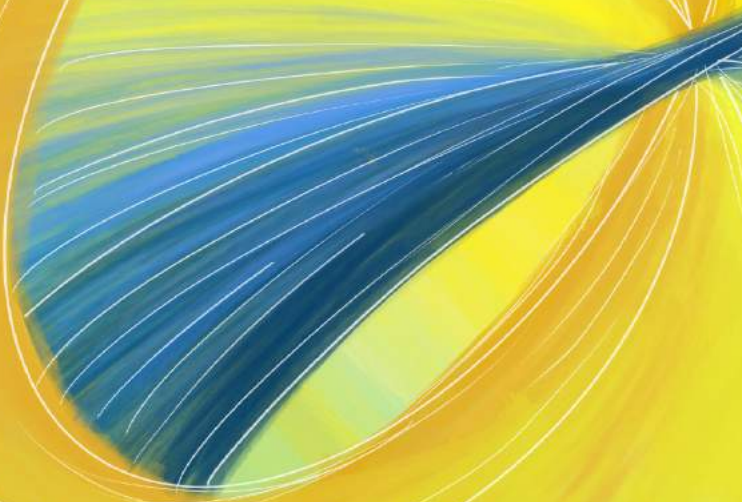
Differential Equation Models in Epidemiology



Marlon M. López-Flores
Dan Marchesin
Vítor Matos
Stephen Schecter

impa

33º Colóquio Brasileiro de Matemática

Differential Equation Models in Epidemiology

Differential Equation Models in Epidemiology
Primeira impressão, julho de 2021

Publicado no Brasil / Published in Brazil.

ISBN 978-65-89124-39-9
MSC (2020) Primary: 92D30, Secondary: 34A26, 34E15, 91A22

Coordenação Geral Carolina Araujo

Produção Books in Bytes **Capa** Izabella Freitas & Jack Salvador

Realização da Editora do IMPA

IMPA
Estrada Dona Castorina, 110
Jardim Botânico
22460-320 Rio de Janeiro RJ

www.impa.br
editora@impa.br

Contents

Acknowledgments **iii**

Preface **iv**

1 SIS Model **1**
1.1 The model . . . 1
1.2 How to do Western science . . . 3
1.3 Differential equations background: basics . . . 3
1.4 Phase line for the SIS system . . . 7
1.5 The constant β and the derivation of the SIS model . . . 9
1.6 The constant γ . . . 11
1.7 The basic reproduction number R_0 . . . 12
1.8 The effective reproduction number R_e . . . 13
1.9 Discussion of the SIS model . . . 14
1.10 Problems . . . 15

2 SIR model **18**
2.1 The model . . . 18
2.2 Differential equations background: vector fields and nullclines . . . 19
2.3 Differential equations background: functions and differential equations . . . 21
2.4 Orbits and phase portrait of the SIR system . . . 23
2.5 Interpretation of the orbits . . . 25

2.6 Changing population size 28
2.7 Discussion of the SIR model 32
2.8 Problems 33

3 SIR model with loss of immunity **35**
3.1 The model 35
3.2 Phase portraits 36
3.3 Differential equations background: linear differential equations . . 38
3.4 Differential equations background: asymptotic stability and linearization 43
3.5 Equilibria of the SIR model with loss of immunity 45
3.6 Differential equations background: planar theory 47
3.7 Global stability 50
3.8 Discussion of the SIR model with loss of immunity 51
3.9 Problems 52

4 A Covid-19 model and the next generation matrix **55**
4.1 The model 55
4.2 Equilibria of the Covid-19 model 58
4.3 Differential equations background: normally hyperbolic manifolds of equilibria 59
4.4 Digression: estimating R_0 at the start of an epidemic 60
4.5 Eigenvalues of equilibria of the Covid-19 model 62
4.6 The next generation matrix 63
4.7 Differential equations background: matrix exponential 68
4.8 Explanation of the entries of V^{-1} 69
4.9 Disease variants 71
4.10 Discussion of the Covid-19 model 73
4.11 Problems 74

5 Spontaneous human behavioral change in epidemiological models **79**
5.1 A model for human behavior in an epidemic 80
5.2 Slow time and fast time 84
5.3 The fast limit system 85
5.4 The slow limit system 87
5.5 Combining solutions of the fast and slow limit systems 87
5.6 Entry-exit integral 88
5.7 Singular solutions 90

5.8 Discussion of human behavioral change 92
5.9 Problems 93

A Solutions to problems **97**
A.1 Solution for Chapter 1 97
A.2 Solution for Chapter 2 100
A.3 Solution for Chapter 3 102
A.4 Solution for Chapter 4 105
A.5 Solution for Chapter 5 110

Bibliography **114**

Index **118**

Acknowledgments

This book originated as the text for a short course at the thirty-third Colóquio Brasieiro de Matemática held in August 2021. We thank the organizers for enthusiastically accepting our proposal to offer the course.

Professor Paulo Ney de Souza's direction and assistance were crucial in creating a quality editorial product.

The instructors, without whom this effort would have not been successful, were Prof. Pablo Castañeda (ITAM, México City), Ciro Campolina and Arthur Bizzi (IMPA, Rio de Janeiro), and Lucas Furtado (CUNY, New York City).

Steve, Dan and Vitor thank their spouses Nancy Schecter, Miyoko Ohtani and Ruth Leão for their patience and support during this project. Marlon thanks his mother, Nora Flores, his late father, Ovidio López, and his sister Jennifer for all the support always given.

Preface

The purpose of this short course is to introduce some models of infectious diseases that are expressed as systems of ordinary differential equations (ODEs). The background expected of the reader is the introductory knowledge of ODEs that students usually acquire in calculus courses, together with matrix theory through eigenvalues and eigenvectors. Three good texts for furthering your knowldge of differential equations are Perko (2001), Schaeffer and Cain (2016), and Sotomayor (1979). However, we will introduce ideas about ODEs that may be new to you as they are needed.

The motivation for the course is the Covid-19 pandemic. During the pandemic the general public has become aware of the importance of mathematical models, both to anticipate the course of the pandemic and to evaluate possible interventions.

Researchers in mathematical epidemiology attempt to model a wide variety of infectious diseases using a variety of mathematical tools. To give some context for the course, we will describe some of this variety of both diseases and tools, and then describe our focus in this course.

Infectious diseases

Infectious diseases are disorders caused by organisms such as bacteria, viruses, fungi, or parasites. They have been responsible for enormous suffering and death throughout recorded history.

New infectious diseases have emerged continually during recorded history and

will continue to emerge in the future. The source is often some sort of transmission of diseases of other species.

Infectious diseases and their spread can be viewed as byproducts of human progress. Domestication of animals and human penetration into all global biomes have helped diseases to migrate from other species to humans. Global trade, which has existed since ancient times, has helped diseases to spread.

Progress in scientific understanding, sanitation, prevention measures, and treatments has led to improved control of many infectious diseases in most parts of the world. Our increased knowledge and experience have given us remarkable tools to bring to bear on the Covid-19 pandemic and on the infectious diseases that will emerge in the future.

Modes of transmission

We will give a few examples of especially deadly infectious diseases and their modes of transmission.

Plague is caused by a bacterium that is typically transmitted by the bite of a flea that previously bit an infected animal. It can also be transmitted from person to person by coughing. Plague pandemics have been among the most devastating episodes in human history (Frith 2012). The Justinian Plague originated in Ethiopia and reached Constantinople (now Istanbul) in 541 AD. It killed some 5,000 to 10,000 people per day in the city, and ultimately killed perhaps 100 million people in Africa, Asia and Europe over the next few years. There were repeated outbreaks over the next 200 years. In Europe, according to Frith (ibid.), "the social and economic disruption caused by the pandemic marked the end of Roman rule and led to the birth of culturally distinctive societal groups that later formed the nations of medieval Europe."

Plague reappeared in Europe in 1347 (the Black Death), brought from Asia Minor to Crimea by a Tartar army. It killed a quarter of the population of Europe, 25 million people, by 1350. Outbreaks continued in Africa, Asia and Europe for over 300 years. The Black Death led to the breakdown of medieval society and the growth of a middle class.

Plague reemerged in China in 1855 and was not fully controlled for a hundred years, by which time it had killed 15 million people, mostly in India.

Smallpox is caused by a virus that is spread by contact with patients' sores, by contact with contaminated objects such as bedding or clothing, and by coughing and sneezing. It was already present in 3rd century BC Egypt. It was brought to

the Americas, where it was unknown and there was no immunity, by Europeans starting in the 1520s. It is estimated that Old World diseases, principally smallpox, killed 90 to 95% of the indigenous population of the Americas. Although vaccination campaigns began in the 19th century, smallpox still killed 300 million to 500 million people during the 20th century. Smallpox was declared eradicated in 1979 (Wikipedia 2021e).

Malaria is caused by a parasite that is transmitted by mosquito bites. There were 229 million cases of malaria in 2019, leading to 409,000 deaths. 94% of cases and deaths were in Africa (CDC 2021).

Cholera is a bacterial disease usually spread through contaminated water. There have been seven cholera pandemics since the 19th century. Cholera currently kills at least 21,000 people per year (WHO 2021). A cholera epidemic in Haiti that began in 2010, following an earthquake, sickened almost 800,000 people (Wikipedia 2021b).

Acquired immunodeficiency syndrome (AIDS) is caused by the human immunodeficiency virus (HIV). It is transmitted sexually, by contact with infected blood or contaminated needles, and from mother to child. AIDS has killed around 33 million people since it was first identified in the 1980s (Wikipedia 2021f). It probably jumped from chimpanzees or gorillas to humans in Central Africa in the 1920s (Wikipedia 2021d).

While mathematical epidemiologists attempt to model all these diseases, in this course we shall be concerned with infectious diseases that are principally transmitted directly from one person to another.

Influenza is the prime example. It is caused by a family of viruses that are spread by coughing or sneezing. The first documented flu pandemic began in Asia in 1510 and spread along trade routes (Wikipedia 2021a).

The so-called Spanish flu pandemic of 1918–1920 killed as many as 100 million people worldwide. It was first observed in the state of Kansas in the United States in January 1918 (Wikipedia 2021h). It rapidly spread to other parts of the United States and Europe, and then around the world, reaching Brazil by August 1918. In Rio de Janeiro, the Spanish flu killed about 15,000 people and sickened another 600,000—about 66% of the city's population. "The city soon saw itself poised on the verge of collapse. There was not enough food, not enough medicine, not enough doctors, and not enough hospitals to take in the sickest. …The city streets gradually were transformed into a sea of unburied bodies, as there were not enough gravediggers to inter the bodies or caskets in which to place them." (Goulart 2005)

Mutations of the 1918 virus are responsible for most influenza cases since then

(Taubenberger and Morens 2006). Flu pandemics in 1957–58 and 1967–68 killed 1 to 4 million people worldwide (Wikipedia 2021g).

Coronaviruses are spread like influenza viruses. SARS-CoV was first reported in China in February 2003 and probably originated from bats. It spread to the Americas, Europe, and Asia and killed almost 800 people. MERS-CoV was first reported in Saudi Arabia in 2012. It emerged from bats via camels as an intermediate host, and has killed over 800 people. SARS-CoV-2, first reported in Wuhan, China, in December 2019, causes the syndrome known as Covid-19, which is presently a global pandemic. It is also generally believed to have emerged from bats. It has caused almost three million deaths as of mid-April 2021 (Wikipedia 2021c).

Models used in mathematical epidemiology

Our course will describe the use of ODEs to model the spread of diseases like influenza and the coronavirus diseases. ODE models are the ones most commonly used to anticipate the spread of these diseases and to explore the likely effect of countermeasures. ODE models divide a population into categories, called compartments, and describe the evolution of the populations fractions in the compartments over time. There may be just two compartments, infected and not infected, or a large number of compartments that divide the population in whatever ways seem important.

Here are some other types of models used in mathematical epidemiology, which we will not discuss.

Stochastic models

Especially at the start of an epidemic, when only a few people are infected, the element of chance is important in whether the epidemic spreads or dies out. ODE models are deterministic. Stochastic models take the probabilistic aspect of epidemics into account. An introductory reference is Allen (2008).

Network models

Both ODE models and stochastic models divide a population into compartments, and assume that members of compartments encounter each other at certain rates. Network models by contrast represent individuals as nodes in a network, and represent their contacts with each other by edges that connect the nodes. Similar to

stochastic models, disease is transmitted across edges probabilistically. Such models achieve added realism but are hard to analyze unless strong assumptions are made. A good reference is Kiss, Miller, and Simon (2017).

Another type of network model uses two types of nodes, one for individuals and one for mixing locations such as workplaces, stores, and schools. Edges connect individuals to mixing locations. These models have become important during the Covid-19 pandemic due to the availability of aggregate cellphone data that records the movement of people from homes to mixing locations (Chang et al. 2021).

Agent-based models

Agent-based models are computer programs that simulate the interactions of individuals (agents) in a given society over a period of time. They can be remarkably realistic.

In 2006 a group at Imperial College (London) created agent-based models to simulate flu epidemics in the United Kingdom and United States, based on data about population density, household size and age structure, schools, workplaces, and commuting; see Ferguson, Cummings, et al. (2006). The models were repurposed in a report of Ferguson, Laydon, et al. (2020) to predict the possible course of the Covid-19 pandemic in the UK and US. This report greatly influenced the response of the UK and US governments to the pandemic (Booth 2020).

COMORBUSS, an agent-based model developed in Brazil, is intended to carefully model a single city in order to advise which disease mitigation efforts would be most effective there (`https://comorbuss.org`, `http://www.cemeai.icmc.usp.br/ModCovid19/comorbuss`).

Problems with agent-based models include the effort required to build them, the time required to run them, and the fact that their interactions are probabilistic, so many runs may be required to get good predictions.

PDE models

In ODE models the variables are functions of time only. In partial differential equation (PDE) models the variables are functions of time and space. Thus PDE models can be used to study the spread of an epidemic in space. For example, Berestycki, Roquejoffre, and Rossi (2021) used a PDE model to study the early spread of Covid-19 by road networks in Italy.

ODE models in mathematical epidemiology

The fundamental ODE model of mathematical epidemiology is the SIR model, whose name represents its compartments, susceptible, infective, and recovered. It was introduced in a 1927 paper by A. G. McKendrick, a Scottish physician with experience fighting malaria in Sierra Leone and dysentery and rabies in India, and W. O. Kermack, a blind Scottish chemist (Kermack and McKendrick 1927). We shall discuss their model in Chapter 2. The SIS model (susceptible, infective, susceptible) is even simpler; we discuss it in Chapter 1.

A basic result underpinning a large part of applied mathematics is the Perron–Frobenius Theorem, which says, roughly speaking, that the principal eigenvalue of a positive matrix is positive and corresponds to a positive eigenvector. It is behind two important results of mathematical epidemiology. One explains why in many epidemiological models, if the susceptible population is renewed by a mechanism such as loss of immunity or births, a disease can become endemic; see Hethcote (1978). Another, the next generation matrix method, shows how to calculate the basic reproduction number in a complicated model.

The Perron–Frobenius Theorem is beyond the scope of this course. However, in Chapter 3, we use simpler arguments to show how renewal of the susceptible population in a simple SIR model can lead to a disease becoming endemic. And in Chapter 4 we explain the next generation matrix and how to use it, without going into proofs. Our main example in that chapter is an extension of the SIR model that represents the main features of Covid-19.

Chapter 5 introduces spontaneous human behavioral change. You know from experience that when infection levels rise, many people who can stay home will do so, and many will practice stricter hygiene and social distancing. When infection levels fall, people relax. This evident fact greatly affects the spread of an infectious disease, but is rarely accounted for in epidemiological models. How to deal with human behavioral change is at the research frontier in mathematical epidemiology. We explain an approach that uses imitation dynamics, an idea from game theory.

1 SIS Model

1.1 The model

In a human population, an infectious disease such as measles, influenza, or Covid-19 spreads due to a combination of pathogen characteristics and human behavior. Pathogen characteristics determine the circumstances under which an infective person can readily infect another. Human behavior determines how frequently those circumstances occur.

A word about English terminology. *Infectious diseases* are disorders caused by organisms such as bacteria, viruses, fungi, or parasites. Those that can be passed from person to person are called *contagious diseases*. This course will consider only contagious diseases. However, in mathematical epidemiology the term "contagious disease" is rarely used; the broader term "infectious disease" is almost always used instead. We will follow this tradition. An infected individual who is able to pass on the disease to another person is called *contagious*, *infectious*, or *infective*. These words all have the same meaning. In ordinary spoken English, only the first two are commonly used, but in mathematical epidemiology, the third is most common. We will follow this tradition by almost always using the word "infective" when referring to an individual or a collection of individuals.

In this chapter we consider an infectious disease for which no one has immu-

nity and immunity is never acquired. The common cold is an example. We assume that the population under consideration can be divided into two groups: the susceptibles (those who do not have the disease) and the infectives (those who have the disease; we assume they are all infective). In epidemiology these groups are called *compartments*.

A susceptible individual can acquire the disease from an infective individual. The susceptible individual then becomes infective, and remains so until the disease runs its course. Once the disease has run its course, the infective individual returns to being susceptible, since there is no immunity.

We assume that the population has a constant size. In Section 2.6 we will discuss epidemiological models for populations with changing size.

Let $S(t)$ denote the fraction of the population that is susceptible at time t, and let $I(t)$ denote the fraction of the population that is infective at time t. We of course have $S(t) \geqslant 0$, $I(t) \geqslant 0$, and $S(t) + I(t) = 1$. A susceptible individual becomes infective due to a contact with an infective individual that has the appropriate characteristics for transmission of the disease. These characteristics may relate to the length of the contact, the closeness of the individuals during the contact, where the contact occurs, whether the sick individual sneezes, etc. Such contacts are sometimes called *adequate contacts*. Of course there is also an element of probability in whether a contact actually results in transmission of the disease. We will address this aspect of adequate contacts in Section 1.5.

It is perhaps reasonable to assume that if we multiply $I(t)$ by a number k, then we will multiply the rate at which such contacts occur by k; and if we multiply $S(t)$ by a number k, then we will also multiply the rate at which such contacts occur by k. This means that the rate at which the disease is transmitted at time t is proportional to the product $S(t)I(t)$.

Similarly, it is perhaps reasonable to assume that the rate at which infectives become well at time t, and return to being susceptible, is proportional to $I(t)$.

These assumptions lead to the following pair of equations for the rates:

$$\frac{dS}{dt} = -\beta S(t)I(t) + \gamma I(t), \tag{1.1}$$

$$\frac{dI}{dt} = \beta S(t)I(t) - \gamma I(t). \tag{1.2}$$

In these equations, β and γ are positive proportionality constants. Notice the sign of each term.

In Sections 1.5 and 1.6 we will give a more careful derivation of (1.1)–(1.2), and we will explain the interpretation of the constants β and γ.

We will often use a dot to denote derivative with respect to t. Thus the equations (1.1) and (1.2) can be rewritten like this:

$$\dot{S}(t) = -\beta S(t)I(t) + \gamma I(t), \tag{1.3}$$

$$\dot{I}(t) = \beta S(t)I(t) - \gamma I(t). \tag{1.4}$$

Usually the equations are written more simply, like this:

$$\dot{S} = -\beta SI + \gamma I, \tag{1.5}$$

$$\dot{I} = \beta SI - \gamma I. \tag{1.6}$$

Equations (1.5)–(1.6) constitute the *SIS model*.

By adding (1.5) and (1.6), we find that $\dot{S} + \dot{I} = 0$, so if $S + I = 1$ initially, then $S + I$ remains 1 always. This makes sense since S and I are population fractions.

1.2 Differential equations background: how to do Western science

Equations (1.5) and (1.6) tell us that if we know the values of S and I at time t, then we know the rates at which S and I change at time t. The formulas just describe in mathematical terms a concept of how disease transmission works.

However, the equations do not tell us what happens over a long time. We do not know if the disease will die out, or spread until the entire population is always infected, or oscillate in prevalence, or tend toward being always present in the population at some intermediate level.

Equations (1.5) and (1.6) constitute a system of differential equations. Solutions of the system will tell us what happens.

The idea that Isaac Newton introduced into Western science in the 17th century is that our understanding of how the world works is usually an understanding about rates. Such an understanding can be expressed as a system of differential equations. The solutions of the system will tell us what happens.

1.3 Differential equations background: basics

Before we go on, we will give some basic notions and facts about differential equations from the rather geometric point of view that we will use.

Suppose $x(t) = (x_1(t), \ldots, x_n(t))$ is a moving point in $\mathbb{R}^n$. At time t, its velocity vector is $\dot{x}(t) = (\dot{x}_1(t), \ldots, \dot{x}_n(t))$. The velocity vector is usually drawn with its tail at the point $x(t)$.

For example, suppose $x(t) = (\cos t, \sin t)$, a moving point in $\mathbb{R}^2$. The point $x(t)$ runs around the circle of radius 1, centered at the origin. We have $\dot{x}(t) = (-\sin t, \cos t)$. Therefore $x(0) = (1, 0)$, $\dot{x}(0) = (0, 1)$, $x(\frac{\pi}{2}) = (0, 1)$, and $\dot{x}(\frac{\pi}{2}) = (-1, 0)$. These facts are illustrated in Figure 1.1.

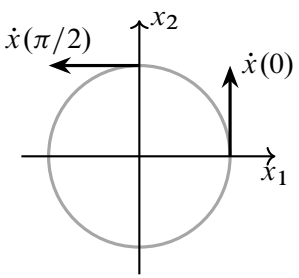


Figure 1.1: The curve $x(t) = (\cos t, \sin t)$ in gray and the velocity vectors $\dot{x}(0)$ and $\dot{x}(\frac{\pi}{2})$ in black.

As we have seen, a scientific idea often tells us that if we know x, a point that represents the state of the system, at some time, then we know $\dot{x}$, how x is changing, at that time. In other words, the velocity vector $\dot{x}$ is a function of the state x, i.e., $\dot{x} = f(x)$ or

$$\dot{x}_1 = f_1(x_1, \ldots, x_n), \tag{1.7}$$

$$\vdots$$

$$\dot{x}_n = f_n(x_1, \ldots, x_n). \tag{1.8}$$

An equation of the form $\dot{x} = f(x)$ is a *first-order autonomous ordinary differential equation*.

- First-order: there are only first derivatives, not higher derivatives.
- Autonomous: the derivative only depends on the state of the system x, not on the time t.
- Ordinary: there are only ordinary derivatives, not partial derivatives.

When a differential equation $\dot{x} = f(x)$ on $\mathbb{R}^n$ with $n > 1$ is written in the form (1.7)–(1.8), it is called a *system of differential equations*.

You surely remember Newton's Second Law of Motion, $F = ma$, or force equals mass times acceleration. Let x be the position of an object at time t, let $v = \dot{x}$ be its velocity, and let $a = \dot{v} = \ddot{x}$ be its acceleration. Newton's Second Law of Motion is the second-order differential equation $\ddot{x} = F/m$. However, we can rewrite it as a system of first-order differential equations in the variables (x, v):

$$\begin{aligned} \dot{x} &= v, \\ \dot{v} &= F/m. \end{aligned}$$

F can depend on x and v in various ways, depending on whether you are considering a spring, a pendulum, gravity, or whatever. Once you have an equation for F, you have a system of differential equations. If x is in $\mathbb{R}$, this is a system of two differential equations. If x is in $\mathbb{R}^3$, we have a system of six differential equations. (If the force depends on time, the system is nonautonomous. We do not consider nonautonomous systems in these notes.)

To use the differential equation $\dot{x} = f(x)$ to make a prediction of what will happen, i.e., to predict $x(t)$, we need to solve an *initial value problem*

$$\dot{x} = f(x), \quad x(t_0) = x_0.$$

In other words, given the differential equation $\dot{x} = f(x)$ and the state of the system at time t_0, $x(t_0) = x_0$, we need to find a function $x(t)$ such that $x(t_0) = x_0$ and, at every time t, $\dot{x}(t) = f(x(t))$.

For example, the system

$$\dot{x}_1 = -x_2, \tag{1.9}$$
$$\dot{x}_2 = x_1, \tag{1.10}$$

with the initial condition $(x_1(0), x_2(0)) = (1, 0)$, has the solution $(x_1(t), x_2(t)) = (\cos t, \sin t)$. To check that this is indeed a solution of the system, just substitute $\dot{x}_1(t)$ and $\dot{x}_2(t)$ into the left side, and substitute $x_1(t)$ and $x_2(t)$ into the right side:

$$\begin{aligned} -\sin t &= -\sin t, \\ \cos t &= \cos t. \end{aligned}$$

To check that $(x_1(0), x_2(0)) = (1, 0)$, just notice that $\cos 0 = 1$ and $\sin 0 = 0$.

The following theorem gathers some fundamental facts about differential equations:

Theorem 1.1. *Let U be an open set in $\mathbb{R}^n$, let $f : U \to \mathbb{R}^n$ be a continuously differentiable function, and let $x_0 \in U$. Then:*

1. *The initial value problem*

$$\dot{x} = f(x), \quad x(t_0) = x_0,$$

has a unique solution.

2. *If $x(t)$ stays in a compact (closed and bounded) subset of U as t increases (respectively decreases), then $x(t)$ is defined for $t_0 \leqslant t < \infty$ (respectively $-\infty < t \leqslant t_0$).*

Our differential equations $\dot{x} = f(x)$ will always have f continuously differentiable, so that this theorem applies.

The set U on which the differential equation is defined is called *phase space*.

The solution that Theorem 1.1 says exists may not be one that you, or anyone, can give an explicit formula for, but it is there, and it can be approximated numerically.

A point x_0 at which $f(x_0) = 0$ is an *equilibrium* of $\dot{x} = f(x)$.

Corollary 1.1. *If x_0 is an equilibrium of $\dot{x} = f(x)$, then the unique solution of the initial value problem*

$$\dot{x} = f(x), \quad x(t_0) = x_0,$$

is $x(t) = x_0$ for $-\infty < t < \infty$.

To prove this, just check that the formula for $x(t)$ gives a solution of the initial value problem, and recall that solutions are unique.

Corollary 1.2. *Let $x(t)$ be a solution of $\dot{x} = f(x)$. Suppose that at one time t_0, the point $x(t_0)$ is not an equilibrium. Then at every time t, the point $x(t)$ is not an equilibrium.*

This is an immediate consequence of the previous corollary.

Theorem 1.1 and its corollaries have important consequences that we will begin to see in the following section.

1.4 Phase line for the SIS system

For the SIS system (1.5)–(1.6), we saw at the end of Section 1.1 that $S(t)+I(t) = 1$ at every time t. Hence we do not really need both equations, since if we can calculate $I(t)$, we can find $S(t)$ from $S(t) = 1-I(t)$. We shall therefore use (1.5) only, after the substitution $S = 1 - I$:

$$\dot{I} = \beta(1-I)I - \gamma I = (\beta-\gamma)I - \beta I^2. \tag{1.11}$$

This is a single differential equation. You can find the general solution by writing it in the form $\frac{dI}{dt} = (\beta-\gamma)I - \beta I^2$ and using separation of variables and partial fractions (or an integral table or your favorite software). We shall do this for particular values of β and γ in the problems at the end of the chapter.

An easier way to see what is going on is to draw the *phase line*, which is the I-axis with dots where equilibria are located and arrows to show where solutions are increasing and decreasing. Where $\dot{I} > 0$, $I(t)$ is increasing; where $\dot{I} < 0$, $I(t)$ is decreasing; and where $\dot{I} = 0$, there is an equilibrium.

To draw the phase line, it may help you to first draw the graph of $\dot{I}$ as a function of I, i.e., draw the graph of

$$\dot{I} = (\beta-\gamma)I - \beta I^2 = I(\beta-\gamma-\beta I). \tag{1.12}$$

It is a parabola. You can use the graph to draw the phase line, since it helps you to see where $\dot{I}$ is positive, negative, and zero. For example, there are two equilibria, at $I = 0$ and $I = \frac{\beta-\gamma}{\beta} = 1 - \frac{\gamma}{\beta}$. See Figure 1.2.

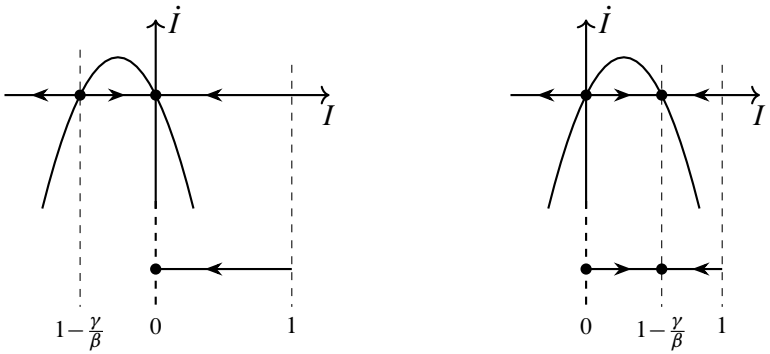
Actually, since I is a population fraction, only the interval $\mathcal{I} = \{I : 0 \leqslant I \leqslant 1\}$ is important. In other words, this interval is our phase space. When we restrict the phase line to the interval $\mathcal{I}$, we obtain the *phase portrait*.

Remembering that β and γ are positive, we see that there are two cases:

- If $\frac{\gamma}{\beta} > 1$, the nonzero equilibrium is not in $\mathcal{I}$.
- If $\frac{\gamma}{\beta} < 1$, the nonzero equilibrium is in $\mathcal{I}$.

(Of course there is a third case, $\frac{\gamma}{\beta} = 1$, but we will always ignore such unlikely intermediate cases.) See Figure 1.2.

By Corollary 1.2, solutions that start away from equilibria cannot pass through equilibria. Theorem 1.1 says that solutions that stay bounded are defined for infinite time. Another important fact is that in one dimension, bounded solutions must approach equilibria. Therefore:



(a) $\gamma/\beta > 1$. (b) $\gamma/\beta < 1$.

Figure 1.2: Phase portrait of the one-dimensional SIS equation (1.11) for $\frac{\gamma}{\beta} > 1$ and $\frac{\gamma}{\beta} < 1$. The parabola (1.12) is used to draw the phase line; then we restrict to the interval $0 \leqslant I \leqslant 1$. The arrows point left where $\dot{I} < 0$, and point right where $\dot{I} > 0$.

- If $\frac{\gamma}{\beta} > 1$, all solutions in $\mathcal{I}$ approach 0 as $t \to \infty$.
- If $\frac{\gamma}{\beta} < 1$, all solutions in $\mathcal{I}$ with $I(0) > 0$ approach $1 - \frac{\gamma}{\beta}$ as $t \to \infty$. Solutions with $0 < I(0) < 1 - \frac{\gamma}{\beta}$ approach 0 as $t \to -\infty$.

The interpretation of the phase portraits is as follows.

- If $\frac{\gamma}{\beta} > 1$ and the disease enters the population, the disease dies out.
- If $\frac{\gamma}{\beta} < 1$ and the disease enters the population, the fraction of the population with the disease tends toward the positive number $I = 1 - \frac{\gamma}{\beta}$, so eventually its prevalence in the population is roughly this value

In the second case the disease is said to be *endemic*, and the equilibrium at $I = 1 - \frac{\gamma}{\beta}$ is called the *endemic equilibrium*. The other possibilities we imagined, that $I(t)$ might tend to 1 or might oscillate, do not occur. (In general we do not consider the case $\gamma = 0$; we assumed $\gamma > 0$. The case $\gamma = 0$ would correspond to a disease from which one never gets well. In the case $\gamma = 0$ we would indeed have $I(t) \to 1$.)

The fact that solutions of differential equations take infinite time to approach equilibria can lead to confusion. In the case $\frac{\gamma}{\beta} > 1$, for example, solutions take infinite time to approach 0. It thus appears that the epidemic never quite dies out. In reality, however, population size is finite, so once the population fraction is sufficiently small, in fact no infected persons remain. It is for this reason that we say "the disease dies out." Similarly, in the case $\frac{\gamma}{\beta} < 1$, we will sometimes say that the disease eventually reaches the endemic equilibrium.

Epidemiologists do not distinguish the cases using the fraction $\frac{\gamma}{\beta}$ as we just did. Instead they use the fraction $\frac{\beta}{\gamma}$; they say that if $\frac{\beta}{\gamma} < 1$, the disease dies out, and if $\frac{\beta}{\gamma} > 1$, the disease is endemic. To explain why epidemiologists prefer $\frac{\beta}{\gamma}$ to $\frac{\gamma}{\beta}$, we will look more deeply into the meaning of the constants β and γ.

The common cold is supposed to be a disease to which the SIS model applies. The model predicts that the prevalence of the common cold is constant throughout the year. Is that not true where you live? Perhaps where you live β is not constant but varies with the season. If we replace β with a time-varying given function $\beta(t)$, then the SIS system is nonautonomous, so our analysis does not apply.

1.5 The constant β and the derivation of the SIS model

In this section we give a careful derivation of the SIS model and explain the meaning of the constant β. To achieve these goals we look at the actual size of the population instead of population fractions.

Recall that we are considering a population of constant size, i.e., we ignore births, deaths, immigration, and emigration. Let N be the population size. Let $s(t)$ denote the number of susceptibles and let $i(t)$ denote the number of infectives, so that $s(t) + i(t) = N$. The disease spreads due to contacts with appropriate characteristics for transmission of the disease. For simplicity we just use the word "contacts" to mean contacts with the right characteristics. If we multiply the rate at which an infected person has contacts with others by the fraction of the population that is susceptible, which is $s(t)/N$, we obtain the rate at which an infective person has contacts with susceptible persons. If time is measured in days, we have

$$\frac{\text{contacted persons}}{\text{infective person} \cdot \text{day}} \cdot \frac{s(t)}{N} = \frac{\text{contacted susceptible persons}}{\text{infective person} \cdot \text{day}}.$$

Multiplying this rate by the probability that contact results in disease, we obtain

the rate at which an infective person creates new infective persons:

$$\frac{\text{new infective persons}}{\text{infective person} \cdot \text{day}} = \frac{\text{contacted persons}}{\text{infective person} \cdot \text{day}} \cdot \frac{s(t)}{N} \cdot \text{probability of transmission.} \quad (1.13)$$

We define

$$\beta = \frac{\text{contacted persons}}{\text{infective person} \cdot \text{day}} \cdot \text{probability of transmission.} \quad (1.14)$$

We use the symbol β because this will turn out to be the meaning of β in the SIS system (1.5)–(1.6). In mathematical epidemiology β is called the *adequate contact rate*. It is the rate of new infections per infective when all contacted persons are susceptible (i.e., when $s(t) = N$), as occurs at the beginning of an epidemic. The term "adequate contact" was already used in Section 1.1.

By definition, β is the product of two terms. Under normal circumstances, when individuals are not taking steps such as mask-wearing to protect against disease, the second term, the probability of transmission, is a property of the disease itself, i.e., how contagious it is. The first term, however, depends of the mode of life of the population. For example, it may be higher in an urban area, in which individuals typically come into contact with many others in the course of a day, than in a rural area.

The rate at which new infective persons in the entire population are created is obtained multiplying (1.13) by $i(t)$:

$$\frac{\text{new infective persons}}{\text{day}} = \beta \cdot \frac{s(t)}{N} \cdot i(t).$$

The rate at which infectives become well is proportional to $i(t)$ with the same proportionality constant γ used previously.

We obtain the system of differential equations

$$\dot{s} = -\frac{\beta}{N} si + \gamma i, \quad (1.15)$$

$$\dot{i} = \frac{\beta}{N} si - \gamma i. \quad (1.16)$$

Remember that S and I are population fractions, so $S = \frac{s}{N}$ and $I = \frac{i}{N}$. To derive the system (1.5)–(1.6) from (1.15)–(1.16), in (1.15)–(1.16) just make the substitutions $s = NS$, $i = NI$, $\dot{s} = N\dot{S}$, and $\dot{i} = N\dot{I}$. Try it!

It is interesting to check the units in (1.15)–(1.16). Let us just look at the first equation. If time is measured in days, then the unit of $\dot{s}$ is persons/day. Since N, s and i all have the unit persons, the unit of β must be 1/days. This is correct: β represents new infected persons per infected person per day when everyone is susceptible. The person units cancel.

Similarly, the unit of γ is 1/days. But what does γ mean exactly?

1.6 The constant γ

A *probability distribution* on an interval J is a function $g(t)$ defined on J such that $g(t) \geqslant 0$ and $\int_J g(t)\ dt = 1$. If K is a subinterval of J, then the probability that t lies in that subinterval is $\int_K g(t)\ dt$. The *mean* of t is $\int_J tg(t)\ dt$. This is analogous to how the mean, or average, of a finite probability distribution is calculated.

Let us ignore recruitment into the infective compartment, and for simplicity we assume that the entire population is infected at time 0. Then (1.6) simplifies to the initial value problem

$$\dot{I} = -\gamma I, \quad I(0) = 1.$$

The solution is $I(t) = e^{-\gamma t}$.

$\dot{I}(t) = -\gamma e^{-\gamma t}$ is the rate at which $I(t)$ changes. It is negative since $I(t)$ decreases as people become well. The rate at which people become well (expressed as a population fraction per unit of time) is $-\dot{I}(t) = \gamma e^{-\gamma t}$, which is positive.

Eventually everyone becomes well:

$$\int_0^\infty -\dot{I}(t)\ dt = \int_0^\infty \gamma e^{-\gamma t}\ dt = 1. \tag{1.17}$$

You should check this calculation. Since the integral is 1, $-\dot{I}(t) = \gamma e^{-\gamma t}$ is a probability distribution on the interval $0 \leqslant t < \infty$.

Since people who become well at time t were ill for time t, the average length of time that people are ill is just the mean value of t on the interval $0 \leqslant t < \infty$. This mean value is

$$\int_0^\infty -t\dot{I}(t)\ dt = \int_0^\infty \gamma t e^{-\gamma t}\ dt = \frac{1}{\gamma}. \tag{1.18}$$

You should check this calculation too.

Thus we have our interpretation of γ: $1/\gamma$ is the average length of time that people are ill with the disease being modeled. Notice that if the unit of γ is 1/days, as in the previous section, then the $1/\gamma$ is time in days, which makes sense.

1.7 The basic reproduction number R_0

The basic reproduction number R_0, pronounced "R naught" in English, is the most important value calculated in epidemiological models. It has become known to the general public during the Covid-19 pandemic through countless news articles.

The basic reproduction number R_0, pronounced "R naught" in English, is the most important value calculated in epidemiological models. It has become known to the general public during the Covid-19 pandemic through countless news articles.

R_0 is the average number of people infected by an infective individual when a disease is introduced into a population, *under the assumption that the entire population, in particular everyone that individual encounters, is susceptible to the disease*. In the SIS model, R_0 is the adequate contact rate (i.e., the number of new infections caused per day by one infective when the entire population is susceptible) times the average number of days that an individual is infective. In symbols,

$$R_0 = \beta \cdot \frac{1}{\gamma} = \frac{\beta}{\gamma}.$$

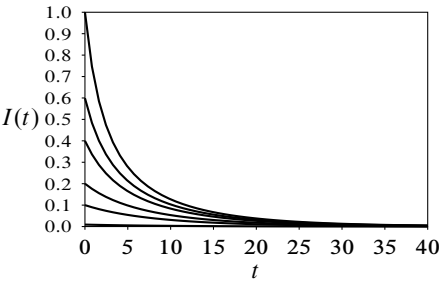
When a disease is introduced into a population, by the time an infective individual recovers, she has replaced herself with R_0 other infective individuals. Thus if $R_0 > 1$, the disease spreads; if $R_0 < 1$, it dies out.

Let us look again at the phase portraits in Figure 1.2. Consider a solution $I(t)$ that starts near $I = 0$, so very few people are infective and almost everyone is susceptible.

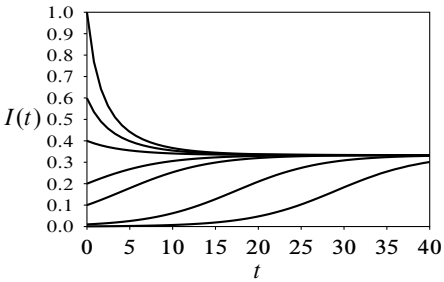
- If $\frac{\gamma}{\beta} > 1$, then $R_0 = \frac{\beta}{\gamma} < 1$. You can see that $I(t)$ declines toward 0.
- If $\frac{\gamma}{\beta} < 1$, then $R_0 = \frac{\beta}{\gamma} > 1$. You can see that if $I(t)$ starts near 0, $I(t)$ increases toward the nonzero equilibrium value, which is

$$I = 1 - \frac{\gamma}{\beta} = 1 - \frac{1}{R_0}.$$

Sample solutions are shown in Figure 1.3.



(a) $\beta = 0.3$, $\gamma = 0.4$, so $R_0 = 3/4$. All solutions approach 0.



(b) $\beta = 0.6$, $\gamma = 0.4$, so $R_0 = 3/2$. There is an equilibrium at $1 - 2/3 = 1/3$. All solutions with $I(0) > 0$ approach 1/3.

Figure 1.3: Solutions of the one-dimensional SIS equation (1.11) for initial values 0.001, 0.01, 0.1, 0.2, 0.4, 0.6, and 1.

You may want to google estimated values of R_0 for infectious diseases that interest you.

1.8 The effective reproduction number R_e

As an epidemic proceeds, the susceptible population fraction changes, so the number of people infected by a single infective individual also changes. The *effective reproduction number* at a given time, denoted R_e, is defined to be the average number of people infected by a single infective individual at that time.

For the SIS model, when the susceptible population fraction is S, equations (1.13) and (1.14) imply that the average rate at which a single infective individual infects others is βS. Therefore at a time when the susceptible population fraction is S,

$$R_e = \beta S \cdot \frac{1}{\gamma} = \frac{\beta S}{\gamma}. \tag{1.19}$$

One expects that when $R_e < 1$, the number of infectives is falling, and when $R_e > 1$, the number of infectives is rising. This is correct for the SIS model in the

region $I > 0$. From (1.19),

$$R_e < 1 \leftrightarrow \frac{\beta S}{\gamma} < 1 \leftrightarrow \frac{\beta(1-I)}{\gamma} < 1 \leftrightarrow \beta - \beta I < \gamma \leftrightarrow \beta - \gamma - \beta I < 0.$$

On the other hand, from (1.11),

$$\dot{I} = (\beta - \gamma - \beta I)I.$$

Therefore in the region $I > 0$, $R_e < 1$ if and only if $\dot{I} < 0$. Similarly, in the region $I > 0$, $R_e > 1$ if and only if $\dot{I} > 0$.

The formula (1.19) can also be used to calculate R_e at an equilibrium. However, the interpretation of R_e at the endemic equilibrium $I = 1 - \frac{\gamma}{\beta}$ differs from the interpretation at the disease-free equilibrium $I = 0$.

At the endemic equilibrium $I = 1 - \frac{\gamma}{\beta}$ we have $R_e = 1$ and, of course, $\dot{I} = 0$. This makes sense: at the endemic equilibrium, each infected individual replaces himself with exactly one new infective individual, so the value of I remains constant.

However, the same logic cannot be applied to the disease-free equilibrium $I = 0$. At the disease-free equilibrium there are no infectives, so it does not make sense to discuss of the number of people infected by each infective individual.

At $I = 0$, $R_e = R_0$, and we saw in the previous section that R_0 can be less than or greater than 1. At the disease-free equilibrium, R_e should be interpreted as the the approximate number of people infected by a single infective individual *near* the equilibrium.

1.9 Discussion of the SIS model

The assumption that a disease spreads at a rate proportional to the product of the infective and susceptible population fractions is sometimes called the *law of mass action*. This law comes from chemistry. If a well-stirred solution contains two reactants, the law of mass action says that the rate of the reaction is proportional to the product of the concentrations of the two reactants. In epidemiology the analog of a well-stirred solution in chemistry is a well-mixed population, in which people encounter each other randomly.

Human populations are rarely well-mixed in the modern world. People preferentially encounter certain subgroups of the population: family or roommates, friends, neighbors, coworkers or fellow students, people who ride the same bus,

etc. Epidemiologists use *network models* and *agent-based models* to capture these complications. They are harder to develop, run, and analyze than differential equations models. Another way to model these complications is to introduce more compartments into differential equations models to represent different social groups, age groups, etc. We will look at a model with subgroups in Problem 4.4.

Another assumption of the SIS model is that no one enters or leaves the population being modeled. A good exercise is to try to figure out other assumptions.

A differential equation model such as the SIS model can be misleading at the start of an epidemic. If $R_0 > 1$, the SIS model predicts that if even a single individual somehow contracts the infectious disease, it will spread until the endemic equilibrium is reached. In fact there is a degree of chance involved in whether a single infective individual infects anyone else, and if so how many. *Stochastic models* are used to quantify the probability that a disease that is initially contracted by a small number of people will actually spread.

According to statistician George Box, "All models are wrong, but some are useful." Extensions of the SIS model that we will explore in subsequent sections are certainly wrong but definitely useful: they have been used by governments throughout the Covid-19 pandemic to predict the progression of the epidemic under possible government policies or changes in population behavior.

1.10 Problems

Problem 1.1 *Phase lines*

Draw the phase lines for the following differential equations.

1. $\dot{x} = (x-1)(x-2)(x-3)$.
2. $\dot{x} = x^2(1-x)$.

Problem 1.2 *Derivation of SIS in dimensionless form*

Derive the system (1.5)–(1.6) from (1.15)–(1.16) by making the substitutions $s = SN$ and $i = IN$ in (1.15)–(1.16).

Problem 1.3 *Probability distribution*

Show that for $\gamma > 0$, $\int_0^\infty \gamma e^{-\gamma t}\, dt = 1$.

Problem 1.4 *Mean time*

Show that for $\gamma > 0$, $\int_0^\infty \gamma t e^{-\gamma t}\ dt = \frac{1}{\gamma}$. (You may want to find a table of integrals on the web.)

Problem 1.5 *Endemic equilibrium*

The endemic equilibrium in the SIS model is $I = 1 - \frac{\gamma}{\beta}$. Thus I depends on β and γ. Suppose you are in charge of public health and you think that the endemic equilibrium is too high, i.e. too many people are sick at any given time. How might you try to reduce the value of the endemic equilibrium?

Problem 1.6 *SIS with disease importation*

In the SIS model (1.1)–(1.2), suppose people sometimes get sick due to visitors from outside the population who arrive with the disease and transmit it. We modify the model as follows:

$$\dot{S} = -\beta SI + \gamma I - \mu S,$$
$$\dot{I} = \beta SI - \gamma I + \mu S.$$

The modified model just says that susceptibles can contract the disease at a rate proportional to their population fraction, independent of the number of infectives in the population, due to visitors.

1. Since $\dot{S} + \dot{I} = 0$, we should only need one equation. Use the $\dot{I}$ equation with the substitution $S = 1 - I$, and multiply out to express the equation as a polynomial in I. Answer:

$$\dot{I} = \mu + (\beta - \gamma - \mu)I - \beta I^2.$$

2. The graph of $\dot{I}$ as a function of I is a parabola. Since the coefficient of I^2 is negative, the parabola opens downward. Show that when $I = 0$, $\dot{I}$ is positive, and when $I = 1$, $\dot{I}$ is negative.

3. Use the facts in part 2 to explain why there is a unique equilibrium in $\mathcal{I}$, and all solutions in $\mathcal{I}$ approach it.

Problem 1.7 *Solutions of the SIS system*

We consider (1.11) with $\beta = 2$ and $\gamma = 1$ (so $R_0 > 1$), and we rewrite $\dot{I}$ as $\frac{dI}{dt}$:

$$\frac{dI}{dt} = I - 2I^2.$$

1. In the interval $0 < I < \frac{1}{2}$, find the general solution using separation of variables and partial fractions. Answer:

$$\ln \frac{I}{1-2I} = t + C.$$

   We chose an interval where you would not have to worry about absolute values when you used the natural logarithm function.

2. We want I as a function of t, so solve for I and write your answer as simply as you can. Answer:

$$I = \frac{A}{2A + e^{-t}}, \quad A > 0.$$

3. Check by substitution into the differential equation that in fact all functions of this form are solutions, whether or not $A > 0$. The constant A is determined by the initial condition $I(0) = I_0$.

4. From the answer in part 2 it appears that $\lim_{t\to-\infty} I(t) = 0$ and that $\lim_{t\to\infty} I(t) = \frac{1}{2}$. From the phase line, this is correct if $0 < I(0) < \frac{1}{2}$. However, if $I(0) > \frac{1}{2}$, we see from the phase line that as t decreases, $I(t) \to \infty$. What is wrong? Suggestion: for definiteness, look at the solution with $I(0) = 1$. (This problem shows that the phase line can be easier to understand than the general solution.)

2 SIR model

2.1 The model

In this chapter we consider an infectious disease that is not fatal and that confers permanent immunity on people who contract it. To model such a disease we divide the population into three compartments: susceptible, infective, and recovered (hence immune). As with the SIS model we consider a population of constant size; we will discuss changing population size in Section 2.6. Let $S(t)$, $I(t)$, and $R(t)$ denote the population fractions in each compartment at time t. We of course have $S(t) \geq 0$, $I(t) \geq 0$, $R(t) \geq 0$, and $S(t) + I(t) + R(t) = 1$. Assuming the law of mass action, the governing system of differential equations is similar to the SIS system:

$$\dot{S} = -\beta SI, \tag{2.1}$$

$$\dot{I} = \beta SI - \gamma I, \tag{2.2}$$

$$\dot{R} = \gamma I. \tag{2.3}$$

The constants $\beta > 0$ and $\gamma > 0$ have the same meaning they had in the SIS model. Equations (2.1)–(2.3) constitute the *SIR model*. The only difference from the SIS model is that when infectives recover they do not return to the susceptible com-

partment; instead they move into the recovered compartment and can no longer contract the disease.

The SIR model can also be used when some fraction of the population is not susceptible to the disease for a reason such as genetics, behavior, previously acquired immunity, etc. This population fraction is included in the recovered compartment.

In addition, the SIR model can be used for a disease such as Covid-19 that is sometimes fatal. Those who die are included in the R compartment, and the compartment is called "removed." The constant $\beta > 0$ has almost the same meaning as in the SIS model, but the constant γ now represents the rate at which infectives either recover or die. We show how to derive the SIR model for this case in Sec. 2.6.

Recall the basic reproduction number R_0, which for the SIS model was given by

$$R_0 = \beta \cdot \frac{1}{\gamma} = \frac{\beta}{\gamma}.$$

For the model (2.1)–(2.3), R_0 has the same formula.

You can check that $\dot{S} + \dot{I} + \dot{R} = 0$, so if $S + I + R = 1$ initially, then $S + I + R$ remains 1 always. Thus we do not need all three equations, since if we can calculate $S(t)$ and $I(t)$, we can find $R(t)$ from $R(t) = 1 - S(t) - I(t)$. We shall therefore use (2.1) and (2.2) only:

$$\dot{S} = -\beta SI, \tag{2.4}$$

$$\dot{I} = \beta SI - \gamma I. \tag{2.5}$$

For the SIS model we could calculate formulas for solutions, but the phase portrait was more helpful. For the SIR model it is not possible to find formulas for solutions. We therefore turn to the phase portrait. The system is 2-dimensional, so we need some more differential equations background.

2.2 Differential equations background: vector fields and nullclines

Geometrically, a differential equation $\dot{x} = f(x)$, with $x \in \mathbb{R}^n$ and f a function from an open set U in $\mathbb{R}^n$ to $\mathbb{R}^n$, defines a *vector field* on U. The vector $f(x)$ at the point x is drawn with its tail at x. When $n = 2$, as for the SIR system (2.4)–(2.5),

it is often not hard to get an idea what the vector field looks like. One can begin by finding the curves where $\dot{x}_1 = 0$ or $\dot{x}_2 = 0$ (the *nullclines*).

For the system (2.4)–(2.5), we see that $\dot{S} = 0$ when $S = 0$ or $I = 0$, i.e., on the two axes is the SI-plane. We also see that $\dot{I} = 0$ when $I = 0$ or $S = \frac{\gamma}{\beta}$. We have equilibria where both $\dot{S} = 0$ and $\dot{I} = 0$. We conclude that the line $I = 0$ (the S-axis) consists of equilibria, and there are no other equilibria.

The vector field is vertical on the $\dot{S} = 0$ nullclines and horizontal on $\dot{I} = 0$ nullclines (except where they intersect). The nullclines divide the plane into open regions on which $\dot{S}$ and $\dot{I}$ each has a constant sign. The signs in each region determine whether the vectors in that region point northeast, northwest, southwest, or southeast. Usually you can tell the direction in the open region by looking at which way the vectors point (up or down, right or left) on the nullclines that bound it.

We are only interested in the triangle

$$\mathcal{T} = \{(S, T) : S \geqslant 0,\ I \geqslant 0,\ S + I \leqslant 1\},$$

which is our phase space. Thus there are two cases: $R_0 = \frac{\beta}{\gamma} < 1$ and $R_0 = \frac{\beta}{\gamma} > 1$. In the first case, $\frac{\gamma}{\beta} > 1$, so the line $S = \frac{\gamma}{\beta}$ does not meet $\mathcal{T}$; in the second case, $\frac{\gamma}{\beta} < 1$, so it does.

The vector field on $\mathcal{T}$ in the two cases is shown in Figure 2.1.

- First notice the equilibria along the S-axis in both cases.
- Next, look at the I-axis in both cases. On it $\dot{S} = 0$ and $\dot{I} < 0$, so the vectors point straight down.
- In the case $R_0 < 1$, the positive part of the triangle

  $$\mathcal{T}_+ = \{(S, T) : S > 0,\ I > 0,\ S + I \leqslant 1\}$$

  lies in a single region, since no nullclines cut it. In this region it is easy to check that $\dot{S} < 0$ and $\dot{I} < 0$, so vectors point southwest (to the left and down).
- Finally, in the case $R_0 > 1$, $\mathcal{T}_+$ is cut in two by the nullcline $S = \frac{\gamma}{\beta}$, on which $\dot{I} = 0$. You can check that $\dot{S} < 0$ throughout $\mathcal{T}_+$. On the other hand, $\dot{I}$ is positive when $S > \frac{\gamma}{\beta}$ and negative when $S < \frac{\gamma}{\beta}$.

As with the SIS model, the sign of $\dot{I}$ is related to the effective reproduction number R_e (see Section 1.8.). The effective reproduction number for the SIR model has the same formula as for the SIS model, $R_e = \frac{\beta S}{\gamma}$. From (2.2) you can easily check that in the region $I > 0$,

- $R_e < 1$ if and only if $\dot{I} < 0$.
- $R_e = 1$ if and only if $\dot{I} = 0$.
- $R_e > 1$ if and only if $\dot{I} > 0$.

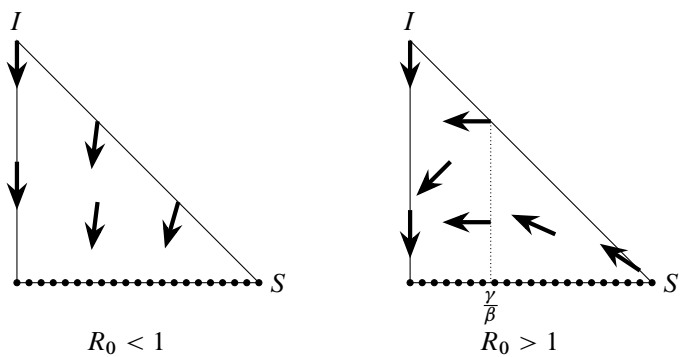


Figure 2.1: Nullclines, equilibria, and vector field for (2.4)–(2.5) in the triangle $\mathcal{T}$.

2.3 Differential equations background: functions and differential equations

Consider the differential equation $\dot{x} = f(x)$ on $\mathbb{R}^n$. Let $x(t)$ be a solution. Let $V : \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable function. Then $V(x(t))$ gives the value of V along the solution as a function of t. According to the chain rule, the rate of change of V along the solution is

$$\dot{V} = \frac{\partial V}{\partial x_1}(x(t))\,\dot{x}_1(t) + \ldots + \frac{\partial V}{\partial x_n}(x(t))\,\dot{x}_n(t) = \nabla V(x(t)) \cdot \dot{x}(t) =$$
$$= \nabla V(x(t)) \cdot f(x(t)).$$

where $\nabla V(x) = \left(\frac{\partial V}{\partial x_1}(x), \ldots, \frac{\partial V}{\partial x_n}(x)\right)$ is the gradient of V at the point x, and $\cdot$ represents dot product.

A nice way to think of this formula is like this: if a solution of $\dot{x} = f(x)$ passes through a point x, then the derivative of V along the solution at that point is

$$\dot{V}(x) = \nabla V(x) \cdot f(x). \tag{2.6}$$

For example, consider the SIR system (2.4)–(2.5) in the case $R_0 > 1$. If a solution starts on the line $S + I = 1$ at a point with $\frac{\gamma}{\beta} < S < 1$, we know that the vector there points northwest. But does it point into $\mathcal{T}$ or out of $\mathcal{T}$? It would be very bad if it pointed out of $\mathcal{T}$. That would mean that solutions that start on $S + I = 1$ soon have $S + I > 1$. This would not make sense since S and I are supposed to be population fractions. We would have a bad model.

To see if this can really happen, we write the SIR system (2.1)–(2.2) in vector form as

$$(\dot{S}, \dot{I}) = f(S, I) = (-\beta SI, \beta SI - \gamma I),$$

and consider the function $V(S, I) = S + I$. We compute:

$$\nabla V(S, I) \cdot f(S, I) = (1, 1) \cdot (-\beta SI, \beta SI - \gamma I) = -\gamma I < 0 \text{ if } I > 0.$$

Thus the function $S + I$ is decreasing along solutions of the SIR system in the region $I > 0$. Hence we can be sure that if a solution starts in $\mathcal{T}_+$ on $S + I = 1$, $S + I$ will immediately decrease, so the solution will enter the interior of $\mathcal{T}$.

Often it is easier not to use the formula (2.6) to compute $\dot{V}$. For the example we just considered, with $V(S, I) = S + I$ and $(\dot{S}, \dot{I}) = (-\beta SI, \beta SI - \gamma I)$, we could just calculute

$$\dot{V} = \dot{S} + \dot{I} = -\beta SI + \beta SI - \gamma I = -\gamma I.$$

As another example, for $V(S, I) = S^2 + I^2$ and the same differential equation, we could just use the chain rule to compute

$$\dot{V} = 2S\dot{S} + 2I\dot{I} = 2S(-\beta SI) + 2I(\beta SI - \gamma I) = (2I - 2S)\beta SI - 2\gamma I^2.$$

Another use of the equation for $\dot{V}$ is the following theorem:

Theorem 2.1. *Suppose that whenever $V(x) = c$, we have $\nabla V(x) \neq 0$ and $\dot{V}(x) = \nabla V(x) \cdot f(x) = 0$. Then the set $V(x) = c$ is invariant under $\dot{x} = f(x)$, i.e., a solution of $\dot{x} = f(x)$ that starts in the set $V(x) = c$ stays in the set $V(x) = c$.*

For example, for the SIR system (2.4)–(2.5), $I = 0$ implies $\dot{I} = 0$. According to Theorem 2.1, it follows that the line $I = 0$ is invariant. (The function V is $V(S, I) = I$.)

2.4 Orbits and phase portrait of the SIR system

An *orbit* of a differential equation is the curve in phase space that is traced out by a solution. The word "curve" in this definition should be interpreted generously; an equilibrium is an orbit.

We can find the orbits of the SIR system (2.4)–(2.5) by a method you learned in calculus. We rewrite the SIR system as

$$\frac{dS}{dt} = -\beta SI, \tag{2.7}$$

$$\frac{dI}{dt} = \beta SI - \gamma I. \tag{2.8}$$

You learned in calculus that you can divide the second equation by the first to get a differential equation for the orbits in SI-space that are traced out by solutions. Dividing, we get

$$\frac{dI}{dS} = -1 + \frac{\gamma}{\beta S}. \tag{2.9}$$

Integrating, we obtain the family of curves

$$I = -S + \frac{\gamma}{\beta} \ln S + C. \tag{2.10}$$

You can check that all these curves attain their maximum value at $S = \frac{\gamma}{\beta}$, as we would expect from Figure 2.1.

Figure 2.2 shows one of these curves in the case $R_0 > 1$. Let us consider a solution $(S(t), I(t))$ that starts at a point on this curve that is in $\mathcal{T}_+$. The curve traced out by the solution is not the entire curve $I = -S + \frac{\gamma}{\beta} \ln S + C$, because by Corollary 1.2, solutions cannot pass through the equilibria on the S-axis. The curve traced out by this solution is just the part of the curve $I = -S + \frac{\gamma}{\beta} \ln S + C$ that is in $I > 0$. The curve $I = -S + \frac{\gamma}{\beta} \ln S + C$ intersects the S-axis in two points $(S_-, 0)$ and $(S_+, 0)$, with $S_+ < \frac{\gamma}{\beta} < S_-$, and

$$\lim_{t \to -\infty} (S(t), I(t)) = (S_-, 0) \text{ and } \lim_{t \to \infty} (S(t), I(t)) = (S_+, 0).$$

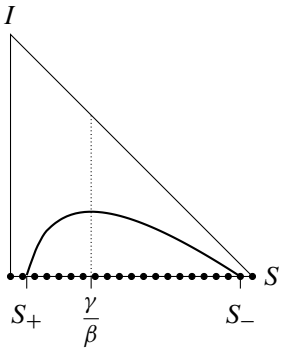
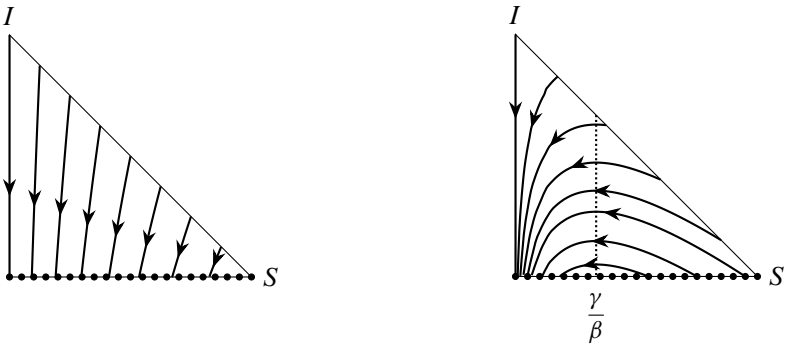


Figure 2.2: One orbit of the SIR system with $R_0 = \frac{\beta}{\gamma} = 3$, so $\frac{\gamma}{\beta} = \frac{1}{3}$. We have chosen $S_- = 0.95$, so S_+ (computed as below) is 0.0672.

Remark 2.2. Thus the curves (2.10) we found by solving the differential equation (2.9) are not orbits of the SIR system (2.7)–(2.8); they are *invariant curves*. A curve is invariant if a solution that starts on the curve stays on the curve. In general, invariant curves are unions of orbits, i.e., they may consist of more than one orbit. For example, each curve (2.10) contains two equilibria where it crosses the S-axis. In Theorem 2.1 we described another way that one can find invariant curves. There we noted that the line $I = 0$ is invariant for the SIR system. It also consists of more than one orbit, since the origin is an equilibrium.

A *phase portrait* of $\dot{x} = f(x)$ is a sketch of phase space that shows all unusual orbits and examples of typical orbits, together with arrows on the orbits that indicate the direction of movement.

For the SIR system, the equilibria on the S-axis and the vertical orbit along the I-axis qualify as unusual; the curves in Equation (2.10) qualify as typical. The formula (2.10) shows that the phase portrait of the SIR system depends only on the ratio $\frac{\gamma}{\beta}$, or equivalently, it depends only on $R_0 = \frac{\beta}{\gamma}$. Figure 2.3 shows the phase portrait of the SIR system in the two cases $R_0 < 1$ and $R_0 > 1$. The second phase portrait uses a value of γ that is reasonable for Covid-19, and a value of β that was considered reasonable for Covid-19 in urban areas near the start of the pandemic.



(a) $R_0 = 1/3$ ($\beta = 0.1, \gamma = 0.3$). (b) $R_0 = 3$ ($\beta = 0.3, \gamma = 0.1$).

Figure 2.3: Phase portraits of the SIR system. In phase portrait (b), $\frac{\gamma}{\beta} = \frac{1}{3}$.

2.5 Interpretation of the orbits

In the case $R_0 > 1$, many of the orbits (2.10) of the SIR system are like the one shown in Figure 2.2: they connect an equilibrium $(S, I) = (S_-, 0)$, with $\frac{\gamma}{\beta} < S_- \leqslant 1$, to an equilibrium $(S, I) = (S_+, 0)$, with $0 < S_+ < \frac{\gamma}{\beta}$. One can interpret such an orbit as follows. An epidemic starts at a population state near $(S, I) = (S_-, 0)$ with $\frac{\gamma}{\beta} < S_- \leqslant 1$. In other words, the population fraction S_- is susceptible and no one is yet infected. The remaining population fraction $R_- = 1 - S_-$ is not susceptible to the disease for a reason such as those mentioned in Section 2.1. When the disease is introduced, so that I becomes slightly positive, the number of infectives increases, and the number of susceptibles drops. Eventually the number of susceptibles falls below $\frac{\gamma}{\beta}$, and the number of infectives starts to drop. The disease then dies out.

In terms of the effective reproduction number $R_e = \frac{\beta S}{\gamma}$, the epidemic begins near an equilibrium $(S_-, 0)$ with $\frac{\gamma}{\beta} < S_- \leqslant 1$; these equilibria have $R_e > 1$. The number of infectives therefore rises, and the number of susceptibles falls, which causes the effective reproduction number R_e to fall. After R_e passes 1 at $S = \frac{\gamma}{\beta}$ and falls below 1, the number of infectives falls, and the epidemic dies out.

At the end of the epidemic the susceptible population fraction is S_+ with $0 < S_+ < \frac{\gamma}{\beta}$. Thus $S_- - S_+$ is the fraction of the population that contracted the disease during the epidemic.

It is therefore of interest of compute S_+ when S_- is given. We can do this as

follows. If the curve (2.10) passes through the point $(S_-, 0)$, then

$$0 = -S_- + \frac{\gamma}{\beta} \ln S_- + C \implies C = S_- - \frac{\gamma}{\beta} \ln S_-.$$

Then if the curve also passes through the point $(S_+, 0)$, we have

$$0 = -S_+ + \frac{\gamma}{\beta} \ln S_+ + C = -S_+ + \frac{\gamma}{\beta} \ln S_+ + S_- - \frac{\gamma}{\beta} \ln S_-$$

More concisely,

$$-(S_+ - S_-) + \frac{\gamma}{\beta}(\ln S_+ - \ln S_-) = 0.$$

Given S_- you can find S_+ by writing

$$F(S) = -(S - S_-) + \frac{\gamma}{\beta}(\ln S - \ln S_-) = 0.$$

and solving for S using a numerical method. In Figure 2.2, $S_- = 0.95$ and $S_+ = 0.0672$.

Of particular interest is the value of S_+ when $S_- = 1$, i.e., when initially the entire population is susceptible to the disease. Then $1 - S_+$ gives the fraction of the population infected during the epidemic. For $S_- = 1$, Figure 2.4 shows $1 - S_+$ as a function of $R_0 = \frac{\beta}{\gamma}$ for $1 < R_0 \leqslant 5$.

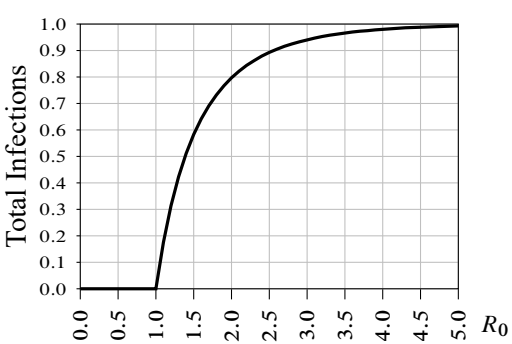


Figure 2.4: Total population fraction infected by the end of the epidemic as a function of R_0 for the SIR model with $S_- = 1$ and $0 \leqslant R_0 \leqslant 5$.

Also of interest is the maximum value of I on the orbit, which gives the maximum fraction of the population infected at one time during the epidemic. This number helps tell whether the hospital system will be overwhelmed at some point. The maximum value of I occurs at $S = \frac{\gamma}{\beta}$. In Figure 2.2, the maximum value of I is 0.2676. For $S_- = 1$, Figure 2.5 shows the maximum value of I as a function of R_0.

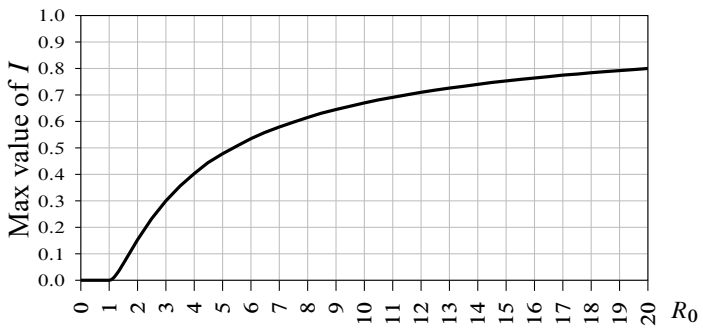


Figure 2.5: Maximum value of I as a function of R_0 for the SIR model with $S_- = 1$ and $0 < R_0 \leqslant 20$.

Let us pause for a moment to consider some implications of Figure 2.5. For $R_0 = 3$, which was considered a reasonable estimate for Covid-19 in urban areas near the start of the pandemic, the maximum value of I is 0.3005. This means that 30% of the population could be infected at one time. The population of Brazil is about 211 million, and 30% of 211 million is about 63 million. About 3.5% of Covid-19 patients require hospitalization; 3.5% of 63 million is 2,205,000. So over two million Brazilians could require hospitalization simultaneously. By comparison, the number of hospital beds in Brazil is about 410,000. One could also ask, if 30% of medical workers were infected with Covid-19 and others were afraid they could be, how many of these hospital beds would actually be available.

Actually, experience tells us that the reality would be somewhat less dire than that described in the previous paragraph. People react to a dramatic epidemic, even in the absence of government orders, by staying home, wearing masks, etc. Their behavior reduces the value of β, and hence the value of R_0. We will further discuss how humans respond to an epidemic in Chapter 5.

2.6 Changing population size

In this section we introduce the subject of epidemiological models with changing population size. The change could be due to any combination of births, natural deaths, deaths due to the disease being modeled, immigration, or emigration.

The key issue is whether and how the adequate contact rate β changes as the population changes. We said in Section 1.5 that β depends on the mode of life of the population being modeled, and would differ in rural and urban areas. Kermack and McKendrick (1927) suggest that β should be related to population density.

With Kermack and McKendrick (ibid.)'s suggestion as motivation, we consider two situations.

1. "Urban Island." On an urban island a change in population would produce a corresponding change in population density, and might be expected to produce a corresponding change in the rate of human contacts. In this case we shall assume that β changes proportionally with population. Thus if $N(t)$ is population at time t and $\beta(t)$ is adequate contact rate at time t, then in the Urban Island case
$$\beta(t) = \frac{N(t)}{N(0)}\beta(0). \tag{2.11}$$

2. "Rural Settlement." In a rural settlement a population increase might result in an expansion of the settled area without a corresponding increase in population density or the rate of human contact. Thus in the rural settlement case we shall assume that β does not change as the population changes.

We shall illustrate these two situations by considering two epidemiological models with changing populations: (1) SIR for a disease that is sometimes fatal on an urban island, and (2) SIS in a rural settlement with an exponentially growing population.

2.6.1 SIR with fatalities on an urban island

In this section we consider the SIR model for a disease that is sometimes fatal on an urban island.

Consider a population of initial size $N(0)$. At time t let $s(t)$ denote the number of susceptibles, let $i(t)$ denote the number of infectives, let $w(t)$ denote the number of people who have recovered and are now immune, and let $d(t)$ denote the number who have died of the disease. Then the population at time t is $N(t) = s(t)+i(t)+w(t)$, and the adequate contact rate $\beta(t)$ is given by (2.11).

As in Sec. 1.5, the rate at which an infective person creates new infective persons at time t is $\beta(t)$ times the fraction of the population at time t that is susceptible, which is $\frac{s(t)}{N(t)}$. To get the rate at which new infective persons in the entire population are created at time t, we then multiply by $i(t)$. We obtain

$$\dot{s}(t) = -\beta(t)\frac{s(t)}{N(t)}i(t) = -\frac{N(t)}{N(0)}\beta(0)\frac{s(t)}{N(t)}i(t) = -\frac{\beta(0)}{N(0)}s(t)i(t).$$

We assume that infectives recover at the rate γi and die at the rate δi. (The relative values of γ and δ could depend on the availability of high-quality medical care.) Then the entire system of differential equations is

$$\dot{s} = -\frac{\beta(0)}{N(0)}si, \tag{2.12}$$

$$\dot{i} = \frac{\beta(0)}{N(0)}si - \gamma i - \delta i, \tag{2.13}$$

$$\dot{w} = \gamma i, \tag{2.14}$$

$$\dot{d} = \delta i. \tag{2.15}$$

If we combine the w and d compartments into a single compartment $r = w + d$, called the removed compartment, we have

$$\dot{r} = \dot{w} + \dot{d} = \gamma i + \delta i = (\gamma + \delta)i.$$

The system becomes

$$\dot{s} = -\frac{\beta(0)}{N(0)}si, \tag{2.16}$$

$$\dot{i} = \frac{\beta(0)}{N(0)}si - (\gamma + \delta)i, \tag{2.17}$$

$$\dot{r} = (\gamma + \delta)i. \tag{2.18}$$

Finally, we note that $\dot{s} + \dot{i} + \dot{r} = 0$, so $s + i + r$ is always $N(0)$. We let S, I, and R denote population fractions: $S = \frac{s}{N(0)}$, $I = \frac{i}{N(0)}$, and $R = \frac{r}{N(0)}$. In (2.16)–(2.18) we make the substitutions

$$s = N(0)S,\ i = N(0)I,\ r = N(0)R,\ \dot{s} = N(0)\dot{S},\ \dot{i} = N(0)\dot{I},\ \dot{r} = N(0)\dot{R}.$$

After a little algebra we obtain

$$\dot{S} = -\beta(0)SI, \tag{2.19}$$

$$\dot{I} = \beta(0)SI - (\gamma + \delta)I, \tag{2.20}$$

$$\dot{R} = (\gamma + \delta)I. \tag{2.21}$$

This system is identical to the SIR system (2.1)–(2.3) if we set $\beta = \beta(0)$ and replace $\gamma + \delta$ by γ. For this reason the SIR system (2.1)–(2.3) can be used for a disease that is sometimes fatal, provided the R compartment is renamed "removed" and we keep in mind that the constants β and γ have new interpretations.

From our study of the SIR system we see that in the system (2.19)–(2.19), epidemics can occur if $R_0 = \frac{\beta(0)}{\gamma+\delta} > 0$.

2.6.2 SIS in a rural settlement with a growing population

We consider the SIS model in a rural settlement with adequate contact rate β, birth rate $\mu > 0$, and death rate $\nu > 0$. We assume $\mu > \nu$, so the population is growing exponentially: $\dot{N} = (\mu - \nu)N$, so $N(t) = e^{(\mu-\nu)t}N(0)$. Recall that in a rural settlement the adequate contact rate β does not grow as the population grows.

As we saw in Section 1.5, if we ignore births and deaths, the system of differential equations is (1.15)–(1.16). This system still correctly models the effects of the disease; however, in interpreting it, we must remember that $N = N(t)$, the population at time t, which is changing. To take into account births and deaths, we assume (1) all newborns are susceptible (not initially infected) and (2) both susceptibles and infectives have the same death rate ν. With these assumptions, three terms are added to (1.15)–(1.16), and we obtain

$$\dot{s} = -\frac{\beta}{N}si + \gamma i + \mu N - \nu s, \tag{2.22}$$

$$\dot{i} = \frac{\beta}{N}si - \gamma i - \nu i. \tag{2.23}$$

Next, we define population fractions $S = \frac{s}{N}$ and $I = \frac{i}{N}$. In (2.22)–(2.23) we make the substitutions $s = NS$ and $i = NI$. When we substitute for $\dot{s}$ and $\dot{i}$ we must remember that the population N is changing:

$$\dot{s} = N\dot{S} + \dot{N}S = N\dot{S} + (\mu - \nu)NS,$$

$$\dot{i} = N\dot{I} + \dot{N}I = N\dot{I} + (\mu - \nu)NI.$$

After making these substitutions we have

$$N\dot{S} + (\mu - \nu)NS = -\frac{\beta}{N}N^2SI + \gamma NI + \mu N - \nu NS,$$
$$N\dot{I} + (\mu - \nu)NI = \frac{\beta}{N}N^2SI - \gamma NI - \nu NI.$$

In each equation, we move the second term on the left hand side to the right hand side, cancel like terms, and divide by N. We obtain

$$\dot{S} = -\beta SI + \gamma I + \mu - \mu S,$$
$$\dot{I} = \beta SI - \gamma I - \mu I.$$

Finally, we recall that the population fractions S and I add up to 1, so that in the first equation, $\mu - \mu S = \mu(S + I) - \mu S = \mu I$. Therefore we have

$$\dot{S} = -\beta SI + \gamma I + \mu I = -\beta SI + (\gamma + \mu)I, \tag{2.24}$$
$$\dot{I} = \beta SI - \gamma I - \mu I = \beta SI - (\gamma + \mu)I. \tag{2.25}$$

Notice that the death rate ν does not appear in the simplified system (2.24)–(2.25). The reason is that both compartments have the same death rate, so deaths do not affect the population fractions in the compartments.

The system (2.24)–(2.25) is identical to the SIS system (1.1)–(1.2) if we replace $\gamma + \mu$ by γ. Therefore the dynamics of the two systems are the same.

In particular, if we reduce to one equation by setting $S = 1 - I$ in (2.25), we obtain

$$\dot{I} = (\beta - \gamma - \mu - \beta I)I. \tag{2.26}$$

There are equilibria at $I = 0$ and $I = 1 - \frac{\gamma+\mu}{\beta}$. If $\frac{\beta}{\gamma+\mu} < 1$, then the second equilibrium is negative, and all solutions in the biologically relevant region $0 \leqslant I \leqslant 1$ approach 0.

There are two points one should be careful of in this problem with a growing population.

1. In the previous paragraph we did not refer to $\frac{\beta}{\gamma+\mu}$ as R_0. That is because it cannot be interpreted as the number of new infections caused by a single infective individual when the entire population is susceptible, which remains $\frac{\beta}{\gamma}$.

2. If $I(t) \to 0$ as in the previous paragraph, one should be careful before concluding the the disease dies out. Since I is a fraction of a growing population, I can approach 0 while the number of infectives actually increases. We will return to this issue in Problem 3.3.

2.7 Discussion of the SIR model

We see from Figure 2.3 that

- The epidemic dies out when $R_0 < 1$.
- The epidemic can grow when $R_0 > 1$.

When $R_0 > 1$, the number of infectives grows until the susceptible population fraction has fallen to $\frac{\gamma}{\beta} = \frac{1}{R_0}$. At this time the remaining population fraction (recovered plus infective) is $1 - \frac{1}{R_0}$, and the number of infectives begins to fall. The population fraction $1 - \frac{1}{R_0}$ is the *herd immunity* fraction. Once this population fraction is no longer susceptible to the disease, the disease begins to die out.

A vaccination program reduces the susceptible population and thereby helps to achieve herd immunity. For Covid-19, if we use the estimate $R_0 = 3$ from Section 2.5, the herd immunity fraction is about $1 - \frac{1}{3} = \frac{2}{3}$.

It is important to point out that the constant β, the adequate contact rate, is under human control. Let us look again at the definition of β in formula in Equation (1.14):

$$\beta = \frac{\text{contacted persons}}{\text{infective person} \cdot \text{day}} \cdot \text{probability of transmission.}$$

The first factor, the number of contacted persons per infective person per day, can be reduced if infective persons realize that they are infective (because they have symptoms) or that they may be infective (because they came into contact with an infective person), and then choose, or are required, to quarantine themselves. If infectives may not have symptoms, as with Covid-19, then they may not realize they are infective. In this case, in order to reduce the number of contacted persons per infected person per day, it may be necessary for the entire population to stay home as much as possible and maintain social distance when they do not. Recall that whether a contact is adequate can depend on the location, for example indoor vs. outdoor. Thus it may be necessary to close businesses, schools, places of

worship, etc., or institute various restrictions, in order to reduce the number of crowded indoor contacts.

The second factor, the probability that a susceptible contacted person is infected, might be reduced by mask-wearing by both possible infectives and susceptibles.

When γ remains constant, reducing β reduces R_0.

Figure 2.4 shows how the ultimate number of individuals infected in an epidemic can be reduced by reducing R_0. Figure 2.5 shows how the maximum number of people infected at one time, and hence the maximum stress on the health care system, can be reduced by reducing R_0.

We shall return to the issue of controlling Covid-19 in Chapter 4 using a more detailed model.

2.8 Problems

Problem 2.1 *Orbits of the SIR system*

For $R_0 > 1$, we have seen that the curve (2.10) that passes through the point $(S, I) = (S_-, 0)$ is $I = -(S - S_-) + \frac{\gamma}{\beta}(\ln S - \ln S_-)$.

1. Find the value of S where I attains its maximum, and obtain a formula for the value of I at the maximum, I_{max}.

2. Your formula gives the maximum population fraction infected at one time during the epidemic. It gives I_{max} as a function of β, γ and S_-. Assuming $S_- > \frac{\gamma}{\beta}$, show that:

   (a) $\frac{\partial I_{max}}{\partial S_-} > 0$. Does this make sense in epidemiological terms? Is it consistent with Figure 2.3?

   (b) $\frac{\partial I_{max}}{\partial \beta} >$ Does this make sense in epidemiological terms?

   (c) $\frac{\partial I_{max}}{\partial \gamma} < 0$. Does this make sense in epidemiological terms?

Problem 2.2 *Practice with phase portraits I*

Consider the system

$$\dot{x} = -x, \tag{2.27}$$

$$\dot{y} = -2y. \tag{2.28}$$

1. Find all equilibria.
2. Divide (2.28) by (2.27) to obtain a differential equation for invariant curves in the xy-plane. Find the general solution and simplify. (Answer: $y = Cx^2$.)
3. Explain why the line $x = 0$ is also invariant.
4. Use your answers to parts 1–3 to draw the phase portrait. Refer back to the differential equation for the direction of the arrows.
5. How many orbits does each invariant curve contain?

Problem 2.3 *Practice with phase portraits II*

Consider the system

$$\dot{x} = -\alpha x - \omega y, \tag{2.29}$$
$$\dot{y} = \omega x - \alpha y, \tag{2.30}$$

with $\alpha > 0$ and $\omega > 0$.

1. Find all equilibria.
2. Let $V(x, y) = x^2 + y^2$. Show that $\dot{V} = -2\alpha V$. It follows that $V(t) = e^{-2\alpha t} V(0)$, which implies that all solutions approach the origin.
3. Let $\theta(x, y) = \arctan \frac{y}{x}$, the usual angular coordinate. Show that $\dot{\theta} = \omega$. Since $\omega > 0$, it follows that all solutions circle counterclockwise around the origin. (Actually $\arctan \frac{y}{x}$ is only defined in the right half plane; you may ignore this technicality.)
4. Parts 2 and 3 imply that all solutions spiral counterclockwise around the origin while approaching the origin. Use this information to sketch the phase portrait.

3 SIR model with loss of immunity

3.1 The model

In this chapter we consider an infectious disease for which immunity can be lost over time. The model we shall use, which is sometimes called the SIRS model, is the SIR model with two additional terms:

$$\dot{S} = -\beta SI + \rho R, \tag{3.1}$$

$$\dot{I} = \beta SI - \gamma I, \tag{3.2}$$

$$\dot{R} = \gamma I - \rho R. \tag{3.3}$$

The new terms indicate that individuals transfer from the recovered compartment, where they are immune to the disease, to the susceptible compartment at a rate proportional to R. The meaning of the proportionality constant ρ is that the average length of time before immunity is lost is $1/\rho$. This model is not yet used for the Covid-19 pandemic because no estimate for $1/\rho$ is known.

You can check that $\dot{S} + \dot{I} + \dot{R} = 0$, so if $S + I + R = 1$ initially, then $S + I + R$ remains 1 always. Thus we do not need all three equations, since if we can calculate $S(t)$ and $I(t)$, we can find $R(t)$ from $R(t) = 1 - S(t) - I(t)$. We

shall therefore use (3.1) and (3.2) only, after the substitution $R = 1 - S - I$:

$$\dot{S} = -\beta SI + \rho(1 - S - I), \tag{3.4}$$

$$\dot{I} = \beta SI - \gamma I. \tag{3.5}$$

We are only interested in the triangle $\mathcal{T}$ defined in the previous chapter, which is our phase space.

As with the SIS and SIR models,

$$R_0 = \beta \cdot \frac{1}{\gamma} = \frac{\beta}{\gamma} \quad \text{and} \quad R_e = \frac{\beta S}{\gamma}.$$

Moreover, as with the SIS and SIR models, in the region $I > 0$,

- $R_e < 1$ if and only if $\dot{I} < 0$.
- $R_e = 1$ if and only if $\dot{I} = 0$.
- $R_e > 1$ if and only if $\dot{I} > 0$.

3.2 Phase portraits

We first find the nullclines:

- $\dot{S} = 0$ when $I = \frac{\rho - \rho S}{\rho + \beta S}$ (a pair of hyperbolas).
- $\dot{I} = 0$ when $I = 0$ or $S = \frac{\gamma}{\beta}$ (two lines).

You learned to graph the pair of hyperbolas in calculus using derivatives, asymptotes, and intercepts.

Because we are only interested in the triangle $\mathcal{T}$, there are two cases, $\frac{\gamma}{\beta} > 1$ (i.e. $R_0 < 1$) and $\frac{\gamma}{\beta} < 1$ (i.e. $R_0 > 1$). Figure 3.1 shows the configuration of the nullclines in the entire SI-plane in the second case.

There are equilibria where $\dot{S} = \dot{I} = 0$, i.e., where the $\dot{S}$ and $\dot{I}$ nullclines intersect. Thus there are two equilibria, $(S, I) = (1, 0)$ and

$$(S, I) = (S_*, I_*) = \left(\frac{\gamma}{\beta}, \frac{\rho(\beta - \gamma)}{\beta(\rho + \gamma)} \right). \tag{3.6}$$

The second is in $\mathcal{T}$ only when $R_0 > 1$, i.e. $\frac{\gamma}{\beta} < 1$.

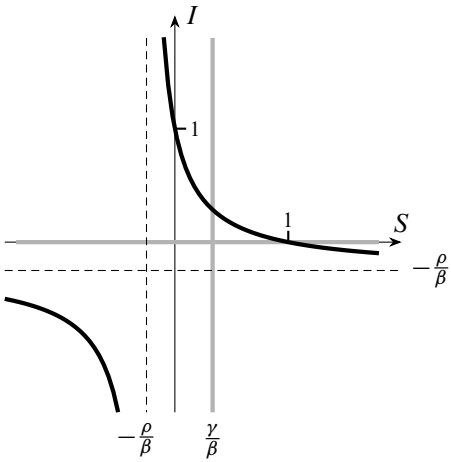


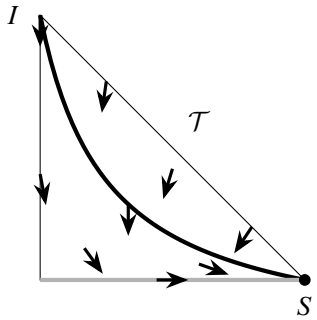
Figure 3.1: Nullclines of the system (3.4)–(3.5) in the case $\frac{\gamma}{\beta} < 1$ (i.e. $R_0 > 1$). The $\dot{S} = 0$ nullclines are black and the $\dot{I} = 0$ nullclines are gray. In the case $\frac{\gamma}{\beta} > 1$ (i.e. $R_0 < 1$) the vertical gray nullcline would be to the right of $S = 1$.

Figure 3.2 shows the nullclines, equilibria, and vector field in the triangle $\mathcal{T}$ in the two cases $R_0 < 1$ and $R_0 > 1$.

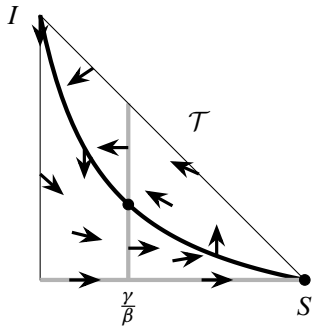
In the case $R_0 < 1$, all solutions in $\mathcal{T}$ with $I > 0$ have I decreasing everywhere. It appears that they all approach the equilibrium $(S, I) = (1, 0)$. We will see shortly that this is correct. At this equilibrium the entire population is susceptible to the disease (because of loss of immunity) but no one is infected; the disease is no longer present in the population.

The case $R_0 > 1$ is less straightforward. Solutions appear to circle around the equilibrium (S_*, I_*), but we cannot tell if they spiral toward (S_*, I_*), spiral away from (S_*, I_*), or join back up with themselves to form periodic solutions. It also appears possible that some solutions could join up with themselves to form periodic solutions, and others could spiral toward or away from those periodic solutions.

We can get more information about our system using *linearization*. Differential calculus is based on the idea that it is helpful to approximate a nonlinear function by a linear function. Similarly, it is sometimes helpful to approximate a nonlinear differential equation by a linear one. Linearization can tell us whether solutions near an equilibrium head toward it or away from it.



(a) $R_0 < 1$, $\beta = 0.6$, $\gamma = 1.2$, $\rho = 0.15$.



(b) $R_0 > 1$, $\beta = 0.6$, $\gamma = 0.2$, $\rho = 0.15$.

Figure 3.2: Nullclines, equilibria, and vector field for (3.4)–(3.5) in the triangle $\mathcal{T}$. The $\dot{S} = 0$ nullclines are black and the $\dot{I} = 0$ nullclines are gray.

3.3 Differential equations background: linear differential equations

A *linear differential equation* is a system of the form

$$\dot{x}_1 = a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n, \tag{3.7}$$
$$\dot{x}_2 = a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n, \tag{3.8}$$
$$\vdots$$
$$\dot{x}_n = a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n, \tag{3.9}$$

with all the a_{ij}'s constants.

Let

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \qquad \dot{x} = \begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \vdots \\ \dot{x}_n \end{pmatrix}, \qquad A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}. \tag{3.10}$$

Then the system (3.7)–(3.9) can be written as the single equation $\dot{x} = Ax$ (matrix product).

One fact we see immediately is that every linear differential equation has an equilibrium at $x = 0$.

In the case $n = 1$, (3.7)–(3.9) reduces to $\dot{x} = ax$, with $x \in \mathbb{R}$ and a a constant. The solution with $x(0) = x_0$ is $x = x_0 e^{at}$. If $a < 0$, every solution approaches the equilibrium at $x = 0$ as $t \to \infty$. If $a > 0$ and $x_0 \neq 0$, solutions grow as t increases, but they approach the equilibrium at $x = 0$ as $t \to -\infty$.

With this example in mind, it is reasonable to ask whether the linear differential equation $\dot{x} = Ax$ has any solutions of the form $x = x_0 e^{\lambda t}$. (Here x and x_0 are in $\mathbb{R}^n$, λ is a constant, and x_0 should be a nonzero vector to get an interesting result.) To answer this question, we substitute $x = x_0 e^{\lambda t}$ into both sides of $\dot{x} = Ax$ and obtain

$$\lambda e^{\lambda t} x_0 = A e^{\lambda t} x_0 \quad \text{or} \quad \lambda x_0 = A x_0 \quad \text{or} \quad (A - \lambda I) x_0 = 0.$$

Here I is the $n \times n$ identity matrix.

$$I = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix},$$

which has the property $Ix = x$ for any $x \in \mathbb{R}^n$.

The equation $(A - \lambda I)x_0 = 0$ has solutions other than $x_0 = 0$ if and only if the determinant of the matrix $A - \lambda I$ is 0, i.e., $\det(A - \lambda I) = 0$. The numbers λ such that $\det(A - \lambda I) = 0$ are called *eigenvalues* of A. If λ is an eigenvalue of A, the nonzero vectors x_0 such that $(A - \lambda I)x_0 = 0$ are called *eigenvectors* for the eigenvalue λ. The set of all solutions of $(A - \lambda I)x_0 = 0$, including $x_0 = 0$, is a subspace of $\mathbb{R}^n$ called the *eigenspace* for the eigenvalue λ. Eigenvalues and eigenvectors may be complex.

The equation $\det(A - \lambda I) = 0$ turns out to be a polynomial equation of degree n (the *characteristic equation* of A), so there are exactly n eigenvalues, counting multiplicity.

Example. Consider the linear system

$$\dot{x}_1 = x_2, \tag{3.11}$$
$$\dot{x}_2 = x_1. \tag{3.12}$$

Written as a matrix equation, it is

$$\begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$$

The characteristic equation is

$$\det\left(\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix} - \lambda\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\right) = \det\begin{pmatrix}-\lambda & 1\\ 1 & -\lambda\end{pmatrix} = \lambda^2 - 1 = 0.$$

Therefore the eigenvalues are $\lambda = -1$ and $\lambda = 1$.

To find eigenvectors for the eigenvalue $\lambda = -1$, we look for solutions to the equation $\big(A - (-1)I\big)x_0 = 0$, with $A = \begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}$:

$$\left(\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix} - (-1)\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\right)\begin{pmatrix}x_1\\ x_2\end{pmatrix} = \begin{pmatrix}0\\ 0\end{pmatrix} \quad \text{or} \quad \begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix}\begin{pmatrix}x_1\\ x_2\end{pmatrix} = \begin{pmatrix}0\\ 0\end{pmatrix}.$$

The solutions of this equation are all multiples of the vector $\begin{pmatrix}-1\\ 1\end{pmatrix}$. They constitute a line, i.e., a 1-dimensional subspace of $\mathbb{R}^2$. This line is the eigenspace for the eigenvalue $\lambda = -1$. If $x(0) = x_0$ is a point on this line (a multiple of $\begin{pmatrix}1\\ -1\end{pmatrix}$), then $x(t) = e^{-t}x_0$. This formula implies that $x(t)$ is always a point on the line, $x(t) \to 0$ as $t \to \infty$, and, if $x_0 \neq 0$, $\|x(t)\| \to \infty$ as $t \to -\infty$.

Similarly, for the eigenvalue $\lambda = 1$, the eigenspace is all multiples of the vector $\begin{pmatrix}1\\ 1\end{pmatrix}$. It is again a line. If $x(0) = x_0$ is a point on this line, then $x(t) = e^t x_0$. This formula implies that $x(t)$ is always a point on the line, $x(t) \to 0$ as $t \to -\infty$, and, if $x_0 \neq 0$, $\|x(t)\| \to \infty$ as $t \to \infty$.

Using this information the phase portrait of the linear system (3.11)–(3.12) can be sketched; see Figure 3.3. The line $x_2 = -x_1$ is the eigenspace for the eigenvalue -1; on it the direction of movement is toward the origin. The line $x_2 = x_1$ is the eigenspace for the eigenvalue 1; on it the direction of movement is away from the origin. Other initial conditions can be regarded as a linear combination of $\begin{pmatrix}-1\\ 1\end{pmatrix}$ and $\begin{pmatrix}1\\ 1\end{pmatrix}$. As t increases, the component in the $\begin{pmatrix}-1\\ 1\end{pmatrix}$ direction decreases, while the component in the $\begin{pmatrix}1\\ 1\end{pmatrix}$ direction increases.

You could also try to sketch the phase portrait of (3.11)–(3.12) using nullclines, or by finding a formula for the invariant curves.

The linear differential equation $\dot{x} = Ax$ is called *hyperbolic* if all eigenvalues of A have nonzero real part. There are three types of hyperbolic linear differential equations.

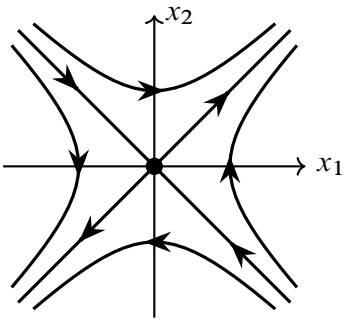


Figure 3.3: Phase portrait of the linear system (3.11)–(3.12).

1. All eigenvalues have negative real part: all solutions approach the origin as $t \to \infty$.

2. All eigenvalues have positive real part: all solutions approach the origin as $t \to -\infty$.

3. Counting multiplicity, k eigenvalues have negative real part and $n-k$ eigenvalues have positive real part. Then there are subspaces E^s of dimension k and E^u of dimension $n-k$ such that:

   - a solution $x(t)$ of $\dot{x} = Ax$ approaches the origin as $t \to \infty$ if and only if $x(0) \in E^s$;
   - a solution $x(t)$ of $\dot{x} = Ax$ approaches the origin as $t \to -\infty$ if and only if $x(0) \in E^u$.

E^s and E^u are called the *stable subspace* and the *unstable subspace* respectively of $\dot{x} = Ax$.

The linear system (3.11)–(3.12) is an example of the third type. The subspaces E^s and E^u both have dimension one.

We will not discuss nonhyperbolic linear differential equations.

Even in just two dimensions, linear differential equations of the first type (all eigenvalues have negative real part) come in several flavors. We will look at examples of the two most common flavors: (a) two different negative real eigenvalues and (b) complex eigenvalues with negative real part.

Linear differential equations in two dimensions of the second type (all eigenvalues have positive real part) have two analogous flavors: (a) two different pos-

itive real eigenvalues and (b) complex eigenvalues with positive real part. Their phase portraits look like those we will draw below with the arrows reversed.

3.3.1 Two different negative real eigenvalues

As a simple example, we consider the system

$$\begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ 0 & -2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}. \tag{3.13}$$

The eigenvalues are -1 and -2. You drew the phase portrait of this system in Problem 2.2. See Figure 3.4.

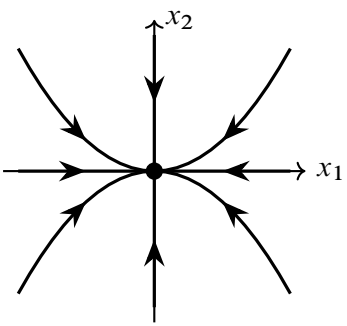


Figure 3.4: Phase portrait of the linear system (3.13).

All the invariant curves you found, except the x_2-axis, are tangent to the x_1-axis. To understand this, notice that the x_2-axis is the eigenspace for the eigenvalue -2, and the x_1-axis is the eigenspace for the eigenvalue -1. Since $-2 < -1 < 0$, the x_2-coordinate of any solution goes to 0 faster than the x_1-coordinate (like e^{-2t} instead of e^{-t}), so solutions end up tangent to the x_1-axis (except for solutions along the x_2-axis, which have no x_1-coordinate).

In general, in two dimensions, if the eigenvalues are $-\lambda_2 < -\lambda_1 < 0$, then almost all solutions approach the origin tangent to the eigenspace for the eigenvalue $-\lambda_1$.

3.3.2 Complex eigenvalues with negative real part

As a simple example, we consider the system

$$\begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \end{pmatrix} = \begin{pmatrix} -\alpha & -\omega \\ \omega & -\alpha \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}. \tag{3.14}$$

with $\alpha > 0$ and $\omega > 0$. The eigenvalues are $-\alpha \pm \omega i$. You drew the phase portrait in Problem 2.3. See Figure 3.5.

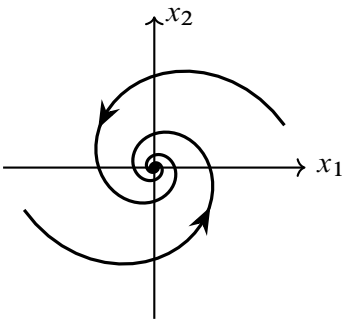


Figure 3.5: Phase portrait of the linear system (3.14).

In general, in two dimensions, if the eigenvalues are $-\alpha \pm \omega i$ with $\alpha > 0$ and $\omega > 0$, then solutions approach the origin while spiraling either clockwise or counterclockwise.

3.4 Differential equations background: asymptotic stability and linearization

An equilibrium x_0 of $\dot{x} = f(x)$ is *asymptotically stable* if solutions that start near x_0 stay near x_0 in future time, and in addition approach x_0 as $t \to \infty$.

Asymptotically stable equilibria are states that one expects to observe persisting in the natural world. If some perturbation takes the state of the system a small distance away from an asymptotically stable equilibrium, the state returns to the equilibrium.

Suppose $\dot{x} = f(x)$ has an equilibrium at x_0. To study solutions near x_0, we make the substitution $x = x_0 + y$. Then small y corresponds to x near x_0. We obtain $\dot{y} = f(x_0 + y)$. By Taylor's Theorem

$$\dot{y} = f(x_0) + Df(x_0)y + \ldots = Df(x_0)y + \ldots$$

because x_0 is an equilibrium. Here $Df(x_0)$ is the $n \times n$ matrix whose ij-entry is $\frac{\partial f_i}{\partial x_j}$ evaluated at the point x_0.

The *linearization* of the differential equation $\dot{x} = f(x)$ at the equilibrium x_0 is the linear differential equation $\dot{y} = Df(x_0)y$. We can determine the phase portrait of $\dot{y} = Df(x_0)y$ by finding eigenvalues and eigenvectors.

The equilibrium x_0 of $\dot{x} = f(x)$ is called *hyperbolic* if the linear differential equation $\dot{y} = Df(x_0)y$ is hyperbolic.

To state the following theorem, we need the notion of a *manifold*. A k-dimensional manifold in $\mathbb{R}^n$ is a subset of $\mathbb{R}^n$ that, near each of its points, looks like a k-dimensional subspace of $\mathbb{R}^n$. A 1-dimensional manifold is just a curve, and a 2-dimensional manifold is just a surface.

Theorem 3.1 (Linearization Theorem)**.** *If x_0 is a hyperbolic equilibrium of $\dot{x} = f(x)$, then the phase portrait of $\dot{x} = f(x)$* near *x_0 looks just like the phase portrait of $\dot{y} = Df(x_0)y$ near the origin. In particular:*

- *If all eigenvalues of $Df(x_0)$ have negative real part, then x_0 is an asymptotically stable equilibrium of $\dot{x} = f(x)$. The equilibrium x_0 is called an* attractor.

- *If all eigenvalues of $Df(x_0)$ have positive real part, then x_0 is an asymptotically stable equilibrium of $\dot{x} = -f(x)$. In other words, for $\dot{x} = f(x)$, all solutions that start near x_0 stay near x_0* in backward time*, and approach x_0 as $t \to -\infty$. The equilibrium x_0 is called a* repeller.

- *If $Df(x_0)$ has k eigenvalues with negative real part and $n-k$ eigenvalues with positive real part ($0 < k < n$), then there are invariant manifolds $W^s(x_0)$ of dimension k and $W^u(x_0)$ of dimension $n-k$ through x_0 such that:*

  - *$W^s(x_0)$ contains all solutions that approach x_0 as $t \to \infty$;*
  - *$W^u(x_0)$ contains all solutions that approach x_0 as $t \to -\infty$.*

  The equilibrium x_0 is called a saddle*. $W^s(x_0)$ and $W^u(x_0)$ are called the* stable manifold *of x_0 and the* unstable manifold *of x_0 respectively. At x_0, $W^s(x_0)$ and $W^u(x_0)$ are tangent respectively to the stable and unstable subspaces of $\dot{y} = Df(x_0)y$, translated to x_0.*

See Figure 3.6.

A particularly simple case is the SIS model $\dot{I} = f(I) = (\beta - \gamma)I - \beta I^2$, for which the dimension is one. We have seen that there are equilibria at $I = 0$ and $I = 1 - \frac{\gamma}{\beta}$. We have $f'(I) = \beta - \gamma - 2\beta I$. You can check that $f'(0) = \beta - \gamma$

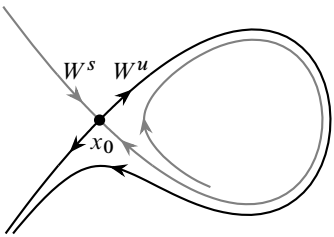


Figure 3.6: Stable manifold of x_0 in gray and unstable manifold of x_0 in black.

and $f'(1-\frac{\gamma}{\beta}) = -(\beta-\gamma)$. Thus we see that in the case $R_0 = \frac{\beta}{\gamma} > 1$, the equilibrium at $I = 0$ is repeller, and the one at $I = 1-\frac{\gamma}{\beta}$ is an attractor. In the case $R_0 = \frac{\beta}{\gamma} < 1$, the equilibrium at $I = 0$ is an attractor, and the one at $I = 1-\frac{\gamma}{\beta}$ is a repeller (but it is not in the interval $0 \leqslant I \leqslant 1$). Compare Figure 1.2

On the other hand, Theorem 3.1 does not apply to the SIR model (2.4)–(2.5), because the equilibria are not hyperbolic. You will check this in the problems.

3.5 Equilibria of the SIR model with loss of immunity

For the SIR model with loss of immunity (3.4) –(3.5), the *linearization matrix* is

$$\begin{pmatrix} \frac{\partial \dot{S}}{\partial S} & \frac{\partial \dot{S}}{\partial I} \\ \frac{\partial \dot{I}}{\partial S} & \frac{\partial \dot{I}}{\partial I} \end{pmatrix} = \begin{pmatrix} -\beta I - \rho & -\beta S - \rho \\ \beta I & \beta S - \gamma \end{pmatrix}. \tag{3.15}$$

We saw that there are equilibria at $(S, I) = (1, 0)$ and, when $R_0 > 1$, at $(S, I) = (S_*, I_*) = \left(\frac{\gamma}{\beta}, \frac{\rho(\beta-\gamma)}{\beta(\rho+\gamma)}\right)$. At $(S, I) = (1, 0)$, you can check by substituting into (3.15) that

$$\begin{pmatrix} \frac{\partial \dot{S}}{\partial S} & \frac{\partial \dot{S}}{\partial I} \\ \frac{\partial \dot{I}}{\partial S} & \frac{\partial \dot{I}}{\partial I} \end{pmatrix} = \begin{pmatrix} -\rho & -\beta - \rho \\ 0 & \beta - \gamma \end{pmatrix}. \tag{3.16}$$

You can easily check that the eigenvalues of (3.16) are $-\rho$ and $\beta - \gamma$. The first is negative. The second is negative when $R_0 = \frac{\beta}{\gamma} < 1$ and positive when $R_0 > 1$. Thus the equilibrium at $(1, 0)$ is an attractor when $R_0 < 1$ and a saddle when $R_0 > 1$. In particular:

Proposition 3.1. *For $R_0 < 1$, solutions of (3.4)–(3.5) that start near $(1,0)$ approach $(1,0)$ as $t \to \infty$.*

At $(S, I) = (S_*, I_*)$, you will find with a little more work that

$$\begin{pmatrix} \frac{\partial \dot{S}}{\partial S} & \frac{\partial \dot{S}}{\partial I} \\ \frac{\partial \dot{I}}{\partial S} & \frac{\partial \dot{I}}{\partial I} \end{pmatrix} = \begin{pmatrix} -\rho\frac{\beta-\gamma}{\rho+\gamma} - \rho & -\gamma - \rho \\ \rho\frac{\beta-\gamma}{\rho+\gamma} & 0 \end{pmatrix}. \tag{3.17}$$

You could try to find the eigenvalues of this matrix, but there is another approach that is a little easier.

You can check that the eigenvalues of

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \tag{3.18}$$

are the solutions $\lambda = \lambda_1$ and $\lambda = \lambda_2$ of the equation

$$\lambda^2 - (a_{11} + a_{22})\lambda + (a_{11}a_{22} - a_{12}a_{21}) = 0. \tag{3.19}$$

Therefore

$$\begin{aligned} \lambda^2 - (a_{11} + a_{22})\lambda + (a_{11}a_{22} - a_{12}a_{21}) &= (\lambda - \lambda_1)(\lambda - \lambda_2) \\ &= \lambda^2 - (\lambda_1 + \lambda_2)\lambda + \lambda_1\lambda_2. \end{aligned}$$

Thus we have

$$a_{11} + a_{22} = \lambda_1 + \lambda_2 \text{ and } a_{11}a_{22} - a_{12}a_{21} = \lambda_1\lambda_2. \tag{3.20}$$

The expression $a_{11} + a_{22}$ is the *trace* of the matrix (3.18), and the expression $a_{11}a_{22} - a_{12}a_{21}$ is its determinant.

Formula (3.20) helps us determine the signs of λ_1 and λ_2 if we can find the signs of $a_{11} + a_{22}$ and $a_{11}a_{22} - a_{12}a_{21}$. For example,

- If $a_{11}a_{22} - a_{12}a_{21} < 0$, then one of λ_1, λ_2 is positive and one is negative.
- If $a_{11}a_{22} - a_{12}a_{21} = 0$, then at least one of λ_1, λ_2 is 0.
- If $a_{11}a_{22} - a_{12}a_{21} > 0$ and $a_{11} + a_{22} < 0$, then either λ_1, λ_2 are both negative, or λ_1, λ_2 are a complex conjugate pair $\alpha \pm \beta i$ with $\alpha < 0$.

For the matrix (3.17), you can easily check that

$$a_{11} + a_{22} = -\rho\frac{\beta - \gamma}{\rho + \gamma} - \rho \text{ and } a_{11}a_{22} - a_{12}a_{21} = \rho(\beta - \gamma).$$

If $R_0 = \frac{\beta}{\gamma} > 1$, we see immediately that

$$a_{11} + a_{22} < 0 \text{ and } a_{11}a_{22} - a_{12}a_{21} > 0.$$

Therefore either λ_1, λ_2 are both negative, or λ_1, λ_2 are a complex conjugate pair $\alpha \pm \beta i$ with $\alpha < 0$. In either case the equilibrium (S_*, I_*) is an attractor.

We conclude:

Proposition 3.2. *For $R_0 > 1$, solutions of (3.4)–(3.5) that start near (S_*, I_*) approach (S_*, I_*) as $t \to \infty$.*

Proposition 3.1 and Proposition 3.2 do not give the whole truth; more solutions approach these equilibria than just those that start near them. To show this we first need some background about differential equations in the plane.

3.6 Differential equations background: planar theory

Recall that for a differential equation in one dimension, solutions that stay bounded approach equilibria. The situation is a little more complicated in two dimensions. To explain the situation in two dimensions, we need a few concepts.

If a solution of a differential equation is periodic in time, the corresponding orbit is called *closed* because it is always a simple closed curve. For example, in Section 1.3 we saw that the differential equation $\dot{x}_1 = -x_2$, $\dot{x}_2 = x_1$ has the solution $(\cos t, \sin t)$, which has period 2π. The corresponding orbit, shown in Figure 1.1, is the circle $x^2 + y^2 = 1$, which is a simple closed curve.

Theorem 3.2 (Poincaré–Bendixson Theorem)**.** *Let $\dot{x} = f(x)$ be a differential equation with $n = 2$ that is defined on an open set U. Then a solution of $\dot{x} = f(x)$ that stays in a compact (closed and bounded) subset of U as t increases approaches either*

- *a set that contains an equilibrium; or*
- *a closed orbit.*

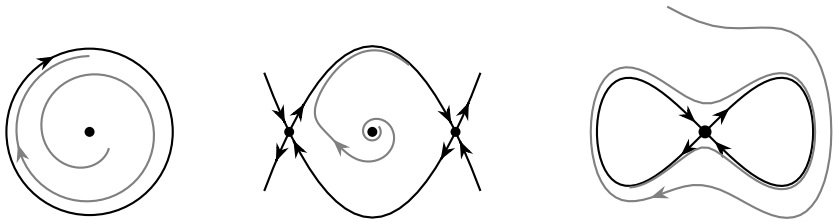


Figure 3.7: (a) A solution approaches a closed orbit. (b) A solution approaches a separatrix cycle that consists of two nontrivial orbits and two equilibria. (c) A solution approaches a graphic that consists of two separatrix cycles. Each separatrix cycle consists of a single nontrivial orbit and a single equilibrium.

A differential equation $\dot{x} = f(x)$ is called *polynomial* if, when we write it as a system

$$\begin{aligned} \dot{x}_1 &= f_1(x_1, \ldots, x_n), \\ &\vdots \\ \dot{x}_n &= f_n(x_1, \ldots, x_n), \end{aligned}$$

all the function $f_1, \ldots, f_n$ are polynomials. The differential equations we encounter in this course are all polynomial.

For polynomial differential equations with $n = 2$, we can give more detail than is in Theorem 3.2.

It can happen that one or more orbits that connect equilibria, together with the equilibria, form a simple closed curve. If the differential equation gives a consistent direction of motion around the closed curve, it is called a *separatrix cycle*. A *graphic* is a connected union of two or more separatrix cycles. See Figure 3.7. Don't worry too much about these notions. We mostly need them to state the following theorem.

Theorem 3.3 (Generalized Poincaré–Bendixson Theorem)**.** *Let $\dot{x} = f(x)$ be a polynomial differential equation with $n = 2$ that has only isolated equilibria. Then a solution of $\dot{x} = f(x)$ that stays in a compact (closed and bounded) set as t increases approaches either*

- *an equilibrium;*

- *a closed orbit;*
- *a separatrix cycle; or*
- *a graphic.*

For both versions of the Poincaré–Bendixson Theorem, see Perko (2001), p. 245.

The existence of separatrix cycles and graphics is often easy to rule out. In addition, one can sometimes rule out the existence of closed orbits using Bendixson's Criterion or Dulac's Criterion, which we shall discuss shortly. Once it is known that there are no separatrix cycles or closed orbits, a solution that stays in a compact set must approach an equilibrium.

Bendixson's Criterion and Dulac's Criterion are based on the 2D Divergence Theorem, a version of Green's Theorem, which you learned in calculus.

Suppose

$$\begin{aligned} \dot{x}_1 &= f_1(x_1, x_2), \\ \dot{x}_2 &= f_2(x_1, x_2) \end{aligned}$$

is defined on an open set U in the plane. We write $\dot{x} = f(x)$ for short. Let C be a simple closed curve in U, oriented counterclockwise. At each x on C, let $n(x)$ be the unit outward-pointing normal vector.

The *divergence* of f, $\nabla \cdot f$, is defined by

$$\nabla \cdot f(x_1, x_2) = \frac{\partial f_1}{\partial x_1}(x_1, x_2) + \frac{\partial f_2}{\partial x_2}(x_1, x_2).$$

Theorem 3.4 (2D Divergence Theorem)**.** *Suppose C and its interior,* IntC*, are contained in U. Then*

$$\iint_{\text{IntC}} \nabla \cdot f(x)\, dA = \int_C f(x) \cdot n(x)\, ds.$$

The first integral is an ordinary double integral of a function over a region in the plane. The second integral is the integral of a function around a curve with respect to arc length.

Corollary 3.1 (Bendixson's Criterion)**.** *Assume (1) the open set U has no holes, and (2) $\nabla \cdot f$ is always positive on U, or $\nabla \cdot f$ is always negative on U. Then $\dot{x} = f(x)$ has no closed orbits in U.*

Proof. If $\dot{x} = f(x)$ had a closed orbit C in U, then by the 2D Divergence Theorem we would have

$$\iint_{\text{IntC}} \nabla \cdot f(x)\, dA = \int_C f(x) \cdot n(x)\, ds = \int_C 0\, ds = 0 \qquad (3.21)$$

since $f(x)$ is tangent to C and $n(x)$ is perpendicular to C. However, if $\nabla \cdot f$ is always positive, then the double integral is positive; and if $\nabla \cdot f$ is always negative, then the double integral is negative. This is a contradiction. □

Remark 3.5. A similar argument shows that under the same assumptions, there are no separatrix cycles or graphics in U. See Perko (2001), p. 266, Problem 1.

Corollary 3.2 (Dulac's Criterion)**.** *Assume (1) the open set U has no holes, and (2) there is a positive function $g(x)$ such that $\nabla \cdot gf$ is always positive on U, or $\nabla \cdot gf$ is always negative on U. Then $\dot{x} = f(x)$ has no closed orbits in U.*

Proof. By Bendixson's Criterion, $\dot{x} = g(x)f(x)$ has no closed orbits in U. But $\dot{x} = f(x)$ and $\dot{x} = g(x)f(x)$ have the same orbits. (The vectors $f(x)$ and $g(x)f(x)$ point in the same direction, they just have different lengths.) Therefore $\dot{x} = f(x)$ has no closed orbits in U. □

Remark 3.6. As with Bendixson's Criterion, a similar argument shows that under the same assumptions, there are no separatrix cycles or graphics in U.

3.7 Global stability of the SIR model with loss of immunity

Our triangle $\mathcal{T}$ is a closed and bounded subset of the plane. For the system (3.4)–(3.5), solutions that start in $\mathcal{T}$ stay in it. To see this one just has to check that solutions cannot escape through the boundary. Solutions that start on $I = 0$ stay on $I = 0$ and approach the equilibrium $(0, 1)$. You can check solutions that start on the other sides of $\mathcal{T}$ using the derivatives of $V(S, I) = S$ and $V(S, I) = S + I$ along solutions.

Dulac's Criterion (Corollary 3.2) and the remark that follows it can be used to show that for any values of the positive parameters α, β, and γ, the system (3.4)–(3.5) has no closed orbits, separatrix cycles or graphics in the open set $I > 0$. The

function $g(S, I) = \frac{1}{I}$ is positive in $I > 0$. If we multiply (3.4)–(3.5) by $\frac{1}{I}$, we obtain

$$\dot{S} = -\beta S + \rho \frac{1 - S - I}{I},$$
$$\dot{I} = \beta S - \gamma.$$

The divergence is

$$\frac{\partial \dot{S}}{\partial S} + \frac{\partial \dot{I}}{\partial I} = -\beta - \frac{\rho}{I}, \tag{3.22}$$

which is negative in $I > 0$. Then Dulac's Criterion (Corollary 3.2) and the remark that follows it imply that (3.4)–(3.5) has no closed orbits, separatrix cycles or graphics in the open set $I > 0$.

By Theorem 3.2, all solutions in $\mathcal{T}$ approach equilibria. More precisely:

1. For $R_0 < 1$, there are no equilibria in $\mathcal{T}$ except $(1, 0)$, so all solutions in $\mathcal{T}$ approach $(1, 0)$ as $t \to \infty$.

2. For $R_0 > 1$, the equilibrium $(S, I) = (1, 0)$ is a hyperbolic saddle. Its stable manifold is the line $I = 0$. Thus solutions that start in $I > 0$ cannot approach it as $t \to \infty$ (see the explanation of Theorem 3.1). The only other equilibrium in $\mathcal{T}$ is (S_*, I_*). Therefore all solutions in $\mathcal{T}$ with $I > 0$ approach the equilibrium (S_*, I_*) as $t \to \infty$.

3.8 Discussion of the SIR model with loss of immunity

The SIR model with loss of immunity has much in common with the SIS model of Chapter 1.

1. For $R_0 < 1$ there is a single equilibrium at which the entire population is susceptible and no one is infected. All solutions approach that equilibrium.

2. For $R_0 > 1$ that equilibrium is still present but is no longer asymptotically stable. A new equilibrium appears in the interior of phase space, at which the disease is present in the population at an intermediate level. The new equilibrium is asymptotically stable, and in fact almost all solutions approach it.

Models with these properties are called *endemic models*, and the interior equilibrium is called an endemic equilibrium. Endemic models often arise because of

some mechanism that replenishes the susceptible compartment, such as loss of immunity or, as we shall see in the problems, births. These processes are not important at the start of an epidemic but become important over the long haul.

By contrast, models such as the SIR model, in which there are many equilibria, are called *epidemic models*. They ignore long-term process such as loss of immunity and births. They are often used at the start of an epidemic of a new disease when length of immunity is not known and births are not important.

In the next chapter and its problems you will see more examples of both types of models.

Let us take another look at the endemic equilibrium for the SIR model with loss of immunity:

$$(S_*, I_*) = \left(\frac{\gamma}{\beta}, \frac{\rho(\beta-\gamma)}{\beta(\rho+\gamma)}\right).$$

You should recognize $\frac{\gamma}{\beta}$, the value of S_*. As we would expect from Section 1.8, it is the value of S at which $R_e = 1$. For the usual SIR model (2.4)–(2.5) with $R_0 > 1$, the line $S = \frac{\gamma}{\beta}$ separates the region where I is rising from the region where I is falling; see Figure 2.3. In the usual SIR model, herd immunity is reached when the susceptible population fraction declines to $\frac{\gamma}{\beta}$. As for I_*, the population fraction that is infected at the endemic equilibrium, notice that as $\rho \to 0$, $I_* \to 0$. The constant ρ is near 0 when the average length of time until immunity is lost is very large. Thus when the average length of time until immunity is lost is very large, the equilibrium (S_*, I_*) is close to $(\frac{\gamma}{\beta}, 0)$. For the usual SIR model (2.4)–(2.5), $(\frac{\gamma}{\beta}, 0)$ is the equilibrium that separates equilibria near which the epidemic grows from equilibria near which the epidemic dies; see Figure 2.3.

3.9 Problems

Problem 3.1 *Equilibria of the SIR system*

1. Compute the linearization matrix of the SIR system (2.4)–(2.5).
2. Evaluate your matrix at an equilibrium $(S, 0)$.
3. Find the eigenvalues of the matrix in part 2. You should get $\lambda = 0,\ \beta S - \gamma$.
4. Are any of the equilibria hyperbolic?

Problem 3.2 *SIR with births and natural deaths*

We consider the SIR model with births and natural deaths. (Natural deaths are those that are not caused by the disease being modeled.) The birth rate is μ, and all newborn individuals are susceptible. We will assume that the death rate is also μ and in each compartment the death rate is μ times the population fraction of that compartment. Thus we get the following system for the population fractions:

$$\dot{S} = \mu - \beta SI - \mu S, \tag{3.23}$$

$$\dot{I} = \beta SI - \gamma I - \mu I, \tag{3.24}$$

$$\dot{R} = \gamma I - \mu R. \tag{3.25}$$

1. Show that if $S + I + R = 1$, then $\dot{S} + \dot{I} + \dot{R} = 0$. Then explain the following statement: therefore by Theorem 2.1, if $S + I + R = 1$ initially, then $S + I + R$ remains 1.

2. Since $S + I + R = 1$ always, we only need the first two equations:

$$\dot{S} = \mu - \beta SI - \mu S, \tag{3.26}$$

$$\dot{I} = \beta SI - \gamma I - \mu I. \tag{3.27}$$

   Find the equilibria of (3.26)–(3.27). Answer: $(1, 0)$ and

$$(S_*, I_*) = \left(\frac{\gamma + \mu}{\beta}, \mu \left(\frac{1}{\gamma + \mu} - \frac{1}{\beta} \right) \right). \tag{3.28}$$

3. Show: If $\frac{\gamma+\mu}{\beta} < 1$, then (S_*, I_*) is in $\mathcal{T}$; if $\frac{\gamma+\mu}{\beta} > 1$, then (S_*, I_*) is not in $\mathcal{T}$.

4. Compute the linearization matrix of (3.26)–(3.27). Answer:

$$\begin{pmatrix} -\beta I - \mu & -\beta S \\ \beta I & \beta S - \gamma - \mu \end{pmatrix}. \tag{3.29}$$

5. Show that the eigenvalues of the linearization matrix at the equilibrium $(1, 0)$ are $-\mu$ and $\beta - \gamma - \mu$. Then explain the following: if $\frac{\gamma+\mu}{\beta} < 1$, so that (S_*, I_*) is in $\mathcal{T}$, then $(1, 0)$ is a saddle; if $\frac{\gamma+\mu}{\beta} > 1$, so that (S_*, I_*) is not in $\mathcal{T}$, then $(1, 0)$ is an attractor.

6. Show that if $\frac{\gamma+\mu}{\beta} < 1$, so that (S_*, I_*) is in $\mathcal{T}$, then (S_*, I_*) is an attractor. Hint: it is enough to show that for the matrix

$$\begin{pmatrix} -\beta I_* - \mu & -\beta S_* \\ \beta I_* & \beta S_* - \gamma - \mu \end{pmatrix}, \tag{3.30}$$

the trace is negative and the determinant is positive.

7. Use Dulac's Criterion (Corollary 3.2) and the remark that follows it as in Section 3.7 to show that for any values of the positive parameters β, γ, and μ, there are no closed orbits, separatrix cycles or graphics in $I > 0$. What conclusions can you draw?

Problem 3.3 *SIS with a growing population revisited*

Recall our discussion in Section 2.6.2 of the SIS model in a rural settlement with a growing population. We assumed birth rate $\mu > 0$ and death rate $\nu > 0$, so that the population at time t was $N(t) = N(0)e^{(\mu-\nu)t}$. We also assumed $\mu > \nu$, so the population is growing. The model reduced to

$$\dot{I} = f(I) = (\beta - \gamma - \mu - \beta I)I.$$

As in Section 2.6.2 we shall assume $\frac{\beta}{\gamma+\mu} < 1$.

1. Show that $f'(0) = \beta - \gamma - \mu$. Explain why this calculation shows that the equilibrium at 0 is an attractor.

2. From part 1, if $I(0)$ is near 0, the solution of $\dot{I} = f(I)$ is approximately $I(t) = I(0)e^{(\beta-\gamma-\mu)t}$. Since I is a population fraction, the actual number of infectives at time t is $i(t) = I(t)N(t)$. Under what condition is $i(t)$ increasing?

4 A Covid-19 model and the next generation matrix

In this chapter we explain the next generation matrix, which is used to calculate R_0 in more complicated compartmental models than those we have considered so far. The next generation matrix can be used to calculate R_e at any disease-free equilibrium. Two nice introductions to the next generation matrix are Blackwood and Childs (2018) and van den Driessche (2017). The main example we will consider is a compartmental model that captures the principal features of Covid-19.

4.1 The model

Our model for Covid-19 is based on the SEIR model, which adds an exposed compartment between the susceptible compartment and the infective compartment. Susceptibles who contract the disease first pass through the exposed compartment before entering the infective compartment. The SEIR model is used when individuals who contract a disease take a while before they develop symptoms and become infective.

We consider a modified version of the SEIR model that includes the most

salient features of Covid-19. Susceptibles who contract Covid-19 are asymptomatic and not infective for about 2.5 days. They then become infective for about 2.5 days before any of them develop symptoms. After this period about a third of Covid sufferers continue to be asymptomatic but remain infective; the other two thirds develop symptoms while remaining infective. The asymptomatic group ceases to be infective after about five days, the symptomatic group after about ten days. (Data from Ngonghala, Iboi, and Gumel (2020).)

For modeling purposes we shall assume that all individuals in both the asymptomatic and symptomatic groups enter the recovered compartment when they cease to be infective. We shall assume that individuals in the recovered compartment have permanent immunity; the actual average length of immunity for Covid-19 is not known.

More detailed Covid-19 models have additional compartments for hospitalized, ICU, and deceased patients. Even more detailed models divide each compartment by age, behavior, social position, or other factors.

The model we shall describe has six compartments:

- S: susceptible.
- E: exposed but not yet infective or symptomatic.
- C: infective but not yet symptomatic.
- I: infective and symptomatic.
- A: infective and asymptomatic.
- R: recovered.

See Figure 4.1.

The system of differential equations for the population fractions is:

$$\dot{S} = -\beta_C SC - \beta_I SI - \beta_A SA, \tag{4.1}$$

$$\dot{E} = \beta_C SC + \beta_I SI + \beta_A SA - \gamma_E E, \tag{4.2}$$

$$\dot{C} = \gamma_E E - \gamma_C C, \tag{4.3}$$

$$\dot{I} = p\gamma_C C - \gamma_I I, \tag{4.4}$$

$$\dot{A} = (1-p)\gamma_C C - \gamma_A A, \tag{4.5}$$

$$\dot{R} = \gamma_I I + \gamma_A A. \tag{4.6}$$

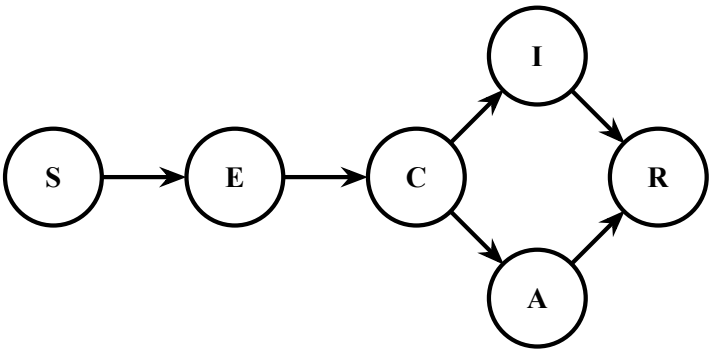


Figure 4.1: SECIAR Model.

The positive constants β_C, β_I, and β_A represent the rate of potential new infections caused directly by individuals in the compartments C, I, and A respectively. For Covid-19, β_C is known to be the largest of the three. β_I is smaller than might be expected since individuals in the I compartment know they are sick and are likely to stay home. The positive constants γ_E, γ_C, γ_I, γ_A are the rates at which individuals move out of various compartments. The constant p, $0 < p < 1$, is the probability that an individual in compartment C develops symptoms.

You can check that

$$\dot{S} + \dot{E} + \dot{C} + \dot{I} + \dot{A} + \dot{R} = 0.$$

Thus if $S + E + C + I + A + R = 1$ initially, then $S + E + C + I + A + R$ remains 1.

As in the earlier chapters we do not need the equation for R. In this chapter, however, we will retain it, since a reduction from six equations to five is not very helpful. The correct phase space for this model is

$$\{(S, E, C, I, A, R) : \\ S \geqslant 0,\ E \geqslant 0,\ C \geqslant 0,\ I \geqslant 0,\ A \geqslant 0,\ R \geqslant 0,\ S+E+C+I+A+R = 1\}.$$

However, we will ignore this and just focus on the six-dimensional system (4.1)–(4.6).

4.2 Equilibria of the Covid-19 model

To find equilibria, we begin by setting the last four equations equal to 0. We obtain a system of four linear equations in the four unknowns E, C, I, A. In matrix form, the system of linear equations is

$$\begin{pmatrix} \gamma_E & -\gamma_C & 0 & 0 \\ 0 & p\gamma_C & -\gamma_I & 0 \\ 0 & (1-p)\gamma_C & 0 & -\gamma_A \\ 0 & 0 & \gamma_I & \gamma_A \end{pmatrix} \begin{pmatrix} E \\ C \\ I \\ A \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

You can check that the determinant of the matrix is $\gamma_E\gamma_C\gamma_I\gamma_A$, which is positive. Therefore the only solution is $E = C = I = A = 0$.

Thus all equilibria have $E = C = I = A = 0$. These values also make the first two equations equal to zero. The set of equilibria of the six-dimensional system (4.1)–(4.6) is therefore

$$\{(S, E, C, I, A, R) : E = C = I = A = 0\}.$$

The infected states in this model are E, C, I and A. The equilibria are precisely the states in which the values of all infected variables are 0. Thus all equilibria are disease-free.

The linearization of the system (4.1)–(4.6) at an equilibrium $(S, 0, 0, 0, 0, R)$ has the matrix

$$\begin{pmatrix} 0 & 0 & -\beta_C S & -\beta_I S & -\beta_A S & 0 \\ 0 & -\gamma_E & \beta_C S & \beta_I S & \beta_A S & 0 \\ 0 & \gamma_E & -\gamma_C & 0 & 0 & 0 \\ 0 & 0 & p\gamma_C & -\gamma_I & 0 & 0 \\ 0 & 0 & (1-p)\gamma_C & 0 & -\gamma_A & 0 \\ 0 & 0 & 0 & \gamma_I & \gamma_A & 0 \end{pmatrix}. \tag{4.7}$$

To find the eigenvalues, you would subtract λ from the diagonal terms and take the determinant. A reasonable first step would be to expand by the first column, and a reasonable second step would be to expand by the last column. If you do this you will find that the eigenvalues of (4.7) are 0, 0, and the eigenvalues of the submatrix

$$K = \begin{pmatrix} -\gamma_E & \beta_C S & \beta_I S & \beta_A S \\ \gamma_E & -\gamma_C & 0 & 0 \\ 0 & p\gamma_C & -\gamma_I & 0 \\ 0 & (1-p)\gamma_C & 0 & -\gamma_A \end{pmatrix}. \tag{4.8}$$

4.3 Differential equations background: normally hyperbolic manifolds of equilibria

The SIR system (2.4)–(2.5) has a line of equilibria. We saw in Problem 1 of the previous section that they all have at least one 0 eigenvalue.

The system (4.1)–(4.6) has a plane of equilibria. We saw above that they all have at least two 0 eigenvalues.

Suppose the differential equation $\dot{x} = f(x)$ on $\mathbb{R}^n$ has a k-dimensional subspace $\mathcal{E}$ of equilibria. An equilibrium x_0 in $\mathcal{E}$ is said to be *normally hyperbolic* if the matrix $Df(x_0)$ has exactly k 0 eigenvalues, and the remaining $n-k$ eigenvalues all have nonzero real part.

Theorem 4.1. *Suppose the differential equation $\dot{x} = f(x)$ on $\mathbb{R}^n$ has a k-dimensional subspace $\mathcal{E}$ of equilibria. Let $\mathcal{E}_0$ be subset of $\mathcal{E}$ such that all equilibria in $\mathcal{E}_0$ are normally hyperbolic. Suppose that for all x_0 in $\mathcal{E}_0$, $Df(x_0)$ has ℓ eigenvalues with negative real part and m eigenvalues with positive real part, $k+\ell+m=n$. (ℓ or m may be zero.) Then*

- *Each x_0 in $\mathcal{E}_0$ has a stable manifold $W^s(x_0)$ of dimension ℓ.*
- *Each x_0 in $\mathcal{E}_0$ has an unstable manifold $W^u(x_0)$ of dimension m.*
- *The union of the manifolds $W^s(x_0)$ in a manifold called $W^s(\mathcal{E}_0)$ of dimension $k+\ell$.*
- *The union of the manifolds $W^u(x_0)$ in a manifold called $W^u(\mathcal{E}_0)$ of dimension $k+m$.*

The SIR system (2.4)–(2.5) provides an example of this theorem. We have $n=2$. There is a 1-dimensional space of equilibria $(S,0)$, so $k=1$. In Problem 3.1 you calculated that at $(S,0)$, one eigenvalue of the linearization matrix is 0, and the other is $\beta S-\gamma$. In the case $R_0=\frac{\beta}{\gamma}>1$, equilibria with $\frac{\gamma}{\beta}<S\leqslant 1$ have their second eigenvalue positive, so $\ell=0$ and $m=1$. Each equilibrium has a 1-dimensional unstable manifold, namely (a portion of) the curve (2.10) that passes through it. These curves fit together to form a manifold of dimension $k+m=2$, namely an open subset of the plane. Similarly, equilibria with $0\leqslant S<\frac{\gamma}{\beta}$ have their second eigenvalue negative, so $\ell=1$ and $m=0$. Each equilibrium has a 1-dimensional stable manifold. These curves fit together to form a manifold of dimension $k+\ell=2$, an open subset of the plane.

Recall from Section 2.2 that for the SIR model, the effective reproduction number is $R_e = \frac{\beta S}{\gamma}$. From this formula you can easily see that equilibria with $R_e > 1$ (those with $\frac{\gamma}{\beta} < S \leqslant 1$) have their second eigenvalue positive, and equilibria with $R_e < 1$ (those with $0 \leqslant S < \frac{\gamma}{\beta}$) have their second eigenvalue negative. This makes sense. As we mentioned in Section 1.8, at a disease-free equilibrium (all the equilibria of the SIR-model have $I = 0$), R_e should be interpreted as the the approximate number of people infected by a single infective individual *near* the equilibrium. It therefore makes sense that an equilibrium with $R_e > 1$ has an unstable manifold, and an equilibrium with $R_e < 1$ has a stable manifold.

The equilibrium $(S, I) = (\frac{\gamma}{\beta}, 0)$ is not normally hyperbolic since both eigenvalues are zero. Many different things can happen at equilibria that are not normally hyperbolic. This is an example of the Anna Karenina Principle in mathematics. The Russian novelist Tolstoy's 1877 novel *Anna Karenina* begins, "All happy families are alike; each unhappy family is unhappy in its own way." The reason is that in happy families, the spouses sufficiently agree about issues such as money, religion, child-raising, division of labor, cleanliness standards, and so on. Disagreement in any of these areas can occur in many ways and lead to an unhappy marriage. Similarly, all normally hyperbolic equilibria are alike (Theorem 4.1), but normal hyperbolicity can fail in many ways, resulting in many different types of behavior.

Theorem 4.1 also applies to differential equations that have a k-dimensional manifold of equilibria, but we do not need this level of generality.

4.4 Digression: estimating R_0 at the start of an epidemic

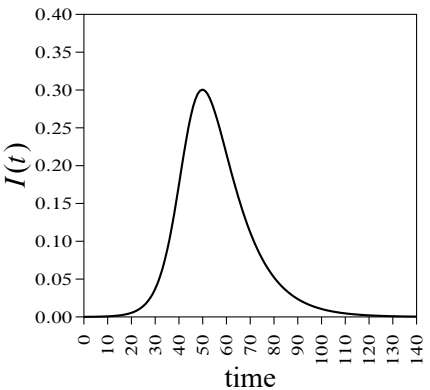
Problem 3.1 showed in particular that in the SIR model, at the equilibrium $(1, 0)$, the nonzero eigenvalue is $\beta - \gamma$. This implies that at the start of an epidemic in which the entire population is susceptible, I should increase and S should decrease (more precisely, $1 - S$ should increase) at the exponential rate $e^{(\beta-\gamma)t}$.

For $\beta = 0.3$ and $\gamma = 0.1$ ($R_0 = 3$), the orbits of the SIR system were graphed in Figure 2.3. The orbit through $(1, 0)$ has the equation

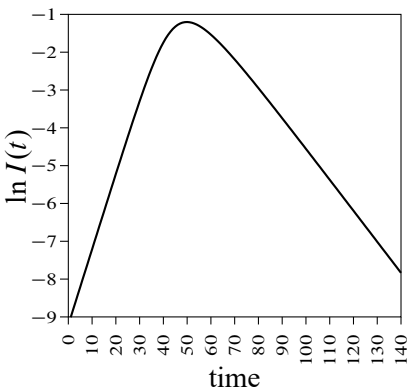
$$I = -(S - 1) + \frac{\gamma}{\beta} \ln S = -(S - 1) + \frac{1}{3} \ln S.$$

Figure 4.2 shows the graph of $I(t)$ for a solution that starts at a point on this orbit near $(1, 0)$. (The point has $S = 0.9999$.) $I(t)$ appears to grow exponentially

at first. This can be checked by plotting $\ln I(t)$; see Figure 4.2b. Initially the graph is approximately linear with upward slope. This is the sign of exponential growth. (The end of the graph is also approximately linear with downward slope. This is the sign of exponential decay to the final state.)



(a) Graph of $I(t)$.



(b) Graph of $\ln I(t)$.

Figure 4.2: SIR model with $\beta = 0.3$ and $\gamma = 0.1$.

The exponential rate of growth can be checked by finding the slope of the line. The portion of the graph in Figure 4.2b from $t = 0$ to $t = 30$ is approximately linear. Using the start and end points of this portion of the graph yields a slope of 0.198, which is approximately equal to $\beta - \gamma = 0.3 - 0.1 = 0.2$.

At the beginning of an epidemic, the number of cases is often observed to grow exponentially. For example, case data for the United Kingdom during the first month of the Covid-19 pandemic, and the natural logarithm of the case data, are shown in Stevens et al. (2020). The authors calculate an exponential growth rate for cumulative cases of 0.251. The rate of growth of cumulative cases is precisely the rate of growth of $1-S$, and so can be compared to the SIR model. We therefore obtain $\beta - \gamma = 0.251$.

Can this information be used to estimate R_0 for Covid-19 in the United Kingdom? Not immediately; since $R_0 = \frac{\beta}{\gamma}$, we need one more piece of information, for example β or γ. For example, if it was estimated at the time that the average duration of infectiousness with Covid-19 was ten days, then we would have $\gamma = 0.1$, so $\beta = 0.351$, and then $R_0 = \frac{\beta}{\gamma} = 3.51$.

Of course, this estimate uses the SIR model, which is not accurate for Covid-19, but has the advantage that it requires little information about the disease.

4.5 Eigenvalues of equilibria of the Covid-19 model

Theorem 4.1 tells us that the nature of the equilibria of the Covid-19 model is determined by their eigenvalues other than the two 0 eigenvalues we found above. Thus we should find the eigenvalues of the matrix K defined in (4.8).

When $S = 0$ (no one in the population is susceptible to the disease), you can easily check that the eigenvalues of K are $-\gamma_E$, $-\gamma_C$, $-\gamma_I$, and $-\gamma_A$. Thus an equilibrium with $S = 0$ has a 4-dimensional stable manifold.

As we increase S in the subspace of equilibria, the eigenvalues of K will change. They all have negative real part until one becomes 0, or a pair become pure imaginary. Let's guess that all eigenvalues of K have negative real part until one becomes 0. We will see later that this guess is correct.

The product of the eigenvalues of any matrix is the determinant of the matrix. (We saw this for 2×2 matrices in Section 3.5.) Therefore an eigenvalue of K becomes 0 if and only if the determinant of K also becomes 0.

After some algebra, one finds that the determinant of K is

$$\gamma_E\,\gamma_C\,\gamma_I\,\gamma_A - \gamma_E\left(\gamma_I\,\gamma_A\,\beta_C + p\,\gamma_C\,\gamma_A\,\beta_I + (1-p)\,\gamma_C\,\gamma_I\,\beta_A\right) S.$$

Thus $\det K$ has the form $\det K = a - bS$ where a and b are positive. Therefore $\det K$ is positive when $S < \frac{a}{b}$, $\det K = 0$ when $S = \frac{a}{b}$, and $\det K$ is negative when $S > \frac{a}{b}$.

Some more algebra tells us that $\det K = 0$ when

$$S = \frac{1}{\frac{\beta_C}{\gamma_C} + \frac{p\,\beta_I}{\gamma_I} + \frac{(1-p)\,\beta_A}{\gamma_A}}. \tag{4.9}$$

When S is less than this value, all four eigenvalues have negative real part. When S is greater than this value, the determinant becomes negative. One expects that this happens because three of the four eigenvalues are negative and one becomes positive. This is indeed what happens.

Based on the analogy of the SIR model, one might expect that the denominator of (4.9) is the basic reproduction number R_0 for the Covid-19 model, so that the value of S in (4.9) would be $\frac{1}{R_0}$. If $R_0 < 1$, then all equilibria with $S \leqslant 1$ are attracting, so an epidemic would die out. If $R_0 > 1$, equilibria with $\frac{1}{R_0} < S \leqslant 1$ have a positive eigenvalue, so epidemics can initially grow.

4.6 The next generation matrix

The next generation matrix is used to calculate the effective reproduction number R_e at disease-free equilibria for the compartmental models used in epidemiology. It is a very general method, but for concreteness we will explain it using our Covid-19 model.

The differential equations for the infected compartments are (4.2)–(4.5). Let us rewrite those four differential equations as follows:

$$\begin{pmatrix} \dot{E} \\ \dot{C} \\ \dot{I} \\ \dot{A} \end{pmatrix} = \mathcal{F} - \mathcal{V} = \begin{pmatrix} \beta_C SC + \beta_I SI + \beta_A SA \\ 0 \\ 0 \\ 0 \end{pmatrix} - \begin{pmatrix} \gamma_E E \\ -\gamma_E E + \gamma_C C \\ -p\gamma_C C + \gamma_I I \\ -(1-p)\gamma_C C + \gamma_A A \end{pmatrix}$$

$\mathcal{F}$ includes all terms that represent transfer *into* an infected compartment from uninfected compartments (in our case just the susceptible compartment). $\mathcal{V}$ includes all terms that represent other transfers *out of* infected compartments; other transfers *into* infected compartments show up with a minus sign. Transfers from one infected compartment to another show up twice, once with a plus sign and once with a minus sign. Transfers from an infected compartment to an uninfected compartment (in our case the recovered compartment) show up only once, with a plus sign.

To compute the next generation matrix, one first calculates the matrix K as we did by finding the linearization matrix at an equilibrium and then extracting the submatrix that has one row and one column for each infected compartment. In our case we obtained K by linearizing at an equilibrium $(S, 0, 0, 0, 0, R)$, and K has one row and one column for each of the infected compartments E, C, I, and A. We then write K as

$$K = F - V = \begin{pmatrix} 0 & \beta_C S & \beta_I S & \beta_A S \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} - \begin{pmatrix} \gamma_E & 0 & 0 & 0 \\ -\gamma_E & \gamma_C & 0 & 0 \\ 0 & -p\gamma_C & \gamma_I & 0 \\ 0 & -(1-p)\gamma_C & 0 & \gamma_A \end{pmatrix}.$$

F_{ij} is just the partial derivative of $\mathcal{F}_i$ with respect to the jth infected state. V_{ij} is just the partial derivative of $\mathcal{V}_i$ with respect to the jth infected state. All entries of F are necessarily nonnegative.

If we multiply F by a vector of population fractions in the infected compartments, we get a linear approximation to the total rate of transfer into all infected compartments from uninfected compartments.

For example, look at the calculation

$$F\begin{pmatrix} E \\ C \\ I \\ A \end{pmatrix} = \begin{pmatrix} 0 & \beta_C S & \beta_I S & \beta_A S \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} E \\ C \\ I \\ A \end{pmatrix} = \begin{pmatrix} \beta_C SC + \beta_I SI + \beta_A SA \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

This calculation finds the approximate total rate of transfer into all infected compartments from uninfected compartments, when the population fractions in the infected compartments are the given vector. There is transfer only into the E compartment. The rate depends on the S-component of the equilibrium where we linearized, in addition to depending on the population fractions in the infected compartments.

Remark 4.2. In general we should only get an approximation to the total rate of transfer, since we are using a linearization make the calculation. However, in our Covid-19 model, since all terms in the differential equations are linear in the variables E, C, I and A for a fixed value of S, the approximation is exact. This is also true for other calculations in this section that involve multiplying by a vector of population fractions in the infected compartments.

Remark 4.3. A linearization matrix is normally multiplied by a vector representing change in input in order to approximate change in output. Thus one would expect F to be multiplied by a vector of the form $(\Delta E \ \Delta C \ \Delta I \ \Delta A)^\top$. However, the point at which we linearized had $E = C = I = A = 0$ (it was a disease-free equilibrium), so $(\Delta E \ \Delta C \ \Delta I \ \Delta A)^\top = (E \ C \ I \ A)^\top$. This is why we can multiply F by the vector $(E \ C \ I \ A)^\top$. Because of this consideration, the next generation matrix method can only be used at disease-free equilibria.

If we multiply V by a vector of population fractions in the infected compartments, we get a linear approximation to the total rate of other transfers out of all infected compartments (i.e. transfers given by $\mathcal{V}$).

For example, look at the calculation

$$\begin{pmatrix} \gamma_E & 0 & 0 & 0 \\ -\gamma_E & \gamma_C & 0 & 0 \\ 0 & -p\gamma_C & \gamma_I & 0 \\ 0 & -(1-p)\gamma_C & 0 & \gamma_A \end{pmatrix}\begin{pmatrix} E \\ 0 \\ I \\ 0 \end{pmatrix} = \begin{pmatrix} \gamma_E E \\ -\gamma_E E \\ \gamma_I I \\ 0 \end{pmatrix}.$$

This calculation finds the approximate total rate of other transfers out of all infected compartments, when the population fractions in the infected compartments

are the given vector. For simplicity we used a vector with population fractions only in compartments E and I. We obtain a transfer out of compartment E, a corresponding transfer into compartment C, and a transfer out of compartment I with no corresponding transfer into an uninfected compartment. (Transfer out of compartment I is into the recovered compartment.) As in our previous calculation, the result is exact, although we only expected an approximation.

The next generation matrix makes use of V^{-1}, the inverse of the matrix V. If you are not familiar with V^{-1}, it is the matrix you multiply V by to get I. For our Covid-19 model,

$$V^{-1} = \begin{pmatrix} \frac{1}{\gamma_E} & 0 & 0 & 0 \\ \frac{1}{\gamma_C} & \frac{1}{\gamma_C} & 0 & 0 \\ \frac{p}{\gamma_I} & \frac{p}{\gamma_I} & \frac{1}{\gamma_I} & 0 \\ \frac{1-p}{\gamma_A} & \frac{1-p}{\gamma_A} & 0 & \frac{1}{\gamma_A} \end{pmatrix}. \tag{4.10}$$

We shall show in Section 4.8 that V^{-1}_{ij} always gives the mean time spent in infected compartment i if you start in infected compartment j. For example, for our Covid-19 model, the first column of (4.10) says that if you start in compartment E, on average you will spend time $\frac{1}{\gamma_E}$ in compartment E, $\frac{1}{\gamma_C}$ in compartment C, $\frac{p}{\gamma_I}$ in compartment I, and $\frac{1-p}{\gamma_A}$ in compartment A. If you remember Section 1.6, you can easily see that this is correct.

The *next generation matrix* is defined to be the matrix product FV^{-1}. For our model,

$$\begin{aligned} FV^{-1} &= \begin{pmatrix} 0 & \beta_C S & \beta_I S & \beta_A S \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\gamma_E} & 0 & 0 & 0 \\ \frac{1}{\gamma_C} & \frac{1}{\gamma_C} & 0 & 0 \\ \frac{p}{\gamma_I} & \frac{p}{\gamma_I} & \frac{1}{\gamma_I} & 0 \\ \frac{1-p}{\gamma_A} & \frac{1-p}{\gamma_A} & 0 & \frac{1}{\gamma_A} \end{pmatrix} \\ &= \begin{pmatrix} \frac{\beta_C S}{\gamma_C} + \frac{p\beta_I S}{\gamma_I} + \frac{(1-p)\beta_A S}{\gamma_A} & \frac{\beta_C S}{\gamma_C} + \frac{p\beta_I S}{\gamma_I} + \frac{(1-p)\beta_A S}{\gamma_A} & \frac{\beta_I S}{\gamma_I} & \frac{\beta_A S}{\gamma_A} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \end{aligned} \tag{4.11}$$

To understand FV^{-1}, it helps to use an intuitive meaning for the entries of the matrix F. Roughly speaking, F_{ik} is rate of transfer from uninfected compart-

ments into infected compartment i that is due to infection by people in infected compartment k. Next, think about how matrix multiplication works. V^{-1}_{kj} is mean time spent in infected compartment k by people initially in infected compartment j. Multiplying the rate F_{ik} by the time V^{-1}_{kj} gives mean transfers from uninfected compartments into infected compartment i that are due to the time spent in infected compartment k by individuals initially in infected compartment j. Summing over k gives the ij entry of FV^{-1}, which therefore is total mean transfer from uninfected compartments to infected compartment i caused by individuals initially in infected compartment j, over the entire time that they remain in infected compartments.

Let's look at a few entries of FV^{-1} in (4.11) to see how this works.

- The 21-entry of FV^{-1} is 0. This means that an individual initially in infected compartment 1, which is E, over the entire time he remains in infected compartments, causes no transfers at all from compartment S or R into infected compartment 2, which is I. The reason is simple: individuals in compartment S, when they become infected, transfer into compartment E, not compartment I; and individuals in compartment R never transfer into infected compartments. This is the reason the last three rows of FV^{-1} are rows of zeros.

- The 11-entry of FV^{-1} is $\frac{\beta_C S}{\gamma_C} + \frac{p\beta_I S}{\gamma_I} + \frac{(1-p)\beta_A S}{\gamma_A}$. This means that an individual initially in infected compartment 1, which is E, over the entire time he remains in infected compartments, causes this many transfers from compartment S into compartment E. The number depends on the equilibrium $(S, 0, 0, 0, 0, R)$ where we are linearizing.

  For $S = 1$, $\frac{\beta_C}{\gamma_C}$ represents infections caused during the time the individual is in compartment C; $\frac{p\beta_I}{\gamma_I}$ represents infections caused during the time the individual is in compartment I, weighted by the probability that he passes through that compartment; $\frac{(1-p)\beta_A}{\gamma_A}$ represents infections caused during the time the individual is in compartment A, weighted by the probability that he passes through that compartment.

  For $0 \leqslant S < 1$, each term must be multiplied by S. For example, the rate at which an individual in compartment C infects others becomes $\beta_C S$, not β_C.

- The 13-entry of FV^{-1} is $\frac{\beta_I S}{\gamma_I}$. This means that an individual initially in infected compartment 3, which is I, over the entire time he remains in in-

fected compartments, causes this many transfers from compartment S into compartment E. An individual initially in compartment I spends mean time $\frac{1}{\gamma_I}$ in that compartment before recovering, which explains the entry.

The 12-entry and 14-entry of FV^{-1} are left to you to explain.

From the meaning of FV^{-1} it follows that if we multiply FV^{-1} by an initial vector of infected population fractions $\mathcal{I}_0$, we get the vector of new infections in the various compartments caused by $\mathcal{I}_0$:

$$\mathcal{I}_{\text{new}} = FV^{-1}\mathcal{I}_0.$$

The size of $\mathcal{I}_{\text{new}}$ is some multiple of the size of $\mathcal{I}_0$ that depends on $\mathcal{I}_0$. For a nonnegative matrix, which FV^{-1} must be, the maximum value of this multiple is the largest eigenvalue of FV^{-1}.

It therefore makes sense to define the effective reproduction number R_e at a disease-free equilibrium to be the largest eigenvalue of the next generation matrix FV^{-1} at that equilibrium. In particular, it makes sense to define the basic reproducton number R_0 to be the largest eigenvalue of FV^{-1} at the equilibrium with $S = 1$.

By the way, in the models treated in the first three chapters, there was only one infected compartment, and the product FV^{-1} at the equilibrium with $S = 1$ simplified to $\beta\gamma^{-1}$.

For our Covid-19 model, the largest eigenvalue of FV^{-1} for $S \geqslant 0$ is

$$R_e = \frac{\beta_C S}{\gamma_C} + \frac{p\beta_I S}{\gamma_I} + \frac{(1-p)\beta_A S}{\gamma_A}. \tag{4.12}$$

When $S = 1$ we get

$$R_0 = \frac{\beta_C}{\gamma_C} + \frac{p\beta_I}{\gamma_I} + \frac{(1-p)\beta_A}{\gamma_A}, \tag{4.13}$$

which agrees with our earlier guess. This value for R_0 is the 11-entry of FV^{-1} for $S = 1$, which we discussed above. It makes intuitive sense. The first summand represents infections caused before any of the infected experience symptoms. The next two summands represent infections caused after this period, by symptomatic and asymptomatic infected individuals. These two terms are weighted by the fraction of individuals in compartment C who pass to each compartment.

Moreover, a basic theorem states that at a disease-free equilibrium of an epidemiological model, if $R_e < 1$, then all eigenvalues of K have negative real part;

and if $R_e > 1$, then at least one eigenvalue of K has positive real part (van den Driessche and Watmaugh 2002).

For our Covid-19 model, $R_e = R_0 S$. Therefore R_e is

- less then 1 if $S < \frac{1}{R_0}$;
- equal to 1 if $S = \frac{1}{R_0}$;
- greater than 1 if $S > \frac{1}{R_0}$.

Thus, as we guessed earlier, if $R_0 < 1$, an epidemic that starts will die out; if $R_0 > 1$, an epidemic that starts near an equilibrium with $\frac{1}{R_0} < S \leqslant 1$ will initially grow.

In the next two sections we will outline a proof that the entries of V^{-1} have the interpretation we gave earlier.

4.7 Differential equations background: matrix exponential

Consider a linear differential equation $\dot{x} = Ax$, with x in $\mathbb{R}^n$ and A an $n \times n$ matrix. The solution is a certain matrix function of t times the initial condition $x(0) = x_0$. That matrix function of t is called e^{tA}.

For example, consider the system of linear differential equations

$$\begin{aligned} \dot{x}_1 &= ax_1, \\ \dot{x}_2 &= bx_2, \end{aligned}$$

with initial conditions $x_1(0) = x_{10}$ and $x_2(0) = x_{20}$. The solution is

$$\begin{aligned} x_1(t) &= e^{at} x_{10}, \\ x_2(t) &= e^{bt} x_{20}. \end{aligned}$$

In matrix terms the system of linear differential equations and the initial conditions become

$$\begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \end{pmatrix} = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad \begin{pmatrix} x_1(0) \\ x_2(0) \end{pmatrix} = \begin{pmatrix} x_{10} \\ x_{20} \end{pmatrix},$$

and the solution is

$$\begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} = \begin{pmatrix} e^{at} & 0 \\ 0 & e^{bt} \end{pmatrix} \begin{pmatrix} x_{10} \\ x_{20} \end{pmatrix}.$$

In terms of the matrix exponential we have

$$A = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}, \quad e^{tA} = \begin{pmatrix} e^{at} & 0 \\ 0 & e^{bt} \end{pmatrix}.$$

In this example, given A, it was easy to find e^{tA}. We will not discuss general methods for finding e^{tA}. The important fact is:

Theorem 4.4. *The solution of $\dot{x} = Ax$, $x(0) = x_0$ is $x(t) = e^{tA}x_0$.*

You may recall that when x is a real number, e^x can be written as the infinite series

$$e^x = 1 + x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \dots.$$

Similarly, when A is a square matrix,

$$e^{tA} = I + tA + \frac{1}{2!}t^2A^2 + \frac{1}{3!}t^3A^3 + \dots. \tag{4.14}$$

In this formula, $A^2 = AA$, $A^3 = AAA$, etc. (matrix multiplication). Thus it is not surprising that e^x and e^{tA} have much in common. For example, $\frac{d}{dt}e^{ta} = ae^{ta}$ and

$$\frac{d}{dt}e^{tA} = Ae^{tA}. \tag{4.15}$$

4.8 Explanation of the entries of V^{-1}

Let

$$\mathcal{I} = \begin{pmatrix} E \\ C \\ I \\ A \end{pmatrix},$$

and consider the differential equation

$$\dot{\mathcal{I}} = -V\mathcal{I}, \quad \mathcal{I}(0) = \mathcal{I}_0.$$

The solution $\mathcal{I}(t)$ gives the vector of population fractions in the infected states at time t, assuming the initial vector of population fractions is $\mathcal{I}_0$, ignoring inputs from the uninfected states. $\mathcal{I}(t) \to 0$ as $t \to \infty$, since eventually all infected individuals transfer to the uninfected recovered state.

On the other hand, $-\dot{\mathcal{I}}(t) = V\mathcal{I}(t)$ is the rate of transfer *out* of the infected compartments at time t. We can calculate

$$\int_0^\infty -\dot{\mathcal{I}}(t)\, dt = [-\mathcal{I}(t)]_{t=0}^{t=\infty} = -\mathcal{I}(\infty) + \mathcal{I}(0) = \mathcal{I}(0) = \mathcal{I}_0. \tag{4.16}$$

The integral gives total transfer out of all compartments, which equals the initial concentrations, since eventually all infected individuals transfer into the recovered compartment. Equation (4.16) is the analog of equation (1.17) in dimensions greater than 1.

Let us take $\mathcal{I}_0$ to be 1 in one infected compartment and 0 in other infected compartments. Then the mean time spent in each infected compartment by individuals who start in that particular infected compartment is $\int_0^\infty -t\dot{\mathcal{I}}(t)\, dt$. This equation is analogous to the integral in equation (1.18).

Theorem 4.5. $\displaystyle\int_0^\infty -t\dot{\mathcal{I}}(t)\, dt = V^{-1}\mathcal{I}_0.$

This formula is analogous to (1.18). To interpret it, note that if $\mathcal{I}_0$ has a 1 in the first infected compartment and 0 in the others, then $V^{-1}\mathcal{I}_0$ is the first column of V^{-1}; if $\mathcal{I}_0$ has a 1 in the second infected compartment and 0 in the others, then $V^{-1}\mathcal{I}_0$ is the second column of V^{-1}; etc. Thus the jth column of V^{-1} gives the mean time spent in each infected compartment by individuals who start in the jth infected compartment.

Proof. We cannot evaluate the integral as we did in equation (1.18) because we do not know an explicit antiderivative. Instead we use integration by parts (which is how the antiderivative we used in (1.18) was found anyway):

$$\int_0^\infty -t\dot{\mathcal{I}}(t)\, dt = [-t\mathcal{I}(t)]_0^\infty + \int_0^\infty \mathcal{I}(t)\, dt.$$

The term $-t\mathcal{I}(t)$ is clearly 0 at $t = 0$. Its value at ∞ is actually a limit,

$$\lim_{t\to\infty} -t\mathcal{I}(t).$$

It turns out $\mathcal{I}(t)$ approaches 0 exponentially as $t \to \infty$, so the limit is 0.

Therefore

$$\int_0^\infty -t\dot{\mathcal{I}}(t)\, dt = \int_0^\infty \mathcal{I}(t)\, dt.$$

Now we use the fact that $\dot{\mathcal{I}} = -V\mathcal{I}$, so $\mathcal{I}(t) = e^{-tV}\mathcal{I}(0)$:

$$\int_0^\infty \mathcal{I}(t)\, dt = \int_0^\infty e^{-tV}\mathcal{I}(0)\, dt = [-V^{-1}e^{-tV}\mathcal{I}(0)]_0^\infty = V^{-1}\mathcal{I}(0).$$

This calculation may look a little mysterious at first glance. You can check that $-V^{-1}e^{-tV}\mathcal{I}(0)$ is the antiderivative by differentiating it using (4.15). Also, $e^{0V} = I$, as you can see from the infinite series (4.14). Finally, it turns out that $e^{-tV} \to 0$ as $t \to \infty$. □

4.9 Disease variants

In this section we present a simple model, based on the SIR model, of how a more infective variant of a disease can overwhelm a less infective variant.

We consider a disease with two variants. Both variants have the same γ, but their values of β, which we denote β_1 and β_2, differ. We will assume that $\beta_2 > \beta_1$, so that the second variant is more easily spread.

In addition to the usual S and R compartments, there are two infected compartments, one for each variant, which we denote I_1 and I_2. The system of differential equations is

$$\dot{S} = -\beta_1 SI_1 - \beta_2 SI_2, \tag{4.17}$$

$$\dot{I}_1 = \beta_1 SI_1 - \gamma I_1, \tag{4.18}$$

$$\dot{I}_2 = \beta_2 SI_2 - \gamma I_2, \tag{4.19}$$

$$\dot{R} = \gamma I_1 + \gamma I_2. \tag{4.20}$$

The set of equilibria is $\{(S, I_1, I_2, R) : I_1 = I_2 = 0\}$. The linearization of the system (4.17)–(4.20) at an equilibrium $(S, 0, 0, R)$ has the matrix

$$\begin{pmatrix} 0 & -\beta_1 S & -\beta_2 S & 0 \\ 0 & \beta_1 S - \gamma & 0 & 0 \\ 0 & 0 & \beta_2 S - \gamma & 0 \\ 0 & \gamma & \gamma & 0 \end{pmatrix}.$$

The eigenvalues are 0, 0, $\beta_1 S - \gamma$, and $\beta_2 S - \gamma$. In particular, if $S = 1$ (entire population susceptible), the eigenvalues are 0, 0, $\beta_1 - \gamma$, and $\beta_2 - \gamma$. We will assume that $\beta_1 - \gamma > 0$, so that $\beta_2 - \gamma > 0$ as well. In this case both variants can spread, but since the second eigenvalue is greater, we expect that early in the epidemic, at least, the second variant will spread faster.

The submatrix K with rows and columns for the infected states is

$$\begin{pmatrix} \beta_1 S - \gamma & 0 \\ 0 & \beta_2 S - \gamma \end{pmatrix}.$$

We rewrite it as

$$K = F - V = \begin{pmatrix} \beta_1 S & 0 \\ 0 & \beta_2 S \end{pmatrix} - \begin{pmatrix} \gamma & 0 \\ 0 & \gamma \end{pmatrix},$$

and calculate the next generation matrix at $S = 1$:

$$FV^{-1} = \begin{pmatrix} \beta_1 & 0 \\ 0 & \beta_2 \end{pmatrix} \begin{pmatrix} \gamma^{-1} & 0 \\ 0 & \gamma^{-1} \end{pmatrix} = \begin{pmatrix} \beta_1 \gamma^{-1} & 0 \\ 0 & \beta_2 \gamma^{-1} \end{pmatrix}.$$

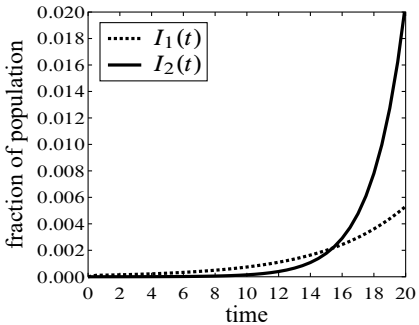
The eigenvalues are $\beta_1 \gamma^{-1}$ and $\beta_2 \gamma^{-1}$. The largest is $\beta_2 \gamma^{-1}$. We therefore have

$$R_0 = \beta_2 \gamma^{-1}.$$

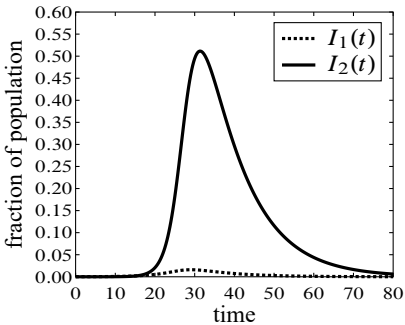
Comparing to the usual SIR model, we see that this value of R_0 is just the basic reproduction number one would expect if the less easily spread variant of the disease did not exist.

In fact, at the start of the epidemic, the more easily spread variant will overwhelm the other variant. Figure 4.3 shows the solution of (4.17)–(4.20) with $\beta_1 = 0.3$, $\beta_2 = 0.6$, $\gamma = 0.1$, and initial condition

$$(S(0), I_1(0), I_2(0), R(0)) = (1 - 10^{-4} - 10^{-6}, 10^{-4}, 10^{-6}, 0). \tag{4.21}$$



(a) Small time.



(b) Large time.

Figure 4.3: Solution of (4.17)–(4.18) with initial condition (4.21). Only the values of $I_1(t)$ and $I_2(t)$ are shown.

Thus initially the first variant is present at 100 times the level of the second. Only the values of $I_1(t)$ and $I_2(t)$ are shown. Figure 4.3a shows that the second variant passes the first after about 16 days. Figure 4.3b shows a longer time period.

The explanation for the dramatic curves in Figure 4.3 is simply the difference in exponential growth rates at the start of an epidemic. If a more easily spread variant is introduced later in the epidemic, the result could be less dramatic.

4.10 Discussion of the Covid-19 model

For our Covid-19 model (4.1)–(4.6), let us look more carefully at the expression (4.13) that we found for R_0:

$$R_0 = \frac{\beta_C}{\gamma_C} + \frac{p\beta_I}{\gamma_I} + \frac{(1-p)\beta_A}{\gamma_A}.$$

Here are some plausible values for the parameters.

- Since an average of 2.5 days are spent in state C, we use $\gamma_C = 1/2.5 = 0.4$. Since individuals are most infective at this time we take $\beta_C = 0.5$, greater than the value $\beta = 0.3$ that we have generally used for Covid-19.
- Since asymptomatic infectives cease to be infective after about five days, we take $\gamma_A = 1/5 = 0.2$, and we take $\beta_A = 0.3$.

- Since symptomatic infectives cease to be infective after about ten days, we take $\gamma_I = 1/10 = 0.1$. Symptomatic infectives are more infective than asymptomatic infectives, but on average they have fewer contacts since their illness usually keeps them at home. We take $\beta_I = \beta_A = 0.3$.
- Since about 2/3 of exposed individuals move into the symptomatic infective compartment I, we take $p = 2/3$.

Using these parameter values and our expression for R_0, we obtain

$$R_0 = (0.5)(2.5) + (2/3)(0.3)(10) + (1/3)(0.3)(5) = 1.25 + 2 + 0.5 = 3.75.$$

One can use this expression to estimate the impact of various control measures for Covid-19. For example, temperature checks are intended to prevent individuals in compartment I from contacting others. Employers can also urge any employee who feels sick (those in compartment I) to stay home. Suppose these measures succeed in reducing β_I from 0.3 to 0.1. We would then have

$$R_0 = (0.5)(2.5) + (2/3)(0.1)(10) + (1/3)(0.3)(5) = 1.25 + 0.67 + 0.5 = 2.42.$$

These measures would not stop the spread of the disease.

The problem is that individuals in compartments C and A probably do not suspect they have the disease. Mask-wearing by a large part of the population could perhaps reduce each β by about a third. This would reduce R_0 by a third, to 1.61.

Other possible control measures are contact tracing and lockdowns. Contact tracing might identify some people in compartment A who do not know they have the disease and encourage them to stay home. It is unlikely to identify many people while they are still in compartment E. A lockdown, like mask wearing, would reduce each β.

4.11 Problems

Problem 4.1 *Why so different?*

According to the data on Covid-19 the United Kingdom presented in Stevens et al. (2020), after 30 days, 12,647 individuals had contracted Covid-19. This is 0.02% of the UK population of 66,650,000. On the other hand, in Figure 4.2a, after 30 days, about 3% of the population is infected. Why the difference?

Problem 4.2 *Could you do better?*

In Section 4.4 we estimated R_0 for Covid-19 in the United Kingdom, at the start of the pandemic, to be 3.51. What are some sources of error in this estimate? Could you have improved it at the time?

Problem 4.3 *SEIR with births and natural deaths*

We will consider the SEIR model mentioned at the start of this chapter, with births and natural deaths included as we did for the SIR model in Problem 3.2. The system of differential equations for the population fractions is

$$\dot{S} = \mu - \beta SI - \mu S, \tag{4.22}$$

$$\dot{E} = \beta SI - \gamma_E E - \mu E, \tag{4.23}$$

$$\dot{I} = \gamma_E E - \gamma_I I - \mu I, \tag{4.24}$$

$$\dot{R} = \gamma_I I - \mu R. \tag{4.25}$$

We could drop the last equation, but we will not. The state space is the four-dimensional region

$$\mathcal{S} = \{(S, E, I, R) : S \geqslant 0,\ E \geqslant 0,\ I \geqslant 0,\ R \geqslant 0,\ S + E + I + R \leqslant 1\}. \tag{4.26}$$

The constants β, γ_E, γ_I, and μ are all positive.

1. Explain each term in the system of equations.

2. For fixed S, show that if

$$(\gamma_E + \mu)(\gamma_I + \mu) - \beta\gamma_E S \neq 0, \tag{4.27}$$

then any equilibrium of (4.22)–(4.25) with that value of S must have $E = I = R = 0$. Suggestion: Set the last three equations equal to 0. For fixed S, you have a system of three linear equations in three unknowns. In matrix form your system of equations is

$$\begin{pmatrix} -(\gamma_E + \mu) & \beta S & 0 \\ \gamma_E & -(\gamma_I + \mu) & 0 \\ 0 & \gamma_I & -\mu \end{pmatrix} \begin{pmatrix} E \\ I \\ R \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Show that (4.27) implies that the determinant of the 3×3 matrix is not zero. Then the only solution is $E = I = R = 0$.

3. The previous problem implies that if

$$\frac{\beta\gamma_E}{(\gamma_E+\mu)(\gamma_I+\mu)} < 1, \tag{4.28}$$

then the only equilibrium of (4.22)–(4.25) in $\mathcal{S}$ is $(1,0,0,0)$. Do you see why?

4. Of course $(1,0,0,0)$ is an equilibrium of (4.22)–(4.25) whether or not (4.28) holds; however it turns out that when

$$\frac{\beta\gamma_E}{(\gamma_E+\mu)(\gamma_I+\mu)} > 1,$$

there is a second equilibrium in $\mathcal{S}$. Instead of pursuing the second equilibrium, we will linearize the system (4.22)–(4.25) at the disease-free equilibrium $(1,0,0,0)$. Check that the linearization of (4.22)–(4.25) at $(1,0,0,0)$ has the matrix

$$\begin{pmatrix} -\mu & 0 & -\beta & 0 \\ 0 & -(\gamma_E+\mu) & \beta & 0 \\ 0 & \gamma_E & -(\gamma_I+\mu) & 0 \\ 0 & 0 & \gamma_I & -\mu \end{pmatrix}. \tag{4.29}$$

5. The submatrix K with rows and columns for infected states only is

$$K = \begin{pmatrix} -(\gamma_E+\mu) & \beta \\ \gamma_E & -(\gamma_I+\mu) \end{pmatrix}.$$

Show that the eigenvalues of (4.29) are $-\mu$ with multiplicity two and the eigenvalues of K.

6. Show that if $\frac{\beta\gamma_E}{(\gamma_E+\mu)(\gamma_I+\mu)} < 1$, then K has two eigenvalues with negative real part; and if $\frac{\beta\gamma_E}{(\gamma_E+\mu)(\gamma_I+\mu)} > 1$, then K has one negative eigenvalue and one positive eigenvalues. Suggestion: see Section 3.5. We conclude that epidemics cannot start near $(1,0,0,0)$ when $\frac{\beta\gamma_E}{(\gamma_E+\mu)(\gamma_I+\mu)} < 1$, but they can when $\frac{\beta\gamma_E}{(\gamma_E+\mu)(\gamma_I+\mu)} > 1$.

7. As we did in this chapter, write K in the form $K = F - V$ where F only contains terms that represent transfers into an infected compartment from a

uninfected compartment. Then calculate the next generation matrix FV^{-1} and find its largest eigenvalue. (The inverse of a 2×2 matrix is an easy formula; if you do not know it, google it.) Answer: the largest eigenvalue, which is R_0, is

$$\frac{\beta\gamma_E}{(\gamma_E + \mu)(\gamma_I + \mu)}.$$

How does this relate to the results of part 6?

Problem 4.4 *SIR-type model with population subgroups*

As we mentioned in Chapter 1, people in a population may preferentially encounter certain subgroups of the population, and this fact can be modeled by introducing more compartments into differential equation models. Let us consider an SIR-type model in which, for simplicity, we divide the population into just two subgroups, subgroup 1 and subgroup 2. The two subgroups might represent young and old, or poor and rich, or liberal and conservative. The contact rates within the two subgroups might differ, and the contact rates between subgroups are expected to be lower than the contact rates within subgroups. In addition, if the two subgroups have different levels of health or different susceptibility to the disease, the probability that a contact results in disease transmission could depend on the subgroup of the susceptible person who is contacted. Thus, if we consider contacts between infectives in subgroup i and susceptibles in subgroup j, the resulting adequate contact rate β_{ij} can depend on both i and j. We shall also allow the two subgroups to have different mean illness times, so that γ_1 and γ_2 may differ.

We consider a compartmental model in which there are two susceptible compartments S_1 and S_2 for individuals in subgroups 1 and 2 respectively, and two infective compartments I_1 and I_2 for individuals in subgroups 1 and 2. We shall only use one removed compartment, which includes individuals from both subgroups. See Figure 4.4.

The resulting system of differential equations is

$$\dot{S}_1 = -\beta_{11} S_1 I_1 - \beta_{21} S_1 I_2, \tag{4.30}$$

$$\dot{S}_2 = -\beta_{12} S_2 I_1 - \beta_{22} S_2 I_2, \tag{4.31}$$

$$\dot{I}_1 = \beta_{11} S_1 I_1 + \beta_{21} S_1 I_2 - \gamma_1 I_1, \tag{4.32}$$

$$\dot{I}_2 = \beta_{12} S_2 I_1 + \beta_{22} S_2 I_2 - \gamma_2 I_2, \tag{4.33}$$

$$\dot{R} = \gamma_1 I_1 + \gamma_2 I_2. \tag{4.34}$$

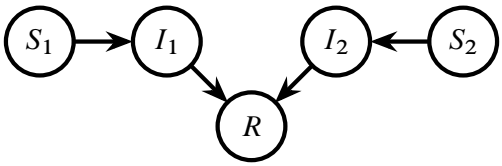


Figure 4.4: SIR-type model with two population subgroups.

The phase space is

$$\{(S_1, S_2, I_1, I_2, R) : S_1 \geqslant 0, S_2 \geqslant 0, I_1 \geqslant 0, I_2 \geqslant 0, R \geqslant 0, \\ S_1 + S_2 + I_1 + I_2 + R = 1\}.$$

The calculation of R_0 in this problem presents a subtlety. R_0 is calculated at the equilibrium at which the entire population is susceptible. In this problem, however, all equilibria of the form $(S_1, S_2, 0, 0, 0)$ have the entire population susceptible. (We of course have $S_1 \geqslant 0$, $S_2 \geqslant 0$, and $S_1 + S_2 = 1$.) The solution is to calculate R_0 at the equilibrium in which S_1 and S_2 represent the population fractions of the two subgroups that are actually present at the start of the epidemic.

1. Find the equilibria and find the matrices F and V.

2. Find the next generation matrix.

3. Using the next generation matrix that you found in part 2, calculate R_0 at the equilibrium with $S_1 = 1$ and $S_2 = 0$. Does the result make sense to you? Why?

4. Assume all β_{ij} are the same number β, and both γ_i are the same number γ. Using the next generation matrix that you found in part 2, calculate R_0 at an arbitrary equilibrium $(S_1, S_2, 0, 0, 0)$ with $S_1 \geqslant 0$, $S_2 \geqslant 0$, and $S_1 + S_2 = 1$. Does the result make sense to you? Why?

5. Assume $\beta_{11} = 2\gamma_1$, $\beta_{21} = 3\gamma_2$, $\beta_{12} = 3\gamma_1$, $\beta_{22} = 10\gamma_2$, and that we start with $S_1 = S_2 = \frac{1}{2}$. Using the next generation matrix that you found in part 2, find R_0 and the corresponding eigenvectors and interpret the result.

5 Spontaneous human behavioral change in epidemiological models

In the models we have considered so far, human behavior is not affected by the course of the epidemic. For example, in all our models the constant β, which represents the rate of adequate contacts per infective individual, does not change as the epidemic proceeds. You know from experience that this is false. As the number of infectives in the population rises, many susceptibles will react by changing their behavior. They stay home, practice improved hygiene and social distancing, and wear masks. The result is a decrease in β. Later, if the number of infectives in the population falls, susceptibles may relax these practices.

Changes in human behavior can be affected by government orders that close businesses or require people to stay home and to wear masks if they do not. However, behavioral change will occur whether or not there are government orders, and compliance with orders is greater if they correspond to what people are inclined

to do anyway.

Similarly, in our brief discussion of vaccines in Section 2.7, we just described what would happen if a sufficient population fraction was vaccinated. We did not discuss how people might react to the vaccination program itself. The reaction may have much in common with behavioral change. If the number of infectives in the population is high, people may suppress any concerns about vaccine safety in order to protect themselves or their children from the disease. If the number of infectives in the population is low, perhaps because of the vaccination program itself, concerns about vaccine safety may lead people to refuse vaccination for themselves or their children.

In this chapter we will look at how an idea from evolutionary game theory, imitation dynamics, can be used to model human behavior in these situations. Game theory is about situations in which your payoff from an action depends not only on your own choices but on the choices of others. Evolutionary game theory is the side of game theory that uses differential equations to model how choices change in response to the changing choices of others.

Payoffs in game theory are not necessary monetary; they include whatever individuals regard as the positive and negative consequences of choices, and may be subjective.

We will discuss human behavior in an epidemic. In the problems we will look at human response to vaccination programs.

Imitation dynamics was introduced into mathematical epidemiology by Bauch and Bhattacharyya (2012) in the context of a model for childhood vaccination. Poletti (2010) introduced the idea of using imitation dynamics to model human response to an epidemic. The analysis of Poletti's model in this chapter comes from Schecter (2021).

5.1 A model for human behavior in an epidemic

We will consider an SIR model in which susceptibles can choose between normal behavior and careful behavior. The resulting values of β are β_n for normal behavior and β_c for careful behavior. Careful behavior results in fewer contacts with infectives, so $\beta_c < \beta_n$.

Each behavior yields an expected payoff to a susceptible who adopts it. For simplicity the baseline payoff of normal behavior in the absence of infectives is taken to be 0. If infectives are present in the population, normal behavior has an additional negative payoff due to the possibility of contracting the disease. In

the SIR model, for each individual susceptible, the probability of contracting the disease is proportional to the population fraction of infectives I. We therefore take the payoff p_n of normal behavior to be

$$p_n = -m_n I$$

where m_n is a positive constant.

Individuals adopting careful behavior still might get the disease, so they have a payoff of $-m_c I$ where m_c is a positive constant but $m_c < m_n$. They also suffer an average negative payoff $-k$, where k is a positive constant, from adopting careful behavior, due to some combination of loss of income, refraining from desired activities, loss of valued social interactions, etc. Thus the payoff p_c of careful behavior is

$$p_c = -k - m_c I.$$

The difference between the two payoffs is

$$p_n - p_c = k - (m_n - m_c)I = k - mI,$$

where m, like k is a positive constant. We see that normal behavior has a higher payoff when $I < \frac{k}{m}$ and a lower payoff when $I > \frac{k}{m}$.

Let x denote the fraction of susceptibles who use normal behavior, $0 \leqslant x \leqslant 1$, so $1 - x$ is fraction of susceptibles who use careful behavior. The idea of imitation dynamics is that susceptibles using normal and careful behavior encounter each other at a rate proportional to the product $x(1 - x)$; this is the law of mass action again. The encounters could be in person or by email, text, or social media. If an individual finds that the individual she encounters is using a behavior that gives a better payoff than her own, it is possible that she will change to the opposite behavior. The rate of change is assumed to be proportional to the difference in payoffs of the two behaviors. Thus

$$\dot{x} = \rho x(1 - x)(p_n - p_c) = \rho x(1 - x)(k - mI). \tag{5.1}$$

Notice that when normal behavior gives a higher payoff, x increases; when normal behavior gives a lower payoff, x decreases.

Imitation dynamics is analogous to transmission of a disease from an infective to a susceptible. If, for example, normal behavior has a higher payoff, we can think of people practicing normal behavior as infectives, and people practicing careful behavior as susceptibles. When people from the two groups meet, those practicing

normal behavior may "infect" those practicing careful behavior and cause them to change their behavior.

(Actually, Equation (5.1) is not quite right. If the size of the susceptible group decreases, for example, then encounters between different types of susceptibles should become less frequent. This fact is not taken into account in Equation (5.1). You may recall that when we looked at the SIR model for a sometimes fatal disease, in Section 2.6.1, we did correctly account for the change in rate of encounters caused by a decrease in a group size. Nevertheless we shall leave Equation (5.1) as it is, since it is simple and commonly used, and changing it does not usually make much difference in the solutions.)

The full SIR model with imitation dynamics, ignoring the equation for the recovered compartment, is

$$\dot{S} = -\big(\beta_n x + \beta_c(1-x)\big)SI, \tag{5.2}$$

$$\dot{I} = \big(\beta_n x + \beta_c(1-x)\big)SI - \gamma I, \tag{5.3}$$

$$\dot{x} = x(1-x)(\beta_c - \beta_n)I + \rho x(1-x)(k - mI). \tag{5.4}$$

Let us unpack this system of equations. The average value of β in the population, given x, is $\beta_n x + \beta_c(1-x)$. That explains the first two equations. The only mystery is the first summand in the $\dot{x}$ equation. Since $\beta_c < \beta_n$, this term is always negative. It just reflects the fact that on its own, x tends to decrease, because susceptibles with normal behavior contract the disease more frequently than susceptibles with careful behavior and transfer to compartment I. The form of this term will be derived in the problems.

The state space is the prism

$$\mathcal{P} = \{(S, I, x) : S \geqslant 0,\ I \geqslant 0,\ S + I \leqslant 1,\ 0 \leqslant x \leqslant 1\}.$$

We shall make two assumptions:

1. $\frac{\beta_c}{\gamma} < 1 < \frac{\beta_n}{\gamma}$.
2. $\frac{k}{m} < 1$.

The first assumption says that with normal behavior, $R_0 > 1$, so the epidemic can spread; but with careful behavior, $R_0 < 1$, so the epidemic should die out. The second assumption says that neither behavior is guaranteed to always give a higher payoff. While normal behavior always gives a higher payoff for sufficiently low I, the second assumption guarantees that for sufficiently large I in the interval $0 \leqslant I \leqslant 1$, careful behavior gives a higher payoff.

It also seems reasonable to assume that ρ is much greater than β_n, β_c and γ, since behavior is capable of changing much faster than most diseases can spread. Figure 5.1, taken from Schecter (2021), shows a typical simulation of the system (5.2)–(5.4). In this simulation the parameter values are

$$\beta_n = 1/2,\ \beta_c = 1/10,\ \gamma = 1/6,\ k = 3/10,\ m_n = 5,\ m_c = 2,\ \rho = 200.$$

From the values of β_n, β_c, and γ, we see that $R_0 = 3$ for normal behavior and $R_0 = 0.6$ for careful behavior. We have $I/(m_n - m_c) = 1/10$, so normal behavior has a higher payoff for $I < 1/10$, and careful behavior has a higher payoff for $I > 1/10$.

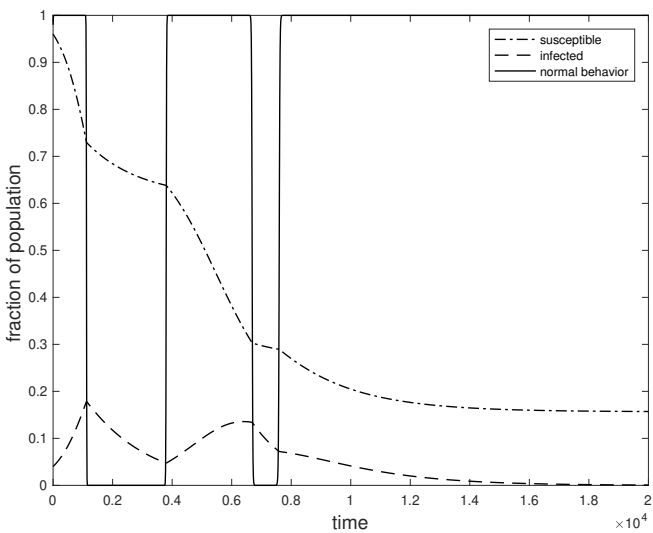


Figure 5.1: A simulation of the system (5.2)–(5.4). If t represents time in days, the time scale of this simulation is $1/200$ of a day. The simulation shows 20,000 time units, or 100 days.

At the start of the simulation, $(S, I, x) = (0.96, 0.04, 0.98)$. In particular, $I < 1/10$. Almost all the population quickly adopts normal behavior. After I rises to about 0.18, the population quickly switches to careful behavior. Because of the careful behavior, I falls to about 0.05. The population then quickly returns to normal behavior, and I rises to about 0.13 (the second wave of the epidemic). After two more fast behavioral switches, the epidemic dies out.

An important characteristic of the model we are studying in this chapter, which we see in the simulation, is that unlike our previous models, it can produce epidemics with several waves.

Another important characteristic of the model is that behavior is "sticky." By this we mean that behavior does not immediately switch from normal to careful, or vice-versa, when I passes 1/10. It takes a while for behavior to change. This is because behavioral change is caused by interaction with other people using a different behavior, which takes time.

In the remainder of this chapter we will try to gain mathematical insight into these aspects of the model using the theory of *slow-fast systems* (Kuehn 2015).

5.2 Slow time and fast time

Since we assume that ρ is large, we write $\rho = \frac{1}{\epsilon}$ with $\epsilon > 0$ small. We then multiply (5.4) by ϵ to remove fractions:

$$\dot{S} = -\big(\beta_n x + \beta_c(1-x)\big)SI, \tag{5.5}$$

$$\dot{I} = \big(\beta_n x + \beta_c(1-x)\big)SI - \gamma I, \tag{5.6}$$

$$\epsilon\dot{x} = \epsilon x(1-x)(\beta_c - \beta_n)I + x(1-x)(k - mI), \tag{5.7}$$

Remember that the dot in (5.5)–(5.7) represents derivative with respect to t, which normally represents time in days. The variable t is called slow time. Behavior can change on a faster time scale. To capture this fact, we define a fast time $\tau = \frac{t}{\epsilon}$. For example, if t is time in days and $\epsilon = \frac{1}{24}$, then τ is time in hours.

By the chain rule,

$$\frac{d}{dt} = \frac{d}{d\tau}\frac{d\tau}{dt} = \frac{1}{\epsilon}\frac{d}{d\tau}$$

We make this substitution in all three equations of (5.5)–(5.7) and then multiply the first two equations by ϵ to remove fractions. If we use $'$ to mean derivative with respect to τ, we end up with the system

$$S' = -\epsilon\big(\beta_n x + \beta_c(1-x)\big)SI, \tag{5.8}$$

$$I' = \epsilon\big(\beta_n x + \beta_c(1-x)\big)SI - \epsilon\gamma I, \tag{5.9}$$

$$x' = \epsilon x(1-x)(\beta_c - \beta_n)I + x(1-x)(k - mI). \tag{5.10}$$

For $\epsilon > 0$, the systems (5.5)–(5.7) and (5.8)–(5.10) have exactly the same orbits. The only difference is the speed of movement along orbits. However, the limits at $\epsilon = 0$ are entirely different.

The slow limit system, given by setting $\epsilon = 0$ in (5.5)–(5.7), is

$$\dot{S} = -\big(\beta_n x + \beta_c(1-x)\big)SI, \tag{5.11}$$
$$\dot{I} = \big(\beta_n x + \beta_c(1-x)\big)SI - \gamma I, \tag{5.12}$$
$$0 = x(1-x)(k-mI). \tag{5.13}$$

The fast limit system, given by setting $\epsilon = 0$ in (5.8)–(5.10), is

$$S' = 0, \tag{5.14}$$
$$I' = 0, \tag{5.15}$$
$$x' = x(1-x)(k-mI). \tag{5.16}$$

Solutions like that shown in the simulation can be approximated by combining solutions of the fast limit system (5.14)–(5.16) and the slow limit system (5.11)–(5.13). We will therefore look at the two systems separately, and then look at how to combine their solutions.

5.3 The fast limit system

The fast limit system (5.14)–(5.16) has three planes of equilibria: $x = 0$, $x = 1$, and $I = \frac{k}{m}$. You can check that the linearization matrix at a point on $x = 0$ or $x = 1$ has the eigenvalues 0, 0, and $\frac{\partial x'}{\partial x} = (1-2x)(k-mI)$. Therefore, equilibria on these two planes are normally hyperbolic provided $I \neq \frac{k}{m}$. You can check that

- On the plane $x = 0$, equilibria are normally repelling (positive eigenvalue) for $I < \frac{k}{m}$ and normally attracting (negative eigenvalue) for $I > \frac{k}{m}$.
- On the plane $x = 1$, the situation is the reverse: equilibria are normally attracting (negative eigenvalue) for $I < \frac{k}{m}$ and normally repelling (positive eigenvalue) for $I > \frac{k}{m}$.

The plane of equilibria $I = \frac{k}{m}$ has no normally hyperbolic equilibria and is of little importance. We will ignore it.

Solutions of the fast limit system with $0 < x < 1$ are easy to understand; see Figure 5.2a:

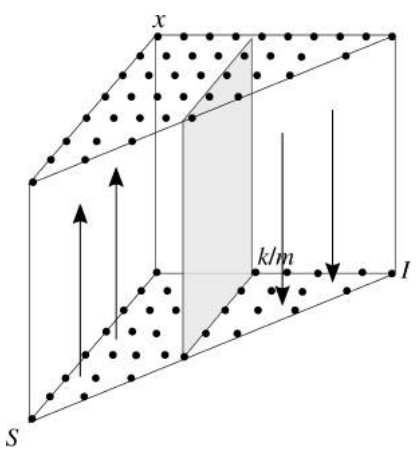
- S and I do not change along solutions.
- If $I < \frac{k}{m}$, then $\dot{x} > 0$, and in fact

$$\lim_{t\to-\infty} x(t) = 0, \quad \lim_{t\to\infty} x(t) = 1.$$

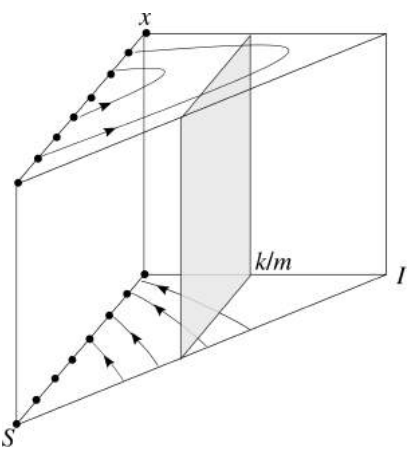
- If $I > \frac{k}{m}$, then $\dot{x} < 0$, and in fact

$$\lim_{t\to-\infty} x(t) = 1, \quad \lim_{t\to\infty} x(t) = 0.$$

The last two statements mean that if I is small, behavior quickly evolves toward everyone behaving normally; if I is large, behavior quickly evolves toward everyone behaving carefully.



(a) For the fast limit system, orbits in $0 < x < 1$ are vertical, and the triangles at $x = 0$ and $x = 1$ consist of equilibria



(b) The slow limit system is only defined on the triangles at $x = 0$ and $x = 1$. On $x = 0$ it is an SIR system with $R_0 < 1$; on $x = 1$ it is an SIR system with $R_0 > 1$.

Figure 5.2: Fast and slow limit systems in the prism $\mathcal{P}$.

5.4 The slow limit system

The slow limit system (5.11)–(5.13) only makes sense where the last equation is satisfied, i.e., on the planes $x = 0$ and $x = 1$ (the top and bottom of the prism), and on the plane $I = \frac{k}{m}$. The last set is of no importance and we will ignore it.

The planes $x = 0$ and $x = 1$ are actually invariant for the slow system (5.5)–(5.7) for every $\epsilon > 0$. Equivalently, they are invariant for the original system (5.2)–(5.4) for every $\rho > 0$.

On the set $x = 0$ (all behavior is careful) the slow limit system reduces to

$$\dot{S} = -\beta_c SI, \tag{5.17}$$

$$\dot{I} = \beta_c SI - \gamma I. \tag{5.18}$$

On the set $x = 1$ (all behavior is normal) the slow limit system reduces to

$$\dot{S} = -\beta_n SI, \tag{5.19}$$

$$\dot{I} = \beta_n SI - \gamma I. \tag{5.20}$$

Both systems are ordinary SIR systems. See Figure 5.2b.

5.5 Combining solutions of the fast and slow limit systems

The idea is that for a small $\epsilon > 0$, a solution of the fast system, or equivalently of the slow system, that starts in $0 < x < 1$ will be near a solution of the fast limit system until x is near 0 or 1. Then the vectors in the fast limit system become small and the slow limit system takes over.

Thus a solution for small $\epsilon > 0$ of the fast or slow system will immediately approach $x = 1$ (everyone behaves normally) if $I < \frac{k}{m}$, or $x = 0$ (everyone behaves carefully) if $I > \frac{k}{m}$. Having arrived near the plane $x = 1$ or $x = 0$ at the attracting part of the plane, the solution will follow a solution of the corresponding SIR system. If that solution eventually moves into the repelling part of the plane $x = 1$ or $x = 0$, the solution may leave and follow a solution of the fast limit system to the other side of the prism. The process then repeats.

This process can produce repeated waves of an epidemic. A wave occurs whenever the slow portion of the solution in the plane $x = 1$ has $I(t)$ increasing. The standard SIR system cannot produce repeated waves.

It turns out that a solution that is tracking a solution on $x = 0$ or $x = 1$ does not immediately leave when the tracked solution enters the repelling part of the

plane. Instead it leaves approximately at the point determined by the *entry-exit integral*. This phenomenon can occur when the slow limit system has a subspace of equilibria that remains invariant for $\epsilon > 0$, like $x = 0$ and $x = 1$ in our system. The entry-exit integral gives the mathematical explanation for the behavior stickiness we observed in the simulation.

5.6 Entry-exit integral

For the fast limit system (5.14)–(5.16), attraction or repulsion toward $x = 0$ at a point $(S, I, 0)$ is governed by the number

$$\frac{\partial x'}{\partial x}(S, I, 0) = k - mI. \tag{5.21}$$

Let $(S_0, I_0, 0)$ be a point of $x = 0$ where (5.21) is negative. Let $(S(t), I(t))$ be the solution of (5.17)–(5.18) with $(S(0), I(0)) = (S_0, I_0)$. Let $t_1 > 0$ be the time such that

$$\int_0^{t_1} \frac{\partial x'}{\partial x}(S(t), I(t), 0)\, dt = 0. \tag{5.22}$$

Let $(S_1, I_1) = (S(t_1), I(t_1))$. Because the solution $(S(t), I(t))$ approaches an equilibrium in the repelling part of $x = 0$ (see the phase portrait), such a time t_1 always exists. The reason is that the solution takes infinite time to reach the equilibrium, so the integral in (5.22) increases without bound as $t_1 \to \infty$.

Theorem 5.1. *For small $\epsilon > 0$, suppose a solution of* (5.8)–(5.10) *(or* (5.5)–(5.7)*) arrives in a small neighborhood of $x = 0$ near the point $(S_0, I_0, 0)$. Then the solution will leave that neighborhood near the point $(S_1, I_1, 0)$.*

Notice that the integral (5.22) is surely negative (representing attraction toward $x = 0$) for small $t_1 > 0$, but the curve $(S(t), I(t), 0)$ will eventually enter the repelling part of $x = 0$, at which point a positive contribution to the integral (representing repulsion from $x = 0$) starts building up. At t_1, repulsion balances attraction, and the solution leaves.

Similarly, for the fast limit system (5.14)–(5.16), attraction or repulsion toward $x = 1$ at a point $(S, I, 1)$ is governed by the number

$$\frac{\partial x'}{\partial x}(S, I, 1) = -(k - mI). \tag{5.23}$$

Let $(S_0, I_0, 1)$ be a point of $x = 1$ where (5.23) is negative. Let $(S(t), I(t))$ be the solution of (5.19)–(5.20) with $(S(0), I(0)) = (S_0, I_0)$. Let t_1 be the smallest positive number, if there is one, such that

$$\int_0^{t_1} \frac{\partial x'}{\partial x}(S(t), I(t), 1)\, dt = 0. \tag{5.24}$$

Let $(S_1, I_1) = (S(t_1), I(t_1))$. Unlike in the plane $x = 0$, in the plane $x = 1$ such a time t_1 does not necessarily exist. If you look at the phase portrait, you will see that the solution in $x = 1$ may never enter the repelling part of the plane, or it may stay in the repelling part of the plane for only a brief time before reentering the attracting part of the plane.

Theorem 5.2. *For small $\epsilon > 0$, suppose a solution of* (5.8)–(5.10) *(or* (5.5)–(5.7)*) arrives in a small neighborhood of $x = 1$ near the point $(S_0, I_0, 1)$.*

- *If there is a time t_1 as defined above, then the solution will leave that neighborhood near the point $(S_1, I_1, 1)$.*
- *If there is no such t_1, then the solution will never leave the neighborhood and will continue to follow $(S(t), I(t), 1)$.*

Let us consider how to calculate the integral in (5.22) or (5.24) for a fixed value of t_1. The integral in (5.22) for example is

$$\int_0^{t_1} \frac{\partial x'}{\partial x}(S(t), I(t), 0)\, dt = \int_0^{t_1} \big(k - mI(t)\big)\, dt. \tag{5.25}$$

The differential equation in the plane $x = 0$ is given by (5.17)–(5.18). According to (2.10), the curve $(S(t), I(t))$ from $(S_0, I_0) = (S(0), I(0))$ to $(S_1, I_1) = (S(t_1), I(t_1))$, as a curve in the SI-plane, has the equation

$$I = -S + \frac{\gamma}{\beta_c} \ln S + C.$$

We can determine C from the fact that the curve passes through (S_0, I_0):

$$C = I_0 + S_0 - \frac{\gamma}{\beta_c} \ln S_0. \tag{5.26}$$

Then, using (5.17), we can convert (5.25) into an integral with the variable S by making the substitution $S = S(t)$, $dS = \dot{S}\, dt$, so that

$$dt = \frac{1}{\dot{S}}\, dS = -\frac{1}{\beta_c SI}\, dS.$$

We obtain

$$\int_0^{t_1} \big(k - mI(t)\big)\, dt = \int_{S_0}^{S_1} -\frac{k - mI}{\beta_c SI}\, dS$$
$$= \int_{S_0}^{S_1} -\frac{k - m(-S + \frac{\gamma}{\beta_c} \ln S + C)}{\beta_c S(-S + \frac{\gamma}{\beta_c} \ln S + C)}\, dS \qquad (5.27)$$

where C is given by (5.26).

The integral (5.27) looks like the sort of integral one studies in calculus. Unfortunately there is no formula for an antiderivative, so it must be evaluated numerically.

Once one has the integral as a numerically computed function of t_1, if it switches from negative to positive as t_1 increases, one can use a numerical method such as bisection to find the value of t_1 where the integral is 0.

5.7 Singular solutions

A sequence of solutions of the fast and slow limit systems that a true solution, for small $\epsilon > 0$, is expected to follow is called a *singular solution*.

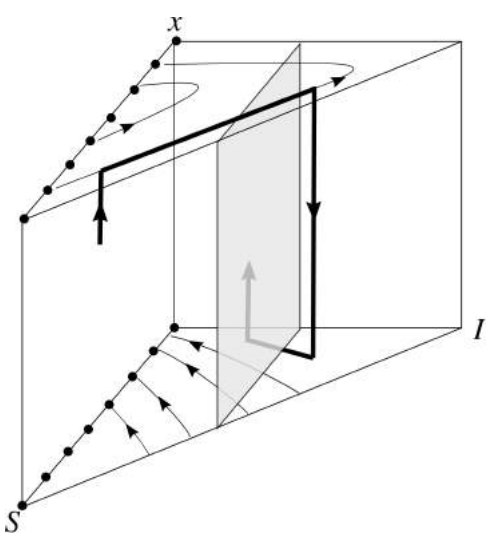


Figure 5.3: A singular solution.

Figure 5.3 shows the beginning of a singular solution.

1. The first fast solution starts at point where S is large, I is small, and most people are using normal behavior. It leads to a point in $x = 1$, i.e., everyone uses normal behavior.

2. The first slow solution is in $x = 1$, i.e., it is a solution of an SIR system with $R_0 > 1$. The number of infectives increases until a point where the entry-exit integral (5.24) is zero.

3. The second fast solution leads to a point in $x = 0$, i.e., everyone uses careful behavior.

4. The second slow solution is in $x = 0$, i.e., it is a solution of an SIR system with $R_0 < 1$. The number of infectives declines until a point where the entry-exit integral (5.22) is zero.

5. The third fast solution then leads to a point in $x = 1$.

6. The third slow solution (not shown) is in $x = 1$. It could represent a second wave of the epidemic (i.e., it could begin with $I(t)$ increasing). In this case there could be a fourth fast solution, or the epidemic could pass a peak and then die out. Alternatively, the third slow solution could have $I(t)$ decreasing from the start, representing the end of the epidemic.

Let us give a numerical example. We use the parameter values

$$\beta_n = 1/2,\ \beta_c = 1/10,\ \gamma = 1/6,\ k = 3/10,\ m_n = 5,\ m_c = 2.$$

The same parameter values were used in the simulation in Section 5.1. As in the simulation we shall use the starting point $(S, I, x) = (0.96, 0.04, 0.98)$. The computations of entry-exit integrals needed to produce the singular solution were done in Schecter (2021). The resulting singular solution is the following. You should compare the singular solution to the simulation, which is actually a simulation of (5.8)–(5.10) with $\epsilon = 1/200$.

1. The first fast solution goes from (0.96, 0.04, 0.98) to (0.96, 0.04, 1).

2. The first slow solution goes from (0.96, 0.04, 1) to (0.720, 0.184, 1), where the entry-exit integral is 0.

3. The second fast solution goes from (0.720, 0.184, 1) to (0.720, 0.184, 0).

4. The second slow solution goes from (0.720, 0.184, 0) to (0.626, 0.045, 0), where the entry-exit integral is 0.

5. The third fast solution goes from (0.626, 0.045, 0) to (0.626, 0.045, 1).

6. The third slow solution goes from (0.626, 0.045, 1) to (0.276, 0.122, 1), where the entry-exit integral is 0.

7. The fourth fast solution goes from (0.276, 0.122, 1) to (0.276, 0.122, 0).

8. The fourth slow solution goes from (0.276, 0.122, 0) to (0.268, 0.081, 0), where the entry-exit integral is 0.

9. The fifth fast solution goes from (0.268, 0.081, 0) to (0.268, 0.081, 1).

10. The fifth slow solution goes from (0.268, 0.081, 1) to (0.146, 0, 1), where the epidemic ends. There is no point where the entry-exit integral is 0.

5.8 Discussion of human behavioral change

The model presented in this chapter is part of the growing discipline of behavioral epidemiology. A good introduction is Bauch, d'Onofrio, and Manfredi (2013).

The model could be changed in various ways. Here are a few.

1. In the model, everyone experiences the same fixed cost $-k$ of careful behavior. This is of course not correct. The costs of staying home to a professional who can work from home and to a casual laborer who must find work every day are completely different. Perhaps the model should have different compartments for different social groups.

2. The fixed cost $-k$ of careful behavior may not actually be fixed. Over time people may simply tire of wearing a mask and staying home, or they may exhaust their savings and need to find work. In these cases the perceived cost of staying home increases.

3. The model assumes that people react to the present state of the epidemic, i.e., to $I(t)$. Another possibility is that people react to the memory of the epidemic, i.e., to $I(s)$ for $s \leqslant t$. This might be one reason why, when $I(t)$ declines, many people still stay home. See Poletti, Ajelli, and Merler (2012) for this approach.

4. In the model people only react to the state of the epidemic and their interactions with other people. In fact people also react to government orders and to information presented in the media. Since different media can present very different information, perhaps the model should have different compartments for different styles of media consumption.

Human behavior is complicated. Different aspects of human behavior might be salient in different situations. Thus simplified models that only capture one or a few aspects of human behavior might correctly account for behavior in some situations but not others.

Mathematical epidemiologists know that spontaneous human behavioral change is important. However, it is rarely included in models in practical use. For example, Ferguson, Cummings, et al. (2006), in their detailed agent-based model of influenza in the US and UK that was mentioned in the Preface, write: "We do not assume any spontaneous change in the behaviour of uninfected individuals as the pandemic progresses, but note that behavioural changes that increased social distance together with some school and workplace closure occurred in past pandemics …and might be likely to occur in a future pandemic even if not part of official policy. …Such spontaneous changes in population behaviour might more easily reduce peak daily case incidence."

There are at least three reasons that behavioral change is not taken into account in most models.

1. As we mentioned above, human behavior is complicated, and one can imagine taking into account many different aspects of human behavior.

2. There is not yet a widely accepted general approach to how to take into account human behavior in epidemic models.

3. Gathering data relevant to human behavior is a challenging problem.

These issues present an important challenge for the future.

5.9 Problems

Problem 5.1 *Completion of the derivation of the SIR model with imitation dynamics*

In this problem we explain the first summand in equation (5.4). The system (5.2)–(5.4) is actually based on a four-compartment model. There are two susceptible

compartments, S_n and S_c. Individuals in S_n use normal behavior with adequate contact rate β_n, and individuals in S_c use careful behavior with adequate contact rate β_c. The other compartments are infectives and recovered. Ignoring imitation dynamics, we have the following system of differential equations for the population fractions:

$$\dot{S}_n = -\beta_n S_n I, \tag{5.28}$$

$$\dot{S}_c = -\beta_c S_c I, \tag{5.29}$$

$$\dot{I} = (\beta_n S_n + \beta_c S_c) I - \gamma I, \tag{5.30}$$

$$\dot{R} = \gamma I. \tag{5.31}$$

Define $x = S_n/(S_n + S_c)$, so $1 - x = S_c/(S_n + S_c)$. Use the quotient rule to show that

$$\dot{x} = x(1 - x)(\beta_c - \beta_n) I.$$

Problem 5.2 *Human response to vaccination programs*

Consider the SIR system with births and natural deaths that we looked at in Problem 3.2. We assume $\frac{\gamma+\mu}{\beta} < 1$. We saw in Problem 3.2 that with this assumption, the disease becomes endemic, i.e., there is an interior equilibrium (S_*, I_*). With the opposite assumption $\frac{\gamma+\mu}{\beta} > 1$, the disease dies out. R_0 for this system is $\frac{\beta}{\gamma+\mu}$.

Since the disease becomes endemic if nothing is done, it is reasonable to develop a vaccination program. Suppose that a fraction x of newborns are vaccinated against the disease, and that the vaccination is completely effective. Thus a fraction x of newborns enter the recovered category, and a fraction $1 - x$ enter the susceptible category. The system in Problem 3.2 becomes

$$\dot{S} = \mu(1 - x) - \beta S I - \mu S, \tag{5.32}$$

$$\dot{I} = \beta S I - \gamma I - \mu I, \tag{5.33}$$

$$\dot{R} = \gamma I - \mu R + \mu x. \tag{5.34}$$

Parents who choose not to vaccinate their newborn have a negative payoff because their child might get the disease. This payoff is proportional to the number of infectives in the population. It is therefore $-mI$ where m is a positive constant.

On the other hand, parents who choose to vaccinate their newborn have a negative payoff due to the possibility of side effects from the vaccine. This payoff is $-k$ where k is a positive constant.

Therefore

- payoff of vaccinating minus payoff of not vaccinating equals $-k + mI$.

For vaccines in current use, k is much smaller than m. Thus when I is not too low ($I > \frac{k}{m}$), vaccinating has a higher payoff, but when I is very low ($I < \frac{k}{m}$), as it will be if the vaccination program has been in effect for a long time, then not vaccinating has a higher payoff.

Assuming imitation dynamics, we have

$$\dot{x} = \rho x(1-x)(-k+mI)$$

with ρ a positive constant. When vaccinating gives a higher payoff, x increases; when vaccinating gives a lower payoff, x decreases.

The complete system is

$$\dot{S} = \mu(1-x) - \beta SI - \mu S, \tag{5.35}$$
$$\dot{I} = \beta SI - \gamma I - \mu I, \tag{5.36}$$
$$\dot{R} = \gamma I - \mu R + \mu x, \tag{5.37}$$
$$\dot{x} = \rho x(1-x)(-k+mI). \tag{5.38}$$

1. Show that if $S + I + R = 1$, then $\dot{S} + \dot{I} + \dot{R} = 0$. Therefore we shall ignore the $\dot{R}$ equation and consider the reduced system

$$\dot{S} = \mu(1-x) - \beta SI - \mu S, \tag{5.39}$$
$$\dot{I} = \beta SI - \gamma I - \mu I, \tag{5.40}$$
$$\dot{x} = \rho x(1-x)(-k+mI). \tag{5.41}$$

   The state space is the prism $\mathcal{P}$.

2. Find all the equilibria. Answer:

$$(S, I, x) = (1, 0, 0),\ (S_*, I_*, 0),\ (0, 0, 1),\ (S_*, \frac{k}{m}, x_*),$$

   with

$$S_* = \frac{\gamma+\mu}{\beta},\ I_* = \mu\left(\frac{1}{\gamma+\mu} - \frac{1}{\beta}\right),\ x_* = 1 - \left(\frac{\beta k}{\mu m} + 1\right) S_*$$

Because of the assumption that $\frac{\gamma+\mu}{\beta} < 1$, S_* and I_* are strictly between 0 and 1; (S_*, I_*) is the endemic equilibrium found in Problem 3.2. We shall also assume that x_* is strictly between 0 and 1. This assumption is reasonable since, as we mentioned, $\frac{k}{m}$ is typically very small.

3. Use the methods of this course to learn whatever additional information you can about the dynamics of this model. The planes $x = 0$ and $x = 1$ are invariant. Can you use linearization and planar theory to determine the flow on these planes? Does linearization at the equilibria help you to understand the flow in $0 < x < 1$? Is it appropriate to view this system as a fast-slow system, and does that help?

A Solutions to problems

A.1 Solution for Chapter 1

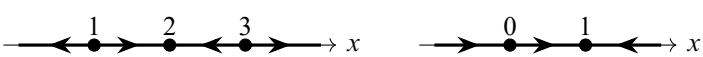


Figure A.1: Phase portraits for Problem 1.1.

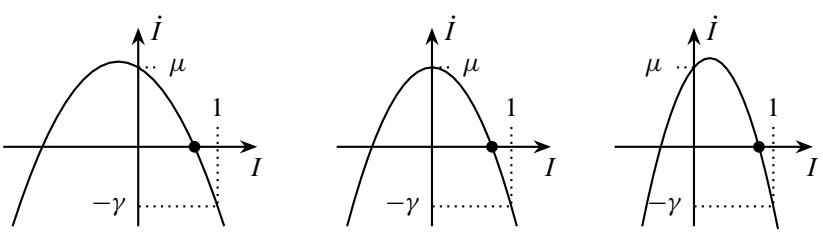


Figure A.2: Problem 1.6 part 3

Solution 1.1

See Figure A.1.

Solution 1.2

$$\dot{s} = -\frac{\beta}{N} si + \gamma i \implies N\dot{S} = -\frac{\beta}{N}(NS)(NI) + \gamma(NI) \implies \dot{S} = -\beta SI + \gamma I.$$

Solution 1.3

$$\int_0^\infty \gamma e^{-\gamma t}\, dt = \lim_{N\to\infty} \int_0^N \gamma e^{-\gamma t}\, dt.$$
$$\int_0^N \gamma e^{-\gamma t}\, dt = -e^{-\gamma t}\,\Big]_0^N = -e^{-\gamma N} + 1.$$
$$\int_0^\infty \gamma e^{-\gamma t}\, dt = \lim_{N\to\infty} \left(-e^{-\gamma N} + 1\right) = 1.$$

Solution 1.4

From an integral table or integration by parts, $\int \gamma t e^{-\gamma t}\, dt = -(\gamma^{-1} + t)e^{-\gamma t}$. Therefore:

$$\int_0^\infty \gamma t e^{-\gamma t}\, dt = \lim_{N\to\infty} \int_0^N \gamma t e^{-\gamma t}\, dt.$$
$$\int_0^N \gamma t e^{-\gamma t}\, dt = -(\gamma^{-1} + t)e^{-\gamma t}\,\Big]_0^N = -(\gamma^{-1} + N)e^{-\gamma N} + \gamma^{-1}.$$
$$\int_0^\infty \gamma e^{-\gamma t}\, dt = \lim_{N\to\infty} \left(-(\gamma^{-1} + N)e^{-\gamma N} + \gamma^{-1}\right) = \gamma^{-1}.$$

Solution 1.5

Since $\frac{\partial}{\partial\beta}\left(1 - \frac{\gamma}{\beta}\right) = \frac{\gamma}{\beta^2} > 0$, if we can reduce β, we will reduce the endemic equilibrium. (You can also see this just by looking at the equation for the endemic equilibrium.) We can reduce β by reducing the frequency of contacts that infectives have with susceptibles, or by reducing the probability of transmission of the

disease when a contact occurs. The first can be reduced by quarrantine of infectives, encouraging social distancing, closure of businesses, schools, etc., lockdowns of the population, and similar measures. The second can be reduced by encouraging mask wearing. You may have additional ideas.

Solution 1.6

1. $\dot{I} = \beta SI - \gamma I + \mu S = \beta(1-I)I - \gamma I + \mu(1-I) = \beta I - \beta I^2 - \gamma I + \mu - \mu I = \mu + (\beta - \gamma - \mu)I - \beta I^2$.

2. When $I = 0$, $\dot{I} = \mu$, which is positive. When $I = 1$, $\dot{I} = \mu + (\beta - \gamma - \mu) - \beta = -\gamma$, which is negative.

3. The apex of the parabola can have I negative, $I = 0$, or I positive. Either way we have essentially the same phase portrait in $0 \leqslant I \leqslant 1$.

See Figure A.2.

Solution 1.7

1.

$$\frac{dI}{dt} = I(1-I)$$
$$\frac{dI}{I(1-I)} = dt \quad \text{Use partial fractions:}$$
$$\frac{1}{I}dI + \frac{2}{1-2I}dI = dt$$

Integrate. Since we assumed $0 < I < \frac{1}{2}$, we don't need absolute value signs when we take the antiderivatives on the left.

$$\ln I - \ln(1-2I) = t + C$$
$$\ln \frac{1}{1-2I} = t + C$$

2. Exponentiate both sides:

$$\frac{1}{1-2I} = Ae^t. \quad A \text{ is positive since } A = e^C. \text{ Solve for } I:$$
$$I = Ae^t - 2Ae^t I$$
$$(2Ae^t + 1)I = Ae^t$$
$$I = \frac{Ae^t}{2Ae^t + 1} = \frac{A}{2A + e^{-t}} \quad \text{(Multiplied top and bottom by } e^{-t}.)$$

3. Check that $I = \frac{A}{2A+e^{-t}}$ is a solution of $\dot{I} = I(1-2I)$ for any A:

$$\dot{I} = -\frac{A}{(2A+e^{-t})^2} \cdot -e^{-t} = \frac{Ae^{-t}}{(2A+e^{-t})^2}$$
$$I(1-I) = \frac{A}{2A+e^{-t}}\left(1 - 2\frac{A}{2A+e^{-t}}\right) = \frac{A}{2A+e^{-t}} \cdot \frac{2A+e^{-t}-2A}{2A+e^{-t}}$$
$$= \frac{Ae^{-t}}{(2A+e^{-t})^2}$$

4. Suppose $I(0) = 1$. Determine A:

$$1 = \frac{A}{2A+1} \implies 2A+1 = A \implies A = -1.$$

So the solution of $\dot{I} = I(1-2I)$, $I(0) = 1$ is $I(t) = \frac{-1}{-2+e^{-t}}$. The numerator is 0 when

$$-2 + e^{-t} = 0 \implies e^{-t} = 2 \implies -t = \ln 2 \implies t = -\ln 2.$$

The solution is only defined on the interval $-\ln 2 < t < \infty$. As $t \to \infty$, $I(t) \to \frac{1}{2}$. As $t \to -\ln 2$ from the right, $I(t) \to \infty$.

A.2 Solution for Chapter 2

Solution 2.1

1. The maximum value of I occurs where $\frac{dI}{dS} = 0$. $\frac{dI}{dS} = -1 + \frac{\gamma}{\beta S} = 0 \implies S = \frac{\gamma}{\beta}$. (We knew this without doing the calculation.) Therefore

$$I_{\max} = I\left(\frac{\gamma}{\beta}\right) = -\left(\frac{\gamma}{\beta} - S_-\right) + \frac{\gamma}{\beta}\left(\ln\frac{\gamma}{\beta} - \ln S_-\right).$$

2. (a) $\frac{\partial I_{\max}}{\partial S_-} = 1 - \frac{\gamma}{\beta S_-}$. If $S_- > \frac{\gamma}{\beta}$ then $\beta S_- > \gamma$, so $\frac{\gamma}{\beta S_-} < 1$, so $1 - \frac{\gamma}{\beta S_-} > 0$. Thus $\frac{\partial I_{\max}}{\partial S_-} > 0$. It makes sense that if the susceptible population fraction as the start of the epidemic is larger, then the maximum number of infected would be larger. Also, we see in Figure 2.3 of the text that if S_- is larger, then the maximum value of I is larger.

   (b) $\frac{\partial I_{\max}}{\partial \beta} = \frac{\gamma}{\beta^2} - \frac{\gamma}{\beta^2}\left(\ln\frac{\gamma}{\beta} - \ln S_-\right) + \frac{\gamma}{\beta}\cdot\frac{\beta}{\gamma}\cdot - \frac{\gamma}{\beta^2} = -\frac{\gamma}{\beta^2}\left(\ln\frac{\gamma}{\beta} - \ln S_-\right) > 0$ because $S_- > \frac{\gamma}{\beta}$. If β increases, the disease is more easily spread. It makes sense that this would lead to an increase in the maximum value of I.

   (c) $\frac{\partial I_{\max}}{\partial \gamma} = -\frac{1}{\beta} + \frac{1}{\beta}\left(\ln\frac{\gamma}{\beta} - \ln S_-\right) + \frac{\gamma}{\beta}\cdot\frac{\beta}{\gamma}\cdot\frac{1}{\beta} = \frac{1}{\beta}\left(\ln\frac{\gamma}{\beta} - \ln S_-\right) < 0$ because $S_- > \frac{\gamma}{\beta}$. If γ increases, then $\frac{1}{\gamma}$, the mean length of time that an infective is ill and can infect others, decreases. It makes sense that this would lead to a decrease in the maximum value of I.

Solution 2.2

1. Solve the system of equations $-x = 0,\ -2y = 0$. You get $(x, y) = (0, 0)$.

2. $\frac{dy}{dx} = \frac{2y}{x} \implies \frac{1}{y}dy = \frac{2}{x}dx \implies \ln y = 2\ln x + C \implies y = Ax^2$.

3. Use Theorem 2.1. Let $V(x, y) = x$. Then $\nabla V(x, y) = (1, 0) \neq (0, 0)$. Also, when $V(x, y) = 0$, we have $\nabla V(0, y)\cdot f(0, y) = (1, 0)\cdot(0, -2y) = 0$. Therefore by Theorem 2.1, the line $x = 0$ is invariant. (A quicker way to do this problem, which involves doing the second check but skipping the first, is just to note that $x = 0 \implies \dot{x} = 0$.)

4. See Figure 3.4.

5. Three.

Solution 2.3

1. Solve the system of equations $-\alpha x - \omega y = 0,\ \omega x - \alpha y = 0$. You get $(x, y) = (0, 0)$.

2. Instead of using (2.6), we'll just use the formula for $\dot{V}$ that precedes it (the chain rule):

$$\dot{V} = \frac{\partial V}{\partial x}\dot{x} + \frac{\partial V}{\partial y}\dot{y} = 2x(-\alpha x - \omega y) + 2y(\omega x - \alpha y) = -2\alpha(x^2 + y^2)$$
$$= -2\alpha V.$$

3.

$$\dot{\theta} = \frac{\partial \theta}{\partial x}\dot{x} + \frac{\partial \theta}{\partial y}\dot{y} =$$
$$= \frac{1}{1 + \left(\frac{y}{x}\right)^2}\left(-\frac{y}{x^2}\right)(-\alpha x - \omega y) + \frac{1}{1 + \left(\frac{y}{x}\right)^2}\left(\frac{1}{x}\right)(\omega x - \alpha y) =$$
$$= -\frac{y}{x^2 + y^2}(-\alpha x - \omega y) + \frac{x}{x^2 + y^2}(\omega x - \alpha y) =$$
$$= \frac{\omega(x^2 + y^2)}{x^2 + y^2} = \omega.$$

4. See Figure 3.5.

A.3 Solution for Chapter 3

Solution 3.1

1.

$$\begin{pmatrix} \frac{\partial \dot{S}}{\partial S} & \frac{\partial \dot{S}}{\partial I} \\ \frac{\partial \dot{I}}{\partial S} & \frac{\partial \dot{I}}{\partial I} \end{pmatrix} = \begin{pmatrix} -\beta I & -\beta S \\ \beta I & \beta S - \gamma \end{pmatrix}$$

2. At an equilibrium $(S, 0)$,

$$\begin{pmatrix} \frac{\partial \dot{S}}{\partial S} & \frac{\partial \dot{S}}{\partial I} \\ \frac{\partial \dot{I}}{\partial S} & \frac{\partial \dot{I}}{\partial I} \end{pmatrix} = \begin{pmatrix} 0 & -\beta S \\ 0 & \beta S - \gamma \end{pmatrix}$$

3. Eigenvalues:

$$\begin{vmatrix} -\lambda & -\beta S \\ 0 & \beta S - \gamma - \lambda \end{vmatrix} = -\lambda(\beta S - \gamma - \lambda) = 0 \implies \lambda = 0,\ \beta S - \gamma.$$

4. No. All have at least one eigenvalue that is 0.

Solution 3.2

1. $\dot{S} + \dot{I} + \dot{R} = (\mu - \beta SI - \mu S) + (\beta SI - \gamma I - \mu I) + (\gamma I - \mu R) = \mu - \mu(S + I + R) = 0$ if $S + I + R = 1$. Let $V(S, I, R) = S + I + R$. $\nabla V = (1\ 1\ 1)$. $\dot{V} = \dot{S} + \dot{I} + \dot{R} = 0$ if $S + I + R = 1$. By Theorem 2.1 the set $S + I + R = 1$ is invariant.

2. To find the equilibria we must solve simultaneously two equations:

$$\begin{aligned} \mu - \beta SI - \mu S &= 0 \\ \beta SI - \gamma I - \mu I &= 0 \end{aligned}$$

The second equation factors: $(\beta S - \gamma - \mu)I = 0$. Therefore $S = \frac{\gamma+\mu}{\beta}$ or $I = 0$. In the first case, the first equation yields $I = \frac{\mu}{\gamma+\mu} - \frac{\mu}{\beta}$. In the second case, the first equation yields $S = 1$.

3. $(S_*, I_*) = \left(\frac{\gamma+\mu}{\beta}, \mu\left(\frac{1}{\gamma+\mu} - \frac{1}{\beta}\right)\right)$. If $\frac{\gamma+\mu}{\beta} > 1$, the $S_* > 1$, so $(S_*, I_*) \notin \mathcal{T}$. Now suppose $\frac{\gamma+\mu}{\beta} < 1$. To show that $(S_*, I_*) \in \mathcal{T}$, we must show that $S_* \geqslant 0$, $I_* \geqslant 0$, and $S_* + I_* \leqslant 1$. (To prove something, it often helps to go back to the definition to see what you need to show!) $S_* > 0$ is obvious since γ, μ and β are positive. Also,

$$\frac{\gamma+\mu}{\beta} < 1 \implies \gamma + \mu < \beta \implies \frac{1}{\gamma+\mu} > \frac{1}{\beta} \implies \frac{1}{\gamma+\mu} - \frac{1}{\beta} > 0.$$

Therefore $I_* > 0$. Finally,

$$\begin{aligned} S_* + I_* = \frac{\gamma+\mu}{\beta} + \frac{\mu}{\gamma+\mu} - \frac{\mu}{\beta} = \frac{\gamma}{\beta} + \frac{\mu}{\gamma+\mu} = \\ = \frac{\gamma(\gamma+\mu) + \mu\beta}{\beta(\gamma+\mu)} < \frac{\gamma\beta + \mu\beta}{\beta(\gamma+\mu)} = 1. \end{aligned}$$

4. $\begin{pmatrix} \frac{\partial \dot{S}}{\partial S} & \frac{\partial \dot{S}}{\partial I} \\ \frac{\partial \dot{I}}{\partial S} & \frac{\partial \dot{I}}{\partial I} \end{pmatrix} = \begin{pmatrix} -\beta I - \mu & -\beta S \\ \beta I & \beta S - \gamma - \mu \end{pmatrix}$

5. At $(1, 0)$, the linearization matrix is

$$\begin{pmatrix} -\mu & -\beta \\ 0 & \beta - \gamma - \mu \end{pmatrix}$$

Eigenvalues:

$$\begin{vmatrix} -\mu-\lambda & -\beta \\ 0 & \beta-\gamma-\mu-\lambda \end{vmatrix} = (-\mu-\lambda)(\beta-\gamma-\mu-\lambda) = 0 \implies \lambda = -\mu,\ \beta-\gamma-\mu.$$

The eigenvalue $-\mu$ is always negative. If $\frac{\gamma+\mu}{\beta} < 1$, then $\gamma+\mu < \beta$, so $\beta-\gamma-\mu = \beta-(\gamma+\mu) > 0$. Therefore $(1,0)$ is a saddle. If $\frac{\gamma+\mu}{\beta} > 1$, then $\gamma+\mu > \beta$, so $\beta-\gamma-\mu = \beta-(\gamma+\mu) < 0$. Therefore $(1,0)$ is an attractor.

6. Consider the matrix

$$\begin{pmatrix} -\beta I_* - \mu & -\beta S_* \\ \beta I_* & \beta S_* - \gamma - \mu \end{pmatrix}.$$

$$\begin{aligned}
\text{Trace} &= -\beta I_* - \mu + \beta S_* - \gamma - \mu \\
&= -\beta\mu\left(\frac{1}{\gamma+\mu} - \frac{1}{\beta}\right) - \mu + \beta\frac{\gamma+\mu}{\beta} - \gamma - \mu \\
&= -\frac{\beta\mu}{\gamma+\mu} + \mu - \mu + \gamma + \mu - \gamma - \mu \\
&= -\frac{\beta\mu}{\gamma+\mu} < 0.
\end{aligned}$$

$$\begin{aligned}
\text{Determinant} &= (-\beta I_* - \mu)(\beta S_* - \gamma - \mu) - \beta I_*(-\beta S_*) \\
&= -\beta^2 I_* S_* + \beta\gamma I_* + \beta\mu I_* - \beta\mu S_* + \gamma\mu + \mu^2 + \beta^2 I_* S_* \\
&= \beta(\gamma+\mu)I_* - \beta\mu S_* + \gamma\mu + \mu^2 \\
&= \beta(\gamma+\mu)\mu\left(\frac{1}{\gamma+\mu} - \frac{1}{\beta}\right) - \beta\mu\frac{\gamma+\mu}{\beta} + \gamma\mu + \mu^2 \\
&= \beta\mu - (\gamma+\mu)\mu - \mu(\gamma+\mu) + \gamma\mu + \mu^2 \\
&= \beta\mu - \gamma\mu - \mu^2 - \gamma\mu - \mu^2 + \gamma\mu + \mu^2 \\
&= \mu(\beta-\gamma-\mu) > 0 \text{ because } \beta > \gamma+\mu.
\end{aligned}$$

7. If we multiply the system (3.26)–(3.27) by $\frac{1}{I}$, which is positive in $I > 0$, we obtain

$$\begin{aligned}
\dot{S} &= \frac{\mu}{I} - \beta S - \mu\frac{S}{I}, \\
\dot{I} &= \beta S - \gamma - \mu.
\end{aligned}$$

The divergence of this system is

$$\frac{\partial \dot{S}}{\partial S} + \frac{\partial \dot{I}}{\partial I} = -\beta - \frac{\mu}{I},$$

which is negative in $I > 0$. By Dulac's Criterion (Corollary 3.2) the system (3.26)–(3.27) has no closed orbits, separatrix cycles or graphics in $I > 0$. Arguing as in Section 3.7, we draw the same conclusions as in that section.

Solution 3.3

1.

$$\begin{aligned} f(I) &= (\beta - \gamma - \mu)I - \beta I^2 \\ f'(I) &= \beta - \gamma - \mu - 2\beta I \\ f'(0) &= \beta - \gamma - \mu \end{aligned}$$

We assumed $\frac{\beta}{\gamma+\mu} < 1$. This implies $\beta < \gamma+\mu$, which implies $\beta-\gamma-\mu < 0$. Therefore $f'(0) < 0$, so 0 is an attractor.

2. $i(t) = I(t)N(t) \approx I(0)e^{(\beta-\gamma-\mu)t}N(0)e^{(\mu-\nu)t} = I(0)N(0)e^{(\beta-\gamma-\nu)t}$. Therefore $i(t)$ is increasing if $\beta - \gamma - \nu > 0$, or $\beta > \gamma + \nu$. We assumed $\beta < \gamma + \mu$. Therefore the condition we need is $\gamma + \nu < \beta < \gamma + \mu$, or $\nu < \beta - \gamma < \mu$.

A.4 Solution for Chapter 4

Solution 4.1

First, let's make sure we're comparing apples to apples and not apples to oranges. In Figure 4.2a, $I(30)$ is the population fraction infected at $t = 30$. In the Stevens chart, if we assume that people stay infected for 10 days, then the number infected on day 30 would be the total number of cases as of day 30 minus the total number of cases as of day 20, or $12647 - 2244 = 10403$. Thus $I(30) = 10403/66,650,000 = .00016$, or .016%. This makes the discrepancy a little worse:

There are many possible sources for the discrepancy, but the most important is probably the difference in initial conditions. In the Stevens chart, $t = 0$ would

be 22/02/2020. A reasonable estimate for cases to date on that day, from the data presented, is nine. If we assume all those people were still infective on that day, we would have $I(0) = 9/66,650,000 = .00000014$.

In Figure 4.2a, $S(0) = .9999$. $I(0)$ can be calculated from the equation $I = -(S-1) + \frac{1}{3}\ln S$:

$$I(0) = -(.9999 - 1) + \frac{1}{3}\ln .9999 = .000067.$$

Thus $I(0)$ in Figure 4.2a is almost 500 times greater than $I(0)$ in the Stevens chart, while $I(30)$ in Figure 4.2a is almost 200 times greater than $I(30)$ in the Stevens chart.

These ratios differ by less than one order of magnitude. Thus most of the discrepancy is explained by the difference in initial conditions.

Solution 4.2

Sources of error: The UK data are incomplete, since they only include hospitalized patients who were found to have covid. We are using an SIR model and assuming that the total number of infectives increases at the same exponential rate as the hospitalized infectives. It is not clear that one can do better with the available data, but you may have some ideas.

Solution 4.3

1. Omitted.

2. Expand the determinant by the last column:

$$\begin{vmatrix} -(\gamma_E+\mu) & \beta S & 0 \\ \gamma_E & -(\gamma_I+\mu) & 0 \\ 0 & \gamma_I & -\mu \end{vmatrix} = -\mu\big((\gamma_E+\mu)(\gamma_I+\mu) - \gamma_E\beta S\big).$$

   Since $\mu > 0$, the determinant is nonzero if and only if $(\gamma_E+\mu)(\gamma_I+\mu) - \gamma_E\beta S \neq 0$. When the determinant is nonzero, the only solution of the linear equation is $E = I = R = 0$.

3. The determinant is zero only for $S = \frac{(\gamma_E+\mu)(\gamma_I+\mu)}{\beta\gamma_E}$ which is greater than 1. So the determinant does not vanishes on $\mathcal{S}$ and the right hand sides of (4.23)–(4.25) are 0 if and only if $E = I = R = 0$. Since $I = 0$, the right

hand side of (4.22) is 0 if and only if $S = 1$. We conclude that the only equilibrium of (4.22)–(4.25) in $\mathcal{S}$ is $(1, 0, 0, 0)$.

4. The linearization matrix is

$$\begin{pmatrix} \frac{\partial \dot{S}}{\partial S} & \frac{\partial \dot{S}}{\partial E} & \frac{\partial \dot{S}}{\partial I} & \frac{\partial \dot{S}}{\partial R} \\ \frac{\partial \dot{E}}{\partial S} & \frac{\partial \dot{E}}{\partial E} & \frac{\partial \dot{E}}{\partial I} & \frac{\partial \dot{E}}{\partial R} \\ \frac{\partial \dot{I}}{\partial S} & \frac{\partial \dot{I}}{\partial E} & \frac{\partial \dot{I}}{\partial I} & \frac{\partial \dot{I}}{\partial R} \\ \frac{\partial \dot{R}}{\partial S} & \frac{\partial \dot{R}}{\partial E} & \frac{\partial \dot{R}}{\partial I} & \frac{\partial \dot{R}}{\partial R} \end{pmatrix} = \begin{pmatrix} -\beta I - \mu & 0 & -\beta S & 0 \\ \beta I & -\gamma_E - \mu & \beta S & 0 \\ 0 & \gamma_E & -\gamma_I - \mu & 0 \\ 0 & 0 & \gamma_I & -\mu \end{pmatrix}.$$

At $(S, E, I, R) = (1, 0, 0, 0)$ the linearization matrix becomes

$$\begin{pmatrix} -\mu & 0 & -\beta & 0 \\ 0 & -(\gamma_E + \mu) & \beta & 0 \\ 0 & \gamma_E & -(\gamma_I + \mu) & 0 \\ 0 & 0 & \gamma_I & -\mu \end{pmatrix}.$$

5. To find the eigenvalues of the previous matrix we write

$$\begin{aligned} &\begin{vmatrix} -\mu - \lambda & 0 & -\beta & 0 \\ 0 & -(\gamma_E + \mu) - \lambda & \beta & 0 \\ 0 & \gamma_E & -(\gamma_I + \mu) - \lambda & 0 \\ 0 & 0 & \gamma_I & -\mu - \lambda \end{vmatrix} \\ &\quad = (-\mu - \lambda) \begin{vmatrix} -(\gamma_E + \mu) - \lambda & \beta & 0 \\ \gamma_E & -(\gamma_I + \mu) - \lambda & 0 \\ 0 & \gamma_I & -\mu - \lambda \end{vmatrix} \\ &\quad = (-\mu - \lambda)^2 \begin{vmatrix} -(\gamma_E + \mu) - \lambda & \beta \\ \gamma_E & -(\gamma_I + \mu) - \lambda \end{vmatrix} = 0. \end{aligned}$$

Thus the eigenvalues are $-\mu$, $-\mu$, and the eigenvalues of

$$K = \begin{pmatrix} -(\gamma_E + \mu) & \beta \\ \gamma_E & -(\gamma_I + \mu) \end{pmatrix}.$$

6. $\det K = (\gamma_E + \mu)(\gamma_I + \mu) - \beta\gamma_E$.

Suppose $\det K < 0$. Then K has one negative eigenvalue and one positive eigenvalue. $\det K < 0$ is equivalent to $\frac{\beta\gamma_E}{(\gamma_E+\mu)(\gamma_I+\mu)} > 1$.

Suppose $\det K > 0$ Since the trace of K is $-(\gamma_E + \gamma_I + 2\mu)$, which is negative, in this case there are two eigenvalues with negative real part. $\det K > 0$ is equivalent to $\frac{\beta\gamma_E}{(\gamma_E+\mu)(\gamma_I+\mu)} < 1$.

7.
$$K = F - V = \begin{pmatrix} 0 & \beta \\ 0 & 0 \end{pmatrix} - \begin{pmatrix} \gamma_E + \mu & 0 \\ -\gamma_E & \gamma_I + \mu \end{pmatrix}.$$

Using the formula $\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$, we find that

$$V^{-1} = \frac{1}{(\gamma_E + \mu)(\gamma_I + \mu)} \begin{pmatrix} \gamma_I + \mu & 0 \\ \gamma_E & \gamma_E + \mu \end{pmatrix}$$

Then matrix multiplication yields the next generation matrix

$$FV^{-1} = \frac{1}{(\gamma_E + \mu)(\gamma_I + \mu)} \begin{pmatrix} \beta\gamma_E & \beta(\gamma_E + \mu) \\ 0 & 0 \end{pmatrix}$$

The eigenvalues are $\frac{\beta\gamma_E}{(\gamma_E+\mu)(\gamma_I+\mu)}$ and 0. The largest is $\frac{\beta\gamma_E}{(\gamma_E+\mu)(\gamma_I+\mu)}$, so that is R_0. According to a theorem mentioned toward the end of Section 4.6, when the largest eigenvalue of FV^{-1} is greater than 1, K has an eigenvalue with positive real part, and when the largest eigenvalue of FV^{-1} is less than 1, all eigenvalues of K have negative real part. This is consistent with part 6.

Solution 4.4

1. Adding (4.30) and (4.32) we get $0 = -\gamma_1 I_1 = 0$, so $I_1 = 0$. Adding (4.31) and (4.33) we get $I_2 = 0$. If $I_1 = 0$ and $I_2 = 0$ we have an equilibrium, so $(S_1, S_2, 0, 0, R)$ are the equilibria.

$$\mathcal{F} = \begin{pmatrix} \beta_{11}S_1I_1 + \beta_{21}S_1I_2 \\ \beta_{12}S_2I_1 + \beta_{22}S_2I_2 \end{pmatrix}, \quad F = \begin{pmatrix} \beta_{11}S_1 & \beta_{21}S_1 \\ \beta_{12}S_2 & \beta_{22}S_2 \end{pmatrix}.$$

$$\mathcal{V} = \begin{pmatrix} \gamma_1 I_1 \\ \gamma_2 I_2 \end{pmatrix}, \quad V = \begin{pmatrix} \gamma_1 & 0 \\ 0 & \gamma_2 \end{pmatrix}, \quad \text{so } V^{-1} = \begin{pmatrix} \frac{1}{\gamma_1} & 0 \\ 0 & \frac{1}{\gamma_2} \end{pmatrix}.$$

2. The next generation matrix is

$$FV^{-1} = \begin{pmatrix} \beta_{11}S_1 & \beta_{21}S_1 \\ \beta_{12}S_2 & \beta_{22}S_2 \end{pmatrix} \begin{pmatrix} \frac{1}{\gamma_1} & 0 \\ 0 & \frac{1}{\gamma_2} \end{pmatrix} = \begin{pmatrix} \dfrac{\beta_{11}S_1}{\gamma_1} & \dfrac{\beta_{21}S_1}{\gamma_2} \\ \dfrac{\beta_{12}S_2}{\gamma_1} & \dfrac{\beta_{22}S_2}{\gamma_2} \end{pmatrix}.$$

3. Let's find R_0 for $S_1 = 1$ and $S_2 = 0$.

$$FV^{-1} = \begin{pmatrix} \dfrac{\beta_{11}S_1}{\gamma_1} & \dfrac{\beta_{21}S_1}{\gamma_2} \\ \dfrac{\beta_{12}S_2}{\gamma_1} & \dfrac{\beta_{22}S_2}{\gamma_2} \end{pmatrix} = \begin{pmatrix} \dfrac{\beta_{11}}{\gamma_1} & \dfrac{\beta_{21}}{\gamma_2} \\ 0 & 0 \end{pmatrix}.$$

Therefore FV^{-1} has eigenvalues 0 and $R_0 = \dfrac{\beta_{11}}{\gamma_1}$. The value for R_0 makes sense since it is the value of R_0 for a standard SIR model in which the entire population is of subgroup 1.

4. Let's find R_0: $FV^{-1} = \begin{pmatrix} \dfrac{\beta S_1}{\gamma} & \dfrac{\beta S_1}{\gamma} \\ \dfrac{\beta(1-S_1)}{\gamma} & \dfrac{\beta(1-S_1)}{\gamma} \end{pmatrix}$.

The sum of the eigenvalues is the trace, $\dfrac{\beta}{\gamma}$, and the product is the determinant, zero, so the greater eigenvalue is $R_0 = \dfrac{\beta}{\gamma}$. R_0 does not depend on the initial population's composition. The value of R_0 makes sense since the entire population is of the same type.

5. $FV^{-1} = \begin{pmatrix} \dfrac{\beta_{11}S_1}{\gamma_1} & \dfrac{\beta_{21}S_1}{\gamma_2} \\ \dfrac{\beta_{12}S_2}{\gamma_1} & \dfrac{\beta_{22}S_2}{\gamma_2} \end{pmatrix} = \dfrac{1}{2}\begin{pmatrix} 2 & 3 \\ 3 & 10 \end{pmatrix}$.

The matrix $\begin{pmatrix} 2 & 3 \\ 3 & 10 \end{pmatrix}$ has the characteristic equation

$$\lambda^2 - 12\lambda + 11 = (\lambda - 1)(\lambda - 11) = 0,$$

so the eigenvalues are 1 and 11. The matrix FV^{-1} has eigenvalues $1/2$ and $R_0 = 11/2$.

The eigenvectors corresponding to R_0 are (a_1, a_2) where

$$\frac{1}{2}\begin{pmatrix} 2 & 3 \\ 3 & 10 \end{pmatrix}\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \frac{11}{2}\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$$
$$\Leftrightarrow \begin{pmatrix} 2-11 & 3 \\ 3 & 10-11 \end{pmatrix}\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Leftrightarrow a_2 = 3a_1.$$

At the beginning of the epidemic, the fraction of infectives I_2 grows three times as fast as the fraction of infectives I_1.

A.5 Solution for Chapter 5

Solution 5.1

$$\dot{x} = \frac{d}{dt}\left(\frac{S_n}{S_n + S_c}\right) = \frac{(S_n + S_c)\dot{S}_n - S_n(\dot{S}_n + \dot{S}_c)}{(S_n + S_c)^2} = \frac{S_c\dot{S}_n - S_n\dot{S}_c}{(S_n + S_c)^2}$$
$$= \frac{-S_c\beta_n S_n I + S_n\beta_c S_c I}{(S_n + S_c)^2} = (\beta_c - \beta_n)\frac{S_n S_c}{(S_n + S_c)^2} I$$
$$= (\beta_c - \beta_n)\frac{S_n}{S_n + S_c}\frac{S_c}{S_n + S_c} I = (\beta_c - \beta_n)x(1-x)I.$$

Solution 5.2

1.
$$\dot{S} + \dot{I} + \dot{R} = \mu(1-x) - \beta SI - \mu S + \beta SI - \gamma I - \mu I + \gamma I - \mu R + \mu x$$
$$= \mu(1-x) - \mu S - \mu I - \mu R + \mu x = \mu - \mu(S+I+R) = 0 \text{ if } S+I+R = 1.$$

2. To find the equilibria we must solve the system of equations
$$\mu(1-x) - \beta SI - \mu S = 0,$$
$$\beta SI - \gamma I - \mu I = 0,$$
$$\rho x(1-x)(-k + mI) = 0.$$

The second equation factors into $(\beta S-\gamma-\mu)I = 0$, so $S = \frac{\gamma+\mu}{\beta}$ or $I = 0$.

The third equation yields $x = 0$ or $x = 1$ or $I = \frac{k}{m}$.

We can pair each of the two possibilities from the second equation with each of the three possibilities from the third equation, which yields six pairs. However, we cannot pair $I = 0$ with $I = \frac{k}{m}$, so there are just five pairs:

$$(S, x) = \left(\frac{\gamma+\mu}{\beta}, 0\right),\ (S, x) = \left(\frac{\gamma+\mu}{\beta}, 1\right),\ (S, I) = \left(\frac{\gamma+\mu}{\beta}, \frac{k}{m}\right),$$
$$(I, x) = (0, 0),\ (I, x) = (0, 1).$$

Substitute each of these five pairs into the first equation and solve for the remaining variable. It turns out that $(S, x) = \left(\frac{\gamma+\mu}{\beta}, 1\right)$ cannot be used since it leads to a negative value for I. We obtain the four equilibria given in the text.

3. In the invariant plane $x = 0$ (no vaccination) we just have the system studied in Problem 3.2. With the assumption that $\frac{\gamma+\mu}{\beta} < 1$, the interior equilibrium (S_*, I_*) is an attractor; in fact, it attracts all solutions in the plane $x = 0$ with $I > 0$. There is a second equilibrium at $(S, I) = (1, 0)$. It is a saddle that attracts solutions in $I = 0$.

   In the invariant plane $x = 1$ (all newborns are vaccinated) the system reduces to

$$\dot{S} = -\beta SI - \mu S,$$
$$\dot{I} = \beta SI - \gamma I - \mu I$$

   This system has one equilibrium, at $(0, 0)$. It is an attractor. You can use Dulac's Criterion, Corollary 3.2, (multiply by $\frac{1}{I}$) to show that all solutions in the plane $x = 1$ are attracted to $(0, 0)$. In other words, the vaccination program is completely successful: eventually there are no susceptibles and no infectives.

   Regarded as points in SIx-space, the equilibria just discussed are $(S_*, I_*, 0)$, $(1, 0, 0)$ and $(0, 0, 1)$. If you calculate the 3×3 linearization matrix at these equilibria, you will see that in addition to the eigenvalues they have as equilibria in the SI-plane, each has an additional eigenvalue $\rho(1 - 2x)(-k +$

mI). At $(S_*, I_*, 0)$ this eigenvalue is $\rho(-k + mI_*)$, which is positive if k is small. Similarly, at $(0, 0, 1)$ this eigenvalue is ρk, which is positive.

Interpretation: at the endemic equilibrium with no one vaccinating, the vaccination rate will increase; at the fully vaccinated equilibrium with no susceptibles and no infectives, the vaccination rate will fall.

Fast-slow structure: It is not clear that it is correct to assume that attitudes toward vaccination change on a faster time-scale than that of the disease. However, we will sketch the analysis on the assumption that they do.

Write $\rho = \frac{1}{\epsilon}$ with $\epsilon > 0$ a small number. Substitute into Equation (5.38) and multiply that equation by ϵ to obtain the slow system:

$$\begin{aligned}\dot{S} &= \mu(1-x) - \beta SI - \mu S,\\ \dot{I} &= \beta SI - \gamma I - \mu I,\\ \epsilon\dot{x} &= x(1-x)(-k + mI).\end{aligned}$$

Let $\tau = \frac{t}{\epsilon}$ (fast time), use prime to denote $\frac{d}{d\tau}$, and write the fast system:

$$\begin{aligned}S' &= \epsilon\big(\mu(1-x) - \beta SI - \mu S\big),\\ I' &= \epsilon\big(\beta SI - \gamma I - \mu I\big),\\ \dot{x}' &= x(1-x)(-k + mI).\end{aligned}$$

Slow limit system:

$$\begin{aligned}\dot{S} &= \mu(1-x) - \beta SI - \mu S,\\ \dot{I} &= \beta SI - \gamma I - \mu I,\\ 0 &= x(1-x)(-k + mI).\end{aligned}$$

Fast limit system:

$$\begin{aligned}S' &= 0,\\ I' &= 0,\\ \dot{x}' &= x(1-x)(-k + mI).\end{aligned}$$

The slow limit system makes sense on $x = 0$ and $x = 1$ (ignore $I = \frac{k}{m}$). On $x = 0$ we have the system studied in Problem 3.2; all solutions in $I > 0$

are attracted to the endemic equilibrium (S_*, I_*). We assume $\frac{k}{m} < I_*$. On $x = 1$ we have the system found in part 3 of this problem; all solutions are attracted to to the equilibrium $(S, I) = (0, 0)$.

Solutions of the fast limit system have S and I constant. For $I < \frac{k}{m}$, x decreases; for $I > \frac{k}{m}$, x increases. In other words, when the number of infectives is very small, the vaccination rate falls; otherwise it increases.

Singular solutions: We'll construct a solution that starts with $I > \frac{k}{m}$ and $0 < x < 1$.

(a) The first fast solution immediately goes to $x = 1$.

(b) The first slow solution is in $x = 1$ (everyone vaccinates). It heads toward $(S, I) = (0, 0)$. Attraction toward $x = 1$ accumulates until the solution passes $I = \frac{k}{m}$. Then repulsion begins to accumulate. Because the solution approaches an equilibrium in the repelling part of the plane, there will always be a time at which the entry-exit integral is 0.

(c) The second fast solution leads to a point in $x = 0$.

(d) The second slow solution is in $x = 0$ (no one vaccinates). It approaches the endemic equilibrium, which is in the repelling part of the plane. Therefore there will be a time when the entry-exit integral is 0.

(e) The third fast solution leads to a point in $x = 1$. The process repeats.

Bibliography

L. J. S. Allen (2008). "An introduction to stochastic epidemic models." In: *Mathematical epidemiology*. Vol. 1945. Lecture Notes in Math. Springer, Berlin, pp. 81–130. MR: 2428373. Zbl: 1206.92022 (cit. on p. viii).

C. Bauch and S. Bhattacharyya (2012). "Evolutionary game theory and social learning can determine how vaccine scares unfold." *PLoS Comput. Biol.* 8.4, e1002452, 12 pp. MR: 2930006 (cit. on p. 80).

C. Bauch, A. d'Onofrio, and P. Manfredi (2013). "Behavioral epidemiology of infectious diseases: an overview." In: *Modeling the interplay between human behavior and the spread of infectious diseases*. Ed. by P. Manfredi and A. d'Onofrio. Springer, pp. 1–19 (cit. on p. 92).

H. Berestycki, J.-M. Roquejoffre, and L. Rossi (2021). "Propagation of epidemics along lines with fast diffusion." *Bull. Math. Biol.* 83.1, Paper No. 2, 34 pp. MR: 4188239 (cit. on p. ix).

J. C. Blackwood and L. M. Childs (2018). "An introduction to compartmental modeling for the budding infectious disease modeler." *Lett. Biomath.* 5.1, pp. 195–221. MR: 3844698 (cit. on p. 55).

W. Booth (Mar. 17, 2020). *A chilling scientific paper helped upend U.S. and U.K. coronavirus strategies*. Washington Post. URL: https://www.washingtonpost.com/world/europe/a-chilling-scientific-paper-helped-upend-us-and-uk-coronavirus-strategies/2020/03/17/aaa84116-6851-11ea-b199-3a9799c54512_story.html (cit. on p. ix).

F. Brauer and C. Castillo-Chavez (2012). *Mathematical models in population biology and epidemiology*. Second ed. Vol. 40. Texts in Applied Mathematics. Springer, New York, pp. xxiv+508. MR: 3024808. Zbl: 1302.92001.

CDC (Feb. 23, 2021). *Malaria*. URL: https://www.cdc.gov/parasites/malaria/index.html (cit. on p. vii).

S. Chang et al. (2021). "Mobility network models of COVID-19 explain inequities and inform reopening." *Nature* 589.1, pp. 82–87 (cit. on p. ix).

N. M. Ferguson, D. A. T. Cummings, C. Fraser, J. Cajka, P. Cooley, and D. Burke (2006). "Strategies for mitigating an influenza pandemic." *Nature* 442, pp. 448–452 (cit. on pp. ix, 93).

N. M. Ferguson, D. Laydon, et al. (Mar. 16, 2020). "Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand." Imperial College COVID-19 Response Team Report Number 9 (cit. on p. ix).

J. Frith (2012). "The History of Plague–Part 1. The Three Great Pandemics." *J. of Military and Veterans' Health* 20, pp. 11–16 (cit. on p. vi).

A. d. C. Goulart (2005). "Revisiting the Spanish flu: the 1918 influenza pandemic in Rio de Janeiro." *História, Ciências, Saúde — Manguinhos* 12.1, pp. 1–41 (cit. on p. vii).

H. W. Hethcote (1978). "An immunization model for a heterogeneous population." *Theoret. Population Biol.* 14.3, pp. 338–349. MR: 520143. Zbl: 0392.92009 (cit. on p. x).

— (2000). "The mathematics of infectious diseases." *SIAM Rev.* 42.4, pp. 599–653. MR: 1814049. Zbl: 0993.92033.

W. O. Kermack and A. G. McKendrick (1927). "A contribution to the mathematical theory of epidemics." *Proc. Roy. Soc. London Ser. A* 115, pp. 700–721. Zbl: 53.0517.01 (cit. on pp. x, 28).

I. Z. Kiss, J. C. Miller, and P. L. Simon (2017). *Mathematics of epidemics on networks: from exact to approximate models*. Vol. 46. Interdisciplinary Applied Mathematics. Springer, Cham, pp. viii+413. MR: 3644065 (cit. on p. ix).

C. Kuehn (2015). *Multiple time scale dynamics*. Vol. 191. Applied Mathematical Sciences. Springer, pp. xiii+814 (cit. on p. 84).

C. N. Ngonghala, E. A. Iboi, and A. B. Gumel (2020). "Could masks curtail the post-lockdown resurgence of COVID-19 in the US?" *Math. Biosci.* 329, pp. 108452, 18. MR: 4143540. Zbl: 07264581 (cit. on p. 56).

L. Perko (2001). *Differential Equations and Dynamical Systems*. Vol. 7. Texts in Applied Mathematics. Springer, pp. xiv+555 (cit. on pp. v, 49, 50).

P. Poletti (2010). PhD thesis. University of Trento (cit. on p. 80).

P. Poletti, M. Ajelli, and S. Merler (2012). "Risk perception and effectiveness of uncoordinated behavioral responses in an emerging epidemic." *Math. Biosci.* 238, pp. 80–89 (cit. on p. 92).

D. G. Schaeffer and J. W. Cain (2016). *Ordinary differential equations: basics and beyond*. Vol. 65. Texts in Applied Mathematics. Springer, New York, pp. xxx+542. MR: 3561103 (cit. on p. v).

S. Schecter (2021). "Geometric Singular Perturbation Theory Analysis of an Epidemic Model with Spontaneous Human Behavioral Change." *J. Math. Biol.* 82 (cit. on pp. 80, 83, 91).

J. Sotomayor (1979). *Lições de Equações Diferenciais Ordinárias*. Vol. 11. Projeto Euclides. Instituto de Matemática Pura e Aplicada, Rio de Janeiro, pp. xvi+327. MR: 0651910 (cit. on p. v).

R. Stevens, R. Perera, C. Heneghan, R. Hobbs, and J. Oke (Sept. 23, 2020). *Exponential growth: what it is, why it matters, and how to spot it*. The Centre for Evidence-Based Medicine. URL: https://www.cebm.net/covid-19/exponential-growth-what-it-is-why-it-matters-and-how-to-spot-it (cit. on pp. 61, 74).

J. K. Taubenberger and D. M. Morens (2006). "1918 influenza: the mother of all pandemics." *Emerging Infectious Diseases* 12.1, pp. 15–22 (cit. on p. viii).

P. van den Driessche (2017). "Reproduction numbers of infectious disease models." *Infectious Disease Modeling* 2, pp. 288–303 (cit. on p. 55).

P. van den Driessche and J. Watmaugh (2002). "Reproduction numbers and subthreshold endemic equilibria for compartmental models of disease transmission." *Math. Biosci.* 180, pp. 29–48 (cit. on p. 68).

WHO (Feb. 5, 2021). *Cholera*. URL: https://www.who.int/news-room/fact-sheets/detail/cholera (cit. on p. vii).

Wikipedia (Mar. 29, 2021a). *1510 influenza pandemic*. URL: https://en.wikipedia.org/wiki/1510_influenza_pandemic (cit. on p. vii).

— (Apr. 8, 2021b). *2010s Haiti cholera outbreak*. URL: https://en.wikipedia.org/wiki/2010s_Haiti_cholera_outbreak (cit. on p. vii).

— (Apr. 10, 2021c). *Coronavirus*. URL: https://https://en.wikipedia.org/wiki/Coronavirus (cit. on p. viii).

— (Apr. 13, 2021d). *History of HIV/AIDS*. URL: https://en.wikipedia.org/wiki/History_of_HIV/AIDS (cit. on p. vii).

— (Mar. 28, 2021e). *History of smallpox*. URL: https://en.wikipedia.org/wiki/History_of_smallpox (cit. on p. vii).

— (Apr. 8, 2021f). *HIV/AIDS*. URL: https://en.wikipedia.org/wiki/HIV/AIDS (cit. on p. vii).

— (Apr. 6, 2021g). *Influenza pandemic*. URL: https://en.wikipedia.org/wiki/Influenza_pandemic (cit. on p. viii).
— (Apr. 23, 2021h). *Spanish flu*. URL: https://en.wikipedia.org/wiki/Spanish_flu (cit. on p. vii).

Index

A
AIDS, vii
Anna Karenina Principle, 60
Asymptomatic, 56
Attractor, 44, 53

B
Basic reproduction number R_0, 12
Behavior
    careful, 92, 94
    human, x, 1, 27, 79, 93
    normal, 94
    sticky, 84, 88
Behavioral epidemiology, 92
Bendixson's Criterion, 49
Births, 30, 52, 53, 75, 94
Box, George, 15
Brazil, vii, ix, 27

C
Chain rule, 21, 22, 84
Characteristic equation, 39
China, viii
Cholera, vii
Common cold, 2, 9
COMORBUSS, ix
Compartment, 2
Contact, 9
    adequate, 2, 10
    rate, adequate, 10, 28, 32, 94
Contact tracing, 74
Coronaviruses, viii
Covid-19, viii, 27, 32, 55, 73–75
    model, 55

D
Deaths, 19, 28, 30
    natural, 53, 75, 94
Determinant, 46, 54, 58, 62, 75
Differential equation, 4
    linear, 38, 68
        hyperbolic, 40
    polynomial, 48
Divergence, 49, 51
Divergence Theorem, 49
Dulac's Criterion, 50, 54

E
Eigenvalues and eigenvectors, 39, 44, 58, 59, 62, 67, 72, 76, 77, 85
Endemic, x, 8
Entry-exit integral, 88
Equilibrium, 6, 20, 23, 24, 36, 38, 51, 53, 58, 59, 72, 75, 85, 95
  asymptotically stable, 43
  disease-free, 14, 58, 60, 63, 64, 67
  endemic, 8, 14, 16, 51
  hyperbolic, 44, 51, 52
  normally hyperbolic, 59, 60, 85
Exponential growth and decay, 30, 54, 61, 73

G
Game theory, 80
Global stability, 50
Gradient, 22
Graphic, 48, 50, 54

H
Health care system, 33
Herd immunity, 32, 52
Hospitals, 27

I
Imitation dynamics, 80, 95
Immunity, 2, 18, 35, 52
India, x
Infectious disease, v, 1
Infective, 1
Influenza, vii
Initial value problem, 5
Integration by parts, 70
Invariant curve, 24, 40
Invariant set, 22
Inverse of a matrix, 65, 77
Italy, ix

K
Kermack, W. O., x

L
Law of mass action, 14
Limit system
  fast, 85
  slow, 85, 87
Linearization, 37, 44, 64
  matrix, 45, 53, 58, 72, 76, 85
Lockdowns, 74

M
Malaria, vii
Manifold, 44
Mask-wearing, 33, 74, 79
Matrix exponential, 68, 69
McKendrick, A. G., x
Mean, 11
Media, 93
Memory, 92
MERS-CoV, viii
Model
  agent-based, ix, 15, 93
  endemic, 51
  epidemic, 52
  network, viii, 15
  ordinary differential equation, x
  partial differential equation, ix
  SEIR, 55, 75
  SIR, x, 18, 59, 60
  SIS, 3, 13
  stochastic, viii, 15

N
Natural logarithm, 17, 61

Newborns, 94
Newton, Isaac, 3
Newton's Second Law of Motion, 5
Next generation matrix, x, 63, 65, 77
Nullcline, 20, 36

O
Orbit, 23, 25, 33, 60
    closed, 47, 50, 54

P
Partial fractions, 17
Payoff, 80, 82, 94
Perron–Frobenius Theorem, x
Phase
    line, 7, 15
    portrait, 7, 24, 33, 34, 40, 42, 43
    space, 6
Plague, vi
Population
    changing, 28
    constant, 2, 9, 18
    growing, 30, 54
Probability distribution, 11, 15

Q
Quarantine, 32

R
Recovered, 18
Removed, 19, 30
Repeller, 44
Reproduction number
    basic R_0, 12, 19, 27, 62, 67, 68, 73–75, 82
    effective R_e, 13, 21, 25, 52, 60, 63, 67
Rio de Janeiro, vii
Rural settlement, 28

S
Saddle, 44, 51, 53
SARS-CoV, viii
Saudi Arabia, viii
Separation of variables, 17
Separatrix cycle, 48, 50, 54
Sierra Leone, x
Singular solution, 90
Slow-fast system, 84
Smallpox, vi
Social distancing, 32, 79
Spanish flu, vii
Stable and unstable
    manifolds, 51, 59
    subspaces, 41
Susceptible, 2

T
Temperature checks, 74
Time
    fast, 84
    mean, 11, 16, 65, 70
    slow, 84
Trace, 46, 54

U
United Kingdom, ix, 61, 74, 75
United States, ix
Units, 11
Urban island, 28

V
Vaccine, 32, 80, 94
Variants, 71
Vector field, 19, 20, 37
Velocity vector, 4
Visitors, 16

W
Waves of an epidemic, 87

Títulos Publicados — 33º Colóquio Brasileiro de Matemática

Geometria Lipschitz das singularidades – *Lev Birbrair e Edvalter Sena*

Combinatória – *Fábio Botler, Maurício Collares, Taísa Martins, Walner Mendonça, Rob Morris e Guilherme Mota*

Códigos Geométricos – *Gilberto Brito de Almeida Filho e Saeed Tafazolian*

Topologia e geometria de 3-variedades – *André Salles de Carvalho e Rafał Marian Siejakowski*

Ciência de Dados: Algoritmos e Aplicações – *Luerbio Faria, Fabiano de Souza Oliveira, Paulo Eustáquio Duarte Pinto e Jayme Luiz Szwarcfiter*

Discovering Euclidean Phenomena in Poncelet Families – *Ronaldo A. Garcia e Dan S. Reznik*

Introdução à geometria e topologia dos sistemas dinâmicos em superfícies e além – *Víctor León e Bruno Scárdua*

Equações diferenciais e modelos epidemiológicos – *Marlon M. López-Flores, Dan Marchesin, Vítor Matos e Stephen Schecter*

Differential Equation Models in Epidemiology – *Marlon M. López-Flores, Dan Marchesin, Vítor Matos e Stephen Schecter*

A friendly invitation to Fourier analysis on polytopes – *Sinai Robins*

PI-álgebras: uma introdução à PI-teoria – *Rafael Bezerra dos Santos e Ana Cristina Vieira*

First steps into Model Order Reduction – *Alessandro Alla*

The Einstein Constraint Equations – *Rodrigo Avalos e Jorge H. Lira*

Dynamics of Circle Mappings – *Edson de Faria e Pablo Guarino*

Statistical model selection for stochastic systems – *Antonio Galves, Florencia Leonardi e Guilherme Ost*

Transfer Operators in Hyperbolic Dynamics – *Mark F. Demers, Niloofar Kiamari e Carlangelo Liverani*

A Course in Hodge Theory Periods of Algebraic Cycles – *Hossein Movasati e Roberto Villaflor Loyola*

A dynamical system approach for Lane–Emden type problems – *Liliane Maia, Gabrielle Nornberg e Filomena Pacella*

Visualizing Thurston's Geometries – *Tiago Novello, Vinícius da Silva e Luiz Velho*

Scaling Problems, Algorithms and Applications to Computer Science and Statistics – *Rafael Oliveira e Akshay Ramachandran*

An Introduction to Characteristic Classes – *Jean-Paul Brasselet*

impa

Instituto de
Matemática
Pura e Aplicada

ISBN 978-65-89124-39-9