start by setting $\frac{dw}{dt} = 0$, which implies that at an equilibrium point,

$$\tilde{v} = \gamma w.$$

Substituting this into the other differential equation and setting it equal to zero we get

$$\begin{aligned} \frac{d\tilde{v}}{dt} &= \tilde{v}(\tilde{v} - a)(1 - \tilde{v}) - w \\ &= \tilde{v}(\tilde{v} - a)(1 - \tilde{v}) - \frac{1}{\gamma} \tilde{v} \\ &= \tilde{v} \left[(\tilde{v} - a)(1 - \tilde{v}) - \frac{1}{\gamma} \right] \end{aligned}$$

and hence

$$\tilde{v} \left[(\tilde{v} - a)(1 - \tilde{v}) - \frac{1}{\gamma} \right] = 0.$$

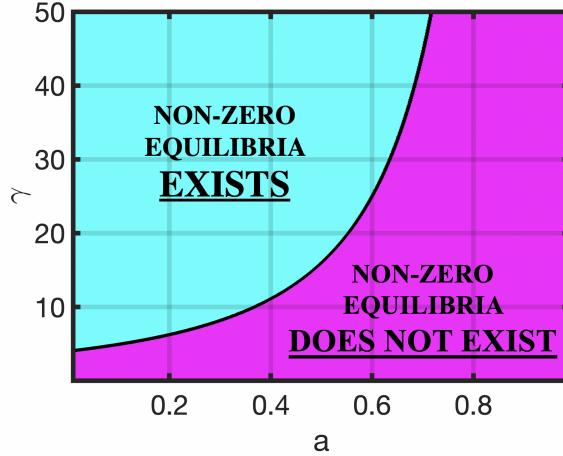


Figure 3.19: Regions for which the non-zero equilibrium *could* exist for varying a and γ values.

Therefore, we find that either $\tilde{v} = 0$ or that $-\tilde{v}^2 + (1+a)\tilde{v} - (a + 1/\gamma) = 0$. Using the quadratic formula we find that the possible equilibrium values are:

$$\tilde{v} \in \left\{ 0, \frac{1+a}{2} + \frac{\sqrt{(a-1)^2 - 4/\gamma}}{2}, \frac{1+a}{2} - \frac{\sqrt{(a-1)^2 - 4/\gamma}}{2} \right\}.$$

From our earlier realization that at an equilibrium point that $\bar{w} = \frac{1}{\gamma}\tilde{v}$, we can deduce what the accompanying \bar{w} value is as well.

Before we go onto our next exciting venture (stability analysis), take a moment and look at those possible equilibrium points. Now...since we are working with real-valued variables, we cannot obtain complex (imaginary) equilibrium values! So, before pressing on, we need to see what conditions lead to real-valued equilibrium. To do this, we get to inspect what is happening under that square root, i.e., we need

$$(a-1)^2 - 4/\gamma > 0 \Rightarrow \gamma(a-1)^2 > 4.$$

To connect this with our solutions from earlier, this could be where that plausible non-zero equilibrium appeared from in Figure 3.18b! However, as we just saw through that inequality above, this equilibrium only potential arises for certain parameter combinations. Figure 3.19 illustrates plausible existence regions for these equilibria for different combinations of parameters a and γ . Now, just because these equilibria could exist for certain parameter combinations doesn't necessarily mean that they are stable. Welp, on this note let's see when these equilibria are stable.

Computing the Jacobian matrix for (3.34), we get

$$J = \begin{bmatrix} -3\tilde{v}^2 + 2(1+a)\tilde{v} - a & -1 \\ \epsilon & -\epsilon\gamma \end{bmatrix}.$$

Let's break this into cases: the zero equilibrium and the non-zero equilibria.

- **(0,0)-Equilibria:**

Evaluating J at the $(0,0)$ equilibrium gives us:

$$J(0,0) = \begin{bmatrix} -a & -1 \\ \epsilon & -\epsilon\gamma \end{bmatrix}.$$

The eigenvalues of the above Jacobian are found by setting the determinant of $|\lambda I - J(0,0)| = 0$. Doing this gives us the following characteristic polynomial:

$$\lambda^2 + (a + \epsilon\gamma)\lambda + \epsilon(a\gamma + 1) = 0.$$

Using our pal, the quadratic formula, gives us the eigenvalues, i.e.,

$$\lambda = \frac{-(a + \epsilon\gamma) \pm \sqrt{(a + \epsilon\gamma)^2 - 4\epsilon(a\gamma + 1)}}{2}.$$

From the form of the above eigenvalues, we note that if the eigenvalues are real-valued, both eigenvalues must be negative since

$$\left| \frac{a + \epsilon\gamma}{2} \right| > \left| \sqrt{(a + \epsilon\gamma)^2 - 4\epsilon(a\gamma + 1)} \right|.$$

Moreover, if the eigenvalues are complex, since we assume that all parameters are positive, implies that $\text{Re}(\lambda) = \frac{-(a + \epsilon\gamma)}{2} < 0$. Therefore, whether the eigenvalues are real-valued or complex-valued, both cases lead to **stability** of this equilibrium.

- **Non-Zero Equilibria:**

We could, of course, evaluate the Jacobian matrix, $J(\bar{v}, \bar{w})$ at the non-zero equilibrium values, either

$$\bar{v} = \frac{1+a}{2} + \frac{\sqrt{(a-1)^2 - 4/\gamma}}{2} \quad \text{or} \quad \bar{v} = \frac{1+a}{2} - \frac{\sqrt{(a-1)^2 - 4/\gamma}}{2},$$

with the associated value of $\bar{w} = \frac{\bar{v}}{\gamma}$, and then compute the resulting eigenvalues. However, while possible, this looks incredibly tedious and it is unclear whether much information may be gained through such analysis - *the resulting parameter restrictions may be rather difficult to analyze.*

To analyze the stability properties of these equilibrium values, we will get to use computational methods. That is, we will use MATLAB to do the heavy lifting for us! We will let $\epsilon = 0.005$ and vary both a and γ and check (1) whether the non-zero equilibrium even could exist and (2) if it exists ($\gamma(a-1)^2 > 4$) what its stability properties are. More specifically, we will vary a between $0 < a < 1$ and $0 < \gamma < 50$ and see what we get! In essence, we are doing the following:

1. Pick a combination of a and γ values.
2. Compute potential non-zero equilibrium value.
 - If it exists:
 - (a) Calculate Jacobian for that specific equilibrium
 - (b) Compute eigenvalues of specific Jacobian matrix
 - (c) Store combination if eigenvalues give stability of the non-zero equilibrium. If they don't store that information as well.
 - If they don't exist, mark region as non-viable for stability analysis.

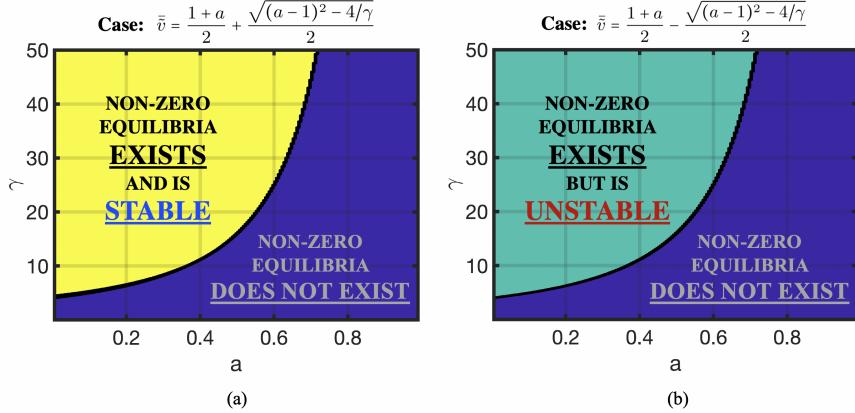


Figure 3.20: Stability regions for different parameter combinations of a and γ when $\epsilon = 0.005$ for (a) $\bar{v} = \frac{1+a}{2} + \frac{\sqrt{(a-1)^2 - 4/\gamma}}{2}$ and (b) $\bar{v} = \frac{1+a}{2} - \frac{\sqrt{(a-1)^2 - 4/\gamma}}{2}$.

This can all be done with a few `for-loops`, `if-statements`, and `2D arrays` in MATLAB. We can visualize the information through a colormap as well.

Figure 3.20 illustrates the *stability regions* for both non-zero equilibria to (3.34). Figure 3.20a shows when the possible equilibrium point associated with $\bar{v} = \frac{1+a}{2} + \frac{\sqrt{(a-1)^2 - 4/\gamma}}{2}$ is stable. Interestingly, when this equilibrium exists, it looks as though it is always stable from our calculations (compare this to Figure 3.19). On the other hand, the equilibrium associated with $\bar{v} = \frac{1+a}{2} - \frac{\sqrt{(a-1)^2 - 4/\gamma}}{2}$ appears to always be unstable whenever it can exist.

To quickly recap, we have seen that when $0 < a < 1$, $\gamma > 0$, and $\epsilon > 0$, the FitzHugh-Nagumo equations in (3.34) can have 3 possible equilibrium values. However, only two may exist, depending on whether $\gamma(a-l)^2 > 4$ is satisfied by choice of parameters a and γ . If all three exist, we see that two are stable and one is unstable. Moreover, we can plot such equilibrium values on a number line for \bar{v} , with their associated stability properties, see Figure 3.21.

Notice that when all three equilibrium exist, the story of equilibrium and stability is reminiscent to that of the *Allee effect*. This makes sense after-all. Since the right hand side of $\frac{d\bar{v}}{dt}$ is a cubic polynomial we would expect three possible equilibrium points. However, such non-zero equilibrium only exist via restrictions on γ , as we saw in Figures 3.19 and 3.20. Moreover, if we plot the nullclines, we can see when all three equilibrium exist more clearly - all three exist when there are 3 intersections of the nullclines associated with each equation!

Recall from Section 2.5 that to plot the nullclines, we are looking for a relationship between the dependent variables \bar{v} and w to plot on the phase-plane \bar{v} vs. w , such that those relationships make the right hand side of the governing differential equations equal to zero. That is, we first find when the right hand sides equal zero of (3.34), i.e.,

$$\begin{aligned} \frac{d\bar{v}}{dt} = \bar{v}(\bar{v} - a)(1 - \bar{v}) - w &= 0 & w &= \bar{v}(\bar{v} - a)(1 - \bar{v}) \\ \frac{dw}{dt} = \epsilon(\bar{v} - \gamma w) &= 0 & \Rightarrow & \\ && w &= \frac{1}{\gamma}\bar{v} \end{aligned} \quad . \quad (3.35)$$

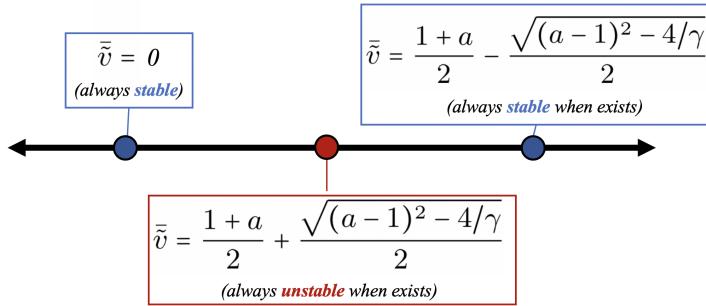


Figure 3.21: Summary of equilibrium and stability properties for the FitzHugh-Nagumo equations given in (3.34).

Now we can plot the relationships in (3.35) showing how \tilde{v} and w relate to one another one each nullcline curve. Plots of these nullclines are given in Figure 3.22. As expected, Figure 3.22a shows that when $\gamma = 5.0$ (and $a = 0.25$ and $\epsilon = 0.005$) that only one equilibrium emerges - the zero equilibrium. This is because the nullclines only intersect at one point! This equilibrium was illustrated in Figure 3.18a. On the other hand, as γ increased, more equilibrium points emerged - the nullcline curves intersected three times. This is shown in Figure 3.22b and corresponds to the case of 3.18b. Note that at this junction we did not prove why such solutions in Figure 3.18a and b go toward specific equilibrium with our nullcline/phase-plane analysis. Notice that large values of γ correspond to more rapid decreases in the blocking mechanism from (3.34). Therefore when γ is large, the blocking mechanism fails at resetting the system, as w continually decreases rapidly. Thus, the case involving a stable non-zero equilibrium, as hinted in Figures 3.22b and 3.18b, could be thought of as the neuron failing to reset after firing due to the blocking mechanism's inability to reduce the flow of positively charged ions into the neuron.

Finally, notice that our neuron governed by (3.34) only may fire once (zero equilibrium), based on the initial values of $\tilde{v}(0)$ and $w(0)$, or stay in a state of constant firing (non-zero equilibrium). There is not another stimulus to jostle the system into either firing again or possibly resetting! Let's next see what happens when we apply an additional current to our neuron.

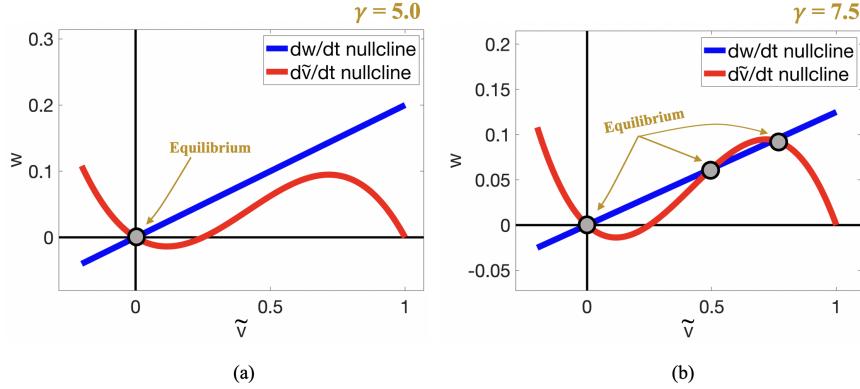


Figure 3.22: Showing the intersection of the nullclines for the FitzHugh-Nagumo system given in (3.34) for (a) $\gamma = 5.0$ and (b) $\gamma = 7.5$. As gamma varies, 2 additional equilibrium points emerge for a total of 3 possible equilibrium points.

3.3.2 Analyzing the FitzHugh-Nagumo Eqns (w/ applied current, no diffusion)

Let's now analyze the FitzHugh-Nagumo equations in which there is an applied current, I_{app} , that increases the voltage potential. The governing equations are given as

$$\begin{cases} \frac{d\tilde{v}}{dt} = \tilde{v}(\tilde{v} - a)(1 - \tilde{v}) - w + I_{app} \\ \frac{dw}{dt} = \epsilon(\tilde{v} - \gamma w). \end{cases} \quad (3.36)$$

We will assume a constant applied (input) current, i.e., $I_{app} \in \mathbb{R}^+$. This case begs the question, *can the neuron convert a constant applied current into a usable action potential output?*. People may also model this case with a square wave activation signal, rather than a constant input [1]. Furthermore, notice that the additional term I_{app} makes our differential equation system **non-homogeneous**. Also, this additional term increases the voltage, \tilde{v} , and hence the positive sign in front of I_{app} .

From a physics perspective, current, I , refers to the flow of charged particles (ions). It has units of charges (*coulombs*) per time, i.e., $[I] = \frac{\text{C}}{\text{s}}$. Since we cast our FitzHugh-Nagumo equations in scaled units (*remember how we derived \tilde{v} from v ?*), the units of I_{app} won't be C/s, but a similarly scaled version as well. Thus, the rate of change of the potential with respect to time should be proportional to the rate ions enter the cell; the latter is quantified by I_{app} in our model.

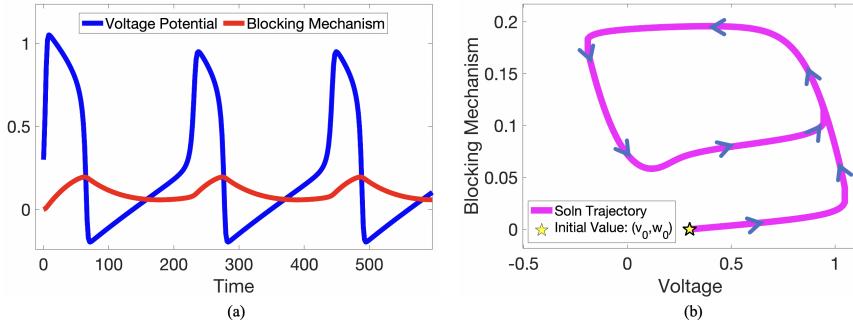


Figure 3.23: (a) A possible solution to the original FitzHugh-Nagumo equations with constant applied current. These solutions were created for $a = 0.25$, $\epsilon = 0.005$, $\gamma = 2.0$, and $I_{app} = 0.1$, with initial values $\tilde{v}(0) = 0.3$ and $w(0) = 0$. The constant applied current can give rise to periodic firing neurons! (b) The phase plane associated with the solutions from (a).

A possible solution to the FitzHugh-Nagumo equations with a constant applied current, as given in (3.36), is shown in Figure 3.23a. The solutions were solved with parameter values $a = 0.25$, $\epsilon = 0.005$, $\gamma = 2.0$, and $I_{app} = 0.075$ and initial values $\tilde{v}(0) = 0.3$ and $w(0) = 0$. This figure illustrates that the addition of a *constant* applied current can lead to the neuron periodically firing! Thus, a neuron can translate even a constant input current into usable output in the form of action potential bursts. Real neurons exhibit the same behavior (*periodic bursts*) when subjected to a constant applied current. Moreover, one could design pacemaker electronics based off these principles, i.e., a constant input current (analog input) gets translated into an output periodic burst (digital output) via these mechanisms. The corresponding phase plane is given in Figure 3.23b. As expected based off of Figure 3.23a, cyclical behavior emerges in the phase plane as the trajectories $(\tilde{v}(t), w(t))$ are trapped in a periodic orbit.

The nullclines of the system are given in Figure 3.24a. Since we only changed the $\frac{d\tilde{v}}{dt}$ equation, the w -Nullclines are not affected by the additional current, I_{app} . In the case involving parameters $a = 0.25$, $\epsilon = 0.005$, $\gamma = 2.0$, and $I_{app} = 0.0075$ and initial values $\tilde{v}(0) = 0.3$ and $w(0) = 0$, the nullclines only intersect once. This gives rise to one equilibrium. We could go through all of the equilibria and stability analysis to show that this single equilibrium is, in fact, unstable. Moreover, Figure 3.24b illustrates that this equilibrium is likely unstable, as the solution trajectory appears to get very close to the equilibrium, but moves away from it... it actually falls into a periodic orbit around it.

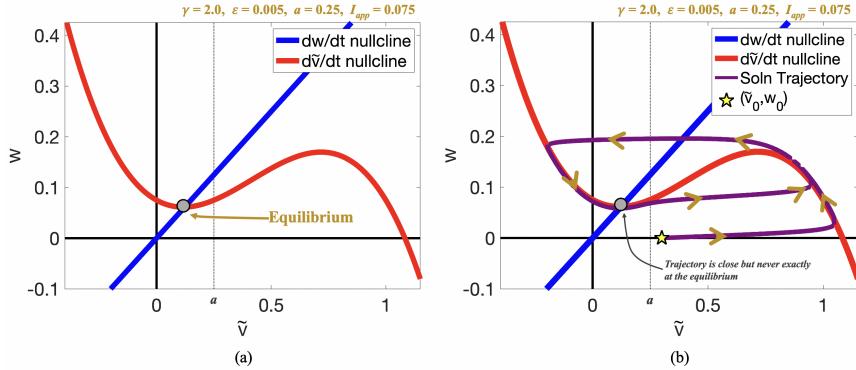


Figure 3.24: (a) The nullclines of the system given in (3.36) when $a = 0.25$, $\epsilon = 0.005$, $\gamma = 2.0$, and $I_{app} = 0.075$, with initial values $\tilde{v}(0) = 0.3$ and $w(0) = 0$. There is one equilibrium value present. (b) The nullclines and solution trajectory $(\tilde{v}(t), w(t))$ for the aforementioned parameters.

The take home message of the FitzHugh-Nagumo equations with a **constant** applied current is that *the neuron is able to take a constant input signal and translate it into a bursting periodic output signal... for some parameter combinations*. So far we've only looked a particular parameter combination that lead us to observing such a lovely periodic action potential bursting process. However, not all combinations of parameters will lead to such dynamics. Naturally, as we turn all the parameter knobs and analyze different parameter combinations we could see three possible equilibrium show up, different stability properties of those equilibrium, and/or other bifurcations in the neuron's system dynamics. For example, if we pick on the γ parameter once again, i.e., the parameter governing how quickly the blocking mechanism dissipates its strength, we can see various scenarios arise when all other parameters are kept constant.

Consider three cases: $\gamma = \{1, 2, 5\}$, for $a = 0.25$, $\epsilon = 0.005$, and $I_{app} = 0.075$, with initial values $\tilde{v}(0) = 0.3$ and $w(0) = 0$. These three different γ values result in the model displaying three very different kinds of behavior, as illustrated in Figure 3.25. When $\gamma = 1$, the neuron fires, but rather than reset back to $\tilde{v} = 0$, the neuron maintains a constant voltage around $\tilde{v} \sim 0.06$. That is, the neuron does not fire again. On the other hand, when $\gamma = 5.0$, the neuron fires, resulting in an action potential spike; however, the neuron never resets back. In this case ($\gamma = 5$), the blocking mechanism is not able to drive $\tilde{v} \rightarrow 0$. Thus, the neuron is basically stuck in a state of constant firing, as \tilde{v} maintains a constant value around $\tilde{v} \sim 0.8$. The case in the middle of those two extremes, when $\gamma = 2.0$, illustrates that these parameter combinations give rise to a neuron that produces a periodic output signal, as discussed previously in this section.

Lastly, as the phase plane suggests, we could analyze each of these cases using nullclines, equilibria, and local stability analysis. Compared to this exact type of analysis performed in Section 3.3.1, here analytical analysis will be a bit more difficult. This is because the equilibrium will have to be computed numerically, rather than analytically, due to the cubic polynomial that arises from the voltage equation in (3.36). That being said, it is still possible to perform such analysis, but we will forfeit it here to move onto perhaps more interesting questions you might be thinking about in regards to applying a current to a neuron.

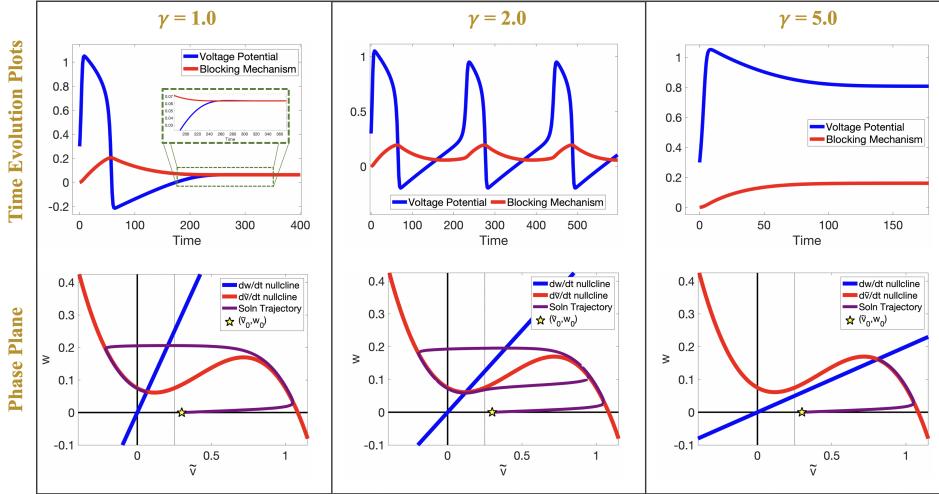


Figure 3.25: A comparison of the time evolution and phase planes when varying γ for $a = 0.25$, $\epsilon = 0.005$, and $I_{app} = 0.075$, with initial values $\tilde{v}(0) = 0.3$ and $w(0) = 0$.

You may be wondering what would happen if we have a time-dependent applied current (I would be)! Let's do a simple experiment in which we use a time-dependent applied current in the form of a simple on/off switch, i.e., a square wave activation signal. Such a signal is shown in Figure 3.26, where the period of the signal is 5 time units and the duration in which the current is on 2 of those 5 time units. This certainly opens up a big can of worms - there are so many different variations on the *duty cycle* [2] that we could consider, i.e., the overall fraction of the period in which the input current is actually turned on (applied to the neuron) compared to the overall period. Let's see what happens if we use the duty cycle shown in Figure 3.26 but with a different I_{app} strength magnitude.

Figure 3.27 gives possible solutions to (3.36) but for a time-varying applied current, with duty cycle given in Figure 3.26, for $a = 0.25$, $\epsilon = 0.005$, and $\gamma = 2.0$, with initial values $\tilde{v}(0) = 0.3$ and $w(0) = 0$ for three different I_{app} values: $I_{app} \in \{0.1, 0.2, 0.625, 0.75\}$. From a quick glance, we can recognize that the solutions look different depending on value of I_{app} . For the aforementioned parameter values, if I_{app} is low enough ($I_{app} = 0.1$), the neuron will fire once, but will not fire again. The magnitude of the applied current is too low to push the neuron over the $\tilde{v} > a$ limit and thus does not fire again.

However, for a slightly higher value of the applied current, $I_{app} = 0.2$, the neuron will repeatedly fire and we see similar periodic bursting action potential dynamics (see Figure 3.27). These dynamics appear similar to the behavior shown in Figure 3.23a for a constant current. However, due to the time-varying nature of the applied current, there are spurious oscillations throughout the entire voltage dynamics. Note that these are directly related to the period and duty cycle of the input signal (Figure 3.26). Moreover, we see that the blocking mechanism also displays periodic behavior as well.

As the I_{app} value increases, periodic action potential bursts continue; however, the behavior begins to look more wave-like, rather than taking the traditionally distinct action potential dynamics. When $I_{app} = 0.625$, this behavior is visible, see Figure 3.27 and compare cases $I_{app} = 0.2$ and 0.625 . Furthermore, as I_{app} continues increasing the periodic bursting behavior ceases and the dynamics are saturated by the strength of the applied

current. That is, the time-varying applied current is what is driving the voltage rather than neuron dictating the overall voltage dynamics. For high enough values of I_{app} , the voltage simply bounces up and down with a large amplitude, rather than display the distinct periodic bursting behavior, which is a signature of action potentials.

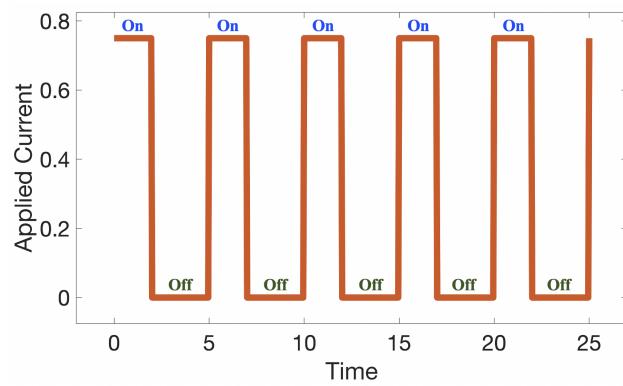


Figure 3.26: An example of an on/off square wave time dependent applied current signal, $I_{app}(t)$.

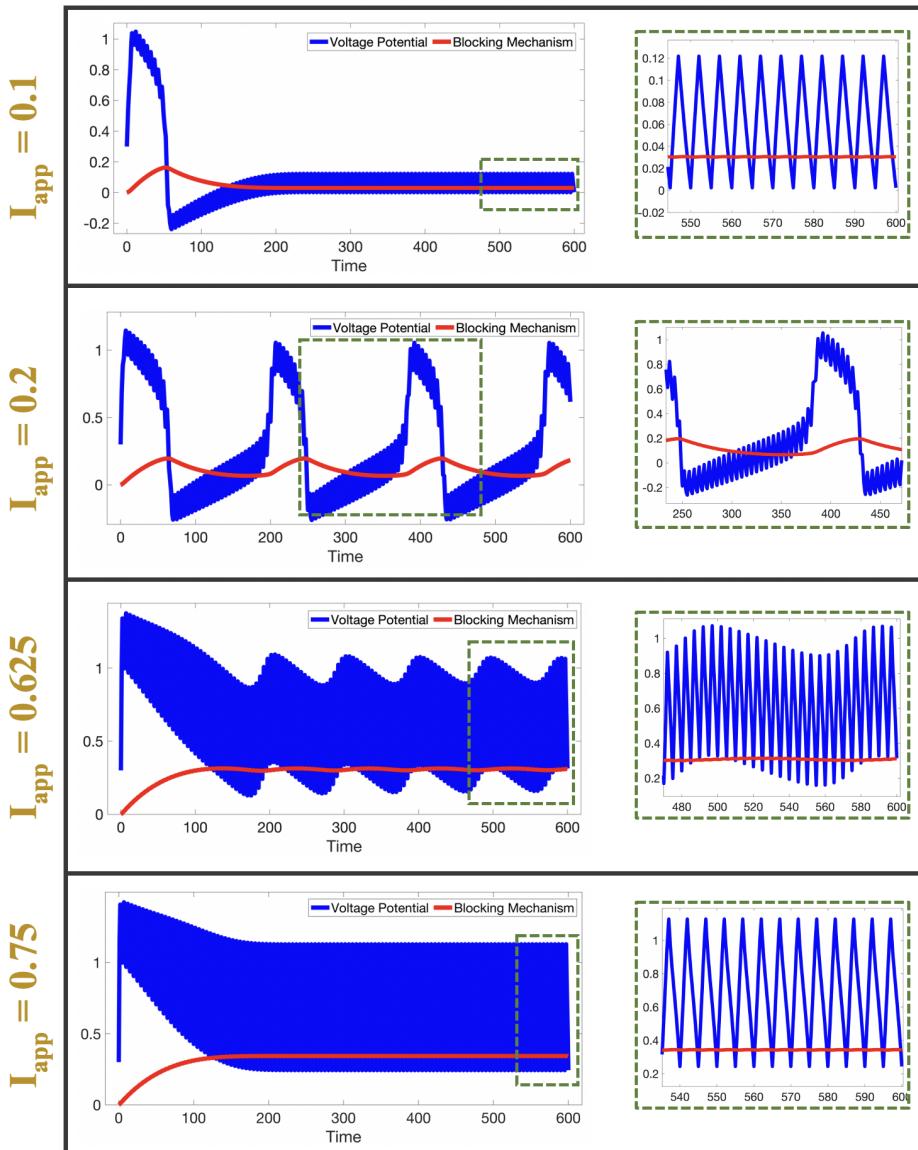


Figure 3.27: Possible solutions to (3.36) for a time-dependent applied current in which each pulse cycle lasts for a total of 5 time units and the duration of applied current in each overall period is 2 time units, for varying I_{app} current magnitudes.