Universitat Autònoma de Barcelona

Mathematical modelling of pathogen specialisation

Anel Nurtay

ADVERTIMENT. L'accés als continguts d'aquesta tesi queda condicionat a l'acceptació de les condicions d'ús establertes per la següent llicència Creative Commons: http://cat.creativecommons.org/?page_id=184

ADVERTENCIA. El acceso a los contenidos de esta tesis queda condicionado a la aceptación de las condiciones de uso establecidas por la siguiente licencia Creative Commons: http://es.creativecommons.org/blog/licencias/

WARNING. The access to the contents of this doctoral thesis it is limited to the acceptance of the use conditions set by the following Creative Commons license: https://creativecommons.org/licenses/?lang=en

Mathematical Modelling of Pathogen Specialisation

Anel Nurtay

Centre de Recerca Matemàtica, Universitat Autònoma de Barcelona,

and Institute for Integrative Systems Biology

(Consejo Superior de Investigaciones Científicas - Universitat de València)

Thesis submitted under the supervision of Prof. Dr. Lluís Alsedà

and co-supervision of Prof. Dr. Santiago F. Elena

in order to be awarded the degree of

Doctor of Philosophy in Mathematics

Barcelona 2019

This thesis is dedicated to my parents.

Acknowledgements

First and foremost, I would like to thank my supervisor, Lluís Alsedà, for being a strong team leader and, what is more important, a kind-hearted friend throughout the last few years; for gifting me with the academic freedom and for gathering the best collaborators the work required. I also thank my other supervisor, Santiago Elena, who has never been second, in anything; for being the kind of biologist every mathematician dreams to work with and, in particular, for his genuine patience with my limited training in virology; for providing me with the most valuable opportunities I could possibly wish for.

I cannot thank Josep Sardanyès enough for all the meticulous feedback and hard work he put in the development of the doctoral thesis. The content of this work would not be as broad and descriptive without his questions and challenges, especially so in Chapters 3, 4 and 5. I am particularly grateful for Josep proposing new routes at a rare moment of doubt by providing an excellent tutorship; for becoming the big academic brother only the luckiest of students can have.

I want to salute Matthew Hennessy for the highest level of academic professionalism he have brought to the table. I thank him for examining this doctoral thesis, several times; for his indisputably fundamental knowledge of mathematical modelling that he generously and patiently shared during the work on Chapters 4, 5, 6, and 7. On a personal level, I would like to thank Matt for teaching me the academic precision and the power of hardworking by his own example; for a friendship of a lifetime.

Many other people have contributed to my scientific development over the course of the training. For instance, I had a great discussion with Antoni Guillamon on the subject of bifurcations of periodic orbits that benefited the work recorded in Chapter 4. Tomás Alarcón had a great impact on the development of a stochastic model that ensured the preliminary results seen in Chapter 8. The work on Chapter 6 was eased due to the expertise of Brian R. Wetton. I had a great conversation with Vladimir Sobolev and learned a lot on the methods of system order reduction. I thank Andrei Korobeinikov whose ideas were explored in Chapters 7 (the main model) and 8 (the stochastic model); I am genuinely grateful to Andrei for giving me the chance to walk this path.

Along with all mentioned above, I treasure my acquaintance with Elisa Beltrán, Timothy Myers, Helena Ribera, Marc Calvo, and Néstor Costa, to all of whom I, most likely, owe a croissant and a big hug for their support and friendship.

La meva formació al CRM compta amb el finançament de l'Obra Social "la Caixa" dins del programa Recerca en Matemàtica Col·laborativa (My training at CRM is funded by "la Caixa" Foundation within the programme Research on Collaborative Mathematics).

Abstract

The occurrence of new disease-causing viruses is tightly linked to the specialisation of viral sub-populations towards new host types. Mathematical modelling provides a quantitative framework that can aid with the prediction of long-term processes such as specialisation. Due to the complex nature of intra- and inter-specific interactions present in evolutionary processes, elaborate mathematical tools such as bifurcation analysis must be employed while studying population dynamics. In this thesis, a hierarchy of population models is developed to understand the onset and dynamics of specialisation and their dependence on the parameters of the system. Using a model for a wild-type and mutant virus that compete for the same host, conditions for the survival of only the mutant sub-population, along with its coexistence with the wild-type strain, are determined. Stability diagrams that depict regions of distinct dynamics are constructed in terms of infection rates, virulence and the mutation rate; the diagrams are explained in terms of the biological characteristics of the sub-populations. For varying parameters, the phenomenon of intersection and exchange of stability between different periodic solutions of the system is observed and described in the scope of the competing wild-type and mutant strains. In the case of several types of hosts being available for competing specialist and generalist strains, regions of bistability exist, and the probabilities of observing each state are calculated as functions of the infection rates. A strange chaotic attractor is discovered and analysed with the use of Lyapunov exponents. This, combined with the stability diagrams, shows that the survival of the generalist in a stable environment is an unlikely event. Furthermore, the case of $N \gg 1$ different strains competing for different types of host cells is studied. For this case, a counterintuitive and non-monotonic dependence of the specialisation time on the burst size and mutation rate is discovered as a result of carrying out a regression analysis on numerically obtained data. Overall, this work makes broad contributions to mathematical modelling and analysis of pathogen dynamics and evolutionary processes.

Resum

L'aparició de nous virus causants de malalties està estretament lligada a l'especialització de subpoblacions virals cap a nous tipus d'amfitrions. La modelització matemàtica proporciona un marc quantitatiu que pot ajudar amb la predicció de processos a llarg termini com pot ser l'especialització. A causa de la naturalesa complexa que presenten les interaccions intra i interespecífiques en els processos evolutius, cal aplicar eines matemàtiques complexes, com ara l'anàlisi de bifurcacions, al estudiar dinàmiques de població. Aquesta tesi desenvolupa una jerarquia de models de població per poder comprendre l'aparició i les dinàmiques d'especialització, i la seva dependència dels paràmetres del sistema. Utilitzant un model per a un virus de tipus salvatge i un virus mutat que competeixen pel mateix amfitrió, es determinen les condicions per a la supervivència únicament de la subpoblació mutant, juntament amb la seva coexistència amb el cep de tipus salvatge. Els diagrames d'estabilitat que representen regions de dinàmiques diferenciades es construeixen en termes de taxa d'infecció, virulència i taxa de mutació; els diagrames s'expliquen en base a les característiques biològiques de les subpoblacions. Per a paràmetres variables, s'observa i es descriu el fenomen d'intersecció i intercanvi d'estabilitat entre diferents solucions sistemàtiques i periòdiques en l'àmbit dels ceps de tipus salvatge i els ceps mutants en competència directa. En el cas de que diversos tipus d'amfitrions estiguin disponibles per a ser disputats per ceps especialitzats i generalistes existeixen regions de biestabilitat, i les probabilitats d'observar cada estat es calculen com funcions de les taxes d'infecció. S'ha trobat un rar atractor caòtic i s'ha analitzat amb l'ús d'exponents de Lyapunov. Això, combinat amb els diagrames d'estabilitat, mostra que la supervivència del cep generalista en un entorn estable és un fet improbable. A més, s'estudia el cas dels diversos ceps $N \gg 1$ que competeixen per diferents tipus de cèl·lules amfitriones. En aquest cas s'ha descobert una dependència no monotònica, contraria al que es preveia, del temps d'especialització sobre la mida inicial i la taxa de mutació, com a conseqüència de la realització d'un anàlisi de regressió sobre dades obtingudes numèricament. En general, aquest treball fa contribucions àmplies a la modelització matemàtica i anàlisi de la dinàmica dels patògens i els processos evolutius.

Resumen

La aparición de nuevos virus causantes de enfermedades está estrechamente ligada a la especialización de las subpoblaciones virales hacia nuevos tipos de anfitriones. La modelización matemática proporciona un marco cuantitativo que puede ayudar a la predicción de procesos a largo plazo como la especialización. Debido a la naturaleza compleja que presentan las interacciones intra e interespecíficas en los procesos evolutivos, aplicar herramientas matemáticas complejas, tales como el análisis de bifurcación, al estudiar dinámicas de población. Esta tesis desarrolla una jerarquía de modelos de población para poder comprender la aparición y las dinámicas de especialización, y su dependencia de los parámetros del sistema. Utilizando un modelo para un virus de tipo salvaje y un virus mutado que compiten por el mismo anfitrión, se determinan las condiciones para la supervivencia únicamente de la subpoblación mutante, junto con su coexistencia con la cepa de tipo salvaje. Los diagramas de estabilidad que representan regiones de dinámicas diferenciadas se construyen en términos de tasa de infección, virulencia y tasa de mutación; los diagramas se explican en base a las características biológicas de las subpoblaciones. Para parámetros variables, se observa y se describe el fenómeno de intersección e intercambio de estabilidad entre diferentes soluciones sistemáticas y periódicas en el ámbito de las cepas de tipo salvaje y las cepas mutantes en competencia directa. En el caso de que varios tipos de anfitriones estén disponibles para ser disputados por cepas especializadas y generalistas existen regiones de biestabilidad, y las probabilidades de observar cada estado se calculan como funciones de las tasas de infección. Se ha encontrado un raro atractor caótico y se ha analizado con el uso de exponentes de Lyapunov. Esto, combinado con los diagramas de estabilidad, muestra que la supervivencia de la cepa generalista en un entorno estable es un hecho improbable. Además, se estudia el caso de los varias cepas $N \gg 1$ que compiten por diferentes tipos de células anfitrionas. En este caso se ha descubierto una dependencia no monotónica, contraria a lo que se preveía, del tiempo de especialización sobre el tamaño inicial y la tasa de mutación, como consecuencia de la realización de un análisis de regresión sobre datos obtenidos numéricamente. En general, este trabajo hace contribuciones amplias a la modelización matemática y el análisis de la dinámica de los patógenos y los procesos evolutivos.

Contents

Part I

Introduction

Chapter 1

Viral dynamics

1.1 Epidemics caused by pathogens

In the fifth century BC, Hippocrates employs the adjective *epidemios* to mean "which circulates or propagates in a country" [108], and, in his legendary Corpus Hippocraticum, titles seven out of 62 treatises Epidemics [65]. Interestingly Hippocrates, the Father of Medicine, does not use for the title of those specific books a well-established term of that period, the word *nosos* - disease. Approximately at the same time, Plato uses *epidemeo* in the context of "to return home after a voyage, to be in town". Knowing this, we can understand why Hippocrates prefers the word epidemics which implies "previous occurrence elsewhere" to describe the wide-spread occurrence of similar clinical symptoms [86].

Throughout history, wide-spread diseases have been following humans, animals, and plants, with repeated occurrences of similar conditions. Attempts to explain the causes of such outbreaks were limited by the technology available at the time and, sometimes, the imagination and general beliefs of people. Before *Vibrio cholerae* was isolated in 1854 [98], epidemics of cholera were commonly explained by blood, phlegm, and yellow and black bile imbalances that were believed to be the consequence of the lifestyle of the sick. Commoners aside, it took great effort to convince the scientific community of the pathogenic causes of cholera. The deliberate consumption of bacteria cultures by opponents of the hypothesis, which was supposed to illustrate an inconsistency of the latter, ended up supporting the pathogenic theory of epidemics at a great cost.

Another big step towards understanding the nature of wide-spread diseases lay in discovering the contagiousness of such epidemics and in the determination of the main carriers. As famously illustrated by Dr John Snow during the repeated occurrence of cholera in London, certain public water pumps appeared to be hotbeds of greater cholera mortality rate

than others. With the use of epidemiological maps, he showed that the water can be a carrier of the disease. The difference in sources that supplied city pumps became a singular factor determining safety of consumable water. It was noticed that pumps supplied from the higher parts of the river delivered safe water, and the river parts containing wastes, which came from sewer channels, supplied the deadliest of pumps [118].

Nowadays, an epidemic stands for any outbreak of a disease that spreads rapidly throughout the population. Some widely common conditions in humans without a known pathogenic origin, such as obesity, are listed as epidemics, too. However, over the last two decades, roughly all of the outbreaks that have been of clinical and public health importance were caused by emerging pathogens and, more specifically, by viruses [24]. These epidemics include the outbreak of Severe Acute Respiratory Syndrome (SARS) coronavirus [110], which in 2003 led to more than 8000 human infections. Severe cases of SARS infection often progress rapidly, leading to respiratory distress and require intensive care. Another recently detected human coronavirus, Erasmus Medical Center (HCoV-EMC) [14, 146, 131], is known to have, firstly, a weak species barrier and, secondly, replicate faster than SARS [63], which are the warnings of a possible epidemic. An epidemic that spreads throughout the world is referred to as pandemic. In 2009, a pandemic occurred after the repetitive emergence of the influenza virus led to a generation of drug-resistant sub-type known as A(H1N1)pdm09 [119, 125]. Since 2013, another sub-type A(H7N9) known as avian influenza virus [43] has infected 1567 people with 615 deaths in China. For a constantly evolving and seasonally re-occurring virus like influenza, the international spread is almost a certainty.

The weak species barrier mentioned above indicates that a virus can infect different species and switch between host types. Most diseases caused by viruses have been transmitted to the human population from surrounding animals either via consumption of untreated animal products or via direct interaction. Due to the overall small dimensions and large quantities of virus particles in one host, transmission of a single virus particle – a **virion** – to a new type of host in a shared habitat is a highly probable event. However, it is the survival of the virion in a new type of host which determines the likelihood of transmission. If a virion infects the new type of host successfully, then the virus has a chance of transmitting between individuals of that species. In medicine and biology, the term transmission refers to the passing of a virus from one individual to another. Depending on the type of virus, the transmission can happen via different routes: direct contact, droplet contact, airborne transmission, *etc.* Some virus types can also be transmitted through vectors – intermediate carriers – such as mosquitoes and ticks. The time between the virion transmission and occurrence of symptoms is called incubation period of viral infection. The incubation period

of some viruses, such as the human immunodeficiency virus (HIV), can be as long as several years which contributes to the wide spread of viral diseases all over the globe.

1.2 Virus and host cell interaction

First discovered as "an exceedingly small parasitic microorganism" during a study of tobacco mosaic disease [59], viruses can be found in almost every living organism on the planet. Most of the virus types are $20 - 300$ nanometres in size with rare exceptions being giants, which slightly exceed the size of a cell and infect amoebas. In general, most of the virions on Earth are found in marine life, including other microorganisms such as bacteria [124]. It is important to understand the mechanisms that lead to the appearance, survival, and evolution of virus populations despite the fact that not all viruses cause a disease. It is believed that **virulence** – the damage-causing property of a pathogen – is a consequence of co-evolution between the immune response of a host and, in our case, the virus [41, 51, 32].

The structure, shape, and composition of a virus varies depending on the type. All virions, however, have a protein shell called the capsid, which surrounds the genome of the virus and accommodates enzymes. The genome is formed of RNA or DNA strains. Based on recent studies, **viral strains** contain slight differences even for the virions of the same type and harvested from the same plant or animal [145]. The differing strains of the same virus are referred to as **variants**. These variants are also referred to as **quasispecies** within the field of evolutionary virology [31, 113, 114]. The enzymes of the virus catalyse processes inside the host cell. For example, an enzyme called retroviral integrase plays an important role in incorporating the viral genetic material, *i.e.* RNA, of a retrovirus into the genome of the host in the form of viral DNA. Some viruses have envelopes, which are mostly obtained from the membrane of the cell they infected; the similarity between the envelope of a virion and the membrane of a cell may allow the virus to avoid the immune system [45]. Envelopes also contain viral proteins that conduct different roles. For instance, glycoproteins identify and connect the virion to the surface of a cell membrane allowing the viral capsid to enter inside the host cell. Since viruses lack autonomous reproduction mechanisms, they rely on their host for survival: after entering a host cell, the virion exploits the machinery and resources of the host cell by the use of the cellular enzymes to replicate its own genome and assemble new virions. The average number of virions produced by one cell is called the **burst size**. A more detailed description of this process is illustrated in Figure 1.1 for the case of the family of retroviruses mentioned above.

Not all viral infections lead to an immediate replication of viral genome; the stage of virion production by an infected cell is called a lytic cycle. A dormant stage after the

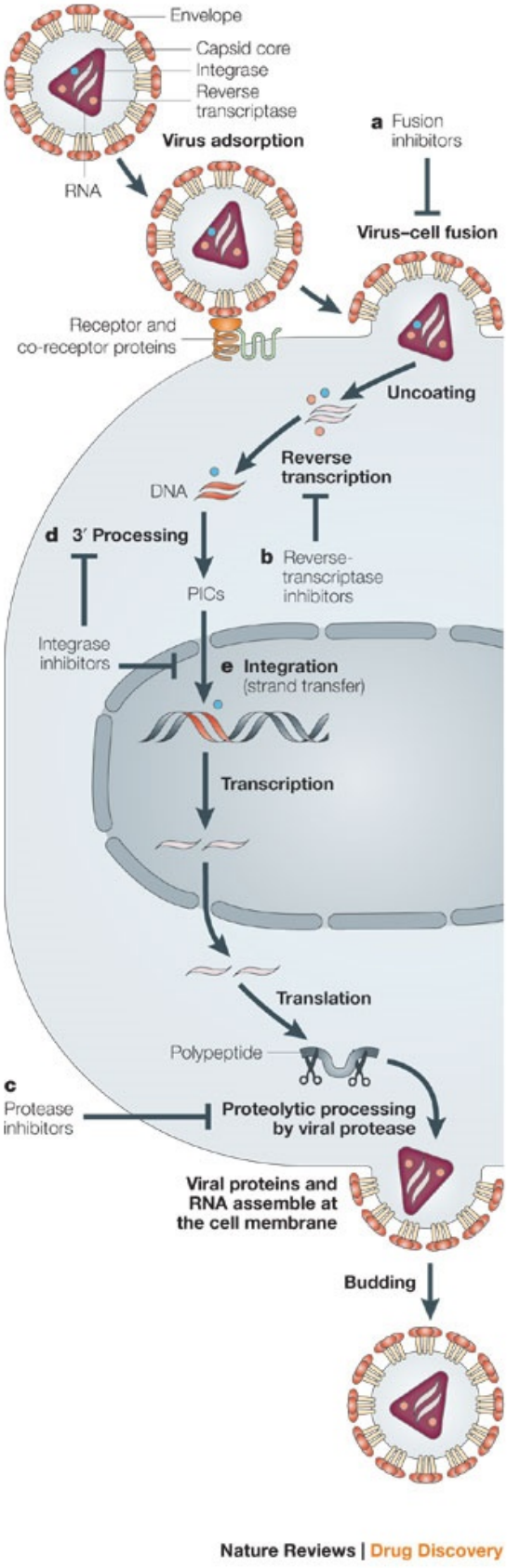


Figure 1.1: Replication of retroviruses, such as human immunodeficiency virus type 1 (HIV-1). After the retrovirus enters a host cell, reverse transcriptase converts the retroviral RNA genome into double-stranded DNA. This viral DNA then migrates to the nucleus and becomes integrated into the host genome. Viral genes are transcribed and translated. New virus particles assemble, exit the cell, and can infect another cell. — The image is used by permission from Springer Nature Customer Service Centre GmbH (License Number: 4505870941555). Original Figure 1 was first published in [103] by Y. Pommier et.al. Integrase inhibitors to treat HIV/AIDS. *Nat. Rev. Drug Discov.*, 4(3):236, 2005.

integration of a viral genome into a genome of the cell is called a lysogenic cycle and is characterised by a normal multiplication of the infected cell. This process is controlled by viral genes, and can be destabilised by *e.g.* a mutagen [123], and lead to a lytic cycle after several generations of cell division [144]. Viruses such as bacteriophage lambda and HIV can propagate in both ways. In case of HIV, the lysogenic cycle contributes to a longer incubation period and might lead to a greater number of transmissions due to lower probability of detection, which is believed to be one of the main difficulties in finding a cure for AIDS [102]. In this manuscript, mostly the lytic cycle is being discussed.

Due to the size difference between virions and host cells, in combination with the abundance of viral particles, single cells can be infected by several virions. An infection of a single cell by more than one virion of the same viral type is called coinfection, and the average number of virions spent to accomplish a successful infection is called the **multiplicity of infection**. However, some animals or plants can be a host for different types of pathogens. The process of sequential infection of a cell of an organism by virions of far origins is called a superinfection. Although this discussion might diverge from classic definitions for host-parasite scale, it does present the idea of coinfection and superinfection and, more importantly, rescale the processes to a single-cell level. In this thesis, only the coinfection of a host cell is considered.

1.3 Virus evolution

For every set of given environmental conditions, there are certain characteristics of an organism that can provide it with a reproductive advantage. Commonly, those characteristics are detectable and help the organism to interact with the environment. The organism can even be classified based on its unique features, also referred to as a **phenotype** of the organism. Natural selection acts on the phenotype. The heritable background – the genome – of the advantageous phenotype may become more common in a population due to the reproductive success. With time, this process can result in populations that are suitable for particular ecological niches (microevolution). Moreover, the ecological segregation may eventually lead to speciation, that is the emergence of new species (macroevolution).

The grand role in evolution of organisms is played by fitness and related ideas such as fitness landscapes and evolutionary optima. Nevertheless, the concept of fitness is often context dependent [28]. It is widely accepted to consider the **fitness** of a species as a measure of overall reproductive success. In the case of viral populations, the fitness of the viral strains can safely be related to infectivity of the virions. Fitness in virology is closely related to the concept of the basic reproduction number, which constitutes the number of

new transmissions that a single transmission can create. In terms of cell-virion dynamics, the **basic reproduction number** can be determined as the average number of infectious virus particles created from a single infected cell. Usually, the basic reproduction number is proportional to the burst size and the infection rate.

The combination of very large population sizes, short generation times, and lack of proof-reading mechanisms during genome replication confer viral populations with an extremely high evolutionary plasticity that allow them to quickly adapt to environmental changes such as new host species, presence of antiviral drugs, new transmission routes, or to new vectors [5]. This tremendous evolvability, coupled with densely populated animal and plant susceptible hosts (in many cases lacking genetic variability for resistance to infection), are the reasons for the persistence and emergence of new viral diseases or the re-emergence of new strains with novel properties of already known diseases. The continuous emergence of new mutants leads to an overlap in existence of wild-type and mutant genotypes of the same virus within individual infected hosts [130, 72]. This cloud of mutants is usually known as a viral quasispecies [5] and it constitutes the target of selection, instead of the individual viral genomes. The reservoir of coexisting genetic variants may lead to the emergence of new genotypes with different host ranges, pathologies, and epidemiological properties that may result in outbreaks [37, 142].

With the development of high-coverage, ultra-deep sequencing techniques, it is now possible to characterise in great detail virus genetic diversity along the course of infection of individual hosts, demonstrating the coexistence of multiple mutant genotypes within individual hosts, some even during long periods of time [11, 58, 67]. Furthermore, some of these studies have also shown that dynamics are highly complex and do not only depend on the differences in replicative fitness among individual variants, but on other parameters such as the size and frequency of within-host bottlenecks, complementation of strains, fixation of additional mutations on the same genotype, epistasis, the availability of beneficial mutations (which indeed depends on the degree of adaptation to the host), the load of deleterious mutations, or clonal interference among coexisting beneficial mutations [2, 26, 78, 69].

Depending on the envelope and capsid structure, some viruses infect only one type of host, while other viruses can infect several different types of hosts. In the scope of this thesis, if a virus can infect only one type of cell, then it is called a **specialist**. In foreign environments, a specialist virus slowly loses its infectivity and brings no harm to its host. When a virus can infect different types of cells, regardless of whether the same or several different host species are being infected, the virus is called a **generalist**. Empirical studies have shown that generalist viruses have a lower success rate with a specific type of a host cell in comparison to a virus which is specialised towards that cell type [61, 36]. This exchange

between proficiency in one trait, *e.g.* a greater ability to infect one specific cell type, and another, *e.g.* average ability to infect many types of cells, is called a fitness trade-off and is a result of the finite energy capacity of organisms including viruses. The generalist has been described as the "jack-of-all trades" which is "master of none" [138], referring to the benefits provided for specialist viruses as a result of the fitness trade-off. However, in cases of scarcity of certain type of cells or competition with fitter specialised viruses, the generalist may switch hosts and find other niches to occupy and, therefore, can have better chances of persisting in the environment. These dynamics lead to different scenarios that can occur in the evolutionary studies of viruses.

1.4 Mathematical models of population dynamics

Population ecology studies the interaction between competing, cooperating, or otherwise related populations along with the effect of the environment on the interaction patterns [64]. By the use of mathematical models, the dynamics of population interactions can be explored using multitude of mathematical and computational tools within the framework of Dynamical Systems Theory. A main attraction of population dynamics is the possibility of long-term realistic and precise predictions that models can provide. Historically, most of the models that describe population dynamics have been based on phenomenology. The classic equation of logistic growth is a great example of a phenomenological model:

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \beta x \left(1 - \frac{x}{K}\right), \tag{1.1}$$

where x is the population size. The intrinsic growth rate β is a cumulative rate of change of the population size. In the absence of external mechanisms, such as artificial harvest of individuals from the studied population, there is always a limit to how large a population can grow without causing the destruction of its environment. That load, K, is called the carrying capacity of the environment, and it leads to the limitation of the population growth. The equation (1.1) is used to describe the self-regulating growth of a population. When it was initially proposed in 1845 by Verhulst [132], no justification of the model was given, which illustrates the "describing but not explaining the mechanisms" nature of phenomenological models. The dependency of the intrinsic growth rate and carrying capacity on the characteristics of individuals in the population is uncertain, and therefore can be misleading in some cases [111].

Although any model is understandably a simplification of reality, some phenomena can be understood and described better when studied by individual-based models [111]. Due

to the use of mechanisms in individual processes, models of such kind are referred to as mechanistic or first principle models [38]. The main reason for switching from solely phenomenological models to more mechanistic models is the idea that selection works through individuals. The survival of a species may depend on differences in reproductive success of individuals. In the case of viral dynamics, where an infection can be started by a single virion, the properties shown by individuals are of great importance and must be taken into account.

In this thesis, the individual-level interaction processes are modelled using the law of mass action [95, 101, 60]. The origin of the mass action principle lays in chemical reactions, where the rate of a reaction is proportional to the product of the concentrations of the reactants. This non-linearity can give rise to complex dynamics such as chaos. Interaction between individuals in population ecology are observed to be quite elaborate and, therefore, can be considered using the law of mass action: virus particles and uninfected cells are reactants, and the infected cells are the product of the reaction [92, 126]. This approach was taken in the famous Lotka-Volterra predator-prey model [82, 133]. Moreover, empirical studies have shown that, for example, elimination of virion targets by immune $CD8^+$ T cells follows the law of mass action [42].

Strictly speaking, the mutation of even the most simple organisms such as viruses is a highly complicated procedure that is driven by various external and internal processes. Therefore, modelling the mutation can be done in various ways [83, 21]. In the scope of this thesis, a mutation is a change in a genome of a virus that is sufficient to differ the **wild-type** strain from all the **mutants** produced in the next generations. Mutation is modelled by the flow of one population (wild-type) to another (mutants) or, in a more general case, via a diffusion operator.

1.5 Goals and hypotheses

The process of specialisation is closely related to the coexistence and competition of different strains of a virus. In order to study the underlying processes and causes of these phenomena, different mathematical and computational models will be developed and analysed. Each system is dedicated to explaining certain processes. The following questions are challenged throughout the thesis.

1.5.1 Specialisation leads to the appearance of new viral subtypes

How do new viruses appear? Despite the absence of proof-reading mechanisms in mutating viral strains, a continuous distribution of all possible genetic variations is never observed in nature. On the contrary, it is common to find specific branching and clustering of variants. To answer the question of specialisation, it is important to understand that the viruses are always subjected to changing ecological environments and are shaped by the interaction with their hosts. Viral populations can adapt to new environments either by selection on standing genetic variation, that is the presence of different combinations of genes resulting in different phenotypic traits in a population, or by selection on new mutations [12]. By means of either, specialisation towards different host cell types can drift the viral strains away from one another. Several generations of selective growth in different hosts and/or host cell types can result in an accumulation of too many suitable mutations for new niches in variants of the same origin. We hypothesise that specialisation, eventually, leads to a phenotypic divergence of the variants and emergence of new viruses.

1.5.2 Low mutation rates lead to intraspecific competition

How does mutation rate affect the process of specialisation? In an abundance of susceptible host cells and different cell types, the specialisation time depends on the parameters of the system such as burst size and mutation rate. Understandably, a high mutation rate must lead to faster specialisation of populations towards different hosts. Likewise, a greater burst size must lead to quicker specialisation too. However, we aim to illustrate that the qualitative dependence of the specialisation time on system parameters, such as the burst size, is not the same for small and large mutation rates. Predictably, the combination of high mutation rates and growing burst sizes leads to a faster specialisation time. On the contrary, viral specialisation slows down with growing burst size for a fixed low mutation rate. We hypothesise that intraspecific competition between variants of close origin [47], *i.e.* clonal interference or, simply put, interspecific competition in terms of viral quasispecies, can slow down the process of specialisation of the virus population with low mutation rates and increasing burst sizes.

1.5.3 Coexistence is possible with a fitter wild-type strain

How can variants in the same environment that have similar properties coexist instead of out-competing one another? The appearance of a mutant usually plays against the wild-type strain because of the intraspecific competition. Moreover, the mutant strains are more fit to the environment due to selection [7]. We will illustrate that for a wild-type virus to

survive the competition with mutant strains, the wild-type either has to be more infectious than the mutant strains or it has to have access to other host cell types. We illustrate that the weak species barrier, that is, a weak host-cell-types barrier of a virus, offers an advantage in the presence of fitter competitors. We also aim to show that, in certain cases, the behaviour of the system with generalist wild-type and specialist mutant-type strains cannot be predicted even in a deterministic setup due to the presence of a strange chaotic attractor.

Alongside with above stated questions, many other queries will be discussed. The main goal of this thesis, however, remains in the development of mathematical tools for understanding mechanisms that drive viral specialisation. In Chapter 2, the relevant mathematical concepts and derivations used to address these issues are described. Examples of simple models for a viral population interacting with a susceptible host cell population are discussed in Chapter 3, where the effect of different growth rates of susceptible cells on dynamics is considered. The most realistic model from Chapter 3 is employed to study the dynamics of populations of two viral strains with a shared host cell population in Chapter 4. By introducing a second type of host cell, the dynamics of a generalist and a specialist virus populations are considered in Chapter 5. The modelling specifics of increasing number of variants are defined and solved in Chapter 6, where a host cell population is shared between different variants of a virus. These modifications enable a model to be developed that considers multiple populations of healthy cell types that can be infected by the different variants, which differ from one another with two phenotypic characteristics. This model is used to study the specialisation time in Chapter 7. The thesis concludes in Chapter 8 with general discussion of main results obtained from this research.

Chapter 2

Mathematical tools and concepts

By a broad definition of it in mathematics, a dynamical system describes units which vary in time, t. The units may be for instance a location in space or a temperature, or as in this thesis, population sizes. A mathematical model which is constructed to depict the change of population quantities in time must exploit some form of dynamical equations, such as difference equations, where time is considered in discrete steps, or differential equations, which treat time as a continuous variable. Moreover, while studying a complicated relationship between different populations, not a single equation but a system of such equations may be involved. When the processes depicted in the model are simple enough, it is possible to obtain solutions of the system of differential equations analytically. However, nonlinear differential equations often cannot be solved analytically and numerical approaches as well as other techniques are required. A large part of dynamical systems theory deals with understanding the long-term behaviour of solutions, which can either tend to stationary values that are constant in time, become periodic in time, or even be highly irregular and aperiodic if the system is chaotic. There are a large number of tools that have been developed to understand the qualitative behaviour of solutions, and the description of the most crucial concepts are provided below.

2.1 Linear stability analysis

The concept of approximating values of a function around a certain point with a tangent vector paves the way for conducting a linear stability analysis. An ideological core of linear stability analysis is in considering a generalised version of a tangent to a function. Assume,

that the following system of differential equations is being considered

$$\frac{\mathrm{d}\boldsymbol{v}(t)}{\mathrm{d}t} = \boldsymbol{M}(\boldsymbol{v}(t)), \tag{2.1}$$

with equilibria $\boldsymbol{v}_i$, which satisfy $\boldsymbol{M}(\boldsymbol{v}_i) = \boldsymbol{0}$ for each $i = 1, \ldots, n$, where n is the number of equilibria. For sufficiently small δ, the system can *be linearized* around $\boldsymbol{v}_i$ by substituting $\boldsymbol{v}(t)$ with $\boldsymbol{v}_i + \delta\,\hat{\boldsymbol{v}}(t)$ in (2.1), for each $i = 1, \ldots, n$. Straightforward substitution on the left hand side (LHS) and a Taylor expansion that follows the substitution on the right hand side (RHS) of the equation (2.1) yields the following:

$$\begin{aligned}
&\text{LHS:} \quad \frac{\mathrm{d}\boldsymbol{v}(t)}{\mathrm{d}t} = \frac{\mathrm{d}(\boldsymbol{v}_i + \delta\,\hat{\boldsymbol{v}}(t))}{\mathrm{d}t} = \frac{\mathrm{d}\boldsymbol{v}_i}{\mathrm{d}t} + \frac{\mathrm{d}\delta\,\hat{\boldsymbol{v}}(t)}{\mathrm{d}t}, \\
&\text{RHS:} \quad \boldsymbol{M}(\boldsymbol{v}_i + \delta\,\hat{\boldsymbol{v}}(t)) = \boldsymbol{M}(\boldsymbol{v}_i) + \boldsymbol{D}\boldsymbol{M}\big|_{\boldsymbol{v}_i}\,\delta\,\hat{\boldsymbol{v}}(t) + \ldots = \boldsymbol{M}(\boldsymbol{v}_i) + J(\boldsymbol{v}_i)\,\delta\,\hat{\boldsymbol{v}}(t) + \ldots,
\end{aligned} \tag{2.2}$$

where J is the matrix of every first-order partial derivative of a vector-valued function $\boldsymbol{M}$, *i.e.* the Jacobian matrix. Since equilibrium $\boldsymbol{v}_i$ satisfies the equation (2.1) for all $i = 1, \ldots, n$, the term $\dot{\boldsymbol{v}}_i$ from the LHS and $\boldsymbol{M}(\boldsymbol{v}_i)$ from the RHS can be cancelled leading to the following simplified form:

$$\frac{\mathrm{d}\,\hat{\boldsymbol{v}}(t)}{\mathrm{d}t} = J(\boldsymbol{v}_i)\,\hat{\boldsymbol{v}}(t) + O(\delta). \tag{2.3}$$

After neglecting the small terms that are $O(\delta)$ in size, the solution of differential equation (2.3) can be obtained by writing

$$\hat{\boldsymbol{v}}(t) = \hat{\boldsymbol{w}}\,e^{\lambda t}, \tag{2.4}$$

whose straightforward substitution in (2.3) results in

$$\lambda\,\hat{\boldsymbol{w}} = J\,\hat{\boldsymbol{w}}. \tag{2.5}$$

By definition, λ is an eigenvalue of J and $\hat{\boldsymbol{w}}$ is an eigenvector of J. The eigenvalues of the Jacobian matrix evaluated at equilibrium $\boldsymbol{v}_i$, that is λ from $\det(J(\boldsymbol{v}_i) - I\lambda) = 0$, are exploited to determine the type and stability of this equilibrium. The **Lyapunov theorem** states that if $Re(\lambda) > 0$, then $\hat{\boldsymbol{v}}(t) = \hat{\boldsymbol{w}}e^{\lambda t}$ grows exponentially in time, so the equilibrium point $\boldsymbol{v}_i$ is unstable, and if $Re(\lambda) < 0$, then $\hat{\boldsymbol{v}}(t) = \hat{\boldsymbol{w}}e^{\lambda t}$ converges to a finite value at $t \to +\infty$, therefore leading to a stable equilibrium $\boldsymbol{v}_i$. Among all, it is worthy of noting a special kind of equilibria, a hyperbolic equilibrium, which is a solution of $\boldsymbol{M}(\boldsymbol{v}_i) = \boldsymbol{0}$ at

which its Jacobian has no eigenvalue with real part equal to zero. Hyperbolic equilibria are known to be "sturdy", *i.e.* a small change in parameter values will not lead to a change in stability (which would be the case if the real part was equal to zero – a change in parameter values could lead to the point becoming stable or unstable depending on how the system shifts) [122]. In relation to local stability of quilibrium points, the **Hartman-Grobman theorem** [50, 48] states that linearization of a system around a hyperbolic equilibrium provides faithful description of its stability. According to this theorem, when all of the eigenvalues of the Jacobian have non-zero real part, the equilibrium can be characterised by the largest (first-order) term of its Taylor expansion and its stability is not affected by nonlinear terms.

Linear stability analysis plays an important role in this thesis, and is employed throughout the manuscript in Chapters 3, 4, 5, and 6.

2.2 Existence and stability of periodic solutions

A periodic solution corresponds to a type of trajectory of a dynamical system which repeats itself in time. More specifically, if $\boldsymbol{v}^*(t)$ is a non-constant solution of (2.1), and if there is a positive and minimal T such that $\boldsymbol{v}^*(t) = \boldsymbol{v}^*(t+T)$ for all t, then $\boldsymbol{v}^*(t)$ is a periodic solution of (2.1). The period of this solution is T. Each periodicity interval $[t, t+T]$ provides $\boldsymbol{v}^*(t)$ in the variable space which is referred to as a periodic orbit or – as an isolated object – a limit cycle. Sometimes it is possible to prove that a limit cycle exists or cannot exist analytically using the techniques discussed below [49]. For an autonomous (time-invariant, *i.e.* $\boldsymbol{M}$ does not explicitly depend on time) planar vector field, **index theory** can be used to show that inside the region enclosed by a periodic orbit there must be at least one equilibrium. In case there is only one equilibrium, it must be a sink (stable node), source (unstable node), or center. If every equilibria inside the periodic orbit is hyperbolic, then there must be an odd number $2m + 1$ of equilibria, of which m are saddles and $m + 1$ are sinks or sources. The theory is beneficial for showing that a periodic orbit does not exist in a region of phase space because a periodic orbit cannot exist without appropriate equilibria. **Dulac's criterion** or the **Bendixson–Dulac theorem** states that if there exists a differentiable function ϕ such that the divergence of the vector field $\phi\boldsymbol{M}$ with respect to the state variables has the same sign everywhere except possibly in a set of measure 0, then the autonomous system $\dot{\boldsymbol{v}} = \boldsymbol{M}(\boldsymbol{v})$ has no non-constant periodic solutions lying entirely within the region. The theorem requests a physical interpretation of periodic solutions and applies Green's theorem for the proof. Dulac's criterion is a generalised form of **Bendixson's criterion**, which corresponds to $\phi = 1$ in the body of the theorem above. These criteria can be

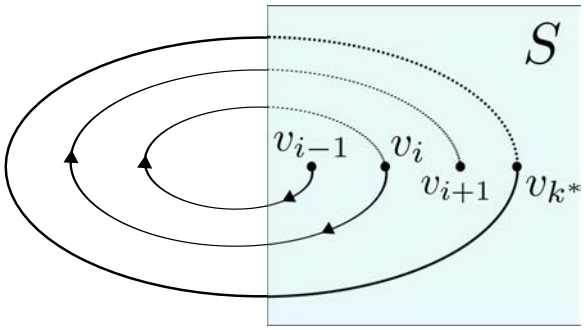


Figure 2.1: Illustration of a Poincaré map P on S: $\boldsymbol{v}_{i+1} = P(\boldsymbol{v}_i)$ and $\boldsymbol{v}_{k^*} = P(\boldsymbol{v}_{k^*})$. Note, that dashed curves mean appearance of trajectories *behind* S, where S is the Poincaré section transversal to the flow.

useful for showing that a periodic orbit cannot exist in a region of a phase plane. When a geometrical approach is plausible, the **Poincaré-Bendixson theorem** can be convenient. It implies that if a trajectory enters and does not leave a closed and bounded region of planar phase space which contains no stationary solution $\boldsymbol{v}$, then the trajectory must approach a periodic orbit as $t \longrightarrow \infty$. This approach can sometimes be used to establish the existence of a stable periodic orbit for a planar vector field.

Poincaré maps can be an intuitive and useful concept to study the stability of periodic solutions. To describe the simplified Poincaré map for an N-dimensional system (2.1), imagine an $(N-1)$-dimensional subspace, S, which is transverse to the flow of trajectories of the system. Following a trajectory, we observe it intersecting S without loss of generality at $\boldsymbol{v}_1$. For long enough time (positive or negative), the same trajectory might intersect S more than once at $\boldsymbol{v}_k$, where $k \geq 1$. The sequence of trajectory intersections with S, $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k$, form the Poincaré map $\boldsymbol{v}_k = P(\boldsymbol{v}_{k-1})$ on S for all k. When these intersections occur for the same trajectory repeatedly, we may have a periodic solution or chaos. If a trajectory starting from $\boldsymbol{v}_{k^*}$ on S returns to $\boldsymbol{v}_{k^*}$ after some time T, that is $\boldsymbol{v}_{k^*} = P(\boldsymbol{v}_{k^*})$, then the system has a periodic solution of period T. Moreover, by checking whether slightly perturbed trajectories cross S and asymptotically approach or diverge from $\boldsymbol{v}_{k^*}$, we can determine the stability of the periodic solution corresponding to $\boldsymbol{v}_{k^*}$ in N-dimensional space, see Figure 2.1. More specifically, we suppose that $\Delta\boldsymbol{v}_0$ is an infinitesimal perturbation that keeps $\boldsymbol{v}_{k^*} + \Delta\boldsymbol{v}_0$ on the same S. After another $i+1$ returns to S, ignoring the small $O(||\Delta\boldsymbol{v}_i||^2)$ terms at each iteration, we have the following mapping on S

$$\boldsymbol{v}_{k^*} + \Delta\boldsymbol{v}_{i+1} = P(\boldsymbol{v}_{k^*} + \Delta\boldsymbol{v}_i) = P(\boldsymbol{v}_{k^*}) + [LP(\boldsymbol{v}_{k^*})]\Delta\boldsymbol{v}_i + O(||\Delta\boldsymbol{v}_i||^2). \tag{2.6}$$

The $LP(\boldsymbol{v}_{k^*})$ is an $(N-1)\times(N-1)$ matrix called the linearized Poincaré map. If C_j are some

scalars, $\boldsymbol{e}_j$ are basis eigenvectors, and λ_j are eigenvalues of $LP(\boldsymbol{v}_{k^*})$ for $j = 1, \ldots, N-1$, then perturbations on S can be found by

$$\Delta \boldsymbol{v}_{i+1} = \sum_{j=1}^{N-1} (\lambda_j)^i C_j \boldsymbol{e}_j. \tag{2.7}$$

These λ_j are called **characteristic** or nontrivial **Floquet multipliers** of a limit cycle, and the limit cycle is linearly stable if and only if $|\lambda_j| < 1$ for all $j = 1, \ldots, N-1$. Classically, a total of N Floquet multipliers are considered, with the last one being $\lambda = 1$, which corresponds to perturbations along the periodic solution. If a limit cycle is stable, then all numerical integration for any initial conditions from the basin of attraction of the limit cycle will converge to the limit cycle for large enough t.

With mathematical models of complicated processes, it is often impossible to find and study equations for periodic solutions $\boldsymbol{v}(t)$ analytically. Therefore, periodic orbits are usually searched for using numerical methods. The problem of locating (stable or unstable) periodic orbits for an autonomous vector field can be reformulated so that a variant of the **Newton-Raphson algorithm** can be used to numerically solve $\varphi_T(\boldsymbol{v}) - \boldsymbol{v} = \boldsymbol{0}$ for $\boldsymbol{v}$ and T, where $\varphi_T(\boldsymbol{v})$ is a trajectory which ends at $\boldsymbol{v}$ after a time T [99]. However, the topology of high dimensional flows is hard to visualise, and even with a decent starting guess for the shape and location of a periodic orbit, methods like the Newton-Raphson algorithm are likely to fail [73]. More robust numerical methods to locate a cycle can be formulated as a periodic **boundary-value problem** (BVP) on a fixed interval. Usually, the period of the cycle, T, is unknown and thus by considering T as a parameter in (2.1) we obtain the following system of equations:

$$\frac{\mathrm{d}\boldsymbol{u}(\tau)}{\mathrm{d}\tau} = T\boldsymbol{M}(\boldsymbol{u}(\tau)), \tag{2.8a}$$

$$\boldsymbol{u}(0) = \boldsymbol{u}(1), \tag{2.8b}$$

$$\Psi[\boldsymbol{u}] = \int_0^1 \langle \boldsymbol{u}(\tau), \dot{\boldsymbol{U}}(\tau) \rangle d\tau = 0, \tag{2.8c}$$

where T is a time-scaling factor, τ is the new time variable on the interval $[0, 1]$, and $\boldsymbol{u}(\tau)$ satisfies the periodic boundary condition, which alone does not define the periodic solution uniquely. When there is a need for selection of the unique representation of the periodic solution, a phase condition, $\Psi[\boldsymbol{u}] = 0$, has to be added, where $\Psi[\boldsymbol{u}]$ is a scalar functional defined on the periodic solution. There are many ways of defining the phase condition.

The phase condition used in (2.8c) is called integral phase condition, where $\boldsymbol{U}(\tau)$ is a reference period-one solution, and operator $\langle \cdot, \cdot \rangle$ is a classical inner product [29]. In order to solve the BVP, it must be reduced to a finite-dimensional problem. This can be done by the use of various techniques. The **method of shooting** reduces the BVP to an initial value problem and calculates different trajectories using a procedure that iterates over various initial conditions until the boundary conditions are satisfied. For N-dimensional systems, the method provides a system of $(N+1)$ scalar equations with N equations for the components of $\boldsymbol{v}(t)$ and one equation for the period T. The system can be solved with a Newton-like method, and a successful result would be the periodic solution $\boldsymbol{v}_{k*}$ and the value of T. On the pessimistic side, the numerically obtained results may be different from exact solutions due to errors caused by strong growth in unstable directions. An attempt to decrease this error accumulation naturally leads to the method of **multiple shooting**, which considers the superposition of n correspondence maps from S_n, where S_i are dividing the time interval into n subintervals. Despite having a better algorithmic stability, increasing n is numerically expensive and even then the method may fail due to the accumulation of error for large systems. When computational efficiency is not a priority, the method of **finite differences** can be applied. This method discretizes the derivative in (2.8a) at τ_j by finite differences for $j = 0, \ldots, n-1$. The system (2.8) turns into $nN+1$ discrete system of equations, which can be solved by Newton-like methods. Versions of this method are often used and they often can provide a sufficiently accurate result. However, finite difference methods require accurate initial guesses and large n. Furthermore, these methods may fail to converge to a periodic solution since they are sensitive to error accumulation from unstable directions in the case of high-dimensional systems. The method used in this thesis allows for the computation of both stable and unstable periodic solutions, and is called **orthogonal collocation** [52]. In summary, this method considers the BVP (2.8) by introducing the discretization of the interval $[0, 1]$ into n sub-intervals:

$$0 = \tau_0 < \tau_1 < ... < \tau_n = 1, \tag{2.9}$$

and then searches for an approximate solution that is a piece-wise differentiable, continuous function which is a vector-polynomial $\boldsymbol{u}^{(j)}(\tau)$ with a maximal degree m within every subinterval $[\tau_j, \tau_{j+1}]$, for $j = 0, \ldots, n-1$. That is, within each sub-interval an approximate solution satisfies the system (2.8a) at m collocation points:

$$\tau_j = \tau_{j,0} < \tau_{j,1} < \tau_{j,2} < ... < \tau_{j,m} = \tau_{j+1}. \tag{2.10}$$

The unknown polynomial $\boldsymbol{u}^{(j)}(\tau)$ can be represented through the interpolation:

$$\boldsymbol{u}^{(j)}(\tau) = \sum_{i=0}^{m} \boldsymbol{u}^{(j)}(\tau_{j,i})\, l_{j,i}(\tau), \tag{2.11}$$

where $l_{j,i}(\tau)$ are Lagrange basis polynomials. The non-trivial placing of the collocation points in the sub-intervals $[\tau_j, \tau_{j+1}]$ is to minimise the approximation error; it is known that the optimal choice is to place them as Gauss points, which are the roots of the m^{th} degree Legendre polynomial relative to the sub-interval. The periodicity condition (2.8b) and the phase condition (2.8c) are also discretized, and the latter employs Lagrange quadrature coefficients in the approximation of the integral. The Legendre polynomials compose the orthogonal system on the interval (thus the name of the method), and collocation at their numerically obtained and known roots provide an extremely accurate approximation of the smooth solution of (2.8):

$$||\boldsymbol{u}(\tau_j) - \boldsymbol{u}^{(j)}(\tau_{j,0})|| = O(h^{2m}), \tag{2.12}$$

as $h = \max_{0\leq j\leq n-1} |\tau_{j+1} - \tau_j| \longrightarrow 0$ [71, Sec. 10.1].

In the scope of the studied models, the limit cycles are of a great importance for biological interpretations. Their existence and stability are vigorously studied in Chapters 4 and 5.

2.3 Bifurcation analysis

The effect of varying parameters on the studied quantities is a main concern of mathematical modelling. The change in dynamics as a function of the model parameters becomes a crucial issue in biological dynamical systems. Sometimes, qualitatively different dynamical behaviours of the model can be observed for different intervals of a parameter. These changes in dynamics of the system are referred to as bifurcations. The goal of bifurcation analysis is to construct bifurcation diagrams. The **bifurcation diagram** divides the parameter space into regions within which the system shows a qualitative behaviour characteristically diverging from the neighbouring areas. The boundaries separating these areas are the bifurcation points. Overall, the bifurcation analysis can be computationally and methodologically challenging. For the necessary numerical simulations of processes discussed in Chapters 3 to 6, the programming language MATLAB and its toolbox MatCont [27] are employed. More theoretical background of the bifurcation analysis is provided below.

The **Kupka-Smale theorem** lists three generic characteristics of a dynamical system: (i) hyperbolic equilibrium points, (ii) hyperbolic periodic equilibria, and (iii) transversal intersections of stable and unstable manifolds of equilibrium points and periodic or-

bits [68, 120]. The changes in these Kupka-Smale conditions lead to different bifurcation types. The codimension of a type of bifurcation is defined as the minimal number of parameter families in which that bifurcation type occurs. Codimension-one bifurcations correspond to a single failure of the Kupka-Smale conditions. For example, the saddle-node, transcritical, and Andronov-Hopf bifurcations (further referred to as Hopf bifurcation) correspond to a change along an equilibrium that occurs for one varying parameter. When all eigenvalues of an equilibrium have strictly negative real parts, then for the given parameters, that equilibrium is stable. For a certain combination of parameters, two equilbria with different stability might non-destructively collide and, moreover, with variation of parameters above the critical combination one of the equilibria can lose its stability while the other one becomes stable. This "stability exchange" around the critical parameter combination corresponds to **the transcritical bifurcation** of intersecting equilibria. **The saddle-node bifurcation** is a collision and disappearance of two equilibria, or appearance of two branches of equilibria depending on the direction of the varying parameter. In N-dimensional phase space, with $N \geq 2$, the Jacobian matrix at the saddle-node bifurcation has a simple zero eigenvalue with the real parts of other eigenvalues being positive or negative. **The Hopf bifurcation** leads to the emergence of a periodic solution, which may be stable or unstable. It is worth mentioning that there are subcases of bifurcation types, too; for instance, the Hopf bifurcation can be subcritical or supercritical [71]. The supercritical Hopf, for instance, changes a stable equilibrium to unstable and provides a stable periodic orbit around the unstable equilibrium. The **fold limit cycle bifurcation** is sometimes called a saddle node of limit cycles because it starts two new, stable and unstable, limit cycles for the varying parameter and corresponds to a position where the stable and unstable limit cycles meet in the phase space. There are also global bifurcations such as homoclinic bifurcation of equilibria, homoclinic tangencies of stable and unstable manifolds of periodic solutions, and heteroclinic bifurcation of equilibria and periodic solutions. Other types of codimension-one bifurcations can be found in systems with quasiperiodic oscillations or chaotic dynamics.

Increasing the codimension makes the classification of bifurcation types even more difficult. However, five types of codimension-two bifurcations of equilibria are considered classic: Bautin bifurcation, Bogdanov-Takens bifurcation, cusp bifurcation, Hopf-Hopf bifurcation, and fold-Hopf bifurcation. The **Bautin bifurcation** is sometimes referred to as a generalised Hopf bifurcation. It is a bifurcation of an equilibrium at which the critical equilibrium has a pair of purely imaginary eigenvalues and the first Lyapunov coefficient for the Hopf bifucation is zero. This bifurcation connects branches of sub- and supercritical Hopf bifurcations in the parameter space. For parameter values close to the Bautin bifurcation, the system has two limit cycles which collide and disappear via a saddle node bifurcation

of periodic orbits. The **Bogdanov-Takens bifurcation** is a bifurcation of an equilibrium point at which the critical equilibrium has two zero eigenvalues. For parameter values close to the standard Bogdanov-Takens bifurcation, the system has two equilibria, one of which is a saddle and another one is not a saddle, that collide and disappear via a saddle node bifurcation. The non-saddle equilibrium undergoes a Hopf bifurcation generating a limit cycle. This cycle degenerates into a homoclinic orbit and disappears via a saddle homoclinic bifurcation. The **cusp bifurcation** is characterised by two branches of saddle-node bifurcation curves meeting tangentially, which on a two-dimensional bifurcation diagram forms a sharp cusp. For parameter values close to cusp bifurcation points, the system can have three equilibria which collide and disappear pairwise via saddle node bifurcations. In general, the cusp bifurcation implies the presence of a hysteresis in the system. The **Hopf-Hopf bifurcation**, or the double-Hopf bifurcation, is a bifurcation of an equilibrium point at which the critical equilibrium has two pairs of purely imaginary eigenvalues. The bifurcation point in the parameter plane lies at a transversal intersection of two curves of Hopf bifurcations. Usually, two branches of torus bifurcations emanate from the Hopf-Hopf bifurcation point. The **fold-Hopf bifurcation** is a bifurcation of an equilibrium at which the critical equilibrium has a zero eigenvalue and a pair of purely imaginary eigenvalues. This bifurcation is also called the zero-Hopf bifurcation, the saddle node Hopf bifurcation or the Gavrilov-Guckenheimer bifurcation. The bifurcation point in the parameter plane lies at a tangential intersection of curves of saddle node bifurcations and Hopf bifurcations. When a branch of torus bifurcations emanate from the zero-Hopf point, a rich variety of additional global bifurcations can take place and can lead to the birth of strange chaotic attractors that exist in localised regions of parameter space.

2.4 Chaos

By definition, stochastic processes are not considered in deterministic systems. Hence, a deterministic model should produce the same output for a given set of parameters and initial values of variables. Moreover, a deterministic system should produce similar outputs for *almost identical* initial conditions. However, not all deterministic models allow for the strict prediction of their future states. Knowing all $\boldsymbol{v}_i$ that turns the right-hand side of (2.1) into zero, finding the conditions for their stability, and mapping basins of attractions in case of a bistability might not be enough to predict the behaviour of the system for certain scenarios. Chaos, in dynamical systems, is a long-term non-periodic behaviour of a deterministic system that has a strong dependence on initial conditions. This means that nearby initial conditions do not lead to trajectories that remain close to each other for large

times. Incidentally, above mentioned Poincaré-Bendixson theorem then implies that chaos cannot exist in planar phase space. Thus, the minimal dimensionality of an autonomous system of nonlinear differential equations that allows the appearance of chaotic behaviour is three.

A classical way of quantifying chaos is by means of the so-called **Lyapunov exponents**. This elegant method monitors the long-term evolution of an infinitesimal N-sphere of initial conditions for a dynamical system. If the system tends to chaotic behaviour for large times, the sphere becomes an N-ellipsoid due to the trajectories diverging exponentially. The i^{th} Lyapunov exponent is then defined in terms of the length of i^{th} principal axis $p_i(t)$ of the ellipsoid:

$$\varepsilon_i = \lim_{t \longrightarrow \infty} \frac{1}{t} \log_2 \frac{p_i(t)}{p_i(0)}, \tag{2.13}$$

where the ε_i are ordered and form a vector of Lyapunov exponents $\boldsymbol{\varepsilon}$. Every positive exponent corresponds to a "stretching direction" of the system which leads to the divergence of close trajectories. This means that the long-term behaviour of an initial condition that is slightly different from an observed trajectory cannot be strictly predicted; thus the chaos. The search for the Lyapunov exponents is typically conducted by numerical integration of the nonlinear equations of the system starting from a given initial condition within the basin of attraction of the chaotic attractor and as described in Ref. [141]. The full nonlinear system (2.1) and the equations that result from its linearization are simultaneously integrated in time using a set of N initial conditions that define a set of N orthonormal vectors. These N orthonormal vectors define the principal axes of the initial N-sphere. The solutions to the linearized equations describe the evolution and growth of the principal axes and enable the Lyapunov exponents to be calculated. However, it is important to recall that, while each of the initial N orthonormal vectors diverges in magnitude with time, in a chaotic system they also tend to align themselves in the local direction of fastest growth. The collapse of the vectors towards a common direction causes the tangent space orientation of all axes vectors to become indistinguishable. Therefore, the repeated use of the Gram-Schmidt reorthonormalization procedure on the vector frame is required. The frequency of reorthonormalization is known to be irrelevant as long as a magnitude and an orientation of divergences does not exceed computer limitations.

Biological and, especially, epidemiological systems are not unfamiliar with chaotic dynamics [34, 25]. An interesting and novel example of chaos identified in the population dynamics of specialist and generalist viral strains is discussed in Chapter 5.

2.5 Solving partial differential equations

Consider the following system of partial differential equations (PDEs):

$$\frac{\partial \boldsymbol{v}(s,t)}{\partial t} = \boldsymbol{M}\left(\boldsymbol{v}, \frac{\partial \boldsymbol{v}}{\partial s}, \frac{\partial^2 \boldsymbol{v}}{\partial s^2}\right), \tag{2.14}$$

with initial condition:

$$\boldsymbol{v}(s,0) = \boldsymbol{v}_0(s), \tag{2.15}$$

and certain boundary conditions. Technically, the solution of a PDE problem is a function that defines the dependent variable $\boldsymbol{v}$ in terms of the independent variables s and t, thus solving the PDE problem means finding a function that satisfies simultaneously (2.14) and all of its auxiliary conditions. The solutions to a PDE can be found using either analytical or numerical methods. While analytical solutions are exact expressions for $\boldsymbol{v}(s,t)$, they are generally difficult to derive for all but a narrow spectrum of PDE problems. As in the case of a complicated nonlinear system of algebraic equations, the solution of elaborate systems of PDEs are usually obtained numerically. The numerical solution is considered to be an approximation to the analytical solution. The **method of lines** (MOL) is a technique used for the calculation of accurate numerical solutions of time-dependent PDEs. The principle of the MOL is in replacing the spatial derivatives in the PDE with algebraic approximations, thus yielding a system of ordinary differential equations. To illustrate the basic idea, allow the spatial domain of the function to be discretized by m (equidistant) inner points, that is $0 = s_0 < s_1 < ... < s_i < ... < s_m < s_{m+1} = 1$ for corresponding $i = 0, \ldots, m+1$. Then, the first partial derivative of $\boldsymbol{v}$ with respect to s can be calculated by the second-order approximation at i as such:

$$\frac{\partial \boldsymbol{v}(s,t)}{\partial s} \approx \frac{\boldsymbol{v}_{i+1}(t) - \boldsymbol{v}_{i-1}(t)}{2\Delta s} + O(\Delta s^2), \tag{2.16}$$

where $\Delta s = 1/(m+1)$ for $i = 1, \ldots, m$. If present in the PDE problem, a second partial derivative is approximated, for example, using a centered difference formula with second-order error:

$$\frac{\partial^2 \boldsymbol{v}(s,t)}{\partial s^2} \approx \frac{\boldsymbol{v}_{i+1}(t) - 2\boldsymbol{v}_i(t) + \boldsymbol{v}_{i-1}(t)}{\Delta s^2} + O(\Delta s^2), \tag{2.17}$$

and so on. Henceforth, the spatial derivatives are discretized with respect to the spatial independent variables, and only the initial-value variable, *time*, remains. Eventually, with only one remaining independent variable, we have an enlarged system of ODEs that approximates the original system of PDEs. The challenge is in performing a correct construction

of the system of ODEs with attention to its initial and boundary conditions. The initial condition requires a fairly straightforward discretization. Boundary conditions differ depending on the processes described in the problem, and thus different approaches might be necessary. For a no-flux boundary condition, values of fictitious points can be obtained from equation (2.16) with the centered finite difference approximation at $i = 0$ and $i = m + 1$:

$$\frac{\partial \boldsymbol{v}(s,t)}{\partial s} \approx \frac{\boldsymbol{v}_1(t) - \boldsymbol{v}_{-1}(t)}{2\Delta s} = 0 \quad \text{and} \quad \frac{\partial \boldsymbol{v}(s,t)}{\partial s} \approx \frac{\boldsymbol{v}_{m+2}(t) - \boldsymbol{v}_m(t)}{2\Delta s} = 0, \tag{2.18}$$

which yields

$$\boldsymbol{v}_{-1} = \boldsymbol{v}_1 \quad \text{and} \quad \boldsymbol{v}_{m+2} = \boldsymbol{v}_m. \tag{2.19}$$

Further, the system of ODEs can be integrated by any suitable numerical method. This shows one of the important features of the MOL, that is the exploitation of well-established numerical procedures for ODEs to solve elaborate systems of PDEs [115]. For more elaborate models, the classical method of finite differences can be exploited, too.

In Chapters 6 and 7, PDEs are used to formulate models of population dynamics. To obtain numerical solutions to these models, the programming languages MATLAB and Python are employed. Certain calculations in Chapter 7 – as the regression analysis described further below – require use of computational clusters, and for that BASH scripts are coded.

2.6 Regression analysis

Numerical simulations can help us to understand the behaviour of a complex phenomenon for any set of introduced parameter values. Finding real-life values of the parameters can become a challenge in its own. Luckily, in most cases the estimation of a relationship between the variables is of more interest rather than individual outcomes of mechanistic models. There is, usually, a question that includes free and dependent variables. The answer to that question requires the quantitative description of an effect that the change in the free variable poses over the dependent variables. Regression analysis provides concentrated information about the general trends in behaviour of the dependent variables. From rather straightforward simulations, we can observe if multiple parameters are affecting the qualitative behaviour of the system. In that case, we can employ regression analysis in an iterative way.

For example, consider a population that changes in order to occupy certain resource niches. The successful niche occupation requires some amount of time, which we denote

by T. Assume, that time T depends on parameters, a and b, and maybe more, which are not given and may have broad domains. First, by fixing all parameters except a at approximate values, we can conduct a regression analysis: a acts as a free variable and T as a dependent variable. That is, we calculate T as a function of a for densely chosen data points $a_1 < a_2 < \ldots < a_p$, for large enough p. Regression analysis can be considered as a highly visual tool, because it implicates plots of studied variables. Depending on the general shape of the scattered plot of $T(a_i)$, it is possible to employ a corresponding regression technique. Assume that the scattered plot vaguely resembles a linear function. In such a case, values of $T(a_i)$ for all $i = 1, \ldots, p$ can be characterised by two numbers in total: the slope and intercept of the linear function that the values of $T(a_i)$ are clustered around. Moreover, depending on the model, the most interesting information can be concentrated singularly in one number, *e.g.* the slope L, which summarises the effect the parameter a has over the niche occupation time T as a general trend. In this case, large and positive L would mean that a slight increase in parameter a leads to a rapid increase in time T, therefore suggesting that a is greatly important parameter. On the other hand, a small and negative value of L would mean that a slight increase in a leads to a slow decrease in T showing robustness and inverse proportionality of the studied variable T towards parameter a. Depending on the specifics of the model, some variations of a least squares method are used to find the linear function; these methods minimise the sum of (weighted) squared differences between the scattered plot and the regression function. Going back to the example: after obtaining the first linear function, we save its slope L as the regression value. Then we choose and change slightly the second parameter of interest, b, and conduct the same process of linear regression with respect to a_i, for $i = 1, \ldots, p$. Each iteration for b_j from $b_1 < b_2 < \ldots < b_q$ for large enough q, provides us, for instance, with a single value of the slope, $L(b_j)$ for $j = 1, \ldots, q$. By doing this, we obtain quantitative representation of the effect that parameters a and b have on the time it takes for a population to occupy a niche.

In the scope of this thesis, we must deal with nonlinear regression problems. That is, the scattered plot of $T(a_i)$ does not necessarily resemble a linear function after straightforward calculations. In such a case, the first approach can be the use of nonlinear regression functions. By doing so, we can obtain very accurate approximations of the scattered plot, but the method requires tracking several parameters that characterise the regression function, instead of just one, *e.g.* L. This can be inconvenient for further iterations. The second approach lays in the linearization of the problem itself. This approach includes altering the studied model or its domain in a way that scattered plot of the dependent variable becomes linear. The most frequently used linerization techniques can be divided into transformation

and segmentation methods. The segmentation method allows the use the linear regression for sub-domains of independent variables of the regression analysis. This method is very suitable for systems that have a threshold value in their independent variables yet may not be very convenient for an iterative approach. Similarly, the transformation methods require caution since even the most appropriate model alterations lead to a change in data values and their interpretations, as well as to a change in the structure of errors and domains. This method can be useful, however, when we observe clear tendency such as exponential growth or trigonometric trend. Then, plotting the logarithm or an inverse trigonometric function of the studied variable may lead to a suitable conditions for the linear regression analysis.

Regression analysis is a very powerful tool with great applications in data analysis. In Chapter 7, we use the regression analysis to determine the time it takes for a population to split into different sub-populations. There, critical parameter combinations are found that change the trend of the split time drastically.

2.7 Stochastic modelling

Stochasticity can occur in population dynamics because of environmental and/or demographic reasons [3]. Environmental stochasticity corresponds to random changes that happen *to* the studied population, while demographic stochasticity refers to population fluctuations due reproductive processes that especially affects small populations. The frequent change of temperature is a perfect example of environmental stochasticity since it may affect the process of cell growth, therefore, leading to a fluctuation in the virus population size due to the unavailability of susceptible cells.

From a mathematical perspective, the stochasticity is modelled by introducing fluctuations into a deterministic system [44]. At the end of this thesis, we overview a combination of the environmental and/or demographic stochasticity types which affects the infection rate of a pathogen population. First, we consider the infection rate as a function of every sub-type s_i of a pathogen, $\alpha(s_i)$ for $i = 1, \ldots, N$. Next, we generate an instance of Brownian motion (*e.g.* by the Wiener process) for every s_i, and thus $\alpha(s_i, t)$ varies with time. The random variable of the infection rate at time t, that is $\alpha(s_i, t)$, has a normal distribution whose mean is the infection rate at time $t = 0$ and whose variance is $\delta^2 t$, where δ is the "speed" of the Brownian motion. Note that the infection rate of each pathogen sub-type is independent of the infection rates of the other sub-types. This approach provides the possibility to control the intensity of infection rate perturbations for each sub-type of a pathogen

population autonomously. The use of the stochastic mathematical models applied in viral dynamics is illustrated in Chapter 8.

Part II

Modelling viral evolution: population dynamics of a single viral strain

Chapter 3

Growth rate-based pursuit for a basic model

In this chapter we consider a basic model of viral dynamics. The choice of a simple model will allow for better understanding of the dynamics of the considered processes. The effects of qualitatively different growth rates of a susceptible, uninfected cell population on the behaviour of the whole system are studied thoroughly. Modelling the growth rate of the uninfected cell population in different ways is sufficient to illustrate the main set of qualitatively different behaviours. Starting from a simple constant influx rate, we develop gradually the growth rate up to a logistic growth. The main purpose of this chapter is to pinpoint the most reliable model for further development of more elaborated systems incorporating further biological complexity.

3.1 Basic model

The basic model of viral dynamics has three state variables: the population of uninfected cells, x; the population of infected cells, y; and the population of free virus particles, also called virions, z. These quantities describe sizes of the populations either *in vivo* or *in vitro*, depending on how the flow, A, of uninfected cells is modelled. The system we are considering in many ways mimics the classical model of viral dynamics [95]. Free virus particles infect uninfected cells at infection rate α. The infection decreases the uninfected cell population and increases the infected cell population at a rate proportional to the product of uninfected cells and viral load abundance, which yields αzx. The infection rate α can be regarded as the probability of successful infection. The average number of virions

required for a successful infection process is called the multiplicity of infection (MOI). In this model, the MOI is chosen to be a constant $\nu \geq 1$. Outflow and death of infected cells occurs at a rate γ, which can be considered as the virulence of the virus. Free virions are produced by the infected cells at a rate $\kappa\gamma y$, for a burst size κ, and are removed from the system at a rate ζz.

The evolution of the interacting populations is described by the following system of ODEs which, in turn, is based on the Nowak and May model of HIV dynamics [95]:

$$\begin{aligned} \dot{x} &= A - \alpha\, z\, x, \\ \dot{y} &= \alpha\, z\, x - \gamma\, y, \\ \dot{z} &= \kappa\gamma\, y - \nu\, \alpha\, z\, x - \zeta\, z. \end{aligned} \tag{3.1}$$

This slightly updated *classic* model of virus dynamics, provides a starting point for modelling the dynamics of specialisation of viral strains. In the following sections, we consider different growth rates, A, of the uninfected cell population and their effect on the dynamics.

3.2 Growth rate as a constant flow

The simplest approach is to consider a combination of influx and outflow as a constant flow of uninfected cells. Justification for usage of constant flow is modelling virus replication in a system where the total population of uninfected cells incomparably surpasses the population of virus and can be controlled (*e.g.* harvested from a Petri dish). Setting $A = \beta$ in (3.1) gives the following ODE system

$$\begin{aligned} \dot{x} &= \beta - \alpha\, z\, x, \\ \dot{y} &= \alpha\, z\, x - \gamma\, y, \\ \dot{z} &= \kappa\gamma\, y - \nu\, \alpha\, z\, x - \zeta\, z. \end{aligned} \tag{3.2}$$

System (3.2) has a single equilibrium defined by the values $\{x^*, y^*, z^*\}$ that yield $\dot{x} = \dot{y} = \dot{z} = 0$, which is given by:

$$x^* = \frac{\zeta}{\alpha(\kappa - \nu)}, \quad y^* = \frac{\beta}{\gamma}, \quad z^* = \frac{\beta(\kappa - \nu)}{\zeta}. \tag{3.3}$$

From the crucial condition of non-negative population sizes, we obtain restrictions for the parameters: $\alpha > 0$, $\kappa - \nu > 0$, $\gamma > 0$, $\zeta > 0$, and $\beta \geq 0$. The case $\kappa = \nu$ corresponds to the absence of an equilibrium solution, *i.e.* in this case, the solution to (3.2) behaves as $x \approx \beta t, y = 0, z = 0$ for large t. To avoid the possibility of an "explosion" of the uninfected cell population and the extinction of the viral populations, in the scope of model (3.2), we consider only κ which is strictly more than ν. It is clear that the y^* and z^* components are linearly proportional to β, meaning that larger β creates an environment with a greater capacity for the viral population.

For the linear stability analysis, we consider the Jacobian of the system (3.2) evaluated at the equilibrium solution obtaining:

$$J = \begin{bmatrix} -\frac{\alpha\,\beta\,(\kappa-\nu)}{\zeta} & 0 & -\frac{\zeta}{\kappa-\nu} \\ \frac{\alpha\,\beta\,(\kappa-\nu)}{\zeta} & -\gamma & \frac{\zeta}{\kappa-\nu} \\ -\frac{\nu\,\alpha\,\beta\,(\kappa-\nu)}{\zeta} & \gamma\,\kappa & -\frac{\zeta\,\kappa}{\kappa-\nu} \end{bmatrix}.$$

However, the eigenvalues of J are too complicated to obtain analytically. Numerical analysis predicts that for large enough β the equilibrium $\{x^*, y^*, z^*\}$ is stable; and, depending on parameters, is either a sink or a stable spiral.

3.3 Constant influx and decay of uninfected cells

The next step is to introduce into the system a death rate for uninfected cells, δ. Although the rest of the model remains the same as in (3.2), there are some changes in the dynamics of the model due to the addition of the death term. The ODEs now read:

$$\begin{aligned} \dot{x} &= \beta - \alpha\,z\,x - \delta\,x, \\ \dot{y} &= \alpha\,z\,x - \gamma\,y, \\ \dot{z} &= \kappa\gamma\,y - \nu\,\alpha\,z\,x - \zeta\,z. \end{aligned} \tag{3.4}$$

System (3.4) has two equilibrium points:

$$\begin{aligned} \boldsymbol{v}_1 &:= \{x = \beta/\delta, \quad y = 0, \quad z = 0\}; \\ \boldsymbol{v}_2 &:= \left\{x = \frac{\zeta}{\alpha\,(\kappa-\nu)}, \quad y = \frac{\beta\,(\kappa-\nu)\alpha - \delta\,\zeta}{\alpha\,(\kappa-\nu)\gamma}, \quad z = \frac{\beta\,(\kappa-\nu)\alpha - \delta\,\zeta}{\alpha\,\zeta}\right\}. \end{aligned} \tag{3.5}$$

Here and further, the asterisks are ignored in equilibrium points that have designated names. The first equilibrium point, $\boldsymbol{v}_1$, corresponds to a virus-free state of the system, when the uninfected cells population attain the maximum carrying capacity of the environment while infected cell population and virus population are zero. The second solution, $\boldsymbol{v}_2$, stands for the virus abundant case. For biological meaningfulness, apart from an essential requirement of positive α, γ, δ, and ζ, it is necessary to keep

$$\kappa > \nu \tag{3.6}$$

and

$$\beta\alpha(\kappa - \nu) \geq \delta\zeta. \tag{3.7}$$

These constraints enforce non-negative population sizes. We see that in order for the virus-abundant solution $\boldsymbol{v}_1$ to be biologically meaningful, the burst size must be strictly more than the multiplicity of infection. Numerical checks on eigenvalues show that the equilibrium states are a stable node and a saddle. A transcritical bifurcation occurs when $\beta\alpha(\kappa-\nu) = \delta\zeta$, and the steady states switch their stability. However, it might be *comme il faut* to admit that one of the equilibria enters the biologically meaningful (*i.e.* positive) phase space when $\beta\alpha(\kappa - \nu) = \delta\zeta$, that is why both equilibria cannot switch their stability when, biologically speaking, the virus abundant steady state never existed "before" the collision of steady states, after which it "becomes" stable.

There is a widely accepted method of linking the basic reproductive ratio, usually noted as R_0, to the growth or decay of viral population. If the basic reproductive ratio is greater than one we observe growth of the studied population. To determine the basic reproductive ratio for this model, we consider inequality (3.7), where the parameter combination that corresponds to equality in (3.7) holds importance for epidemiological discussion. Let us call that parameter combination R_0:

$$R_0 = \frac{\beta\alpha(\kappa - \nu)}{\delta\zeta} \tag{3.8}$$

In terms of inequality (3.7), for this mathematical model of viral dynamics, both equilibria exist in biologically meaningful phase space when $R_0 \geq 1$. By numerically checking the signs of eigenvalues when $R_0 > 1$ and $R_0 < 1$, we confirm that R_0 is the basic reproductive ratio. Indeed, it resembles the basic reproductive ratio from the classic models and has the same properties: it is proportional to all quantities that promote the growth of the viral population in the equations of the system and inversely proportional to almost all quantities that suppress the growth of the viral population. Most importantly, if $R_0 > 1$, then system (3.4) has a stable viral abundant state.

3.4 Birth and death of uninfected cells

This section considers the case when uninfected cells multiply and die at rates $\beta > 0$ and $\delta > 0$, respectively. This is an intuitively realistic version of an independent, undisturbed system. The population dynamics are modelled as follows:

$$\begin{aligned} \dot{x} &= \beta\, x - \alpha\, z\, x - \delta\, x, \\ \dot{y} &= \alpha\, z\, x - \gamma\, y, \\ \dot{z} &= \kappa\gamma\, y - \nu\, \alpha\, z\, x - \zeta\, z. \end{aligned} \tag{3.9}$$

This model has two equilibria:

$$\begin{aligned} \boldsymbol{v}_0 &:= \{x = 0, \quad y = 0, \quad z = 0\}\,; \\ \boldsymbol{v}_1 &:= \left\{ x = \frac{\zeta}{\alpha\,(\kappa - \nu)}, \quad y = \frac{\zeta\,(\beta - \delta)}{\alpha\,(\kappa - \nu)\gamma}, \quad z = \frac{\beta - \delta}{\alpha} \right\}. \end{aligned} \tag{3.10}$$

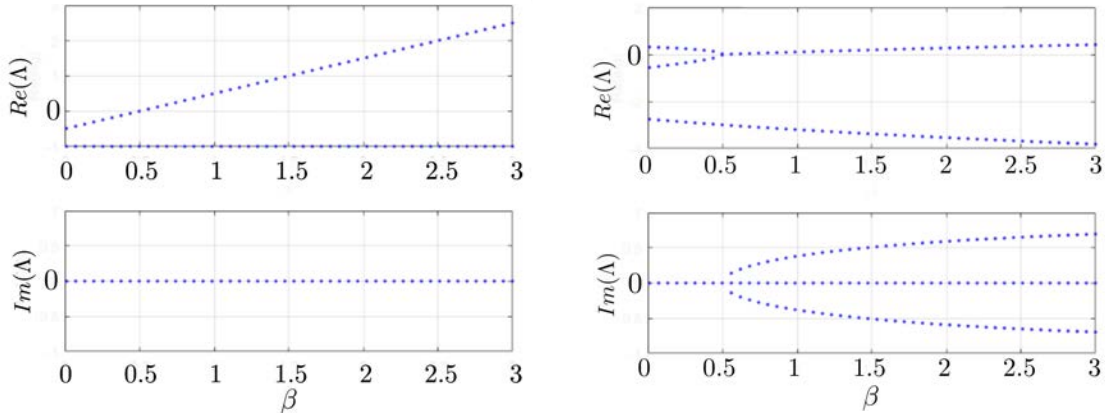
By introducing the exponential growth rate, we can model a system where all populations can become extinct. It is biologically justified, and we must allow the possibility of this pessimistic scenario. Along with the trivial equilibrium $\boldsymbol{v}_0$, there is a virus-abundant state $\boldsymbol{v}_1$. To study the local behaviour of the system around the equilibria, let us consider the linearization of the model. The Jacobian matrix of model (3.9) is

$$J = \begin{bmatrix} \beta - \alpha\, z - \delta & 0 & -\alpha\, x \\ \alpha\, z & -\gamma & \alpha\, x \\ -\alpha\, \nu\, z & \gamma\, \kappa & -\alpha\, \nu\, x - \zeta \end{bmatrix}.$$

For the trivial solution, $\boldsymbol{v}_0$, we can find the eigenvalues (which are the diagonal elements of the triangular Jacobian matrix at the solution):

$$\Lambda = \begin{bmatrix} \beta - \delta \\ -\gamma \\ -\zeta \end{bmatrix}.$$

The eigenvalues at $\boldsymbol{v}_1$ are rather complicated, so we will calculate them numerically further below. Based on Λ and Figure 3.1 (a), we can be sure that the trivial solution is always



(a) Λ at the trivial solution $\boldsymbol{v}_0$

(b) Λ at the nontrivial solution $\boldsymbol{v}_1$

Figure 3.1: Real and imaginary parts of eigenvalues at equilibria for the system (3.9). Other parameters are fixed at $\alpha = 1, \delta = 0.5, \gamma = 1, \kappa = 4, \nu = 2, \zeta = 1$.

a saddle with a one-dimensional unstable manifold ($\dim W^u(\boldsymbol{v}_0) = 1$) if the birth rate of uninfected cells, β, is greater than their death rate, δ. The zero solution becomes stable if there are more uninfected cells dying out than appearing in the system. So, for increasing β, the trivial solution changes from a stable node to a saddle point, while the virus-abundant solution changes from a saddle to an unstable spiral. For $\beta > \delta$, there are no stable equilbria and, therefore, the solution does not have a steady state in this parameter regime.

The virus abundant solution is biologically relevant, that is, all components of $\boldsymbol{v}_1$ are non-negative, if:

- $\beta \geq \delta$, uninfected cell's birth rate is larger than its death rate;
- $\kappa > \nu$, the burst size of a virus is greater than its MOI;
- $\alpha > 0$, virus infects the uninfected cells;
- $\gamma > 0$, infected cells also have positive death rate;
- $\zeta \geq 0$, virus is removed from the system at a non-negative rate.

However, it is possible to obtain a combination of parameters corresponding to different conditions than those listed above for which the populations will still remain positive, *e.g.* see the second steady state. Although it is mathematically relevant, biologically it would drive us away from purposeful discussion. Say, the first condition does not hold, *i.e.* the birth rate of uninfected cells is smaller than their death rate, $\beta \leq \delta$. This would force the infection rate (or fitness) α to be negative in order to have non-negative z. Furthermore, if one comes up with an explanation for *negative fitness*, we will face either ζ or $\kappa - \nu$ being non-positive/negative, in order to have positive population size of uninfected cells.

A "negative death rate of viral population" is another way of saying that there is a total amount of virus based growth of the population, which is admittedly more suitable for cells, because the virus alone lacks the machinery for self-production. In comparison to the negative death rate, "burst size of a virus is smaller than multiplicity of infection" would be a more reasonable path to take, although the burst size is usually greater than the MOI. If we agree on $\beta \leq \delta$, $\alpha < 0$, and $\kappa < \nu$, for non-negative ζ, in order to keep y non-negative, γ has to be less than zero. Some viruses are known for increasing the fitness of their host by infecting it [112], which mathematically lead to a negative death rate of infected cells. Since we do not understand what, in given terms, is *negative fitness*, we stick with the itemised list of conditions above.

Biological relevance by no means implies stability of the equilibrium. So far, we noted the stability of the trivial solution for small β. For the chosen parameters, the real part of at least one eigenvalue at $\boldsymbol{v}_0$ for all β is non-negative, see in Figure 3.1 (b). In order to understand if the nontrivial equilibrium becomes a locally asymptotically stable equilibrium at any parameter combination, consider the characteristic polynomial of the system, which is defined as $\det(J - \lambda I) = 0$. For $\det(J - \lambda I) = 0$, one gets:

$$P(\lambda) = \lambda^3 + \frac{(\kappa\,\gamma - \gamma\,\nu + \zeta\,\kappa)}{\kappa - \nu}\lambda^2 - \frac{\nu\,(\beta - \delta)\,\zeta}{\kappa - \nu}\lambda + (\beta - \delta)\,\gamma\,\zeta = 0. \tag{3.11}$$

From the simulations in Figure 3.1, we see that a real and positive eigenvalue exists for $\beta < \delta$ with the other two eigenvalues being real and negative. In order for the system to gain stability around its nontrivial equilibrium, the positive eigenvalue must cross zero and the others must remain negative. It is worth noting, that there are actually two ways in which $\boldsymbol{v}_1$ could become stable: either $\lambda = 0$ or $Re(\lambda) = 0$ with $Im(\lambda) \neq 0$, assuming that the other eigenvalues have negative real part. The second case is equivalent to search for, as shown further, nonexistent Hopf bifurcation: consider eigenvalue $\lambda = i\,\omega$, where ω must be real; $Re(P(i\,\omega) = 0)$ and $Im(P(i\,\omega) = 0)$ yield contradicting conditions for ω to be real, *i.e.* $\beta > \delta$ and $\delta > \beta$, respectively. Therefore, the virus-free state $\boldsymbol{v}_1$ does not undergo a Hopf bifurcation. Going back to $\lambda = 0$, we obtain from equation (3.11) the following:

$$(\beta - \delta)\gamma\zeta = 0. \tag{3.12}$$

This combination is out of the agreed biologically meaningful parameter space. Therefore, for all biologically meaningful parameter combinations the virus abundant state is unstable. To be able to observe stable nontrivial equilibria, the model must be altered.

3.5 Logistic growth

The previous model is intuitively easy to understand: uninfected cells proliferate and die with a constant rate and also serve as a resource for the production of virus copies. However, real-life systems usually have restrictions on the availability of resources [22]. For instance, the uninfected cells share space and nutrients with each other and thus there is a competition between them. Exponential growth models can be useful only at the early stages of a growth of the population when few cells are present and thus intraspecific competition is negligible. Due to the absence of biological mechanisms restricting the growth of uninfected cells in the previous modelling approach, the solution of model (3.9) "explodes" for most cases, meaning that the number of uninfected cells would go to infinity. In order to avoid uninteresting results, we introduce competition among uninfected cells, which can be considered as intraspecific competition. For the kind of processes being modelled here, the carrying capacity of the environment, K, needs to be introduced. The carrying capacity might be an amount of an environmental characteristic available for uninfected cells, for example, *in vitro* added nutrients and/or available space in a Petri dish. We will assume that the reproduction of uninfected cells is proportional to the growth rate, β, population size, x, and unused fraction of the carrying capacity, $(1 - x/K)$. Under these assumptions, and accounting for the death rate of competing uninfected cells, δx, the current model becomes:

$$\begin{aligned} \dot{x} &= \beta\, x(1 - x/K) - \alpha\, z\, x - \delta\, x, \\ \dot{y} &= \alpha\, z\, x - \gamma\, y, \\ \dot{z} &= \kappa\gamma\, y - \nu\, \alpha\, z\, x - \zeta\, z. \end{aligned} \tag{3.13}$$

The system (3.13) has three equilibria:

$$\begin{aligned} \boldsymbol{v}_0 &:= \{x = 0, \quad y = 0, \quad z = 0\}\,; \\ \boldsymbol{v}_1 &:= \left\{x = \frac{(\beta - \delta)\, K}{\beta}, \quad y = 0, \quad z = 0\right\}; \\ \boldsymbol{v}_2 &:= \left\{x = \frac{\zeta}{\alpha\,(\kappa - \nu)}, \quad y = \frac{\zeta\,((\beta - \delta)(\kappa - \nu)\alpha\, K - \beta\,\zeta)}{\alpha^2(\kappa - \nu)^2\gamma\, K}, \quad z = \frac{(\beta - \delta)(\kappa - \nu)\alpha\, K - \beta\,\zeta}{\alpha^2(\kappa - \nu)\, K}\right\}. \end{aligned}$$

The trivial state $\boldsymbol{v}_0$ stands for the case when no population survives; a virus-free state, $\boldsymbol{v}_1$, leads to an abundance of uninfected cells growing to the maximum carrying capacity of the system; a virus-abundant state, $\boldsymbol{v}_2$, has nonzero values for all three populations. There are a couple of obvious parameter relations for which the equilibria would intersect

or coalesce. For example, if the growth rate and the death rate of uninfected cells are equal, $\beta = \delta$, then the virus-free state would match with the trivial state, *i.e.* $\boldsymbol{v}_0 = \boldsymbol{v}_1$; if $(\beta - \delta)(\kappa - \nu)\alpha K = \beta\zeta$, then the virus-abundant state would coalesce with the virus-free state, *i.e.* $\boldsymbol{v}_1 = \boldsymbol{v}_2$. These combinations are strong candidates for bifurcation points involving collisions of fixed points. To study the dynamics of the model, we investigate the stability of its equilibria using linear stability analysis. The Jacobian of the model (3.13) is the following matrix:

$$J = \begin{bmatrix} \beta - \alpha z - \delta - 2\beta x/K & 0 & -\alpha x \\ \alpha z & -\gamma & \alpha x \\ -\alpha \nu z & \gamma \kappa & -\alpha \nu x - \zeta \end{bmatrix}.$$

Even though the Jacobian is quite simple, the eigenvalues, needed for the linear stability analysis, are complicated for the nonzero solutions. However, as in the previous case, we have a simple expression of the eigenvalues at the trivial solution:

$$\Lambda = \begin{bmatrix} \beta - \delta \\ -\gamma \\ -\zeta \end{bmatrix}.$$

The trivial solution becomes unstable when the growth rate is larger than the death rate of the uninfected cell population. In order to know more about the stability of the solutions, we study the dependency of the eigenvalues on parameters numerically. Our first attempt is to vary β fixing other parameters. Over a biologically meaningful range, the virus-free solution begins as a node and changes from stable to unstable at increasing β. At the same range, the second solution changes from a saddle to a stable node.

Let us fix β large, resulting in the presence of a stable $\boldsymbol{v}_1$, and study the system tuning the infection rate, α. We investigate the system by numerically calculating all eigenvalues for varying α, and simultaneously tracking the value of viral load component, z. As it is clear from Figure 3.2, for increasing α, the virus-free solution loses its stability (illustrated in Figure 3.3 (a)) and becomes a saddle, see the red curve of $\mathrm{Re}(\Lambda)$ crossing zero from below. At that exact value of α, the virus-abundant solution (in black) changes from a locally asymptotically saddle to a stable node (illustrated in Figure 3.3 (b)). Further increases in α lead to the eigenvalues at $\boldsymbol{v}_2$ becoming imaginary, transforming this equilibrium into a spiral attractor (illustrated in Figure 3.3 (c)). The stability of $\boldsymbol{v}_2$ changes again when $\alpha = 1.7$ and the real part of the complex eigenvalues becomes positive. This transition from

a stable spiral into a stable periodic orbit is called a Hopf bifurcation, which is characterised by the Jacobian matrix having a single pair of complex conjugate eigenvalues with zero real part, and results in the creation of a stable periodic orbit.

Two types of bifurcations were found to occur in system (3.13): transcritical and Hopf bifurcations. To look at the changes occurring in the dynamics of the model, a two-dimensional bifurcation diagram has been constructed in the parameter space (α, γ). The lines inside this space represent the Hopf (blue) and transcritical (orange) bifurcations (see Figure 3.4). By checking eigenvalues of the Jacobian, it is feasible to get the exact values of bifurcation points. The bifurcation diagram shown in Figure 3.4 is divided by two curves into three regions, conditionally named left, middle, and right. The left area satisfies the case when for all values of parameters, the system exhibits an abundance of uninfected cells in the absence of virus and infected cells, *i.e.* the equilibrium $\boldsymbol{v}_1$ is stable. The middle region is separated from left one by a transcritical bifurcation curve, meaning that by crossing that curve the steady states $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ interchange their stability. The virulence of the virus (interpreted as γ) seems to play a little role if its fitness is lower than some threshold value and the population of virus is destined to extinction. The threshold values is given by $\alpha = \beta\,\zeta / (\beta - \delta)(\kappa - \nu) K$, whose independence from γ determines its constant appearance on the bifurcation diagram. The middle region has a stable virus-abundant solution. Moreover, if the parameter pair (α, γ) is close to the line of transcritical bifurcations, then state $\boldsymbol{v}_2$ is a stable node, and if the pair is close to the right region then the the steady state is an asymptotically stable point which is approached via damped oscillations. In the right region, which is separated from the middle region by a Hopf bifurcation curve, all equilibria are unstable and there exists a stable periodic orbit around the virus-abundant solution. This means that for all parameters in the right region, all populations are positive and change periodically. For $\gamma = 1$, the four mentioned behaviours of system (3.13), namely stable virus-free state, stable virus-abundant node, stable virus-abundant spiral, and stable periodic orbit, are illustrated in Figure 3.5.

3.6 Nondimensionalization

Before moving to more complicated models, it is useful to first select a base model from those considered in this chapter, which will be further developed in subsequent chapters. We then show how the number of parameters can be reduced in the base model by performing a nondimensionalisation. The most rich, yet realistic, dynamics are provided by model (3.13). It is also the most epidemiologically truthful model listed so far. Reducing the number of free parameters and finding the persistent relations among them can be done

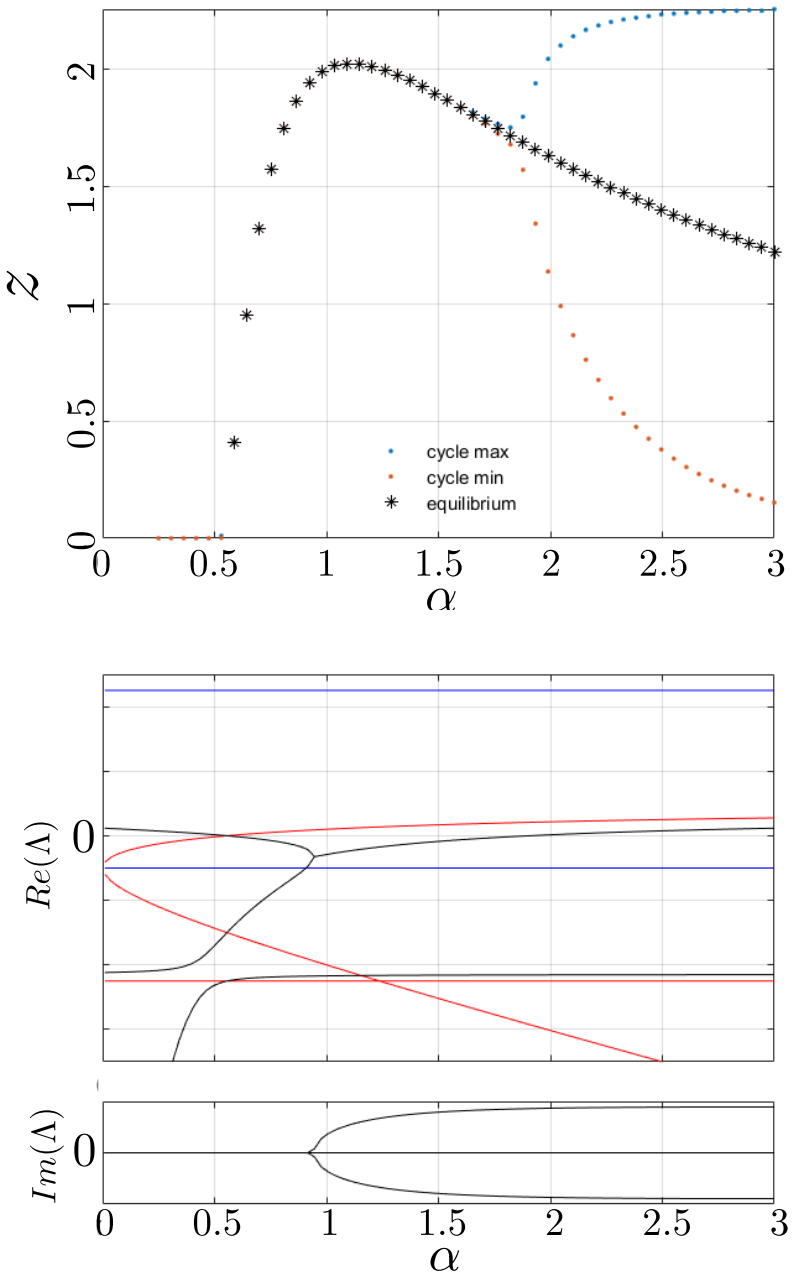
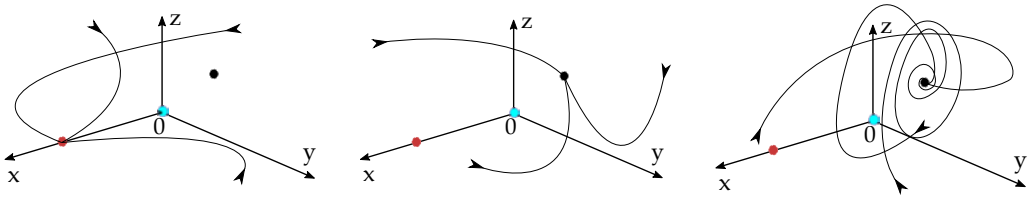


Figure 3.2: Upper plot: numerically obtained values of z (viral load) component of the solution for $t \to \infty$ with initial condition $x = 1, y = 0, z = 0.5$ for different values of α. For small values of α, the virus-free state persists. For $0.5 < \alpha < 1.7$, the stable virus-abundant state is shown in stars. For greater values of α, the virus-abundant state loses stability gaining a stable periodic orbit. Blue and orange points correspond to maximum and minimum values of the stable periodic orbit, respectively. Lower plot: Blue lines correspond to the eigenvalues of the Jacobian at $\boldsymbol{v}_0$, red lines stand for the eigenvalues at the virus-free solution, $\boldsymbol{v}_1$, and black lines correspond to the eigenvalues of the virus-abundant solution of (3.13), given by $\boldsymbol{v}_2$. Fixed parameters used for these simulations: $\beta = 5, \delta = 0.5, \gamma = 1, \kappa = 4, \nu = 2, \zeta = 1$.



(a) $\boldsymbol{v}_1$ is an attractor (b) $\boldsymbol{v}_2$ is a stable node (c) $\boldsymbol{v}_2$ is a stable spiral

Figure 3.3: Solutions of model (3.13) schematically represented in the biologically meaningful phase space for three different values of α increasing from panel (a) to panel (c), and corresponding to stability of different equilibrium points, $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$. The blue, red, and black circles represent the trivial, virus-free, and virus-abundant solutions, respectively. Three initial conditions for each plot are chosen randomly. The arrows along curves indicate the direction of the trajectories.

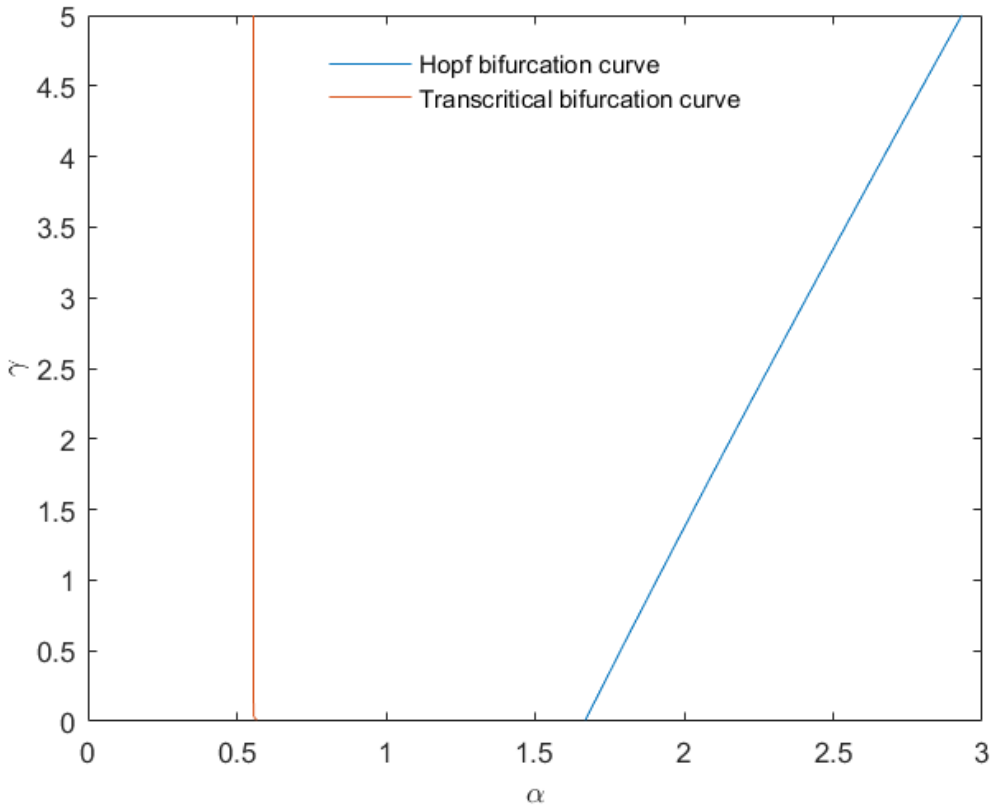
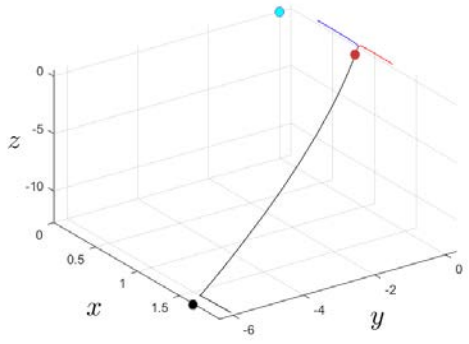
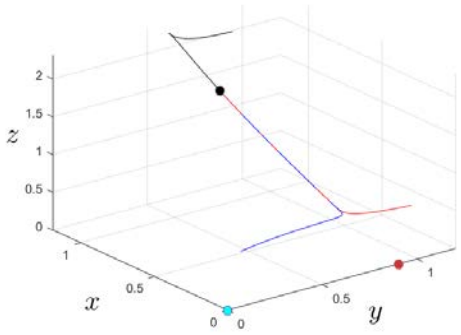


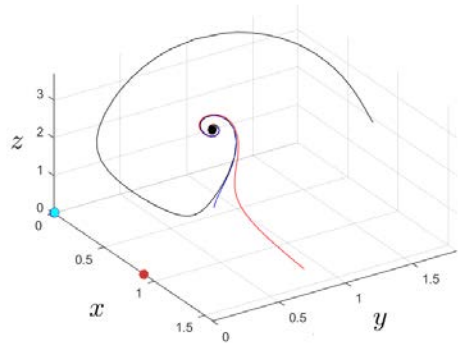
Figure 3.4: Two-dimensional bifurcation diagram for model (3.13), using α and γ as control parameters. Fixed parameters used for these simulations: $\beta = 5, \delta = 0.5, \kappa = 4, \nu = 2$, and $\zeta = 1$.



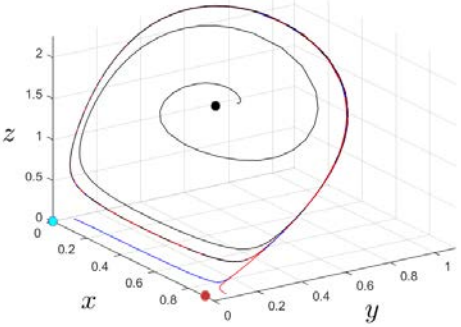
(a) $\boldsymbol{v}_1$ is a sink, $\alpha = 0.5$



(b) $\boldsymbol{v}_2$ is a sink, $\alpha = 0.8$



(c) $\boldsymbol{v}_2$ is a stable spiral, $\alpha = 1.5$



(d) $\boldsymbol{v}_2$ has a stable periodic orbit, $\alpha = 3$

Figure 3.5: Numerical solutions of model (3.13) for four values of α, corresponding to four different behaviours of the system. Three initial conditions for each plot are chosen randomly. The blue, red, and black circles represent the trivial, virus-free, and virus-abundant solutions, respectively. Fixed parameters used for these simulations: $\beta = 5, \delta = 0.5, \gamma = 1, \kappa = 4, \nu = 2, \zeta = 1$.

by a nondimensionalization of the model (3.13). In terms of the old variables t, x, y, z, new variables $\bar{t}, \bar{x}, \bar{y}, \bar{z}$ are defined in the following way:

$$t = (\beta - \delta)^{-1}\bar{t}, \quad x = \left(1 - \frac{\delta}{\beta}\right) K\,\bar{x}, \quad y = \left(1 - \frac{\delta}{\beta}\right) K\,\bar{y}, \quad z = \left(1 - \frac{\delta}{\beta}\right) \kappa K\,\bar{z}.$$

New non-dimensional parameters substitute the corresponding old parameters as follows:

$$\bar{\alpha} = \frac{\kappa\, K}{\beta}\alpha, \quad \bar{\gamma} = \frac{\gamma}{\beta - \delta}, \quad \bar{\nu} = \frac{\nu}{\kappa}, \quad \bar{\zeta} = \frac{\zeta}{\beta - \delta}.$$

Ignoring bars merely for notational convenience, we obtain the nondimensionalized system

$$\begin{aligned} \dot{x} &= x(1 - x) - \alpha\, z\, x, \\ \dot{y} &= \alpha\, z\, x - \gamma\, y, \\ \dot{z} &= \gamma\, y - \nu\, \alpha\, z\, x - \zeta\, z. \end{aligned} \tag{3.14}$$

The system (3.14) has three equilibria:

$$\begin{aligned} \boldsymbol{v}_0 &:= \{x = 0, \quad y = 0, \quad z = 0\}\,; \\ \boldsymbol{v}_1 &:= \{x = 1, \quad y = 0, \quad z = 0\}\,; \\ \boldsymbol{v}_2 &:= \left\{x = \frac{\zeta}{\alpha\,(1 - \nu)}, \quad y = \frac{\zeta\,((1 - \nu)\alpha - \zeta)}{\alpha^2(1 - \nu)^2\gamma}, \quad z = \frac{(1 - \nu)\alpha - \zeta}{\alpha^2(1 - \nu)}\right\}. \end{aligned} \tag{3.15}$$

As expected, by undergoing the same analysis, we find that model (3.14) leads to results quantitatively matching with those of model (3.13) for ranges of parameters and variables chosen according to the nondimensionalization. A discussion of these results is omitted to avoid repetition. However, in comparison to its superior – in the sense of the straightforwardness of biological interpretation – parent model, the number of free parameters of system (3.14) decreased to four. Keeping the number of free parameters minimised is a useful trait for the extended systems based on model (3.13).

Chapter 4

Coexistence of wild-type and mutant viral strains

The main results of this chapter are published in Nurtay A, Hennessy MG, Sardanyés J, Alsedà L, Elena SF. 2019 Theoretical conditions for the coexistence of viral strains with differences in phenotypic traits: a bifurcation analysis. *R. Soc. open sci.* 6: 181179. http://dx.doi.org/10.1098/rsos.181179. The more detailed discussion of this work is presented here, below.

In this chapter, two different strains of a virus are introduced: the wild-type (wt) and mutant strains. This allows us to model the evolutionary mechanisms which determine the long-term coexistence of different genetic variants and strains that, in principle, compete for the same resources (*e.g.* susceptible cells). Understanding the evolutionary forces of such intraspecific competition or strain coexistence are a crucial component for understanding the long-term fate and composition of viral populations and for a thoughtful design of more robust control strategies for known and future outbreaks [121]. Consequently, the coexistence of evolving pathogens has been the target of extensive research [79, 39, 104, 117].

In the mathematical theory of population genetics, mutation, which is understood as any change in the genome of an organism, is often modelled as a "flow" between populations of initial wt individuals and emerging mutant individuals [8, 18, 109, 30]. Here, we introduce mutation into a mathematical model in a similar way although avoiding forceful restrictions put upon mutant strains and allowing the mutant virus to have similar characteristics as the wt virus. The difference between mutant and wt strains occurs when focusing the study on specific phenotypic characteristics. One of the main characteristics to focus on when investigating the survival of a population in dynamical systems that present competition or coexistence [94, 93, 75], is fitness. In the case of virus evolution, fitness is often considered to be proportional to the infection rate of the virus [80, 140, 136]. It is known that infection rates differ between strains, a fact that has significant implications for the evolution of virulence and strain coexistence in nature [137]. Even the balance between genetic diversity and competition is believed to be achieved due to the possibility of coexistence among strains with differences in infection rates [9]. In other words, a direct competition for infecting available cells mediates the stable coexistence only when competitive abilities in viral clones satisfy certain pairwise asymmetries [46]. A secondary phenotypic characteristic that can differ between viral strains is their strategy for exploiting the host cell, *i.e.* their virulence [107]. The evolution of virulence has received great attention from theoreticians, particularly on the coevolution between resistance and virulence traits and their combined effect on host and virus dynamics [15, 139]. However, understanding the evolution of virulence for coexisting viral strains still requires attention due to the complexity of the underlying evolutionary and dynamical processes, being inherently nonlinear. Many of the models brought forward to explain the evolution of virulence take into consideration the processes of coinfection and superinfection [96, 23, 33, 17], where the host is infected simultaneously or sequentially by more than one pathogen strain. Here and further, infection will be modelled at the single-cell level and coinfection will therefore be neglected. The benefit of this approach is that analytical insights into coexistence of viral strains can be obtained.

The purpose of this chapter is to illustrate, by means of a dynamical mathematical model, the conditions for coexistence of viral strains that considers both a wt viral strain and its mutants. We present analytical and numerical results focusing on the parameters related to the differential phenotypic traits of the wt and mutant strains. Conditions for coexistence and invasion have previously been studied using a mathematical model of one host shared between two competing parasites [54]. However, this model did not incorporate the mutation of viruses as a factor, and thus neglected the input of new strains into the system. Nearly all mathematical models in epidemiology that detect various dynamical behaviours with multi-strain infections illustrate the necessity of numerical approaches and the dependency of such models on a large number of parameters [62, 91, 10, 1, 16]. A classical approach used in epidemiology [84] to circumvent this difficulty is to introduce dimensionless parameter groups, such as the basic reproduction number R_0, as in Ref. [135]. However, the number of dimensionless groups can still become large as additional complexity is introduced into the model, as is the case here. Thus, we perform a bifurcation analysis to systematically track how the dynamics of the system change as multiple parameters are varied. In general, the study of parameters in terms of their effect on the stability of certain states of a model is a highly effective way to gain important insights into the investigated system [81, 53, 85].

4.1 Modelling mutation and coexistence

Assume that the wt strain of a virus, z, infects a susceptible cell, x, and creates an infected cell, y, which is linked to the wild type of the pathogen. The infected cell y mostly generates the wt strains, but with rate μ it produces other types of virus. Since strains can change and mutations occur predominantly during the replication and production of daughter virions by an infected cell, the mutation of virus is considered as a population decay/growth of infected cells. All viral strains differing from the wt are called mutant strains, $z^{\square}$. Now, both virus types, $z^{\square}$ and z, infect the susceptible cell population, x, and infected cells become linked to the viral strain, $y^{\square}$ and y. Via infected cells the virus replicates and produces its own type of virions. Moreover, as cells from the infected population y mutate with rate μ into cells of population $y^{\square}$, the size of the latter population increase at the same rate μ. This, implicitly, affects the growth of the population of mutant strain $z^{\square}$. The other processes are the same as in model (3.13), which hold for each type of strain. The time evolution of the

interacting population is described as follows:

$$\begin{aligned}
\dot{x} &= \beta\, x(1 - x/K) - \alpha^{\square}\, z^{\square}\, x - \alpha\, z\, x - \delta\, x,\\
\dot{y}^{\square} &= \alpha^{\square}\, z^{\square}\, x + \mu y - \gamma^{\square}\, y^{\square},\\
\dot{y} &= \alpha\, z\, x - \mu y - \gamma\, y,\\
\dot{z}^{\square} &= \kappa^{\square}\, \gamma^{\square}\, y^{\square} - \nu^{\square}\, \alpha^{\square}\, z^{\square}\, x - \zeta^{\square}\, z^{\square},\\
\dot{z} &= \kappa\, \gamma\, y - \nu\, \alpha\, z\, x - \zeta\, z.
\end{aligned} \tag{4.1}$$

Initial conditions for the system (4.1) are non-negative values:

$$x(0) = X, \quad y^{\square}(0) = Y^{\square}, \quad y(0) = Y, \quad z^{\square}(0) = Z^{\square}, \quad z(0) = Z.$$

Short-hand notation for the five-dimensional vector of initial conditions is $\boldsymbol{v}(0) = \{X, Y^{\square}, Y, Z^{\square}, Z\}$. If not mentioned otherwise, all the numerical simulations will have $Y^{\square} = 0$ and $Y = 0$ due to the initial portion of the infected cells in a simulated batch being zero. Usually, we introduce only healthy cells and the necessary strains of virus allowing the infected cells to appear during the simulations. This system has four equilibria:

$$\{x = 0, \quad y^{\square} = 0, \quad y = 0, \quad z^{\square} = 0, \quad z = 0\};$$

$$\left\{x = \frac{(\beta - \delta)K}{\beta}, \quad y^{\square} = 0, \quad y = 0, \quad z^{\square} = 0, \quad z = 0\right\};$$

$$\left\{x = \frac{\zeta^{\square}}{\alpha^{\square}\,(\kappa^{\square} - \nu^{\square})}, \quad y^{\square} = \frac{(K\,(\beta - \delta)\,(\kappa^{\square} - \nu^{\square})\,\alpha^{\square} - \beta^{\square}\,\zeta)\,\zeta^{\square}}{K\,(\kappa^{\square} - \nu^{\square})^2\,(\alpha^{\square})^2\gamma^{\square}}, \quad y = 0,\right.$$
$$\left. z^{\square} = \frac{K\,(\beta - \delta)\,(\kappa^{\square} - \nu^{\square})\,\alpha^{\square} - \beta^{\square}\,\zeta^{\square}}{K\,(\kappa^{\square} - \nu^{\square})\,(\alpha^{\square})^2}, \quad z = 0\right\};$$

$$\left\{x = \frac{\zeta(\gamma + \mu)}{A\alpha}, \quad y^{\square} = \frac{B\mu\zeta(C + \alpha^{\square}\gamma\kappa^{\square}\zeta)}{A^2\,C(\alpha)^2\gamma^{\square}K}, \quad y = \frac{B\zeta(C - \alpha^{\square}\mu\kappa^{\square}\zeta)}{A^2C(\alpha)^2K},\right.$$
$$\left. z^{\square} = \frac{B\mu\kappa^{\square}\zeta}{AC\alpha K}, \quad z = \frac{B(C - \alpha^{\square}\mu\kappa^{\square}\zeta)}{AC(\alpha)^2K}\right\};$$

where

$$A = \gamma^{\square}\kappa - (\mu+\gamma)\nu;$$

$$B = \left(\gamma^{\square}\kappa - (\mu+\gamma)\nu\right)(\beta-\delta)K\alpha - \beta\zeta\ (\mu+\gamma)\,;$$

$$C = \alpha\zeta^{\square}\left(\gamma^{\square}\kappa - (\mu+\gamma)\nu\right) - \alpha^{\square}\zeta\left(\gamma(\kappa^{\square}-\nu^{\square}) - \mu\nu^{\square}\right).$$

Nondimensionalization simplifies model (4.1), and old variables in terms of new non-dimensional variables would be as follows:

$$t = (\beta-\delta)^{-1}\bar{t}, \quad x = \left(1-\frac{\delta}{\beta}\right)K\,\bar{x}, \quad y^i = \left(1-\frac{\delta}{\beta}\right)K\,\bar{y}^i, \quad z^i = \left(1-\frac{\delta}{\beta}\right)\kappa^{\square}K\,\bar{z}^i,$$

and new non-dimensional parameters substitute corresponding old parameters:

$$\bar{\alpha}^i = \frac{\kappa^{\square}\,K\alpha^i}{\beta}, \quad \bar{\nu}^i = \frac{\nu^i}{\kappa^{\square}}, \quad \bar{\kappa} = \frac{\kappa}{\kappa^{\square}}, \quad \bar{\mu} = \frac{\mu}{\beta-\delta}, \quad \bar{p}^i = \frac{p^i}{\beta-\delta},$$

where p stands for all other parameters, namely: $\gamma^{\square}, \gamma, \zeta^{\square}$, and ζ. Here, i over variables and parameters stands either for a square $^{\square}$ or its absence.

The nondimensionalization puts its restrictions onto the model, but these restrictions are biologically justified. For example, taking $\beta > \delta$, as shown later, forces the trivial equilibrium to be unstable. However, letting the growth rate of uninfected cells being strictly greater than death rate of uninfected cells is an acceptable trade. Ignoring bars, as before with system (3.14), for a better appearance of the current model, we obtain the nondimensionalized system

$$\begin{aligned}
\dot{x} &= x(1-x) - \alpha^{\square}\,z^{\square}\,x - \alpha\,z\,x,\\
\dot{y^{\square}} &= \alpha^{\square}\,z^{\square}\,x + \mu y - \gamma^{\square}\,y^{\square},\\
\dot{y} &= \alpha\,z\,x - \mu y - \gamma\,y,\\
\dot{z^{\square}} &= \gamma^{\square}\,y^{\square} - \nu^{\square}\,\alpha^{\square}\,z^{\square}\,x - \zeta^{\square}\,z^{\square},\\
\dot{z} &= \kappa\gamma\,y - \nu\,\alpha\,z\,x - \zeta\,z.
\end{aligned} \tag{4.2}$$

The equilibria of the model remains qualitatively the same and in terms of non-dimensional

parameters, who have lost their bars for aesthetic purposes, look as follows:

$$\boldsymbol{v}_0 := \{x = 0, \quad y^{\square} = 0, \quad y = 0, \quad z^{\square} = 0, \quad z = 0\};$$

$$\boldsymbol{v}_1 := \{x = 1, \quad y^{\square} = 0, \quad y = 0, \quad z^{\square} = 0, \quad z = 0\};$$

$$\boldsymbol{v}_2 := \left\{ x = \frac{\zeta^{\square}}{\alpha^{\square}\,(1-\nu^{\square})}, \quad y^{\square} = \frac{\left(\alpha^{\square}\,(1-\nu^{\square}) - \zeta^{\square}\right)\zeta^{\square}}{(\alpha^{\square})^2\,(1-\nu^{\square})^2\,\gamma^{\square}}, \quad y = 0, \right.$$
$$\left. z^{\square} = \frac{\alpha^{\square}\,(1-\nu^{\square}) - \zeta^{\square}}{(\alpha^{\square})^2\,(1-\nu^{\square})}, \quad z = 0 \right\};$$

$$\boldsymbol{v}_3 := \left\{ x = \frac{\zeta(\gamma+\mu)}{A\alpha}, \quad y^{\square} = \frac{B\mu\zeta(C+\alpha^{\square}\gamma\zeta)}{A^2\,C(\alpha)^2\gamma}, \quad y = \frac{B\zeta(C-\alpha^{\square}\mu\zeta)}{A^2 C(\alpha)^2}, \right.$$
$$\left. z^{\square} = \frac{B\mu\zeta}{AC\alpha}, \quad z = \frac{B(C-\alpha^{\square}\mu\zeta)}{AC(\alpha)^2} \right\};$$

where

$$A = \gamma\kappa - (\mu+\gamma)\nu;$$

$$B = -\left(\left((\nu-\kappa)\,\alpha + \zeta\right)\gamma - \mu\,\left(\alpha\,\nu + \zeta\right)\right);$$

$$C = -\left(\zeta\,\left(\nu-\kappa\right)\alpha - \alpha\,\zeta\,\left(\nu-1\right)\right)\gamma - \mu\,\left(-\alpha\,\nu\,\zeta + \alpha\,\nu\,\zeta\right).$$

The trivial state, $\boldsymbol{v}_0$, corresponds to the case when all populations go extinct. At the virus-free state, $\boldsymbol{v}_1$, only the population of uninfected cells have a positive size with other populations being zero. The third solution, which is the wt-free state $\boldsymbol{v}_2$, describes a case when the mutant strain outcompetes the wt strain, the population of which is zero, and the infected cells corresponding to wt strain is also zero with all other populations having nonzero values. The coexistence state, $\boldsymbol{v}_3$, is the most elaborate one because neither of the populations at this state is a constant zero. Further we study the stability of these four states.

4.1.1 Differences in infection rates

Linear stability analysis close to the equilibrium points can elucidate the overall dynamics of the system (4.2). First, consider the Jacobian matrix for a vector $\boldsymbol{v} := \{x, y^{\square}, y, z^{\square}, z\}$:

$$J = \begin{bmatrix} -\alpha^{\square} z^{\square} - \alpha z - 2x + 1 & 0 & 0 & -\alpha x & -\alpha^{\square} x \\ \alpha z & -\gamma - \mu & 0 & \alpha x & 0 \\ \alpha^{\square} z^{\square} & \mu & -\gamma^{\square} & 0 & \alpha^{\square} x \\ -\alpha \nu z & \kappa \gamma & 0 & -\alpha \nu x - \zeta & 0 \\ -\alpha^{\square} \nu^{\square} z^{\square} & 0 & \gamma^{\square} & 0 & -\alpha^{\square} \nu^{\square} x - \zeta^{\square} \end{bmatrix}$$

Straightforward calculations for the trivial solution yield eigenvalues

$$\boldsymbol{\Lambda}_0 = (1, -\zeta, -\zeta^{\square}, -\gamma^{\square}, -\gamma - \mu)^T . \tag{4.3}$$

For all the agreed values of parameters, the trivial solution $\boldsymbol{v}_0$, as mentioned before, is always unstable, *i.e.* a saddle. The instability of $\boldsymbol{v}_0$ hints on possible stability of the next equilibrium. Although it is still possible to find analytical expressions for the eigenvalues of the equilibrium $\boldsymbol{v}_1$, they are not as presentable as $\boldsymbol{\Lambda}_0$. In order to perform the stability analysis for this equilibrium, it is more useful to consider the characteristic polynomial of the Jacobian at $\boldsymbol{v}_1$. The polynomial $\det(J - \lambda \boldsymbol{I}) = 0$ can be regrouped in the form of three factors $P_1(\lambda) \cdot P_2(\lambda) \cdot P_3(\lambda) = 0$, where λ is an eigenvalue:

$$\begin{aligned} &\underbrace{(\lambda + 1)}_{P_1} \cdot \\ &\cdot \underbrace{(\lambda^2 + (\alpha\,\nu + \gamma + \mu + \zeta)\,\lambda - \alpha\,\gamma\,\kappa + \alpha\,\gamma\,\nu + \alpha\,\mu\,\nu + \gamma\,\zeta + \mu\,\zeta)}_{P_2} \cdot \\ &\cdot \underbrace{(\lambda^2 + (\alpha^{\square}\,\nu^{\square} + \gamma^{\square} + \zeta^{\square})\,\lambda + \alpha^{\square}\,\gamma^{\square}\,\nu^{\square} - \alpha^{\square}\,\gamma^{\square} + \gamma^{\square}\,\zeta^{\square})}_{P_3} = 0 \end{aligned} \tag{4.4}$$

The first factor provides a constant eigenvalue, $\lambda = -1$, which reserves the possibility for stability of $\boldsymbol{v}_1$. Although the next two factors P_2 and P_3 do not provide obvious eigenvalues, they can lead to a hint for bifurcation points. Based on our knowledge of two kinds of bifurcations that appear in the previous models, which are similar to this one, we may expect to find transcritical and Hopf bifurcations. A Hopf bifurcation for the equilibrium

$\boldsymbol{v}_1$ is impossible because this would require a complex conjugate pair of eigenvalues with zero real part, which in terms of P_2 and P_3 means that their linear coefficients must be zero. However, both linear coefficients, $\alpha\,\nu + \gamma + \mu + \zeta$ of P_2 and $\alpha^{\square}\,\nu^{\square} + \gamma^{\square} + \zeta^{\square}$ of P_3, are strictly positive. A transcritical bifurcation occurs when the Jacobian has a zero eigenvalue and two equilibria non-destructively collide. In terms of the quadratic expressions for eigenvalues, which have positive linear coefficients, we equate their free terms (*i.e.* their constant coefficients) to zero to obtain a pair of eigenvalues, one of which is zero and the other is nonzero, for each of the P_2 and P_3. In other words, by solving $P_2(\lambda = 0) = 0$ and $P_3(\lambda = 0) = 0$, we obtain conditions for the transcritical bifurcation:

$$\begin{aligned} -\alpha\,\gamma\,\kappa + \alpha\,\gamma\,\nu + \alpha\,\mu\,\nu + \gamma\,\zeta + \mu\,\zeta = 0, \\ \alpha^{\square}\,\gamma^{\square}\,\nu^{\square} - \alpha^{\square}\,\gamma^{\square} + \gamma^{\square}\,\zeta^{\square} = 0. \end{aligned} \tag{4.5}$$

These conditions can be rewritten as $R_0 = 1$, where

$$R_0 := \frac{\alpha}{\zeta}\left(\frac{\kappa}{1 + \mu\,\gamma^{-1}} - \nu\right), \tag{4.6}$$

and, similarly, as $R_0^{\square} = 1$, where

$$R_0^{\square} := \frac{\alpha^{\square}}{\zeta^{\square}}\left(1 - \nu^{\square}\right). \tag{4.7}$$

It will be shown below that R_0 and $R_0^{\square}$ are the basic reproductive numbers for the wt and mutant strains, respectively. When $R_0 = 1$, the virus-free state $\boldsymbol{v}_1$ and the coexistence state $\boldsymbol{v}_3$ intersect. By expanding $\boldsymbol{v}_3$ around $R_0 = 1$, we find that it has negative components when $R_0 < 1$ and thus lies outside of the biologically meaningful phase space. However, all of the components of $\boldsymbol{v}_3$ become positive when $R_0 > 1$. Likewise, expanding the equilibrium around $R_0^{\square} = 1$ shows that this condition corresponds to points where the virus-free state $\boldsymbol{v}_1$ and wt-free state $\boldsymbol{v}_2$ intersect. For $R_0^{\square} < 1$, some components of $\boldsymbol{v}_2$ are negative; for $R_0^{\square} > 1$, all components are positive. In the case when both $R_0 = R_0^{\square} = 1$ hold, there is a triple intersection of $\boldsymbol{v}_1$, $\boldsymbol{v}_2$, and $\boldsymbol{v}_3$. Furthermore, the Jacobian has a double zero eigenvalue at this point, indicating the onset of a non-degenerate or degenerate Bogdanov-Takens bifurcation. As will be shown in Section 4.3, the Bogdanov-Takens bifurcations in this model are of degenerate type. More generally, we find that $\boldsymbol{v}_2$ and $\boldsymbol{v}_3$ intersect when

$R_0^{\square} = R_0$, that is, when

$$\frac{\alpha}{\zeta}\left(\frac{\kappa}{1+\mu\,\gamma^{-1}} - \nu\right) = \frac{\alpha^{\square}}{\zeta^{\square}}(1-\nu^{\square}). \tag{4.8}$$

Thus, there are three curves of transcritical bifurcations defined by $R_0 = 1$, $R_0^{\square} = 1$, and $R_0^{\square} = R_0$, all of which simultaneously intersect at the DBT point. Whenever either of these combinations holds, there is a bifurcation point. Expressing from (4.6) and (4.7) α and $\alpha^{\square}$, respectively, yields

$$\alpha_{\mathrm{tr}} = \frac{\gamma\,\zeta + \mu\,\zeta}{\gamma(\kappa-\nu) - \mu\,\nu}, \qquad \alpha_{\mathrm{tr}}^{\square} = \frac{\zeta^{\square}}{1-\nu^{\square}}. \tag{4.9}$$

The obtained expressions would be equivalent if there was no μ (keeping in mind that κ was taken as a ratio of κ and $\kappa^{\square}$). By substituting the expression for $\alpha^{\square}$ into $\boldsymbol{v}_2$ we see it becoming equal to $\boldsymbol{v}_1$. Similarly, substituting the expression for α into $\boldsymbol{v}_3$, we see it becoming equal to $\boldsymbol{v}_1$, too. By checking numerically the signs of eigenvalues around these points, we can conclude that (4.9) are the transcritical bifurcation points between $\boldsymbol{v}_1$ and $\boldsymbol{v}_3$ and between $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$, respectively.

Based on these findings, we can begin to schematically sketch the one- and two-dimensional bifurcation diagrams: the constant equilibrium $\boldsymbol{v}_1$ crosses $\boldsymbol{v}_2$ at value of $\alpha_{\mathrm{tr}}^{\square}$ and the transcritical bifurcation leads to these equilibria exchanging their stability. Along another parameter axis, $\boldsymbol{v}_1$ crosses $\boldsymbol{v}_3$ at value of α_{tr} and it also swaps their stability. The two-dimensional bifurcation diagram with α and $\alpha^{\square}$ on the axis, shown in Figure 4.1, depicts four regions separated from one another by a horizontal and vertical line representing curves of transcritical bifurcations defined by the expressions in (4.9). The lines separating the four regions in Figure 4.1 are horizontal and vertical because the expressions for α_{tr} and $\alpha_{\mathrm{tr}}^{\square}$ in (4.9) are mutually independent. The values of the other parameters are set to

$$\mu = 0.1, \quad \kappa = 1, \quad \gamma^{\square} = \gamma = 0.25, \quad \nu^{\square} = \nu = 0.5, \quad \zeta^{\square} = \zeta = 0.2222. \tag{4.10}$$

By numerically calculating the eigenvalues near the intersection of the two bifurcation curves, we find that region I with smaller values of both parameters is the region where the equilibrium $\boldsymbol{v}_1$ is stable. Right after crossing the boundary into region II, $\boldsymbol{v}_3$ becomes stable. According to the eigenvalues of the Jacobian for parameters of region III, $\boldsymbol{v}_2$ is stable here . Note, that these stability are calculated close to the bifurcation curves, and stability and type of equilibria might change with further increasing parameters as we see below.

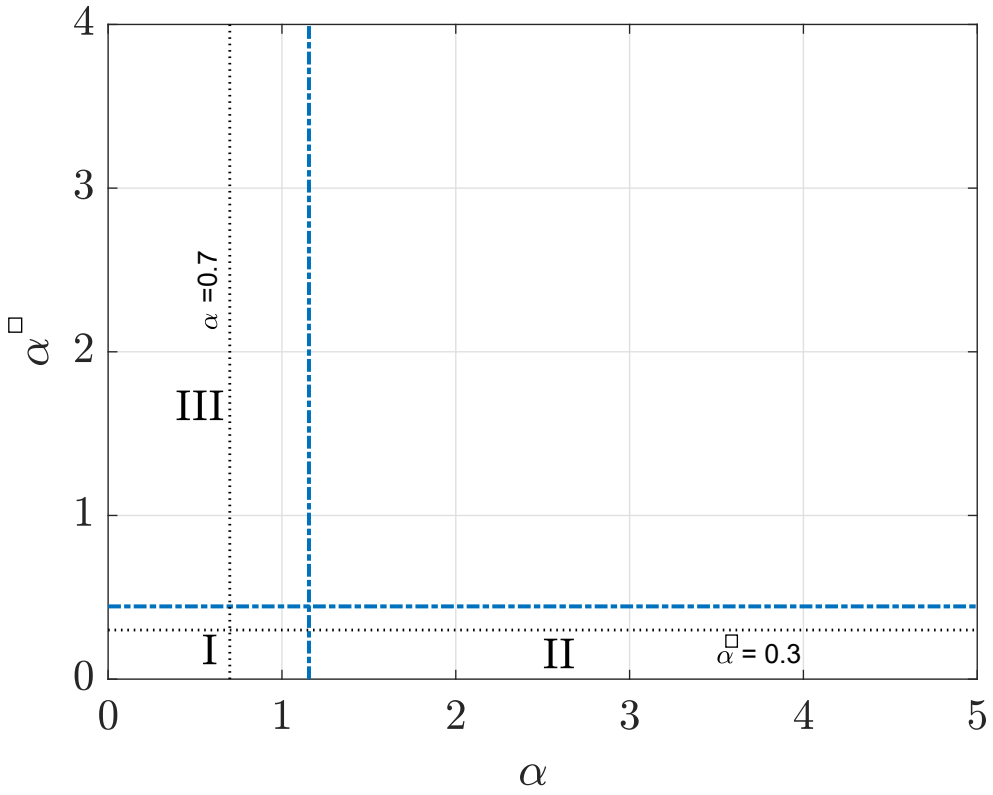


Figure 4.1: Two-dimensional bifurcation diagram with analytically obtained transcritical bifurcation (dashed) lines. The feasible region of parameter space is divided into four subregions by the transcritical bifurcation. Thin dotted lines indicate the chosen values of parameters for the one-dimensional bifurcation diagrams displayed below in Figure 4.2.

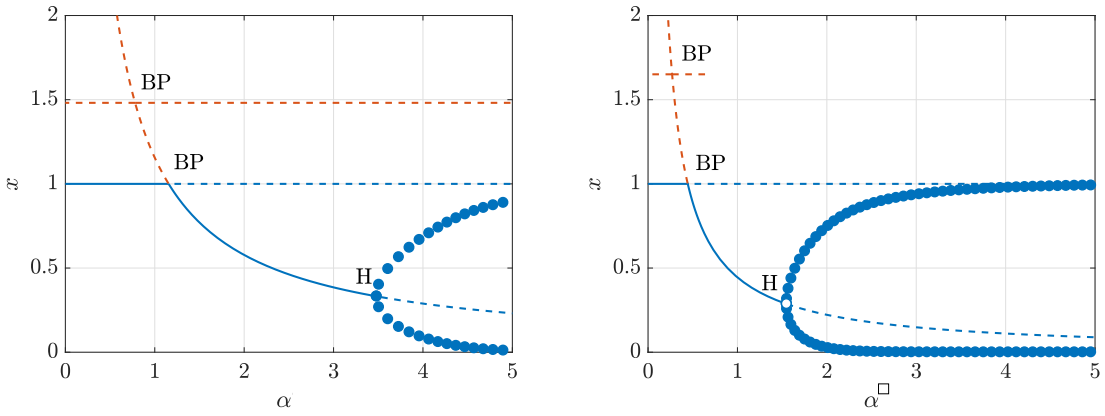
To examine the stability of wt-free state $\boldsymbol{v}_2$, we obtain the eigenvalues from the characteristic polynomial computed from $\det(J(\boldsymbol{v}_2) - \lambda \boldsymbol{I}) = 0$. This equation can be factorised and rewritten as $Q_2(\lambda) \cdot Q_3(\lambda) = 0$, where Q_2 and Q_3 are quadratic and cubic polynomials in λ, respectively. The exact form of Q_2 is not required here. We write $Q_3(\lambda) = a\lambda^3 + b\lambda^2 + c\lambda + d = 0$, where

$$\begin{aligned} a &= \alpha^{\square}(1-\nu^{\square})^2, \\ b &= (1-\nu^{\square})\left(\alpha^{\square}\gamma^{\square}(1-\nu^{\square}) + \zeta^{\square}(\alpha^{\square}+1)\right), \\ c &= \zeta^{\square}\left((\nu^{\square})^2\alpha^{\square} + (\zeta^{\square} - \alpha^{\square} - \gamma^{\square})\nu^{\square} + \zeta^{\square} + \gamma^{\square}\right), \\ d &= (1-\nu^{\square})\zeta^{\square}\gamma^{\square}(\alpha^{\square}(1-\nu^{\square}) - \zeta^{\square}). \end{aligned} \tag{4.11}$$

Following the analysis scheme discussed earlier, to obtain conditions for Hopf bifurcations of $\boldsymbol{v}_2$, we search for a purely imaginary pair of eigenvalues. Setting $Q_3(\lambda = \pm i\omega) = 0$, we obtain the critical condition for a Hopf bifurcation, $ad = bc$, with $\omega = \sqrt{d/b}$ being the angular frequency of the emerging periodic orbits (POs). Interestingly, for this Hopf bifurcation, there is no dependence on parameters associated with the wt virus strain. For all biologically meaningful solutions of $ad = bc$, the mutant viral strain has the potential to gain periodic behaviour through the creation of a stable PO. The other factor, $Q_2(\lambda)$, provides no possibility for a Hopf bifurcation due to a strictly positive linear coefficient in the quadratic polynomial.

In order to determine the behaviour of all the equilibria, let us consider cross sections of Figure 4.1: one-dimensional bifurcation diagrams. The indicators (thin, dotted lines) on Figure 4.1 show chosen cross sections of the bifurcation diagram at fixed value of parameters. The difference between the four states can be easily illustrated by considering only the first component of solutions, x, see Figure 4.2. This justifies the use of the uninfected cell population for all one-dimensional bifurcations of this Chapter, if not specified otherwise. That being said, there is an equilibrium which is not presented in these bifurcation diagrams. The trivial equilibrium is unstable for all values of parameters.

In Figure 4.2 (a, b) one-dimensional bifurcation diagrams are plotted for fixed values of $\alpha^{\square} = 0.3 < \alpha^{\square}_{\mathrm{tr}}$ and $\alpha = 0.7 < \alpha_{\mathrm{tr}}$. For small values of these parameters there is always a stable equilibrium, and for both of the one-dimensional bifurcation diagrams, that stable equilibrium is $\boldsymbol{v}_1$. This matches with the results of linear stability analysis and shows that with parameters $\alpha^{\square} \longrightarrow 0^+, \alpha \longrightarrow 0^+$, the stability of $\boldsymbol{v}_1$ stays the same in the whole region I. Although there is a transcritical bifurcation between $\boldsymbol{v}_2$ and $\boldsymbol{v}_3$ in region I, this bifurcation happens to biologically meaningless equilibria and, moreover, does not



(a) $\alpha^{\square} = 0.3$ (b) $\alpha = 0.7$

Figure 4.2: One-dimensional bifurcation diagrams as cross sections of Figure 4.1 with specified transcritical bifurcation points ('BP') and Hopf bifurcation points ('H'). These diagrams show the change in a population of uninfected cells (first components of equilibria). Dashed lines indicate unstable equilibria, while solid lines correspond to stable equilibria. An equilibrium or a periodic orbit is plotted in blue when all the components of it have biologically meaningful (positive) values, otherwise it is plotted in orange. Round markers show maximum and minimum of periodic orbits.

change the stability of $\boldsymbol{v}_2$ and $\boldsymbol{v}_3$. With growing values of parameters there are standard transcritical bifurcations (marked as BP), which are determined analytically in (4.9). As mentioned above, at increasing α, the stability of $\boldsymbol{v}_1$ shifts to $\boldsymbol{v}_3$. The same dynamics can be observed in the bifurcation diagram increasing $\alpha^{\square}$. In this case, stable $\boldsymbol{v}_1$ meets at the transcritical point the unstable equilibrium $\boldsymbol{v}_2$, and both change their stability. Increasing the values of parameters even further along the stable $\boldsymbol{v}_3$ and $\boldsymbol{v}_2$, we find supercritical Hopf bifurcations. These Hopf bifurcations provide stable periodic orbits around the equilibria for all the values of parameters after the bifurcations and inevitably affect the two-dimensional bifurcation diagram. The bifurcations found in these one-dimensional diagrams can be continued in the two-dimensional bifurcation diagram, which more qualitatively different regions to the parameter space.

Figure 4.3 is an improved version of the previous two-dimensional bifurcation diagram. Additionally to the transcritical bifurcations (dashed lines) from Figure 4.1, new-found continuations of Hopf bifurcations (solid curves) are plotted. Part of the horizontal Hopf bifurcation line divides region III into two regions: $IIIa$, for which $\boldsymbol{v}_2$ changes from a stable node to a stable spiral at increasing $\alpha^{\square}$, and $IIIb$, for which there are stable periodic orbits around $\boldsymbol{v}_2$. The shorter Hopf bifurcation curve separates regions II and IV into two subregions each. Region IIa contains parameters for which equilibrium $\boldsymbol{v}_3$ is stable, and with increasing α turns from a stable node into a stable spiral. As soon as the Hopf

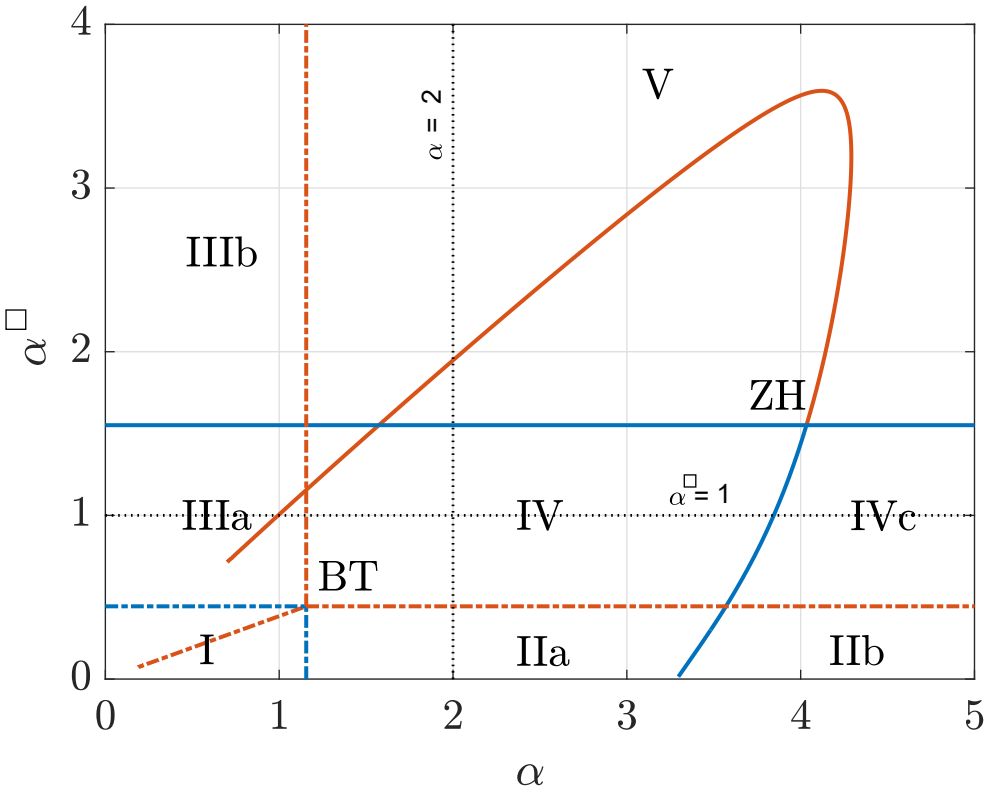
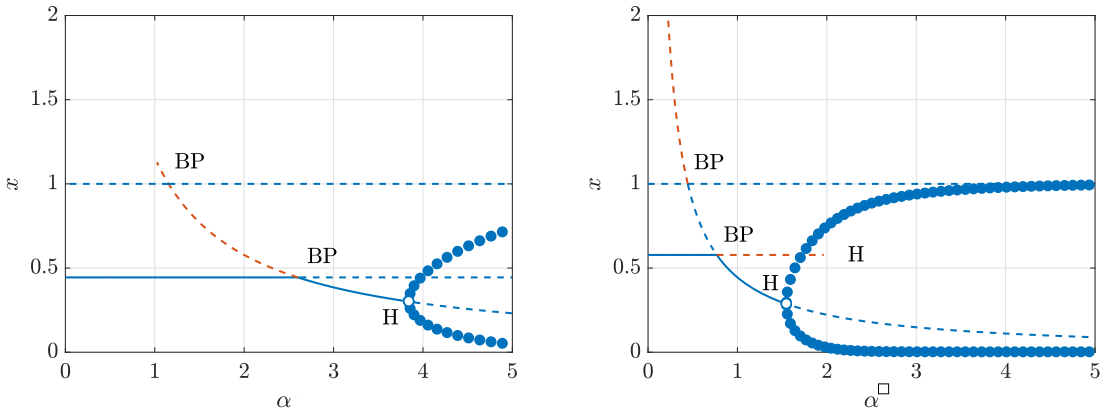


Figure 4.3: Two-dimensional bifurcation diagram with transcritical (dashed) and Hopf (solid) bifurcations. The parameter space is divided into eight sub-regions. The thin dotted lines indicate the values of the parameters chosen to build the bifurcation diagram of Figure 4.4. Blue and orange represent curves of bifurcations that occur inside and outside of biologically meaningful phase space.



(a) $\alpha^{\square} = 1$ (b) $\alpha = 2$

Figure 4.4: One-dimensional bifurcation diagrams where transcritical bifurcation points ('BP') and Hopf bifurcation points ('H') are marked. Blue lines and blue markers stand for biologically meaningful solutions, otherwise, equilibria and periodic orbits (POs) are plotted in orange. Round markers show maximum and minimum of POs. Solid lines and filled markers stand for stable equilibrium and stable POs, respectively.

curve is crossed into region IIb, $\boldsymbol{v}_3$ develops, stable periodic orbits around itself. At this stage of the analysis, we cannot be certain about the behaviour of the system in other regions without running simulations. To illustrate the non-trivial dynamics that can occur consider, for example, region IV. Remember that in $IIIa$ the equilibrium $\boldsymbol{v}_2$ is stable, and that $IIIa$ and IV are separated by a transcritical bifurcation. We also know that in IIa the equilibrium $\boldsymbol{v}_3$ is stable, and IIa and IV are also separated by a transcritical bifurcation. This knowledge would seem enough to guess that IV should accommodate parameters for which $\boldsymbol{v}_1$ is stable again. However, this scenario is far from being the case (see below).

Let us consider one-dimensional bifurcation diagrams for parameter values of new regions. In Figure 4.4 (a), we see a bifurcation diagram for a fixed value of $\alpha^{\square}_{\text{tr}} < \alpha^{\square} = 1.0 < \alpha^{\square}_{\text{H}}$, and varying α we cross regions $IIIa$, IV, and IVc. This bifurcation diagram shows that for smaller values of α (read 'region $IIIa$'), the equilibrium $\boldsymbol{v}_2$ is stable. To describe the bifurcations of the plot in a sequenced manner, we have to recall that equilibrium $\boldsymbol{v}_1$ is unstable for region $IIIa$. At the analytically found transcritical bifurcation, (as before) $\boldsymbol{v}_1$ intersects newly biologically meaningless $\boldsymbol{v}_3$, and this does not affect the overall dynamics of the system. Going back to the stable $\boldsymbol{v}_2$, we see that it later undergoes the transcritical bifurcation with $\boldsymbol{v}_3$. This transition is happening inside region IV; following the direction of growth of α, the stable $\boldsymbol{v}_3$ becomes unstable in region IVc while developing stable periodic orbits around itself, and $\boldsymbol{v}_3$ has the same Hopf bifurcation it develops transitioning from IIa to IIb (see Figure 4.2 (a)). The transcritical bifurcation between $\boldsymbol{v}_2$ and $\boldsymbol{v}_3$ must occur after α_{tr} and before α_{H}, *i.e.* in region IV.

In Figure 4.4 (b), a bifurcation diagram for a fixed value of $\alpha_{\text{tr}} < \alpha = 2.0 < \alpha_{\text{H}}$ is shown. Similar to Figure 4.4 (a), but varying $\alpha^{\square}$ throughout the regions IIa, IV, and V, we study the stability and behaviour of the equilibria. The unstable equilibrium $\boldsymbol{v}_1$ at first undergoes a transcritical bifurcation with $\boldsymbol{v}_2$, which is unstable and not biologically meaningful. This point matches with an orange – biologically meaningless – transcritical bifurcation between regions IIa and IV in Figure 4.3, and does not affect the stability of the equilibria due to high dimensionality of the equilibrium points. Although this bifurcation does not change the stability of either of the equilibria, the transition brings $\boldsymbol{v}_2$ into the biologically meaningful variable space. After that, the same unstable $\boldsymbol{v}_2$ has a transcritical bifurcation with $\boldsymbol{v}_3$, which has been stable throughout IIa and IV. At the latter bifurcation, $\boldsymbol{v}_3$ becomes biologically meaningless and unstable. Meanwhile, $\boldsymbol{v}_2$ turns stable (first – node, then – spiral), and eventually becomes unstable after its Hopf bifurcation and gains a stable periodic orbit in region V. Note that the second Hopf bifurcation marked along negative $\boldsymbol{v}_3$ on Figure 4.4 (b) has biologically meaningless orbits which were dropped while plotting. Mathematically speaking, periodic orbits are there, and parameter values for the bifurcations can be found along the orange curve of Hopf bifurcation in Figure 4.3.

With these findings we improve the two-dimensional bifurcation diagram, see Figure 4.5. There is indeed a transcritical bifurcation between equilibria $\boldsymbol{v}_2$ and $\boldsymbol{v}_3$ separating the region IV into IVa and IVb. The behaviour of the system at parameter values of region IV can be summed up in the following way: Due to the biologically meaninglessness of the transcritical bifurcation between the regions $IIIa$ and IVa, parameters from both of those regions affect the dynamics of the system in the same manner, *i.e.* the equilibrium $\boldsymbol{v}_2$ is stable in $IIIa$ and IVa. For similar reasons, the equilibrium $\boldsymbol{v}_3$ is stable in regions IIa and IVb.

One end of the continuation of the transcritical bifurcation between $\boldsymbol{v}_2$ and $\boldsymbol{v}_3$ meets two other transcritical bifurcations at a Bogdanov-Taken (BT) point and prolongs into region I. Region I does not gain any new features because neither $\boldsymbol{v}_2$ nor $\boldsymbol{v}_3$ are biologically meaningful here (orange, dashed line), that is region I stays a region where equilibrium $\boldsymbol{v}_1$ is stable. The other end of the continuation of the transcritical bifurcation between $\boldsymbol{v}_2$ and $\boldsymbol{v}_3$ at larger values of parameters meets two Hopf bifurcations at the Zero-Hopf (ZH) point and prolongs even further into region V. At this stage of discussion, we try to gain insights of the dynamics in region V by considering cross sections of the bifurcation diagram as was done before. We will examine one-dimensional bifurcation diagrams for parameters crossing regions $IIIa$, V, Va, and regions IIb, IVc, Va, and V.

By increasing α for the fixed value of $\alpha^{\square} = 2.32$, we obtain one-dimensional bifurcation diagram of Figure 4.6 (a). Neither of the equilibria is stable throughout the plot. However, from Figure 4.2 (b), we know that for large values of $\alpha^{\square}$, the unstable equilibrium $\boldsymbol{v}_2$ has

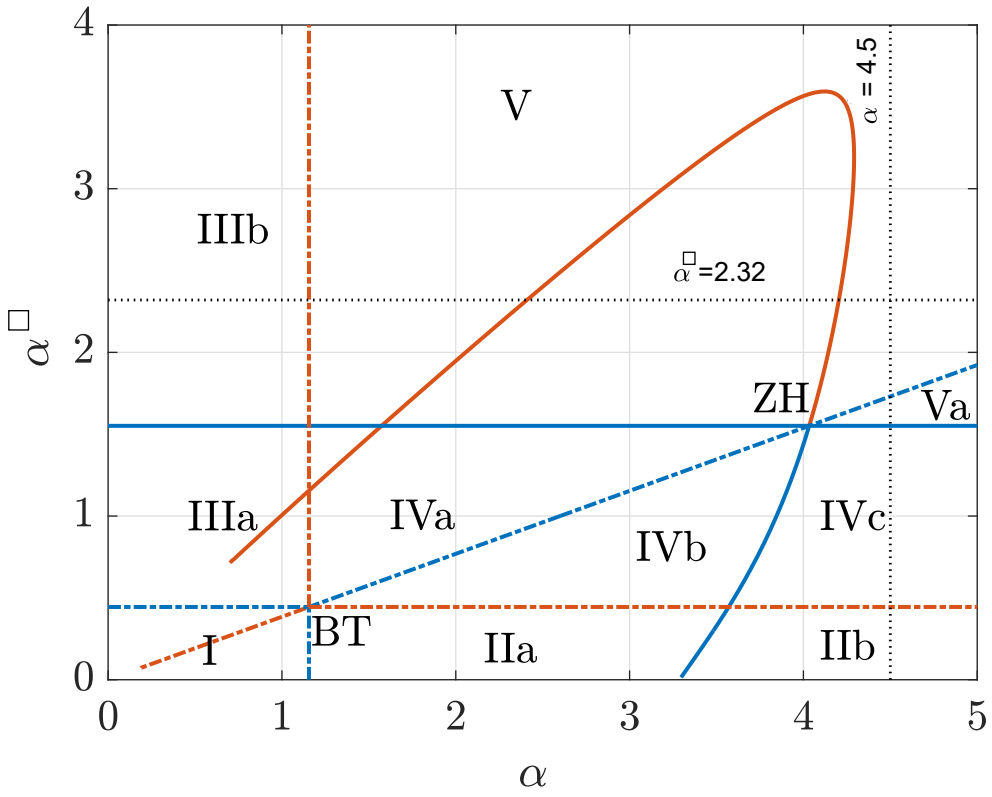
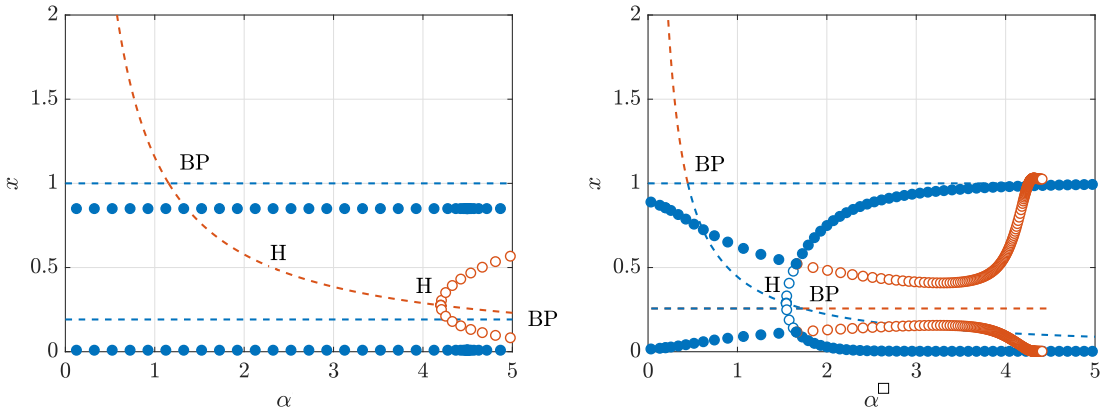


Figure 4.5: Two-dimensional bifurcation diagram with transcritical bifurcations (dashed) and Hopf bifurcations (solid). The parameter space is divided into at least ten sub-regions. Thin dotted lines indicate the chosen values of parameters for the following one-dimensional bifurcation diagrams. Blue and orange represent curves of bifurcations that occur inside and outside of biologically meaningful phase space.



(a) $\alpha^{\square} = 2.32$ (b) $\alpha = 4.5$

Figure 4.6: One-dimensional bifurcation diagrams with at least two periodic orbits in each. Neither of the equilibria is stable (dashed curves and lines), however, there are always stable and biologically meaningful periodic orbits throughout the plots (filled and blue markers).

periodic orbits around it. Therefore, the unstable $\boldsymbol{v}_2$ has the stable (filled) and biologically meaningful (blue) periodic orbit markers around it. Increasing α, we notice a transcritical bifurcation between $\boldsymbol{v}_1$ and $\boldsymbol{v}_3$. Further, there are two consecutive biologically meaningless Hopf bifurcations along the homogeneously unstable $\boldsymbol{v}_3$. These Hopf bifurcations occur exactly at parameter values of α in the biologically meaningless (orange) curve of Hopf bifurcation for the fixed $\alpha^{\square}$ in Figure 4.5. Increasing α after the last Hopf point, we find a transcritical bifurcation of $\boldsymbol{v}_2$ and $\boldsymbol{v}_3$. This transcritical bifurcation does not change the stability, but after it, both $\boldsymbol{v}_2$ and $\boldsymbol{v}_3$ are in the biologically meaningful variable space. Following the growing α even further, we finally find a transcritical bifurcation of periodic orbits (not shown: occurs for much greater values of α). The stable periodic orbit around $\boldsymbol{v}_2$ meets the unstable periodic orbit created at the last Hopf bifurcation of biologically meaningless $\boldsymbol{v}_3$. The transcritical bifurcation of periodic orbits change stability of intersecting periodic orbits. For all values of α after the critical value, there are only stable and biologically meaningful periodic orbits around $\boldsymbol{v}_3$.

In Figure 4.6 (b), we consider a vertical cross section of the two-dimensional bifurcation diagram (see Figure 4.5) for a large value of α fixed. Gradually increasing $\alpha^{\square}$, we can observe a transcritical bifurcation between unstable equilibria $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ that does not lead to an exchange of stability. Meanwhile, *i.e.* for small values of $\alpha^{\square}$, $\boldsymbol{v}_3$ has stable periodic orbits. At the critical value of $\alpha_H^{\square}$, the unstable equilibrium $\boldsymbol{v}_2$ gains unstable yet biologically meaningful periodic orbits around itself. Short after that, the stable periodic orbits of $\boldsymbol{v}_3$ intersect unstable periodic orbits around $\boldsymbol{v}_2$ and the periodic orbits exchange stability. Moreover, after the transcritical bifurcation of periodic orbits, some components of unstable periodic orbits of $\boldsymbol{v}_3$ become negative, making it biologically meaningless. By increasing the value of $\alpha^{\square}$ a little more, we see an intersection of $\boldsymbol{v}_2$ and $\boldsymbol{v}_3$. This transcritical bifurcation puts unstable $\boldsymbol{v}_3$ (and its recently unstable periodic orbits) out of the biologically meaningful variable space. This leads us back to the previously noted matter, that for very large values of $\alpha^{\square}$ (read region V), there are stable and biologically relevant periodic orbits around $\boldsymbol{v}_2$.

By analysing the last two bifurcation diagrams, we found intersections of periodic orbits. From the available results, we can deduce the existence of a curve of transcritical bifurcations of periodic orbits between the horizontal Hopf bifurcation line (of equilibrium $\boldsymbol{v}_2$) and transcritical bifurcation of equilibria $\boldsymbol{v}_2$ and $\boldsymbol{v}_3$. By numerically continuing the intersection of periodic orbits on the two-dimensional bifurcation diagram, we see in Figure 4.7 that the curve of transcritical bifurcations of periodic orbits, which dividing the region into Va and Vb, collide with the degenerate Zero-Hopf point. So far the α and $\alpha^{\square}$ parameter space is divided into following regions:

- Region I: $\boldsymbol{v}_1$ is a stable node;

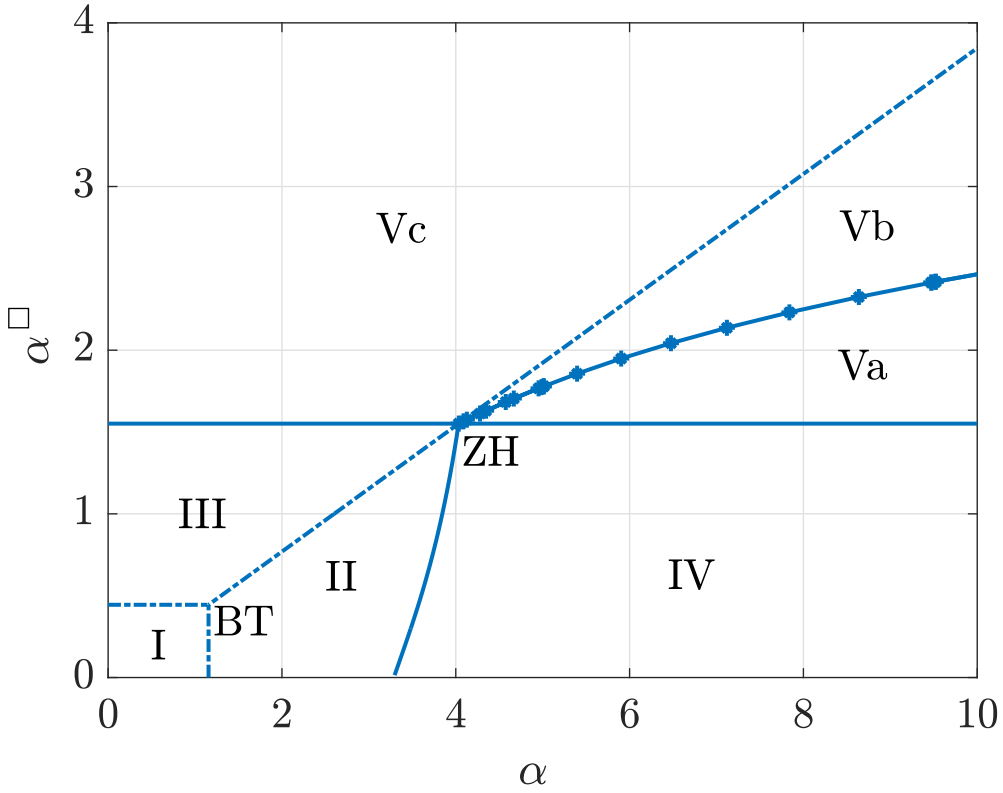
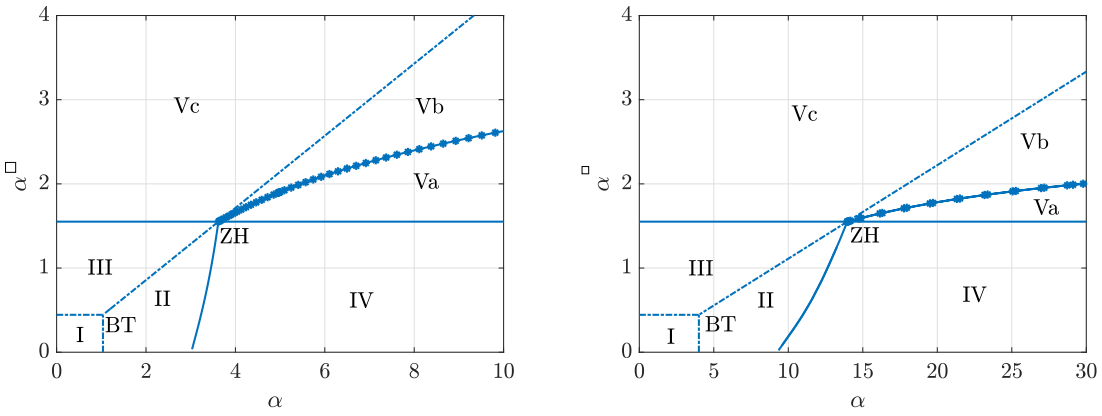


Figure 4.7: Two-dimensional bifurcation diagram with biologically differing regions. Solid lines stand for supercritical Hopf bifurcations, dashed lines are transcritical bifurcations of equilibria, and a star-marked curve is the transcritical bifurcations of periodic orbits. Region I has $\boldsymbol{v}_1$ as a stable node; region II has stable $\boldsymbol{v}_3$; region III has stable $\boldsymbol{v}_2$; regions IV and Va have stable periodic orbits around $\boldsymbol{v}_3$; regions Vb and Vc have stable periodic orbits around $\boldsymbol{v}_2$.

- Region II: $\boldsymbol{v}_3$ is a stable node or stable spiral;
- Region III: $\boldsymbol{v}_2$ is a stable node or stable spiral;
- Region IV: $\boldsymbol{v}_3$ has stable periodic orbits;
- Region Va: $\boldsymbol{v}_3$ has stable periodic orbits, $\boldsymbol{v}_2$ has unstable periodic orbits;
- Region Vb: $\boldsymbol{v}_2$ has stable periodic orbits, $\boldsymbol{v}_3$ has unstable periodic orbits;
- Region Vc: $\boldsymbol{v}_2$ has stable periodic orbits.

The wt strain of virus throughout this section had parameter values that were 10% less than those of the mutant strain of virus$^\square$. In order to see how this affected the overall dynamics of the system, we compare the bifurcation diagram on Figure 4.7 with a bifurcation diagram (see Figure 4.8 (a)) for the case when all the parameters are equal. Two parameters are



(a) $\mu = 0.1$ (b) $\mu = 0.2$

Figure 4.8: Two-dimensional bifurcation diagram for system (4.2) with biologically differing regions for equal parameters. Solid lines indicate supercritical Hopf bifurcations, dashed lines are transcritical bifurcations of equilibria, and a star-marked curve is the transcritical bifurcations of periodic orbits. Region I has $\boldsymbol{v}_1$ as a stable node; region II has stable $\boldsymbol{v}_3$; region III has stable $\boldsymbol{v}_2$; regions IV and Va have stable periodic orbits around $\boldsymbol{v}_3$; regions Vb and Vc have stable periodic orbits around $\boldsymbol{v}_2$.

worth mentioning separately: κ is a ratio of the dimensional κ of the wild-type virus to $\kappa^{\square}$ and in case of equal parameter values the dimensionless κ should be 1. The constant μ is a single uncoupled parameter. In fact, two strains of virus can be differentiated only by a flow of the population from one strain ("generalist") to another strain ("specialist") at rate μ. Almost no qualitative effect of different values of coupled parameters can be seen from comparison of Figures 4.7 and 4.8 (a). A slight change in the curvature of the Hopf bifurcation curve and an increase in the slope of transcritical bifurcation of equilibria are worth noting, although are not crucial for the discussion. The next intuitive question would lie in studying the effect of mutation parameter μ. The slight change in μ leads to quantitative differences in the bifurcation diagram (see Figure 4.8 (b)), but qualitatively the bifurcation diagrams remain the same.

4.1.2 Virulence-based study

We obtained a two-dimensional bifurcation diagram illustrating a transcritical bifurcation of periodic orbits. This study was based on our interest in the role of the infection rate in the population dynamics. We now study the effect of the strains' virulence (parametrised by γ and $\gamma^{\square}$) on the dynamics of the system by building bifurcation diagrams. The infection rates are chosen to be from the different regions of the two-dimensional bifurcation diagram shown in Figure 4.8 (a), which are based on the reference values $\gamma^{\square} = \gamma = 0.25$. The other

parameters are fixed to

$$\mu = 0.1, \quad \kappa = 1, \quad \nu^{\square} = \nu = 0.5, \quad \zeta^{\square} = \zeta = 0.2222. \tag{4.12}$$

We focus on Regions I, II, and V of Figure 4.8 (a) as a starting point, covering all the other regions from there.

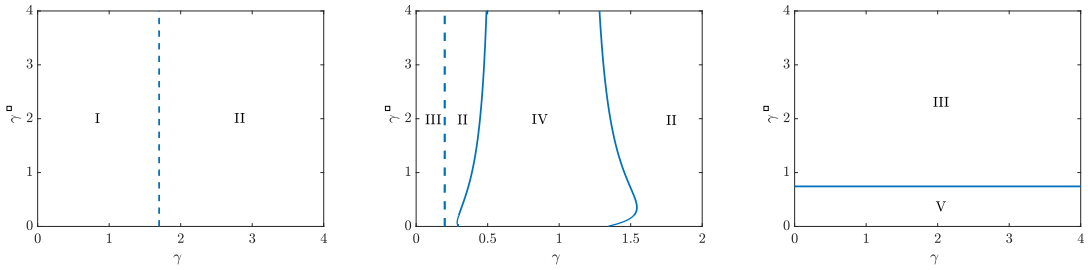
We begin with the case where the infection rates are chosen to coincide with Region I at the virulence reference values. These infection rates therefore correspond to basic reproduction numbers that are less than one. From the definition of $R_0^{\square}$, we see that the basic reproduction number of the mutant is independent of the virulence. Thus, changes in $\gamma^{\square}$ or γ cannot increase $R_0^{\square}$ beyond one. Therefore, Regions III and V, where the mutant-type virus persists at the expense of the wt virus becoming extinct, cannot be entered. The basic reproduction number for the wt virus is an increasing function of the virulence of the wt virus. In the limit of very large virulence, $\gamma \to \infty$, we find that $R_0 \to \alpha(\kappa - \nu)/\zeta$. Thus, if the infection rate α is so small that the limit of R_0 is less than one, then it will not be possible to leave Region I and both strains of the virus will always become extinct. However, if α is sufficiently large and the basic reproduction number increases beyond one, then a change in dynamics will be observed as γ is increased. By solving $R_0 = 1$, a critical value of the virulence of the wt virus is obtained:

$$\gamma_{\text{crit}} = \frac{\mu}{\kappa\left(\zeta\alpha^{-1} + \nu\right)^{-1} - 1}. \tag{4.13}$$

Equation (4.13) is the condition for a transcritical bifurcation between the virus-free state $\boldsymbol{v}_1$ and the coexistence state $\boldsymbol{v}_3$ and defines the boundary between Regions I and II. Thus, only for values of $\gamma > \gamma_{\text{crit}}$ will the virus persist in the system for this choice of infection rates.

The stability diagram for $\alpha = 0.5$ and $\alpha^{\square} = 0.1$ has been numerically computed and is shown in Figure 4.9 (a). These values of the infection rate correspond to Region I at the reference values of the virulence; see Figure 4.8 (a). As predicted, Regions III and V are absent from the stability diagram. However, Region IV is also missing. Thus, for this choice of infection rates, only Regions I and II can be entered by changing the values of the virulence. Regions I and II are separated by a straight dash line given by the critical condition (4.13).

We now consider infection rates given by $\alpha = 3$, $\alpha^{\square} = 1$, which correspond to Region II at the reference virulence; see Figure 4.8 (a). The resulting stability diagram in terms of the virulence is shown in Figure 4.9 (b). The choice of $\alpha^{\square} = 1$ along with the values



(a) $\alpha = 0.5$, $\alpha^{\square} = 0.1$ (b) $\alpha = 3$, $\alpha^{\square} = 1$ (c) $\alpha = 0.5$, $\alpha^{\square} = 2$

Figure 4.9: Two-dimensional bifurcation diagrams for γ versus $\gamma^{\square}$ at different sets of infection rates. Parameter values are given by (4.12) and (a) $\alpha = 0.5$, $\alpha^{\square} = 0.1$; (b) $\alpha = 3$, $\alpha^{\square} = 1$; and (c) $\alpha = 0.5$, $\alpha^{\square} = 2$. Dashed and solid lines correspond to transcritical and Hopf bifurcations, respectively. Region I: stability of trivial solution; Region III: stability of wt-free state; Region II: stability of coexistence state; Region V: stable periodic orbit (PO) around the wt-free state; Region IV: stable PO around the coexistence state.

of the parameters in (4.12) leads to $R_0^{\square} = 2.25$. Thus, variations in the virulence cannot bring the system to Region I and at least one type of virus will always persist. A transition between Regions III and II can occur via the transcritical bifurcation between the wt-free and coexistence states $\boldsymbol{v}_2$ and $\boldsymbol{v}_3$. The critical condition for this transcritical bifurcation depends only on the virulence of the wt viral strain. Thus, the transition between Regions III and II appears as the vertical dashed line in Figure 4.9 (b). Interestingly, this figure shows that as the virulence of the wt virus is increased, the system transitions from Region II to Region IV and then back to Region II. These transitions occur via supercritical Hopf bifurcations, denoted by solid lines. This scenario corresponds to the so-called bubble bifurcation (found in other epidemiological systems see *e.g.* [74]), in which the system is in a stationary state, then enters into an oscillating one, and finally goes back to a stationary state as the control parameter is changed.

To understand why Region V does not appear in the stability diagrams on Figure 4.9 (a) and (b), we first recall that Region V is separated from Region III by a Hopf bifurcation curve of the wt-free state $\boldsymbol{v}_2$. This Hopf curve has been calculated analytically from the equality $ad = bc$, where a, b, c, and d are given in (4.11), and is a quadratic expression for $\gamma^{\square}$ independent of γ and α. Solving $ad = bc$ to find $\gamma^{\square}_{\text{crit}}$ (not shown due to complexity of the expression) at the values of infection rates chosen for Figures 4.9 (a) and (b), we notice they are outside of the positive parameters space. In fact, only for values of the mutant-type infection rate that satisfy

$$\alpha^{\square} > \frac{\zeta^{\square}}{1-\nu^{\square}}\left(1+\frac{1}{\nu^{\square}}\right), \tag{4.14}$$

equivalent to $\alpha^{\square} > 1.3332$ for the chosen parameters, does the Hopf curve of $\boldsymbol{v}_2$ appear on

a γ versus $\gamma^{\square}$ stability diagram; see Figure 4.9 (c). Greater values of $\alpha^{\square}$ lead to increases in the area of Region V under Region III.

The stability diagrams of Figure 4.9 indicate the impact of virulence on the complexity of the dynamics. Clearly, the virulence of the mutant strain does not affect the survival of the wt strain. That is, increases in $\gamma^{\square}$ do not lead to an appearance or disappearance of the wt strain. However, the survival of the mutant strain does not appear to depend on $\gamma^{\square}$ either. Importantly, the virulence of the wt strain, γ, strongly controls the dynamics and the persistence of the wt strain. As can be seen for small values of γ, this strain can become extinct as shown in Figure 4.9 (b). In other words, even for a superior infection rate of the wt viral strain, there is a threshold of γ that must be surpassed in order for this strain to exist. For inferior values that are below this threshold, the survival of the wt strain is impossible because the rate of wt virion production, *i.e.* the death of rate infected cells, is insufficient. The qualitative behaviour of solutions, *i.e.* whether the populations of wt strains undergo oscillations or stabilise to a certain value, depends on the virulence in a nontrivial way. As shown in Figure 4.9 (b), there is an interval of values for γ for which the system has a stable PO governing the coexistence of strains, marked as Region IV. Outside of this interval, the populations reach stable stationary states.

4.1.3 The effect of heterogeneity in mutation rates

Most of the previous analyses were conducted for the mutation rate fixed at $\mu = 0.1$. We now consider the effect of μ on the dynamics. This is a key parameter that has been largely investigated within the framework of the error threshold [35, 20] and lethal mutagenesis [19] in quasispecies theory. In general, the coexistence of two strains requires the basic reproduction number of the wt virus to be greater than one, $R_0 > 1$, which can be interpreted as a condition on the mutation rate:

$$\mu < \gamma \left(\frac{\kappa}{\zeta \alpha^{-1} + \nu} - 1 \right). \tag{4.15}$$

Hence, for two strains to coexist, the mutation rate must be sufficiently small. This result supports a conjecture that excess mutation exhausts the population of the wt strain, thereby leading to a process similar to the well-known error catastrophe [20]. We would expect that a gradual increase of the mutation rate contributes to a better success of the mutant strain as the frequency of mutants generated *de novo* increases. The right-hand side of (4.15) is an increasing function of α, reflecting the fact that a greater rate of infection by the wt virus will offset a greater mutation rate. A stability diagram in terms of the mutation rate μ and the wt infection rate α is shown in Figure 4.10 (a) for the case of $\alpha^{\square} = 0.1$. The dashed line

marking the boundary between Regions I and II, and also defining the region of coexisting populations, has been obtained by replacing the inequality with equality in (4.15). As the infection rate α increases, the boundary between Regions I and II reaches a horizontal asymptote given by

$$\mu_c = \gamma \left(\frac{\kappa}{\nu} - 1 \right). \tag{4.16}$$

For mutation rates that satisfy $\mu < \mu_c$, Regions II and IV can be entered from Region I by increases in the infection rate, promoting coexistence. However, for $\mu > \mu_c$, only the wt-free state can occur. Hence, Eq. (4.16) defines a critical, finite mutation rate at which coexistence no longer becomes possible due to extinction of the wt virus due to the outcompetition by the mutant strains. To explore this in more detail, we have repeated the stability diagram shown in Figure 4.8 (a) using a value of $\mu = 0.5 > \mu_c$. The stability diagram changes drastically, becoming that shown in Figure 4.10 (b), and contains only three regions: I, III, and V. The stability diagrams in Figures 4.8 (a) and 4.10 (b) are linked through the fact that as the mutation rate increases, the BT bifurcation shifts to the right, eventually tending to infinity as the critical value is approached.

Finally, Figures 4.10 (c) and (d) illustrate how the equilibrium populations of infected cells and viral strains change with increasing mutation rate μ. Specifically, we have used $\alpha = 3$ and $\alpha^{\square} = 1$, in both panels. As μ increases beyond the point where (4.15) is satisfied, the population of mutants (virions and infected cells) outcompetes the wt populations, see Figure 4.10 (c) and (d). This phenomenon is similar to the error threshold defined in quasispecies theory. Here we show that mutation is not only involved in this shift, but also depends (at the infection-cell level) on virulence, burst size, and multiplicity of infection.

4.2 Modelling backward mutation

Ideally, we would like to have a model that is simple and flexible enough to be able to study and observe as many variations of the dynamics as possible. These kind of models are usually impossible to study analytically, which would provide a deeper understanding effects of the underlying processes. However, throughout the development of the model by keeping continuance between two consecutive schemes, we can always rely on the patterns in results of a simpler model as a hint. There are many ways to model mutation. This section approaches the task by considering the mutation as a partial, possibly implicit, transfer from one population of viral strains to another. The directed flow, as in models (4.1) and (4.2), can be designed as simple as an exponential addend with the same rate in two equations, one – positive – and another – negative. This implies strict limits on the behaviour of every

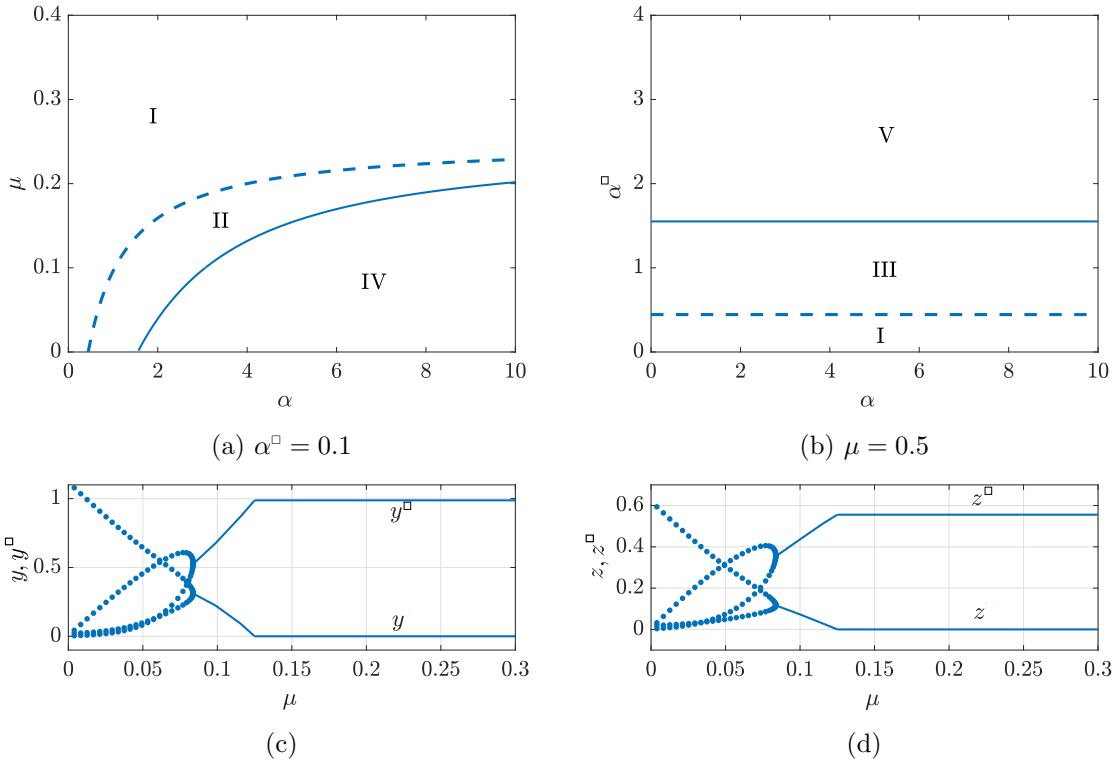


Figure 4.10: (a,b) Two-dimensional stability diagrams with a pair of bifurcation curves: transcritical bifurcations (dashed) and supercritical Hopf bifurcations (solid). We use $\alpha^{\square} = 0.1$ and consider stability diagram over parameters α versus μ. (b) We set $\mu = 0.5 > \mu_c$ and rebuild the α versus $\alpha^{\square}$ stability diagram. As before, Region I: stability of virus-free solution; Region III: stability of wt-free state; Region II: stability of coexistence state; Region V: stable periodic orbit (PO) about wt-free state; Region IV: stable PO about coexistence state. In the lower panels we display one-dimensional bifurcation diagrams for (c) infected cell and (d) viral strain populations versus μ; note, that only stable equilibria (solid curves) and stable POs (filled markers) are shown. We set $\alpha = 3.0$ and $\alpha^{\square} = 1.0$ and track the outcompetition of the wt strains by the mutant ones at increasing μ, resulting in a type of error-threshold found after the oscillatory and static coexistence scenarios for μ satisfying (4.15).

strain, and although this approach has all the rights to exist, we try to relax the artificial differentiation of wt and mutant strains by giving them similar equations. By allowing for backwards mutation in viral population, modelled implicitly through infected cells, we obtain the following model:

$$\begin{aligned}
\dot{x} &= \beta\, x(1 - x/K) - \alpha^{\square}\, z^{\square}\, x - \alpha\, z\, x - \delta\, x,\\
\dot{y^{\square}} &= \alpha^{\square}\, z^{\square}\, x + \mu y - \mu y^{\square} - \gamma^{\square}\, y^{\square},\\
\dot{y} &= \alpha\, z\, x + \mu y^{\square} - \mu y - \gamma\, y,\\
\dot{z^{\square}} &= \kappa^{\square}\gamma^{\square}\, y^{\square} - \nu^{\square}\, \alpha^{\square}\, z^{\square}\, x - \zeta^{\square}\, z^{\square},\\
\dot{z} &= \kappa\gamma\, y - \nu\, \alpha\, z\, x - \zeta\, z,
\end{aligned} \tag{4.17}$$

Model (4.17), being similar to (4.1), describes the dynamics of one uninfected cell population, x, two strains of virus, $z^{\square}$ and z, and infected cell populations linked to their strains of virus, $y^{\square}$ and y. Nonetheless, there is a difference: the mutation is modelled as a primitive, two-points discretization of the diffusion operator. The Laplace operator is widely used to model heat transfer along a physical space. There is an intuitive way to imagine mutation as a transfer of population (heat) along not physical space, as in case of the heat transfer, but along *phenotype space*. If there are in total N strains of virus in a model, then mutation for each strain could be defined as $\Delta y_i \approx y_{i+1} - 2y_i + y_{i-1}$ for each $i = 2, \ldots, N-1$, $\Delta y_i \approx y_{i+1} - y_i$ for $i = 1$, and $\Delta y_i \approx y_i - y_{i-1}$ for $i = N$. More on this approach will be discussed in Section 6.1. Yet, while considering only two strains, mutations have form $\Delta y \sim y^{\square} - y$ and $\Delta y^{\square} \sim y - y^{\square}$, and multiplying them by the mutation rate we obtain addends for the right hand side of $\dot{y}$ and $\dot{y^{\square}}$. Here, distances between strains in denominators of discretization of Laplace operator are ignored for simplicity of discussion.

As it has been shown on previous models, correct nondimensionalization simplifies the equations without loss in quality of the results. The old variables in terms of new non-dimensional variables are

$$t = (\beta - \delta)^{-1}\bar{t}, \quad x = \left(1 - \frac{\delta}{\beta}\right) K\, \bar{x}, \quad y^i = \left(1 - \frac{\delta}{\beta}\right) K\, \bar{y}^i, \quad z^i = \left(1 - \frac{\delta}{\beta}\right) \kappa^{\square} K\, \bar{z}^i,$$

and new non-dimensional parameters replace corresponding old parameters as follows

$$\bar{\alpha}^i = \frac{\kappa^{\square}\, K \alpha^i}{\beta}, \quad \bar{\nu}^i = \frac{\nu^i}{\kappa^{\square}}, \quad \bar{\kappa} = \frac{\kappa}{\kappa^{\square}}, \quad \bar{\mu} = \frac{\mu}{\beta - \delta}, \quad \bar{p}^i = \frac{p^i}{\beta - \delta},$$

where p, as before, stands for parameters $\gamma^{\square}, \gamma, \zeta^{\square}$, and ζ; i over variables and parameters corresponds to a square ($^{\square}$) or its absence. These are the exact substitutions for obtaining non-dimensional parameters and variables as in (4.2); non-dimensional model also look very similar to (4.2):

$$\begin{aligned}
\dot{x} &= x(1-x) - \alpha^{\square}\, z^{\square}\, x - \alpha\, z\, x, \\
\dot{y}^{\square} &= \alpha^{\square}\, z^{\square}\, x + \mu y - \mu y^{\square} - \gamma^{\square}\, y^{\square}, \\
\dot{y} &= \alpha\, z\, x + \mu y^{\square} - \mu y - \gamma\, y, \\
\dot{z}^{\square} &= \gamma^{\square}\, y^{\square} - \nu^{\square}\, \alpha^{\square}\, z^{\square}\, x - \zeta^{\square}\, z^{\square}, \\
\dot{z} &= \kappa\gamma\, y - \nu\, \alpha\, z\, x - \zeta\, z.
\end{aligned} \tag{4.18}$$

The introduction of a slight change into the model, as in addition of a contrary flow of the population instead of one-way flow, affects tremendously, as we will see, the ability to study the system analytically. Although solutions of model (4.18) cannot be found analytically, with the knowledge from the previous model, we can check whether equilibria $\boldsymbol{v}_0$ and $\boldsymbol{v}_1$ of (4.2) fit as solution of (4.18). Simple substitution yields that model (4.18) indeed has at least two equilibria:

$$\begin{aligned}
\boldsymbol{v}_0 &:= \{x = 0, \quad y^{\square} = 0, \quad y = 0, \quad z^{\square} = 0, \quad z = 0\}; \\
\boldsymbol{v}_1 &:= \{x = 1, \quad y^{\square} = 0, \quad y = 0, \quad z^{\square} = 0, \quad z = 0\};
\end{aligned}$$

To study these two equilibria, consider the Jacobian matrix of (4.18) given by:

$$J = \begin{bmatrix}
1 - \alpha^{\square}\, z^{\square} - \alpha\, z - 2\, x & 0 & 0 & -\alpha\, x & -\alpha^{\square}\, x \\
\alpha^{\square}\, z^{\square} & \mu & -\gamma^{\square} - \mu & 0 & \alpha^{\square}\, x \\
\alpha\, z & -\gamma - \mu & \mu & \alpha\, x & 0 \\
-\alpha^{\square}\, \nu^{\square}\, z^{\square} & 0 & \gamma^{\square} & 0 & -\alpha^{\square}\, \nu^{\square}\, x - \zeta^{\square} \\
-\alpha\, \nu\, z & \kappa\, \gamma & 0 & -\alpha\, \nu\, x - \zeta & 0
\end{bmatrix}. \tag{4.19}$$

The eigenvalues of Jacobian matrix at the trivial equilibrium $\boldsymbol{v}_0$ can easily be obtained analytically, being:

$$\mathbf{\Lambda}_0 = \begin{bmatrix} 1 \\ \sqrt{\zeta^{\square}\,\zeta} \\ -\sqrt{\zeta^{\square}\,\zeta} \\ \mu + \sqrt{(\mu+\gamma)\,(\mu+\gamma^{\square})} \\ \mu - \sqrt{(\mu+\gamma)\,(\mu+\gamma^{\square})} \end{bmatrix}. \tag{4.20}$$

The positive components of $\mathbf{\Lambda}_0$ guarantee that the trivial equilibrium remains unstable for all values of the parameters. Unfortunately, the eigenvalues of the Jacobian at $\boldsymbol{v}_1$ are not so easy to obtain. Although obtaining eigenvalues from the polynomial $\det(J(\boldsymbol{v}_1) - \lambda I) = 0$ is not feasible, it is good to keep in mind that we are interested in eigenvalues of certain type. For example, there are a number of bifurcations that may occur if the eigenvalues are equal to zero. By forcing λ to be zero, we obtain certain combination of parameters from $\det(J) = 0$. By solving that combination for a parameter, we obtain a surface in parameter space which depicts a bifurcation curve:

$$\alpha^{\square} = \frac{K_1\alpha + K_2}{K_3\alpha + K_4}, \tag{4.21}$$

where

$$\begin{aligned} K_1 &= ((\nu-\kappa)(\mu+\gamma^{\square})\gamma + \gamma^{\square}\mu\nu)\zeta^{\square}, \\ K_2 &= \zeta\zeta^{\square}((\mu+\gamma)\gamma^{\square} + \gamma\mu), \\ K_3 &= ((\nu-\kappa)\gamma + \mu\nu)(1-\nu^{\square})\gamma^{\square} - \mu\nu^{\square}\gamma(\nu-\kappa), \\ K_4 &= \zeta((1-\nu^{\square})(\mu+\gamma)\gamma^{\square} - \gamma\mu\nu^{\square}). \end{aligned} \tag{4.22}$$

In an attempt of recreating critical values for $\alpha^{\square}$ and α as in (4.9), we obtain (4.21). However, in comparison to (4.9), critical values of $\alpha^{\square}$ and α for this model are not independent (see Figure 4.11).

Let us fix the parameters as follows:

$$\begin{gathered} \mu = 0.1, \qquad \gamma^{\square} = 0.25, \qquad \gamma = 0.25, \qquad \kappa = 1.0, \\ \nu^{\square} = 0.5, \qquad \nu = 0.5, \qquad \zeta^{\square} = 0.2222, \qquad \zeta = 0.2222. \end{gathered} \tag{4.23}$$

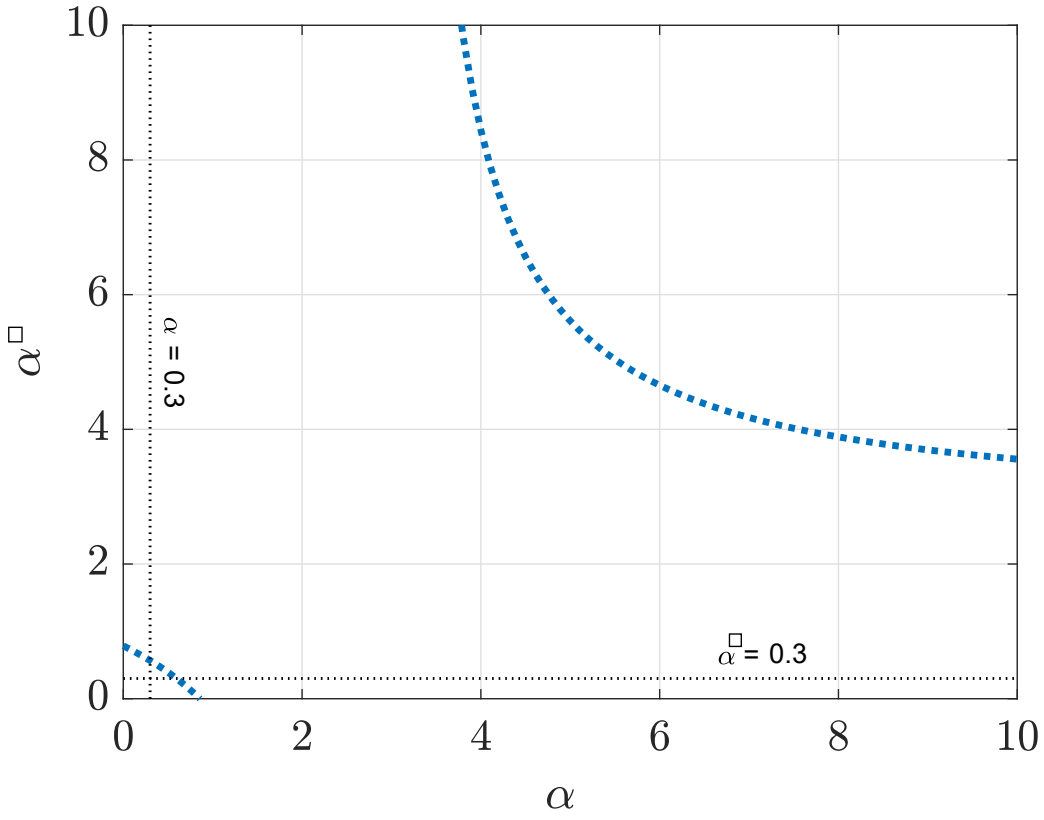


Figure 4.11: Two-dimensional bifurcation diagram for system (4.18) illustrating (4.21). Thin, black dotted lines show cross sections for one-dimensional bifurcation diagrams. Blue dashed curves correspond to the transcritical bifurcation.

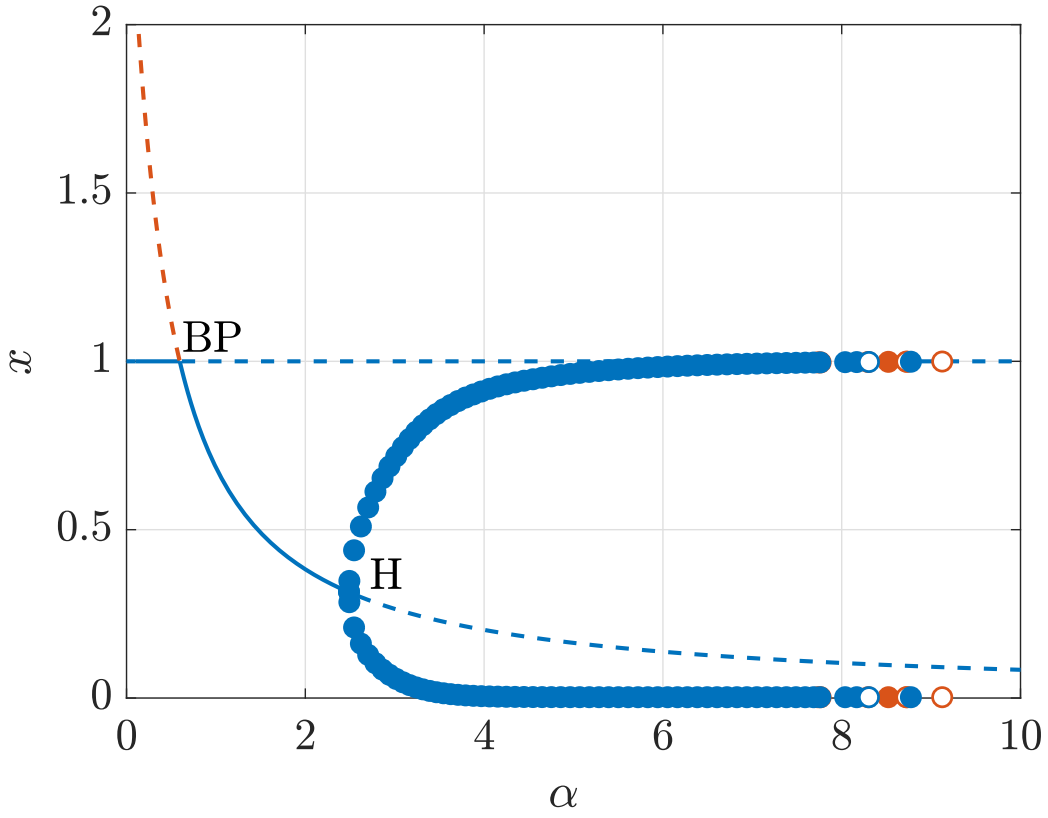


Figure 4.12: One-dimensional bifurcation diagram of model (4.18) for $\alpha^{\square} = 0.3$. Other parameters are fixed as in (4.23). Dashed curves correspond to unstable equilibria, while solid curves correspond to stable equilibria. Round markers indicate maximum and minimum of periodic orbits, and, if filled, markers show the stability of the periodic orbits. Colours blue and orange indicate all-positive and some-negative components, respectively, of both equilibria and periodic orbits.

For the set of parameters chosen above, one-dimensional cross sections are analysed as before. Figure 4.12 displays a one-dimensional bifurcation diagram at $\alpha^{\square} = 0.3$ and varying α. Notice that there is a qualitative resemblance with previous bifurcation diagrams. In this case, as for three-equation model (3.13), the bifurcation diagrams reveal a transcritical bifurcation (marked BP – branching point) and a Hopf bifurcation (labelled as H). The transcritical bifurcation uncovers the existence of a third equilibrium, $\boldsymbol{v}_2$, whose components are all non-zero. Moreover, from this bifurcation diagram, we conclude that (4.21) is, in fact, a transcritical bifurcation between equilibria $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$. Further increasing the bifurcation parameters, equilibrium $\boldsymbol{v}_2$ undergoes the Hopf bifurcation, which creates a stable periodic orbit.

Continuing critical values of parameters at Hopf points in a two-dimensional bifurcation diagram, we obtain Figure 4.13. The studied parameter space is divided into four regions.

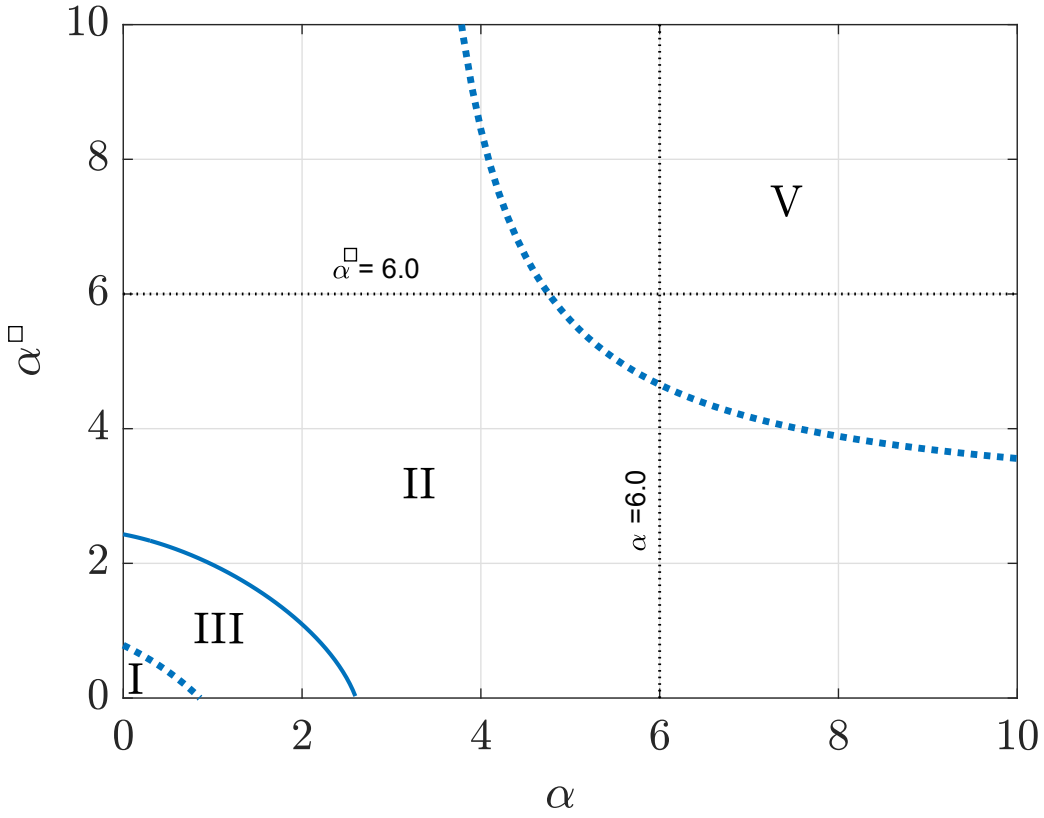


Figure 4.13: Two-dimensional bifurcation diagram for system (4.18) showing transcritical bifurcation in dotted curves and Hopf bifurcation in solid. Thin dotted lines show the cross sections chosen for one-dimensional bifurcation diagrams. Other parameters are fixed as in (4.23).

So far we know how the system behaves at regions I, III and partially at region II. Region I contains parameters for which the equilibrium $\boldsymbol{v}_1$ is a stable node, and for the parameters of region III, $\boldsymbol{v}_2$ is a stable node or a stable spiral. For all values of parameters close to the Hopf curve separating regions III and II, the equilibrium $\boldsymbol{v}_2$ develops a periodic orbit. Let us consider one-dimensional bifurcation diagrams for larger fixed value of α or $\alpha^{\square}$, due to the symmetry they should mirror each other.

The one-dimensional bifurcation diagram for fixed $\alpha^{\square}$ and growing α consists of three equilibria excluding the trivial solution, see Figure 4.14. All equilibria are unstable (dashed curves), and $\boldsymbol{v}_2$ has a stable periodic orbit around it for the smallest values of α. The stability of the periodic orbit appears to be interrupted around the point where $\boldsymbol{v}_1$ has a transcritical bifurcation (BP) with a new equilibrium, named $\boldsymbol{v}_3$. We know it is not $\boldsymbol{v}_2$, because $\boldsymbol{v}_3$ has negative components (indicated with orange as in the previous plots) and has its own unstable orbit created at Hopf bifurcation. The transcritical bifurcation of $\boldsymbol{v}_1$ and

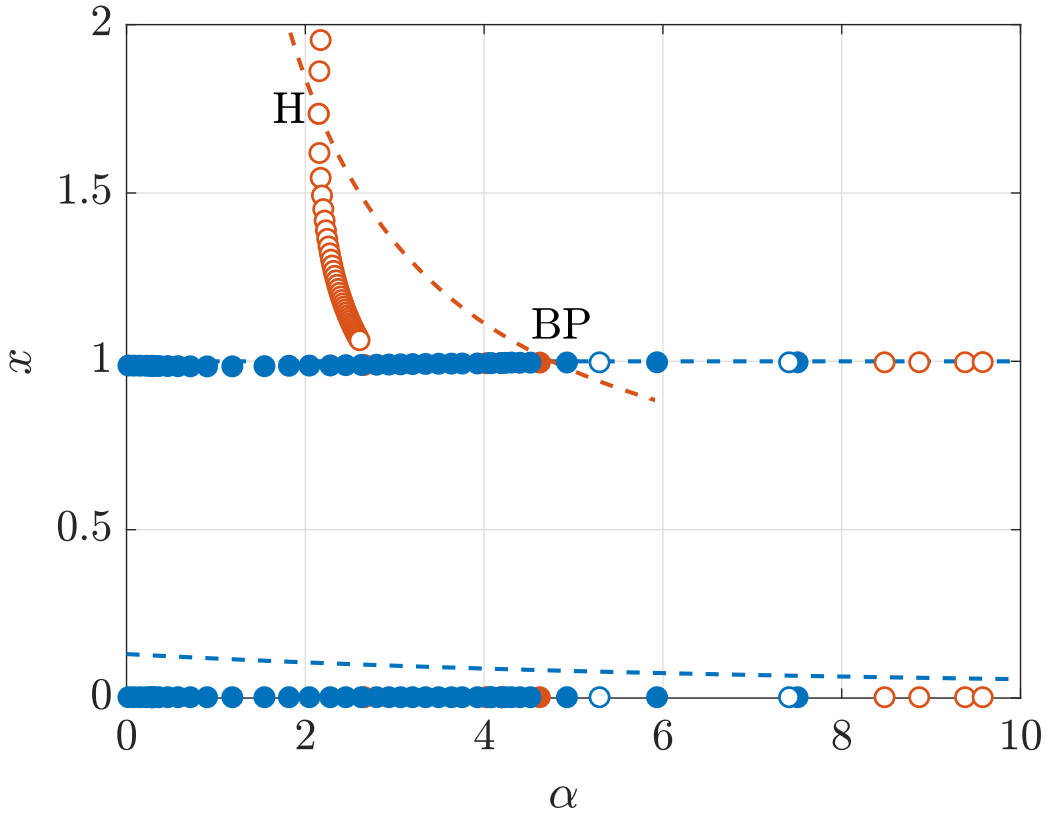
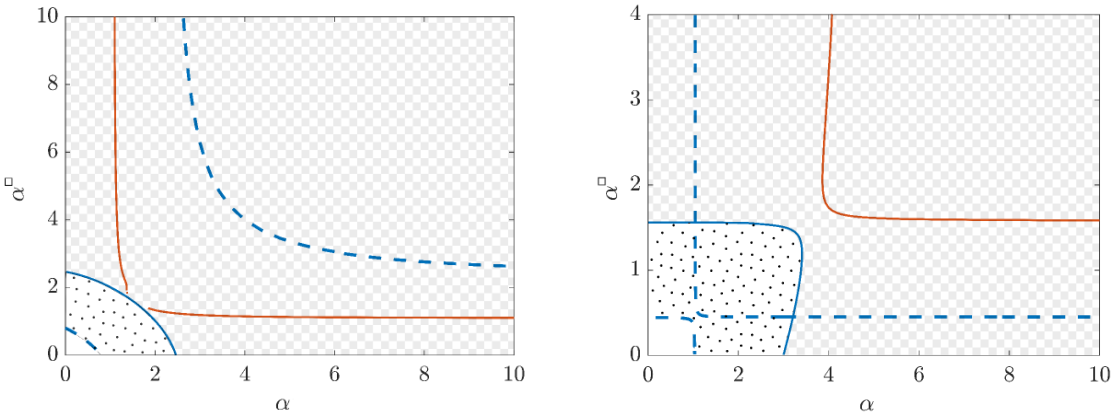


Figure 4.14: One-dimensional bifurcation diagram as cross sections of Figure 4.13 at $\alpha^{\square} = 6$ and other parameters are fixed as in (4.23). Dashed curves correspond to equilibria, round markers indicate maximum and minimum of periodic orbits. Filled markers state stable periodic orbits. Colours blue and orange indicate all-positive and some-negative components respectively of both equilibria and periodic orbits.

$\boldsymbol{v}_3$ does not change their stability. However, the equilibrium $\boldsymbol{v}_3 = \{x, y^{\square}, y, z^{\square}, z\}$ undergoing the transcritical bifurcation with $\boldsymbol{v}_1 = \{1, 0, 0, 0, 0\}$, at the point of bifurcation changes signs of its components from $sign(\boldsymbol{v}_3) = \{+, +, -, +, -\}$ to $sign(\boldsymbol{v}_3) = \{+, -, +, -, +\}$.

To explore how the two-dimensional bifurcation diagram given by Figure 4.8 (a) transforms into Figure 4.15 (a) as backwards mutation is added, we now consider a model whereby the "forwards" and "backwards" mutation rates are unequal. In dimensionless terms, this



(a) $\mu^{\square} = 10^{-1}$ and $\mu = 10^{-1}$ (b) $\mu^{\square} = 10^{-3}$ and $\mu = 10^{-1}$

Figure 4.15: Codimension two bifurcation diagrams in terms of the wt and mutant virus infection rates α and $\alpha^{\square}$ when backwards mutation is possible. The mutation rates are (a) $\mu^{\square} = 10^{-1}$ and $\mu = 10^{-1}$ and (b) $\mu^{\square} = 10^{-3}$ and $\mu = 10^{-1}$. Dashed and solid lines correspond to transcritical and Hopf bifurcations, respectively. Blue denotes biologically feasible equilibria undergoing the bifurcations (all components non-negative), while colour orange denotes bifurcations of equilibria with negative components. The non-patterned, polka-dot, and chequerboard pattern correspond to virus-free, stationary coexisting, and oscillatory coexisting states, respectively.

model is given by

$$\begin{aligned}
\dot{x} &= x(1-x) - \alpha^{\square}\, z^{\square}\, x - \alpha\, z\, x, \\
\dot{y}^{\square} &= \alpha^{\square}\, z^{\square}\, x + \mu y - \mu^{\square} y^{\square} - \gamma^{\square}\, y^{\square}, \\
\dot{y} &= \alpha\, z\, x + \mu^{\square} y^{\square} - \mu y - \gamma\, y, \\
\dot{z}^{\square} &= \gamma^{\square}\, y^{\square} - \nu^{\square}\, \alpha^{\square}\, z^{\square}\, x - \zeta^{\square}\, z^{\square}, \\
\dot{z} &= \kappa\gamma\, y - \nu\, \alpha\, z\, x - \zeta\, z.
\end{aligned} \tag{4.24}$$

We consider the case when $\mu^{\square} = 10^{-3} \ll \mu$ so that backwards mutation is slow and can be considered as a small perturbation to the original dimensionless system (4.1). The resulting two-dimensional bifurcation diagram is shown in Figure 4.15 (b). We see that the vertical and horizontal lines of transcritical bifurcations defined by $R_0 = R_0^{\square} = 1$, which intersected at right angles in Figure 4.8 (a) have now merged into two separate branches that do not intersect each other. The curves of Hopf bifurcations, which also intersected in the case of uni-directional mutation, have also merged into two distinct non-intersecting branches. The role of backwards mutation can be studied by re-building the two-dimensional bifurcation diagram shown in Figure 4.7 with different mutation rates.

The two-dimensional bifurcation diagrams in Figures 4.15 (a) and (b) reveal that backward mutation results in several important changes to the dynamics. Importantly, the behaviour of the system can be described using three main states: the virus-free state (analogous to Region I and displayed in white in Figure 4.15), stationary coexistence all populations (polka dot pattern in Figure 4.15, analogous to Region II), and oscillatory coexistence of all the populations (chequerboard pattern in Figure 4.15, analogous to Region IV). Furthermore, the BT and the ZH bifurcations have vanished, the latter of which implies that the curve of TPO bifurcations no longer exists either. Finally, the transcritical bifurcation between the wt-free and coexistent states $\boldsymbol{v}_2$ and $\boldsymbol{v}_3$ has vanished as well. The resulting stability diagram, shown in Figure 4.15 by patterns, is now considerably simpler and involves only analogues of Regions I, II, and IV. Thus, the only virus-persistent states are those in which both virus strains coexist. The wt-free state can no longer occur due to the creation of the wt virus through backwards mutation.

In summary, we have studied a mathematical model of population dynamics of two viral strains infecting a single population of host cells. Unlike previous models of parasite-host-like interaction (*e.g.* [95]), we consider the emergence of a second strain by mutation from a single strain introduced into the system. The mechanism of virus replication forces the consideration of two types of infected cells, one for each viral strain, in the model. We analysed the system of differential equations and its solutions. By studying parameters space, we identified five different regions, each characterised by distinct dynamics: (Region I) the virus-free state which maximises the population of host cells, (Region III) the stationary existence of mutant virus with nonzero cell populations supporting it, (Region II) the stationary persistence of both viral strains with nonzero cell populations, (Region V) the oscillating existence of the mutant virus and corresponding cell populations but extinct wt virus population, and finally, (Region IV) the oscillating coexistence of all populations.

The population of the mutant-type virus exists as long as the mutation rate of the wt strain is positive. The broken symmetry between the wt and mutant strains is clear from the stability diagrams based on infection rates. From our results, we observe that survival of the wt virus is essential for coexistence. However, the growth of wt virus population jeopardises its persistence in the system by creating its own competitor: the mutant-type virus. Moreover, the larger the mutation rate, the greater the infection rate of the original wt strain should be in order to remain in the system. Interestingly, in the context of the model, we discovered there is a maximum, critical value of the mutation rate which allows for the persistence of the wt strain and hence coexistence. Values of the mutation rate that are higher than the critical value make coexistence impossible even for the most infectious wt strain. Furthermore, our model shows that the concept of the error threshold may not be

considered as a one-parameter driven effect, *i.e.* based solely on the mutation rate [35, 20]. The critical value of mutation is shown here to be proportional to the virulence and burst size of the wt virus and proportionally inverse to its multiplicity of infection. Hence, our results extend the phenomenon of the error threshold at the infected cell population level.

4.3 Degenerate Bogdanov-Takens and zero-Hopf bifurcations

This is an additional section to discuss the technical background of the codimension-two bifurcations mentioned in this chapter. Normal form theory provides a means of predicting the local dynamics that occur near a given bifurcation. This is possible because the local structure of a bifurcation is model independent and thus can be immediately determined from pre-existing analyses of simpler systems of equations. The simplest system of equations that completely captures the dynamics of a particular bifurcation is called the normal form. The conducted numerical studies on the system (4.2) revealed that the Bogdanov-Takens (BT) and zero-Hopf (ZH) bifurcations are non-standard in the sense that the local dynamics do not agree with those predicted by their normal forms. Such a mismatch can occur when one of the terms in the normal form equations vanishes due to its coefficient being equal to zero, implying that an extended system of equations needs to be considered. We now examine the normal forms of the BT and ZH bifurcations in order to rationalise the discrepancies between the two-dimensional bifurcation diagram shown in Figure 4.7 and those predicted from normal form theory.

The local bifurcation diagram near a BT point can be determined from the normal form equations given by:

$$\dot{\xi}_0 = \xi_1, \tag{4.25a}$$

$$\dot{\xi}_1 = \beta_1 + \beta_2\xi_0 + a_2\xi_0^2 + b_2\xi_0\xi_1, \tag{4.25b}$$

provided that the coefficients a_2 and b_2 are not equal to zero, $a_2b_2 \neq 0$, [71, Sec. 8.4]. The quantities β_0 and β_1 in Eqs. (4.25) play the role of bifurcation parameters, with $\beta_0 = \beta_1 = 0$ corresponding to the BT point. The bifurcation diagram of system (4.25) consists of curves of saddle-node and Hopf bifurcations, neither of which were detected near the BT point in the local analysis of models considered in this chapter, as well as curves of global bifurcations involving homoclinic orbits. The procedure outlined by Kuzetnsov [70] enables the normal form coefficients to be related to the model studied here. Remarkably, the coefficient a_2 can be calculated analytically and is found to be equal to zero for all parameter combinations.

Thus, the BT bifurcation occurring in our model is always degenerate and the local dynamics will be different from those of Eqs. (4.25).

The vanishing of a_2 can be rationalised in terms of the number of equilibria at the DBT point. An equilibrium analysis shows that (4.25) will only have two equilibria at the DBT point $\beta_0 = \beta_1 = 0$ if $a_2 \neq 0$; however, the model under study has three. Thus, $a_2 = 0$ is needed to capture the triple equilibrium at the DBT point in our model. This also suggests that higher-order terms must be included in (4.25) in order for it to capture the DBT bifurcation in the model investigated here. While an extended system of normal form equations for degenerate BT bifurcations is proposed in Kuznetsov [70], it cannot capture the transcritical bifurcations found in the original model, suggesting that an alternative form is required. However, due to the simple nature of the local dynamics near the degenerate BT point, which can be obtained analytically from the full model, we do not pursue this point further.

The Poincaré normal form of a ZH bifurcation can be written as [71, Sec. 8.5]

$$\dot{v} = \gamma(\beta) + \frac{1}{2}G_{200}(\beta)v^2 + G_{011}(\beta)|w|^2 + \frac{1}{6}G_{300}(\beta)v^3 + G_{111}(\beta)v|w|^2, \tag{4.26a}$$

$$\dot{w} = \Lambda(\beta)w + H_{110}(\beta)vw + \frac{1}{2}H_{210}(\beta)v^2w + \frac{1}{2}H_{021}(\beta)w|w|^2, \tag{4.26b}$$

where $\beta = (\beta_1, \beta_2)$ is a vector of bifurcation parameters. Analyses of the normal form equations for the ZH bifurcation show that local dynamics depend on the values of the normal form coefficients H_{ijk} and G_{ijk}. A common feature between the various cases is that the ZH bifurcation occurs at the tangential intersection of curves of saddle-node and Hopf bifurcations. However, Figure 4.7 shows that the zero-Hopf bifurcation in our model lies at a transversal intersection of curves of transcritical and Hopf bifurcations. Furthermore, none of the normal forms predict that a curve of TPO bifurcations should emanate from a ZH point. The calculation of the three normal form coefficients for the DZH bifurcation is rather involved and must be performed numerically. We find that the normal form coefficient $G_{011}(0) = 0$ across a range of parameter values, indicating the ZH bifurcation is degenerate. The normal form equations for degenerate ZH bifurcations are not well known and previous studies have focused on different cases [127, 128]. Tigan *et al.* [129] showed that a degenerate ZH bifurcation with $G_{011}(0) = 0$ occurs in a Rössler-type system and used averaging theory to detect periodic orbits, leaving the structure of global bifurcations unresolved. We believe, as far as we know, that this is the first time when a curve of TPO bifurcations has been observed to emanate from a degenerate ZH bifurcation of this type. Determining the corresponding normal form is an interesting task for future work.

Chapter 5

Dynamical system with different cell types and mutant viral strains

For specialisation to exist as a concept of model, diverse sources must be presented. In this chapter, specialisation is introduced as adaptation towards a type of cell the viral strains infect predominantly. Specifically, we introduce two types of susceptible cells for an evolving viral population with different infection rates. When virus infects *mostly* one type of cell it is called a specialist, while if it infects different kinds of cells in comparable amounts then the virus is called a generalist. We study the dynamics of a simplified model for a generalist strain brought into system with two types of cells with the possibility to adapt towards a better niche. In extension of Chapter 4, bifurcation analysis is applied.

5.1 The model

Consider two different types of cells, x_1 and x_2, coexisting in a finite system with carrying capacity K. The uninfected cells x_1 and x_2 have growth and death rates β_1, δ_1 and β_2, δ_2, respectively. Throughout this chapter, $\beta_1 \neq \beta_2$ and $\delta_1 \neq \delta_2$ to differentiate two types of cell population; allowing β_1 and δ_1 to be equal to β_2 and δ_2 would mean considering the same type of cell population twice, which we avoid. Both types of cells can be infected by a strain of virus z, hereinafter referred to as a generalist, with a rate α, while only one of the cell types, x_1, can be infected by $z^{\square}$, a specialist, with a rate $\alpha^{\square}$. The rate of infection is being considered as a property of the virus strain, and is different from the incident rate

in a population of susceptible organisms. Consider these α and $\alpha^{\square}$ as the probabilities of a virion meeting and infecting a susceptible cell in the blood flow, in a tissue or *in vitro*. It is assumed that the generalist may specialise towards the first cell type once the cell is infected. Meanwhile, the second cell type remains susceptible for the generalist strain only. The part of the system describing dynamics of cell populations is modelled as follows:

$$\dot{x_1} = \beta_1 x_1 \left(1 - \frac{x_1 + x_2}{K}\right) - x_1 \alpha z - x_1 \alpha^{\square} z^{\square} - \delta_1 x_1, \tag{5.1a}$$

$$\dot{x_2} = \beta_2 x_2 \left(1 - \frac{x_1 + x_2}{K}\right) - x_2 \alpha z - \delta_2 x_2. \tag{5.1b}$$

Cells become linked to the strain of virus they were infected by, *i.e.* uninfected cells of type x_1 after being infected by $z^{\square}$ become infected cells $y_1^{\square}$, uninfected cells of type x_1 and x_2 after being infected by z become infected cells y_1 and y_2, respectively. The growth of the population of infected cells is assumed to be driven by the infection only, that is, the infected cells do not proliferate. From a biological perspective, these kind of infections are called lytic and they lead to exploitation of an infected cell by a virus after infection opposing the lysogenic infections. During the lysogenic cycle, the virus incorporates its genome into the genetic code of the infected cell and the infected cell grows by division as a healthy cell for several generations. The effect of infection on the death rate of the cell population is depicted by the change in death rates of infected cells. Moreover, the effect of particular virus is also preserved and, overall, the death rates of infected cells $y_1^{\square}$, y_1, and y_2 are $\gamma_1^{\square}$, γ_1, and γ_2, respectively. We also consider mutation of z into $z^{\square}$ with rate μ described by the transfer from population of y_1 to $y_1^{\square}$ since most of the mutations of viral strains occur during replication and transcription processes inside an infected cell. This flow corresponds to mutation of a generalist strain z into a specialist strain $z^{\square}$. The dynamics of infected cells can be modelled as follows:

$$\dot{y_1^{\square}} = \alpha^{\square} z^{\square} x_1 + \mu y_1 - \gamma_1^{\square} y_1^{\square}, \tag{5.1c}$$

$$\dot{y_1} = \alpha z x_1 - \mu y_1 - \gamma_1 y_1, \tag{5.1d}$$

$$\dot{y_2} = \alpha z x_2 - \gamma_2 y_2. \tag{5.1e}$$

The specialist infected cells $y_1^{\square}$ has burst size $\kappa^{\square}$, while generalist infected cells y_1 and y_2 have burst sizes κ_1 and κ_2, respectively. We let $\nu^{\square}$ and ν denote the multiplicity of infection (MOI) of $z^{\square}$ and z, respectively. Viral populations decrease at rates ζ for the generalist strain and $\zeta^{\square}$ for the specialist strain. Under these assumptions, the changes in sizes of the viral populations can be described as follows:

$$\dot{z^{\square}} = \kappa^{\square}\gamma_1^{\square}\, y_1^{\square} - \nu^{\square}\,\alpha^{\square}\, z^{\square}\, x_1 - \zeta^{\square}\, z^{\square}, \tag{5.1f}$$

$$\dot{z} = \kappa_1\,\gamma_1\, y_1 + \kappa_2\gamma_2\, y_2 - \nu\,\alpha\, z(x_1 + x_2) - \zeta\, z. \tag{5.1g}$$

Initial conditions of nonnegative values for $\boldsymbol{v} = \{x_1, x_2, y_1^{\square}, y_1, y_2, z^{\square}, z\}$ must be provided:

$$x_1(0) = X_1, x_2(0) = X_2, y_1^{\square}(0) = Y_1^{\square}, y_1(0) = Y_1, y_2(0) = Y_2, y_1^{\square}(0) = Z^{\square}, y_1(0) = Z. \tag{5.2}$$

The nondimensionalization of system (5.1) is performed by introducing dimensionless variables that are based on characteristic timescales and population sizes. The quantity $\beta_1 - \delta_1$ describes the effective growth rate of first uninfected cell type and its inverse, $(\beta_1 - \delta_1)^{-1}$, is used to define the characteristic time scale of the system. In a virus-free environment, the total maximum size of the uninfected cell populations x_1 is $\tilde{x}_{\max} = (1 - \delta_1/\beta_1)\, K$. Hence, characteristic population scales for both uninfected and infected cell populations are directly related to $\tilde{x}_{\max}$. Moreover, dimensionless variables for virus population sizes are proportional to $\tilde{x}_{\max}$, yet they are re-scaled with respect to the burst size. The latter is required due to a difference in sizes and measurement units of the viral loads and the cell populations in the system. We therefore nondimensionalize the variables according to

$$t = (\beta_1 - \delta_1)^{-1}\bar{t}, \quad x_i = \frac{\beta_1 - \delta_1}{\beta_1} K\, \bar{x}_i, \quad y_i^j = \frac{\beta_1 - \delta_1}{\beta_1} K\, \bar{y}_i^j, \quad z^j = \frac{\beta_1 - \delta_1}{\beta_1}\kappa^{\square} K\, \bar{z}^j,$$

where $i = 1, 2$ and j stands for the viral strain type. New non-dimensional parameters replace corresponding old parameters:

$$\bar{\beta}_1 = \frac{\beta_2 - \delta_2}{\beta_1 - \delta_1}, \quad \bar{\beta}_2 = \frac{\beta_2}{\beta_1}, \quad \bar{\alpha}^j = \frac{\kappa^{\square} K}{\beta_1}\alpha^j, \quad \bar{\nu}^j = \frac{\nu^j}{\kappa^{\square}}, \quad \bar{\kappa}_i = \frac{\kappa_i}{\kappa^{\square}}, \quad \bar{p} = \frac{p}{\beta_1 - \delta_1},$$

where p stands for all other parameters, namely: μ, $\gamma_1^{\square}$, γ_1, γ_2, $\zeta^{\square}$, ζ, and j shows the type of the strain. By omitting the bars for simplicity, we obtain the following nondimensionalized

Table 5.1: Schematic representation of the nine equilibria solutions to (5.3).

Equilibrium	**name**	x_1	x_2	$y_1^{\square}$	y_1	y_2	$z^{\square}$	z
$\boldsymbol{v}_0$	trivial	0	0	0	0	0	0	0
$\boldsymbol{v}_1$	v-free-1	x_1	0	0	0	0	0	0
$\boldsymbol{v}_2$	gen-free-1	x_1	0	$y_1^{\square}$	0	0	$z^{\square}$	0
$\boldsymbol{v}_3$	coex-1	x_1	0	$y_1^{\square}$	y_1	0	$z^{\square}$	z
$\boldsymbol{v}_4$	v-free-2	0	x_2	0	0	0	0	0
$\boldsymbol{v}_5$	spec-free	0	x_2	0	0	y_2	0	z
$\boldsymbol{v}_6$	gen-free-2	x_1	x_2	$y_1^{\square}$	0	0	$z^{\square}$	0
$\boldsymbol{v}_7$	coex-2	x_1	x_2	$y_1^{\square}$	y_1	y_2	$z^{\square}$	z
$\boldsymbol{v}_8$	coex-3	x_1	x_2	$y_1^{\square}$	y_1	y_2	$z^{\square}$	z

system:

$$\dot{x_1} = x_1\,(1 - x_1 - x_2) - x_1\,\alpha\,z - x_1\,\alpha^{\square}\,z^{\square}, \tag{5.3a}$$

$$\dot{x_2} = x_2\,(\beta_1 - \beta_2(x_1 + x_2)) - x_2\,\alpha\,z, \tag{5.3b}$$

$$\dot{y_1^{\square}} = \alpha^{\square}\,z^{\square}\,x_1 + \mu y_1 - \gamma_1^{\square}\,y_1^{\square}, \tag{5.3c}$$

$$\dot{y_1} = \alpha\,z\,x_1 - \mu y_1 - \gamma_1\,y_1, \tag{5.3d}$$

$$\dot{y_2} = \alpha\,z\,x_2 - \gamma_2\,y_2, \tag{5.3e}$$

$$\dot{z^{\square}} = \gamma_1^{\square}\,y_1^{\square} - \nu^{\square}\,\alpha^{\square}\,z^{\square}\,x_1 - \zeta^{\square}\,z^{\square}, \tag{5.3f}$$

$$\dot{z} = \kappa_1\,\gamma_1\,y_1 + \kappa_2\gamma_2\,y_2 - \nu\,\alpha\,z(x_1 + x_2) - \zeta\,z, \tag{5.3g}$$

The model (5.3) has nine solutions, which are only summarised in Table 5.1 due to the excess complexity of the full expressions. The first four solutions match the solutions of system (4.2) with the size of the four new populations being zero. This similarity illustrates that the model (5.3) contains the model (4.2). Indeed, if the newly introduced second type of cell x_2 and the corresponding generalist infected cell y_2 are zero, then the model (5.3) is equivalent to (4.2). We will exploit the results from the previous chapter to understand the dynamics of the extended model focusing on the determining how the addition of new cell and virus types affect the behaviour of the system.

5.2 Linear stability analysis and bifurcations

To gain insights into the general behaviour of model (5.3), we examine the local stability of the equilibria solutions and identify points where bifurcations occur. We note that

due to the high dimensionality of the system, these bifurcation points may not lead to changes in the behaviour of the system. For instance, if two equilibria intersect for a certain parameter set and first eigenvalues $\lambda_1[1]$ and $\lambda_2[1]$ of those equilibria $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ exchange signs, the bifurcation is detected; however, this might not lead to exchange of stability between the two equilibria $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ because other eigenvalues, *e.g.* $\lambda_1[4]$ or $\lambda_2[2]$, can remain positive after the bifurcation preventing the change of stability. The further use of numerical analysis enables the location of the bifurcation points that lead to qualitative changes to be determined. The local stability of the solutions may provide the initial structure of the stability diagram. A classical way to study the stability of equilibria is by means of linear stability analysis. Linearising the system around the equilibrium $\boldsymbol{v}$ we obtain the Jacobian of system (5.3), which is shown in (5.6). There is an elegant way to find possible boundaries of stability regions on a parameter space and these boundaries match the transcritical bifurcations. Therefore, alongside with the investigation of local stability, we will consider candidates for transcritical bifurcations, *i.e.* an intersection of solutions for certain parameter sets, which may lead to the exchange of stability between the involved solutions.

At the trivial solution $\boldsymbol{v}_0$, a straightforward calculation provides the following eigenvalues:

$$\Lambda_0 = (1, -\zeta^{\square}, -\zeta, -\gamma_1^{\square}, -\gamma_1 - \mu, -\gamma_2, \beta_2)^T . \tag{5.4}$$

The first eigenvalue $\lambda = 1$ is constant due to our choice of nondimensionalization. The biologically justified method of nondimensionalization forces the trivial solution to be a saddle for all biologically relevant values of the parameters.

At the first virus-free state $\boldsymbol{v}_1$ (v-free-1), we obtain the following characteristic polynomial for the eigenvalues λ: $P_1(\lambda) \cdot P_2(\lambda) \cdot P_3(\lambda) \cdot P_4(\lambda) \cdot P_5(\lambda) = 0$, where

$$\begin{aligned}
P_1(\lambda) &= \lambda + 1, \\
P_2(\lambda) &= \lambda + \gamma_2, \\
P_3(\lambda) &= \beta_1 - \beta_2 - \lambda, \\
P_4(\lambda) &= -\lambda^2 + (-\alpha\nu - \gamma_1 - \mu - \zeta)\lambda + ((\kappa_1 - \nu)\gamma_1 - \mu\,\nu)\alpha + \zeta(\gamma_1 + \mu), \\
P_5(\lambda) &= \lambda^2 + (\alpha^{\square}\nu^{\square} + \gamma_1^{\square} + \zeta^{\square})\lambda + (\alpha^{\square}(\nu^{\square} - 1) + \zeta^{\square})\gamma_1^{\square}.
\end{aligned} \tag{5.5}$$

The first two factors yield negative eigenvalues $\lambda = -1$ and $\lambda = -\gamma_2$, giving rise to the possibility of stability for the first virus-free state. Third eigenvalue $\lambda = \beta_1 - \beta_2$ leads to a case where growth rates of uninfected cells must be compared. If $\beta_2 > \beta_1$ then

$$\begin{bmatrix}
1-\alpha^{\square}z^{\square}-\alpha z-2\,x_1-x_2 & -x_1 & 0 & 0 & 0 & -\alpha^{\square}\,x_1 & -\alpha\,x_1 \\
\beta_2\,x_2 & \beta_1-\alpha z-\beta_2\,x_1-2\,\beta_2\,x_2 & 0 & 0 & 0 & 0 & -\alpha\,x_2 \\
\alpha^{\square}\,z^{\square} & 0 & -\gamma_1^{\square} & \mu & 0 & \alpha^{\square}\,x_1 & 0 \\
\alpha\,z & 0 & 0 & -\mu-\gamma_1 & 0 & 0 & \alpha\,x_1 \\
0 & \alpha\,z & 0 & 0 & -\gamma_2 & 0 & \alpha\,x_2 \\
-\alpha^{\square}\,\nu^{\square}\,z^{\square} & 0 & \gamma_1^{\square} & 0 & 0 & -\alpha^{\square}\,\nu^{\square}\,x_1-\zeta^{\square} & 0 \\
-\nu\,\alpha\,z & -\nu\,\alpha\,z & 0 & \gamma_1\,\kappa_1 & \gamma_2\,\kappa_2 & 0 & -\nu\,\alpha\,(x_1+x_2)-\zeta
\end{bmatrix} \tag{5.6}$$

The Jacobian matrix J for the system (5.3) at $\boldsymbol{v}=\{x_1,\,x_2,\,y_1^{\square},\,y_1,\,y_2,\,z^{\square},\,z\}$.

$\boldsymbol{v}_1$ can be stable, if $\beta_2 < \beta_1$ then $\boldsymbol{v}_1$ cannot be stable, and if $\beta_2 = \beta_1$ then $\boldsymbol{v}_1$ must undergo one of the zero-eigenvalue bifurcations (*e.g.* saddle-node, transcritical, pitchfork). The fourth, quadratic, factor $P_4(\lambda)$ provides combination of parameters corresponding to a zero-eigenvalue bifurcation:

$$\alpha_{\boldsymbol{v}_{1,1}} := \alpha = \frac{\gamma_1 + \mu}{(\kappa_1 - \nu)\gamma_1 - \mu\nu}\zeta. \tag{5.7}$$

The last factor, $P_5(\lambda)$, yields another parameter set for a zero-eigenvalue bifurcation:

$$\alpha^{\square}_{\boldsymbol{v}_{1,2}} := \alpha^{\square} = \frac{\zeta^{\square}}{1 - \nu^{\square}}. \tag{5.8}$$

The zero-eigenvalue bifurcations of $\boldsymbol{v}_1$, $\alpha_{\boldsymbol{v}_{1,1}}$ and $\alpha^{\square}_{\boldsymbol{v}_{1,2}}$, to a great extent resemble the transcritical bifurcations of the virus-free state from model (4.2). As will be shown further, these parameter combinations are indeed two of the three curves which intersect at a degenerate Bogdanov-Takens bifurcation.

For the first generalist-free state $\boldsymbol{v}_2$ (gen-free-1), we conduct a similar analysis and find that it has three parameter combinations that lead to a zero eigenvalue and does not have strictly positive constant eigenvalues. The first zero-eigenvalue bifurcation of the gen-free-1 state is

$$\alpha^{\square}_{\boldsymbol{v}_{2,1}} := \alpha^{\square} = \frac{\zeta^{\square}}{1 - \nu^{\square}}. \tag{5.9}$$

Notice, that at $\alpha^{\square}_{\boldsymbol{v}_{2,1}}$, the gen-free-1 state undergoes zero-eigenvalue bifurcation when at the same parameter combination, $\alpha^{\square}_{\boldsymbol{v}_{1,2}}$, the v-free-1 state is also subjected to a zero-eigenvalue bifurcation. Moreover, the values of $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ are equal when $\alpha^{\square} = \zeta^{\square}/(1 - \nu^{\square})$. This is a strong evidence that $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ intersect and exchange stability at $\alpha^{\square}_{\boldsymbol{v}_{2,1}}$, however, this conclusion can be drawn only after numerical check of signs of all eigenvalues of both equilibria before and after $\alpha^{\square}_{\boldsymbol{v}_{2,1}}$. The gen-free-1 state has a second

$$\alpha^{\square}_{\boldsymbol{v}_{2,2}} := \alpha^{\square} = \frac{\beta_2}{\beta_1}\frac{\zeta^{\square}}{1 - \nu^{\square}}, \tag{5.10}$$

and third

$$\alpha^{\square}_{\boldsymbol{v}_{2,3}} := \alpha^{\square} = \alpha\frac{\zeta^{\square}}{\zeta}\frac{(\kappa_1 - \nu)\gamma_1 - \mu\nu}{(\gamma_1 + \mu)(1 - \nu^{\square})}. \tag{5.11}$$

set of parameters which lead to a zero eigenvalue. Further analysis will reveal that other equilibria undergo a zero-eigenvalue bifurcation at these exact combinations of parameters.

We now consider the first coexistence state $\boldsymbol{v}_3$ (coex-1) that has a complicated characteristic polynomial. However, it is still possible to recover analytical parameter combinations which correspond to zero eigenvalues. The coex-1 state undergoes zero-eigenvalue bifurcations at

$$\alpha_{\boldsymbol{v}_{3,1}} := \alpha = \frac{\gamma_1+\mu}{(\kappa_1-\nu)\gamma_1-\mu\nu}\zeta, \qquad \alpha^{\square}_{\boldsymbol{v}_{3,2}} := \alpha^{\square} = \alpha\frac{\zeta^{\square}}{\zeta}\frac{(\kappa_1-\nu)\gamma_1-\mu\nu}{(\gamma_1+\mu)(1-\nu^{\square})}, \tag{5.12}$$

and

$$\alpha^{\square}_{\boldsymbol{v}_{3,3}} := \alpha^{\square} = \frac{\zeta^{\square}}{\zeta}\frac{K_0^2(\beta_1-1)\alpha - K_0\,\zeta(\beta_2-1)(\gamma_1+\mu)}{K_0\,K_1\,\alpha+K_2}, \tag{5.13}$$

where

$$\begin{aligned} K_0 &= (\kappa_1-\nu)\gamma_1-\mu\nu,\\ K_1 &= (\beta_1-1)(1-\nu^{\square})\gamma_1+(1+(\beta_1-1)\,\nu^{\square})\mu,\\ K_2 &= (\gamma_1+\mu)((1+\nu^{\square}(\beta_2-1))\mu-(1-\nu^{\square})(\beta_2-1)\gamma_1)\zeta. \end{aligned} \tag{5.14}$$

We see that $\alpha_{\boldsymbol{v}_{1,1}}$ matches $\alpha_{\boldsymbol{v}_{3,1}}$. This implies that both $\boldsymbol{v}_1$ and $\boldsymbol{v}_3$ undergo a zero-eigenvalue bifurcation at $\alpha_{\boldsymbol{v}_{3,1}}$; by substituting $\alpha_{\boldsymbol{v}_{3,1}}$ into $\boldsymbol{v}_3$ and $\boldsymbol{v}_1$, we observe that $\boldsymbol{v}_3$ and $\boldsymbol{v}_1$ intersect. It is also clear that the coex-1 state, $\boldsymbol{v}_3$, undergoes another zero-eigenvalue bifurcation at $\alpha^{\square}_{\boldsymbol{v}_{3,2}}$ when the gen-free-1 state, $\boldsymbol{v}_2$, undergoes a similar bifurcation at $\alpha^{\square}_{\boldsymbol{v}_{2,3}}$. Checking the values of $\boldsymbol{v}_3$ and $\boldsymbol{v}_2$ at $\alpha^{\square}_{\boldsymbol{v}_{3,2}}$, we hypothesize that $\boldsymbol{v}_3$ and $\boldsymbol{v}_2$ have a transcritical bifurcation at $\alpha^{\square}_{\boldsymbol{v}_{3,2}}$. At this point, we know that $\boldsymbol{v}_3$ might have transcritical bifurcations with $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$, and if the two transcritical bifurcations intersect, then a codimension-two bifurcation can take place. From the previous occurrence of codimension-two bifurcations in system (4.2), we know that this bifurcation is an analogue of the degenerate Bogdanov-Takens bifurcation. This bifurcation exists, in terms of infection rates, when the following equations hold simultaneously:

$$\alpha = \frac{\gamma_1+\mu}{(\kappa_1-\nu)\gamma_1-\mu\nu}\zeta, \qquad \alpha^{\square} = \frac{\zeta^{\square}}{1-\nu^{\square}}. \tag{5.15}$$

Considering the second virus-free state $\boldsymbol{v}_4$ (v-free-2), we learn that, despite its simple appearance, eigenvalues of the system at $\boldsymbol{v}_4$ are not convenient to study. Thus, we construct

a characteristic polynomial: $\beta_2^{-2} \cdot Q_1(\lambda) \cdot Q_2(\lambda) \cdot Q_3(\lambda) \cdot Q_4(\lambda) \cdot Q_5(\lambda) \cdot Q_6(\lambda) = 0$, where

$$\begin{aligned}
Q_1(\lambda) &= \lambda + \zeta^{\square}, \\
Q_2(\lambda) &= \lambda + \gamma_1 + \mu, \\
Q_3(\lambda) &= \gamma_1^{\square} + \lambda, \\
Q_4(\lambda) &= \beta_1 + \lambda, \\
Q_5(\lambda) &= \beta_2\,\lambda + \beta_1 - \beta_2, \\
Q_6(\lambda) &= \beta_2\lambda^2 + ((\zeta + \gamma_2)\beta_2 + \nu\,\beta_1\alpha)\lambda + \gamma_2\,(\zeta\,\beta_2 - \beta_1(\kappa_2 - \nu)\alpha).
\end{aligned} \tag{5.16}$$

The first four eigenvalues, $\lambda = -\zeta^{\square}$, $\lambda = -\gamma_1 - \mu$, $\lambda = -\gamma_1^{\square}$, and $\lambda = -\beta_1$, provide a possibility for the v-free-2 state to be stable. The first compound factor is $\beta_2\,\lambda + \beta_1 - \beta_2$, and it stipulates that for $\boldsymbol{v}_4$ to be stable, β_2 must be less than β_1. Here, we may have a "switch" between v-free-1 and v-free-2 states. Recall that v-free-1 could become stable only if β_2 is greater than β_1. The last factor of the characteristic polynomial for $\boldsymbol{v}_4$ provides the following candidate for a zero-eigenvalue bifurcation:

$$\alpha_{\boldsymbol{v}_{4,1}} := \alpha = \frac{\beta_2}{\beta_1}\frac{\zeta}{\kappa_2 - \nu}. \tag{5.17}$$

The next equilibrium, the specialist-free state $\boldsymbol{v}_5$ (spec-free) depicts the only case of the generalist population surviving alone. The existence of this state is the major difference between systems (5.3) and (4.2). The spec-free state has a nonzero generalist population z, which due to the specifics of modelling, mutates into specialist population $z^{\square}$. As we observed before, survival of the wild-typed strain (an analogue of the generalist in model (4.2)) without a mutant strain was unlikely. However, introducing the second, only-generalist susceptible type of cell entirely changed the outcome. Whenever $\boldsymbol{v}_5$ is stable, the system predicts the survival of the generalist in the long run with an extinct specialist population. In order to find when $\boldsymbol{v}_5$ is stable, we consider the characteristic polynomial for the eigenvalues λ:

$$\frac{1}{(\alpha)^2(\kappa_2-\nu)^3}(\lambda+\zeta^{\square})(\lambda+\gamma_1+\mu)(\gamma_1^{\square}+\lambda)\cdot$$
$$\Big(\alpha\,(\kappa_2-\nu)^2\lambda^3+(\kappa_2-\nu)((\alpha\,\kappa_2+\beta_2)\zeta+\gamma_2\alpha\,(\kappa_2-\nu))\lambda^2+$$
$$(\beta_2(\kappa_2+\nu)\zeta-(\kappa_2-\nu)(\alpha\,\beta_1\nu-\beta_2\gamma_2))\zeta\lambda+(\kappa_2-\nu)(\beta_1\alpha\,(\kappa_2-\nu)-\zeta\beta_2)\gamma_2\zeta\Big)\cdot \tag{5.18}$$
$$\big(\alpha\,(\kappa_2-\nu)\,\lambda+(\kappa_2-\nu)(\beta_1-1)\alpha-\zeta(\beta_2-1)\big).$$

Neither the fraction nor the following three factors affect the stability of the spec-free state, $\boldsymbol{v}_5$. However, the next, cubic factor implies that a zero-eigenvalue bifurcation occurs when

$$\alpha_{\boldsymbol{v}_{5,1}} := \alpha = \frac{\beta_2}{\beta_1}\frac{\zeta}{\kappa_2-\nu}. \tag{5.19}$$

Matching $\alpha_{\boldsymbol{v}_{5,1}}$ with $\alpha_{\boldsymbol{v}_{4,1}}$, we expect a transcritical bifurcation of v-free-2 and spec-free states at points of parameter space corresponding to $\alpha_{\boldsymbol{v}_{5,1}}$. The last, linear expression for the eigenvalue crosses zero when the following condition holds:

$$\alpha_{\boldsymbol{v}_{5,2}} := \alpha = \frac{\beta_2-1}{\beta_1-1}\frac{\zeta}{\kappa_2-\nu}. \tag{5.20}$$

The last analytically computable equilibrium solution, the generalist-free state $\boldsymbol{v}_6$ (gen-free-2), has a complicated characteristic polynomial. Nevertheless, it is possible to recover two parameter combinations for zero-eigenvalue bifurcations:

$$\alpha^{\square}_{\boldsymbol{v}_{6,1}} := \alpha^{\square} = \alpha\,\frac{\beta_2\zeta^{\square}\,(\kappa_1\,\gamma_1-\kappa_2(\mu+\gamma_1))}{(1-\nu^{\square})(\mu+\gamma_1)(\beta_1\zeta-\beta_1\alpha(\kappa_2-\nu))} \tag{5.21}$$

and

$$\alpha^{\square}_{\boldsymbol{v}_{6,2}} := \alpha^{\square} = \frac{\beta_2}{\beta_1}\frac{\zeta^{\square}}{1-\nu^{\square}}. \tag{5.22}$$

The condition $\alpha^{\square}_{\boldsymbol{v}_{6,2}}$ is equal to $\alpha^{\square}_{\boldsymbol{v}_{2,2}}$ meaning that gen-free-2 and gen-free-1 state intersect and undergo the zero-eigenvalue bifurcation simultaneously. All analytically found parameter combinations are schematically plotted in Figure 5.1. Here and further, we call a curve of transcritical bifurcations between equilibria $\boldsymbol{v}_i$ and $\boldsymbol{v}_j$ in the two-dimensional bifurcation diagram Tij.

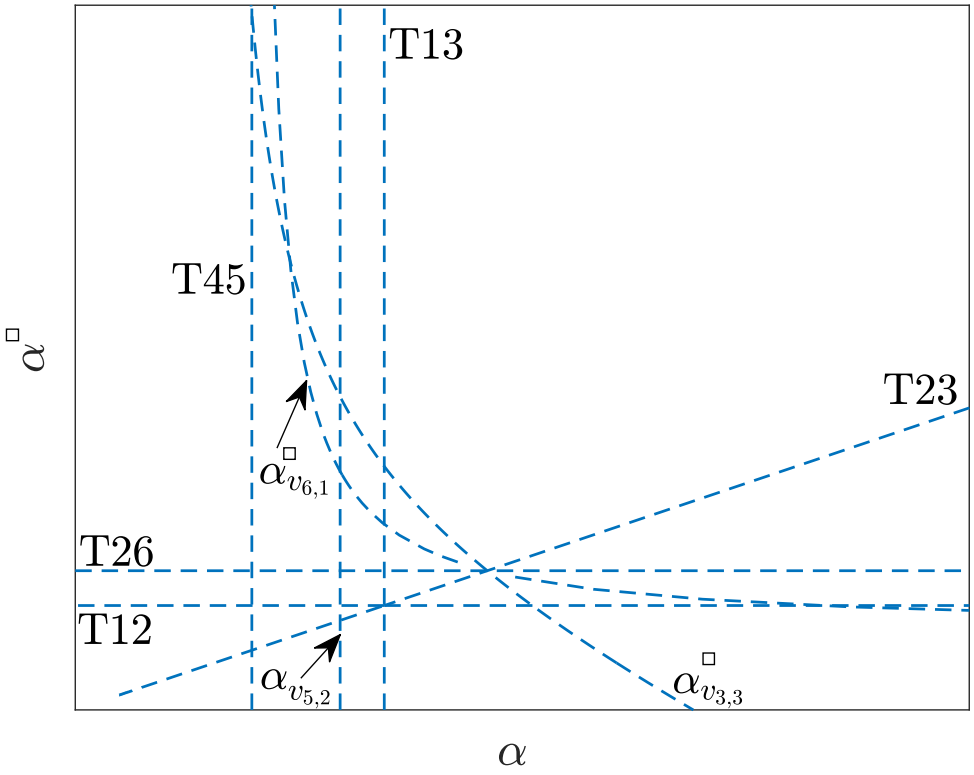


Figure 5.1: Schematic representation of analytically found parameter combinations leading to a zero-eigenvalue bifurcation. Cases when two equilibria have the same parameter combination are the strong candidates for a transcritical bifurcation, and are marked Tij on the plot, where i and j are vector numbers of the corresponding equilibria from Table 5.1.

Table 5.2: List of transcritical bifurcations with analytically obtained expressions.

bifurcation marker Tij	**combination of parameters**	**analytical**	**names**
T12	$\alpha^{\square} = \frac{\zeta^{\square}}{1-\nu^{\square}}$	$\alpha^{\square}_{\boldsymbol{v}_{1,2}}$	$\alpha^{\square}_{\boldsymbol{v}_{2,1}}$
T13	$\alpha = \frac{\gamma_1+\mu}{(\kappa_1-\nu)\gamma_1-\mu\nu}\zeta$	$\alpha_{\boldsymbol{v}_{1,1}}$	$\alpha_{\boldsymbol{v}_{3,1}}$
T23	$\alpha^{\square} = \alpha\frac{\zeta^{\square}}{\zeta}\frac{(\kappa_1-\nu)\gamma_1-\mu\nu}{(\gamma_1+\mu)(1-\nu^{\square})}$	$\alpha^{\square}_{\boldsymbol{v}_{2,3}}$	$\alpha^{\square}_{\boldsymbol{v}_{3,2}}$
T26	$\alpha^{\square} = \frac{\beta_2}{\beta_1}\frac{\zeta^{\square}}{1-\nu^{\square}}$	$\alpha^{\square}_{\boldsymbol{v}_{2,2}}$	$\alpha^{\square}_{\boldsymbol{v}_{6,2}}$
T37	$\alpha^{\square} = \frac{\zeta^{\square}}{\zeta}\frac{K_0^2(\beta_1-1)\alpha-K_0\,\zeta(\beta_2-1)(\gamma_1+\mu)}{K_0\,K_1\,\alpha+K_2}$	$\alpha^{\square}_{\boldsymbol{v}_{3,3}}$	-
T45	$\alpha = \frac{\beta_2}{\beta_1}\frac{\zeta}{\kappa_2-\nu}$	$\alpha^{\square}_{\boldsymbol{v}_{4,1}}$	$\alpha^{\square}_{\boldsymbol{v}_{5,1}}$
T57	$\alpha = \frac{\beta_2-1}{\beta_1-1}\frac{\zeta}{\kappa_2-\nu}$	$\alpha_{\boldsymbol{v}_{5,2}}$	-
T67	$\alpha^{\square} = \alpha\,\frac{\beta_2\zeta^{\square}\,(\kappa_1\,\gamma_1-\kappa_2(\mu+\gamma_1))}{(1-\nu^{\square})(\mu+\gamma_1)(\beta_1\zeta-\beta_1\alpha(\kappa_2-\nu))}$	$\alpha^{\square}_{\boldsymbol{v}_{6,1}}$	-

5.3 Bifurcation analysis

In the previous section, some of the parameter combinations were left without a match, *e.g.* $\alpha^{\square}_{\boldsymbol{v}_{6,2}}$ and $\alpha^{\square}_{\boldsymbol{v}_{3,3}}$. This is because we did not consider the solutions which correspond to $\boldsymbol{v}_7$ coex-2 and $\boldsymbol{v}_8$ coex-3 states due to their complicated appearance. The parameter combinations which lead to zero-eigenvalue bifurcations of $\boldsymbol{v}_7$ and $\boldsymbol{v}_8$ are analysed numerically. In order to proceed, first, let us fix all the parameters except the infection rates as follows:

$$\begin{gathered}\beta_1 = 1.5, \qquad \beta_2 = 2, \qquad \mu = 0.1, \qquad \gamma_1^{\square} = \gamma_1 = \gamma_2 = 0.25, \\ \kappa_1 = \kappa_2 = 1, \qquad \nu^{\square} = \nu = 0.5, \qquad \zeta^{\square} = \zeta = 0.2222.\end{gathered} \tag{5.23}$$

With these parameters fixed, we find that $\alpha^{\square}_{\boldsymbol{v}_{3,3}}$, $\alpha_{\boldsymbol{v}_{5,2}}$, and $\alpha^{\square}_{\boldsymbol{v}_{6,1}}$ correspond to intersections of the equilibrium solutions $\boldsymbol{v}_3$, $\boldsymbol{v}_5$, and $\boldsymbol{v}_6$ with $\boldsymbol{v}_7$, the coex-2 state. The list of transcritical bifurcations and their analytically obtained expressions can be found in Table 5.2. All transcritical bifurcations are plotted in Figure 5.2. In order to understand which of these bifurcations change stability of the states, we construct one-dimensional bifurcation diagrams as cross sections along fixed infection rates at the regions bounded by the transcritical bifurcations. In case of model (4.2) from the previous chapter, the x component was different for all of its four equilibria. Conversely, the model considered in this chapter does not allow the equilibrium vectors to be distinguished by tracking a single component along

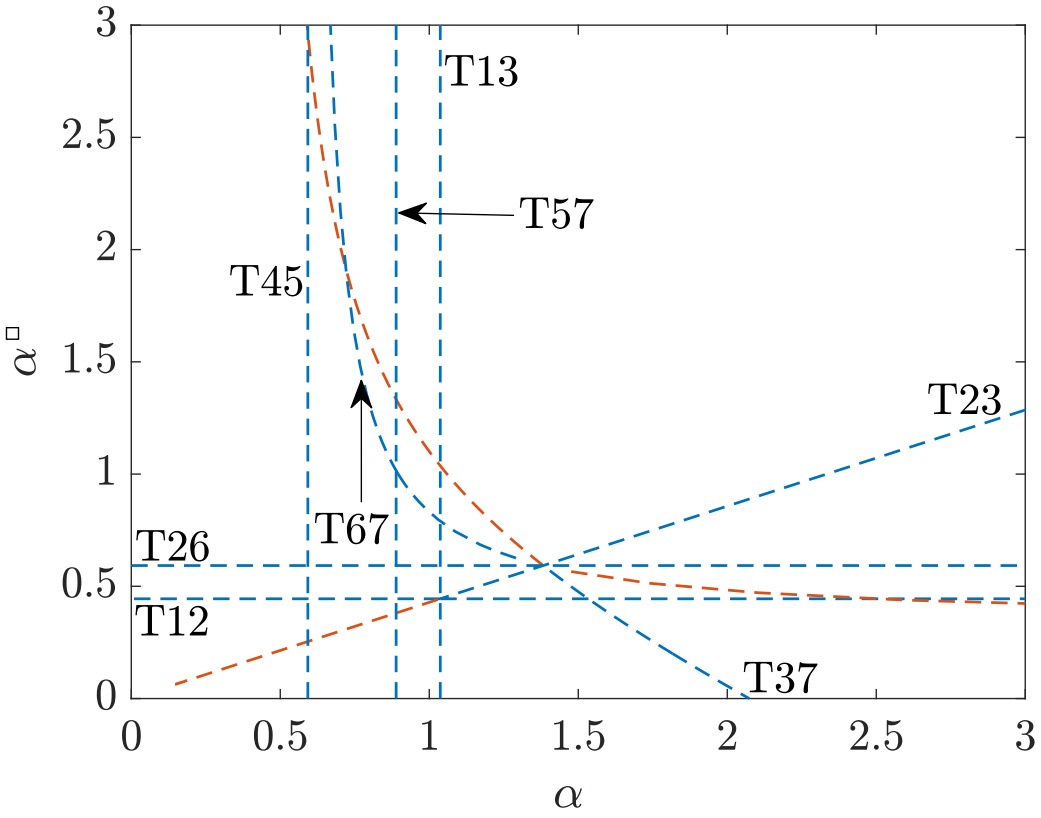


Figure 5.2: Analytically and numerically found parameter combinations leading to a zero-eigenvalue bifurcation. Transcritical bifurcation are marked Tij on the plot, where i and j are vector numbers of the corresponding equilibria from Table 5.1. Orange colour indicates biologically meaningless values of corresponding equilibria with negative coordinates. Parameters are fixed as in (5.23).

a parameter. Thus we must either track several variables at once or consider a measure to describe the vectors. The populations studied in this model form a vector in a normed vector space with, for example, the following norm:

$$V_i = ||\boldsymbol{v}_i|| = \sqrt{x_1^2 + x_2^2 + (y_1^{\square})^2 + y_1^2 + y_2^2 + (z^{\square})^2 + z^2}, \tag{5.24}$$

where $i = 0, \ldots, 8$. Therefore, in further plots if not mentioned otherwise V_i with $i = 0, \ldots, 8$ stands for the length of the corresponding vector from Table 5.1, $\boldsymbol{v}_i$. The definition of the norm might lead to seemingly intersecting curves on a plot, which may happen when the vectors lay on a sphere of the same radius in a seven-dimensional variable space. This will be recognised and will be handled accordingly. When plotting periodic orbits we adapt the same norm for maxima and minima of the periodic orbit.

In terms of Figure 5.2, by fixing $\alpha^{\square}$ at different values and varying α we can construct one-dimensional bifurcation diagrams. The first diagram illustrates the behaviour in the regions below T12, it corresponds to a fixed value of $\alpha^{\square}$ at 0.2 and varying values of α (see Figure 5.3). Due to the expected complexity of the bifurcation diagram, we only show biologically meaningful values. When $\alpha^{\square} = 0.2$, there are three equilibria which at some point become stable. For small values of α, the v-free-1 state $\boldsymbol{v}_1$ remains stable until it crosses and exchanges stability with coex-1 state $\boldsymbol{v}_3$ at $\alpha = 1.04$. This matches with our analytically obtained prediction and corresponds to T13. The stable equilibrium $\boldsymbol{v}_3$ then intersects and exchanges stability with the coex-2 state $\boldsymbol{v}_7$ at $\alpha = 1.82$ where coex-2 gains stability but some of its components become negative and thus it becomes biologically meaningless, and therefore is not shown in Figure 5.3. The intersection point of $\boldsymbol{v}_3$ and $\boldsymbol{v}_7$ was predicted by $\alpha_{\boldsymbol{v}_{3,3}}$ and corresponds to a maximum of α for the bistable interval. Following the same unstable (dashed) curve of $\boldsymbol{v}_7$ backwards, we notice that it intersects and changes stability of spec-free state $\boldsymbol{v}_5$ at $\alpha = 0.89$. This was also predicted by $\alpha_{\boldsymbol{v}_{5,2}}$ and corresponds to the minimum α of the bistable interval. Between the minimum and maximum values of the bistable interval we witness two different types of bistability: v-free-1 state $\boldsymbol{v}_1$ and spec-free state $\boldsymbol{v}_5$ are simultaneously stable, and then coex-1 state $\boldsymbol{v}_3$ and spec-free state $\boldsymbol{v}_5$ are both stable. Furthermore, the spec-free state $\boldsymbol{v}_5$ gains stable periodic orbits at a supercritical Hopf bifurcation. The Hopf bifurcation of the coex-1 state $\boldsymbol{v}_3$, as we have seen in the more simple model (4.2), does exist but its periodic orbits is unstable and does not intersect any other periodic orbits.

To explore the regions between T12 ($\alpha^{\square} = 0.45$) and T26 ($\alpha^{\square} = 0.59$) of Figure 5.2, let us fix $\alpha^{\square} = 0.53$ and consider a one-dimensional bifurcation diagram varying α, which is shown in Figure 5.4. This one-dimensional bifurcation diagram is understandably similar

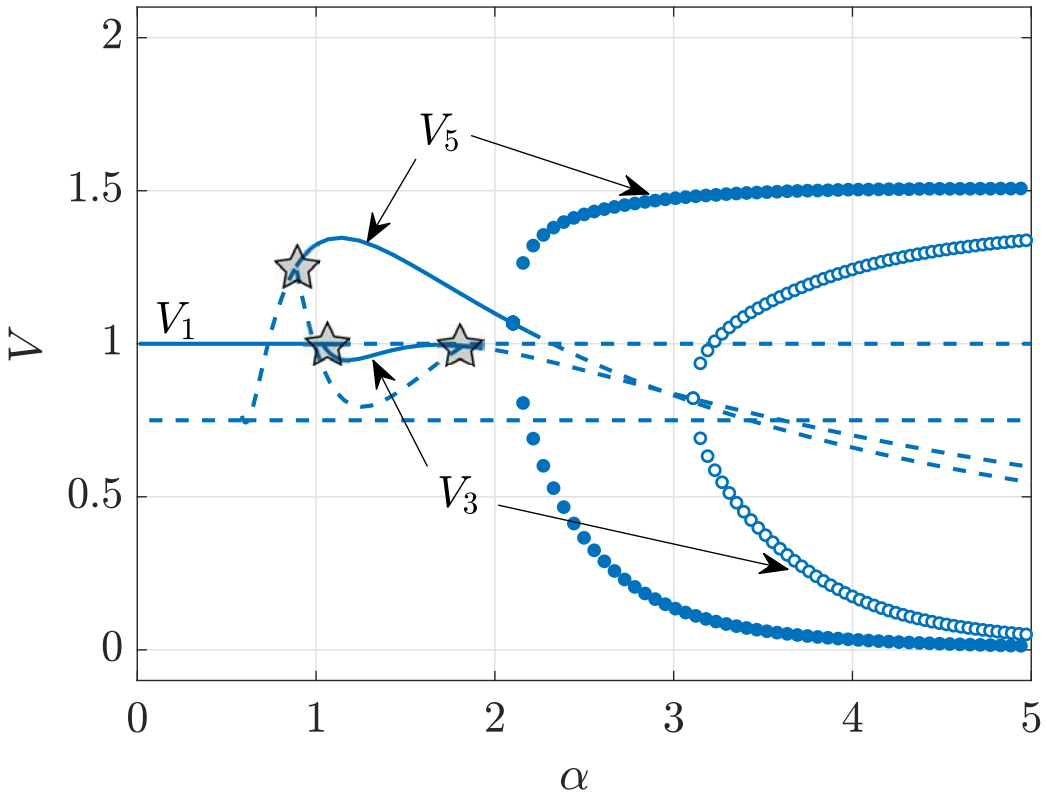


Figure 5.3: One-dimensional bifurcation diagram depicting the norm of all biologically meaningful equilibria, V_i for $i = 0, \ldots, 8$, for a fixed value of $\alpha^{\square} = 0.2$. Solid and dashed lines correspond to stable and unstable equilibria, respectively. Filled and empty markers represent stable and unstable periodic orbits, which do not intersect at increasing values of α. The three stars at $\alpha = 0.89$, $\alpha = 1.04$, $\alpha = 1.82$ are transcritical bifurcations corresponding to T57, T13, and T37 of Figure 5.2, respectively.

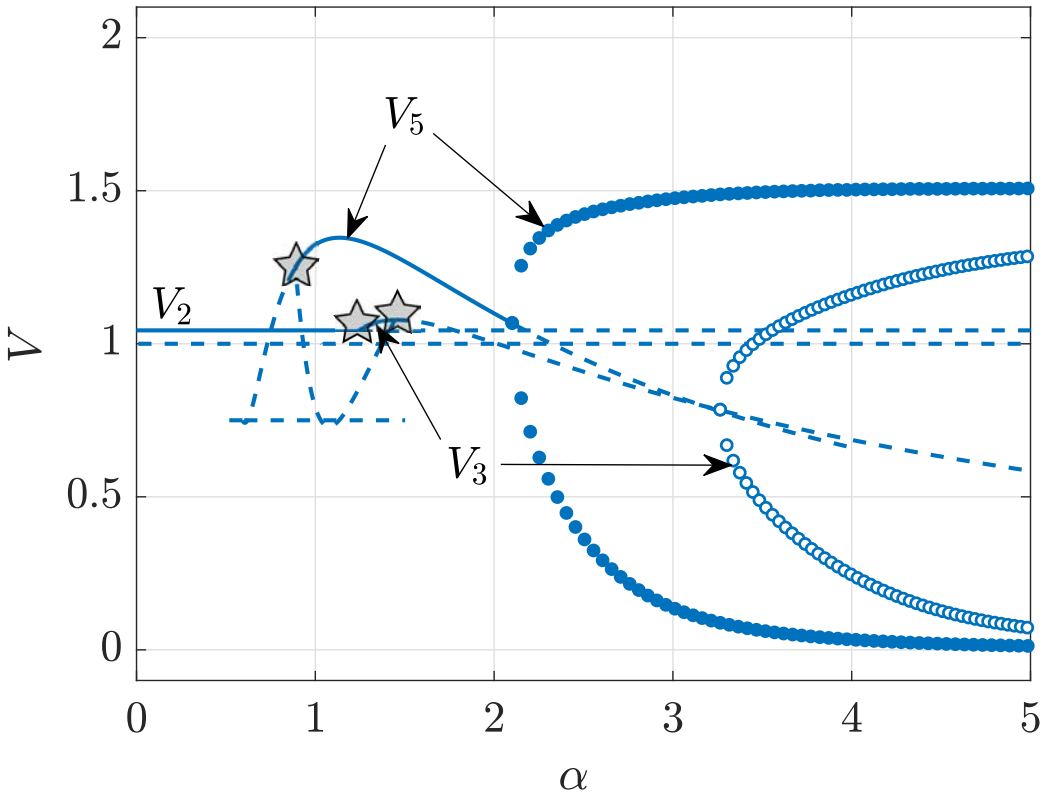


Figure 5.4: One-dimensional bifurcation diagram depicting the norm of biologically meaningful equilibria, V_i for $i = 0, \ldots, 8$, for fixed $\alpha^{\square} = 0.53$. Solid and dashed curves stand for stable and unstable equilibria, respectively. Filled markers and empty markers correspond to stable and unstable periodic orbits. It is worth noting that the periodic orbits do not intersect for larger values of α. The three stars at $\alpha = 0.89$, $\alpha = 1.23$, $\alpha = 1.44$ are transcritical bifurcations corresponding to T57, T23, and T37 of Figure 5.2, respectively.

to the one in Figure 5.3. However, there is one difference: the gen-free-1 state $\boldsymbol{v}_2$ is stable for small values of α, whereas the v-free state $\boldsymbol{v}_1$ is now unstable. With increasing α, the spec-free state $\boldsymbol{v}_5$ also becomes stable at $\alpha_{\boldsymbol{v}_{5,2}}$, which is equal to 0.89 for parameter values fixed at (5.23). This bistability is a peculiar finding which states that at these parameter values, depending on the initial conditions, the system may either lose all generalists or it may lose all specialists. Further, at T23, which corresponds to $\alpha = 1.23$, the bistability shifts towards the stable coex-1 state $\boldsymbol{v}_3$ and the stable spec-free state $\boldsymbol{v}_5$. The bistability interval is limited by the transcritical bifurcation of $\boldsymbol{v}_3$ and $\boldsymbol{v}_7$ at $\alpha_{\boldsymbol{v}_{3,3}}$ and followed by distinctly stable $\boldsymbol{v}_5$. The spec-free state $\boldsymbol{v}_5$ still produces the stable periodic orbit through a supercritical Hopf bifurcation.

Crossing T26 upwards, *i.e.* increasing $\alpha^{\square}$, leads to a gain of stability by the gen-free-2 state $\boldsymbol{v}_6$ and lost of stability by the gen-free-1 state $\boldsymbol{v}_2$. The difference between the two is that gen-free-1 state has only x_1 uninfected cell population to support the existence of viral population while gen-free-2 state allows existence of both x_1 and x_2 in the equilibrium. The reason why an increase in the infection rate of the specialist, $\alpha^{\square}$, from 0.53 to 0.62 leads to appearance of x_2 in the equilibrium state can be explained by recalling that x_1 and x_2 share the same carrying capacity of the system. When specialist gets better at infecting its host cell, that is x_1, the number of uninfected cells of type x_1 decreases which allows x_2 uninfected cell population to survive. That is why T26 can be considered as a critical value of infection rate of the specialist strain which allows the existence of the specialist-insusceptible cell population. In order to check for other significant changes brought by T26 ($\alpha^{\square} = 0.59$), we fix value of $\alpha^{\square} = 0.62$ and vary α. The one-dimensional bifurcation diagram depicting the dynamics of the system is shown in Figure 5.5. For small values of α, the gen-free-2 state $\boldsymbol{v}_6$ is stable. Further, we notice that the stable gen-free-2 state $\boldsymbol{v}_6$ is accompanied by the stable spec-free state $\boldsymbol{v}_5$ after $\alpha = 0.89$ which corresponds to T57. The bistability interval ends at $\alpha = 1.305$ where the gen-free-2 state $\boldsymbol{v}_6$ undergoes a transcritical bifurcation with the coex-2 state $\boldsymbol{v}_7$. Increasing α further, we observe the same outcome as before: the stable spec-free state $\boldsymbol{v}_5$ undergoes a supercritical Hopf bifurcation and produces the stable periodic orbit around itself for large values of α.

The bifurcation analysis will be continued at the value of $\alpha^{\square} = 1$. The one-dimensional bifurcation diagram is shown in Figure 5.6. This cross section has a very rich set of dynamics which are seen more clearly when enlarged in Figure 5.7. At small values of α, none of the equilibria are stable. The similarly small values of α previously (for $\alpha^{\square} = 0.62$, see Figure 5.5) led to the stable gen-free-2 state $\boldsymbol{v}_6$, however now (at $\alpha^{\square} = 1$) the gen-free-2 state is unstable and instead has a stable periodic orbit. This means that somewhere between $\alpha^{\square} = 0.62$ and $\alpha^{\square} = 1$, the gen-free-2 state undergoes a Hopf bifurcation. Following the

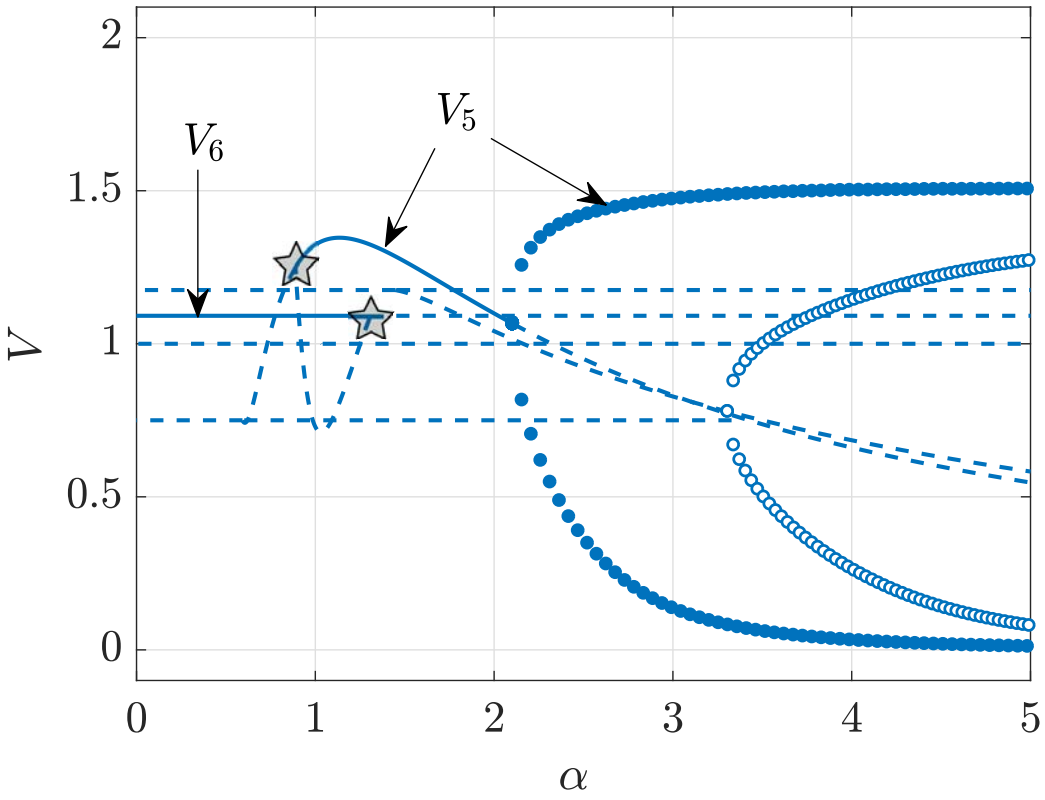


Figure 5.5: One-dimensional bifurcation diagram illustrating the norm of biologically meaningful equilibria, V_i for $i = 0, \ldots, 8$, for fixed $\alpha^{\square} = 0.62$. Solid and dashed curves stand for stable and unstable equilibria, respectively. Filled markers and empty markers stand for stable and unstable periodic orbits, respectively, that do not intersect for larger values of α. Two stars at $\alpha = 0.89$ and $\alpha = 1.305$ are transcritical bifurcations corresponding to T57 and T67 of Figure 5.2, respectively.

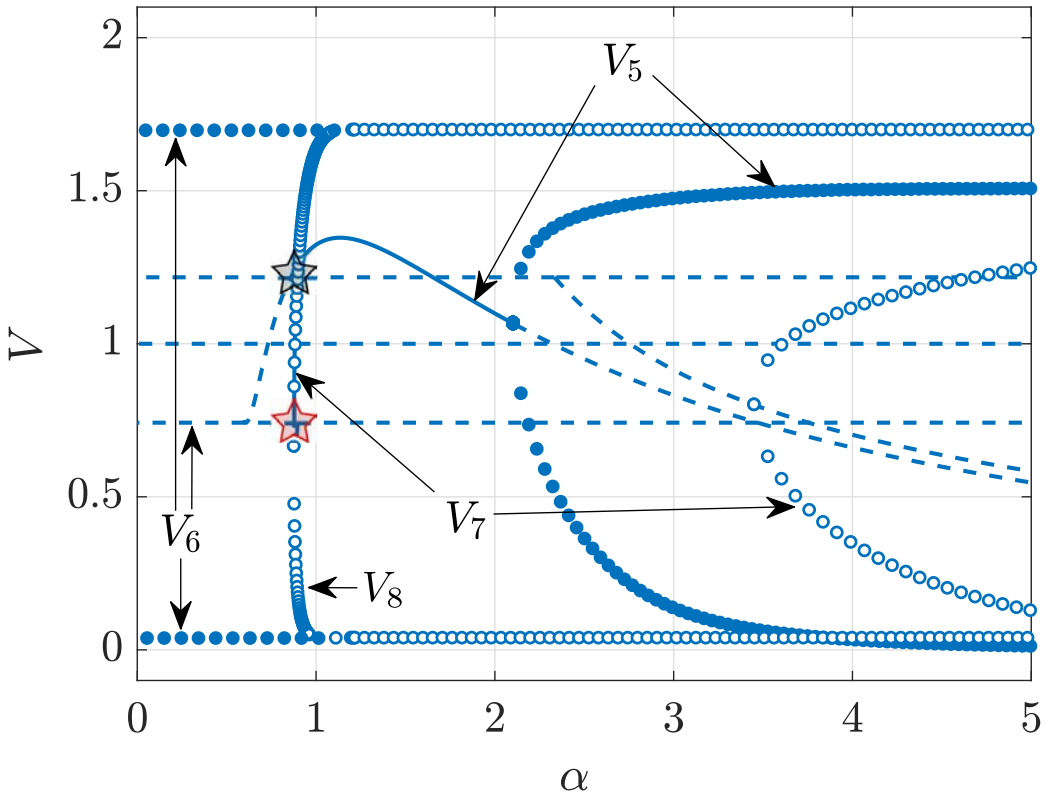


Figure 5.6: One-dimensional bifurcation diagram for the norm of biologically meaningful equilibria, V_i for $i = 0, \ldots, 8$, at fixed $\alpha^{\square} = 1.0$. Solid and dashed curves stand for stable and unstable equilibria, respectively. Filled markers and empty markers stand for stable and unstable periodic orbits, respectively, which do not intersect at increasing values of α. The red star at $\alpha = 0.877$ corresponds to a saddle-node bifurcation of coex-2 and coex-3 states. The black star at $\alpha = 0.89$ is the transcritical bifurcation corresponding to T57.

growth of α in Figure 5.6, we observe a saddle-node bifurcation, which creates coex-2 and coex-3 states at $\alpha = 0.877$, marked with a red star. Here, two branches of equilibria, $\boldsymbol{v}_7$ (stable) and $\boldsymbol{v}_8$ (unstable), appear. The stable branch, $\boldsymbol{v}_7$, undergoes a transcritical bifurcation with the spec-free state $\boldsymbol{v}_5$ at $\alpha = 0.89$ which leads to a stability of $\boldsymbol{v}_5$. The second branch, $\boldsymbol{v}_8$, soon after the saddle-node bifurcation, undergoes a subcritical Hopf bifurcation at $\alpha = 0.879$ which creates an unstable periodic orbits around $\boldsymbol{v}_8$. The unstable equilibrium $\boldsymbol{v}_8$ further intersects $\boldsymbol{v}_6$ where some of its components leave the biologically meaningful variable space. Meanwhile, for all values of α roughly less than 1, the gen-free-2 state $\boldsymbol{v}_6$ has its stable periodic orbit. Therefore, between $\alpha = 0.877$ and $\alpha = 0.89$ there is a bistable region where depending on initial conditions the system may stabilise at either periodic orbits of $\boldsymbol{v}_6$ or the equilibrium $\boldsymbol{v}_7$. Starting at $\alpha = 0.89$ and until the instability of periodic orbits, the bistability region accommodates the equilibrium $\boldsymbol{v}_5$ and the periodic orbit of $\boldsymbol{v}_6$. We notice that the periodic orbit of $\boldsymbol{v}_6$ loses its stability approximately around an intersection point with periodic orbit of $\boldsymbol{v}_8$. This will be discussed in Section 5.4. Further increase of α leads to individually stable spec-free state $\boldsymbol{v}_5$ undergoing the supercritical Hopf bifurcation.

An elaborate set of dynamics is observed when the infection rate of the specialist strain is increased above 1. This case is important from a biological perspective because specialisation is known to be accompanied by the rise of fitness of a population [116, 76]. The one-dimensional bifurcation diagram presented in Figure 5.8 corresponds to a value of $\alpha^{\square} = 2$. As is shown in Figure 5.8, there is neither distinctly stable equilibria nor stable periodic orbits for small values of α. As will be discussed in Section 5.4, a strange chaotic attractor exists in this region. For α approximately between 0.75 and 0.78, stable periodic orbit around the coex-2 state $\boldsymbol{v}_7$ exists, see Figure 5.9. The appearance of the periodic orbits in the one-dimensional bifurcation diagram occurs via a saddle-node bifurcation of periodic orbits (also called a limit point of cycles, fold, or saddle-node of cycles) at $\alpha = 0.757$ and indicated with triangles in Figure 5.9. When increasing α to the critical value of 0.757, two periodic orbits emerge: one stable and one unstable. The stable periodic orbit converges towards a supercritical Hopf bifurcation of the coex-2 state $\boldsymbol{v}_7$, while the unstable periodic orbit approaches the extrema of the periodic orbit around gen-free-2 state $\boldsymbol{v}_6$. The supercritical Hopf bifurcation of $\boldsymbol{v}_7$ leads to this equilibria becoming stable at the point where the stable periodic orbit vanishes. As α is further increased, the stable coex-2 state $\boldsymbol{v}_7$ undergoes the well-known transcritical bifurcation with the spec-free state $\boldsymbol{v}_5$, marked with a black star. For all values of α following $\alpha = 0.89$, the spec-free state $\boldsymbol{v}_5$ and, later, its periodic orbit is stable, see Figure 5.8.

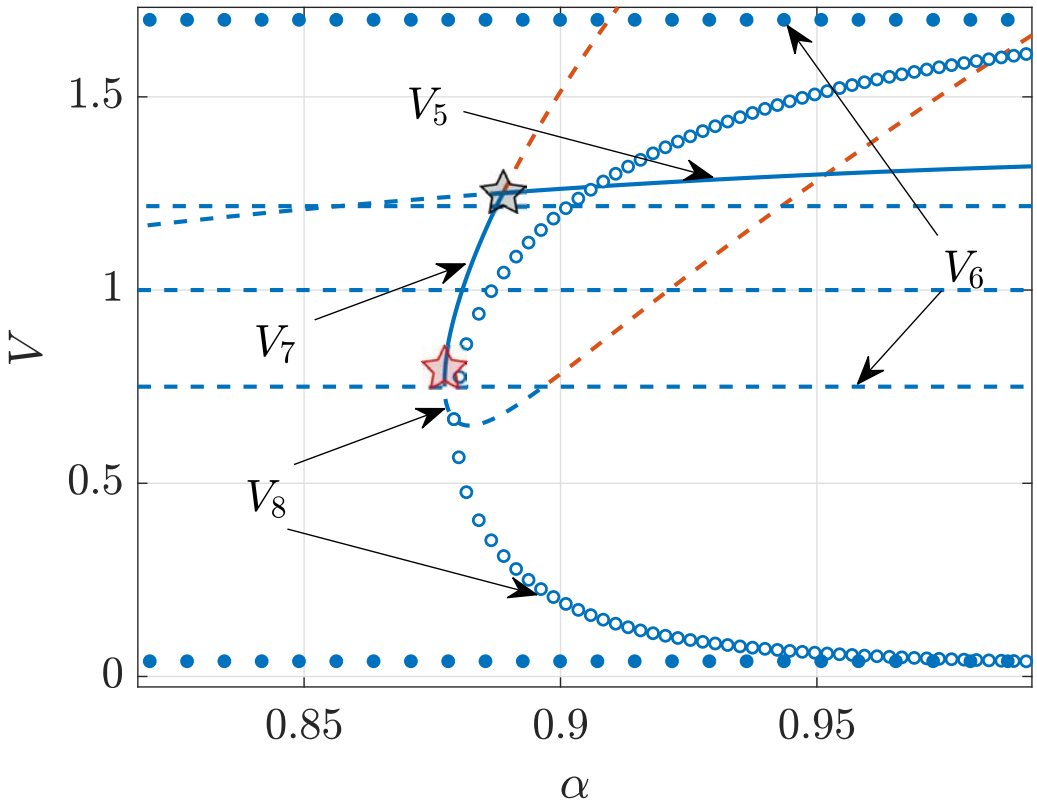


Figure 5.7: Enlarged view of Figure 5.6. The red star indicates the saddle-node bifurcation of coexistence states $\boldsymbol{v}_7$ and $\boldsymbol{v}_8$, while the black star correspond to the transcritical bifurcation of the spec-free state $\boldsymbol{v}_5$ and the coex-2 state $\boldsymbol{v}_7$.

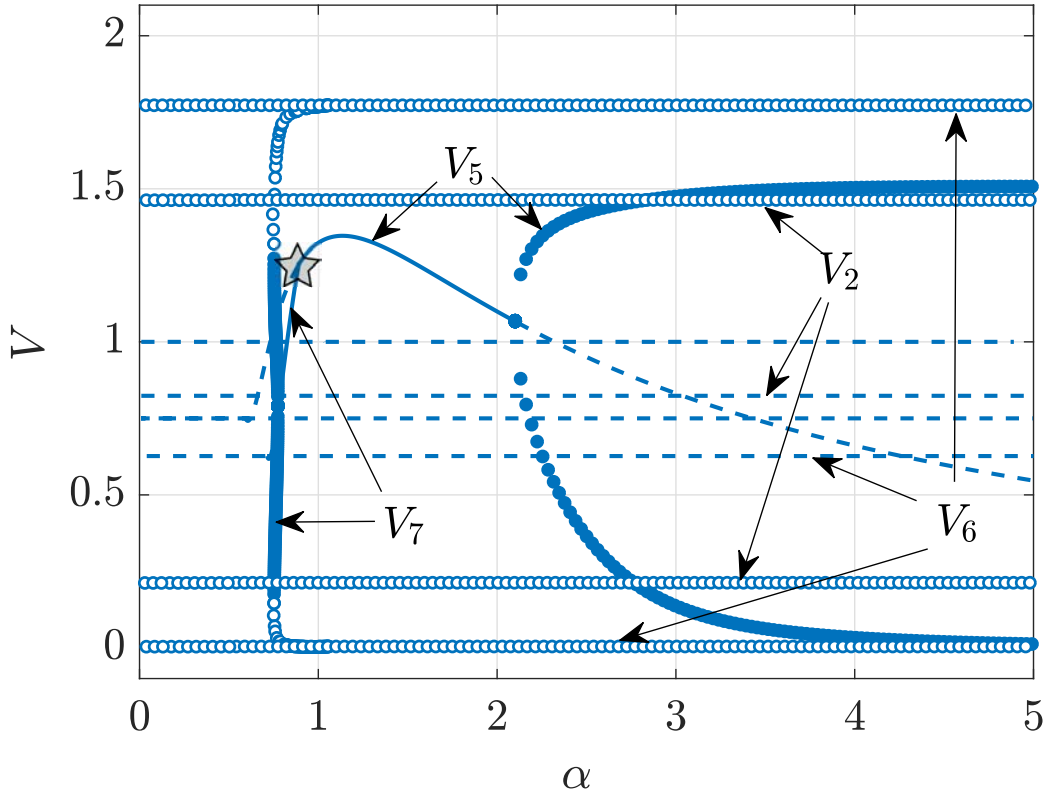


Figure 5.8: One-dimensional bifurcation diagram for a fixed value of $\alpha^{\square} = 2$. Solid curves and filled markers stand for stable equilibria and stable periodic orbits, respectively. Dashed curves and empty markers illustrate unstable equilibria and unstable periodic orbits. The black star at $\alpha = 0.89$ is the transcritical bifurcation of $\boldsymbol{v}_5$ and $\boldsymbol{v}_7$. For enlarged part around $\alpha = 0.89$ see Figure 5.9.

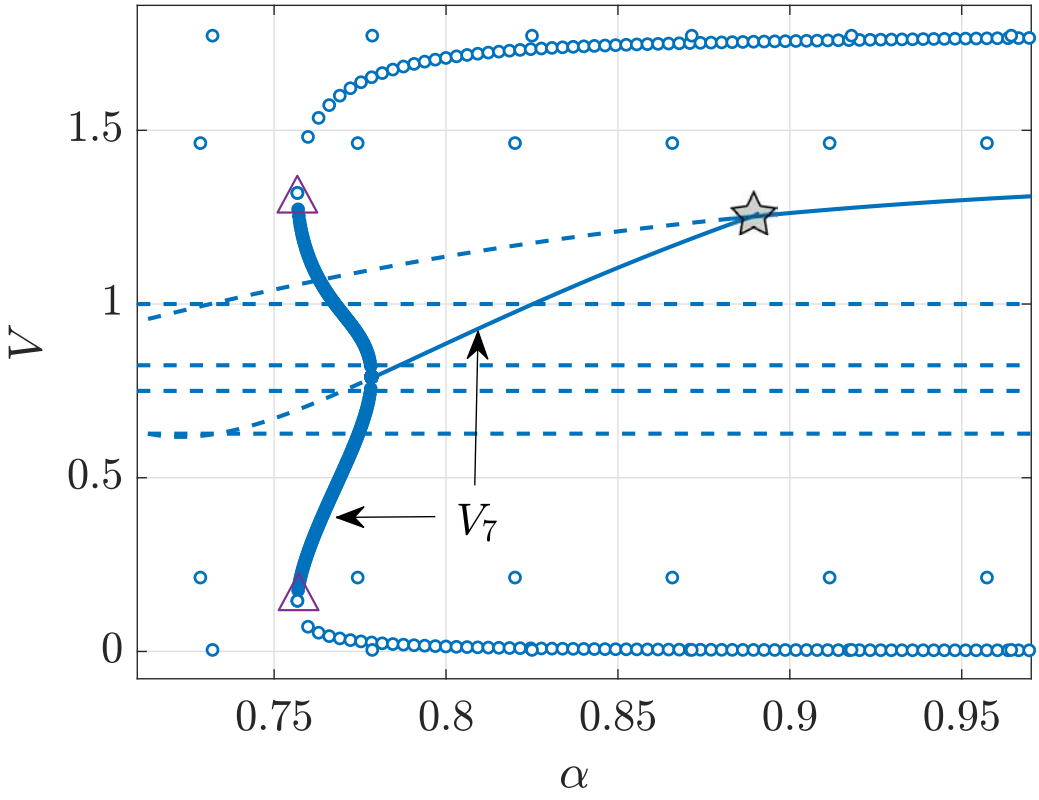


Figure 5.9: Enlarged view of Figure 5.8. Triangles denote the saddle-node bifurcation of periodic orbits of the coex-2 state $\boldsymbol{v}_7$.

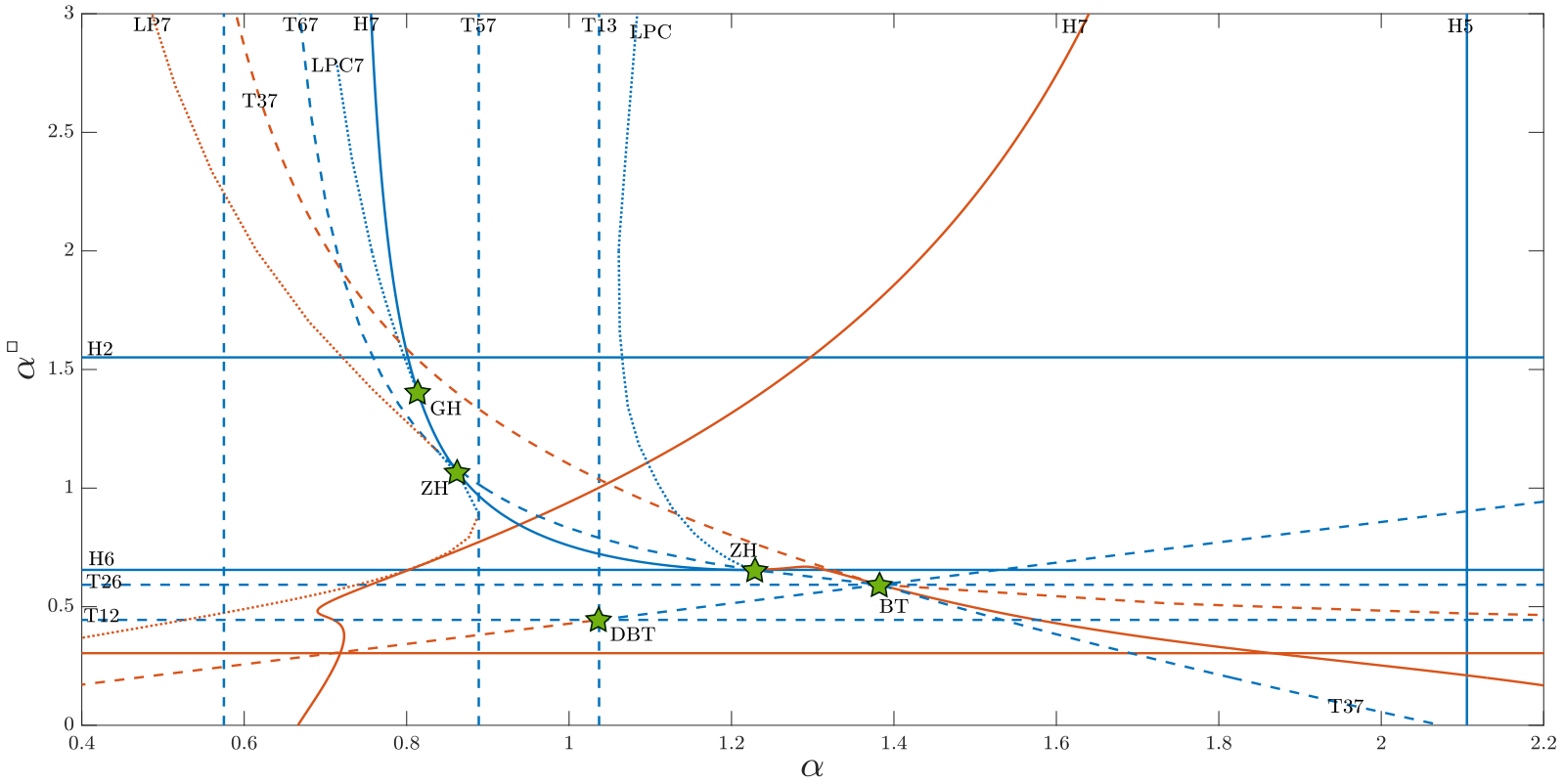


Figure 5.10: Two-dimensional bifurcation diagram for system (5.3) with bifurcation parameters α and $\alpha^{\square}$. Other parameters are fixed as in (5.23). Dashed curves correspond to transcritical bifurcations. Solid curves represent Hopf bifurcations. Dotted curves correspond to limit points of equilibria (LP) and cycles (LPC). Colour blue stands for biologically meaningful values of equilibria and periodic orbits undergoing the bifurcations, while orange curves correspond to bifurcations occurring in negative phase space. A bifurcation marked with Tij is a transcritical bifurcation of $\boldsymbol{v}_i$ and $\boldsymbol{v}_j$ equilibria. Bifurcation curves marked with Hk are curves of Hopf bifurcation of equilibrium $\boldsymbol{v}_k$. LP7 and LPC7 are limit point of $\boldsymbol{v}_7$ and limit point of cycles around $\boldsymbol{v}_7$, respectively. The green stars locate the codimension-two bifurcations such as generalised Hopf bifurcation (GH), zero-Hopf bifurcation (ZH), Bogdanov-Takens bifurcation (BT), and degenerate Bogdanov-Takens bifurcation (DBT).

Information gathered from the bifurcation analysis and the continuation of all bifurcations along both α and $\alpha^{\square}$ allow the construction of a two-dimensional bifurcation diagram for system (5.3), which is shown in Figure 5.10. To clarify the nature of this and following plots, we must recall that there are fourteen parameters and seven independent variables involved in the dynamics of the system. Fixing other parameters than α and $\alpha^{\square}$ is a generous yet necessary sacrifice of generality, which allows us to approach mapping the qualitatively different dynamics on a feasible diagram. The two-dimensional bifurcation diagrams are, in fact, two-dimensional cross sections of the fourteen-dimensional parameter space. There are three main bifurcations considered on the plot in Figure 5.10: the transcritical, the Hopf, and the limit point (of cycles and equilibria). The transcritical bifurcation curves are shown in dashed, and as before, have a label, Tij, which shows that the bifurcation affects equilibria $\boldsymbol{v}_i$ and $\boldsymbol{v}_j$. Bold, solid curves stand for Hopf bifurcations and have labels, Hk, depicting equilibria $\boldsymbol{v}_k$ undergoing the bifurcation. The thin dotted curve labelled LP7 is a theoretical threshold of coex-2 and coex-3 states since it is a saddle-node bifurcation of $\boldsymbol{v}_7$ and $\boldsymbol{v}_8$. The thin dotted curve labelled LPC7 is a limit point of cycles that appear around $\boldsymbol{v}_7$. In Figure 5.10, as before, blue depicts biologically relevant values of variable components of all equilibria involved in a bifurcation. Note that there is a fair resemblance with bifurcation diagram of (4.2), in fact all behavioural traits which are seen in (4.2) model can be found in (5.3).

One interesting feature brought upon the dynamics by the introduction of the second, generalist-only susceptible cell is bistability. Regions bounded by the bifurcation curves may contain more than one stable equilibria and periodic orbit. With this result, the dependency of the long-term behaviour of the system on initial conditions must be considered, and usually basins of attractions are studied. The full basins of attraction for the considered model must be plotted on a seven-dimensional phase space. Besides, for the classical construction of basins of attraction, a seven-dimensional hypercube of initial conditions must be studied. For the purpose of this discussion, instead of computing classical basins of attraction, we take a probabilistic approach to study the dynamics over two-dimensional parameter space. We consider initial conditions for susceptible cell populations and infected cell populations that remain constant and alter initial values of viral loads. Initial values of uninfected cell populations are fixed to be at maximum of carrying capacity of the system, that is $x_1(0) = 1.0$ and $x_2(0) = 0.75$. Infected cell populations are nonexistent at time zero. The viral strain populations z and $z^{\square}$ are taken from a certain grid from 0 to 1 each. The probability of becoming attracted towards each equilibrium is mapped in Figure 5.11 with different colours. Additional to the mapped stability of equilibria as $p(\boldsymbol{v}_i)$, there is

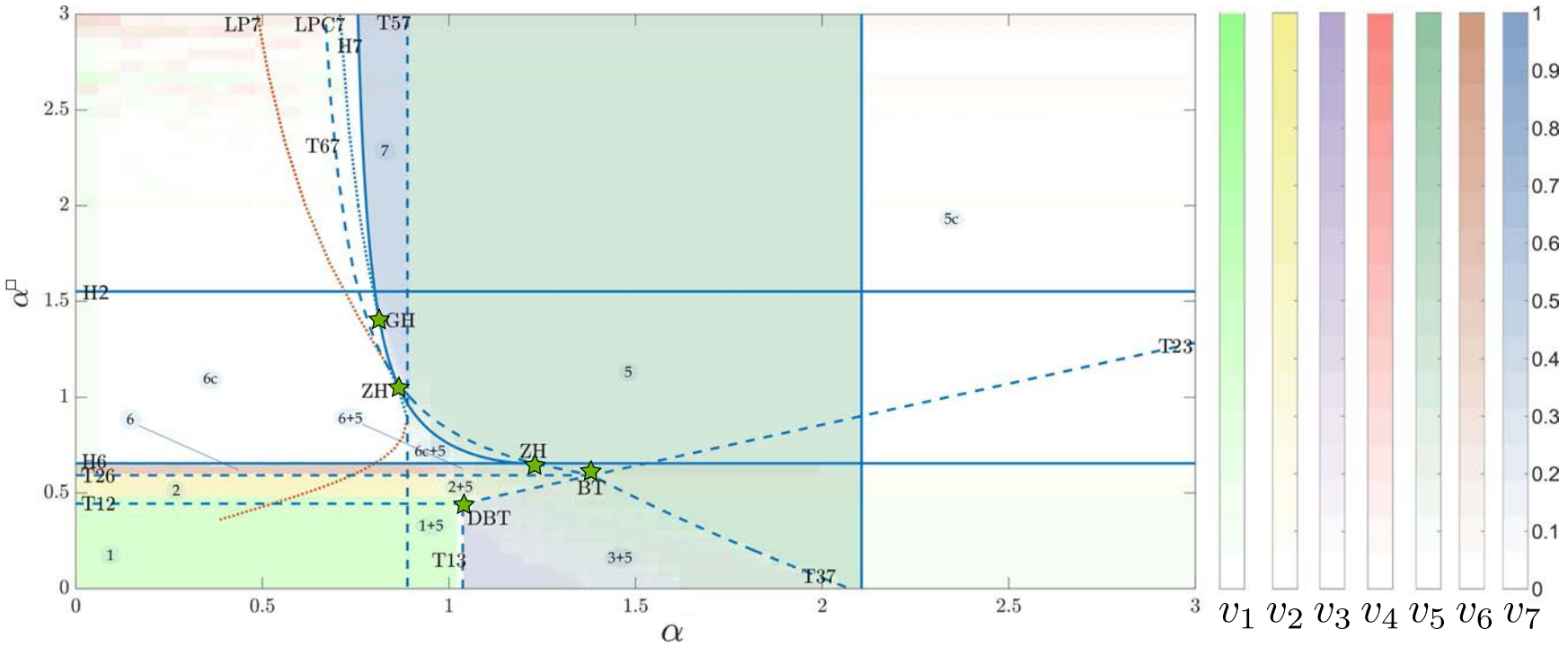


Figure 5.11: Stability map for equilibria $\boldsymbol{v}_i$, $i = 0, \ldots, 7$. Colours of regions illustrate the probability of the system stabilising at the given equilibria when started from different initial conditions (only $z^{\square}(0)$ and $z(0)$ were sampled). White areas correspond to stability of periodic orbits or chaos.

a non-periodic (*e.g.* chaotic) behaviour whose probability is considered as $1 - p(\boldsymbol{v}_i)$ for appropriate $i = 1, \ldots, 8$ and is left in Figure 5.11 white for simplicity reasons.

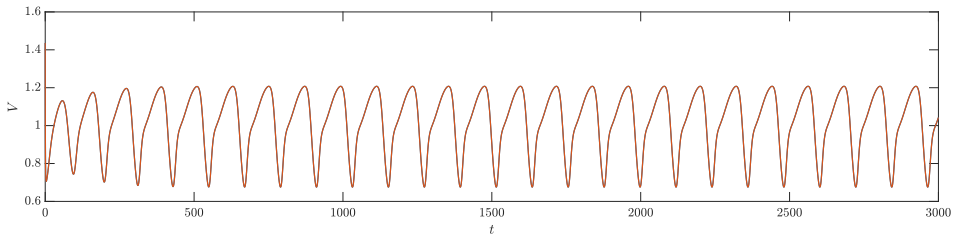
In Figure 5.11, region (1) contains values of infection rates, α and $\alpha^{\square}$, that ensure stability of the v-free state $\boldsymbol{v}_1$ for all possible initial conditions. The system does not allow persistence of the virus strains since neither of them is fit enough to survive. Upon increasing the infection rate of the specialist strain, $\alpha^{\square}$, the behaviour of the system from region (1) shifts into regions (2) and (6) for smaller values of infection rate of the generalist strain, and into regions (2+5) and (6+5) for slightly larger α. Both (2) and (6) accommodate the equilibria corresponding to scenarios where the specialist persists and the generalist dies out, *i.e.* the gen-free-1 state $\boldsymbol{v}_2$ and the gen-free-2 state $\boldsymbol{v}_6$. The threshold for survival of the specialist was found and lays between regions (1) and (2). The bistability regions (2+5) and (6+5) provide almost equal chances of system tending towards $\boldsymbol{v}_2$ or $\boldsymbol{v}_5$ and $\boldsymbol{v}_6$ or $\boldsymbol{v}_5$, respectively. The combination of (2+5) and (6+5), forming a comparably thin, hexagonal shaped region, can be of great epidemiological interest. These regions provide almost equal chances for the survival of a solitary specialist population with generalist population going extinct and, *vice versa*, the survival of the generalist strain and elimination of the specialist population. By examining Figure 5.11, it may seem obvious that in regions (2+5) and (6+5), the spec-free state $\boldsymbol{v}_5$ has higher probability of becoming the main equilibria on narrow, right side of the hexagon, that corresponds to higher infection rates of the generalist. On the contrary, the left side of the hexagon, with lower values of α, provides higher chances for survival of gen-free-1 and gen-free-2 states. One of the boundaries of the hexagon is defined by a line of transcritical bifurcations which has DBT and BT bifurcations at the ends of the segment. This bifurcation exchanges stability between $\boldsymbol{v}_2$ and $\boldsymbol{v}_3$, and leads to a larger hexagon of region (3+5). In region (3+5) the coex-1 state $\boldsymbol{v}_3$ and the spec-free state $\boldsymbol{v}_5$ share chances of becoming the main stable equilibria depending on initial conditions. Only the lower left corner of region (3+5) guarantees the stability of $\boldsymbol{v}_3$ for all initial conditions. For increasing values of both α and $\alpha^{\square}$, the probability of coexistence of both strains decreases and the probability of the solitary survival of the generalist population increases. The latter probability reaches 100% at the transcritical bifurcation between $\boldsymbol{v}_3$ and $\boldsymbol{v}_7$, where coex-1 state $\boldsymbol{v}_3$ loses its stability to biologically meaningless coex-2 state $\boldsymbol{v}_7$. For values of α and $\alpha^{\square}$ above the transcritical bifurcation T37 lays the region (5). The spec-free state $\boldsymbol{v}_5$ occupies a large region which depends strongly on the infection rate of the generalist strain. The maximum value of α for region (5) is defined by the Hopf bifurcation of $\boldsymbol{v}_5$. For values greater than the Hopf bifurcation value of α, in region (5c), the periodic orbit around $\boldsymbol{v}_5$ is stable for all values of $\alpha^{\square}$. The minimum value of α for region (5) is at the transcritical bifurcation with the coex-2 state $\boldsymbol{v}_7$. The transcritical bifurcation between $\boldsymbol{v}_5$

and $\boldsymbol{v}_7$ separates regions (5) and (7). The thin, vertical region (7) holds α and $\alpha^{\square}$ that lead to the stability of the coex-2 state for all initial conditions. Curiously, regions of coexistence occupy comparably small parts of the stability map. The first example of a region with stable periodic orbits (5c) was discussed above. The second example is located at values of $\alpha^{\square}$ greater than Hopf bifurcation of $\boldsymbol{v}_6$, the H6 line, and for values of α less than of the curve LP7. There, a periodic orbit around $\boldsymbol{v}_6$ is stable in all of the region (6c) for any nontrivial initial conditions. The region (6c) *may* exceed over H2 line. The region above Hopf bifurcation of $\boldsymbol{v}_2$, the H2 line, for smaller values of α has low yet positive probability of hosting stable $\boldsymbol{v}_1$ and $\boldsymbol{v}_4$ in a seemingly random manner. The dynamics of this region is discussed in Section 5.4. Interestingly, $\boldsymbol{v}_8$ is not stable for any pair of (α, $\alpha^{\square}$).

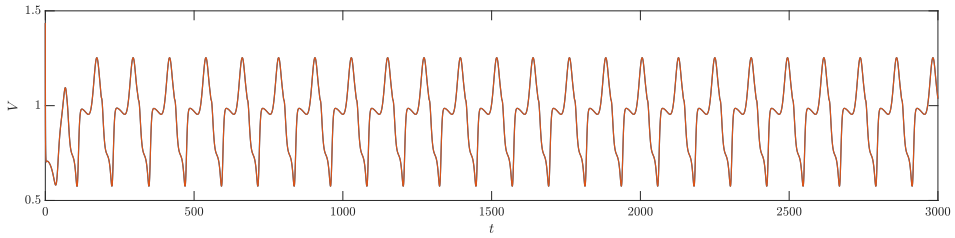
5.4 Chaos in the model of virus specialisation

The dynamics of the system in particular areas of the stability map are uncertain, *e.g.* for small values of α in Figure 5.8 where none of the equilibria or periodic orbits are stable. From the results of taking horizontal cross sections for one-dimensional bifurcation diagrams, we can conclude there is an absence of stable objects for low infection rates of the generalist, α, and high infection rates of the specialist, $\alpha^{\square}$. Vertical cross-sections reveal loss of stability by all equilibria and periodic orbits in the region, too. In order to understand the dynamics of the system for small α and large $\alpha^{\square}$, we compute the time series for a fixed value of $\alpha = 0.5$ and four, gradually increasing, values of $\alpha^{\square}$ at 0.7, 1.2, 1.6, and $\alpha^{\square} = 2$, the results of which are shown in Figure 5.12. Each plot illustrates the time series starting at two, biologically meaningful initial conditions where the second initial condition is a very small perturbation of the first initial condition. At $\alpha^{\square} = 0.7$, the system corresponds to region (6c) of Figure 5.11. These parameters lead to a stable periodic orbit around the gen-free-2 state $\boldsymbol{v}_6$, see Figure 5.12 (a). At $\alpha^{\square} = 1.2$, the periodic orbit shows irregularity which is evident on its emerging second peaks, and at $\alpha^{\square} = 1.6$ a period doubling becomes conspicuous, see Figure 5.12 (b) and (c). Further at $\alpha^{\square} = 2.0$, two major changes are observed: first, a hypothesised cascade of period doubling is affecting the periodic orbit in an intricate way leading to aperiodic cycles, while, second, close initial conditions of two simulations are leading to cycles of dissimilar shape and period starting at the time horizon of $t = 1500$, see Figure 5.12 (d). The obtained numerical results thus seem to indicate the presence of chaotic dynamics, which emerge for high values of the infection rate of the specialist and comparably low infection rate of the generalist.

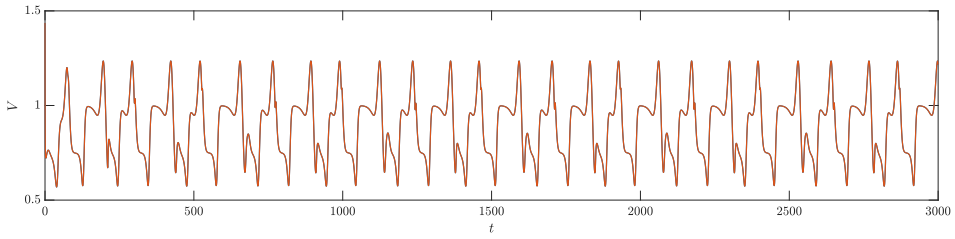
To study chaos in more detail, we calculate Lyapunov exponents. Assume, that first and second initial conditions provide $\boldsymbol{v}_{\text{i.c.}1}(t)$ and $\boldsymbol{v}_{\text{i.c.}2}(t)$, respectively, and in Figure 5.12 (d),



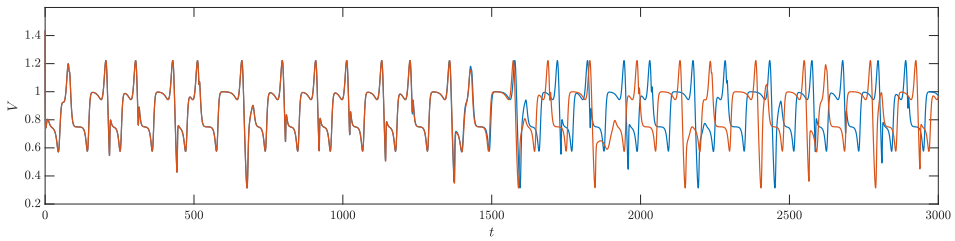
(a) $\alpha^{\square} = 0.7$



(b) $\alpha^{\square} = 1.2$



(c) $\alpha^{\square} = 1.6$



(d) $\alpha^{\square} = 2$

Figure 5.12: Time series of system (5.3) for four different values of $\alpha^{\square}$. Infection rate $\alpha = 0.5$ and other parameters are fixed as (5.23). Initial conditions used for simulations: $\{x_1(0) = 1, x_2(0) = 0.75, y_1^{\square}(0) = 0, y_1(0) = 0, y_2(0) = 0, z^{\square}(0) = 0.5, z(0) = 0.5\}$ and $\{x_1(0) = 1, x_2(0) = 0.75, y_1^{\square}(0) = 0, y_1(0) = 0, y_2(0) = 0, z^{\square}(0) = 0.5001, z(0) = 0.5001\}$.

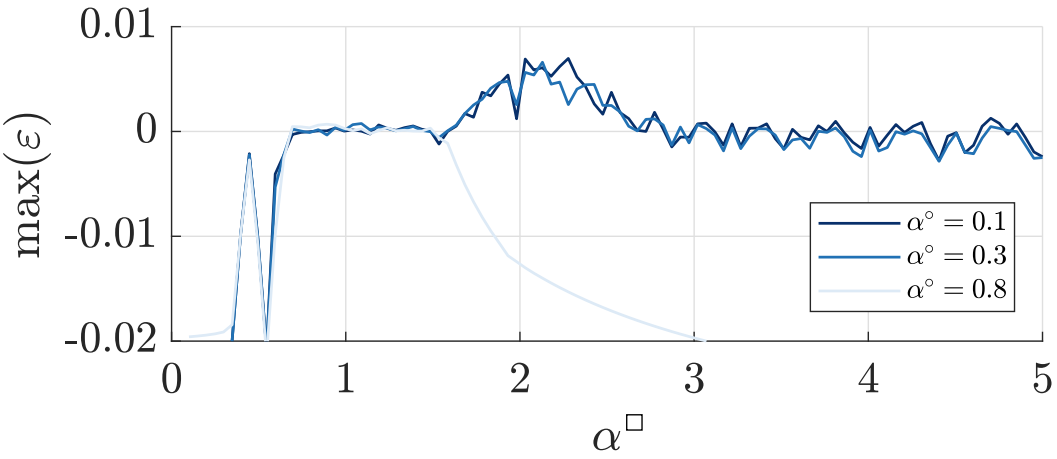


Figure 5.13: Maximal Lyapunov exponents for the dynamical system (5.3) at different infection rates. Parameters of the system are fixed as in (5.23). All simulations have the same time length, $t = 5000$. Chaotic behaviour of the system corresponds to approximate values of infection rates $1.5 \leq \alpha^{\square} \leq 3$.

they diverge as $|\boldsymbol{v}_{\text{i.c.}1}(t) - \boldsymbol{v}_{\text{i.c.}2}(t)| \approx e^{\max(\boldsymbol{\varepsilon})t}|\boldsymbol{v}_{\text{i.c.}1}(0) - \boldsymbol{v}_{\text{i.c.}2}(0)|$, where $\boldsymbol{\varepsilon}$ is a vector of Lyapunov exponents and thus has the same dimension as the vector $\boldsymbol{v}(t)$. The measure of the diverging solutions $\boldsymbol{\varepsilon}$ from a very close initial conditions is used to quantitatively illustrate the existence of chaos. In our notations, when any component of the vector $\boldsymbol{\varepsilon}$ is positive, the solutions $\boldsymbol{v}_{\text{i.c.}1}(t)$ and $\boldsymbol{v}_{\text{i.c.}2}(t)$ diverge exponentially. Finding positive Lyapunov exponents for each parameter combination provides approximate boundaries of the parameter space which lead the system to chaotic behaviour. First, we fix the infection rate of the generalist strain at a certain value, and then calculate all Lyapunov exponents for infection rates of the specialist for 100 values of $\alpha^{\square} \in (0, 5)$. By changing the value of the fixed infection rate of the generalist and repeating the procedure, we obtain and plot the maximal Lyapunov exponent in Figure 5.13. As shown in the plot, three values of infection rate of the generalist are chosen to be $\alpha = 0.1$, $\alpha = 0.3$, $\alpha = 0.8$ and different colours correspond to each value. Stochasticity along the trajectories of maximal Lyapunov exponents is the result of variation of the strength of the exponential divergence. For $\alpha = 0.1$ and $\alpha = 0.3$ the calculations show a range of $\alpha^{\square}$ that lead to positive maximal Lyapunov exponents. However, the light blue trajectory of $\alpha = 0.8$ leads to stable non-chaotic behaviour in the system, since the largest Lyapunov exponents are non-positive. The subintervals of both α and $\alpha^{\square}$ that lead to existence of positive Lyapunov exponents are approximate boundaries of infection rates for chaos in the system. In order to determine the chaotic areas, we map the positive values of maximal Lyapunov exponents along different values of both α and $\alpha^{\square}$, see Figure 5.14. Interestingly, the certain positive values are present only for $0 < \alpha \leq 0.7$ and $1.5 \leq \alpha^{\square} \leq 3$,

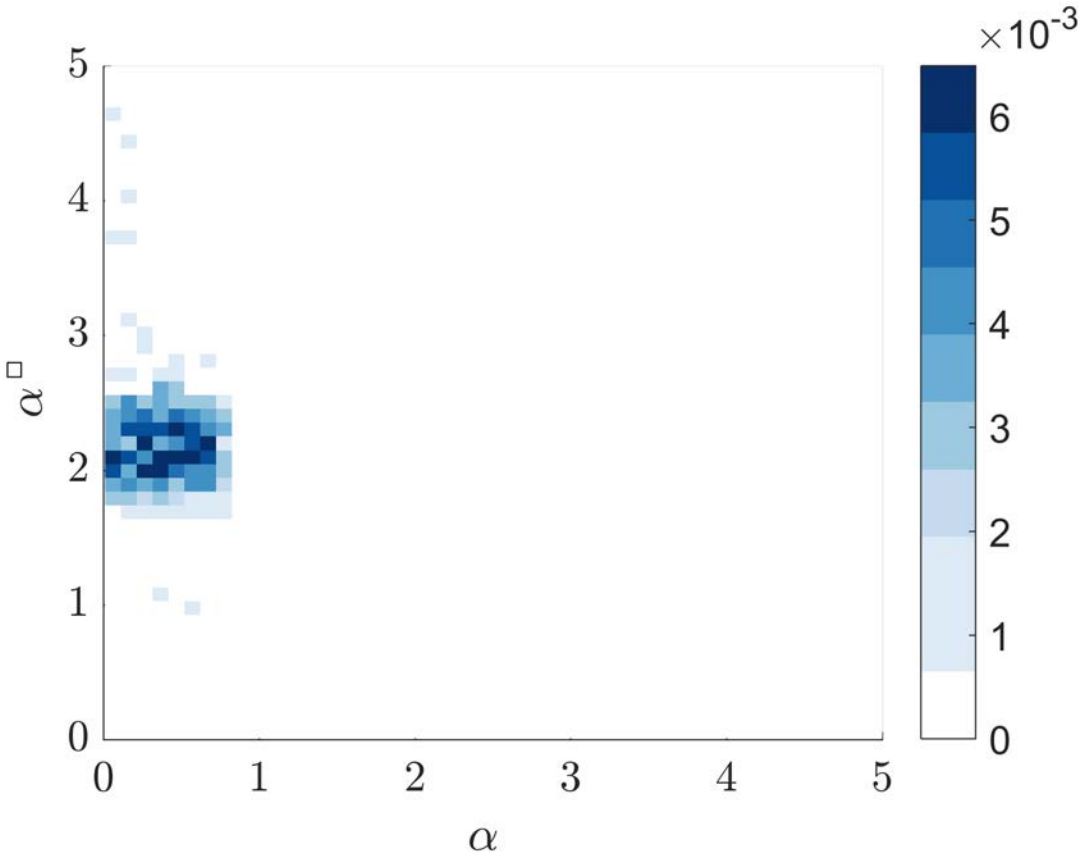


Figure 5.14: Values of the maximal Lyapunov exponents (MLE) for the dynamical system (5.3) at different infection rates are depicted with colour. Parameters of the system are fixed as in (5.23). All simulations have the same time length, $t = 5000$. Chaotic behaviour of the system corresponds to only positive values of the MLE, which are depicted on this plot for approximate values of infection rates $0 < \alpha \leq 1$ and $1.5 \leq \alpha^{\square} \leq 3$. In the next figure we display a zoom inside this region.

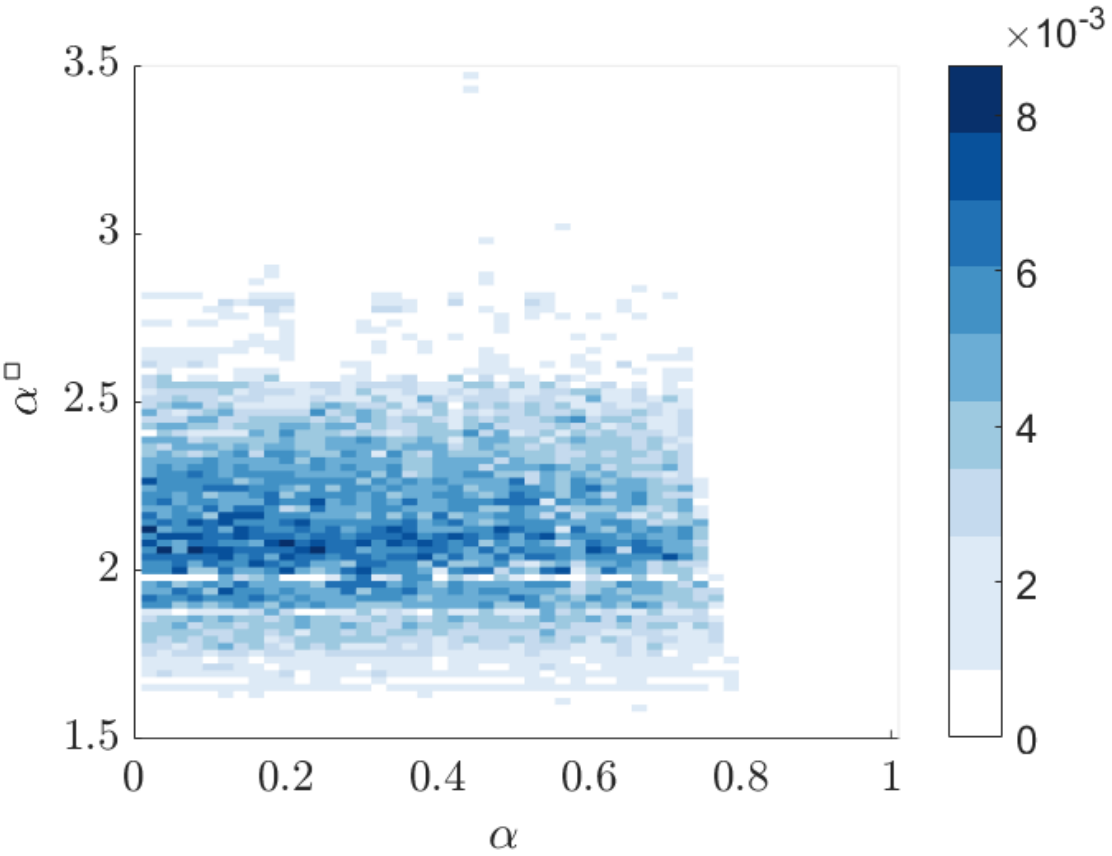
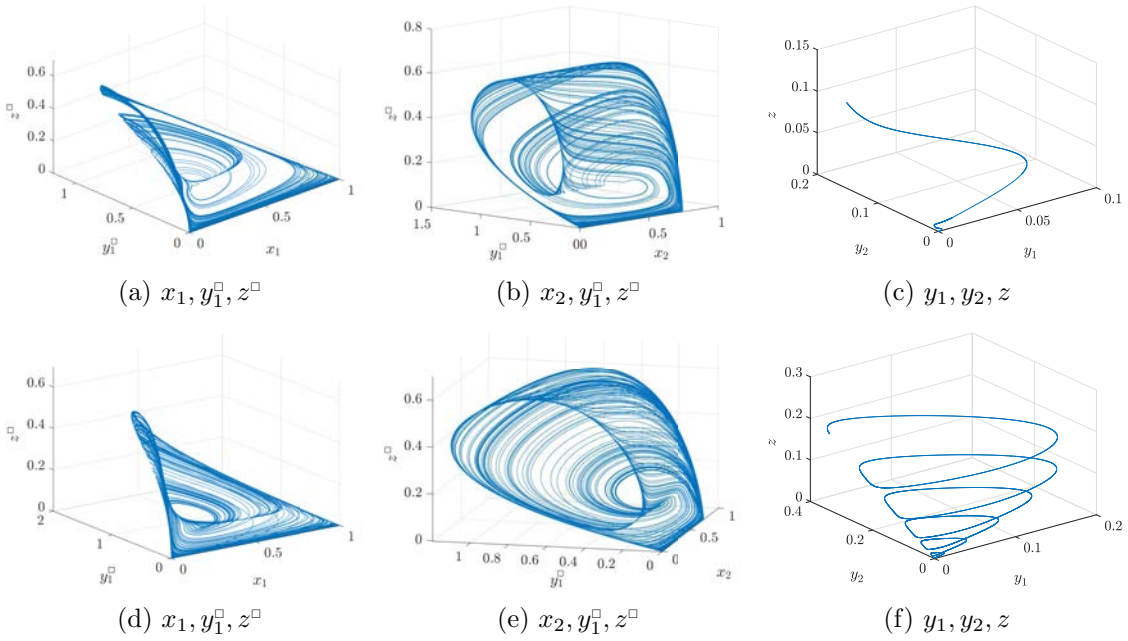


Figure 5.15: Enlarged part of maximal Lyapunov exponents for the dynamical system (5.3). Parameters of the system are fixed as in (5.23). All simulations have the same time length, $t = 5000$.

which can be considered *a chaotic window* which is followed by an existence of a periodic behaviour (see Figure 5.15). Since the Lyapunov exponents are trustworthy only as $t \longrightarrow \infty$, we repeated the calculations for $\alpha^{\square} > 3$ fixing larger time intervals for simulations, which changed the results of Figure 5.14 only negligibly. The strange chaotic attractor is shown in several projections of the phase space in Figure 5.16.

The system of equations (5.3) exhibits the most diverse behavioural traits studied so far in this work. Nevertheless, the results of this chapter are consistent with the previous findings. For small values of infection rates, the infection-free states are dominant. This agrees with the understanding of the viral infectiousness and basic reproductive number: until the infection rates increase the basic reproductive number above one, the virus-free state remains stable. As soon as the infection rate of a generalist or a specialist crosses the critical value, the corresponding population of the generalist or the specialist prevails. The generalist-specialist model is similar to the wild-type and mutant model of Chapter 4: the main similarity being the assumption of directional mutation from the wild-type to the



(a) $x_1, y_1^{\square}, z^{\square}$ (b) $x_2, y_1^{\square}, z^{\square}$ (c) y_1, y_2, z

(d) $x_1, y_1^{\square}, z^{\square}$ (e) $x_2, y_1^{\square}, z^{\square}$ (f) y_1, y_2, z

Figure 5.16: Three-dimensional projections of the attractors for system (5.3) at parameter values fixed as in (5.23). First row, (a), (b), (c), correspond to $\alpha = 0.5$, $\alpha^{\square} = 2.0$. Second row, (d), (e), (f), correspond to $\alpha = 0.7$, $\alpha^{\square} = 2.2$. Notice that strange attractors are found in (a), (b), (d), and (e).

mutant strain and from the generalist to the specialist strain. The key difference between the two models is that the introduction of the generalist requires the addition of the second type of a target cell so that a diversity of resources can be available. Recalling the results of the model of the wild-type and mutant strains sharing the same target cell population, we know that the mutant strain remains stable or coexists with the wild-type strain whenever the basic reproductive ratio for the mutant and wild-type strains exceeds one. Drawing the parallel of those results with the outcome of current generalist and specialist model, we see the notable effects of the introduction of the second target cell type. The existence of the second cell population that is susceptible to the generalist type, as the study have shown, allows the stable and lone existence of the generalist. There are still regions of coexistence on the stability map, however in comparison to the results of the model for wild-type and mutant strains, the generalist can exists without the specialist despite the mutational "leak" of the population.

Part III

Specialisation of viral strains

Chapter 6

Modelling viral diversity

Viral diversity is known to be a key obstacle in the way of curing and finally eradicating diseases caused by almost all viruses [134, 105], for example, HIV [77] and Hepatitis C virus [106]. Understanding the mechanisms driving viral heterogeneity is crucial for determining effective ways of preventing outbreaks. From a mathematical perspective, there are many ways to model and analyse the patterns of genetic diversity observed in viral populations. In this chapter, we consider a method of modelling the viral diversity in scope of the population dynamics, where $m+1$ different variants are considered. The population of different viral strains is placed in an environment where a population of susceptible cells is provided. In this framework, strains differ from one another by their infectiousness. The goals of this chapter are (i) generalise the previously developed model of two strains to $m+1$ strains, (ii) derive a model based on partial integro-differential equations by taking the continuum limit as $m \to \infty$, and (iii) compare the dynamics of the derived model, which describes an infection-supplied evolution process of the heterogeneous viral population, with the dynamics of the model with two viral strains.

6.1 Mutation as a diffusion process in phenotypic space

First, recall the nondimensional model of symmetrical mutation of two different strains and one shared host considered in Chapter 4, and given by:

$$\begin{aligned}
\dot{x} &= x(1-x) - \alpha^{\square}\, z^{\square}\, x - \alpha\, z\, x, \\
\dot{y}^{\square} &= \alpha^{\square}\, z^{\square}\, x + \mu y - \mu y^{\square} - \gamma^{\square}\, y^{\square}, \\
\dot{y} &= \alpha\, z\, x + \mu y^{\square} - \mu y - \gamma\, y, \\
\dot{z}^{\square} &= \gamma^{\square}\, y^{\square} - \nu^{\square}\, \alpha^{\square}\, z^{\square}\, x - \zeta^{\square}\, z^{\square}, \\
\dot{z} &= \kappa\gamma\, y - \nu\, \alpha\, z\, x - \zeta\, z.
\end{aligned} \tag{6.1}$$

This nondimensionalized model was extensively studied in Section 4.2. Here, the mutation of strains $z^{\square}$ and z is modelled via flow in infected cells $y^{\square}$ and y with rate μ. For greater number of strains $m \gg 1$, a similar technique can be used. We assume that a population of virus which consists of different strains is diverse enough for the different strains to be distributed along some characteristic axis. Each value on this characteristic axis identifies a specific strain of virus. Allowing the sequencing of different viral strains over an axis or interval is a very generous assumption, however if the property has a measurable nature, then sorting a physical measurement along an interval becomes a task of mapping from large yet finite subset of $\mathbb{R}$ to $[0,1] \in \mathbb{R}$. By increasing the number of measurements describing the strains, the interval will appear more and more dense. Without loss of generality and for clarity in the proceeding discussion, we call the interval enabling the identification of all strains a phenotype space. For example, consider the envelope glycoprotein (Env) of HIV-1, whose gene product consists of a complex of two subunits, gp120 and gp41. This protein is incorporated into virions that bud to the cell surface of infected T cells. The Env is crucial to viral infectivity. The density of Env on virions can be measured by immunofluorescence microscopy and subcellular membrane fractionation techniques [90]. Different infectivity may be considered as a characteristic which differs one strain from another and thus the density of mature gp120-gp41 products can be mapped along a unit interval $S = [0,1]$.

Having defined the phenotype space, we are now able to incorporate mutation into a general model of $m+1$ strains. On this phenotype space the population of infected cells, $y(s,t)$, and phenotype dependent viral load, $z(s,t)$, can be defined. This is a generalisation of the two-strain case given by (6.1): the two types of viral strains $z(t)$ and $z^{\square}(t)$ now correspond to values of $z(s,t)$ at two specific points $s, s^{\square} \in S$, that is $z = z(s,t)$ and $z^{\square} = z(s^{\square}, t)$. Figure 6.1 illustrates a scheme by which the mutation of the strains is

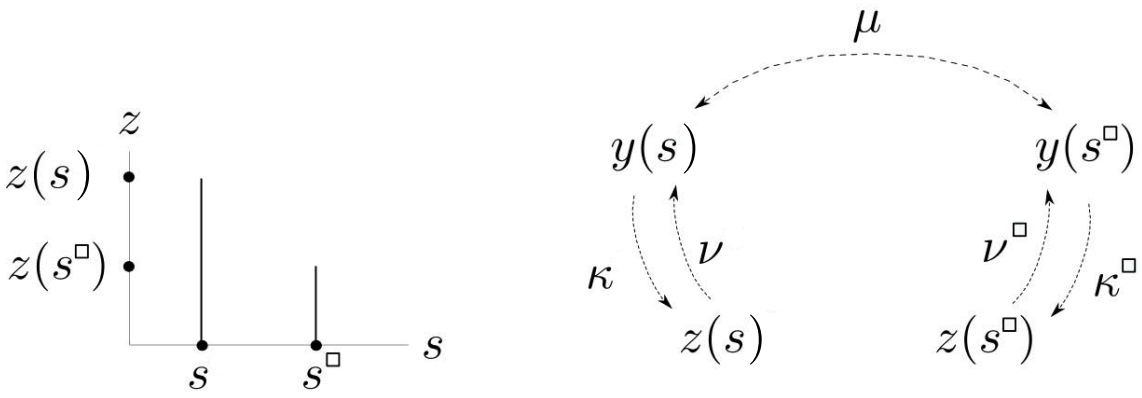


Figure 6.1: Scheme of a model with two phenotypes of virus $z(s)$ and $z(s^{\square})$. The mutation of strains into each other occurs within the infected cells $y(s)$ and $y(s^{\square})$.

modelled in (6.1). In this context, the mutation for each of the strains is proportional to the difference in infected cell populations $y(s)$ and $y(s^{\square})$:

$$\begin{aligned} \Delta y(s^{\square}) &\approx y(s) - y(s^{\square}), \\ \Delta y(s) &\approx y(s^{\square}) - y(s). \end{aligned} \tag{6.2}$$

The difference arises from a conservation argument: the change in y is due to flow in minus flow out; the flow (or mutation) rate is μ for both strains. The addition of a third phenotype, $s^{\star} \in S$, follows a similar rule, see Figure 6.2. The mutation rate of every phenotype is proportional to the change in sizes of infected cell populations $y(s)$, $y(s^{\square})$, and $y(s^{\star})$:

$$\begin{aligned} \Delta y(s) &\approx y(s^{\star}) - 2y(s) + y(s^{\square}), \\ \Delta y(s^{\square}) &\approx y(s) - 2y(s^{\square}) + y(s^{\star}), \\ \Delta y(s^{\star}) &\approx y(s) - 2y(s^{\star}) + y(s^{\square}), \end{aligned} \tag{6.3}$$

that obey the conservation principle for the total size of infected cell populations. That is, $\Delta y(s) + \Delta y(s^{\square}) + \Delta y(s^{\star}) = 0$, so populations are conserved by mutation. Increasing the number of strains considered in the model up to $m+1$ provides the following finite system

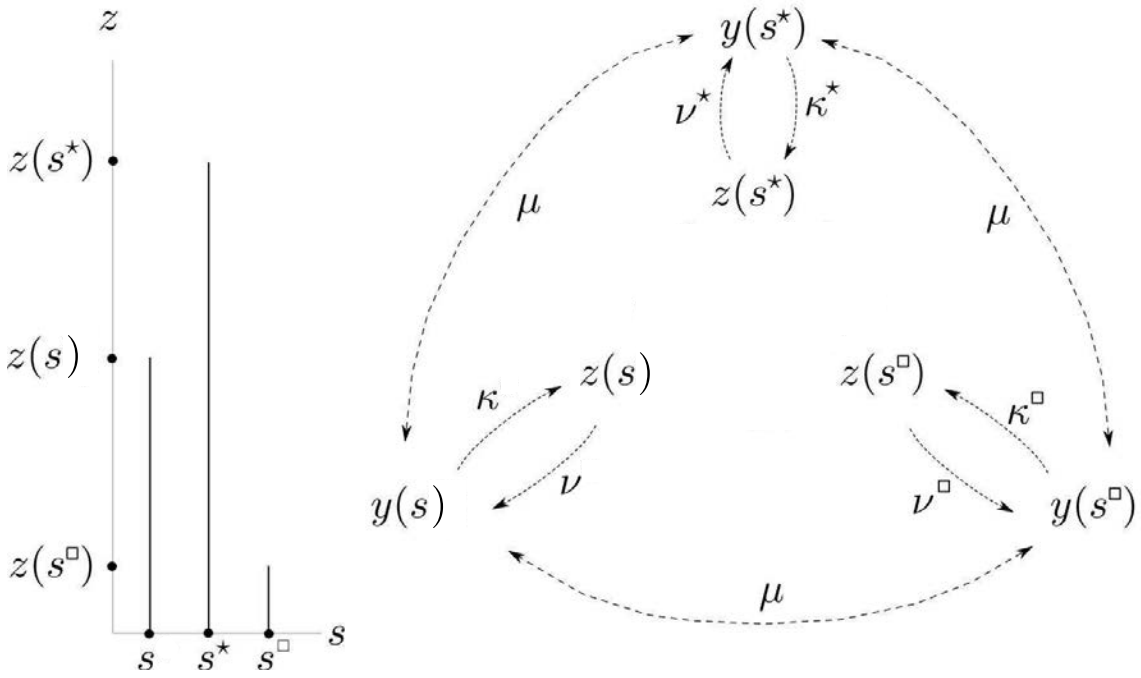


Figure 6.2: Schematic illustration of the interactions between populations in the case of three phenotypes.

of equations due to the discrete phenotype space:

$$\begin{aligned}
\dot{x} &= x(1-x) - x\sum_{i=0}^{m} \alpha_i\, z_i, \\
\dot{y_i} &= \alpha_i\, z_i\, x + \mu(y_{i+1} - 2y_i + y_{i-1}) - \gamma_i\, y_i, \\
\dot{z_i} &= \kappa_i\, \gamma_i\, y_i - \nu_i\, x\, \alpha_i\, z_i - \zeta_i\, z_i,
\end{aligned} \tag{6.4}$$

where $i = 0, 1, \ldots, m$. In order to describe the general case of $m+1$ strains along the phenotype space, we can define $s = i/m$ so that $i = 0$ corresponds to $s = 0$ and $i = m$ gives $s = 1$. Therefore, the lattice spacing in the phenotype space can be measured as

$$h = \frac{1}{m}, \tag{6.5}$$

and stand for the "phenotypic distance" between adjacent subpopulations of virus or infected cells. Consequently, the phenotypic space contains m equidistant subintervals: $0 = s_0 < s_1 < \ldots < s_i < \ldots < s_{m-1} < s_m = 1$. The discrete populations are grouped into p_i, where $p = y$ for infected cells and $p = z$ for virus. The susceptible cell population does not depend on the viral strains and thus is not discretized with respect to i.

Now we consider the limit $m \to \infty$, which turn discrete populations p_i into continuous populations $\bar{p}(s)$. The relationship between p_i and $\bar{p}(s)$ is given by

$$p_i = \frac{1}{m}\bar{p}\left(\frac{i}{m}\right) = h\,\bar{p}(ih) = h\,\bar{p}(s). \tag{6.6}$$

Definition (6.6) approximates values of p_i by the area of a rectangle with height $\bar{p}(s)$ and width h. The factor h that appears in front of $\bar{p}$ ensures that the total size of the population p is conserved in the continuum limit. Suppose, that

$$\sum_{i=0}^{m} p_i = P. \tag{6.7}$$

Then

$$P = \sum_{i=0}^{m} p_i = \sum_{i=0}^{m} \bar{p}(ih)\cdot h \to \int_0^1 \bar{p}(s)ds, \tag{6.8}$$

when $m \to \infty$ and thus $h \to 0$. Therefore, $\int_0^1 \bar{p}(s)ds = P$, and the discrete and continuous populations have the same total size. In other words, h is an approximation of the differential. Based on the discussed notations, the following is true:

$$\begin{aligned} p_{i+1} &= h\bar{p}((i+1)h) = h\bar{p}(s+h), \\ p_{i-1} &= h\bar{p}((i-1)h) = h\bar{p}(s-h). \end{aligned} \tag{6.9}$$

This allows us to find the continuum limit of $p_{i+1} - 2p_i + p_{i-1}$:

$$\begin{aligned} p_{i+1} - 2p_i + p_{i-1} = h\,(\bar{p}(s+h) - 2\bar{p}(s) + \bar{p}(s-h)) \to \\ \to h\left(h^2\frac{\partial^2\bar{p}(s)}{\partial s^2}\right) = h^3\frac{\partial^2\bar{p}(s)}{\partial s^2}, \end{aligned} \tag{6.10}$$

where $h \to 0$. Recalling that expressions above for $\bar{p}(s)$ are true for both $\bar{y}(s)$ and $\bar{z}(s)$, and assuming that the infection rate α_i is the only parameter that varies from strain to strain, we find that the continuum limit for model (6.4) when $m \to \infty$ is

$$\begin{aligned} \frac{\mathrm{d}x}{\mathrm{d}t} &= x(1-x) - x\int_0^1 \alpha(s)\bar{z}(s)ds, \\ \frac{\partial\bar{y}}{\partial t} &= x\,\alpha(s)\,\bar{z}(s) + \bar{\mu}\,\frac{\partial^2\bar{y}(s)}{\partial s^2} - \gamma\,\bar{y}(s), \\ \frac{\partial\bar{z}}{\partial t} &= \kappa\,\gamma\,\bar{y}(s) - \nu\,x\,\alpha(s)\,\bar{z}(s) - \zeta\,\bar{z}(s), \end{aligned} \tag{6.11}$$

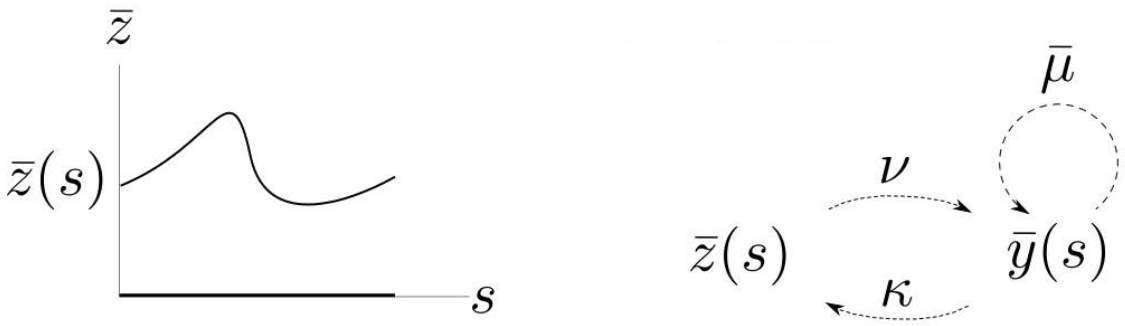


Figure 6.3: Left: Population of viral strains defined on a phenotype space continuously. Right: Populations interaction scheme for virus population of phenotypes $s \in S$, $\bar{z}(s)$, and the infected cell population, $\bar{y}(s)$.

where $\bar{\mu} = \mu\, h^2$, see Figure 6.3. Moreover, when the no-flux boundary condition $\partial\bar{y}(s,t)/\partial s = 0$ is used at $s = 0$ and $s = 1$, we have

$$\int_0^1 \frac{\partial^2 \bar{y}(s,t)}{\partial s^2} ds = \left[\frac{\partial \bar{y}(s,t)}{\partial s}\right]_0^1 = 0, \tag{6.12}$$

which ia the continuous analogue of the discrete conservation law, $\sum_i \Delta y(s_i) = 0$.

Therefore, by applying the diffusion operator to model mutation, we generalise the ODE-based model for the case of $m > 2$ different evolving viral phenotypes. All parameters of the discrete model except the mutation rate remain the same in the continuous model; the mutation rate is scaled with respect to the second order of the phenotype distance. Exploiting the phenotype space S allows us to replace certain parameters of fixed values with functions that vary over S. For instance, the infectiousness for every strain may vary and thus the infection rate can be considered a function, $\alpha(s)$, for every entry of the phenotype space $s \in S$. In terms of (6.1), $\alpha(s)$ is a phenotype-dependent function which remains 0 everywhere except two values of s, forming a profile with two spikes of infection rate values. In the next section, we show that this logical transition from ODE-based systems into PDEs does not alter the qualitative picture of the dynamics nor change the bifurcation structure. Simplified models can provide meaningful insights into elaborate models and the complicated systems we consider further can rely on discretization that secures the specifics of processes which are being studied.

6.2 Modelling of evolving viral strains with one susceptible cell type

By means of the phenotype space, we model the change in size of a viral population, $z(s,t)$, for all phenotypes $s \in S$. A virus cannot replicate and evolve without the machinery of a host cell. Due to the specific replication cycle, the model has to include the interactions of at least three populations. First, the population of uninfected susceptible cells, $x(t)$, which has a growth rate β limited by a carrying capacity K. The population of uninfected cells decreases either when it becomes infected by viral strains $z(s,t)$, due to the competition for the resources with the second population, or by natural decay at rate δ. Infections occur at rate $\alpha(s)$ and depends on the virus strains. The viral population infects the susceptible cell population and, third, the infected cell population inherits the phenotype of the strain, $y(s,t)$. We state that infection is the only process that increases the population of infected cells, that is, after an infection the cells do not multiply. The infection also affects the decay of susceptible cells, which is dependent on the virus virulence, γ. The diffusion term represents the mutations of viral strains which occur during the replication and production of virions by the infected cell. In the scope of this model, we assume that the production of virions is proportional to the virulence and happens at burst rate κ for each infected cell. On average, it takes ν free virus particles of the same strain to guarantee the infection process. In the absence of coinfection by a different strain, this rate ν can be seen as a multiplicity of infection (MOI). Finally, free virions lose their infectivity (and/or are flushed from the system) at a rate ζ.

Based on the mechanisms described above, the interaction of viral strains with a host cell can be studied by the following system of partial integro-differential equations (PIDEs):

$$\begin{aligned}
\frac{\mathrm{d}x(t)}{\mathrm{d}t} &= \beta x(t)\left(1-\frac{x(t)}{K}\right) - x(t)\int_0^1 \alpha(s)\, z(s,t)\, ds - \delta x(t),\\
\frac{\partial y(s,t)}{\partial t} &= x(t)\,\alpha(s)\, z(s,t) + \mu\,\Delta y(s,t) - \gamma\, y(s,t),\\
\frac{\partial z(s,t)}{\partial t} &= \kappa\,\gamma\, y(s,t) - \nu x(t)\alpha(s)\, z(s,t) - \zeta\, z(s,t).
\end{aligned} \tag{6.13}$$

The initial condition for susceptible cell population is chosen to be equal to the equilibrium value of the population at the absence of any pathogen in the system:

$$x(0) = \left(1-\frac{\delta}{\beta}\right) K, \tag{6.14}$$

while the infected cell population is absent at the beginning, *i.e.* $y(s,0) = 0$. The initial condition for the viral population is a constant:

$$z(s,0) = z_0. \tag{6.15}$$

The virus population size and the corresponding infected cell population size are spatial-variable-dependent functions and, thus, the boundary conditions must be introduced. Acknowledging the fact that phenotypic mapping to the interval may have various restraints, for the sake of discussion we apply no-flux boundary conditions for the infected cell population:

$$\left.\frac{\partial y(s,t)}{\partial s}\right|_{s=0,1} = 0. \tag{6.16}$$

The choice of the boundary condition ensures that mutation does not lead to a net loss or gain of infected cells.

The model can be simplified by introducing dimensionless variables that are based on characteristic timescales and population sizes. The quantity $\beta - \delta$ describes the effective growth rate of uninfected cells and its inverse, $(\beta-\delta)^{-1}$, is used to define the characteristic time scale of the system. The population variables are scaled with respect to the maximum size of the uninfected cells population:

$$t = (\beta-\delta)^{-1}\bar{t}, \quad x = \left(1-\frac{\delta}{\beta}\right) K\,\bar{x}, \quad y = \left(1-\frac{\delta}{\beta}\right) K\,\bar{y}, \quad z = \left(1-\frac{\delta}{\beta}\right) \kappa K\,\bar{z},$$

where $\bar{t}$, $\bar{x}$, $\bar{y}(s)$, and $\bar{z}(s)$ are non-dimensional variables. The non-dimensional parameters replace previous parameters as follows

$$\bar{\alpha}(s) = \frac{\kappa\, K\, \alpha(s)}{\beta}, \quad \bar{\mu} = \frac{\mu}{\beta-\delta}, \quad \bar{\gamma} = \frac{\gamma}{\beta-\delta}, \quad \bar{\zeta} = \frac{\zeta}{\beta-\delta}, \quad \bar{\nu} = \frac{\nu}{\kappa}. \tag{6.17}$$

Upon omitting the bars for convenience, the nondimensional system looks as follows:

$$\begin{aligned}
\frac{\mathrm{d}x(t)}{\mathrm{d}t} &= x(t)\,(1-x(t)) - x(t)\int_0^1 \alpha(s)\, z(s,t)\, ds,\\
\frac{\partial y(s,t)}{\partial t} &= x(t)\,\alpha(s)\, z(s,t) + \mu\,\Delta y(s,t) - \gamma\, y(s,t),\\
\frac{\partial z(s,t)}{\partial t} &= \gamma\, y(s,t) - \nu x(t)\alpha(s)\, z(s,t) - \zeta\, z(s,t).
\end{aligned} \tag{6.18}$$

The model (6.18) matches the continuum limit of ODE-based system of equations obtained earlier in (6.11) with $\kappa = 1$ and μ being a small parameter. This allows us to use

the same parameters and initial conditions in the PDE model (6.18) and the ODE model of Section 4.2, allowing for a systematic comparison if the models. To illustrate that, more detailed study of the model (6.18) is required.

6.3 Analyses and results

A first approach to understand the dynamics of a system of equations lays in analysing its equilibria. Finding the equilibria of (6.18) analytically is not feasible yet two of its steady solutions can easily be guessed. The trivial state, $\{x, y, z\} = \{0, 0, 0\}$, and the virus-free state, $\{x, y, z\} = \{1, 0, 0\}$, both turn the right-hand side of system (6.18) into zero. To determine the stability of the trivial state, we consider a slightly perturbed equilibrium and check if the solution tends towards the original trivial state when the perturbation coefficient approaches zero. As shown below, for determining the stability of the trivial equilibrium, a study of the first equation, for $x(t)$, is sufficient. For the equilibrium point $\{x_0, y_0, z_0\} = \{0, 0, 0\}$, consider a point infinitely close to it: $\{\tilde{x}, \tilde{y}, \tilde{z}\} = \{x_0 + \varepsilon\, x, y_0 + \varepsilon\, y, z_0 + \varepsilon\, z\} = \{\varepsilon\, x, \varepsilon\, y, \varepsilon\, z\}$, where $\varepsilon \to 0$. Substitution of the perturbed equilibrium $\{\tilde{x}, \tilde{y}, \tilde{z}\}$ into (6.18), where the equation for x looks as follows:

$$\frac{\mathrm{d}\tilde{x}(t)}{\mathrm{d}t} = \tilde{x}(t)\,(1 - \tilde{x}(t)) - \tilde{x}(t) \int_0^1 \alpha(s)\, \tilde{z}(s,t)\, ds, \tag{6.19}$$

leads to

$$\frac{\mathrm{d}\varepsilon\, x(t)}{\mathrm{d}t} = \varepsilon\, x(t)\,(1 - \varepsilon\, x(t)) - \varepsilon\, x(t) \int_0^1 \alpha(s)\, \varepsilon\, z(s,t)\, ds, \tag{6.20}$$

Simplifying (6.20), gives

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = x(t) - \varepsilon\,(x(t))^2 - \varepsilon\, x(t) \int_0^1 \alpha(s)\, z(s,t)\, ds. \tag{6.21}$$

In taking the limit $\varepsilon \to 0$, we obtain

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = x(t). \tag{6.22}$$

The first component of the solution, x, illustrates how the system behaves along time in the direction of x: $x(t) = e^t + C$, for all $C \in \mathbb{R}$. In this case of trivial equilibrium, the exponential dependency on time of either components of the solution implies the instability of the studied equilibrium. Moderately perturbing the equilibrium we obtained an exponential divergence from the trivial equilibrium, which by all means classifies the equilibrium as unstable. Moreover, the trivial equilibrium of (6.18) is unstable for all possible values

of all parameters. Predictably, the result matches with instability of the trivial equilibrium of the model involving two strains, (6.1). All other equilibria must be studied numerically due to their complexity.

Let us examine how the system behaves for certain fixed values of parameters. For simplicity, we assume the infection rate of strains $\alpha(s)$ are zero everywhere except at two different values of s, that is:

$$\alpha(s) = \begin{cases} \alpha_1 & \text{when } s = s_1, \\ \alpha_2 & \text{when } s = s_2, \\ 0 & \text{otherwise,} \end{cases} \tag{6.23}$$

where $s_1 \neq s_2$. The other parameters are equal to those used for the bifurcation analysis of the model (6.1), given by:

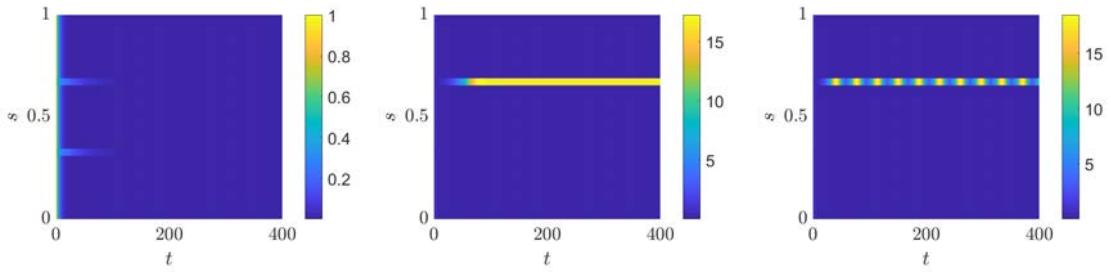
$$\mu = 0.1 \cdot h^2, \qquad \gamma = 0.25, \qquad \nu = 0.5, \qquad \zeta = 0.2222,$$

where h is the phenotype distance and states that the mutation rate is a small parameter for the system (6.18). For a fixed value of $\alpha_1 = 0.3$ at $s = 0.33$ and for three different values of α_2 at $s = 0.66$ from (6.23), we ran numerical simulations using the initial conditions

$$x(0) = 1, \qquad y(s,0) = 0, \qquad z(s,0) = 1. \tag{6.24}$$

Figure 6.4 displays that the viral load $z(s,t)$ is changing in time. When the infection rates of the strains corresponding to $s = 0.33$ and $s = 0.66$ are small, the virus cannot persist in the system (see Figure 6.4 (a)). Increasing the infection rate of the strain corresponding to $s = 0.66$ leads to a survival of the strain (see Figure 6.4 (b)). Further increase in the infection rate brings a periodic behaviour of the fittest strain while the other strain fails as before (see Figure 6.4 (c)). In a deliberate attempt to mimic the interaction of two strains studied in Section 4.2, we achieved the predictably similar outcome: increasing the infection rate brings the system from a virus-free state to a stationary existence and, further, to periodic existence of the strains which are fit enough for the system.

In order to find the parameter intervals corresponding to different behaviours of the fittest strains, we construct a one-parameter bifurcation diagram. Figure 6.5 (a) illustrates the shift in stability of the system from one equilibrium to another due to the varying parameter, α_2. We compare this with the bifurcation diagram of system (6.1) displayed in



(a) $\alpha_1 = 0.3$, $\alpha_2 = 0.3$ (b) $\alpha_1 = 0.3$, $\alpha_2 = 1$ (c) $\alpha_1 = 0.3$, $\alpha_2 = 2$

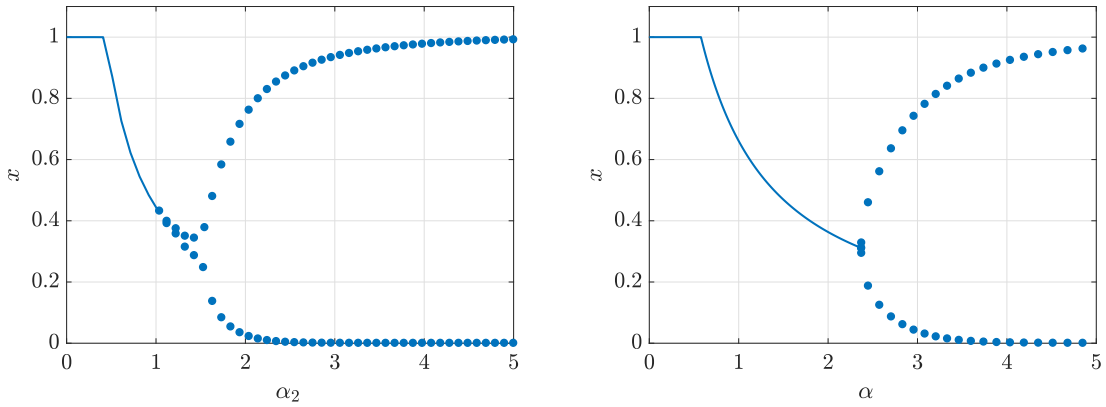
Figure 6.4: Three different time series are shown for the virus population, $z(s,t)$, where the population size is depicted by colour. The initial and boundary conditions are the same for all three cases. In all cases, the strain corresponding to $s = 0.33$ does not persist at the end of the simulations. The system stabilises at the following states: (a) virus-free state, (b) steady existence of the fittest viral strain, and (c) periodic existence of viral strain with a higher infection rate.

Figure 6.5 (b). The value of $\alpha^{\square} = 0.3$ is fixed at $\alpha_1 = 0.3$. The qualitative similarity between the two models is obvious. Apart from a smother transition between steady equilibrium and periodic solution, the behaviour of the PDE- (see Figure 6.5 (a)) and ODE-based (see Figure 6.5 (b)) models are qualitatively identical.

6.4 Fitness landscape analysis

The reproduction rate of a virus is proportional to the number of successful infections that virions accomplish. Depending on phenotypic characteristics of viral strains, different strains can have different infectiousness. Thus, every viral strain defined on the phenotype space can be assigned a fitness value represented by the infection rate. The concept of visualising reproductive success started in early thirties of the XX$^{\text{th}}$ century [143] and was named as fitness landscapes or adaptive landscapes. In the previous section, we explored the behaviour of the model considering two strains with nonzero infection rates in the whole phenotype space. Broadening this concept, consider a distribution of infection rates. This distribution is defined over the phenotype space and has local maxima which are consistent with the most successful strains. Here and further, we use concepts of the fitness landscape and function of infection rate interchangeably.

Our goal in this section is to observe how the behaviour of the system changes when the fitness landscape is a distribution of values rather than a couple of single spikes. As an infection rate for the following test, we first consider the Gaussian approximation of two,



(a) PDEs (b) ODEs

Figure 6.5: One-dimensional bifurcation diagrams for systems (6.18) and (6.1) shown on plots (a) and (b), respectively. Fixed parameters are $\alpha_1 = 0.3$ and $\alpha^{\square} = 0.3$, and bifurcation parameters are α_2 and α. Solid lines correspond to stable equilibria while markers stand for maximum and minimum of periodic orbits. All other corresponding parameters of the models are fixed and equal.

scaled Dirac's delta functions centred at s_1 and s_2:

$$\alpha(s) = \alpha_1\delta(s - s_1) + \alpha_2\delta(s - s_2) \approx \frac{\alpha_1}{\sqrt{2\pi}\varepsilon} e^{\frac{-(s-s_1)^2}{2\varepsilon^2}} + \frac{\alpha_2}{\sqrt{2\pi}\varepsilon} e^{\frac{-(s-s_2)^2}{2\varepsilon^2}}, \tag{6.25}$$

where ε is small and controls the width of the peaks. Dirac's delta function is chosen because the area underneath it is always equal to one. This property of the delta function provides an easy scaling to control the total infectiousness of the studied virus population. The fitness landscape, with $\alpha_1 = 0.03$ and $\alpha_2 = 0.1$ is plotted in Figure 6.6. Here, for visual separation of two peaks, ε is defined as $1.1m^{-1}$, where $m = 60$ is a discretization along the phenotype space used in the method of lines. Approximations of the Dirac's delta functions are centred around $s_1 = 0.33$ for α_1 and $s_2 = 0.66$ for α_2. If we fix $\alpha_1 = 0.03$ and vary α_2, we observe the change of behaviour for strains defined around s_2.

Using the same parameters, initial and boundary conditions as in the previous section, we run simulation for system (6.18), the result of which is shown in Figure 6.7. In comparison to the dynamics of the system with infection rates defined as single spikes, the time series profile of the system is much smoother along s. For small values of infection rates, the virus-free state persists. For greater α_2, steady and periodic viral persistence can be observed. Although there are several surviving strains in Figures 6.7 (b) and (c) that are clustered around $s = 0.66$, we prefer not to use term "coexistence" because the surviving strains are too similar phenotypically to one another. For a fixed value of α_1 and varying α_2, the one-dimensional bifurcation diagram can be constructed, as shown in in Figure 6.8.

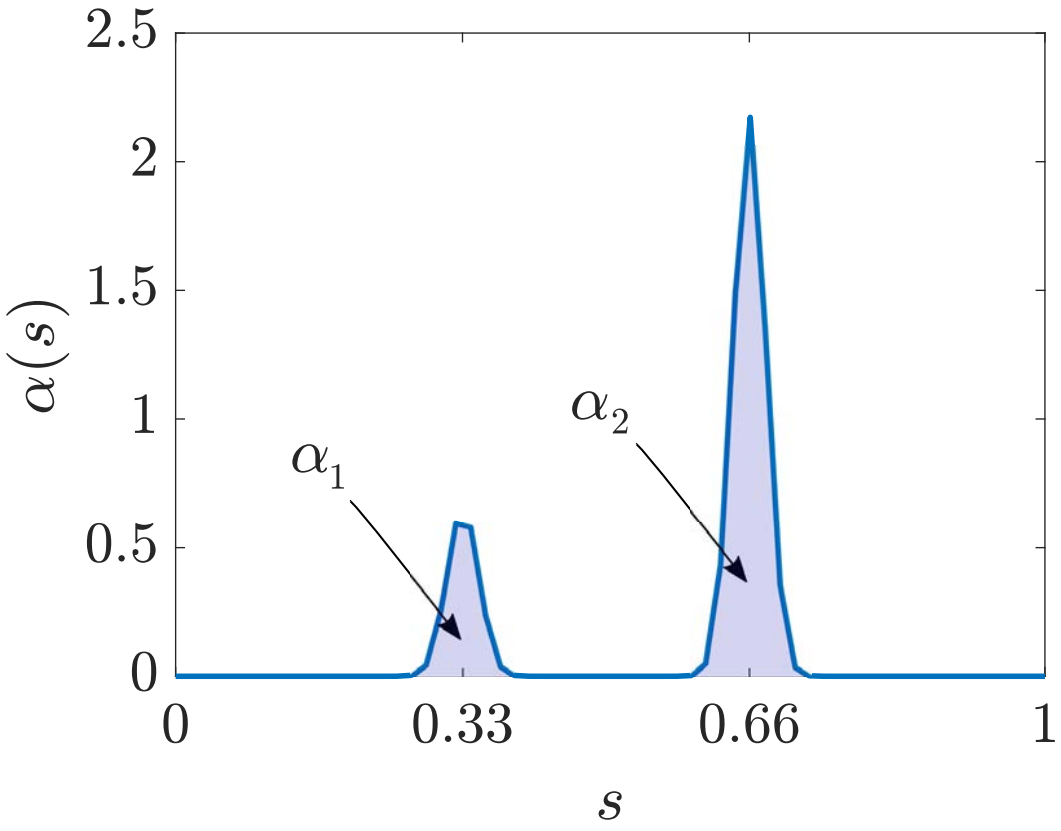
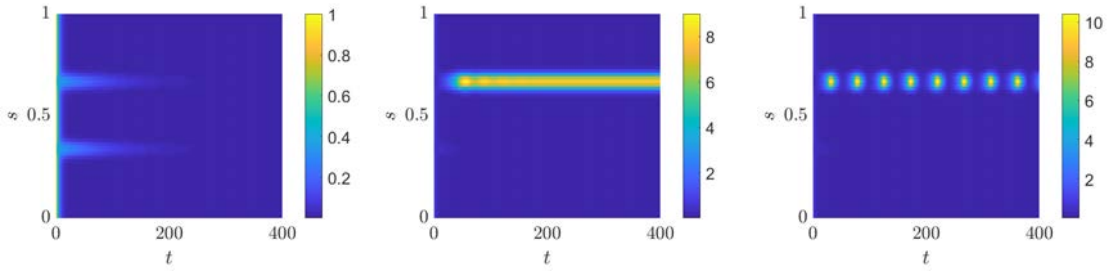


Figure 6.6: Distribution of infection rate with two peaks.



(a) $\alpha_1 = 0.03$, $\alpha_2 = 0.03$ (b) $\alpha_1 = 0.03$, $\alpha_2 = 0.1$ (c) $\alpha_1 = 0.03$, $\alpha_2 = 0.2$

Figure 6.7: Three different time series are shown for the virus population, $z(s,t)$, where the population size is depicted by colour. The initial and boundary conditions are the same for all three cases. In all cases, the strain corresponding to $s = 0.33$ does not persist at the end of the simulations. The system stabilises at the following states: (a) virus-free state, (b) steady existence of the fittest viral strains, and (c) periodic existence of viral strain with the highest infection rates.

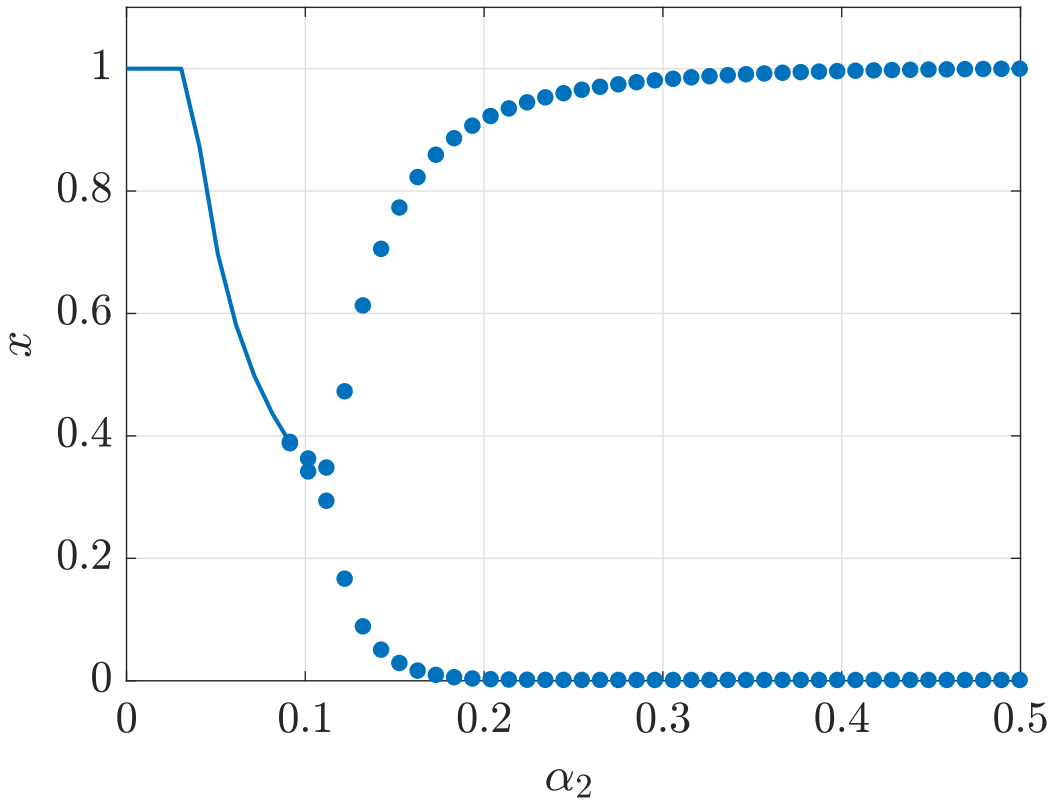


Figure 6.8: One-dimensional bifurcation diagram of the model (6.18) with infection rate depicted in (6.25). Here, the scaling parameter $\alpha_1 = 0.03$. To illustrate the bifurcations, x component of equilibria was chosen. Solid lines stand for steady equilibrium, and round markers correspond to maxima and minima of the periodic orbits.

As in the previous case, we see three distinct states depending on the value of α_2. The transcritical bifurcation slightly above $\alpha_2 = 0.03$ includes an exchange of stability between the virus free-state and the virus-persistent state. Further, the Hopf bifurcation brings a periodic orbit around freshly unstable virus-persistent state. The stable, periodic behaviour is continuous for all values of α_2 exceeding $\alpha_2 = 0.0918$.

To understand how the system behaves when both peaks vary, we build a two-dimensional bifurcation diagram, see Figure 6.9. As was predicted before, small values of infection rate do not allow the persistence of virus, leading to the existence of a region I, where the virus-free state is stable. Dashed lines in blue corresponding to the transcritical bifurcations separate the virus-free state from the virus-persistent states. There are two regions corresponding to this behaviour, both marked as II. Depending on the larger infection rate of phenotypes, the viral strains survive around that phenotype. These steady states border the periodic states and are separated by the Hopf bifurcation in solid blue lines. Periodic

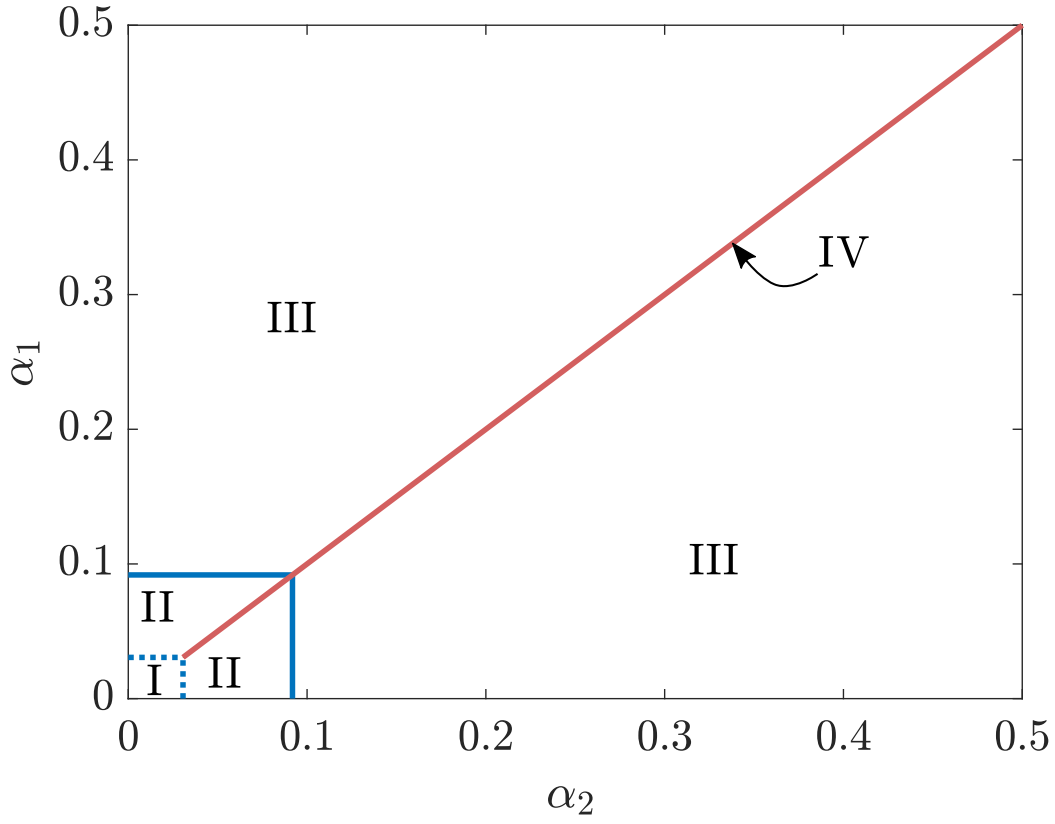
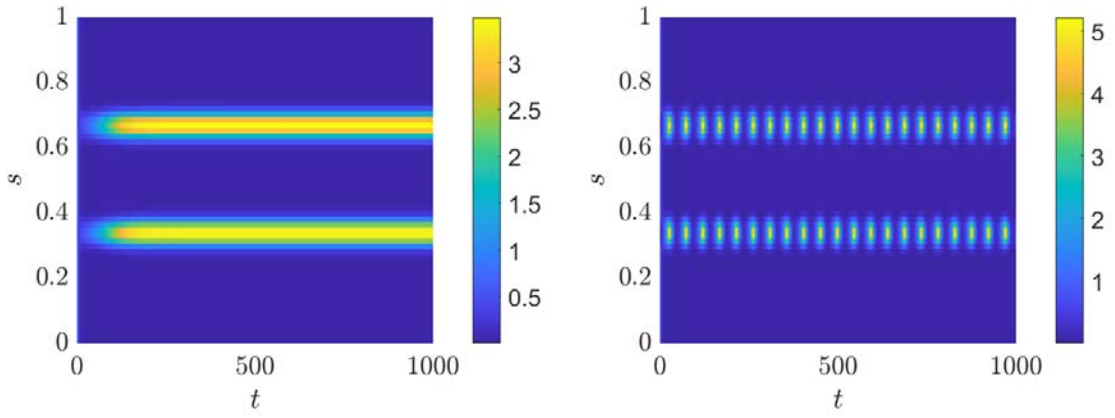


Figure 6.9: Two-dimensional bifurcation diagram for model (6.18) using the infection rate given by (6.25). Region I stand for virus free state, regions III are virus-persistent stationary state, and regions III are oscillatory persistence of virus population. Red line, region IV, corresponds to coexistence of distinct clusters of viral strains.



(a) $\alpha_1 = \alpha_2 = 0.05$ (b) $\alpha_1 = \alpha_2 = 0.2$

Figure 6.10: Coexistence of two clusters of phenotypically distinct strains of virus. Colours indicate the sub-population size of viral strains.

states of virus persistence, marked as III, occupy the largest areas of the plot. The red line separating the coupled regions correspond to the only values of α_1 and α_2 that lead to coexistence of the strains around phenotypes s_1 and s_2. Moreover, the coexistence style repeats the style of viral persistence it is adjacent to. For instance, the infection rate coefficients $\alpha_1 = \alpha_2 = 0.05$ lead to two phenotypically different clusters of strains, see Figure 6.10 (a). The larger infection rate coefficients $\alpha_1 = \alpha_2 = 0.2$ bring two clusters of phenotypically far strains oscillating in the positive phase space, see Figure 6.10 (b). The most counter-intuitive result of these simulations is the strong manifestation of competition between different strains. Since all of the strains share susceptible cells, the success of one strain immediately leads to a failure of all the other clusters of strains in the system. The slightest deviation from the diagonal line, brings an extinction of all strain clusters except the cluster of the fittest strains. Clustered survival also has an important consequence: the strains closest to the fittest one in a load tend to win the competition with more fit, yet distant, strains under some assumptions. That is, imagine that distant phenotypes have infection rates which obey the following assumption: $\alpha(s_1) < \alpha(s_2)$. Then viral strain $z(s_2 + \epsilon)$ for small ϵ may survive in the long run, while $z(s_1)$ vanishes, although $\alpha(s_1) > \alpha(s_2 + \epsilon)$.

Overall, this test shows the consistency between "continuous" and discrete models of viral dynamics. Interestingly, the general dynamics are dictated by the maximum of the infection rate yet every strain performs with respect to its individual infection rate. Moreover, the model (6.18) is consistent with respect to the discretization used for the numerical simulations. That is, for different values of m in a discretization, the one-dimensional bifurcations do not change noticeably.

Chapter 7

Specialisation of viral strains among multiple host species

7.1 Background

In this chapter, we consider viral evolution by adaptive radiation in complex fitness landscapes using the framework of host-virus systems. Most of the research done on host-parasite associations has been formulated in terms of a co-evolutionary scheme in which parasites were given the role of a passive follower agent [6, 88, 100]. However, parasites, also being under Darwinian selection, have their own evolutionary capabilities. Moreover, viruses reproduce at a faster rate than their hosts. It means more generations of viruses compare to fewer generations of host species, which yields a prevailing quantitative appearance of genetic variations in the viral population over the host population. Differences in generation time and high mutation rates motivate the necessity of studying parasite evolution as a main subject.

The occurrence of specialisation by a virus appears to be common: specialisation of HIV-1 [147, 57] and strong connections of bacteriophages to a specific bacterial species [55, 13, 40, 56, 89] are examples of this. In the case of HIV-1, there is a hypothesis explaining the long transmission period in most cases of infection by viral evolution. More precisely, specialisation of the HIV-1 towards specifics of the T-cells within a host. In order to illustrate this process, we can use a mathematical formulation recently suggested for modelling viral evolution [66], which can be extended to study the onset of specialisation. Although experimental data on the subject exists, theories that explain host-range evolution in viruses are not well developed. The development of a simple and convenient mathematical approach

that can be used to illustrate viral evolution is the aim of this chapter. Furthermore, the model is used to explore how the number of generations necessary for specialisation to occur depends on the parameters of the system. Simulations were written in Python using finite differences to approximate derivatives and Simpson's rule to perform numerical integration. The chapter is concluded with an elaborate use of regression analysis on multiple parameters to define a qualitative change in the functional dependence of specialisation time of a virus population.

7.2 Modelling specialisation towards a cell

The model that we formulate in this chapter is initially an extension of a previous model of viral evolution [66], which, in turn, is based on the Nowak and May model of HIV-1 dynamics [95]. The Nowak and May model utilises the concept of a parasite with a free infectious stage [4] and was specifically developed to describe the dynamics of HIV-1 within a host. It is noteworthy that this model is based on a set of clearly identified assumptions and hypotheses (so called "first principles") and, therefore, the results and parameters can be easily and immediately interpreted. In other words, the model is mechanistic, and for this reason constitutes a sound base for further development. In terms of the previous chapter, we are developing the model (6.13). The new model contains the following partial integro-differential equations (PIDEs):

$$\begin{aligned}
\frac{\mathrm{d}x_i(t)}{\mathrm{d}t} &= \lambda_i - x_i(t)\int_{s\in S}\alpha_i(s)\,z(s,t)\,ds - \delta_i\,x_i(t),\\
\frac{\partial y_i(s,t)}{\partial t} &= \alpha_i(s)\,x_i(t)z(s,t) - \gamma_i\,y_i(s,t) + \mu\,\Delta y_i(s,t),\\
\frac{\partial z(s,t)}{\partial t} &= \sum_{i=1}^{N}\kappa_i\,y_i(s,t) - \nu\,\alpha_i(s)\,x_i(t)z(s,t) - \zeta\,z(s,t),
\end{aligned} \tag{7.1}$$

where $i = 1, \ldots, N$ correspond to different types of susceptible and potentially infected cells. By introducing several cell types, we create a possibility for viral strains to exploit one or all of the susceptible cells, and thus specialise or generalise, respectively. The influx of susceptible cells are λ_i, and the per capita retreat (natural death) rates are δ_i for $i = 1, \ldots, N$. This simplified growth rate of susceptible cells is sufficient for the purpose of this chapter. The background studies of the effective growth rate $\lambda - \delta x$ for one susceptible cell type and one viral strain were conducted in Section 3.3. Here, N types of susceptible cells, x_i, are available for all strains of viral population, $z(s)$, to infect according to the corresponding infection rates $\alpha_i(s)$ for $i = 1, \ldots, N$ and viral phenotypes $s \in S$. Separate α_i allow the

model to capture the unique characteristics of each type of susceptible cell while the dependency of each $\alpha_i(s)$ on phenotype space determines the success rate of viral strains in infecting each type of available susceptible cell. Cross infections are left out of the scope of discussion, and thus the average number of virions of the same strain required to accomplish a successful infection is the multiplicity of infection, ν. Each infected cell carries a memory about the strain it was infected by, $y(s,t)$. This allows the mutation of viral strains in the population of infected cells, where most of the mutations occur to be modelled, with a mutation rate μ. The natural apoptosis rate of cells is altered after the infection, and therefore the death rate of the infected cell population, γ_i, is different from the death rate of the corresponding uninfected cells. Since the process of virion release does not necessarily destroy the host cell, we do not specify the connection between decay rate of infected cells γ_i and the exact amount of infected cells involved in the production of virions. We assume that $\gamma_i = \gamma_i^{\text{vir.}} + \gamma_i^{\text{sys.}} + \gamma_i^{\text{apop.}}$, which represents the combination of all possible decay scenarios for a population of infected cells: it might be either due to virulence or as a result of an immune system attack or, as before, because of the pre-programmed natural death of cells. In this model, we consider a lytic cycle of viral reproduction, and thus the burst size, κ_i, is associated with every infected cell type. The average outflow of the virus from the system occurs at rate ζ. For the sake of clarity, the descriptions of the parameters and variables of (7.1) with arbitrarily fixed parameters and initial conditions are listed in the table below:

Symbol	Short description	(Initial) value	Units
t	time	0	day
s	phenotype (space) variable	between $0-1$	-
$x_i(t)$	susceptible cell type i	$\sim 10^2$	$\frac{cells}{ml}$
$y_i(s,t)$	infected cell type i	0	$\frac{cells}{ml}$
$z(s,t)$	viral load	10^0 for $s = s*$	$\frac{copies}{ml}$
λ_i	influx rate of susceptible cells	$\sim 10^8$	$\frac{cells}{ml \cdot day}$
$\alpha_i(s)$	infection rate	between $\sim 10^1$	$\frac{ml}{copies \cdot day}$
δ_i	death rate of susceptible cells	$\sim 10^6$	$\frac{1}{day}$
γ_i	death rate of infected cells	$\sim 10^0$	$\frac{1}{day}$
μ	mutation rate of virus via inf. cells	$\sim 10^{-3}$	$\frac{1}{day}$
κ_i	birth rate of virus by inf. cell i	$\sim 10^3$	$\frac{copies}{cells \cdot day}$
ν	multiplicity of infection	$\sim 10^{-4}$	$\frac{copies}{cells}$
ζ	outflow rate of virus	10^3	$\frac{1}{day}$

From the previous chapter we know that the mutation rate μ is a small parameter. Nondimensionalization of system (7.1) gives the following:

$$\begin{aligned}
\frac{\mu}{\delta_i}\frac{\mathrm{d}x_i(t)}{\mathrm{d}t} &= 1 - \frac{\alpha_{max_i}}{\alpha_{max_1}}\frac{\delta_1}{\delta_i}x_i(t)\int_{s\in S}\alpha_i(s)\,z(s,t)\,ds - \,x_i(t), \\
\frac{\partial y_i(s,t)}{\partial t} &= \frac{\gamma_i}{\mu}(x_i(t)\alpha_i(s)\,z(s,t) - y_i(s,t)) + \Delta y_i(s,t), \\
\frac{\mu}{\zeta}\frac{\partial z(s,t)}{\partial t} &= \sum_{i=1}^{N}\frac{\lambda_i\,\alpha_{max_i}\,\kappa_i}{\zeta\,\delta_i\,\gamma_i}\,y_i(s,t) - \left(\frac{\nu}{\zeta}\sum_{i}^{N}\frac{\lambda_i\,\alpha_{max_i}}{\delta_i}\alpha_i(s)x_i(t) + 1\right)z(s,t),
\end{aligned} \tag{7.2}$$

where $0 \leq \alpha_i(s) \leq 1$ and $[\alpha_i] = 1$, while $[\alpha_{max_i}] = \frac{ml}{copies \cdot day}$. Here, the model is written on the time-scale of mutation. Information about the magnitude of the studied parameters can affect the analysis, and sometimes simplify the model drastically. For example, using the parameters in the table above, we find that the mutation rate is much smaller than the death rates of susceptible cells and the virus,

$$\mu/\delta_i \ll 1, \quad \mu/\zeta \ll 1. \tag{7.3}$$

These inequalities show that the left-hand sides of the first and third equations in (7.2) can be set to zero, resulting in a simplified model given by

$$\begin{aligned} x_i(t) &= \frac{1}{A_i \int_{s\in S} \alpha_i(s)\, z(s,t)\, ds + 1}, \\ \frac{\partial y_i(s,t)}{\partial t} &= B_i \left(x_i(t)\alpha_i(s)\, z(s,t) - y_i(s,t)\right) + \Delta y_i(s,t), \\ z(s,t) &= \frac{\sum_{i=1}^{N} C_i\, y_i(s,t)}{\sum_i^N D_i \alpha_i(s) x_i(t) + 1}, \end{aligned} \tag{7.4}$$

where

$$A_i = \frac{\alpha_{max_i}}{\alpha_{max_1}} \cdot \frac{\delta_1}{\delta_i}, \quad B_i = \frac{\gamma_i}{\mu}, \quad C_i = \frac{\lambda_i\, \alpha_{max_i}\, \kappa_i}{\zeta\, \delta_i\, \gamma_i}, \quad D_i = \frac{\nu}{\zeta}\, \frac{\lambda_i\, \alpha_{max_i}}{\delta_i} \tag{7.5}$$

for all $i = 1, \ldots, N$. However, due to the validity of the reduction depending strongly on other parameter values, we leave analysing this model for future discussion. Without specific knowledge about parameter values, in general for model (7.2) it is possible to obtain a scaled basic reproduction number to characterise the main growth and evolution rate of the viral population:

$$\tilde{R}_0 = \sum_i^N \tilde{R}_{0_i} = \sum_i^N \frac{\lambda_i\, \alpha_{max_i}\, \kappa_i}{\zeta\, \delta_i\, \gamma_i}. \tag{7.6}$$

This combination of parameters is found by nondimensionalization of the initial system. It is worth noting that with proper adjustments due to the change in the burst size, this expression is equivalent to the basic reproduction number (3.8) found in Chapter 3. As a main difference, the expression (7.6) now depends on the number of populations described in the model. Further in this chapter, for simplicity in simulations and illustrations, we limit the model to two types of susceptible cells. In this case, the interaction between variables is described by the following (dimensional) equations:

$$\begin{aligned} \frac{\mathrm{d}x_1(t)}{\mathrm{d}t} &= \lambda_1 - x_1(t) \int_{s\in S} \alpha_1(s)\, z(s,t)\, ds - \delta_1\, x_1(t), \\ \frac{\mathrm{d}x_2(t)}{\mathrm{d}t} &= \lambda_2 - x_2(t) \int_{s\in S} \alpha_2(s)\, z(s,t)\, ds - \delta_2\, x_2(t), \\ \frac{\partial y_1(s,t)}{\partial t} &= x_1(t)\, \alpha_1(s)\, z(s,t) - \gamma_1\, y_1(s,t) + \mu\, \Delta y_1(s,t), \\ \frac{\partial y_2(s,t)}{\partial t} &= x_2(t)\, \alpha_2(s)\, z(s,t) - \gamma_2\, y_2(s,t) + \mu\, \Delta y_2(s,t) \\ \frac{\partial z(s,t)}{\partial t} &= \kappa_1\, y_1(s,t) + \kappa_2\, y_2(s,t) - (\nu_1 x_1(t)\, \alpha_1(s) \ + \nu_2 x_2(t)\, \alpha_2(s) + \zeta)\, z(s,t). \end{aligned} \tag{7.7}$$

Initial conditions for model (7.7) are

$$x_1(0) = \frac{\lambda_1}{\delta_1}, \qquad x_2(0) = \frac{\lambda_2}{\delta_2}, \qquad y_1(s,0) = 0, \qquad y_2(s,0) = 0. \qquad z(s,0) = Z(s), \quad (7.8)$$

where the susceptible cells are given the maximum steady state value which can be achieved in the absence of the virus; and the standard no-flux boundary conditions

$$\left.\frac{\partial y_1(s,t)}{\partial s}\right|_{\bar{S}} = 0 \qquad \text{and} \qquad \left.\frac{\partial y_2(s,t)}{\partial s}\right|_{\bar{S}} = 0, \quad (7.9)$$

are applied for functions $y_1(s,t)$ and $y_2(s,t)$ defined on $S = [0,1] \times [0,1]$ with boundaries $\bar{S}$. Apart from fixed parameters, the model should be complemented with a fitness landscape, that is a weighted sum of functions $\alpha_i(s)$, $i = 1, 2$. We stipulate without loss of generality that $\alpha_i(s)$, $i = 1, 2$, are Gaussian functions defined on the two-dimensional phenotype space S with different means, *e.g.* $(s_1, s_2) = (0.25, 0.75)$ and $(s_1, s_2) = (0.75, 0.25)$, and a variance σ^2 that allows to distinguish the peaks. This assumption provides the possibility for the virus to increase its efficiency of exploiting one of the susceptible cells at the expenses of decreasing the efficacy of exploiting the other. Therefore, the costs of adaptation to one or another cell type can be regulated by defining the parameters $\alpha_i(s)$, $i = 1, 2$. Based on the goal of every particular case, the process of niche occupation can be defined in different ways. In this chapter, we assume that there are two niches – corresponding to the number of provided cell types. Susceptible cells x_1 and x_2 can be infected at a rate $\alpha_1(s)$ and $\alpha_2(s)$ by each viral phenotype $s = (s_1, s_2)$. By defining the $\alpha_i(s)$ as a normal distribution with different means we introduce possible niches. The process of "niche being occupied" should be understood in a way that there is a big enough sub-population of the virus, *e.g.* $z(\hat{s}, t)$ where connected $\hat{s} \subset S$, infecting mostly one type of the cells. Specifically, we arrange a percentage in order to test whether a niche is occupied or not. For instance, if at least 15% of total viral strains are found around the mean of $\alpha_1(s)$, then the niche corresponding to the type of susceptible cells x_1 is considered occupied. Moreover, if the sum of the viral sub-populations which occupy both niches is greater than the rest, then we claim that specialisation of virus towards the susceptible cells has occurred.

7.3 Results

The parameter values for the numerical simulations are fixed as:

$$\begin{aligned} \lambda_1 = \lambda_2 = 1.2 \cdot 10^8, \quad \delta_1 = \delta_2 = 1.04 \cdot 10^6, \quad \gamma_1 = \gamma_2 = 0.71, \\ \mu = 0.001, \quad \kappa_1 = \kappa_2 = 1.5 \cdot 10^3, \quad \nu_1 = \nu_2 = 10^{-4}, \quad \zeta = 10^3. \end{aligned} \tag{7.10}$$

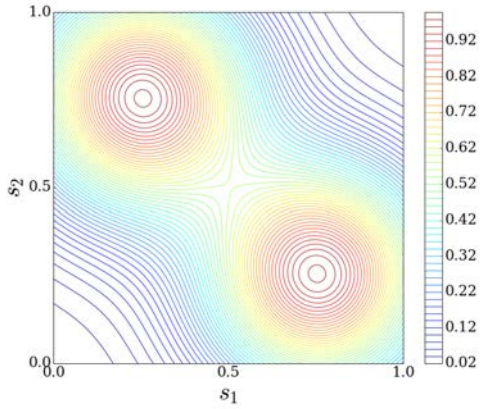
For comprehensibility, we consider niches defining efficiencies $\alpha_1(s)$ and $\alpha_2(s)$ that have peaks which are symmetric with respect to the diagonal $s_1 = s_2$. That is, for the case $N = 2$ with:

$$\begin{aligned} \alpha_1(s_1, s_2) &= \exp\left(-0.001 \cdot ((s_1 - 0.25)^2 + (s_2 - 0.75)^2)\right), \\ \alpha_2(s_1, s_2) &= \exp\left(-0.001 \cdot ((s_1 - 0.75)^2 + (s_2 - 0.25)^2)\right). \end{aligned} \tag{7.11}$$

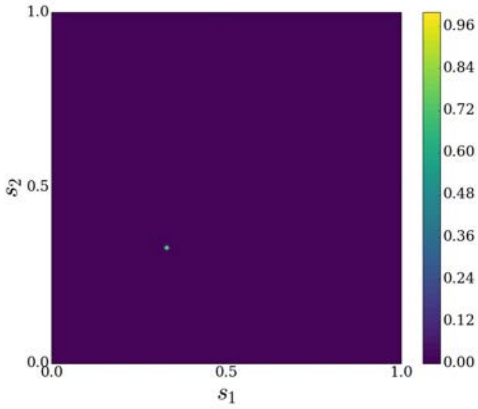
The no-flux boundary conditions are applied at all boundaries. The initial condition for susceptible and infected cells are chosen to be $x_1(0) = \lambda_1/\delta_1\, cells/ml$, $x_2(0) = \lambda_2/\delta_2\, cells/ml$, $y_1(s_1, s_2, 0) = y_2(s_1, s_2, 0) = 0\, cells/ml$, and initial viral population is equal to zero everywhere but at a point:

$$Z(s_1, s_2, 0) = \begin{cases} 1\, \frac{copies}{ml} & \text{for } (s_1, s_2) = (1/3, 1/3), \\ 0\, \frac{copies}{ml} & \text{otherwise.} \end{cases}$$

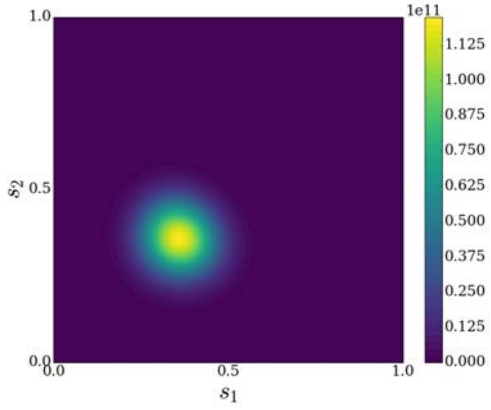
For the parameter values (7.10), that lead to the stable coexistence of phenotypically separate sub-populations, the dynamics are depicted in Figure 7.1. The infection rate (7.11) is shown in Figure 7.1 (a). Figures 7.1(b) to 7.1(f) illustrate the distribution of the viral population $z(s_1, s_2, t)$ for different times: 1 (7.1(b)), 250000 (7.1(c)), 500000 (7.1(d)), 1000000 (7.1(e)), and 1800000 (7.1(f)) for this landscape. The segregation of the population into two sub-populations which then evolve towards further specialisation in consuming a specific resource is clearly seen in Figures 7.1(d) $-$ 7.1(f). From previous chapters, the behaviour of diverging viral sub-populations and the dependency of the type of equilibria on parameters such as infection rate became known. We understand that the maximum value of the infection rate, when other parameters are fixed, dictates the type of equilibrium the system stabilises to. The numerical simulations shown in Figure 7.1 indicate that the phenotypes corresponding to local maxima of the fitness landscape, along with the neighbouring strains, become the most successful strains. This leads to an interesting question



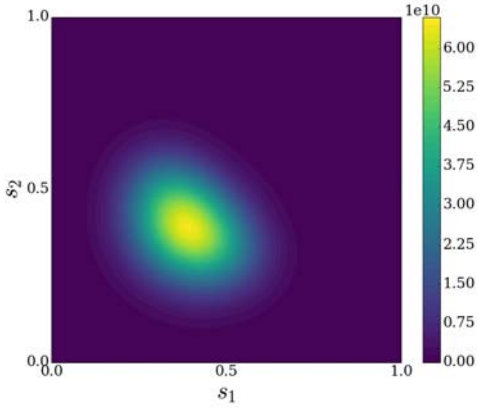
(a) $\alpha_1(s_1, s_2) + \alpha_2(s_1, s_2)$



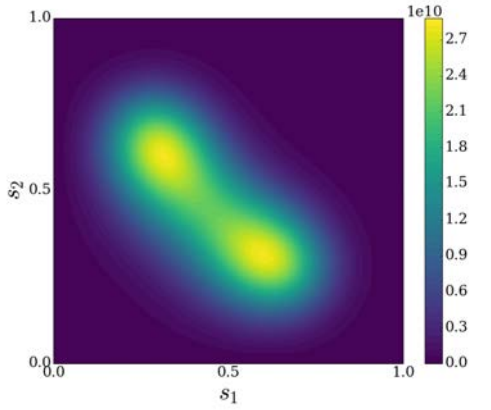
(b) t = 1



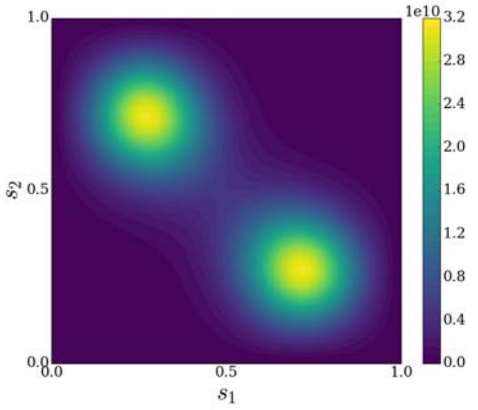
(c) t = 250000



(d) t = 500000



(e) t = 1000000



(f) t = 1800000

Figure 7.1: The infection rate used in simulations (Figure 7.1 (a)), and densities of distribution of viral variants $z(s_1, s_2, t)$ in the phenotype space after given numbers of time generations, t (Figure 7.1 (b)-7.1 (f)). Segregation of the viral population into two subpopulations according to target cells type can be seen in panels (b) to (f). Note, that the maximum values of $z(s_1, s_2, t)$ in each plot are different.

about the dependence of the specialisation time on the parameters in the system. Due to its importance, the question of specialisation time is the main interest of this chapter. However, the specialisation time depends on many parameters that are not available in most cases. In the following section, we consider a useful tool that allows a quantitative determination of critical parameter combinations that lead to a qualitative change in the functional dependence of the specialisation time.

7.4 Regression analysis to characterise specialisation processes

In scope of this chapter, specialisation is the process of the viral population occupying different niches (host types). The population of virus is introduced into the system as a few virions of the same strain in the middle of the phenotype space. The virions are provided with two types of susceptible cells to infect. The infection rates for each susceptible cell type are predetermined and fixed for all possible strains of the system. With sufficient infection rates, it is possible to set the conditions in such a way that the persistence of a non-zero virus population is a robust feature that is not destroyed when the parameter are varied. Depending on the values of parameters, specialisation occurs at finite yet different time ranges. In this section, we study the qualitative change in specialisation time brought by quantitative change in parameters.

The phenotype space is defined as a square with two axes s_1 and s_2. The fitness landscape, which has two local maxima, is defined on the phenotype space. The simplified division of the phenotype space into three parts (two for peaks and one for the valley between) allows the sub-populations of virus to be tracked according to their phenotype. The local maxima of the fitness landscape are enclosed in the upper-left and lower-right triangles, leaving the valley between the peaks to occupy a prolonged hexagon. Since each of the maxima correspond to the largest infection of the susceptible cells, undoubtedly, the upper-left and lower-right triangles host the phenotypes of the most specialised strains. Therefore, the phenotype space $S = [0, 1] \times [0, 1]$ is divided into three regions: the specialist-1 region being upper-left triangle, the specialist-2 region being lower-right triangle, and the stripe of the phenotype space separating them which corresponds to generalist strains. We calculate the time required for the specialist-1 and specialist-2 regions to each accommodate at least 15% of the total viral population and define this to be the specialisation time.

Regression analysis allows the relationship between a dependent variable and an independent variable to be approximately determined in terms of a prescribed function. As the dependent variable we take the studied quantity of specialisation time. As for the independent variable any parameter of the system, defined in a relevant range, is suitable.

As a result of the regression analysis, we gain a quantifiable measure of the dependence of the specialisation time on various parameters: this can be a single equation summarising the relationship between a parameter and the specialisation time or even a single number stating the slope of a linear dependence, which, in general, is called a regression value. Furthermore, the regression analysis can be employed iteratively. We can study a possible change in the regression value by varying another parameter. This leads to concentrated information on the effect of two-parameter variations on the specialisation time.

System (7.7) has fourteen parameters in total, two of which – the infection rates – are functions of the phenotype space. With an acceptable loss of generality, we fix all parameters except κ_1 and κ_2 and let burst sizes of both infected cells to be equal to an average value $\kappa_1 = \kappa$ and $\kappa_2 = \kappa$. In this first iteration, we calculate the specialisation time, T, for more than ten different average burst sizes of infected cells, κ, which are shown as the blue dots in Figure 7.2 (a). Note that axes are in log scale, that is, we plot $y = \log T$ as a function of $x = \log \kappa$. The green curve in the panel of Figure 7.2 (a) is a non-linear function that has been fitted to the obtained data. This function is:

$$y = y_0 + bx + cx^d + lx^{-d}. \tag{7.12}$$

The values of y_0, b, c, l, and d are found to minimise the difference between the scattered plot and the fitting function.

In Figure 7.2 (a), with an increasing value of κ, the specialisation takes more generations to occur. There is an approximately linear dependency between $\log \kappa$ and $\log T$ with a positive slope. However, this is only true for the given set of parameters. To explore how the specialisation time varies with the burst size at different points in parameter space, we treat the mutation rate as a second free parameter. More specifically, we sample ten different values of the mutation rate, μ, in the next iteration of the regression analysis. For each value of the mutation rate, we conduct the first iteration of the analysis with the regression variable κ. By calculating the specialisation time for varying κ, we find a fitting function for every μ, which are shown as green lines in Figure 7.2. For the small values of $\mu = 10^{-9}, 10^{-8}, 10^{-7}, 10^{-6}, 10^{-5}$, and even 10^{-4}, the specialisation time is proportional to the burst size: the dependence is almost linear having a positive slope. Interestingly, around $\mu = 10^{-3}$ this correlation changes. For values of the mutation rate exceeding $\mu = 10^{-3}$, there is a clear linear, inverse dependence between $\log \kappa$ and $\log T$. For faster mutation rates, increase in the burst size leads to a quicker specialisation.

To graph this change, the regression parameters y_0, b, c, l, and d of the fitting function with respect to μ are plotted in Figures 7.3 (a) - (e). Please note, that the horizontal

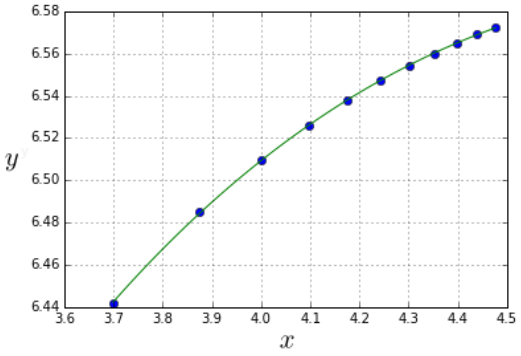
axes correspond to μ. Perfectly fitting an elaborate function to data is computationally expensive. Fortunately, sometimes a simple linear fitting can give as much information as a perfectly fit curve. Figure 7.3 (f) plots the slope of a straight-line fit through the data shown in Figure 7.2 versus the different mutation rates. We can uniquely identify the critical value of the mutation rate where the qualitative change in the T and κ relation occurs. When the curve in Figure 7.3 (f) crosses zero, the proportionality switches from direct to inverse. Figures 7.3 (a) - (e) illustrate the parameters of the fitting function (7.12), and are a very good example of the excess but unnecessary information that can be obtained from a regression analysis.

For very small μ, the dynamics are limited by the mutation rate, and increasing the burst size of the population does not contribute to faster evolution and, in fact, delays the onset of specialisation. Although this phenomenon may seem counter-intuitive at first glance, we suggest that it might be explained by diversity factors causing interspecific competition. For larger μ, other biological processes are limiting the dynamics, and we see expected inverse proportionality between burst size of the population and the specialisation time. The effect of the viral burst on the specialisation time varies depending on the mutation rate. However, the gradual increase of the mutation rate matches with overall decrease in the specialisation time.

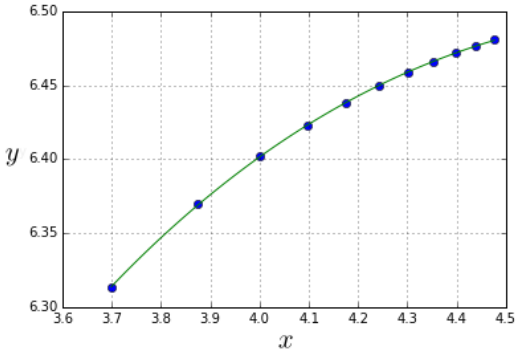
The motivation for this research was to illustrate the evolution of host-specialist viruses. Among the thousands of known viruses, we chose HIV-1 for our model because it is one of the most virulent parasites, which causes approximately million deaths annually according to the World Health Organisation and with almost 37 million people persistently infected by HIV-1. Throughout the history of mathematical biology a great number of models have been suggested to describe the evolution of viruses. Prevailing research is done, however, using phenomenological models. However, flexible frameworks are required in order to connect the large amount of data collected on the subject with mathematical tools. The ideal framework should allow for the possibility of interpreting both input parameters and results. The model developed in this chapter is mechanistic; it allows an easy change of components according to the goal of the research.

The main question leading to this research was about the transmission period of HIV-1 observed all through its history. It is expected that the transmission period of an infection should decrease after years of existence of the virus, if not defeated completely. It is believed that the natural selection of viral strains should increase the efficiency of virions. However in the case of HIV-1/AIDS, even after decades of existence of the virus, HIV-1 still has the same long post-incubation period. A hypothesis aimed to explain this phenomena is based on an assumption that the virion evolves within a host; that is, a virus increasing its efficiency

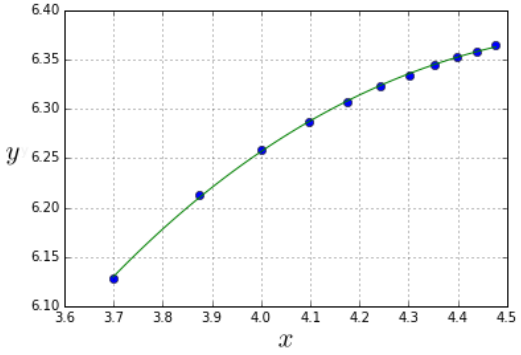
through host specialisation has a slower efficiency after transmission. We constructed a method to identify the qualitative change in specialisation time using a symmetric fitness landscape and parameters which are chosen for illustrative purposes only and are unrelated to a real-life situation. Although the parameters were chosen for illustrative purposes, the results of the model provide interesting insights into the dynamics of viral specialisation and can form the basis for more comprehensive future studies.



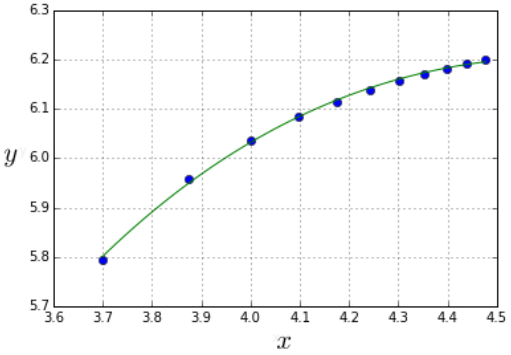
(a) $\mu = 10^{-9}$



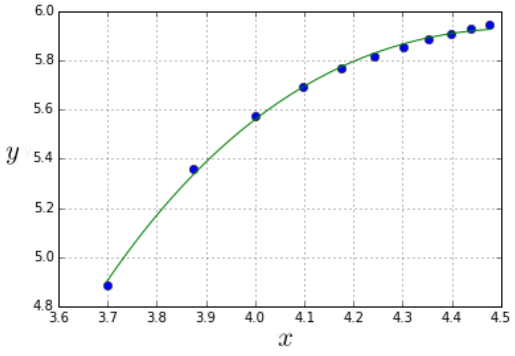
(b) $\mu = 10^{-8}$



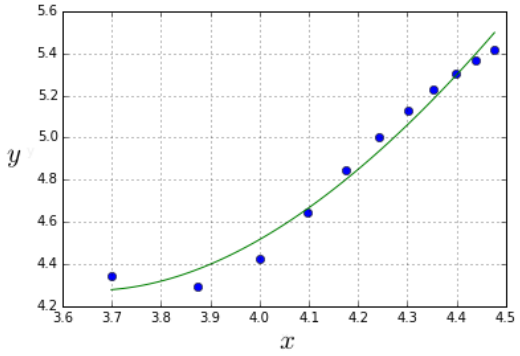
(c) $\mu = 10^{-7}$



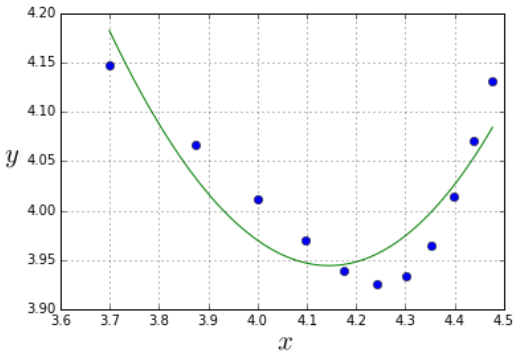
(d) $\mu = 10^{-6}$



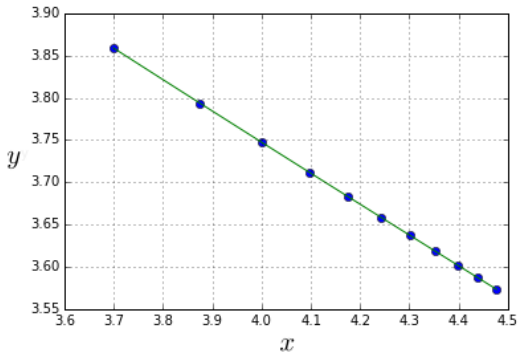
(e) $\mu = 10^{-5}$



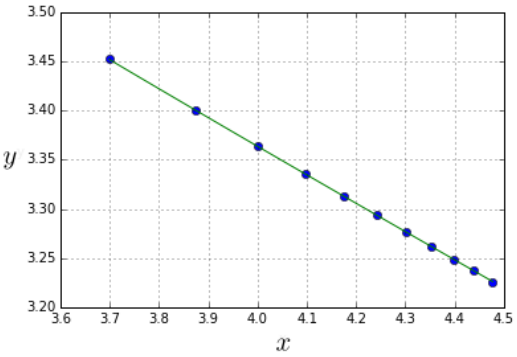
(f) $\mu = 10^{-4}$



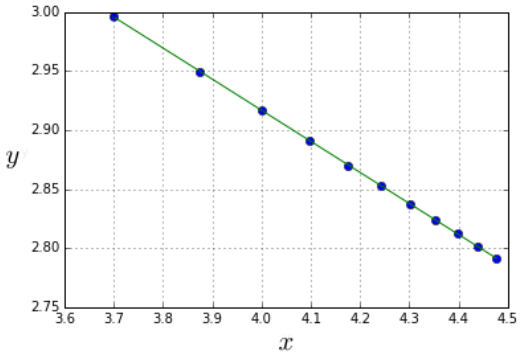
(g) $\mu = 10^{-3}$



(h) $\mu = 10^{-2}$

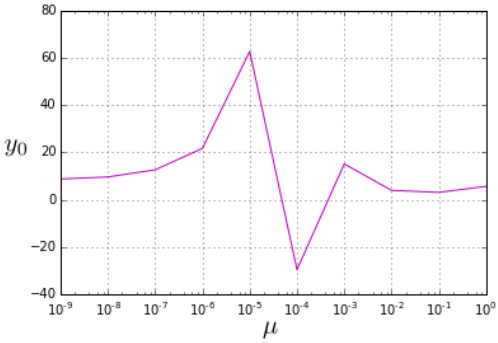


(i) $\mu = 10^{-1}$

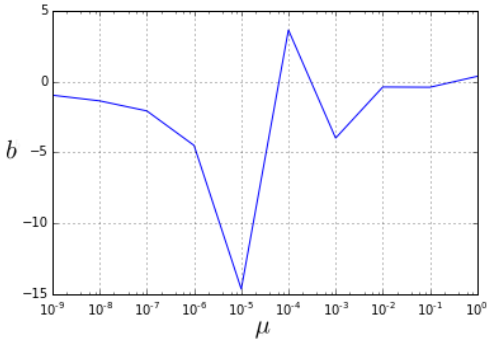


(j) $\mu = 10^{0}$

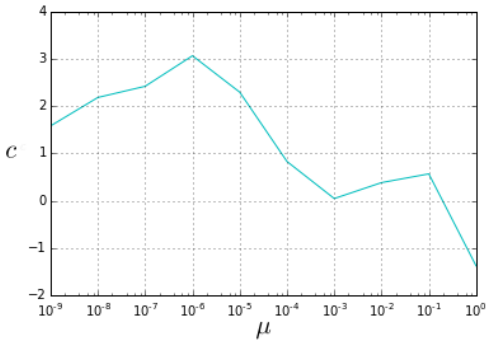
Figure 7.2: Ordinate axis of all plots measure $y = \log T_\mu(\kappa)$ and abscissa axis of all plots is $x = \log \kappa$. Each plot is calculated for a different value of the parameter μ. Blue dots are numerically obtained data, and green curves correspond to the fitting function $y = y_0 + bx + cx^d + lx^{-d}$.



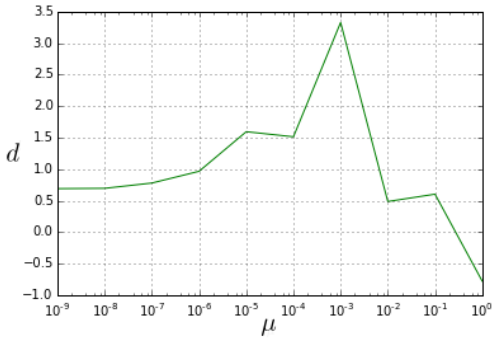
(a) y_0



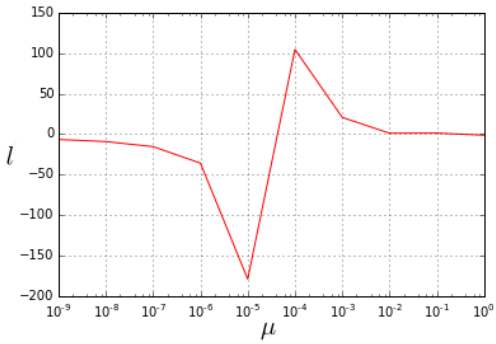
(b) b



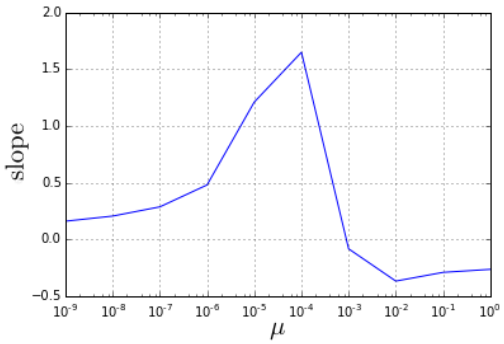
(c) c



(d) d



(e) l



(f) b for linear (only) fitting

Figure 7.3: (a) - (e) the parameters for the regression function (7.12); (f) the slope of the linear regression function.

Part IV

General conclusions

Chapter 8

Discussion and conclusions

8.1 Discussion

In this work, mathematical models of viral evolution were developed and studied by employing the powerful discipline of dynamical systems. It was shown that a wide range of questions in population dynamics can be answered by means of deterministic models based on ordinary differential equations. Sophisticated mathematical tools such as bifurcation theory were employed to analyse the behaviour of the proposed systems. To the best of our knowledge, the iterated regression technique, which was developed in the scope of the thesis for the analysis of numerical data, has not been previously implemented in the field. More importantly, the systematic approach developed in this work can now be used for other similar setups.

In Chapter 3, the sensitivity of a classic mathematical model of viral dynamics to a change in a specific term was studied. The virus greatly depends on its host cell for survival and evolution. However, the host cell itself is subjected to a certain environment and the population size of host cells depends on external factors too. General growth and death rates of the susceptible cell population were altered in the search for the most realistic and biologically meaningful model. It was shown that the combination of phenomenological and mechanistic approaches led to dynamics that most closely mimic those observed in real life [87]. This part of the study was crucial for determining the core model that was to be used to explore the hypotheses stated at the beginning of the thesis.

In general, the idea behind competing species rests on scarcity of common resources. The concept of competing viral strains is associated with a scarcity of susceptible cells and the level of infectivity that the strains display. Since the work focuses on the evolution of one virus, the competition that occurs between different variants of the population has to be

taken into account. In the case when a viral strain competes against its own mutants, the dynamics are more complicated in comparison to the case of two independent competitors. By means of a mathematical model, the conditions for coexistence of mutant and wild-type strains were found and described in Chapter 4 (see also [97]). It was shown that the survival and persistence of the wild-type strain might crucially depend on the infectivity of mutants. Due to the mutation-related decay in the size of wild-type strain population, the mutant population gets an advantage if the infection rates of mutant and wild-type strains are equal. Therefore, coexistence of wild-type and mutant variants is possible only when the infection rate of mutants is low. Moreover, the model predicts the impossibility of coexistence when the mutation rates exceed a critical value. This critical value depends proportionally on the virulence and burst size of the wild-type strain and it is inversely proportional to the multiplicity of infection of the wild-type variant. The fact that bigger burst sizes and smaller MOIs might allow a higher mutation rate in the system of coexisting mutant and wild-type strains is not surprising. The counter-intuitive part is that greater virulence allows coexistence to occur for greater mutation rates. This combination of parameters provides an upper bound for the mutant and wild-type coexistence that clearly depends on the characteristics of the wild-type strain itself. Therefore, it has been illustrated that coexistence, that is also a persistence of the wild-type, is possible only for the wild-type strains that possess certain strict characteristics. This aligns with the ever-changing nature of species, including viruses, and is embodied in a simple mathematical model.

If another type of susceptible cell is present in the system and only one of the strains has the ability to infect it, then the competition dynamics change. The model for this setup was constructed and studied in Chapter 5. Specialists and generalists were forcefully introduced by allowing only one of the two strains to infect both cell types (generalist), while the other strain, called the specialist, can infect only one cell type. We allow the wild-type strain to be the generalist and the mutant strain to be the specialist. Interestingly, even in the scope of this model, specialisation is persistent. There is one equilibrium where "the generalist" outcompetes the specialist, however, it only happens after the cell type susceptible to the specialist becomes extinct. Consequently, the specialist dies out. It can be explained partially by the excess infectivity of the specialist strain, that exhausts the only resource available to it. For lower infection rates of the specialist, the generalist can outcompete the specialist if the infection rate of the generalist is high enough. Then, "the generalist" persists alone, where quotes are referring to the fact that there is only one type of cell available for it to infect. Therefore, the specialist is present as long as its susceptible cell type exists in the system. On the other hand, the generalist can become extinct (be outcompeted by the specialist) even when there is an abundance of cells that can be infected

by the generalist but not the specialist. It is noteworthy that the remarkable performance of the specialist, in this case, does not require its high infectivity rate. Overall, an adaptation towards the host cell in a stable environment leads to the specialisation of viral strains.

The mutation rate not only limits the possibility of coexistence from above, as mentioned earlier, but it also affects the specialisation time. It has been shown in Chapter 7, that the combination of burst size and mutation rate has a nontrivial effect on specialisation speed. As expected, a higher mutation rate speeds up the adaptation process. Moreover, increase of the burst size creates a bigger viral population which should, in theory, lead to faster specialisation. However, when a low mutation rate is combined with a gradually increasing burst size, instead of causing the expected faster specialisation, the specialisation time gradually increases. In other words, specialisation slows down with higher burst sizes when the mutation rate is low. This can be explained by the competition that occurs between virions of the same variants. When the fixed mutation rate does not allow a quick change along the gradient of the fitness landscape, an accumulation of the virions of the same strain creates the delay. Therefore, as it was hypothesised, a low mutation rate of the virus population leads to interspecific competition and this competition slows down the overall specialisation time. Moreover, the method used to determine this phenomenon can provide a quantitative information about the critical values of the parameter combinations studied here.

Other results were obtained in relation to the existence of bistability and chaos in the model considering specialisation from Chapter 5. Although neither of these phenomena is rare in population dynamics, discovering windows of chaos and bistability in the stability diagrams provided deeper insights into specialist-generalist interactions. For example, it was determined that when the infection rate of a specialist is approximately three times larger than the infection rate of a generalist, in scope of the model, it is impossible to predict the behaviour of the system in the long term. When, *vice versa*, the infection rate of the generalist is three times larger than the infection rate of the specialist, then the system becomes bistable: system might stabilise at either the coexistence or specialist-free state. This bistability indicates a strong dependence on the initial sizes of the populations. With increasing infection rate of the generalist strain in the bistable region, a gradual increase in the probability of the generalist outcompeting the specialist was found. A decrease of the infection rate of the generalist in the bistable region leads to higher probability of coexistence. This decrease occurs until the infection rate of the generalist strain reaches a critical lower value, below which only the virus-free state can be stable.

8.2 Future work

In this part of the discussion, some models and their preliminary results are listed with the purpose of illustrating important possible further developments. Due to the complex processes evolved in viral specialisation and depending on future research aims, these models can be modified subsequently.

8.2.1 Lysogenic cycle

Throughout the thesis, the consideration of the lytic cycle has been a matter of choice. The case of lysogenic infection, nevertheless, can be easily incorporated into the system. Consider the familiar model of the three populations x, y, and z from Chapter 3:

$$\begin{aligned} \dot{x} &= \beta\, x(1 - x/K) - \alpha\, z\, x - \delta\, x, \\ \dot{y} &= \alpha\, z\, x - \gamma\, y, \\ \dot{z} &= \kappa\gamma\, y - \nu\, \alpha\, z\, x - \zeta\, z, \end{aligned} \tag{8.1}$$

where γ has been considered a virulence. The rate γ, in fact, states the fraction of all infected cells that are involved in virion production, which is the lytic cycle. The remainder of the infected cell population is either involved in mutation process, as it was shown in more complex models from Chapters 4 and 5, or conveniently ignored. In case of lysogenic infection, the infected cell can behave as uninfected cell for several generations. Under certain assumptions, the model of such phenomenon, therefore, can be written as follows:

$$\begin{aligned} \dot{x} &= \beta\, x(1 - x/K) + (1 - \gamma)\, y - \alpha\, z\, x - \delta\, x, \\ \dot{y} &= \alpha\, z\, x - \gamma\, y - (1 - \gamma)\, y, \\ \dot{z} &= \kappa\gamma\, y - \nu\, \alpha\, z\, x - \zeta\, z. \end{aligned} \tag{8.2}$$

Here, all infected cells either produce virions, with a rate γ, or join the uninfected cell population, with a rate $1-\gamma$. This change introduced into the model leads to a modification of the equilibria. Predictably, the model (8.2) has three equilibria, which mimic the equilibria of model (8.1): trivial equilibrium, virus-free equilibrium, and virus-abundant equilibrium. There are two slight differences between equilibria of two models. First, all burst size entries in the equilibria of (8.2) are scaled by factor of γ. Second, the virus population size in the virus-abundant equilibrium of (8.2) is also scaled by factor of γ. Despite the obvious

quantitative difference, the introduction of the lysogenic cycle in this simplified way does not affect the qualitative outcome of system (8.1). However, the model is still valid, and, depending on the goals of the studies, can be used for deeper analyses. Another approach can be taken to model the lysogenic cycle. By introducing separate populations of proliferating infected cells, w, it is possible to describe the interaction of populations with dormant infections more precisely:

$$\begin{aligned}
\dot{x} &= \beta\, x \left(1 - \frac{x+w}{K}\right) - \alpha\, x\, z - \delta\, x, \\
\dot{y} &= \alpha\, (x+w)\, z + \rho w - y, \\
\dot{w} &= \beta\, w \left(1 - \frac{x+w}{K}\right) + (1-\gamma)\, y - \rho w - \alpha\, w\, z - \delta\, w, \\
\dot{z} &= \kappa\, \gamma\, y - \nu\, \alpha\, z\, (x+w) - \zeta\, z.
\end{aligned} \tag{8.3}$$

In this case, uninfected cells share the environmental resources with proliferating infected cells. After infection, the cells either lyse as before with rate γ or increase the size of proliferating infected cell population with rate $1-\gamma$. Depending on the presence of certain chemicals or radiation, the dormant infection can switch to lytic cycle, which happens with probability ρ. This model has four different equilibria $\{x, y, w, z\}$, which can be labelled as the trivial equilibrium $\{0, 0, 0, 0\}$, virus-free equilibrium $\{x^*, 0, 0, 0\}$, uninfected cell-free equilibrium $\{0, y^*, w^*, z^*\}$, and the virus-abundant equilibrium $\{x^*, y^*, w^*, z^*\}$, where starred population sizes are not constant zeros and can be found analytically. This model and its equilibria are ideologically similar to the model from Chapter 4, which studied coexistence of two different strains. In this case, however, it is the uninfected cell population and passively infected cell population that show either independent persistence in the system or the coexistence. Further analysis is left out of scope of this thesis.

8.2.2 Stochastic models

Stochasticity is important in virus dynamics due to finite population sizes; demographic noise can play a crucial role in the survival of populations, especially when the population size is small. Moreover, due to general environmental noise and not fully understood robustness of the involved processes, some interactions in virus dynamics are subjected to perturbations of the surrounding conditions. The uncertainty in the interaction of populations can be studied by means of stochastic models. Up to this point, only deterministic models have been considered. The deterministic approach allows for better understanding

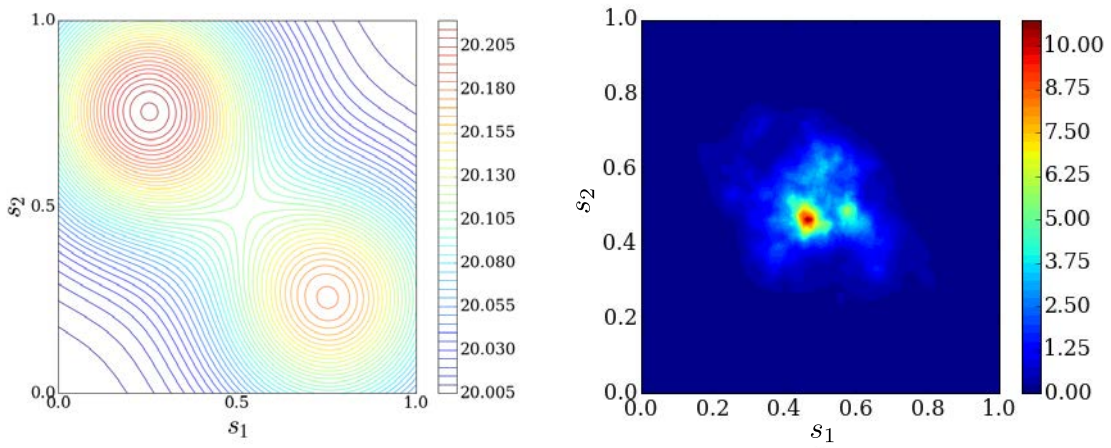


Figure 8.1: Left: Fixed part of the fitness landscape defined as a sum of two infection rate functions (Gaussians). Right: The development of clustered generalist strains when the fitness landscape is stochastically perturbed.

of the dynamics, however the obtained models and the observed results can be oversimplified. For instance, as a result of the model from Chapter 7, the viral population specialises towards the niches by clustering at evolutionary optima of the fitness landscape. In that case, the infection rate $\alpha(s)$ was defined as a fixed function on a phenotype space s. Owing to an average performance of separate virions at each phenotype, the infection rate, in fact, might considerably fluctuate every few generations. Therefore, to capture these perturbations, the infection rate could be modified accordingly as:

$$\alpha(s,t) = \alpha(s) + W_\tau(s,t), \tag{8.4}$$

where W_τ is a Wiener process of comparatively small amplitude approximated by a random walk with biologically meaningful time-step τ. Interestingly, the specialisation process might suffer from the random perturbations: instead of occupying the fitness maxima on the phenotype space, the virus population might reside in valleys of the fitness landscape, as shown in Figure 8.1. These variants are called generalists because, in the scope this thesis, the valleys of the fitness landscape correspond to modest infection rates of the variants that nevertheless infect several cell types simultaneously.

It was impossible to obtain the persistence of only the real generalists under the conditions of the deterministic model. It has been illustrated, that the appearance of specialists is inevitable, and that is due to the inner competition and the spread of the mutation along a higher gradient of the fitness landscape. Despite the competition and the mutation along the fitness gradient that is present in nature, the generalists do exist. This can be explained

by the hypothesis that generalism can be more profitable for a population under unstable environmental conditions. The preliminary results of the stochastic simulations coincide with this explanation. More detailed analysis on this topic is left for future work.

8.3 Summary

The appearance of species has been one of the most intriguing questions of the natural sciences for centuries, and is nowadays one of the most promising challenges of mathematical modelling. The rise of new species, including disease-causing pathogens, can be studied by means of population ecology. Predictability and comparative cost-effectiveness of mathematical models make the latter appealing for studies of population dynamics. In this work, I tried to broaden the spectrum of tools used for the analysis of biological processes and illustrated the use of such tools to answer diverse questions. The possibility of obtaining biological interpretations from the results played an important role, and the outcomes of the work produced both methodological and conceptual insights into mathematical modelling for viral evolution.

In this thesis, examples of simple ordinary differential equation (ODE) models for a viral population interacting with a susceptible host cell population were discussed first, where the effect of different growth and death rates of the susceptible cell population on dynamics of the system was considered. Then, the most realistic model was employed to explore the dynamics of two viral populations which shared a host cell population. Further, by introducing the second type of host cell population, the generalist-specialist dynamics were considered. The ODE system was then converted into a system of partial integro-differential equations (PIDEs) in order to model a viral population with a large number of variants. The ODE and PIDE systems were shown to have the same qualitative dynamics. The PIDE model was then used to describe a system composed to several variants and two types of susceptible cells. The variants in this case differed with respect to more than one phenotypic characteristic. The specialisation time was computed as a function of the mutation rate and burst size. Resultantly, exotic types of bifurcations were discovered and explained in terms of biological phenomena, and regions of chaos and bistability of a specialising viral population were discovered. All the hypotheses were shown to be relevant and confirmed to be true. This work, therefore, makes worthwhile contributions to mathematical modelling and analysis of evolutionary processes and substantially to the field of mathematical biology.

Bibliography

[1] L. J. Abu-Raddad and N. M. Ferguson. The impact of cross-immunity, mutation and stochastic extinction on pathogen diversity. *Proc. R. Soc. Lond. B Biol. Sci.*, 271 (1556):2431–2438, 2004.

[2] A. Acevedo, L. Brosdsky, and R. Andino. Mutational and fitness landscapes of an RNA virus revealed through population sequencing. *Nature*, 505:686–690, 2014.

[3] H. K. Alexander. *The roles of stochasticity and life history in the evolutionary population dynamics of pathogens*. PhD thesis, ETH Zurich, 2014.

[4] R. M. Anderson and R. M. May. The population dynamics of microparasites and their invertebrate hosts. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, 291(1054):451–524, 1981.

[5] R. Andino and E. Domingo. Viral quasispecies. *Virology*, 479-480:46–51, 2015.

[6] S. B. Araujo, M. P. Braga, D. R. Brooks, S. J. Agosta, E. P. Hoberg, F. W. von Hartenthal, and W. A. Boeger. Understanding host-switching by ecological fitting. *PLoS One*, 10(10):e0139225, 2015.

[7] M. Arribas, J. Aguirre, S. Manrubia, and E. Lázaro. Differences in adaptive dynamics determine the success of virus variants that propagate together. *Virus Evol.*, 4(1): vex043, 2018.

[8] F. Bagnoli and M. Bezzi. Speciation as pattern formation by competition in a smooth fitness landscape. *Phys. Rev. Lett.*, 79(17):3302, 1997.

[9] U. Bastolla, M. Lässig, S. C. Manrubia, and A. Valleriani. Biodiversity in model ecosystems, I: coexistence conditions for competing species. *J. Theor. Biol.*, 235(4): 521–530, 2005.

[10] A. D. Bazykin, F. S. Berezovskaya, and S. L. Zudin. Bifurcation approach to the predator-prey population models (Version of the computer book). Technical report, International Centre for Theoretical Physics, 1993.

[11] B. Beerenwinkel and O. Zagordi. Ultra-deep sequencing for the analysis of viral populations. *Curr. Opin. Virol.*, 1:413–418, 2011.

[12] K. D. Behrman and M. Kirkpatrick. Species range expansion by beneficial mutations. *J. Evol. Biol.*, 24(3):665–675, 2011.

[13] E. Beretta and Y. Kuang. Modeling and analysis of a marine bacteriophage infection. *Math. Biosci.*, 149(1):57–76, 1998.

[14] A. Bermingham, M. Chand, C. Brown, E. Aarons, C. Tong, C. Langrish, K. Hoschler, K. Brown, M. Galiano, R. Myers, et al. Severe respiratory illness caused by a novel coronavirus, in a patient transferred to the United Kingdom from the Middle East, September 2012. *Eurosurveillance*, 17(40):20290, 2012.

[15] A. Best, A. White, and M. Boots. The implications of coevolutionary dynamics to host-parasite interactions. *Am. Nat.*, 173(6):779–791, 2009.

[16] K. B. Blyuss. Analysis of symmetries in models of multi-strain infections. *J. Math. Biol.*, 69(6-7):1431–1459, 2014.

[17] B. Boldin, S. A. Geritz, and É. Kisdi. Superinfections and adaptive dynamics of pathogen virulence revisited: a critical function analysis. *Evol. Ecol. Res.*, 11(2): 153–175, 2009.

[18] S. Bonhoeffer and M. A. Nowak. Mutation and the evolution of virulence. *Proc. R. Soc. Lond. B Biol. Sci.*, 258(1352):133–140, 1994.

[19] J. Bull, R. Sanjuán, and C. Wilke. Theory of lethal mutagenesis for viruses. *J. Virol.*, 81(6):2930–2939, 2007.

[20] J. J. Bull, L. A. Meyers, and M. Lachmann. Quasispecies made simple. *PLoS Comput. Biol.*, 1(6):e61, 2005.

[21] À. Calsina and S. Cuadrado. Small mutation rate and evolutionarily stable strategies in infinite dimensional adaptive dynamics. *J. Math. Biol.*, 48(2):135–159, 2004.

[22] À. Calsina and J. Saldaña. A model of physiologically structured population dynamics with a nonlinear individual growth rate. *J. Math. Biol.*, 33(4):335–364, 1995.

[23] C. Castillo-Chavez and J. X. Velasco-Hernandez. On the relationship between evolution of virulence and host demography. *J. Theor. Biol.*, 192(4):437–444, 1998.

[24] C. Y. Chiu. Viral pathogen discovery. *Curr. Opin. Microbiol.*, 16(4):468–478, 2013.

[25] D. S. Coffey. Self-organization, complexity and chaos: the new biology for medicine. *Nat. Med.*, 4(8):882–885, 1998.

[26] J. M. Cuevas, A. Willemsen, J. Hillung, M. P. Zwart, and S. F. Elena. Temporal dynamics of intrahost molecular evolution for a plant RNA virus. *Mol. Biol. Evol.*, 32(5):1132–1147, 2015.

[27] A. Dhooge, W. Govaerts, Y. A. Kuznetsov, H. G. E. Meijer, and B. Sautois. New features of the software matcont for bifurcation analysis of dynamical systems. *Math. Comput. Model. Dyn. Syst.*, 14(2):147–175, 2008.

[28] M. Doebeli, Y. Ispolatov, and B. Simon. Point of view: Towards a mechanistic foundation of evolutionary theory. *Elife*, 6:e23804, 2017.

[29] E. Doedel, H. B. Keller, and J. P. Kernevez. Numerical analysis and control of bifurcation problems (i): bifurcation in finite dimensions. *Int. J. Bifurcat. Chaos*, 1 (03):493–520, 1991.

[30] H. M. Doekes, C. Fraser, and K. A. Lythgoe. Effect of the latent reservoir on the evolution of HIV at the within-and between-host levels. *PLoS Comput. Biol.*, 13(1): e1005228, 2017.

[31] E. Domingo. Quasispecies theory in virology. *J. Virol.*, 76(1):463–465, 2002.

[32] N. K. Duggal, A. Bosco-Lauth, R. A. Bowen, S. S. Wheeler, W. K. Reisen, T. A. Felix, B. R. Mann, H. Romo, D. M. Swetnam, A. D. Barrett, et al. Evidence for co-evolution of West Nile virus and house sparrows in North America. *PLoS Negl. Trop. Dis.*, 8(10):e3262, 2014.

[33] C. Dye, B. G. Williams, M. A. Espinal, and M. C. Raviglione. Erasing the world's slow stain: strategies to beat multidrug-resistant tuberculosis. *Science*, 295(5562): 2042–2046, 2002.

[34] D. J. Earn, P. Rohani, and B. T. Grenfell. Persistence, chaos and synchrony in ecology and epidemiology. *Proc. R. Soc. Lond. B Biol. Sci.*, 265(1390):7–10, 1998.

[35] M. Eigen. Selforganization of matter and evolution of biological macromolecules. *Naturwissenschaften*, 58:465–523, 1971.

[36] S. F. Elena, P. Agudelo-Romero, and J. Lalić. The evolution of viruses in multi-host fitness landscapes. *Open Virol. J*, 3:1, 2009.

[37] S. F. Elena, A. Fraile, and F. García-Arenal. Evolution and emergence of plant viruses. *Adv. Virus Res.*, 88:161–191, 2014.

[38] H. Eskola et al. *Mechanistic population models in biology: model derivation and application in evolutionary studies*. PhD thesis, University of Turku, 2009.

[39] P. W. Ewald. *Evolution of Infectious Disease*. Oxford University Press on Demand, 1994.

[40] S. M. Faruque, M. M. Rahman, K. N. Islam, J. J. Mekalanos, et al. Lysogenic conversion of environmental Vibrio mimicus strains by CTXϕ. *Infect. Immun.*, 67 (11):5723–5729, 1999.

[41] A. Fenton, J. Lello, and M. Bonsall. Pathogen responses to host immunity: the impact of time delays and memory on the evolution of virulence. *Proc. R. Soc. Lond. B Biol. Sci.*, 273(1597):2083–2090, 2006.

[42] V. V. Ganusov, D. L. Barber, and R. J. De Boer. Killing of targets by CD8+ T cells in the mouse spleen follows the law of mass action. *PLoS One*, 6(1):e15959, 2011.

[43] R. Gao, B. Cao, Y. Hu, Z. Feng, D. Wang, W. Hu, J. Chen, Z. Jie, H. Qiu, K. Xu, et al. Human infection with a novel avian-origin influenza A (H7N9) virus. *N. Engl. J. Med.*, 368(20):1888–1897, 2013.

[44] J. García-Ojalvo and J. Sancho. *Noise in Spatially Extended Systems*. Springer Science & Business Media, 2012.

[45] H. R. Gelderblom. Structure and classification of viruses. 1996.

[46] E. Gjini, C. Valente, R. Sa-Leao, and M. G. M. Gomes. How direct competition shapes coexistence and vaccine effects in multi-strain pathogen systems. *J. Theor. Biol.*, 388:50–60, 2016.

[47] A. Grande-Pérez, E. Lázaro, P. Lowenstein, E. Domingo, and S. C. Manrubia. Suppression of viral infectivity through lethal defection. *PNAS*, 102(12):4448–4452, 2005.

[48] D. M. Grobman. Topological classification of neighborhoods of a singularity in n-space. *Matem. Sbor.*, 98(1):77–94, 1962.

[49] J. Guckenheimer and P. Holmes. *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*, volume 42. Springer Science & Business Media, 2013.

[50] P. Hartman. On local homeomorphisms of Euclidean spaces. *Bol. Soc. Mat. Mex.*, 5 (2):220–241, 1960.

[51] D. M. Hawley, E. E. Osnas, A. P. Dobson, W. M. Hochachka, D. H. Ley, and A. A. Dhondt. Parallel patterns of increased virulence in a recently emerged wildlife pathogen. *PLoS Biol.*, 11(5):e1001570, 2013.

[52] J. D. Hedengren, R. A. Shishavan, K. M. Powell, and T. F. Edgar. Nonlinear modeling, estimation and predictive control in APMonitor. *Comput. Chem. Eng.*, 70:133–148, 2014.

[53] H. W. Hethcote. The mathematics of infectious diseases. *SIAM Review*, 42(4):599–653, 2000.

[54] M. E. Hochberg and R. D. Holt. The coexistence of competing parasites. I. The role of cross-species infection. *Am. Nat.*, 136(4):517–541, 1990.

[55] K. H. Hoffmann, B. Rodriguez-Brito, M. Breitbart, D. Bangor, F. Angly, B. Felts, J. Nulton, F. Rohwer, and P. Salamon. Power law rank–abundance models for marine phage communities. *FEMS Microbiol. Lett.*, 273(2):224–228, 2007.

[56] K. Holmfeldt, M. Middelboe, O. Nybroe, and L. Riemann. Large variabilities in host strain susceptibility and phage host range govern interactions between lytic marine phages and their flavobacterium hosts. *Appl. Environ. Microbiol.*, 73(21):6730–6739, 2007.

[57] G. Huang, Y. Takeuchi, and A. Korobeinikov. HIV evolution and progression of the infection to AIDS. *J. Theor. Biol.*, 307:149–159, 2012.

[58] O. Isakov, A. Bordería, D. Golan, A. Hamenahem, G. Celniker, L. Yoff, H. Blanc, M. Vignuzzi, and N. Shomron. Deep sequencing analysis of viral infection and evolution allows rapid and detailed characterization of viral mutant spectrum. *Bioinformatics*, 31:2141–2150, 2015.

[59] D. Iwanowski. Ueber die Mosaikkrankheit der Tabakspflanze. *Zeitschrift für Pflanzenkrankheiten*, 13(1):1–41, 1903.

[60] M. D. Johnston, C. M. Edwards, W. F. Bodmer, P. K. Maini, and S. J. Chapman. Mathematical modeling of cell population dynamics in the colonic crypt and in colorectal cancer. *PNAS*, 104(10):4008–4013, 2007.

[61] T. J. Kawecki. Accumulation of deleterious mutations and the evolutionary cost of being a generalist. *Am. Nat.*, 144(5):833–838, 1994.

[62] J. P. Keener. Oscillatory coexistence in the chemostat: a codimension two unfolding. *SIAM J. Appl. Math.*, 43(5):1005–1018, 1983.

[63] E. Kindler, H. R. Jónsdóttir, D. Muth, O. J. Hamming, R. Hartmann, R. Rodriguez, R. Geffers, R. A. Fouchier, C. Drosten, M. A. Müller, et al. Efficient replication of the novel human betacoronavirus EMC on primary human epithelium highlights its zoonotic potential. *MBio*, 4(1):e00611–12, 2013.

[64] S. E. Kingsland. *Modeling Nature*. University of Chicago Press, 1995.

[65] K. F. Kiple, R. R. Graham, D. Frey, and A. Browne. *The Cambridge World History of Human Disease*, volume 1. Cambridge University Press, 1993.

[66] A. Korobeinikov and C. Dempsey. A continuous strain-space model of viral evolution within a host. *Math. Biosci. Eng.*, 11(4):919–927, 2014.

[67] S. R. Krishnamurthy and D. Wang. Origins and challenges of viral dark matter. *Virus Res.*, 239:136–142, 2017.

[68] I. Kupka. Contributiona la théorie des champs génériques. *Cont. Diff. Eq.*, 2:457–484, 1963.

[69] D. Kutnjak, S. F. Elena, and M. Ravnikar. Time-sampled population sequencing reveals the interplay of selection and genetic drift in experimental evolution of Potato virus Y. *J. Virol.*, 91:e00690–1, 2017.

[70] Y. A. Kuznetsov. Practical computation of normal forms on center manifolds at degenerate Bogdanov–Takens bifurcations. *Int. J. Bifurcat. Chaos*, 15(11):3535–3546, 2005.

[71] Y. A. Kuznetsov. *Elements of Applied Bifurcation Theory*, volume 112. Springer, 2013.

[72] O. Lada, Y. Benhamou, T. Poynard, and V. Thibault. Coexistence of hepatitis B surface antigen (HBs Ag) and anti-HBs antibodies in chronic hepatitis B virus carriers: influence of "a" determinant variants. *J. Virol.*, 80(6):2968–2975, 2006.

[73] Y. Lan and P. Cvitanović. Variational method for finding periodic orbits in a general flow. *Phys. Rev. E*, 69(1):016217, 2004.

[74] V. LeBlanc. A degenerate Hopf bifurcation in retarded functional differential equations, and applications to endemic bubbles. *J. Nonlinear Sci.*, 26:1–25, 2016.

[75] L. Lehmann, C. Mullon, E. Akcay, and J. Cleve. Invasion fitness, inclusive fitness, and reproductive numbers in heterogeneous populations. *Evolution*, 70(8):1689–1702, 2016.

[76] N. Leiby. *Adaptation and specialization in the evolution of bacterial metabolism.* PhD thesis, Harvard University, 2014.

[77] P. Lemey, O. G. Pybus, B. Wang, N. K. Saksena, M. Salemi, and A.-M. Vandamme. Tracing the origin and history of the HIV-2 epidemic. *PNAS*, 100(11):6588–6592, 2003.

[78] A. Leonard, M. McClain, G. Smith, D. Wentworth, R. Halpin, X. Lin, A. Ransier, T. Stockwell, S. Das, A. Gilbert, R. Lambkin-Williams, G. Ginsburg, C. Woods, and K. Koella. Deep sequencing of influenza A virus from a human challenge study reveals a selective bottleneck and only limited intrahost genetic diversification. *J. Virol.*, 90: 24, 11247-11258.

[79] S. Levin and D. Pimentel. Selection of intermediate rates of increase in parasite-host systems. *Am. Nat.*, 117(3):308–315, 1981.

[80] D. N. Levy, G. M. Aldrovandi, O. Kutsch, and G. M. Shaw. Dynamics of HIV-1 recombination in its natural target cells. *Proc. Natl. Acad. Sci. U. S. A.*, 101(12): 4204–4209, 2004.

[81] W. Liu, H. W. Hethcote, and S. A. Levin. Dynamical behavior of epidemiological models with nonlinear incidence rates. *J. Math. Biol.*, 25(4):359–380, 1987.

[82] A. J. Lotka. Analytical note on certain rhythmic relations in organic systems. *PNAS*, 6(7):410–415, 1920.

[83] F. Lutscher, A. Stevens, et al. Emerging patterns in a hyperbolic model for locally interacting cell systems. *J. Nonlinear Sci.*, 12(6):619–640, 2002.

[84] G. Macdonald. The analysis of equilibrium in malaria. *Trop. Dis. Bull.*, 49(9):813, 1952.

[85] L. V. Madden, G. Hughes, and F. Bosch. *The Study of Plant Disease Epidemics.* American Phytopathological Society (APS Press), 2007.

[86] P. M. Martin and E. Martin-Granel. 2,500-year evolution of the term epidemic. *Emerg. Infect. Dis.*, 12(6):976, 2006.

[87] R. Martinez-Corral, J. Liu, G. M. Süel, and J. Garcia-Ojalvo. Bistable emergence of oscillations in growing Bacillus subtilis biofilms. *PNAS*, 115(36):E8333–E8340, 2018.

[88] E. Mazé-Guilmo, S. Blanchet, K. D. McCoy, and G. Loot. Host dispersal as the driver of parasite genetic structure: a paradigm lost? *Ecol. Lett.*, 19(3):336–347, 2016.

[89] M. Mehrotra, G. Wang, and W. M. Johnson. Multiplex PCR for detection of genes for Staphylococcus aureus enterotoxins, exfoliative toxins, toxic shock syndrome toxin 1, and methicillin resistance. *J. Clin. Microbiol.*, 38(3):1032–1035, 2000.

[90] L. R. Miranda, B. C. Schaefer, A. Kupfer, Z. Hu, and A. Franzusoff. Cell surface expression of the HIV-1 envelope glycoproteins is directed from intracellular CTLA-4-containing regulated secretory granules. *PNAS*, 99(12):8031–8036, 2002.

[91] S. Muratori and S. Rinaldi. Remarks on competitive coexistence. *SIAM J. Appl. Math.*, 49(5):1462–1472, 1989.

[92] J. D. Murray. *Mathematical biology: I. An Introduction (Interdisciplinary Applied Mathematics)(Pt. 1)*. New York, Springer, 2007.

[93] R. Nelson. *Plant Disease: An Advanced Treatise*, volume 4, chapter 2, pages 23–46. Academic Press, 2012.

[94] A. E. Noble, N. M. Temme, W. F. Fagan, and T. H. Keitt. A sampling theory for asymmetric communities. *J. Theor. Biol.*, 273(1):1–14, 2011.

[95] M. Nowak and R. M. May. *Virus Dynamics: Mathematical Principles of Immunology and Virology*. Oxford University Press, UK, 2000.

[96] M. A. Nowak and R. M. May. Superinfection and the evolution of parasite virulence. *Proc. R. Soc. Lond. B Biol. Sci.*, 255(1342):81–89, 1994.

[97] A. Nurtay, M. G. Hennessy, J. Sardanyés, L. Alsedà, and S. F. Elena. Theoretical conditions for the coexistence of viral strains with differences in phenotypic traits: a bifurcation analysis. *R. Soc. Open Sci.*, 6(1):181179, 2019.

[98] F. Pacini. *Osservazioni Microscopiche e Deduzioni Patologiche sul Cholera Asiatico*. tip. di F. Bencini, 1854.

[99] T. S. Parker and L. Chua. *Practical Numerical Algorithms for Chaotic Systems*. Springer Science & Business Media, 2012.

[100] R. M. Penczykowski, A.-L. Laine, and B. Koskella. Understanding the ecology and evolution of host–parasite interactions across scales. *Evol. Appl.*, 9(1):37–52, 2016.

[101] A. S. Perelson. Modelling viral and immune system dynamics. *Nat. Rev. Immunol.*, 2(1):28, 2002.

[102] M. C. Pitman, J. S. Lau, J. H. McMahon, and S. R. Lewin. Barriers and strategies to achieve a cure for HIV. *The Lancet HIV*, 5(6):e317–e328, 2018.

[103] Y. Pommier, A. A. Johnson, and C. Marchand. Integrase inhibitors to treat HIV/AIDS. *Nat. Rev. Drug Discov.*, 4(3):236, 2005.

[104] A. Pugliese. On the evolutionary coexistence of parasite strains. *Math. Biosci.*, 177: 355–375, 2002.

[105] O. G. Pybus and A. Rambaut. Evolutionary analysis of the dynamics of viral infectious disease. *Nat. Rev. Genet.*, 10(8):540, 2009.

[106] O. G. Pybus, M. A. Charleston, S. Gupta, A. Rambaut, E. C. Holmes, and P. H. Harvey. The epidemic behavior of the hepatitis C virus. *Science*, 292(5525):2323–2325, 2001.

[107] A. F. Read. The evolution of virulence. *Trends Microbiol.*, 2(3):73–76, 1994.

[108] A. Rey. Dictionnaire historique de la langue française, Dictionnaires Le Robert, Paris 1992. *Zitierweise: Le Robert, Dictionnaire historique de la langue française.*

[109] M. Roberts, V. Andreasen, A. Lloyd, and L. Pellis. Nine challenges for deterministic epidemic models. *Epidemics*, 10:49–53, 2015.

[110] P. A. Rota, M. S. Oberste, S. S. Monroe, W. A. Nix, R. Campagnoli, J. P. Icenogle, S. Penaranda, B. Bankamp, K. Maher, M.-h. Chen, et al. Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science*, 2003.

[111] C. Rueffler, M. Egas, and J. A. Metz. Evolutionary predictions should be based on individual-level traits. *Am. Nat.*, 168(5):E148–E162, 2006.

[112] L. Salvaudon, V. Héraudet, and J. A. Shykoff. Parasite-host fitness trade-offs change with parasite identity: genotype-specific interactions in a plant-pathogen system. *Evolution*, 59(12):2518–2524, 2005.

[113] J. Sardanyés. *Dynamics, evolution and information in nonlinear dynamical systems of replicators*. PhD thesis, Universitat Pompeu Fabra, 2009.

[114] J. Sardanyés and S. F. Elena. Quasispecies spatial models for RNA viruses with different replication modes and infection strategies. *PLoS One*, 6(9):e24884, 2011.

[115] W. E. Schiesser. *The Numerical Method of Lines: Integration of Partial Differential Equations*. Elsevier, 2012.

[116] M. Schwartz and C. Hornych. Specialisation versus diversification: perceived benefits of different business incubation models. *IJEIM*, 15(3):177–197, 2012.

[117] E. W. Seabloom, E. T. Borer, K. Gross, A. E. Kendig, C. Lacroix, C. E. Mitchell, E. A. Mordecai, and A. G. Power. The community ecology of pathogens: coinfection, coexistence and community composition. *Ecol. Lett.*, 18(4):401–415, 2015.

[118] S. Shapin. Sick city: maps and mortality in the time of cholera. *New Yorker (NY: 1925)*, page 110, 2006.

[119] V. Shinde, C. B. Bridges, T. M. Uyeki, B. Shu, A. Balish, X. Xu, S. Lindstrom, L. V. Gubareva, V. Deyde, R. J. Garten, et al. Triple-reassortant swine influenza A (H1) in humans in the United States, 2005–2009. *N. Engl. J. Med.*, 360(25):2616–2625, 2009.

[120] S. Smale. Stable manifolds for differential equations and diffeomorphisms. *Ann. Sc. Norm. Super. Pisa Cl. Sci.*, 17(1-2):97–116, 1963.

[121] B. Spellberg, R. Guidos, D. Gilbert, J. Bradley, H. W. Boucher, W. M. Scheld, J. G. Bartlett, J. Edwards Jr, and I. D. S. of America. The epidemic of antibiotic-resistant infections: a call to action for the medical community from the Infectious Diseases Society of America. *Clin. Infect. Dis.*, 46(2):155–164, 2008.

[122] S. H. Strogatz. *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, and Engineering*. CRC Press, 2018.

[123] R. Sun, S.-F. Lin, L. Gradoville, Y. Yuan, F. Zhu, and G. Miller. A viral gene that activates lytic cycle expression of Kaposi's sarcoma-associated herpesvirus. *PNAS*, 95 (18):10866–10871, 1998.

[124] C. A. Suttle. Marine viruses—major players in the global ecosystem. *Nat. Rev. Microbiol.*, 5(10):801, 2007.

[125] N. S.-O. I. A. H. V. I. Team. Emergence of a novel swine-origin influenza a (h1n1) virus in humans. *N. Engl. J. Med.*, 360(25):2605–2615, 2009.

[126] H. R. Thieme. *Mathematics in Population Biology*. Princeton University Press, 2018.

[127] G. Tigan. Analysis of degenerate fold–Hopf bifurcation in a three-dimensional differential system. *Qual. Theor. Dyn. Syst.*, pages 1–16, 2017.

[128] G. Tigan. Degenerate with respect to parameters fold-Hopf bifurcations. *Discrete Contin. Dyn. Syst. Ser. A*, 37(4):2115–2140, 2017.

[129] G. Tigan, J. Llibre, and L. Ciurdariu. Degenerate Fold–Hopf bifurcations in a Rössler-Type system. *Int. J. Bifurcat. Chaos*, 27(05):1750068, 2017.

[130] H. Tu, S.-D. Xiong, C. Trepo, and Y.-M. Wen. Frequency of hepatitis B virus e-minus mutants varies among patients from different areas of China. *J. Math. Biol.*, 51(2): 85–89, 1997.

[131] S. van Boheemen, M. de Graaf, C. Lauber, T. M. Bestebroer, V. S. Raj, A. M. Zaki, A. D. Osterhaus, B. L. Haagmans, A. E. Gorbalenya, E. J. Snijder, et al. Genomic characterization of a newly discovered coronavirus associated with acute respiratory distress syndrome in humans. *MBio*, 3(6):e00473–12, 2012.

[132] P.-F. Verhulst. Recherches mathématiques sur la loi d'accroissement de la population. *Nouveaux mémoires de l'académie royale des sciences et belles-lettres de Bruxelles*, 18: 14–54, 1845.

[133] V. Volterra. *Variazioni e Fluttuazioni del Numero D'individui in Specie Animali Conviventi*. C. Ferrari, 1927.

[134] B. D. Walker and B. T. Korber. Immune control of HIV: the obstacles of HLA and viral diversity. *Nat. Immunol.*, 2(6):473, 2001.

[135] X. Wang and W. Wang. An HIV infection model based on a vectored immunoprophylaxis experiment. *J. Theor. Biol.*, 313:127–135, 2012.

[136] J. S. Weitz, G. Li, H. Gulbudak, M. H. Cortez, and R. J. Whitaker. Viral fitness across a continuum from lysis to latency. *bioRxiv*, page 296897, 2018.

[137] S. M. White, J. P. Burden, P. K. Maini, and R. S. Hails. Modelling the within-host growth of viral infections in insects. *J. Theor. Biol.*, 312:34–43, 2012.

[138] M. C. Whitlock. The red queen beats the jack-of-all-trades: the limitations on the evolution of phenotypic plasticity and niche breadth. *Am. Nat.*, 148:S65–S77, 1996.

[139] G. Wild, A. Gardner, and S. A. West. Adaptation and the evolution of parasite virulence in a connected world. *Nature*, 459(7249):983, 2009.

[140] D. Wodarz and D. N. Levy. Multiple HIV-1 infection of cells and the evolutionary dynamics of cytotoxic T lymphocyte escape mutants. *Evolution*, 63(9):2326–2339, 2009.

[141] A. Wolf, J. B. Swift, H. L. Swinney, and J. A. Vastano. Determining Lyapunov exponents from a time series. *Physica D*, 16(3):285–317, 1985.

[142] M. Woolhouse, D. Haydon, and R. Antia. Emerging pathogens: the epidemiology and evolution of species jump. *Trends Ecol. Evol.*, 20:238–244, 2005.

[143] S. Wright. *The Roles of Mutation, Inbreeding, Crossbreeding, and Selection in Evolution*, volume 1. The American Naturalist, 1932.

[144] T.-T. Wu, E. J. Usherwood, J. P. Stewart, A. A. Nash, and R. Sun. Rta of murine gammaherpesvirus 68 reactivates the complete lytic cycle from latency. *J. Virol.*, 74 (8):3659–3667, 2000.

[145] C. Wymant, M. Hall, O. Ratmann, D. Bonsall, T. Golubchik, M. de Cesare, A. Gall, M. Cornelissen, C. Fraser, STOPHCV Consortium, The Maela Pneumococcal Collaboration, and The BEEHIVE Collaboration. Phyloscanner: inferring transmission from within-and between-host pathogen genetic diversity. *Mol. Biol. Evol.*, 35(3): 719–733, 2017.

[146] A. M. Zaki, S. van Boheemen, T. M. Bestebroer, A. D. Osterhaus, and R. A. Fouchier. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N. Engl. J. Med.*, 367(19):1814–1820, 2012.

[147] R. Zorenos dos Santos and S. Coutinho. On the dynamics of the evolution of the HIV infection. Technical report, 2000.